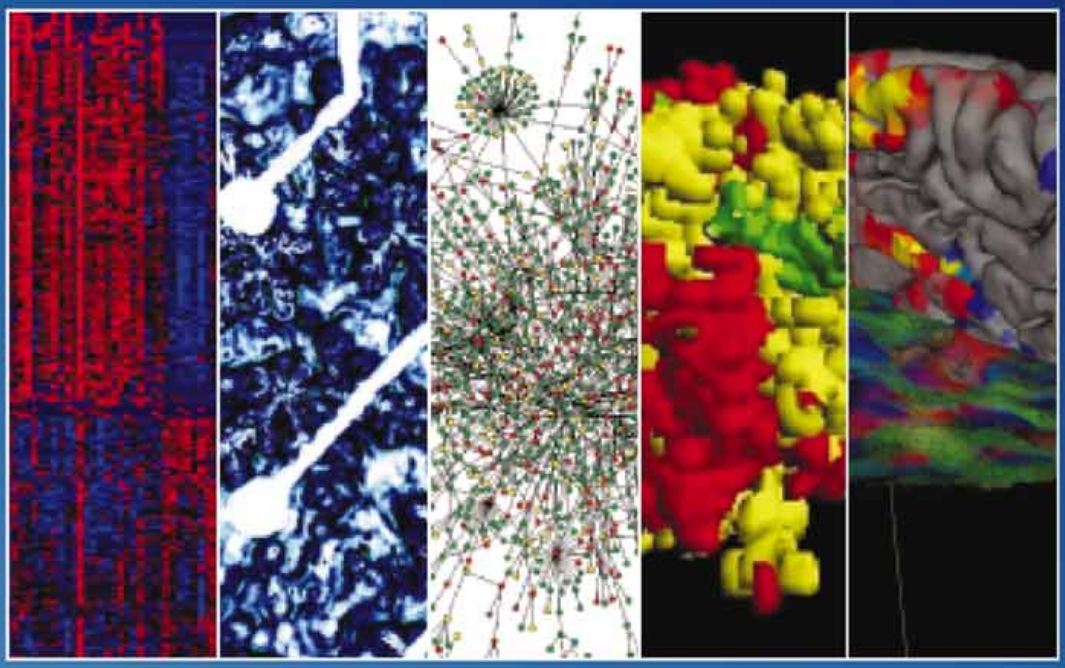
Topics in
Biomedical Engineering
International Book Series
Series Editor: Evangelia Micheli-Tzanakou

Complex Systems Science in Biomedicine



Edited by
Thomas S. Deisboeck
J. Yasha Kresh

Springer

Complex Systems Science in Biomedicine

TOPICS IN BIOMEDICAL ENGINEERING INTERNATIONAL BOOK SERIES

Series Editor: Evangelia Micheli-Tzanakou
Rutgers University
Piscataway, New Jersey

Signals and Systems in Biomedical Engineering:
Signal Processing and Physiological Systems Modeling
Suresh R. Devasahayam

Models of the Visual System
Edited by George K. Hung and Kenneth J. Ciuffreda

PDE and Level Sets: Algorithmic Approaches to Static and Motion Imagery
Edited by Jasjit S. Suri and Swamy Laxminarayan

Frontiers in Biomedical Engineering:
Edited by Ned H.C. Hwang and Savio L-Y. Woo

Handbook of Biomedical Image Analysis:
Volume I: Segmentation Models Part A
Edited by Jasjit S. Suri, David L. Wilson, and Swamy Laxminarayan

Handbook of Biomedical Image Analysis:
Volume II: Segmentation Models Part B
Edited by Jasjit S. Suri, David L. Wilson, and Swamy Laxminarayan

Handbook of Biomedical Image Analysis:
Volume III: Registration Models
Edited by Jasjit S. Suri, David L. Wilson, and Swamy Laxminarayan

Complex Systems Science in Biomedicine
Edited by Thomas S. Deisboeck and J. Yasha Kresh

Complex Systems Science in Biomedicine

Edited by

Thomas S. Deisboeck
Department of Radiology
Massachusetts General Hospital, and
Harvard Medical School
Boston, Massachusetts

and

J. Yasha Kresh
Department of Cardiothoracic Surgery and Medicine
Drexel University College of Medicine
Philadelphia, Pennsylvania

Springer

Thomas S. Deisboeck, M.D.
Assistant Professor of Radiology (HMS, MGH, HST)
Director, Complex Biosystems Modeling Laboratory
Harvard–MIT (HST) Athinoula A. Martinos Center for Biomedical Imaging
Massachusetts General Hospital–East Bldg. 149, 13th Street, Charlestown, MA 02129
deisboec@helix.mgh.harvard.edu

J. Yasha Kresh, Ph.D., F.A.C.C.
Professor and Research Director
Dept. of Cardiothoracic Surgery and
Professor of Medicine
Director, Cardiovascular Biophysics
Drexel Univ. College of Medicine
215 N. 15th Street, MS# 111
Philadelphia, PA 19102-1192
jkresh@drexelmed.edu

Front cover: The first figure appears courtesy of Gustavo Stolovitzky (IBM T. J. Watson Research Center). The second appears courtesy of J. Yasha Kresh (Drexel University College of Medicine). The third appears with permission from *Nature* http://www.nature.com/ and originally appeared in print as Figure 1 in *Nature* **411**:41–42, 2001 ''Lethality and centrality in protein networks,'' by H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai. The fourth appears courtesy of Ricard V. Solé (ICREA Complex Systems Lab, Universitat Pompeu Fabra). The right-hand figure appears courtesy of Josh Snyder, David Tuch, Nouchine Hadjikhani, and Bruce Fischl (Athinoula A. Martinos Center for Biomedical Imaging, Harvard Medical School).

Library of Congress Control Number: 2005934914

ISBN-10: 0-387-30241-7
ISBN-13: 978-0387-30241-6

Printed in the United States of America

9 8 7 6 5 4 3 2 1

springer.com

ACKNOWLEDGMENTS

We gratefully acknowledge the participation of everyone involved in the making of this textbook. Our special thanks go to the contributing authors, whose expertise and enthusiastic commitments made this volume a reality. We also thank our colleagues, whose insights helped shape this book, in particular Tom Kepler, Stuart Kauffman, Ary Goldberger, and Bernard Blickman, as well as Yuri Mansury, Chaitanya Athale, Brian Gregor, Meg Etherington, and Pam Fried. We especially appreciate the energy and excitement of the Springer publishing team (Aaron Johnson, Tim Oliver, Jasmine Benzvi, Shoshana Sternlicht, and Krista Zimmer), whose unwavering patience and tenacity ensured that the project go the distance. Finally, our deepest thanks to our families, who encouraged us with their love and support through the years of continuous intensity and concentration that this effort required. We could not have done it without you: Lizette M. Pérez-Deisboeck and Myrna P. Kresh.

Thomas S. Deisboeck, MD
Boston, Massachusetts

J. Yasha Kresh, PhD
Philadelphia, Pennsylvania

PREFACE

Work on Deisboeck and Kresh's *Complex Systems Science in BioMedicine* started years ago. In fact, thoughts and ideas leading up to this textbook date back to our first conversation, sometime in the fall of 1996. We quickly found common ground, and talked about emergence and self-organization and their relevance for medicine. We were both fascinated by the idea of complexity and marveled about its tremendous possibilities for cancer research, which was then and still is Tom's main scientific interest. Much has happened in science and technology since we first discussed our vision. For instance, in a remarkable international effort the human genome has been deciphered, nanotechnology has become a household name, and computing infrastructure, a critical enabler, is as powerful and affordable as ever before.

It is exactly *because* of this unprecedented progress that *Complex Systems Science in BioMedicine* is now making a case for a new approach in the life sciences. So let us start then with the obvious question first: why do we need a new fresh approach to ensure continued progress in the biomedical sciences? Did decades of methodically thorough research not yield great accomplishments and trigger an unparalleled productivity, with each year seeing thousands of scientific papers published in peer-reviewed journals? Certainly. *Reductionism* has led to ever-growing knowledge about isolated molecular pathways and selected portions of disease processes. We concede, dissecting biological mechanisms into bits and pieces has been utterly successful—if the number of fragmented discoveries is to be the decisive parameter. However, if we take understanding connectivity across scales, or better yet, *function* as the yardstick for measuring scientific achievements, much less progress can be claimed. Neither the vision nor the technical tools necessary to achieve these goals are "mainstream" yet. But there are signs in the biomedical sciences that things are changing—clear signs.

Indeed, most of the field involved in mapping the human genome in the 1990s is now engaged in *functional* genomics. Beginning to realize that the sum of its genes and proteins will not be able to explain a single cell's behavior, much less cell–cell interaction dynamics, let alone entire organ systems, we remember Aristotle, who had already argued that "*The whole is more than the sum of its parts*." For biomedicine it means that, no matter how many more

details we enthusiastically discover on ever smaller scales, we fail in deducing the complexity of a cell or multicellular tissue on the basis of this fragmented knowledge alone. In other words, piecing it together afterwards will not work. We need a new scientific approach, one that takes the nonlinearity of the majority of biological processes as much into account as their multi-scaled character. We believe that we are at a crucial bifurcation, where we need to integrate knowledge rather than dissect it, where we need to collaborate intensely across disciplines, theoretically and experimentally, in order to move forward. Complex systems science can match this challenge. Intrinsically multidisciplinary, it comprises concepts and quantitative tools that enable us to investigate how multiple biological elements interact and how molecular networks guide cell behavior and ultimately determine tissue function.

You might wonder how this is any different from, say physiology, a cornerstone of classic biomedical training. Indeed, physiology, the science of how living organisms function, may well be regarded as a predecessor of what many in the computational biology community now call "systems biology" and which clearly overlaps with complexity science in its goals. Where they differ, however, is in the *approach* to get there. Complex systems science applies a set of concepts and quantitative tools that are based on analogy and commonality, if not universality, between distinctively different systems, biologically or otherwise. Let us give you an example. The reason my, i.e., Tom's, laboratory developed an agent-based model to study cancer cell migration was an admittedly rather tired look out of a window while approaching London's Heathrow Airport by night several years back. What caught my attention was that, from above, the busy suburbs and streets resembled the cellular clusters and path patterns of a growing biosystem where single cells rather than people represent the system's individual "agents." Could one possibly investigate the metabolism-driven interaction of a rapidly evolving multicellular system, internally and with its microenvironment, in a way similar to how social scientists analyze the adaptive, economically driven behavior seen in expanding human societies? If so, then why not try an urban-planning approach for cancer research in an effort to better understand the dynamics of growth, migration and aggregation in tumor cell populations? Chapter 6.3 (Part III) summarizes some of the intriguing results arising from this line of work. This example illustrates how complex systems science approaches the problem at hand with tools adapted from nonlinear dynamics, applying sometimes rather abstract modeling and simulation techniques ranging from network theory to agent-based frameworks. It follows a "top–down" concept based on the claim that abstraction, not simplification, is the key to understanding the complexity of interaction between multiple parts on and across various scales of interest. That, however, is distinctively different from classic physiology, which uses biophysics and engineering concepts to describe the biological entity of interest in as much detail as available and, thus, "bottom–up." Let us emphasize that tackling the very same scientific problem from two seemingly opposing sides should not be seen as much as a case of competing approaches but as an exciting opportunity to exploit their mutual strengths in going forward.

Complex Systems Science in BioMedicine presents some of the fundamental theoretical basics of this rapidly emerging field and exemplifies the potential of the new approach by studying such diverse areas as molecular networks and developmental processes, the immune and nervous systems, the heart, cancer, and multi-organ failure. In this effort, the book itself follows a *multi-scaled* approach from molecular to macroscopic, thereby discussing both the normal and diseased states in selected topics. The invited contributions intentionally represent the dynamic state of the field in that biophysics, bioengineering, and computational biology modeling works are put side by side with complex systems-driven approaches. We believe that such juxtaposition not only anchors the new approach properly in established terrain but also helps showcase the differences.

A section on *emergent* technologies, no matter how long, can hardly ever be complete and, since the book was started years back, must run the risk of being outdated by the time of publication. By taking this risk we show by example that this novel approach has already led to and will continue to inspire design and development of cutting edge technology, ranging from micro-fluidics and innovative database management to multi-scale bioengineering, neuromorphic systems, functional MR imaging, and even operating room design. Undoubtedly, these and other techniques will feedback vital data and thus help complex systems science achieve its goals.

Finally, is there something like complex systems *science* at all or is it merely a powerful tool kit? As stated earlier and as reviewed in the book, there are certain techniques that are ubiquitous for the study of complex systems in economics, population dynamics, and biology. The title of the book reveals that we advocate the application of these techniques *also* to relevant areas in biomedicine where reductionism may have reached its limits. Nothing more, nothing less. As such, this book presents visionary ideas and their potential impact on future directions in biomedical research. It is not and cannot be definitive. Rather, we let the reader judge how far this, our field, has come, and if the presented work at this stage represents merely a promising, fresh approach or if it already signals the dawn of a new and yet to be fully defined science.

As described in detail in Yasha Kresh's introductory chapter, the origins of applying systems ideas in one form or another to the life sciences date back at least several decades. And while initial efforts to move complex systems further into the center of mainstream medicine were undertaken by a few pioneers, this has certainly changed. Over the last years, many colleagues have embraced the necessity of moving in this new direction, also documented by the enthusiastic feedback we received when we asked for participation in this multi-authored book. The newly established multidisciplinary graduate and postgraduate training curricula, sprouting complex systems-related academic centers as well as novel crosscutting grant funding programs, are testimony that these ideas are starting to catch on. What counts now are the steps we take in order to further foster this nascent development. As such, if *Complex Systems Science in BioMedicine* can help draw more attention to the application of complexity techniques to important questions in biomedicine and thus help support ongoing

and upcoming scientific, teaching, and training efforts, we will consider it successful.

The quest for novel ways of thinking was what brought us together back in 1996, first as colleagues, now also as friends. It is the immense potential of complex systems science that provided a source of relentless energy for this textbook and that continues to fuel our scientific work.

Thomas S. Deisboeck, MD
Boston, Massachusetts

Stuart A. Kauffman, MD
Santa Fe, New Mexico
2004

CONTENTS

Part I: Introduction

Part II: Complex Systems Science: The Basics

Section III.6: Cancer: A Systems Approach

Section III.7: The Interaction of Complex Biosystems

Part IV: Enabling Technologies

Part I

INTRODUCTION

INTEGRATIVE SYSTEMS VIEW OF LIFE: PERSPECTIVES FROM GENERAL SYSTEMS THINKING

J. Yasha Kresh

Departments of Cardiothoracic Surgery and Medicine,
Drexel University College of Medicine, Philadelphia

The application of systems thinking and the principles of general systems science to problems in the life sciences is not a new endeavor. In the 1960s systems theory and biology attracted the interest of many notable biologists, cyberneticists, mathematicians, and engineers. The avalanche of new quantitative data (genome, proteome, physiome) incited by the boundless advances in molecular and cellular biology has reawakened interest in and kindled rediscovery of formal model-building techniques. The manifold perspectives presented in many ways is a re-embodiment of the general theory of organismic systems and serves as an impetus to suggest that organized complexity can be understood. The particular affinity expressed in this essay is a reflection of how closely my thinking is associated with the thoughts of Ludwig von Bertalanffy, Ervin Laszlo, and Robert Rosen. We are, by all accounts, at the threshold of a postgenomic era that truly belongs to the biology of systems.

> *Thus, the task is not so much to see what no one yet has seen, but to think what nobody yet has thought about that which everybody sees.*
>
> —Schopenhauer

Systems here ... systems there ... systems everywhere

Address correspondence to: J. Yasha Kresh, Departments of Cardiothoracic Surgery and Medicine, Drexel University College of Medicine, 245 North 15th Street, MS#111, Philadelphia, PA 19102-1192 (JKresh@DrexelMed.edu).

1. INTRODUCTION

The historical framework and ideas presented here feature the disciplines that spawned the science of complex systems (e.g., self-organizing, autopoietic networks, dissipative structures, chaos, fractals). In particular, we use general systems theory (GST), control system theory (i.e., cybernetics, homeodynamics), and dynamical systems theory (nonlinear, chaotic), the forerunners of creative systems thinking, to formulate a coherent theory and elucidate the essential properties of biological phenomena such as structural and functional organization, regulatory control mechanisms, and robustness and fragility.

The defining aims of systems thinking:

— The Believing: why do I see what I see?
— The Being: why do things stay the same?
— The Becoming: why do things change?

The notion of a system comprised of interdependent elements has been the subject of human concern and inquiry for centuries. Man has explored the solar system and the constellations since the beginning of recorded time. We, as a species, have struggled with the complicated array of interconnected elements that control our internal and external world. The more formal understanding of a system, offered by systems science, as a complex of components and their interactions has not changed dramatically through the years.

An inkling of systems science was anticipated by the *Gestalten* in physics, a natural worldview proposed in the 1920s. According to the great leader in the field of GST, Ludwig von Bertalanffy, the ideas of physical *Gestalten* were the precursors intended to elaborate the most general properties of inorganic compared with organic systems. It is worth mentioning that physicists study closed systems, as compared with real systems, that communicate and exchange energy (information) with the environment and thus self-organize, learn, and adapt. Of particular note is the historical precedence that gave rise to the genesis of systems theory as a reaction to the confinement of reductionism and motivated by a keen desire to reestablish the unity of science. Some aspects of intellectual tradition and scientific history are worthy of repetition.

Systems was and remains a fashionable catchword. In the introduction to his seminal book, *General System Theory* (1), von Bertalanffy wrote in 1967 that the concept of systems permeated all fields of science as well as popular thinking, jargon, and mass media. Common parlance continues to include concepts such as adaptation, control, differentiation, dynamic behavior, hierarchy, robustness, reliability, and sensitivity.

The reader is encouraged to visit the Principia Cybernetica website (http://pespmc1.vub.ac.be), an extensive condensed repository of historical and contemporary thinking addressing the age-old philosophical question—What is the meaning of life?—by starting with a formal definition:

Systems Theory:

The transdisciplinary study of the abstract organization of phenomena, independent of their substance, type, or spatial or temporal scale of existence. It investigates both the principles common to all complex entities, and the (usually mathematical) models that can be used to describe them (2).

2. GENERAL SYSTEM THEORY: THE LAWS OF INTEGRATED WHOLES

Von Bertalanffy (1) developed the tenets of system theory in the late 1920s (when he himself was in his twenties). He drew attention to a new perspective as a method, which he called "organismic biology," that assigns a self-organizational dynamics to biological systems. To this end he developed the kinetic theory of open systems, characterized by equifinality and steady state. His main goal was to unite metabolism, growth and morphogenesis, and sense physiology into a dynamic theory of stationary open systems. He spoke of it as an attempt at explanation, calling it "The System Theory of the Organism." It was not until the late 1940s that he recognized that "there exist models, principles and laws that apply to generalized systems or their subclasses irrespective of their particular kind, the nature of the component elements, and the relations or 'forces' between them. We postulate a new discipline called General System Theory." What sustains this systems view is the recognition that one cannot compute the behavior of the whole from the behavior of its parts. More importantly, the preservation of the multitude of interacting atoms, molecules, cells, tissues, and organs is valued by the complex of relationships that entail the organization and not by the individuality of their participation.

When we try to pick up anything by itself
we find it is attached to everything in the universe.
—John Muir

This grand unification concept was criticized as pseudoscience and said to be an attempt to connect things holistically. Such criticisms would have dissi-

pated with the recognition that GST is merely a perspective or paradigm and that such basic conceptual frameworks are central to the development of exact scientific theory and a new way of doing science. GST was not meant to be a single overarching theory (which history tells us has a short-lived existence). Above all, it is a system-theory; it deals with systemic phenomena—organisms, groups, and the like (e.g., nations, economies, biosphere, astronomical universe). It views a system as an integrated whole of its subsidiary components, not a mechanistic aggregate of parts in isolable causal relations (3).

Some of the concepts and principles are rigorous enough to be considered laws in addition to providing a general framework for theory construction. "If this be considered not enough, the reader would do well to remember that a true general theory of all such varieties of systems would constitute a master science that would make Einstein's attempt at a unified field theory pale by comparison" (from Foreword by Ervin Laszlo for a collection of essays gathered together and published in honor of von Bertalanffy two years after his death in 1972). As it was then and remains now, the science of systems is not restricted to a particular level of biological order or set of relationships. This perspective is all inclusive; it allows us to look at a gene network or a cell as an integrated system or to look at the organ, the organism, the family unit, the community, nation, and the biosphere as an organized system (see Figure 1). The concept of a holon (from the Greek *holos* = whole) is used to explain the unity of greater purpose. Arthur Koestler popularized this term to describe the hybrid nature of subwholes/parts in living systems (4). A natural byproduct of this view of a system is the holarchy that is formed in which systems are simultaneously self-contained wholes in relation to their subordinated parts and dependent parts when viewed by the overarching whole (Figure 2). The manifestation of a relationally distributed control structure is the creation of autonomous, self-reliant functional modules that can handle contingencies without central control or intervention.

3. SYSTEMIC PRINCIPLES OF CYBERNETICS

> *Information is information not matter or energy. No materialism which does not admit this can survive at the present day.*
>
> —Norbert Weiner

A special branch of general systems theory that studies systems that can be mapped using loops or looping structure became known as cybernetics. The term cybernetics stems from the Greek *kybernetes* (meaning steersman, governor, or pilot as in autopilot). It became known as a theory of the communication and control of regulatory feedback (information loop). The modern abstract view of cybernetics encompasses the study of systems (subsystems) and their

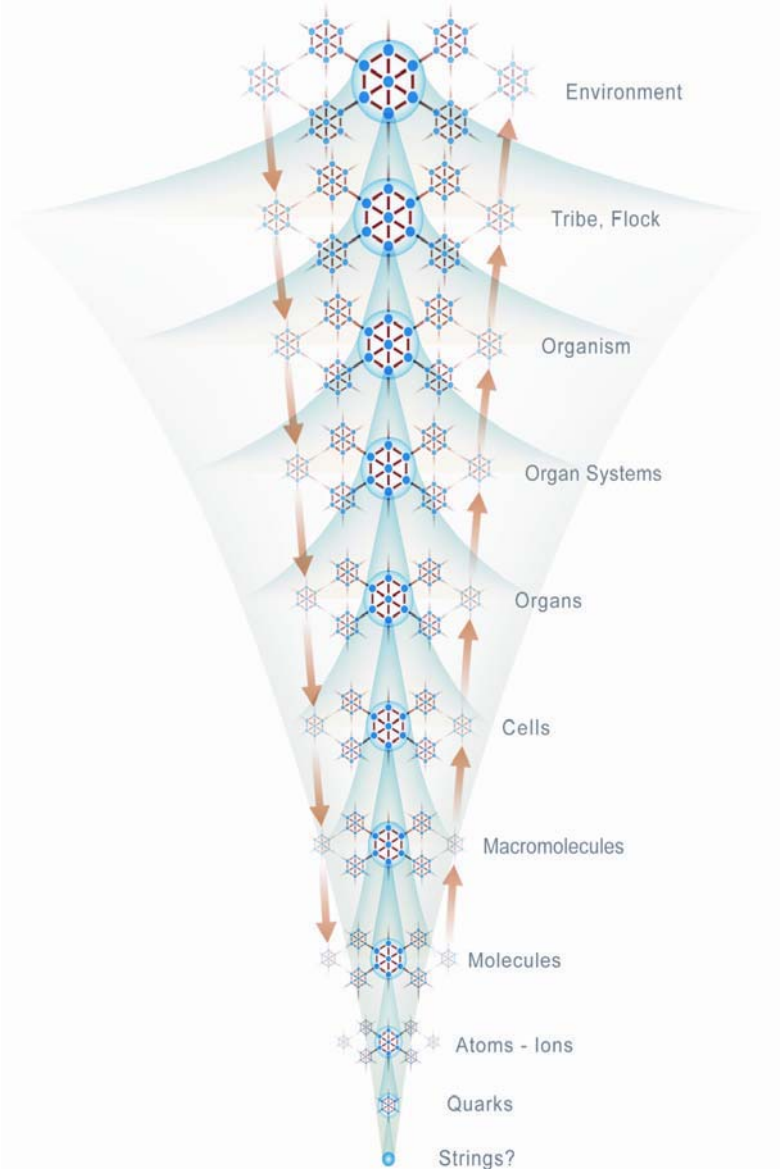


Figure 1. Holarchies and the order of nature: hierarchical structures/units of life leading to complexification in organizational order. The notion of entities that are "independent wholes" and "dependent" parts seen as an overarching assimilation of lower order "parts" into the adjoining level of "wholes." The part–whole Holon dualism allows for concurrent upward–downward causality (arrows) to coexist (4). The overarching levels of interconnected and interdependent continuum suggest an integrated worldview perspective and thinking. The basic causal tension between parts (i.e., mechanistic, reductionist, atomistic) and whole (i.e., organistic, systemic, ecological) is depicted by arrows. (Artwork by M. Clemens.)

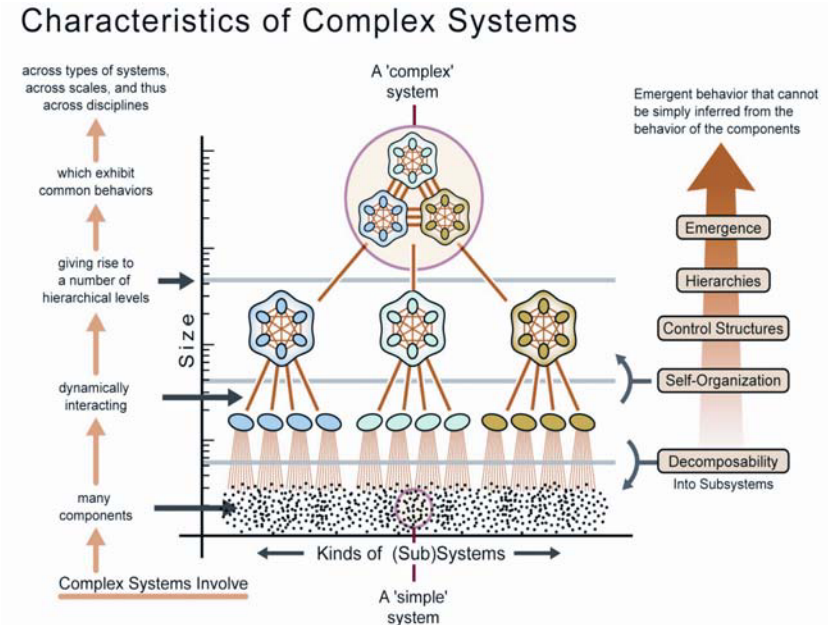


Figure 2. Abstract representation and synthesis of "Complex Systems," including their constituent hierarchical organization entities (components/elements) and dynamic relations (relational emergent function). Note that order and common behavior may arise from both self-organization and control structures. (Artwork by M. Clemens.)

control (Figure 3). Emphasis is placed on the mechanistic relations that hold between the different parts of a system (i.e., input-sensors, controller-centers, output-effectors). The basic premise of cybernetics is the transfer of information and the circular relations that define feedback, self-regulation, and autopoiesis. Cybernetics contributed the understanding of goal-directedness or purpose made possible by a negative feedback loop that minimizes the deviation between outcome and desired goal (Figure 3). The brain–body coupling plays a prominent role in the cybernetic model of regulation and control. The foundation of closed-loop autonomic control is information transmission and the enabling communication pathways, facilitated by sensors (i.e., chemoreceptors, mechanoreceptors/pressoreceptors) and effectors (e.g., sympathetic drive, endocrine release) that couple neural processes (e.g., medulla) to myriad regulatory processes (Figures 3 and 5). Considerable overlap exists between regulatory cycles and centers of the limbic system and the various homeostats that constitute the endocrine, immune, and nervous systems. A disturbance to organismic regulation can occur at multiple levels and is prone to modulation by sleep, wakefulness, and emotional states. These closely coupled interactions give rise to a dynamic equilibrium of the controlled process, manifested as homeodynamic stability, an

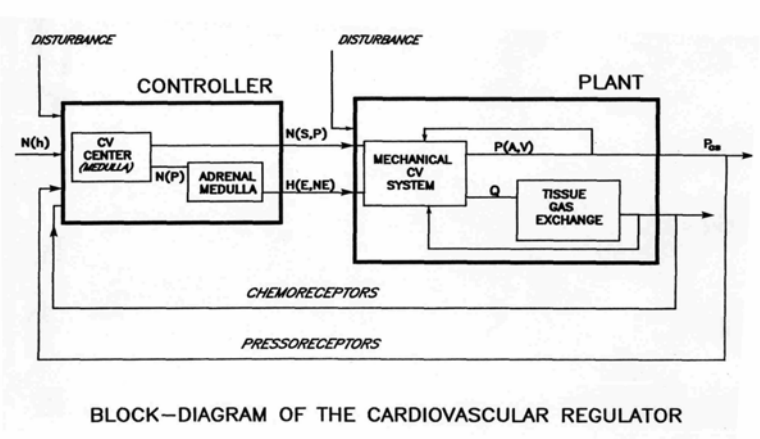


Figure 3. Canonical *closed-loop* control system organization depicting the flow of information as part of the conceptual (cybernetic) model of homeostatic (linear and nonlinear) regulation (systemic blood pressure and oxygenation). The block diagram generalizes the structure and function of the "Controller" (e.g., brain) and "Plant" (e.g., cardiovascular system); feedback is facilitated via chemomechanical sensors (receptors) and other "smart elements" (not easily localized) that can read signals and appraise status. Self-regulation can be achieved in the presence of noise or imposed disturbance (e.g., blood loss, posture/altitude changes).

expression and capacity of complex systems to withstand fluctuational changes from internal and external environments (5,6).

4. BIOLOGICAL SYSTEMATICS: UNDERSTANDING WHOLE SYSTEMS

Cybernetics also deals with how living systems/subsystems regulate, control, and reproduce themselves and how, in turn, they can produce other subsystems that are goal-directed, self-regulating, or self-reproducing. Cybernetics is concerned with understanding the self-organization of human, artificial, and natural systems including the understanding of its own functioning. Importantly, cybernetic systems do not have the means to evolve from a lesser to a more differentiated state. Cybernetics was part of the systems thinking movement and an essential component in the growth of scientific knowledge in the 1940s, motivated by a desire to understand life in its entirety.

W. Ross Ashby (7), Norbert Wiener (8), and Warren McCulluch (9) are credited (albeit, Ashby is less known) with the early formulation of cybernetics inquiry; they emphasized communication and control, the processes of self-

organization and self-regulation, and circular causal feedback mechanisms in the animal and the machine (e.g., robots). Some of these systemic principles and perspectives were assimilated by computer/cognitive sciences and are credited with being at the core of neural network approaches in computing. In addition to the early emphasis placed upon the observed system, the importance of the observer (see Figure 3) has to be considered.

Who will integrate the integrators?
—Margaret Mead

Heinz Foerster (10) recognized the need for a theory of the observer, i.e., description of the describer (see Figure 6). A strong case was made for the need of a transdisciplinary synthesis of a representational framework that can consolidate the concept of self-reference and the meaning of cognition and communication within the natural and social sciences, the humanities, and information science. Because the structure and function of a system cannot be understood in isolation, cybernetics and systems theory should be viewed as two facets of a single approach.

General system theory encompasses the cybernetic theory of feedback, which represents a special class of self-regulating systems. In both cases, the parts entail the structure and function of the whole and as such are not isolable. Nonetheless, a fundamental difference exists between GST and cybernetics, whereby the feedback mechanisms (see Figure 3) are controlled by local constraints in contrast to the free multilevel interplay of the network of reactions in dynamic living systems. Moreover, the regulative mechanisms of cybernetic systems are based on predetermined (fixed) structural feedback. This implies that they are closed systems with respect to exchange of energy and matter and as such do not have the essential characteristics of living systems whose components undergo growth, development, and differentiation, which "shows the existence of a general systems theory that deals with formal characteristics of systems, concrete facts appearing as their special applications by defining variables and parameters. In still other terms, such examples show a formal uniformity of nature" (1). The concept embraced by GST is a broader one and is responsible for the development of the modern studies of nonstationary structures and the dynamics of self-organization in our attempt to understand how the pattern formation functions (see Part II, chapter 4, by Wuchty, Ravasz, and Barabási).

In biology (as well as in behavioral and social sciences), one often encounters phenomena that are poorly explained by the inanimate system of physical laws. When analyzing living objects (or behavior), the tendency is to use functional attributes of the component parts and biochemical processes that are hierarchically organized to maintain the integrity, development, and progression of the system in question. This is not to suggest some vitalistic or metaphysical

purposiveness is at play, dedicated to preserving the omnipresent biological order. A good example of organized complexity (i.e., superposition of system upon system) is the human immune system (see chapter 4.1 by Segel, Part III, this volume), which is comprised of nearly a trillion cells and hundreds of signaling chemicals that regulate with exquisite precision the myriad pathogens that roam the body. The immune system parts are engaged without a central organizer (albeit signals from the brain can modulate its action) to control the detailed action plan.

> *Everything should be as simple as possible*,
> *but not simpler*.
>
> —Albert Einstein

To put the self-organized, parts-collective in perspective (see Figure 4), it is inviting to look at the complexity of information and energy processing that must take place in the human body as a whole. The human organism consists of roughly 50 trillion subsidiary component parts (cells), 40,000 different types of proteins, and a genetic code of approximately 1.5 GBytes (6 billion base pairs or 3 million nucleotides/haploid genome). It is revealing to note that the average

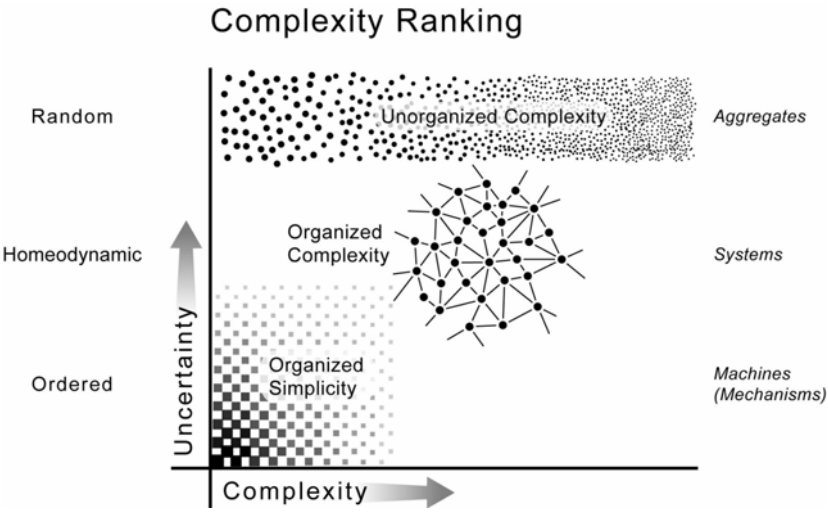


Figure 4. Types of systems with respect to methods of study and complexity ranking, as given by Weinberg (11). Conceptual representation and generalized systems view of "organized complexity." Note the ranking and association between uncertainty and complexity (11) and their mutual interaction in defining ranges of systems (from machines to random aggregates).

person is a carrier of some 40 billion fat cells all in pursuit of a collective (hips, thighs) or a community (organs, abdomen) to inhabit. These too are self-sufficient part–whole cellular entities. The total power consumption of an adult human is that of a 100-watt light bulb. Each individual cell is an intricate self-contained chemical computer that can perform over 10 million chemical reactions per second. Correspondingly, the cerebral cortex of the human brain contains nearly 20 billion neurons, each with over 2,000 synapses, ready to communicate and exchange signals with each other and the rest of the body. The power consumption of this subsystem is surprisingly high (~33% of body total), i.e., 1000 times greater energy utilization than other cell types. (Our mothers were quite insightful, insisting that we cover our heads on a cold day.) The unlikely comparison of neurons to their electronic equivalent translates into the sum total of transistors comprising 500 Pentium-4 microprocessors. The corresponding processing power of the brain is estimated to be 50 terabits per second (compared to ~25 gigabits per second for a Pentium-4). For the brain the emergent complex systems properties manifest attributes such as consciences, memory, and ability to learn. These system-derived properties cannot be understood by studying the neurons or their topological distribution.

Clearly, organismic processes are deliberately ordered to maintain and preserve the integrity of the system. In contrast, the physicochemical processes occurring in an organism that has been impaired (by disease, pathologic condition) still follow the conventional laws of physics but differ profoundly in terms of principles of relational organization and order from the identifiably normal (healthy) system. Molecular biology is not going to give us all the information we need. The information about the whole (collective behavior) is larger than the sum of the information about the parts, i.e., the missing link in the pervasive reductionism practiced today. What is especially needed is a coherent picture of how this information is being used to carry out biological functions.

4.1. Distributed and Shared Regulation

The classical concept of cardiac neural regulation presumes that the neural efferent signals originate from extracardiac centers and, in particular, from the central nervous system (CNS). A byproduct of this supposition is that the cardiac afferent information is considered relevant and meaningful only if it is transmitted directly to the cardiovascular regulatory centers residing within the CNS. In this view, information processing is delegated exclusively to the CNS, whereas the intracardiac ganglia are assigned the passive role of a relay station. This limiting perspective of cardiac neural control is no longer tenable, particularly because it does not make allowances for the existence of the intrinsic components of neural regulation. In studies of patients undergoing heart

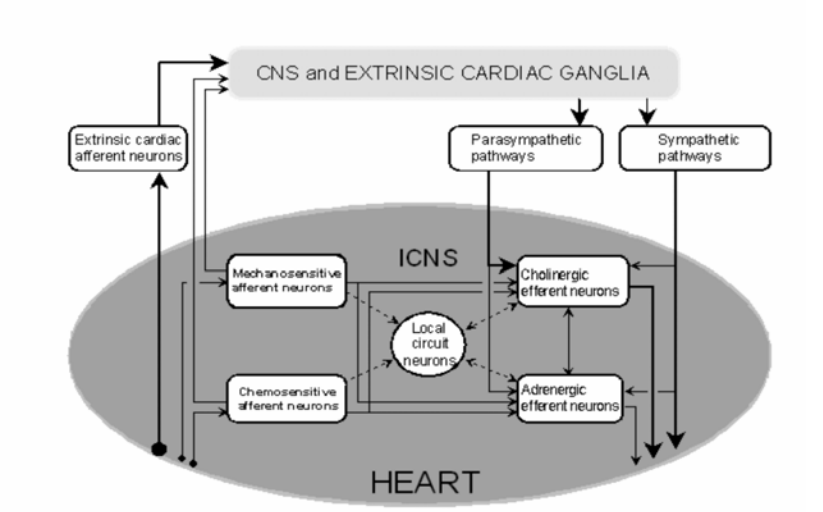


Figure 5. Conceptual scheme of the intrinsic cardiac nervous system (ICNS). Intracardiac afferent neurons provide MECHANOsensitive and CHEMOsensitive input from atrial and ventricular tissues to the intrinsic efferent adrenergic and cholinergic cardiac neurons. CNS = central nervous system. For simplicity, the known sympathetic–parasympathetic interactions and other known efferent intracardiac neurons are not shown. Bold lines represent the pathways of extrinsic cardiac neural feedback control. Thin lines represent intrinsic cardiac neural pathways, the functional role of which remains to be established.

transplantation, it was observed that, although cardiac allografts are extrinsically decentralized, they retain a viable intrinsic neuronal system (12).

An increasing body of evidence has accumulated (5,12) identifying a variety of neural cells residing in the heart and having distinct and significant effects on cardiac performance. The premise that the heart is not merely a muscular pump but is endowed with a level of self-organized neuroendocrine self-regulation is very compelling. In broader terms, the concept of self-regulation is based on the axiom that the heart is a regulatory system, integrating many components, including endothelium-mediated control and afferent/efferent neural mechanisms, and thereby provides feedback of its beat-to-beat performance as a muscular pump (Figure 5). This view would suggest the existence of an intrinsic neural network processor. In fact, the intrinsic neural network is organized such that it functions as a neural center (heart brain) and can facilitate local control of the disparate heart functions and integrate them such that their responses are not merely parallel but tuned (optimized) to accommodate the varied influences on the heart. This local processor might behave as a functional intrinsic cardiac nervous system (ICNS). The conceptual understanding of the functional structures embodied by heart brain is schematized in Figure 8.

The ICNS has many of the complex attributes associated with the CNS, incorporating afferent and efferent components mediating its activity. This form of functional organization provides the heart with the ability to fine tune its adaptive organ-system response. The inherent capacity of the cardiac ganglionic neurons to respond to local mechanical and chemical stimuli may facilitate the means for the adaptive intrinsic mechanism to operate under stress and pathophysiologic conditions (i.e., neuropathy, transplantation, and aging). Ultimately, the intrinsic regulatory neurogenic mechanisms of the ICNS along with the CNS mutually negotiate the functional role that the autonomic nervous system plays in the control of the automaticity, electrical propagation, and contractility of the heart (12).

Likewise, there is compelling evidence of distributed feedbacks with multiple overlapping and surprisingly conflicting short-term goals in the immune system (see chapters 4.1 [by Segel] and 4.2 [by Kepler], Part III, this volume). Information about and progress toward goals are being monitored from sensor detection and are broadcast to the system via vectors of signaling chemicals (i.e., cytokines). This sensor-driven strategy of distributed feedbacks helps improve the performance of a preferentially selected effector cell.

4.2. Multilevel (Hierarchical/Heterarchical) and Distributed Organization

Living systems are organized such that they manifest operational features ascribed to hierarchical and heterarchical structures. The functional organization is inherently a heterarchy of interrelations and as such has no obvious or fixed order rank. Unlike machines and/or mechanisms, the functional hierarchy does not dictate level and importance of cooperativity.

> *No man is an island—he is a holon. A Janus-faced entity who, looking inward, sees himself as a self-contained unique whole, looking outward as a dependent part.*
>
> —Arthur Koestler

The basic rules of distributed cooperation (i.e., superposition of system upon system) are inspired by precepts of holarchy, defined (4) as a hierarchical organization of self-regulating entities (holons) that function as autonomous wholes in supraordination to their parts and as dependent parts in subordination to controls on higher levels defining their function. This superposition perspective implies that natural systems are organized such that every system level (see Figure 1) is constrained by the immediate next level above it and similarly by the supporting level below it. This arrangement, in coordination with the local

environment, promotes stability, robustness, and adaptation. Evolution seems to favor the building design of hierarchical order. The apparent advantages of a multilevel pyramid (see Figure 2), with simple systems at the bottom and more complex ones at the top, are the interfaces and linkages that are created by the intermediates. The nature of these subsystems is dualistic: they behave as integrating wholes to their respective parts and as parts to their respective higher level wholes. The hierarchically organized benefit of this arrangement is inherent in this modularity, whereby the decomposition into subsidiary parts does not ruin or unbalance the entirety of evolutionary organization. Herbert Simon (13) showed mathematically that complex systems evolve from simple systems with greater rapidity if stable intermediate forms exist than if they do not.

All in all, evolution keeps the conserved biological (sub)systems in check and thus robust to uncertainty in the local environment and to failure of the component wholes. It would seem that reductionism, in its current incarnation, is not likely to concatenate the fractionated parts together so as to make the selectively disintegrated living organism whole again.

4.3. Heterarchy (Def: The Other, the Alien + to Reign, to Govern)

Organizational features embodied by heterarchical systems and the topologic character of nested closed circuits, which were introduced nearly half a century ago by the neurophysiologist and cybernetician Warren McCulloch (9), can be considered a superset of the ordinary hierarchical forms. The concept of heterarchy captures the essence of networked dynamic structures, in which the center of control (authority) is redirected to whichever point is most relevant and useful to accomplish the purposive activities. This form of organizational diversity is particularly prevalent in brain function and autonomic function.

> *Today the network of relationships linking the human race to itself and to the rest of the biosphere is so complex that all aspects affect all others to an extraordinary degree. Someone should be studying the whole system, however crudely that has to be done, because no gluing together of partial studies of a complex nonlinear system can give a good idea of the behavior of the whole.*
>
> —Murray Gell-Mann

The disintegration of a multilevel system can come about as a result of dysregulation in the level of communication (downward/upward causation). One can argue that the abnormal growth of individual cells (certain types of cancers) might be the result of loss of optimal amounts of communication (excessive,

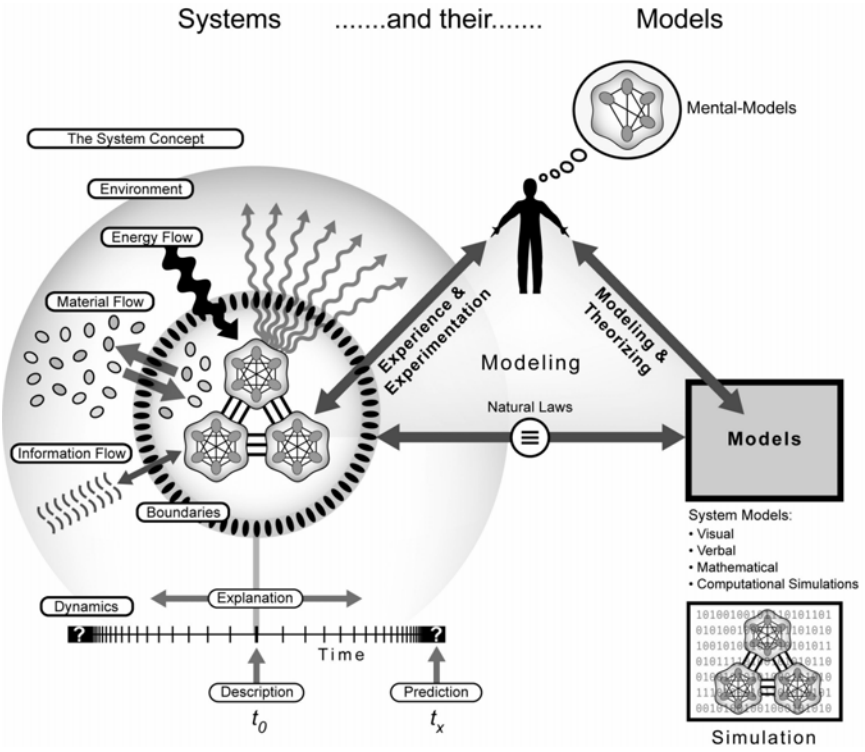


Figure 6. Systems Methodology. The associated steps involved in constructing a systems theoretical model (e.g., visual, verbal, mathematical, computational simulation). The environment acting through its operational agents (information, material flow, energy flow) alters the system's program and thereby the identifiable variables. Thereafter, the observer's specified program and the inherent limitations (e.g., cognitive, conceptual, and inferential) dictate the resultant (emergent) model. (Artwork by M. Clemens.)

diminished, interrupted) between the subsystems (see chapters 6.1 [by Pienta] and 6.2 [by Solé, Gonzáles García, and Costa], Part III, this volume) on the same level or across a heterarchical (9) network. Life is an emergent property of complex systems. Disorganization and disorder in biological systems are manifestations of general system failure (see chapter 7.2 by Buchman, Part III) and lie at the root of acute trauma, diseases, and senescence (see chapter 7.3 by Lipsitz, Part III). This observation must be considered in a context that recognizes the truism that biological disorder is functional disorder order (and the converse). It remains unclear whether senescence is a passive process, brought about by loss in structural and metabolic integrity, or a direct consequence of changes in the epigenetic driving programs (reinforcement, reinitialization/rebooting) that ordinarily perpetuate the dynamic equilibrium (homeody-

namic (14) regulation) via the neural, endocrine, and immune communicative feedback/feedforward subsystems. From the thermodynamic perspective, if one follows the path that leads to system disorganization, the outcome is inevitable. The terminal state (death) is reached when a critical breakdown in the signaling network and in the connectivity of interacting organs, tissue, and cellular processes is breached. In the realm of scientific superstition, the observed changes in regulatory nets commit causation to a sure death.

5. SYSTEMS BIOLOGY AND MATHEMATICAL MODELING

The construction of mathematical models that realistically simulate whole-organ function, or a signaling pathway that regulates cell processes such as cell replication, would be extremely complex and computationally intractable. The time scales of life events range from microseconds (molecular motion) to years (life span). The needed spatial resolution is equally enormous (1 nm ion channel pore size to 1 m body dimension). Clearly, a large assembly of models would be needed to cover the full span of biological hierarchy and order, each in turn able to couple to respective levels of known association (e.g., Human Physiome Project).

For example, it would seem straightforward that a simplified model system (e.g., myocardial tissue) that is restricted to only three cell types (neurocytes, endothelial, myocytes) can be fractionated into various two- and one-cell systems, each of which is suitably simplified to be understood in isolation. On closer examination (see Figures 7 and 8), it becomes obvious that the crucial integrity (stability) and information are completely lost. The simplest explanation offered by system theory is that the fractionation methodologies and analysis techniques used do not commute (see Figure 11) with the dynamic properties of the system as a whole. Moreover, the participation of the environment (local milieu) complicates the fractionation aftermath, destroying the dynamics of the system and its function, thereby preserving mutual information (e.g., genetic mutations and environment contribute to disease manifestation). It may interest the reader to know that this classic three-body problem is universally difficult to reconcile and that the acronym for the neural, endothelial, and muscle (NEM) cell arrangement means "No" in Hungarian. Surely, this is precisely what Ervin Laszlo (born in Budapest, 1932) would have said with regard to fractionation and coarse reductionism. The building of predicative models of cells, organs, and ultimately organisms need not be the sole course or salvation of reductionism. Acquiring analytical data and methods for integrating the network of genes with cells and whole organism remains an important endeavor of systems biology. Importantly, the quantitative understanding of the entire subcellular, cellular, or multicellular systems would dramatically alter the approach to and course of drug discovery and personalized medicine.

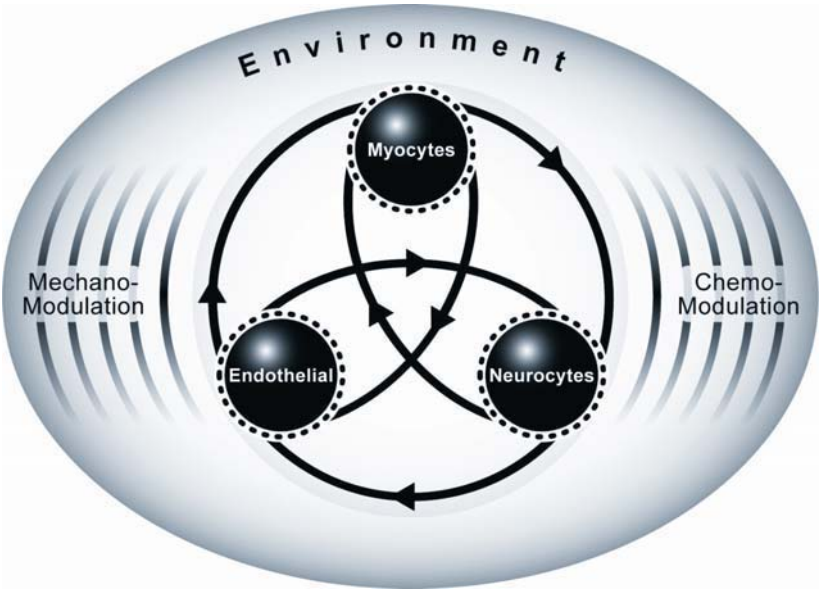


Figure 7. Three-Body Problem. The analytical solution to this problem is universally irresolvable (not "integrable"). In this example the mutual interaction between three cell types (neurocytes, endothelial, and myocytes) cannot be predicted/understood in terms of reduced sets of interactions, further confounded by the "surrounding" environment (mechano- and chemomodulation). This form of autopoiesis suggests a mutual coevolution in function. The joint interaction is figuratively localized at the intersection of the virtual dynamic "orbits" (arrows). This is a well-known problem of celestial mechanics that remains unsolvable unless the interaction is confined to a single plane.

As regards the relations among levels in the vast hierarchy-heterarchy of living order, researchers have expressed interest in simulating biological systems, i.e., bottom–up, commencing with single genes and protein molecules, or top–down, starting with large-scale physiological behavior (devoid of gene and protein–protein interactions). Notable modelers (see chapter 3.2 by Winslow, Part III, this volume) advocate a compromise, working in both directions from the middle (middle–out approach) because of the two levels of data-rich simulation available using this approach (15). A limitation of mathematical and Newtonian paradigms implies that living systems are in fact state-determined, i.e., explicit values can be specified that relate state variables to rates at all levels of organization.

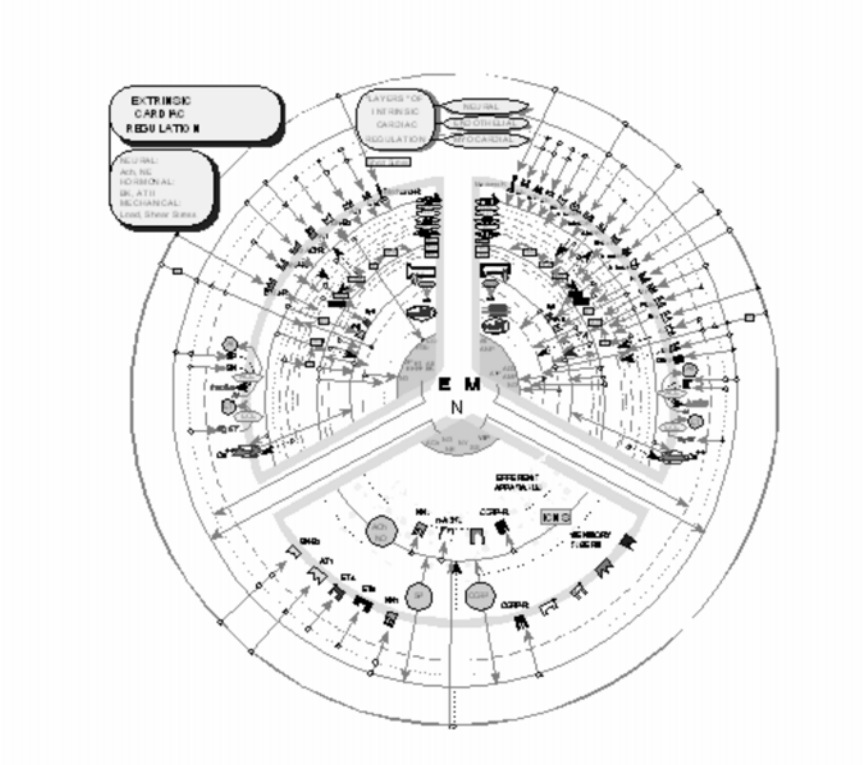


Figure 8. A web of intracellular–extracellular signaling pathways organized as complex regulatory networks. Shown for example purposes only is the intrinsic cardiac nested layers of self-regulation (see Figure 6). Diagrammatic generalization of the network of signaling events (receptor-ligand type) that mediate neuronal, hormonal, and mechanically dependent interaction between Neural (N), Endothelial (E), and Myocardial (M) cells. The emergent extrinsic function is manifested as a finite and ordered (mechanistic) expression, limited in the degrees of operational freedom, i.e., much of the "internal complexity" is not made evident in the externalized homeokinetic functions of regulated stable systems.

5.1. Transfer Function and Organizational Analysis

In those cases (of which there are many) in which there is no biodynamic theory to explain system behavior, it is nonetheless possible to gain some understanding and derive an empirical inference as to the complex structure of system integrity by observing the outputs generated by the system itself (see Part II, chapters 1 [Shalizi] and 2 [Socolar], this volume):

$$\text{Input} \longrightarrow \{\ \textbf{PROCESS}\ \} \longrightarrow \text{Output}$$

These measurable signal surrogates can be related to some relevant feature(s) of the system that generated them. In particular, an expression of the interplay between perturbation (internal/external) to system function and the dynamic response of the regulatory processes, i.e., homeodynamic processes, can be inferred using nonlinear time-series signal analysis techniques (12). Historically, the cardiovascular system (heart rate, blood pressure fluctuations) has been the beneficiary of this approach, primarily because of the ease and accessibility of system variables and the relatively well-characterized modulation of the autonomic response. This approach has gained considerable attention, not only in deciphering the dynamic structure that constitutes cardiovascular regulation but also as a window onto the genesis (conception, birth, puberty) and span (maturation, senescence, death) of human life.

The changes in physiological and functional decline accompanying aging (see chapters 3.2 [by Winslow] and 3.3 [by Glass], Part III, this volume) are an expression of the losses in the organizational integrity (loss of network connectivity, signaling regimes). This form of organismic dysregulation of hierarchical (feedback and feedforward circuits) organization can be conceptualized by a complexification score that is intimately dependent on the degradation, instability, and dropout of homeodynamic regulatory processes governing the trajectory of life including pathologic states, aging, and death.

Reconstituting the functional integrity of a biological system is not a simple act of replacing or putting the constituent parts back together. The main focus of biologists for the better part of the twentieth century was the disassembly of living systems to glean an understanding of the workings of the parts as members of the whole. This reductionist approach started with the cell and systematically descended to the genome itself. Not surprisingly, this exuberance of effort gave rise to a monumental amount of information that is now begging for reintegration into a systematic whole. The GST concepts promulgated throughout the 1960s have been resurrected in a reincarnated form—systems biology (16). Indeed, the lessons of our youthful past are visited upon us again.

Mihajlo Mesarovic (17) anticipated this disconnect in 1968: "It has been said too often, but has been really taken into account too seldom, that the theory and applications are intimately related and none can make significant progress without the other. Actually systems theorists tend to disregard this altogether and take the position that all that is needed next is that the biologists learn and apply systems theory. However, I would like to suggest that one of the many reasons for the existing lag is that *systems theory has not been directly concerned with some of the problems of vital importance in biology*."

The opportunity is ripe to revisit systems theory, its application to biology, and the lessons that can be learned from the early developments, the goal being to see how a more evolved perspective of living systems can provide a fresh look in the postgenome era of the transcriptome and proteome. Interest in formal mathematical models of biological hierarchical processes is increasing. The new

impetus is to model and treat the organization and regulation of genetic pathways as dynamic systems in which causation is a relationship, not between components, but between changes of states of a system (16). The general systems science paradigm, with its tool set for studying collective organization and emergent behavior, is a fitting natural conceptual framework for putting Humpty-Dumpty back together again.

Humpty-Dumpty sat on a wall.
Humpty-Dumpty had a great fall.

All the king's horses and all the king's men,
Couldn't put Humpty-Dumpty together again.
—Lewis Carroll

System science is not a ready-made collection of defined principles of biological organization. It is a construct for building formal models that need not be mathematical. The system approach entails the application of system theory methodology to the analysis and scientific explanation of biological phenomena. Linus Pauling said, "Life is a relationship among molecules and not a property of any molecule." A coherent framework for studying multilevel systems and their relational interaction is indispensable if any progress is to be made in understanding complex biological organization. It would also seem logical that these relational interactions have to be accounted for in space (nanometer to meter span) and time (microsecond–years events).

An added systems complexity results when an attempt to alter and/or manipulate biological organization is formalized such that it manifests a mutually reinforcing "Systems Medicine" architecture (Figure 9). It is an inevitable fact that the human condition (problem solving skills) and perception, coupled with the accumulated knowledge and information processing skills (see Figure 9) contribute to the effective complexity of "manmade" cascading systems, i.e., man–machine, man–man interaction. A case in point is the resultant feedback and feedforward control loops of a clinician interacting (sensing, measuring) with the ailing patient while testing and/or modifying systems performance (see Figure 9). The apparent regulatory interactions and compounding of systems effectors creates the conditions where the "sensor-driven therapy" increases both the robustness and fragility of the integrated system. The manifested dynamics of a high-gain system are both powerful (can be curative) and dangerous (medical errors).

6. EMERGENCE: COMPLEX ADAPTIVE SYSTEMS

Many systems in nature comprise a large number of autonomous parts (subsystems) that interacting locally, in the absence of a high-level global controller,

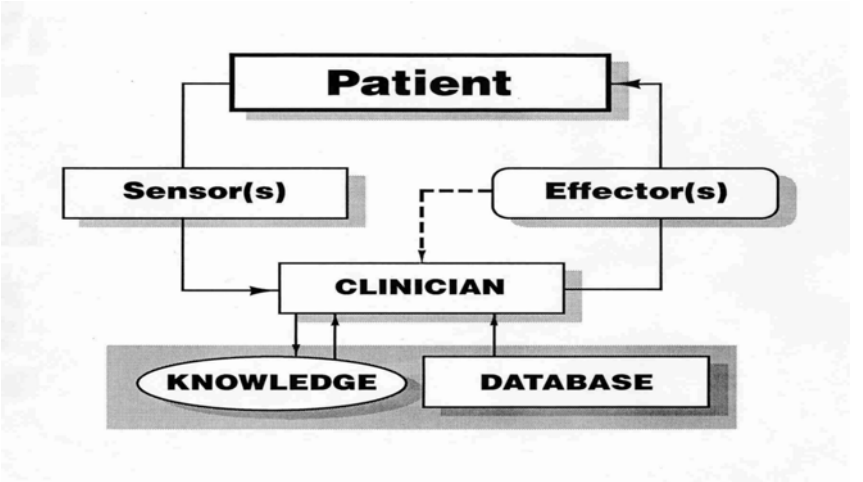


Figure 9. Systems-medicine conceptual framework: intertwined Man (Patient)–Machine (Sensors/Devices) and Patient–Clinician interaction and communication. The confluence of complexity and systems robustness gives rise to a mutually reinforcing state of fragility and risk.

and that can give rise to highly coordinated and optimized behavior. The complex adaptive behavior of global-level structures that emerges is a consequence of nonlinear spatiotemporal interactions of local-level processes or subsystems (see Figure 10). This form of nested cooptation (across levels of organization) is evident in isolated cells, organisms, societies, and ecologies. Systems of this type are nonlinear, nonstationary, nonequlibrium, and nonreductionist, and are governed by universal principles of adaptation and self-organization in which control and order are emergent rather than predetermined.

From a system-theory standpoint, a system that is endowed with a greater number of degrees of freedom is more robust and has a greater ability to accommodate imposed disturbances. In general, biological systems, independent of hierarchical organization (molecular to multicellular), normally operate such that a finite number of regulatory modes can be invoked. Chaotic systems are extremely susceptible to changes in initial conditions, i.e., small changes in a parameter of a chaotic system can produce a large change in the output, i.e., poised at the "edge of chaos" (18). This ability allows the system to switch quickly from one state to another. It may be that the chaotic regime enables a subsystem to exert its function such that regulatory changes can be achieved with minimal external input, reminiscent of self-organized criticality seen in other physical phenomena. From a standpoint of economy of performance (energy use, responsiveness), some upper limit must be set on the number of active degrees of freedom (control variables) that can or need be summoned. Most

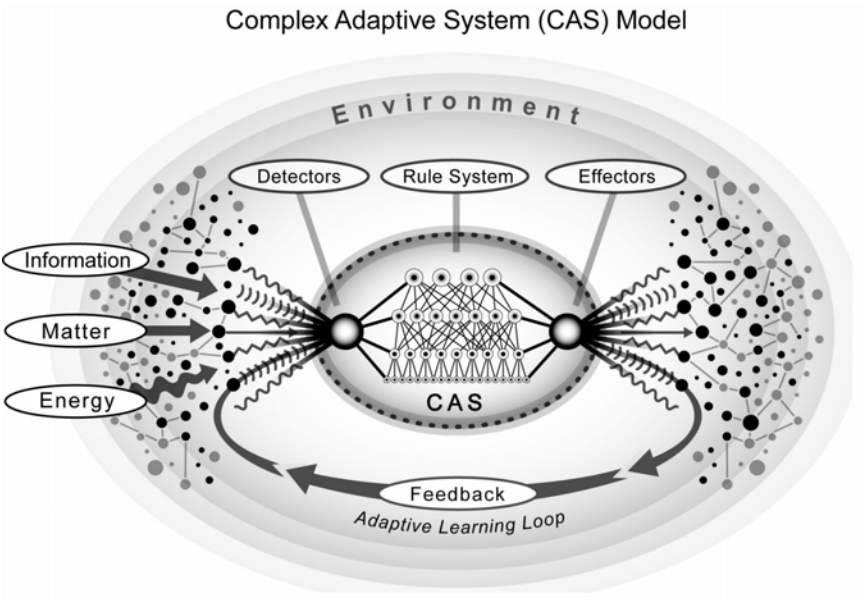


Figure 10. Complex Adaptive Systems (CAS) Schema. Many natural systems (e.g., immune, nervous system) are characterized by a behavior that is emergent as a result of often nonlinear spatiotemporal (complex) interactions among a large number of component "parts" at different levels of organization. The three principal inputs that drive the dynamics of the CAS are information, matter, and energy. These systems tend to be nested and open to the environment (have history); the parts cannot contain the whole; they are relational (short-range acting); nonlinear; contain adaptive feedback loops (anticipatory) with boundaries that are difficult to demarcate before. (Artwork by M. Clemens.)

physiological time-series data seem to be restricted in dimensional complexity to 3 to 6 degrees of freedom (see chapter 3.3 [by Glass], part III, this volume). The whole–part dualistic nature of subsystems (e.g., heart), comprised of multiple nested loops of nonlinear interacting regulators (homeostats), make them especially suitable/prone to chaotic behavior and thus amenable to finer/rapid control/adaptation (12).

6.1. The Living Organism as a Unit of Biological Organization

The questions that persist about biological functions and their interrelationships are:

- Does life get more complicated? i.e., *more effective functions are associated with more complex structures whereas diseases and death are the converse*.

- Are complex organisms better (more) adapted (adaptive) than simple organisms? i.e., is biology a relational science (not merely a complex system)-*throwing away the physics and keeping the organization; or as Robert Rosen noted "throwing away the polypeptide and keeping the active site*" (19).

It is noteworthy that living systems are not known to be structurally stable (e.g., during morphogenesis and evolution); they are not highly ordered, i.e., the order is not great but it is special; they are not program-driven or state-variable determined, i.e., they are execution-driven; they are not digital-computational, and equally surprising is that they are not far removed informationally from 1/f noise communication (6,14).

An intriguing example of a subsystem assimilation challenge arises when a new organ (heart, kidney, liver, lung, tissue, cells) is transplanted to a recipient host. The explanatory model of the newly emergent functional order attributable to graft–host interaction may benefit by evoking organizing principles of coevolution. The heart as an autonomous organ system is endowed with an adaptive plasticity (genotypic/phenotypic memory) and capacity to assimilate ("fitness capacity") within the host and in the process modify the environment, determining the fate of the body system as a whole. The principles by which emergent properties and the functional order of a self-organizing system, such as the heart, achieve homeodynamic stability provide a non-reductionist framework for understanding how biological systems adapt to the imposed internal and external stresses (e.g., organ/tissue replacement). The newly emergent dynamics arising after an organ (heart) transplantation may represent a more stable, versatile, and adaptive bipartite whole. The law of requisite variety, originally conceived by Ross Ashby (7), dictates that the variety in the control system must be equal to or larger than the variety of the perturbations in the local (body) environment, i.e., the larger the variety of actions available to a control system, the larger the variety of perturbations it is able to compensate/regulate.

The integrative collective behavior of living organisms cannot be reconstituted using the traditional reductionist approach, i.e., a mere assembly of parts or their subunits. The whole emanates/evolves relationally from the emergent internal requirements of the constitutive whole–part subsystem organization (3,11). A case in point is the dynamic rhythm patterns that evolve after cardiac transplantation. Over time, the generated patterns resulting from the reunification that takes place within the milieu of the recipient are not specified in the equations of motion (5,12). The behavior of living systems must be treated as a special case in which the description and representation must be inferred directly from the system in question, much like general complex systems. This, of course, implies that no identifiable single mathematical theory of complex systems exists that begs for completely new scientific methodologies and experi-

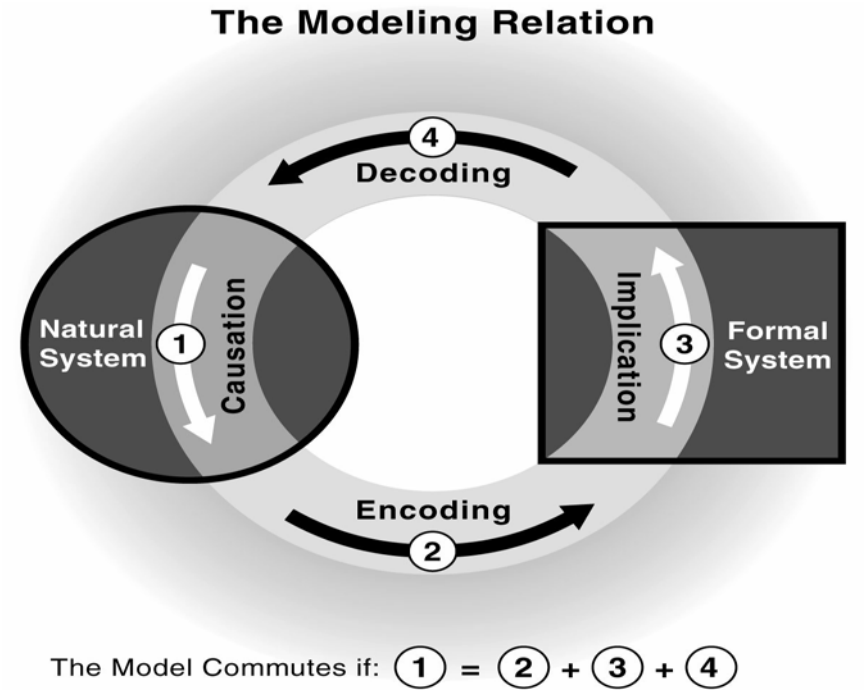


Figure 11. Modeling/Abstraction Formalism: methodology of applying systems theory in biology, i.e., construction of a system and the constructive specifications (conceptual or otherwise). A model only exists when the "modeling relation" commutes (1 = 2 + 3 + 4). The modeling relation involves a "Natural System" undergoing causal events, in the form of a "Formal System" undergoing, implication while connected by an encoding and decoding step. (Artwork by M. Clemens.)

mental epistemologies, a new guide for thinking with emphasis placed on its relativity or situation-dependence approach. Specifically, Robert Rosen's modeling (19) relation (see Figure 11) is a tool for thinking about the relationship between natural systems and structures created (scientific process) for understanding such systems. Simple systems entail a largest model representation. Conversely, complex systems are not bounded and therefore no largest syntactic model can be found. The mechanistic metaphor simply will not do.

There is compelling evidence that malignant tumors behave as self-organizing networks with properties of adaptive multicellular biosystems rather than as unorganized cell masses (see chapter 6.3, by Mansury and Deisboeck, part III, this volume). The emergence of networks with complex topology as diverse as the cell or the Internet that are driven by similar self-organizing processes suggests that they are governed by simple but generic laws. Analysis of the metabolic and protein networks of various organisms shows that cells and com-

plex manmade networks (i.e., World Wide Web) share the same large-scale topology (see Part II, chapter 4 [by Wuchty, Ravasz, and Barabási], this volume). The emergence of networks with complex topology as diverse as the cell or the Internet that are driven by similar self-organizing processes suggests that they are governed by simple but generic laws. Uncovering these organizing principles and the role they play in living processes is one of the major goals of postgenomic biology.

In short, von Bertalanffy gave us a new paradigm for transdisciplinary thinking and synthesis. Considering the fact that many of our practical problems have to do with systemic phenomena (development, preservation, function), a means for system thinking is an essential component of the integration of scientific knowledge. In that respect, systems theory is not meant to be another finished, unified theory to be verified or falsified and fitted either into the spectrum of valid scientific data or placed on the shelves of the history of science and left to the memory of time.

Perhaps this fear is unfounded, as demonstrated by the opening remarks made at the 3rd International Conference on Systems Biology held at the Karolinska Institute in 2003 (Måns Ehrenberg, chair):

> The Human Genome Project and recent advances in proteomics and DNA microarray technology highlight the need for systems-level integration of experiments and theory in order to understand the logic of life. This is the ambitious goal for systems biology, the quantitative study of biological processes as integrated systems rather than as isolated parts (http://www.genome.org/cgi/doi/10.1101/gr.1765703). What is especially needed is a coherent picture of how this information is being used to carry out biological functions.

Parenthetically, the resurgence of a systems approach to biology has been largely based on the premise, as articulated by Leroy Hood, that systems biology must be able to capture the digital informational content of the genes and integrate them together into networks so that we may begin to understand the logic of life.

7. THE COMPLEX SYSTEMS IN SYSTEMS BIOLOGY

It would seem that we are back at the beginning, but this time with an earnest desire to reassemble the complex collection of molecular-cellular pathways and networks in order to gain an understanding (i.e., prediction, control, design) of biology at progressively higher levels of organization. This integrative approach to systems biology, encompassing genomics, transcriptomics, pro-

teomics, and metabolomics is now possible because of advances in high-throughput technologies (e.g., microarrays, microfluidic devices) and a general acceptance of computational and mathematical models by the biology community at large. The ability to analyze and simulate pathways, networks, and the spatial and temporal relationships between genes, transcripts, proteins, metabolites, and cells is an important step in the attempt to attribute cause and effect in living systems (see chapters 1.1 [by Pedraza and Oudenaarden], 1.3 [by Wagner], and 1.4 [by Dhar and Tomita], this volume). A quantitative understanding of entire subcellular, cellular, or multicellular systems could significantly alter the approach taken to personalized medicine and drug discovery. Clearly, the motivation to pursue the analysis of molecular pathways, networks, and regulation of how the cell works as a whole system has been with us for some time. What has changed is the recognition that the parts list (i.e., human genome) or its sequence is a necessary but not sufficient ingredient for the synthesis. Systems biology is not an entirely new disciplinary domain of knowledge. In many ways, it is the recasting of cell physiology with a DNA-twist, i.e., a desire to dissect the emergent relational mechanisms that arise when traversing from the molecular to the systems level. The harnessing of interdisciplinary talents such as mathematics, computer science, and engineering is needed to develop the theoretical framework for complex systems biology (and medicine) problems.

7.1. The Last Word on the Science of Organized Complexity

Much of science in the past century has been concerned with analysis of parts; relatively little emphases has been placed on the synthesis and integration of concepts that constitute the logic of life.

Each chapter contribution to follow, by design, reflects the system view of biology or medicine. The chosen style and manor of system inquiry construction contribute and frame a scheme of ideas that are unflinchingly imaginative and connected without the burden of sustaining a unitary theory. The coalesced book is a system itself; as such, it has four interrelated and internally consistent aspects acting as a whole: systems philosophy, systems theory, systems methodology (e.g., nanotechnology and microfluidics), and systems application.

> *Any theory of wider scope implies a worldview ... any major development in science changes the world outlook and its "natural philosophy" ...*
>
> —Ludwig von Bertalanffy, *Robots, Men and Minds*

8. ACKNOWLEDGMENTS

This work (and thought experiments) has been generously supported over the intervening years by the National Institutes of Health, the American Heart Association, the Defense Advanced Research Projects Agency (DOD), and the State of Pennsylvania. I am grateful to have had the exposure and close collaboration with academic faculty spanning the spectrum of engineering and biomedical sciences. I am especially indebted to Dr. Andrew Wechsler, Professor and Chair of Cardiothoracic Surgery, whose unwavering dedication to scholarship made it possible to take time away from assigned duties and think outside the cardiac and/or surgical box.

9. REFERENCES

1. von Bertalanffy L. 1968. *General system theory*. George Braziller, New York.
2. Heylighen F. 2004. Principia Cybernetica Web (http://pespmc1.vub.ac.be) Center Leo Apostel, Vrije Universiteit Brussel, Krijgskundestraat 33, B-1160 Brussels, Belgium.
3. Laszlo E, ed. 1972. *The relevance of general systems theory*. George Braziller, New York.
4. Koestler A, Smythes JR, eds. 1968. *Beyond reductionism*. Hutchinson, London.
5. Kresh JY, Armour JA. 1997. The heart as a self-regulating system: integration of homeodynamic mechanisms. *Technol Health Care* **5**(1):159–169.
6. Yates E, ed. 1987. *Self-organizing systems: the emergence of order*. Plenum, New York
7. Ashby R. 1956. *An introduction to cybernetics*. Methuen, New York.
8. Wiener N. 1948. *Cybernetics or control and communication in the animal and the machine*. Hermann Editions in Paris. MIT Press, Cambridge (2nd ed. 1961, Wiley & Sons, New York).
9. McCulloch WS. 1965. *Embodiments of mind*. MIT Press, Cambridge.
10. Foerster H v, Zopf G., eds. 1962. *Principles of self-organization*. Pergamon, New York
11. Weinberg GM. 1975. *An introduction to general systems thinking*. Wiley & Sons New York.
12. Kresh JY, Izrailtyan I. 1998. Evolution in functional complexity of heart rate dynamics: a measure of cardiac allograft adaptability. *Am J Physiol* **275**:R720–R727.
13. Simon HA. 1969. *The sciences of the artificial*. MIT Press, Cambridge.
14. Yates FE. 1993. Self-organizing systems. In *The logic of life: the challenge of integrative physiology*. Ed. CAR Boyd, D Noble, pp. 189–218. Oxford UP, New York.
15. Noble D. 2000. Modelling the heart: insights, failures and progress. *Bioessays* **24**:1155–1163.
16. Wolkenhauer O. 2001. Systems biology: the reincarnation of systems theory applied in biology? *Briefings in Bioinformatics* **2**(3):258–270.
17. Mesarovic MD, ed. 1968. *System theory and biology*. Springer, New York (see chapter on "System theory and biology: view of a theoretician," pp. 59–87).
18. Kauffman SA. 1993. *The origin of order: self-organization and selection in evolution*. Oxford UP, New York.
19. Rosen R. 1999. *Life itself*. Columbia UP, New York.

Selected Reading

Shannon CE, Weaver W. 1964. *Mathematical theory of communication*. U Illinois P, Urbana.
Foerster H v, ed. 1974. *The cybernetics of cybernetics*. Biological Computer Laboratory Report 73.38, University of Illinois at Urbana-Champaign.

Hood L, Ideker T, Galitski T. 2001. A new approach to decoding life: systems biology. In *Annual review of genomics and human genetics 2001*, Vol. 2, pp. 343–372. Annual Reviews, Palo Alto.

Kitano H. 2001. *Foundations of systems biology*. MIT Press, Cambridge.

von Bertalanffy L. 1972. *Perspectives on general system theory*. Ed. E Taschdjian, with forewords by M v Bertalanffy, E Laszlo. George Braziller, New York.

Leibler S, Elowitz, MB. 2000. A synthetic oscillatory network of transcriptional regulators. *Nature* 403:335–338.

Lauffenburger DA. 2000. Cell signaling pathways as control modules: complexity for simplicity? *Proc Natl Acad Sci USA* **97**(10):5031–5033.

Doyle JC. 2000. Robustness and dynamics in biological networks. In *The first international conference on systems biology*. MIT Press, New York.

Prigogine I. 1980. *From being to becoming*. Freeman, San Francisco.

Gell-Mann M. 1994. *The quark and the jaguar, adventures in the simple and the complex*. Freeman, New York.

Carpa F. 1997. *The web of life: a new scientific understanding of living systems*. Random House, New York.

Human Genome Organization. *Human genome project* (HGP). http://www.ornl.gov/hgmis/.

Part II

COMPLEX SYSTEMS SCIENCE: THE BASICS

1

METHODS AND TECHNIQUES OF COMPLEX SYSTEMS SCIENCE: AN OVERVIEW

Cosma Rohilla Shalizi

Center for the Study of Complex Systems,
University of Michigan, Ann Arbor

In this chapter, I review the main methods and techniques of complex systems science. As a first step, I distinguish among the broad patterns which recur across complex systems, the topics complex systems science commonly studies, the tools employed, and the foundational science of complex systems. The focus of this chapter is overwhelmingly on the third heading, that of tools. These in turn divide, roughly, into tools for analyzing data, tools for constructing and evaluating models, and tools for measuring complexity. I discuss the principles of statistical learning and model selection; time series analysis; cellular automata; agent-based models; the evaluation of complex-systems models; information theory; and ways of measuring complexity. Throughout, I give only rough outlines of techniques, so that readers, confronted with new problems, will have a sense of which ones might be suitable, and which ones definitely are not.

1. INTRODUCTION

A complex system, roughly speaking, is one with many parts, whose behaviors are both highly variable and strongly dependent on the behavior of the other parts. Clearly, this includes a large fraction of the universe! Nonetheless, it is not vacuously all-embracing: it excludes both systems whose parts just cannot do very much, and those whose parts are really independent of each other. "Complex systems science" is the field whose ambition is to understand complex systems. Of course, this is a broad endeavor, overlapping with many even larger,

Address correspondence to: Prof. Cosma Rohilla Shalizi, Statistics Department, Carnegie Mellon University, Pittsburgh, PA 15213 (cahalizi@stat.cmu.edu).

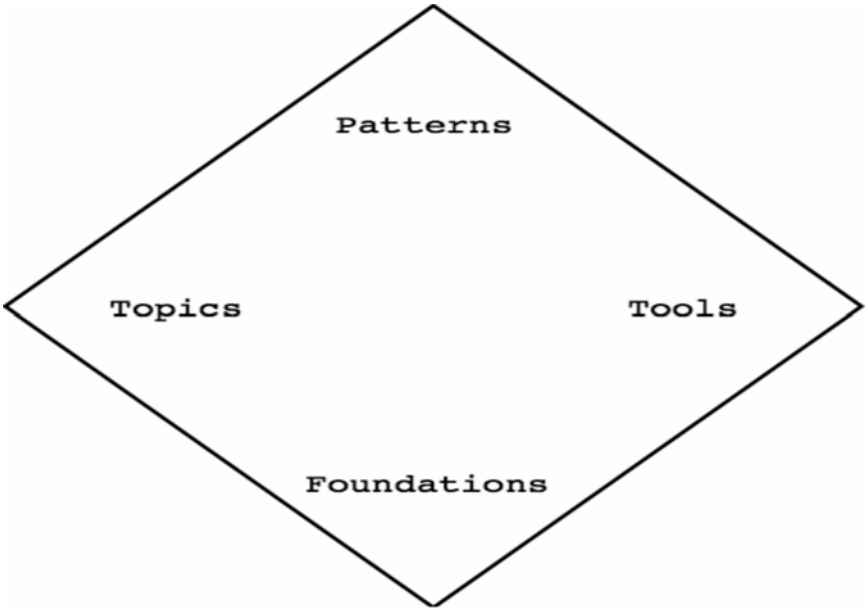


Figure 1. The quadrangle of complex systems. See text.

better-established scientific fields. Having been asked by the editors to describe its methods and techniques, I begin by explaining what I feel does *not* fall within my charge, as indicated by Figure 1.

At the top of Figure 1 I have put "patterns." By this I mean more or less what people in software engineering do (1): a pattern is a recurring theme in the analysis of many different systems, a cross-systemic regularity. For instance, bacterial chemotaxis can be thought of as a way of resolving the tension between the exploitation of known resources, and costly exploration for new, potentially more valuable, resources (Figure 2). This same tension is present in a vast range of adaptive systems. Whether the exploration–exploitation tradeoff arises among artificial agents, human decision-makers or colonial organisms, many of the issues are the same as in chemotaxis, and solutions and methods of investigation that apply in one case can profitably be tried in another (2,3). The pattern "tradeoff between exploitation and exploration" thus serves to orient us to broad features of novel situations. There are many other such patterns in complex systems science: "stability through hierarchically structured interactions" (4), "positive feedback leading to highly skewed outcomes" (5), "local inhibition and long-rate activation create spatial patterns" (6), and so forth.

At the bottom of the quadrangle is "foundations," meaning attempts to build a basic, mathematical science concerned with such topics as the measurement of

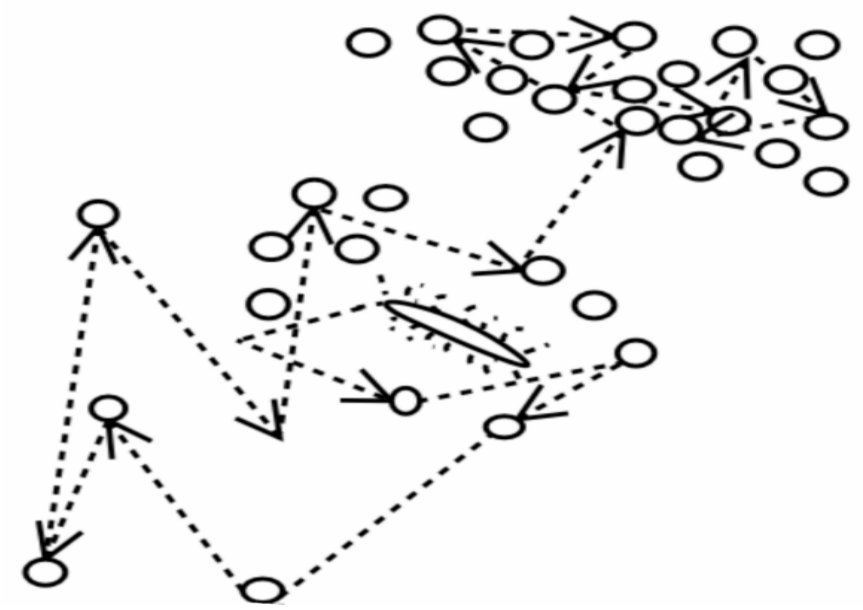


Figure 2. Bacterial chemotaxis. Should the bacterium (center) exploit the currently available patch of food, or explore, in hopes of finding richer patches elsewhere (e.g., at right)? Many species solve this problem by performing a random walk (jagged line), tumbling randomly every so often. The frequency of tumbling increases when the concentration of nutrients is high, making the bacterium take long steps in resource-poor regions, and persist in resource-rich ones (7–9).

complexity (10), the nature of organization (11), the relationship between physical processes and information and computation (12), and the origins of complexity in nature and its increase (or decrease) over time. There is dispute whether such a science is possible, and if so whether it would be profitable. I think it is both possible and useful, but most of what has been done in this area is very far from being applicable to *biomedical* research. Accordingly, I shall pass it over, with the exception of a brief discussion of some work on measuring complexity and organization that is especially closely tied to data analysis.

"Topics" go in the left-hand corner. Here are what one might call the "canonical complex systems," the particular systems, natural, artificial and fictional, which complex systems science has traditionally and habitually sought to understand. Here we find networks (see Part II, chapter 4, by Wuchty, Ravasz, and Barabási, this volume), turbulence (13), physicochemical pattern formation and biological morphogenesis (14,15), genetic algorithms (16,17), evolutionary dynamics (18,19), spin glasses (20,21), neuronal networks (see Part III, section 5, this volume), the immune system (see Part III, section 4, this volume), social insects, ant-like robotic systems, the evolution of cooperation, evolutionary eco-

nomics, etc.[1] These topics all fall within our initial definition of "complexity," though whether they are studied together because of deep connections, or because of historical accidents and tradition, is a difficult question. In any event, this chapter will *not* describe the facts and particular models relevant to these topics.

Instead, this chapter is about the right-hand corner, "tools." Some are procedures for analyzing data, some are for constructing and evaluating models, and some are for measuring the complexity of data or models. In this chapter I will restrict myself to methods which are generally accepted as valid (if not always widely applied), *and* seem promising for biomedical research. These still demand a book, if not an encyclopedia, rather than a mere chapter! Accordingly, I will merely try to convey the essentials of the methods, with pointers to references for details. The goal is for you to have a sense of which methods would be good things to try on your problem, rather than to tell you everything you need to know to implement them.

1.1. Outline of This Chapter

As mentioned above, the techniques of complex systems science can, for our purposes, be divided into three parts: those for analyzing data (perhaps without reference to a particular model), those for building and understanding models (often without data), and those for measuring complexity as such. This chapter will examine them in that order.

The first part, on **data**, opens with the general ideas of **statistical learning and data mining** (§2), namely developments in statistics and machine learning theory that extend statistical methods beyond their traditional domain of low-dimensional, independent data. We then turn to **time series analysis** (§3), where there are two important streams of work, inspired by statistics and nonlinear dynamics.

The second part, on **modeling**, considers the most important and distinctive classes of models in complex systems. On the vital area of **nonlinear dynamics**, let the reader consult Socolar (Part II, chapter 2, this volume). **Cellular automata** (§4) allow us to represent spatial dynamics in a way that is particularly suited to capturing strong local interactions, spatial heterogeneity, and large-scale aggregate patterns. Complementary to cellular automata are **agent-based models** (§5), perhaps the most distinctive and most famous kind of model in complex systems science. A general section (§6) on **evaluating complex models**, including analytical methods, various sorts of simulation, and testing, closes this part of the chapter.

The third part of the chapter considers ways of measuring complexity. As a necessary preliminary, §7 introduces the concepts of **information theory**, with some remarks on its application to biological systems. Then §8 treats **complex-**

ity measures, describing the main kinds of complexity measure, their relationships, and their applicability to empirical questions.

The chapter ends with a guide to further reading, organized by section. These emphasize readable and thorough introductions and surveys over more advanced or historically important contributions.

2. STATISTICAL LEARNING AND DATA-MINING

Complex systems, we said, are those with many strongly interdependent parts. Thanks to comparatively recent developments in statistics and machine learning, it is now possible to infer reliable, predictive models from data, even when the data concern thousands of strongly dependent variables. Such **data mining** is now a routine part of many industries, and is increasingly important in research. While not, of course, a substitute for devising valid theoretical models, data mining *can* tell us what kinds of patterns are in the data, and so guide our model-building.

2.1. Prediction and Model Selection

The basic goal of any kind of data mining is prediction: some variables, let us call them X, are our inputs. The output is another variable or variables Y. We wish to use X to predict Y, or, more exactly, we wish to build a machine which will do the prediction for us: we will put in X at one end, and get a prediction for Y out at the other.[2]

"Prediction" here covers a lot of ground. If Y are simply other variables like X, we sometimes call the problem **regression**. If they are X at another time, we have **forecasting**, or prediction in a strict sense of the word. If Y indicates membership in some set of discrete categories, we have **classification**. Similarly, our predictions for Y can take the form of distinct, particular values (**point predictions**), of ranges or intervals we believe Y will fall into, or of entire probability distributions for Y, i.e., guesses as to the conditional distribution $\Pr(Y|X)$. One can get a point prediction from a distribution by finding its mean or mode, so distribution predictions are in a sense more complete, but they are also more computationally expensive to make, and harder to make successfully.

Whatever kind of prediction problem we are attempting, and with whatever kind of guesses we want our machine to make, we must be able to say whether or not they are good guesses; in fact we must be able to say just how much bad guesses cost us. That is, we need a **loss function** for predictions.[3] We suppose that our machine has a number of knobs and dials we can adjust, and we refer to these parameters, collectively, as θ. The predictions we make, with inputs X and parameters θ, are $f(X,\theta)$, and the loss from the error in these predictions, when

the actual outputs are Y, is $L(Y,f(X,\theta))$. Given *particular* values y and x, we have the empirical loss $L(y,f(x,\theta))$, or $\hat{L}(\theta)$ for short.[4]

Now, a natural impulse at this point is to twist the knobs to make the loss small: i.e., to select the θ that minimizes $\hat{L}(\theta)$; let's write this as follows: $\hat{\theta} = \text{argmin}_{\theta} \hat{L}(\theta)$. This procedure is sometimes called **empirical risk minimization**, or ERM. (Of course, doing that minimization can itself be a tricky nonlinear problem, but I will not cover optimization methods here.) The problem with ERM is that the $\hat{\theta}$ we get from *this* data will almost surely not be the same as the one we'd get from the *next* set of data. What we really care about, if we think it through, is not the error on any particular set of data, but the error we can *expect* on new data, $\mathbf{E}[L(\theta)]$. The former, $\hat{L}(\theta)$, is called the **training** or **in-sample** or **empirical** error; the latter, $\mathbf{E}[L(\theta)]$, the **generalization** or **out-of-sample** or **true** error. The difference between in-sample and out-of-sample errors is due to sampling noise, the fact that our data are not *perfectly* representative of the system we're studying. There will be quirks in our data which are just due to chance, but if we minimize $\hat{L}$ blindly, if we try to reproduce every feature of the data, we will be making a machine that reproduces the random quirks, which do not generalize, along with the predictive features. Think of the empirical error $\hat{L}(\theta)$ as the generalization error, $\mathbf{E}[L(\theta)]$, plus a sampling fluctuation, ε. If we look at machines with low empirical errors, we will pick out ones with low true errors, which is good, but we will also pick out ones with large negative sampling fluctuations, which is not good. Even if the sampling noise ε is very small, $\hat{\theta}$ can be very different from θ_{min}. We have what optimization theory calls an **ill-posed problem** (22).

Having a higher-than-optimal generalization error because we paid too much attention to our data is called **over-fitting**. Just as we are often better off if we tactfully ignore our friends' and neighbors' little faults, we want to ignore the unrepresentative blemishes of our sample. Much of the theory of data mining is about avoiding over-fitting. Three of the commonest forms of tact it has developed are, in order of sophistication, **cross-validation**, **regularization** (or **bold penalties**) and **capacity control**.

2.1.1. Validation

We would never over-fit if we *knew* how well our machine's predictions would generalize to new data. Since our data is never perfectly representative, we always have to estimate the generalization performance. The empirical error provides one estimate, but it's biased towards saying that the machine will do well (since we built it to do well on that data). If we had a second, independent set of data, we could evaluate our machine's predictions on it, and that would give us an unbiased estimate of its generalization. One way to do this is to take our original data and divide it, at random, into two parts, the **training set** and the

test set or **validation** set. We then use the training set to fit the machine, and evaluate its performance on the test set. (This is an instance of **resampling** our data, which is a useful trick in many contexts.) Because we've made sure the test set is independent of the training set, we get an unbiased estimate of the out-of-sample performance.

In **cross-validation**, we divide our data into random training and test sets many different ways, fit a different machine for each training set, and compare their performances on their test sets, taking the one with the best test-set performance. This reintroduces some bias—it could happen by chance that one test set reproduces the sampling quirks of its training set, favoring the model fit to the latter. But cross-validation generally *reduces* over-fitting, compared to simply minimizing the empirical error; it makes more *efficient* use of the data, though it cannot get rid of sampling noise altogether.

2.1.2. Regularization or Penalization

I said that the problem of minimizing the error is **ill-posed**, meaning that small changes in the errors can lead to big changes in the optimal parameters. A standard approach to ill-posed problems in optimization theory is called **regularization**. Rather than trying to minimize $\hat{L}(\theta)$ alone, we minimize

$$\hat{L}(\theta) + \lambda d(\theta), \qquad [1]$$

where $d(\theta)$ is a **regularizing** or **penalty** function. Remember that $\hat{L}(\theta) = \mathbf{E}[L(\theta)] + \varepsilon$, where ε is the sampling noise. If the penalty term is well-designed, then the θ which minimizes

$$\mathbf{E}[L(\theta)] + \varepsilon + \lambda d(\theta) \qquad [2]$$

will be close to the θ that minimizes $\mathbf{E}[L(\theta)]$—it will cancel out the effects of favorable fluctuations. As we acquire more and more data, $\varepsilon \to 0$, so λ, too, goes to zero at an appropriate pace, and the penalized solution will converge on the machine with the best possible generalization error.

How then should we design penalty functions? The more knobs and dials there are on our machine, the more opportunities we have to get into mischief by matching chance quirks in the data. If one machine has fifty knobs and another fits the data just as well but has only a single knob, we should (the story goes) chose the latter—because it's *less* flexible the fact that it does well is a good indication that it will still do well in the future. There are thus many regularization methods that add a penalty proportional to the number of knobs, or, more formally, the number of parameters. These include the Akaike information criterion or AIC (23) and the Bayesian information criterion or BIC (24,25). Other

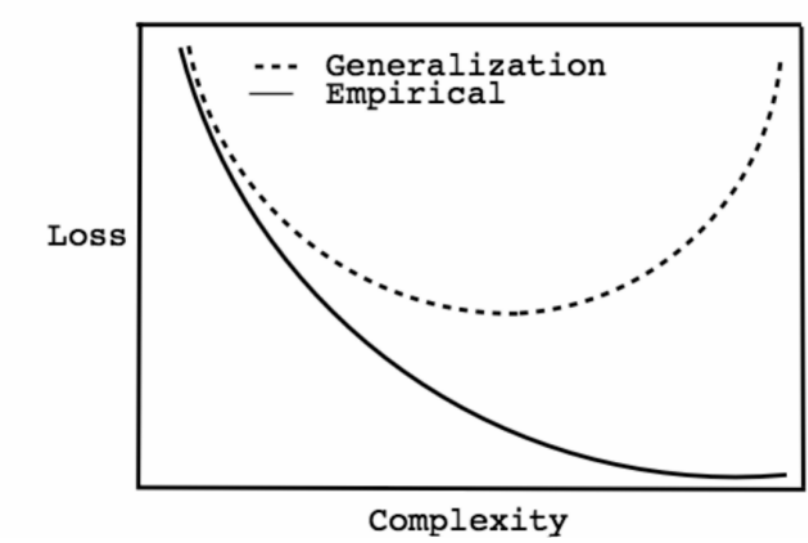


Figure 3. Empirical loss and generalization loss as a function of model complexity.

methods penalize the "roughness" of a model, i.e., some measure of how much the prediction shifts with a small change in either the input or the parameters (26, ch. 10). A smooth function is less flexible, and so has less ability to match meaningless wiggles in the data. Another popular penalty method, the **minimum description length** principle of Rissanen, will be dealt with in §8.3 below.

Usually, regularization methods are justified by the idea that models can be more or less complex, and more complex ones are more liable to over-fit, all else being equal, so penalty terms should reflect complexity (Figure 3). There's something to this idea, but the usual way of putting it does not really work; see §2.3 below.

2.1.3. Capacity Control

Empirical risk minimization, we said, is apt to over-fit because we do not know the generalization errors, just the empirical errors. This would not be such a problem if we could *guarantee* that the in-sample performance was close to the out-of-sample performance. Even if the exact machine we got this way was not particularly close to the optimal machine, we'd then be guaranteed that our *predictions* were nearly optimal. We do not even need to guarantee that *all* the

empirical errors are close to their true values, just that the *smallest* empirical error is close to the smallest generalization error.

Recall that $\hat{L}(\theta) = \mathbf{E}[L(\theta)] + \varepsilon$. It is natural to assume that as our sample size N becomes larger, our sampling error ε will approach zero. (We will return to this assumption below.) Suppose we could find a function $\eta(N)$ to bound our sampling error, such that $|\varepsilon| \leq \eta(N)$. Then we could guarantee that our choice of model was **approximately correct**; if we wanted to be sure that our prediction errors were within ε of the best possible, we would merely need to have $N(\varepsilon) = \eta^{-1}(\varepsilon)$ data-points.

It should not be surprising to learn that we cannot, generally, make approximately correct guarantees. As the eminent forensic statistician C. Chan remarked, "Improbable events permit themselves the luxury of occurring" (27), and one of these indulgences could make the discrepancy between $\hat{L}(\theta)$ and $\mathbf{E}[L(\theta)]$ very large. But if something like the law of large numbers holds, or the ergodic theorem (§3.2), then for every choice of θ,

$$\Pr(|\hat{L}(\theta) - \mathbf{E}[L(\theta)]| > \varepsilon) \to 0, \quad [3]$$

for every positive ε.[5] We should be able to find some function δ such that

$$\Pr(|\hat{L}(\theta) - \mathbf{E}[L(\theta)]| > \varepsilon) \leq \delta(N, \varepsilon, \theta), \quad [4]$$

with $\lim_N \delta(N,\varepsilon,\theta) = 0$. Then, for any particular θ, we could give **probably approximately correct** (28) guarantees, and say that, e.g., to have 95% confidence that the true error is within 0.001 of the empirical error requires at least 144,000 samples (or whatever the precise numbers may be). If we can give probably approximately correct (PAC) guarantees on the performance of one machine, we can give them for any *finite* collection of machines. But if we have infinitely many possible machines, might not there always be *some* of them which are misbehaving? Can we still give PAC guarantees when θ is continuous?

The answer to this question depends on how flexible the set of machines is—its **capacity**. We need to know how easy it is to find a θ such that $f(X,\theta)$ will accommodate itself to any Y. This is measured by a quantity called the Vapnik-Chervonenkis (VC) dimension (22).[6] If the VC dimension d of a class of machines is finite, one can make a PAC guarantee that applies to *all* machines in the class simultaneously:

$$\Pr\left(\max_\theta |\hat{L}(\theta) - \mathbf{E}[L(\theta)]| \geq \eta(N, d, \delta)\right) \leq \delta, \quad [5]$$

where the function $\eta(N,d,\delta)$ expresses the rate of convergence. It depends on the particular kind of loss function involved. For example, for binary classification, if the loss function is the fraction of inputs misclassified,

$$\eta(N,d,\delta) = \frac{1}{\sqrt{N}}\left(\sqrt{d(1+\ln\frac{2N}{d}) + \ln\frac{4}{\delta}}\right). \quad [6]$$

Notice that θ is not an argument to η, and does not appear in [6]. The rate of convergence is the same across all machines; this kind of result is thus called a **uniform law of large numbers**. The really remarkable thing about [5] is that it holds no matter what the sampling distribution is, so long as samples are independent; it is a **distribution-free** result.

The VC bounds lead to a very nice learning scheme: simply apply empirical risk minimization, for a fixed class of machines, and then give a PAC guarantee that the one picked is, with high reliability, very close to the actual optimal machine. The VC bounds also lead an appealing penalization scheme, where the penalty is equal to our bound on the over-fitting, η. Specifically, we set the term $\lambda d(\theta)$ in [1] equal to the η in [5], ensuring, with high probability, that the ε and $\lambda d(\theta)$ terms in [2] cancel each other. This is **structural risk minimization** (SRM).

It's important to realize that the VC dimension is not the same as the number of parameters. For some classes of functions, it is much *lower* than the number of parameters, and for others it's much *higher*. (There are examples of one-parameter classes of functions with infinite VC dimension.) Determining the VC dimension often involves subtle combinatorial arguments, but many results are now available in the literature, and more are appearing all the time. There are even schemes for experimentally estimating the VC dimension (29).

Two caveats are in order. First, because the VC bounds are distribution-free, they are really about the rate of convergence under the worst possible distribution, the one a malicious adversary out to foil our data mining would choose. This means that in practice, convergence is often much faster than [5] would indicate. Second, the usual proofs of the VC bounds all assume independent, identically distributed samples, though the relationship between X and Y can involve arbitrarily complicated dependencies.[7] Recently, there has been much progress in proving uniform laws of large numbers for dependent sequences of samples, and structural risk minimization has been extended to what are called "mixing" processes (30), in effect including an extra term in the η function appearing in [5] that discounts the number of observations by their degree of mutual dependence.

2.2. Choice of Architecture

The basic idea of data mining is to fit a model to data with minimal assumptions about what the correct model should be, or how the variables in the data are related. (This differs from such classical statistical questions as testing

specific hypotheses about specific models, such as the presence of interactions between certain variables.) This is facilitated by the development of extremely flexible classes of models, which are sometimes, misleadingly, called **non-parametric**; a better name would be **megaparametric**. The idea behind mega-parametric models is that they should be capable of approximating any function, at least any well-behaved function, to any desired accuracy, given enough capacity.

The polynomials are a familiar example of a class of functions which can perform such universal approximation. Given any smooth function f, we can represent it by taking the Taylor series around our favorite point x_0. Truncating that series gives an approximation to f:

$$f(x) = f(x_0) + \sum_{k=1}^{\infty} \frac{(x - x_0)^k}{k!} \left. \frac{d^k f}{dx^k} \right|_{x_0} \quad [7]$$

$$\approx f(x_0) + \sum_{k=1}^{n} \frac{(x - x_0)^k}{k!} \left. \frac{d^k f}{dx^k} \right|_{x_0} \quad [8]$$

$$= \sum_{k=0}^{n} a_k \frac{(x - x_0)^k}{k!}. \quad [9]$$

In fact, if f is an nth-order polynomial, the truncated series is exact, not an approximation.

To see why this is *not* a reason to use only polynomial models, think about what would happen if $f(x) = \sin x$. We would need an *infinite*-order polynomial to completely represent f, and the generalization properties of finite-order approximations would generally be lousy: for one thing, f is bounded between –1 and 1 everywhere, but any finite-order polynomial will start to zoom off to ∞ or $-\infty$ outside some range. Of course, this f would be really easy to approximate as a superposition of sines and cosines, which is another class of functions which is capable of universal approximation (better known, perhaps, as Fourier analysis). What one wants, naturally, is to choose a model class which gives a good approximation of the function at hand, *at low order*. We want low-order functions, both because computational demands rise with model order *and* because higher-order models are more prone to over-fitting (VC dimension generally rises with model order).

To adequately describe all of the *common* model classes, or **model architectures**, used in the data mining literature would require another chapter ((31) and (32) are good for this.) Instead, I will merely name a few.

Splines are piecewise polynomials, good for regression on bounded domains; there is a very elegant theory for their estimation (33).

Neural networks or **multilayer perceptrons** have a devoted following, both for regression and classification (32). The application of VC theory to them is quite well-advanced (34,35), but there are many other approaches, including ones based on statistical mechanics (36). It is notoriously hard to understand *why* they make the predictions they do.

Classification and regression trees (CART), introduced in the book of that name (37), recursively subdivide the input space, rather like the game of "twenty questions" ("Is the temperature above 20 centigrade? If so, is the glucose concentration above one millimole?," etc.); each question is a branch of the tree. All the cases at the end of one branch of the tree are treated equivalently. The resulting decision trees are easy to understand, and often similar to human decision heuristics (38).

Kernel machines (22,39) apply nonlinear transformations to the input, mapping it to a much higher dimensional "feature space," where they apply linear prediction methods. This trick works because the VC dimension of linear methods is low, even in high-dimensional spaces. Kernel methods come in many flavors, of which the most popular, currently, are **support vector machines** (40).

2.2.1. Predictive Versus Causal Models

Predictive and descriptive models both are not necessarily causal. PAC-type results give us reliable prediction, *assuming* future data will come from the *same* distribution as the past. In a causal model, however, we want to know how *changes* will propagate through the system. One difficulty is that these relationships are one-way, whereas prediction is two-way (one can predict genetic variants from metabolic rates, but one cannot change genes by changing metabolism). The other is that it is hard (if not impossible) to tell if the predictive relationships we have found are **confounded** by the influence of other variables and other relationships we have neglected. Despite these difficulties, the subject of **causal inference** from data is currently a very active area of research, and many methods have been proposed, generally under assumptions about the absence of feedback (41–43). When we have a causal or generative model, we can use very well-established techniques to infer the values of the hidden or latent variables in the model from the values of their observed effects (41,44).

2.3. Occam's Razor and Complexity in Prediction

Often, regularization methods are thought to be penalizing the *complexity* of the model, and so implementing some version of Occam's Razor. Just as Occam said "entities are not to be multiplied beyond necessity,"[8] we say "parameters

should not be multiplied beyond necessity," or, "the model should be no rougher than necessary." This takes complexity to be a property of an *individual* model, and the hope is that a simple model that can predict the training data will also be able to predict new data. Under many circumstances, one can prove that as the size of a sample approaches infinity regularization will converge on the correct model, the one with the best generalization performance (26). But one can often prove exactly the same thing about ERM without any regularization or penalization at all; this is what the VC bounds [5] accomplish. While regularization methods often do well in practice, so, too, does straight ERM. If we compare the performance of regularization methods to straight empirical error minimization on artificial examples, where we can calculate the generalization performance exactly, regularization sometimes conveys *no clear advantage at all* (45).

Contrast this with what happens in structural risk minimization. There our complexity penalty depends solely on the VC dimension of the *class* of models we're using. A simple, inflexible model which we find only because we're looking at a complex, flexible class is penalized just as much as the most wiggly member of that class. Experimentally, SRM *does* work better than simple ERM, or than traditional penalization methods.

A simple example may help illuminate why this is so. Suppose we're interested in binary classification, and we find a machine θ that correctly classifies a million independent data points. If the real error rate (= generalization error) for θ was one in a hundred thousand, the chance that it would correctly classify a million data points would be $(0.99999)^{10^6} \approx 4.5 \cdot 10^{-5}$. If θ was the very first parameter setting we checked, we could be quite confident that its true error rate was much less than 10^{-5}, no matter how complicated the function $f(X,\theta)$ looked. But if we've looked at ten million parameter settings before finding θ, then the odds are quite good that, among the machines with an error rate of 10^{-5}, we'd find several that correctly classify all the points in the training set, so the fact that θ does is not good evidence that it's the best machine.[9] What matters is not how much algebra is involved in making the predictions once we've chosen θ, but how many alternatives to θ we've tried out and rejected. The VC dimension lets us apply this kind of reasoning rigorously and without needing to know the details of the process by which we generate and evaluate models.

The upshot is that the kind of complexity which matters for learning, and so for Occam's Razor, is the complexity of *classes of models*, not of individual models nor of the system being modeled. It is important to keep this point in mind when we try to measure the complexity of systems (§8).

2.4. Relation of Complex Systems Science to Statistics

Complex systems scientists often regard the field of statistics as irrelevant to understanding such systems. This is understandable, since the exposure most

scientists have to statistics (e.g., the "research methods" courses traditional in the life and social sciences) typically deal with systems with only a few variables and with explicit assumptions of independence, or only very weak dependence. The kind of modern methods we have just seen, amenable to large systems and strong dependence, are rarely taught in such courses, or even mentioned. Considering the shaky grasp many students have on even the basic principles of statistical inference, this is perhaps wise. Still, it leads to even quite eminent researchers in complexity making disparaging remarks about statistics (e.g., "statistical hypothesis testing, that substitute for thought"), while actually reinventing tools and concepts which have long been familiar to statisticians.

For their part, many statisticians tend to overlook the very *existence* of complex systems science as a separate discipline. One may hope that the increasing interest from both fields on topics such as bioinformatics and networks will lead to greater mutual appreciation.

3. TIME SERIES ANALYSIS

There are two main schools of time series analysis. The older one has a long pedigree in applied statistics (46), and is prevalent among statisticians, social scientists (especially econometricians), and engineers. The younger school, developed essentially since the 1970s, comes out of physics and nonlinear dynamics. The first views time series as samples from a stochastic process, and applies a mixture of traditional statistical tools and assumptions (linear regression, the properties of Gaussian distributions) and the analysis of the Fourier spectrum. The second school views time series as distorted or noisy measurements of an underlying dynamical system, which it aims to reconstruct.

The separation between the two schools is in part due to the fact that, when statistical methods for time series analysis were first being formalized, in the 1920s and 1930s, dynamical systems theory was literally just beginning. The real development of nonlinear dynamics into a powerful discipline has mostly taken place since the 1960s, by which point the statistical theory had acquired a research agenda with a lot of momentum. In turn, many of the physicists involved in experimental nonlinear dynamics in the 1980s and early 1990s were fairly cavalier about statistical issues, and some happily reported results which should have been left in their file-drawers.

There are welcome signs, however, that the two streams of thought are coalescing. Since the 1960s, statisticians have increasingly come to realize the virtues of what they call "state-space models," which are just what the physicists have in mind with their dynamical systems. The physicists, in turn, have become more sensitive to statistical issues, and there is even now some cross-disciplinary work. In this section, I will try, so far as possible, to use the state-space idea as a common framework to present both sets of methods.

3.1. The State-Space Picture

The **state** is a vector-valued function of time, x_t. In discrete time, this evolves according to some map,

$$x_{t+1} \equiv F(x_t, t, \varepsilon_t), \qquad [10]$$

where the map F is allowed to depend on time t and a sequence of independent random variables ε_t. In continuous time, we do not specify the evolution of the state directly, but rather the rates of change of the components of the state,

$$\frac{dx}{dt} = F(x, t, \varepsilon_t). \qquad [11]$$

Since our data are generally taken in discrete time, I will restrict myself to considering that case from now on; almost everything carries over to continuous time naturally. The evolution of x is, so to speak, self-contained, or more precisely Markovian: all the information needed to determine the future is contained in the *present* state x_t, and earlier states are irrelevant. (This is basically how physicists *define* "state" (46).) Indeed, it is often reasonable to assume that F is independent of time, so that the dynamics are **autonomous** (in the terminology of dynamics) or **homogeneous** (in that of statistics). If we could look at the series of states, then, we would find it had many properties which made it very convenient to analyze.

Sadly, however, we do not observe the state x; what we observe or measure is y, which is generally a noisy, nonlinear function of the state: $y_t = h(x_t, \eta_t)$, where η_t is measurement noise. Whether y, too, has the convenient properties depends on h, and usually y is *not* convenient. Matters are made more complicated by the fact that we do not, in typical cases, know the observation function h, nor the state-dynamics F, nor even, really, what space x lives in. The goal of time-series methods is to make educated guess about all these things, so as to better predict and understand the evolution of temporal data.

In the ideal case, simply from a knowledge of y, we would be able to identify the state space, the dynamics, and the observation function. As a matter of pure mathematical possibility, this can be done for essentially arbitrary time series (48,49). Nobody, however, knows how to do this with complete generality in practice. Rather, one makes certain assumptions about, say, the state space, which are strong enough that the remaining details can be filled in using y. Then one checks the result for accuracy and plausibility, i.e., for the kinds of errors which would result from breaking those assumptions (50).

Subsequent parts of this section describe classes of such methods. First, however, I describe some of the general properties of time series, and general measurements which can be made upon them.

Notation. There is no completely uniform notation for time series. Since it will be convenient to refer to sequences of consecutive values. I will write all the measurements starting at s and ending at t as y_s^t. Further, I will abbreviate the set of all measurements up to time t, y_{t+1}^{∞}, as y_{t}^{-}, and the future starting from t, y^{∞}_{t+1}, as y_{t}^{+}.

3.2. General Properties of Time Series

One of the most commonly assumed properties of a time series is **stationarity**, which comes in two forms: **strong** or **strict** stationarity, and **weak**, **wide-sense** or **second-order** stationarity. Strong stationarity is the property that the probability distribution of sequences of observations does not change over time. That is,

$$\Pr(Y_t^{t+h}) = \Pr(Y_{t+\tau}^{t+\tau+h}) \tag{12}$$

for all lengths of time h and all shifts forwards or backwards in time τ. When a series is described as "stationary" without qualification, it depends on context whether strong or weak stationarity is meant.

Weak stationarity, on the other hand, is the property that the first and second moments of the distribution do not change over time.

$$\mathbf{E}[Y_t] = \mathbf{E}[Y_{t+\tau}], \tag{13}$$

$$\mathbf{E}[Y_t Y_{t+h}] = \mathbf{E}[Y_{t+\tau} Y_{t+\tau+h}]. \tag{14}$$

If Y is a Gaussian process, then the two senses of stationarity are equivalent. Note that both sorts of stationarity are statements about the true distribution, and so cannot be simply read off from measurements.

Strong stationarity implies a property called **ergodicity**, which is much more generally applicable. Roughly speaking, a series is ergodic if any sufficiently long sample is representative of the entire process. More exactly, consider the **time-average** of a well-behaved function f of Y,

$$\left\langle f \right\rangle_{t_1}^{t_2} \equiv \frac{1}{t_2 - t_1} \sum_{t=t}^{t=t_2} f(Y_t). \tag{15}$$

This is generally a random quantity, since it depends on where the trajectory started at t_1, and any random motion which may have taken place between then and t_2. Its distribution generally depends on the precise values of t_1 and t_2. The series Y is ergodic if almost all time-averages converge eventually, i.e., if

$$\lim_{T\to\infty} \left\langle f \right\rangle_t^{t+T} = \overline{f} \qquad [16]$$

for some constant $\overline{f}$ independent of the starting time t, the starting point Y_t, or the trajectory Y_t^{∞}. **Ergodic theorems** specify conditions under which ergodicity holds; surprisingly, even completely deterministic dynamical systems can be ergodic.

Ergodicity is such an important property because it means that statistical methods are very directly applicable. Simply by waiting long enough one can obtain an estimate of any desired property that will be closely representative of the future of the process. Statistical inference *is* possible for non-ergodic processes, but it is considerably more difficult, and often requires multiple time series (51,52).

One of the most basic means of studying a time series is to compute the **autocorrelation function** (ACF), which measures the linear dependence between the values of the series at different points in time. This starts with **autocovariance function**:

$$C(s,t) \equiv \mathbf{E}[(y_s - \mathbf{E}[y_s])\,(y_t - \mathbf{E}[y_t])]. \qquad [17]$$

(Statistical physicists, unlike everyone else, call *this* the "correlation function.") The autocorrelation itself is the autocovariance, normalized by the variability of the series:

$$\rho(s,t) \equiv \frac{C(s,t)}{\sqrt{C(s,s)C(t,t)}}, \qquad [18]$$

ρ is ± 1 when y_s is a linear function of y_t. Note that the definition is symmetric, so $\rho(s,t) = \rho(t,s)$. For stationary or weakly stationary processes, one can show that ρ depends only on the difference τ between t and s. In this case one just writes $\rho(\tau)$, with one argument. $\rho(0) = 1$, always. The time t_c such that $\rho(t_c) = 1/e$ is called the **(auto)correlation time** of the series.

The correlation function is a **time-domain** property, since it is basically about the series considered as a sequence of values at distinct times. There are also **frequency-domain** properties, which depend on reexpressing the series as a sum of sines and cosines with definite frequencies. A function of time y has a Fourier transform that is a function of frequency, $\tilde{y}$:

$$\tilde{y} = \mathcal{F}y\,, \qquad [19]$$

$$\tilde{y}_\nu = \sum_{t=1}^{T} e^{-i\frac{2\pi\nu t}{T}} y_t\,, \qquad [20]$$

assuming the time series runs from $t = 1$ to $t = T$. (Rather than separating out the sine and cosine terms, it is easier to use the complex-number representation, via $e^{i\theta} = \cos\theta + i \sin\theta$.) The inverse Fourier transform recovers the original function:

$$y = \mathcal{F}^{-1}\tilde{y}, \tag{21}$$

$$y_t = \frac{1}{T}\sum_{\nu=0}^{T-1} e^{i\frac{2\pi\nu t}{T}} \tilde{y}_\nu . \tag{22}$$

The Fourier transform is a linear operator, in the sense that $\mathcal{F}(x + y) = \mathcal{F}x + \mathcal{F}y$. Moreover, it represents series we are interested in as a sum of trigonometric functions, which are themselves solutions to linear differential equations. These facts lead to extremely powerful frequency-domain techniques for studying linear systems. Of course, the Fourier transform is always *valid*, whether the system concerned is linear or not, and it may well be useful, though that is not guaranteed.

The squared absolute value of the Fourier transform, $f(\nu) = |\tilde{y}_\nu|^2$, is called the **spectral density** or **power spectrum**. For stationary processes, the power spectrum $f(\nu)$ is the Fourier transform of the autocovariance function $C(\tau)$ (a result called the Wiener-Khinchin theorem). An important consequence is that a Gaussian process is completely specified by its power spectrum. In particular, consider a sequence of independent Gaussian variables, each with variance σ^2. Because they are perfectly uncorrelated, $C(0) = \sigma^2$, and $C(\tau) = 0$ for any $\tau \neq 0$. The Fourier transform of such a $C(\tau)$ is just $f(\nu) = \sigma^2$, independent of ν—every frequency has just as much power. Because white light has equal power in every color of the spectrum, such a process is called **white noise**. Correlated processes, with uneven power spectra, are sometimes called **colored noise**, and there is an elaborate terminology of red, pink, brown, etc., noises (53, ch. 3).

The easiest way to estimate the power spectrum is simply to take the Fourier transform of the time series, using, e.g., the fast Fourier transform algorithm (54). Equivalently, one might calculate the autocovariance and Fourier transform in that manner. Either way, one has an estimate of the spectrum, which is called the **periodogram**. It is unbiased, in that the expected value of the periodogram at a given frequency is the true power at that frequency. Unfortunately, it is not consistent—the variance around the true value does not shrink as the series grows. The easiest way to overcome this is to apply any of several well-known smoothing functions to the periodogram, a procedure called **windowing** (55). (Standard software packages will accomplish this automatically.)

The Fourier transform takes the original series and decomposes it into a sum of sines and cosines. This is possible because *any* reasonable function can be represented in this way. The trigonometric functions are thus a **basis** for the space of functions. There are many other possible bases, and one can equally

well perform the same kind of decomposition in any other basis. The trigonometric basis is particularly useful for stationary time series because the basis functions are themselves evenly spread over all times (56, ch. 2). Other bases, localized in time, are more convenient for nonstationary situations. The most well-known of these alternate bases, currently, are wavelets (57), but there is, literally, no counting the other possibilities.

3.3. The Traditional Statistical Approach

The traditional statistical approach to time series is to represent them through linear models of the kind familiar from applied statistics.

The most basic kind of model is that of a **moving average**, which is especially appropriate if x is highly correlated up to some lag, say q, after which the ACF decays rapidly. The moving average model represents x as the result of smoothing $q + 1$ independent random variables. Specifically, the MA(q) model of a weakly stationary series is

$$y_t = \mu + w_t + \sum_{k=1}^{q} \theta_k w_{t-k}, \qquad [23]$$

where μ is the mean of y, the θ_i are constants and the w_t are white noise variables. q is called the **order** of the model. Note that there is no direct dependence between successive values of y; they are all functions of the white noise series w. Note also that y_t and y_{t+q+1} are completely independent; after q time-steps, the effects of what happened at time t disappear.

Another basic model is that of an **autoregressive process**, where the next value of y is a linear combination of the preceding values of y. Specifically, an AR(p) model is

$$y_t = \alpha + \sum_{k=1}^{p} \phi_k y_{t-k} + w_t, \qquad [24]$$

where ϕ_i are constants and $\alpha = \mu + \sum_{k=1}^{p} \phi_k$. The order of the model, again is p. This is the multiple regression of applied statistics transposed directly on to time series, and is surprisingly effective. Here, unlike the moving average case, effects propagate indefinitely—changing y_t can affect all subsequent values of y. The remote past only becomes irrelevant if one controls for the last p values of the series. If the noise term w_t were absent, an AR(p) model would be a pth order linear difference equation, the solution to which would be some combination of exponential growth, exponential decay and harmonic oscillation. With noise, they become oscillators under stochastic forcing (58).

The natural combination of the two types of model is the **autoregressive moving average model**, ARMA(p,q):

$$y_t = \alpha + \sum_{k-1}^{p} \phi_k y_{t-k} + w_t + \sum_{k=1}^{q} \theta_k w_{t-k} . \quad [25]$$

This combines the oscillations of the AR models with the correlated driving noise of the MA models. An AR(p) model is the same as an ARMA(p,0) model, and likewise an MA(q) model is an ARMA(0,q) model.

It is convenient, at this point in our exposition, to introduce the notion of the **back-shift operator** B,

$$By_t = y_{t-1}, \quad [26]$$

and the **AR and MA polynomials**,

$$\phi(z) = 1 - \sum_{k=1}^{p} \phi_k z^k , \quad [27]$$

$$\theta(z) = 1 + \sum_{k=1}^{q} \theta_k z^k , \quad [28]$$

respectively. Then, formally speaking, in an ARMA process is

$$\phi(B) y_t = \theta(B) w_t . \quad [29]$$

The advantage of doing this is that one can determine many properties of an ARMA process by algebra on the polynomials. For instance, two important properties we want a model to have are **invertibility** and **causality**. We say that the model is invertible if the sequence of noise variables w_t can be determined uniquely from the observations y_t; in this case we can write it as an MA(∞) model. This is possible just when $\theta(z)$ has no roots inside the unit circle. Similarly, we say the model is causal if it can be written as an AR(∞) model, without reference to any *future* values. When this is true, $\phi(z)$ also has no roots inside the unit circle.

If we have a causal, invertible ARMA model, with known parameters, we can work out the sequence of noise terms, or **innovations** w_t associated with our measured values y_t. Then, if we want to forecast what happens past the end of our series, we can simply extrapolate forward, getting predictions $\hat{y}_{T+1}, \hat{y}_{T+2}$, etc. Conversely, if we knew the innovation sequence, we could determine the parameters ϕ and θ. When both are unknown, as is the case when we want to fit a model, we need to determine them jointly (55). In particular, a common proce-

dure is to work forward through the data, trying to predict the value at each time on the basis of the past of the series; the sum of the squared differences between these predicted values $\hat{y}_t$ and the actual ones y_t forms the empirical loss:

$$L = \sum_{i=1}^{T} (y_t - \hat{y}_t)^2 . \qquad [30]$$

For this loss function, in particular, there are very fast standard algorithms, and the estimates of ϕ and θ converge on their true values, provided one has the right model order.

This leads naturally to the question of how one determines the order of ARMA model to use, i.e., how one picks p and q. This is precisely a model selection task, as discussed in §2. All methods described there are potentially applicable; cross-validation and regularization are more commonly used than capacity control. Many software packages will easily implement selection according to the AIC, for instance.

The power spectrum of an ARMA(p,q) process can be given in closed form:

$$f(\nu) = \frac{\sigma^2}{2\pi} \frac{(1 + \sum_{k=1}^{q} \theta_k e^{-i\nu k})^2}{(1 + \sum_{k=1}^{p} \phi_k e^{-\nu k})^2} . \qquad [31]$$

Thus, the parameters of an ARMA process can be estimated directly from the power spectrum, if you have a reliable estimate of the spectrum. Conversely, different hypotheses about the parameters can be checked from spectral data.

All ARMA models are weakly stationary; to apply them to nonstationary data one must transform the data so as to make it stationary. A common transformation is **differencing**, i.e., applying operations of the form

$$\nabla y_t = y_t - y_{t-1}, \qquad [32]$$

which tends to eliminate regular trends. In terms of the back-shift operator,

$$\nabla y_t = (1 - B) y_t, \qquad [33]$$

and higher-order differences are

$$\nabla^d y_t = (1 - \mathrm{B})^d y_t. \qquad [34]$$

Having differenced the data to our satisfaction, say d times, we then fit an ARMA model to it. The result is an **autoregressive integrated moving average** model, ARIMA(p,d,q) (59), given by

$$\phi(B)(1-B)^d y_t = \theta(B) w_t, \qquad [35]$$

As mentioned above (§3.1), ARMA and ARIMA models can be recast in state space terms, so that our y is a noisy measurement of a hidden x (60). For these models, both the dynamics and the observation functions are linear, that is, $x_{t+1} = \mathbf{A}x_t + \varepsilon_t$ and $y_t = \mathbf{B}x_t + \eta_t$, for some matrices $\mathbf{A}$ and $\mathbf{B}$. The matrices can be determined from the θ and ϕ parameters, though the relation is a bit too involved to give here.

3.3.1. Applicability of Linear Statistical Models

It is often possible to describe a nonlinear dynamical system through an effective linear statistical model, provided the nonlinearities are cooperative enough to appear as noise (61). It is an under-appreciated fact that this is at least sometimes true even of turbulent flows (62,63); the generality of such an approach is not known. Certainly, if you care only about predicting a time series, and not about its structure, it is always a good idea to try a linear model first, even if you *know* that the real dynamics are highly nonlinear.

3.3.2. Extensions

While standard linear models are more flexible than one might think, they do have their limits, and recognition of this has spurred work on many extensions and variants. Here I briefly discuss a few of these.

Long Memory. The correlations of standard ARMA and ARIMA models decay fairly rapidly, in general exponentially; $\rho(t) \propto e^{-t/t_c}$, where τ_c is the correlation time. For some series, however, τ_c is effectively infinite, and $\rho(t) \propto t^{-\alpha}$ for some exponent α. These are **long-memory processes**, because they remain substantially correlated over very long times. These can still be accommodated within the ARIMA framework, formally, by introducing the idea of *fractional* differencing, or, in continuous time, fractional derivatives (64,53). Often long-memory processes are self-similar, which can simplify their statistical estimation (65).

Volatility. All ARMA and even ARIMA models assume constant variance. If the variance is itself variable, it can be worthwhile to model it. **Autoregressive conditionally heteroscedastic** (ARCH) models assume a fixed mean value for y_t, but a variance which is an auto-regression on y_t^2. **Generalized ARCH** (GARCH) models expand the regression to include the (unobserved) earlier variances. ARCH and GARCH models are especially suitable for processes that display **clustered volatility**, periods of extreme fluctuation separated by stretches of comparative calm.

Nonlinear and Nonparametric Models. Nonlinear models are obviously appealing, and when a particular parametric form of model is available, reasonably straightforward modifications of the linear machinery can be used to fit, evaluate and forecast the model (55, chap. 9). However, it is often impractical to settle on a good parametric form beforehand. In these cases, one must turn to nonparametric models, as discussed in §2.2; neural networks are a particular favorite here (35). The so-called **kernel smoothing methods** are also particularly well-developed for time series, and often perform almost as well as parametric models (66). Finally, information theory provides **universal prediction methods**, which promise to asymptotically approach the best possible prediction, starting from exactly no background knowledge. This power is paid for by demanding a long initial training phase used to infer the structure of the process, when predictions are much worse than many other methods could deliver (67).

3.4. The Nonlinear Dynamics Approach

The younger approach to the analysis of time series comes from nonlinear dynamics, and is intimately bound up with the state-space approach described in §3.1 above. The idea is that the dynamics on the state space can be determined *directly* from observations, at least if certain conditions are met.

The central result here is the Takens Embedding Theorem (68); a simplified, slightly inaccurate version is as follows. Suppose the d-dimensional state vector x_t evolves according to an unknown but continuous and (crucially) deterministic dynamic. Suppose, too, that the one-dimensional observable y is a smooth function of x, and "coupled" to all the components of x. Now at any time we can look not just at the present measurement $y(t)$, but also at observations made at times removed from us by multiples of some lag τ: $y_{t-\tau}$, $y_{t-2\tau}$, etc. If we use k lags, we have a k-dimensional vector. One might expect that, as the number of lags is increased, the motion in the lagged space will become more and more predictable, and perhaps in the limit $k \to \infty$ would become deterministic. In fact, the dynamics of the lagged vectors become deterministic at a finite dimension; not only that, but the deterministic dynamics are completely equivalent to those of the original state space! (More exactly, they are related by a smooth, invertible change of coordinates, or **diffeomorphism**.) The magic **embedding dimension** k is at most $2d + 1$, and often less.

Given an appropriate reconstruction via embedding, one can investigate many aspects of the dynamics. Because the reconstructed space is related to the original state space by a smooth change of coordinates, any geometric property that survives such treatment is the same for both spaces. These include the dimension of the attractor, the Lyapunov exponents (which measure the degree of sensitivity to initial conditions), and certain qualitative properties of the autocorrelation function and power spectrum ("correlation dimension"). Also preserved

is the relation of "closeness" among trajectories—two trajectories that are close in the state space will be close in the embedding space, and vice versa. This leads to a popular and robust scheme for nonlinear prediction, the **method of analogs**: when one wants to predict the next step of the series, take the current point in the embedding space, find a similar one with a known successor, and predict that the current point will do the analogous thing. Many refinements are possible, such as taking a weighted average of nearest neighbors, or selecting an analog at random, with a probability decreasing rapidly with distance. Alternately, one can simply fit non-parametric predictors on the embedding space. (See (69) for a review.) Closely related is the idea of **noise reduction**, using the structure of the embedding-space to filter out some of the effects of measurement noise. This can work even when the statistical character of the noise is unknown (see (69) again).

Determining the number of lags, and the lag itself, is a problem of model selection, just as in §2, and can be approached in that spirit. An obvious approach is to minimize the in-sample forecasting error, as with ARMA models; recent work along these lines (70,71) uses the minimum description length principle (described in §8.3.1 below) to control over-fitting. A more common procedure for determining the embedding dimension, however, is the **false nearest neighbor method** (72). The idea is that if the current embedding dimension k is sufficient to resolve the dynamics, $k + 1$ would be too, and the reconstructed state space will not change very much. In particular, points which were close together in the dimension-k embedding should remain close in the dimension-$k + 1$ embedding. Conversely, if the embedding dimension is too small, points that are really far apart will be brought artificially close together (just as projecting a sphere on to a disk brings together points on the opposite side of a sphere). The particular algorithm of Kennel et al. (72), which has proved very practical, is to take each point in the k-dimensional embedding, find its nearest neighbor in that embedding, and then calculate the distance between them. One then calculates how much further apart they would be if one used a k+1-dimensional embedding. If this extra distance is more than a certain fixed multiple of the original distance, they are said to be "false nearest neighbors." (Ratios of 2 to 15 are common, but the precise value does not seem to matter very much.) One then repeats the process at dimension $k + 1$, stopping when the proportion of false nearest neighbors becomes zero, or at any rate sufficiently small. Here, the loss function used to guide model selection is the number of false nearest neighbors, and the standard prescriptions amount to empirical risk minimization. One reason simple ERM works well here is that the problem is intrinsically finite-dimensional (via the Takens result).

Unfortunately, the data required for calculations of quantities like dimensions and exponents to be reliable can be quite voluminous. Approximately $10^{2+0.4D}$ data-points are necessary to adequately reconstruct an attractor of dimension D (73, pp. 317–319). (Even this is more optimistic than the widely quoted,

if apparently pessimistic, calculation of (74), that attractor reconstruction with an *embedding* dimension of k needs 42^k data-points!) In the early days of the application of embedding methods to experimental data, these limitations were not well appreciated, leading to many calculations of low-dimensional deterministic chaos in EEG and EKG series, economic time series, etc., which did not stand up to further scrutiny. This in turn brought some discredit on the methods themselves, which was not really fair. More positively, it also led to the development of ideas such as **surrogate-data methods**. Suppose you have found what seems like a good embedding, and it appears that your series was produced by an underlying deterministic attractor of dimension D. One way to test this hypothesis would be to see what kind of results your embedding method would give if applied to similar but *non*-deterministic data. Concretely, you find a stochastic model with similar statistical properties (e.g., an ARMA model with the same power spectrum), and simulate many time series from this model. You apply your embedding method to each of these **surrogate data** series, getting the approximate distribution of apparent "attractor" dimensions when there really is no attractor. If the dimension measured from the original data is not significantly different from what one would expect under this null hypothesis, the evidence for an attractor (at least from this source) is weak. To apply surrogate data tests well, one must be very careful in constructing the null model, as it is easy to use over-simple null models, biasing the test towards apparent determinism.

A few further cautions on embedding methods are in order. While *in principle* any lag τ is suitable, in practice both very long and very short lags lead to pathologies. A common practice is to set the lag to the autocorrelation time (see above), or the first minimum of the mutual information function (see §7 below), the notion being that this most nearly achieves a genuinely "new" measurement (75). There is some evidence that the mutual information method works better (76). Again, while in principle almost any smooth observation function will do, given enough data, in practice some make it much easier to reconstruct the dynamics; several **indices of observability** try to quantify this (77). Finally, it strictly applies only to deterministic observations of deterministic systems. Embedding approaches are reasonably robust to a degree of noise in the observations. They do not cope at all well, however, to noise in the dynamics itself. To anthropomorphize a little, when confronted by apparent non-determinism, they respond by adding more dimensions, and so distinguishing apparently similar cases. Thus, when confronted with data that really are stochastic, they will infer an infinite number of dimensions, which is correct in a way, but definitely not helpful. These remarks should not be taken to belittle the very real power of nonlinear dynamics methods. Applied skillfully, they are powerful tools for understanding the behavior of complex systems, especially for probing aspects of their structure which are not directly accessible.

3.5. Filtering and State Estimation

Suppose we have a state-space model for our time series, and some observations y, can we find the state x? This is the problem of **filtering** or **state estimation**. Clearly, it is not the same as the problem of finding a model in the first place, but it is closely related, and also a problem in statistical inference.

In this context, a **filter** is a function which provides an estimate $\hat{x}_t$ of x_t on the basis of observations up to and including[10] time t: $\hat{x}_t = f(y_0^t)$. A filter is **recursive**[11] if it estimates the state at t on the basis of its estimate at $t-1$ and the new observation: $\hat{x}_t = f(\hat{x}_{t-1}, y_t)$. Recursive filters are especially suited to online use, since one does not need to retain the complete sequence of previous observations, merely the most recent estimate of the state. As with prediction in general, filters can be designed to provide either point estimates of the state, or distributional estimates. Ideally, in the latter case, we would get the conditional distribution, $\Pr(X_t = x | Y_1^t = y_1^t)$, and in the former case the conditional expectation, $\int_x x \Pr(X_t = x | Y_1^t = y_1^t) dx$.

Given the frequency with which the problem of state estimation shows up in different disciplines, and its general importance when it does appear, much thought has been devoted to it over many years. The problem of optimal *linear* filters for stationary processes was solved independently by two of the "grandfathers" of complex systems science, Norbert Wiener and A.N. Kolmogorov, during the Second World War (78,79). In the 1960s, Kalman and Bucy (80–82) solved the problem of optimal recursive filtering, assuming linear dynamics, linear observations and additive noise. In the resulting **Kalman filter**, the new estimate of the state is a weighted combination of the old state, extrapolated forward, and the state that would be inferred from the new observation alone. The requirement of linear dynamics can be relaxed slightly with what's called the "extended Kalman filter," essentially by linearizing the dynamics around the current estimated state.

Nonlinear solutions go back to pioneering work of Stratonovich (83) and Kushner (84) in the later 1960s, who gave optimal, recursive solutions. Unlike the Wiener or Kalman filters, which give point estimates, the Stratonovich-Kushner approach calculates the complete conditional distribution of the state; point estimates take the form of the mean or the most probable state (85). In most circumstances, the strictly optimal filter is hopelessly impractical numerically. Modern developments, however, have opened up some very important lines of approach to practical nonlinear filters (86), including approaches that exploit the geometry of the nonlinear dynamics (87,88), as well as more mundane methods that yield tractable numerical approximations to the optimal filters (89,90). Noise reduction methods (§3.4) and hidden Markov models (§3.6) can also be regarded as nonlinear filters.

3.6. Symbolic or Categorical Time Series

The methods we have considered so far are intended for time series taking continuous values. An alternative is to break the range of the time series into discrete categories (generally only finitely many of them); these categories are sometimes called **symbols**, and the study of these time series **symbolic dynamics**. Modeling and prediction then reduces to a (perhaps more tractable) problem in discrete probability, and many methods can be used that are simply inapplicable to continuous-valued series (10). Of course, if a bad discretization is chosen, the results of such methods are pretty well meaningless, but sometimes one gets data that are already nicely discrete—human languages, the sequences of biopolymers, neuronal spike trains, etc. We shall return to the issue of discretization below, but for the moment we will simply consider the applicable methods for discrete-valued, discrete-time series, however obtained.

Formally, we take a continuous variable z and **partition** its range into a number of discrete **cells**, each labeled by a different symbol from some **alphabet**; the partition gives us a discrete variable $y = \phi(z)$. A **word** or **string** is just a sequence of symbols, $y_0 y_1 \ldots y_n$. A time series z_0^n naturally generates a string $\phi(z_0^n) \equiv \phi(z_0)\ \phi(z_1) \ldots \phi(z_n)$. In general, not every possible string can actually be generated by the dynamics of the system we're considering. The set of allowed sequences is called the **language**. A sequence that is never generated is said to be **forbidden**. In a slightly inconsistent metaphor, the rules that specify the allowed words of a language are called its **grammar**. To each grammar there corresponds an abstract machine or **automaton** that can determine whether a given word belongs to the language, or, equivalently, generate all and only the allowed words of the language. The generative versions of these automata are stochastic, i.e., they generate different words with different probabilities, matching the statistics of $\phi(z)$.

By imposing restrictions on the forms the grammatical rules can take, or, equivalently, on the memory available to the automaton, we can divide all languages into four nested classes, a hierarchical classification due to Chomsky (91). At the bottom are the members of the weakest, most restricted class, the **regular languages** generated by automata within only a fixed, finite memory for past symbols (**finite state machines**). Above them are the **context free** languages, whose grammars do not depend on context; the corresponding machines are **stack automata**, which can store an unlimited number of symbols in their memory, but on a strictly first-in, first-out basis. Then come the **context-sensitive** languages; and at the very top, the unrestricted languages, generated by universal computers. Each stage in the hierarchy can simulate all those beneath it.

We may seem to have departed very far from dynamics, but actually this is not so. Because different languages classes are distinguished by different kinds of memories, they have very different correlation properties (§3.2), mutual in-

formation functions (§7), and so forth—see (10) for details. Moreover, it is often easier to determine these properties from a system's grammar than from direct examination of sequence statistics, especially since specialized techniques are available for grammatical inference (92,93).

3.6.1. Hidden Markov Models

The most important special case of this general picture is that of regular languages. These, we said, are generated by machines with only a finite memory. More exactly, there is a finite set of states x, with two properties:

1. The distribution of y_t depends solely on x_t, and
2. The distribution of x_{t+1} depends solely on x_t.

That is, the x sequence is a Markov chain, and the observed y sequence is a noisy function of that chain. Such models are very familiar in signal processing (94), bioinformatics (95), and elsewhere, under the name of **hidden Markov models** (HMMs). They can be thought of as a generalization of ordinary Markov chains to the state-space picture described in §3.1. HMMs are particularly useful in filtering applications, since very efficient algorithms exist for determining the most probable values of x from the observed sequence y. The **expectation-maximization** (EM) algorithm (96) even allows us to simultaneously infer the most probable hidden states and the most probable parameters for the model.

3.6.2. Variable-Length Markov Models

The main limitation of ordinary HMMs methods, even the EM algorithm, is that they assume a fixed **architecture** for the states, and a fixed relationship between the states and the observations. That is to say, they are not geared towards inferring the structure of the model. One could apply the model-selection techniques of §2, but methods of direct inference have also been developed. A popular one relies on **variable-length Markov models**, also called **context trees** or **probabilistic suffix trees** (97–100).

A suffix here is the string at the end of the y time series at a given time, so, for example, the binary series *abbabbabb* has suffixes *b*, *bb*, *abb*, *babb*, etc., but not *bab*. A suffix is a **context** if the future of the series is independent of its past, given the suffix. Context-tree algorithms try to identify contexts by iteratively considering longer and longer suffixes, until they find one that seems to be a context. For instance, in a binary series, such an algorithm would first try

whether the suffices a and b are contexts, i.e., whether the conditional distribution $\Pr(Y_{t+1}|Y_t = a)$ can be distinguished from $\Pr(Y_{t+1}|Y_t = a, Y^-_{t-1})$, and likewise for $Y_t = b$. It could happen that a is a context but b is not, in which case the algorithm will try ab and bb, and so on. If one sets x_t equal to the context at time t, x_t is a Markov chain. This is called a *variable-length* Markov model because the contexts can be of different lengths.

Once a set of contexts has been found, they can be used for prediction. Each context corresponds to a different distribution for one-step-ahead predictions, and so one just needs to find the context of the current time series. One could apply state-estimation techniques to find the context, but an easier solution is to use the construction process of the contexts to build a decision tree (§2), where the first level looks at Y_t, the second at Y_{t-1}, and so forth.

Variable-length Markov models are conceptually simple, flexible, fast, and frequently more accurate than other ways of approaching the symbolic dynamics of experimental systems (101). However, not every regular language can be represented by a finite number of contexts. This weakness can be remedied by moving to a more powerful class of models, discussed next.

3.6.3. Causal-State Models, Observable-Operator Models, and Predictive-State Representations

In discussing the state-space picture in §3.1 above, we saw that the state of a system is basically defined by specifying its future time-evolution, to the extent that it can be specified. Viewed in this way, a state X_t corresponds to a distribution over future observables Y^+_{t+1}. One natural way of finding such distributions is to look at the *conditional* distribution of the future observations, given the previous history, i.e., $\Pr(Y^+_{t+1}|Y^-_t = y^-_t)$. For a given stochastic process or dynamical system, there will be a certain characteristic family of such conditional distributions. One can then consider the distribution-valued process generated by the original, observed process. It turns out that the former is always a Markov process, and that the original process can be expressed as a function of this Markov process plus noise. In fact, the distribution-valued process has all the properties one would want of a state-space model of the observations (48,49). The conditional distributions, then, can be treated as states.

This remarkable fact has led to techniques for modeling discrete-valued time series, all of which attempt to capture the conditional-distribution states, and all of which are strictly more powerful than VLMMs. There are at least three: the **causal-state models** or **causal-state machines** (CSMs),[12] introduced by Crutchfield and Young (102), the **observable operator models** (OOMs) introduced by Jaeger (103), and the **predictive state representations** (PSRs) introduced by Littman, Sutton, and Singh (104). The simplest way of thinking of such objects is that they are VLMMs where a context or state can contain more

than one suffix, adding expressive power and allowing them to give compact representations of a wider range of processes. (See (105) for more on this point, with examples.)

All three techniques—CSMs, OOMs and PSRs—are basically equivalent, though they differ in their formalisms and their emphases. CSMs focus on representing states as classes of histories with the same conditional distributions, i.e., as suffixes sharing a single context. (They also feature in the "statistical forecasting" approach to measuring complexity, discussed in §8.3.2 below.) OOMs are named after the operators that update the state; there is one such operator for each possible observation. PSRs, finally, emphasize the fact that one does not actually need to know the probability of every possible string of future observations, but just a restricted subset of key trajectories, called "tests." In point of fact, all of them can be regarded as special cases of more general prior constructions due to Salmon ("statistical relevance basis") (106,107) and Knight ("measure-theoretic prediction process") (48,49), which were themselves independent. (This area of the literature is more than usually tangled.)

Efficient **reconstruction algorithms** or **discovery procedures** exist for building CSMs (105) and OOMs (103) directly from data. (There is currently no such discovery procedure for PSRs, though there are parameter-estimation algorithms (108).) These algorithms are reliable, in the sense that, given enough data, the probability that they build the wrong set of states becomes arbitrarily small. Experimentally, selecting an HMM architecture through cross-validation never does better than reconstruction, and often much worse (105).

While these models are more powerful than VLMMs, there are still many stochastic processes that cannot be represented in this form; or, rather, their representation requires an infinite number of states (109,110). This is mathematically unproblematic, though reconstruction will then become much harder. (For technical reasons, it seems likely to be easier to carry through for OOMs or PSRs than for CSMs.) In fact, one can show that these techniques would work straightforwardly on continuous-valued, continuous-time processes, if only we knew the necessary conditional distributions (48,111). Devising a reconstruction algorithm suitable for this setting is an extremely challenging and completely unsolved problem; even parameter estimation is difficult, and currently only possible under quite restrictive assumptions (112).

3.6.4. Generating Partitions

So far, everything has assumed that we are either observing truly discrete quantities, or that we have a fixed discretization of our continuous observations. In the latter case, it is natural to wonder how much difference the discretization makes. The answer, it turns out, is *quite a lot*; changing the partition can lead to

completely different symbolic dynamics (113–115). How then might we choose a *good* partition?

Nonlinear dynamics provides an answer, at least for deterministic systems, in the idea of a **generating partition** (10,116). Suppose we have a continuous state x and a deterministic map on the state F, as in §3.1. Under a partitioning ϕ, each point x in the state space will generate an infinite sequence of symbols, $\Phi(x)$, as follows: $\phi(x)$, $\phi(F(x))$, $\phi(F^2(x))$, ... The partition ϕ is generating if each point x corresponds to a *unique* symbol sequence, i.e., if Φ is invertible. Thus, no information is lost in going from the continuous state to the discrete symbol sequence.[13] While one must know the continuous map F to determine exact generating partitions, there are reasonable algorithms for approximating them from data, particularly in combination with embedding methods (75,117,118). When the underlying dynamics are stochastic, however, the situation is much more complicated (119).

4. CELLULAR AUTOMATA

Cellular automata are one of the more popular and distinctive classes of models of complex systems. Originally introduced by von Neumann as a way of studying the possibility of mechanical self-reproduction, they have established niches for themselves in foundational questions relating physics to computation in statistical mechanics, fluid dynamics, and pattern formation. Within that last, perhaps the most relevant to the present purpose, they have been extensively and successfully applied to physical and chemical pattern formation, and, somewhat more speculatively, to biological development and to ecological dynamics. Interesting attempts to apply them to questions like the development of cities and regional economies lie outside the scope of this chapter.

4.1. A Basic Explanation of CA

Take a board, and divide it up into squares, like a chess- or checkerboard. These are the cells. Each cell has one of a finite number of distinct colors—red and black, say, or (to be patriotic) red, white, and blue. (We do not allow continuous shading, and every cell has just one color.) Now we come to the "automaton" part. Sitting somewhere to one side of the board is a clock, and every time the clock ticks the colors of the cells change. Each cell looks at the colors of the nearby cells, and its own color, and then applies a definite rule, the **transition rule**, specified in advance, to decide its color in the next clock-tick; and all the cells change at the same time. (The rule can say "stay the same.") Each cell is a sort of very stupid computer—in the jargon, a **finite-state**

automaton—and so the whole board is called a **cellular automaton**, or CA. To run it, you color the cells in your favorite pattern, start the clock, and stand back.

Let us follow this concrete picture with one more technical and abstract. The cells do not have to be colored, of course; all that's important is that each cell is in one of a finite number of states at any given time. By custom they're written as the integers, starting from 0, but any "finite alphabet" will do. Usually the number of states is small, under ten, but in principle any finite number is allowed. What counts as the "nearby cells," the **neighborhood**, varies from automaton to automaton; sometimes just the four cells on the principal directions, sometimes the corner cells, sometimes a block or diamond of larger size; in principle any arbitrary shape. You do not need to stick to a chessboard; you can use any regular pattern of cells that will fill the plane (or "tessellate" it; an old name for cellular automata is **tessellation structures**). And you do not have to stick to the plane; any number of dimensions is allowed. There are various tricks for handling the edges of the space; the one which has "all the advantages of theft over honest toil" is to assume an infinite board.

Cellular Automata as Parallel Computers. CA are synchronous massively parallel computers, with each cell being a finite state transducer, taking input from its neighbors and making its own state available as output. From this perspective, the remarkable thing about CA is that they are computationally universal, able to calculate any (classically) computable function; one can use finite-state machines, the least powerful kind of computer, to build devices equivalent to Turing machines, the most powerful kind of computer. The computational power of different physically motivated CA is an important topic in complex systems (120,121), though it must be confessed that CA with very different computational powers can have very similar behavior in most other respects.

Cellular Automata as Discrete Field Theories. From the perspective of physics, a CA is a "digitized" classical field theory, in which space, time, and the field (state) are all discrete. Thus fluid mechanics, continuum mechanics, and electromagnetism can all be simulated by CA[14]; typically, however, the physical relevance of a CA comes not from accurately simulating some field theory at the microscopic level, but from the large-scale phenomena they generate.

Take, for example, simulating fluid mechanics, where CA are also called **lattice gases** or **lattice fluids**. In the "HPP" (122) rule, a typical lattice gas with a square grid, there are four species of "fluid particle," which travel along the four principal directions. If two cells moving in opposite directions try to occupy the same location at the same time, they collide, and move off at right angles to their original axis Figure 4). Each cell thus contains only an integer number of particles, and only a discrete number of values of momentum are possible. If one takes averages over reasonably large regions, however, then density and

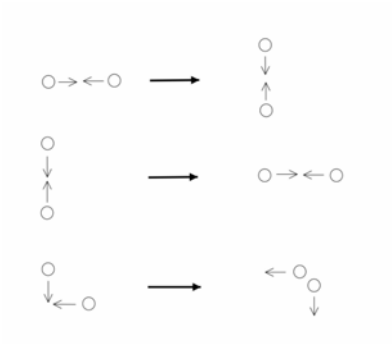


Figure 4. Collisions in the HPP lattice gas rule. Horizontal collisions produce vertically moving particles (top) and vice versa (middle). Particles moving at right angles pass by each other unchanged (bottom, omitting the reflections and rotations of this figure).

momentum approximately obey the equations of continuous fluid mechanics. Numerical experiments show that this rule reproduces many fluid phenomena, such as diffusion, sound, shockwaves, etc. However, with this rule, the agreement with fluid mechanics is only approximate. In particular, the square lattice makes the large-scale dynamics anisotropic, which is unphysical. This in turn can be overcome in several ways—for instance, by using a hexagonal lattice (123). The principle here—get the key parts of the small-scale "microphysics" right, and the interesting "macrophysics" will take care of itself—is extensively applied in studying pattern formation, including such biologically relevant phenomena as phase separation (124), excitable media (125), and the self-assembly of micelles (126,127).

5. AGENT-BASED MODELS

If there is any *one* technique associated with complex systems science, it is agent-based modeling. An agent-based model is a computational model that represents individual agents and their collective behavior. What, exactly, do we mean by "agent"? Stuart Kauffman has offered[15] the following apt definition: "An agent is a thing which does things to things." That is, an agent is a persistent thing that has some *state* we find worth representing, and which interacts with other agents, mutually modifying each others' states. The components of an agent-based model are a collection of agents and their states, the rules governing the interactions of the agents, and the environment within which they live. (The environment need not be represented in the model if its effects are constant.) The state of an agent can be arbitrarily simple, say just position, or the color of a cell in a CA. (At this end, agent-based models blend with traditional stochastic

models.) States can also be extremely complicated, including, possibly, sophisticated internal models of the agent's world.

Here is an example to make this concrete. In epidemiology, there is a classic kind of model of the spread of a disease through a population called an "SIR" model (128, §4). It has three classes of people—the susceptible, who have yet to be exposed to the disease; the infected, who have it and can pass it on; and the resistant or recovered, who have survived the disease and cannot be reinfected. A traditional approach to an SIR model would have three variables, namely the number of people in each of the three categories, $S(t)$, $I(t)$, $R(t)$, and would have some deterministic or stochastic dynamics in terms of those variables. For instance, in a deterministic SIR model, one might have

$$S(t+1) - S(t) = -a\left(\frac{I(t)}{S(t)+I(t)+R(t)}\right)S(t)\,, \qquad [36]$$

$$I(t+1) - I(t) = \left[a\frac{S(t)}{S(t)+I(t)+R(t)} - b - c\right]I(t)\,, \qquad [37]$$

$$R(t+1) - R(t) = bI(t), \qquad [38]$$

which we could interpret by saying that (i) the probability of a susceptible person being infected is proportional to the fraction of the population which is already infected, (ii) infected people get better at a rate b, and (iii) infected people die at a rate c. (This is not a particularly *realistic* SIR model.) In a stochastic model, we would treat the right-hand sides of [36]–[38] as the mean changes in the three variables, with (say) Poisson-distributed fluctuations, taking care that, e.g., the fluctuation in the $aI/(R+S+I)$ term in [36] is the same as that in [37]. The thing to note is that, whether deterministic or stochastic, the whole model is cast in terms of the aggregate quantities S, I and R, and those aggregate variables are what we would represent computationally.

In an agent-based model of the same dynamics, we would represent *each* individual in the population as a distinct agent, which could be in one of three states, S, I, and R. A simple interaction rule would be that at each time-step, each agent selects another from the population entirely at random. If a susceptible agent (i.e., one in state S) picks an infectious agent (i.e., one in state I), it becomes infected with probability a. Infectious agents die with probability b and recover with probability c; recovered agents never change their state. So far, we have merely reproduced the stochastic version of [36]–[38], while using many more variables. The power of agent-based modeling only reveals itself when we implement more interesting interaction rules. For instance, it would be easy to assign each agent a position, and make two agents more likely to interact if they are close. We could add visible symptoms that are imperfectly associated with

the disease, and a tendency not to interact with symptomatic individuals. We could make the degree of aversion to symptomatic agents part of the agents' state. All of this is easy to implement in the model, even in combination, but *not* easy to do in a more traditional, aggregated model. Sometimes it would be all but impossible; an excellent case in point is the highly sophisticated model of HIV epidemiology produced by Jacquez, Koopman, Simon, and collaborators (129,130), incorporating multiple routes of transmission, highly non-random mixing of types, and time-varying infectiousness.

Agent-based models steer you towards representing individuals, their behaviors and their interactions, rather than aggregates and their dynamics. Whether this is a good thing depends, of course, on what you know, and what you hope to learn. If you know a lot about individuals, agent-based models can help you leverage that knowledge into information about collective dynamics. This is particularly helpful if the population is heterogeneous, since you can represent the different types of individuals in the population by different states for agents. This requires a bit of effort on your part, but often not nearly so much as it would to represent the heterogeneity in an aggregated model. Conversely, if you think you have the collective dynamics down, an ABM will let you check whether a candidate for an individual-level mechanism really will produce them. (But see §6, below.)

Ideally, there are no "mass nouns" in an ABM, nothing represented by a smeared-out "how much": everything should be represented by some definite number of distinctly located agents. At most, some aggregate variables may be stuffed into the environment part of the model, but only simple and homogeneous ones. Of course, the *level* of disaggregation at which it is useful to call something an agent is a matter for particular applications, and need not be the same for every agent in a model. (E.g., one might want to model an entire organ as a single agent, while another, more interesting organ is broken up into multiple interacting agents, along anatomical or functional lines.) Sometimes it's just not practical to represent everything which we know is an individual thing by its own agent: imagine trying to do chemical thermodynamics by tracking the interactions of a mole of molecules. Such cases demand either giving up on agent-based modeling (fortunately, the law of mass action works pretty well in chemistry), or using fictitious agents that represent substantial, but not too large, collections of individuals.

Models describing the collective dynamics of aggregate variables are sometimes called "equation-based models," in contrast to agent-based models. This is sloppy, however: it is always possible, though generally tedious and unilluminating, to write down a set of equations that describe the dynamics of an agent-based model. Rather than drawing a false contrast between agents and equations, it would be better to compare ABMs to "aggregate models," "collective models," or perhaps "factor models."

5.1. Computational Implementation: Agents are Objects

The nicest way to computationally implement the commitment of distinctly representing each agent is to make agents **objects**, which are, to oversimplify slightly, data structures that have internal states, and interact with each other by passing messages. While objects are not necessary for agent-based models, they do make programming them *much* easier, especially if the agents have much more state than, say, just a position and a type. If you try to implement models with sophisticated agents without using objects, the odds are good that you will find yourself reinventing well-known features of object-oriented programming. (Historically, object-oriented programming *began* with languages for simulation modeling (131).) You might as well save your time, and do those things *right*, by using objects in the first place.

Generally speaking, computational implementations of ABMs contain many non-agent objects, engaged in various housekeeping tasks, or implementing the functions agents are supposed to perform. For instance, an agent, say a rat, might be supposed to memorize a sequence, say, of turns in a maze. One way of implementing this would be to use a linked list, which is an object itself. Such objects do not represent actual features of the *model*, and it should be possible to vary them without interfering with the model's behavior. Which objects are picked out as agents is to some degree a matter of convenience and taste. It is common, for instance, to have mobile agents interacting on astatic environment. If the environment is an object, modelers may or may not speak of it as an "environment agent," and little seems to hinge on whether or not they do.

There are several programming environments designed to facilitate agent-based modeling. Perhaps the best known of these is (www.swarm.org), which works very flexibly with several languages, is extensively documented, and has a large user community, though it presently (2004) lacks an institutional home. REPAST, while conceptually similar, is open-source (repast.sourceforge.net) and is associated with the University of Chicago. STARLOGO, and its successor, NETLOGO (ccl.sesp.northwestern.edu/netlogo), are extensions of the popular LOGO language to handle multiple interacting "turtles," i.e., agents. Like Logo, children can learn to use them (132), but they are fairly easy for adults, too, and certainly give a feel for working with ABMs.

5.2. Three Things Which Are Not Agent-Based Models

Not everything which involves the word "agent" is connected to agent-based modeling.

Representative agent models are not ABMs. In these models, the response of a population to environmental conditions is found by picking out a *single* typical or representative agent, determining its behavior, and assuming that eve-

ryone else does likewise. This is sometimes reasonable, but it's clearly diametrically opposed to what an ABM is supposed to be.

Software agents are not ABMs. Software agents are a very useful and rapidly developing technology (133, ch. 2); an agent, here, is roughly a piece of code that interacts with other software and with pieces of the real world autonomously. Agents index the Web for search engines, engage in automated trading, and help manage parts of the North American electrical power grid, among other things. Some agent software systems are *inspired* by ABMs (134). When one wants to model their behavior, an ABM is a natural tool (but not the only one by any means: see (135)). But a set of software agents running the Michigan power grid is not a *model* of anything, it's *doing* something.

Finally, **multi-agent systems** (136) and **rational agents** (137) in artificial intelligence are not ABMs. The interest of this work is in understanding, and especially *designing*, systems capable of sophisticated, autonomous cognitive behavior; many people in this field would restrict the word "agent" to apply only to things capable, in some sense, of having "beliefs, desires, and intentions." While these are certainly complex systems, they are not usually intended to be *models* of anything else. One can, of course, press them into service as models (138), but generally this will be no more than a heuristic device.

5.3. The Simplicity of Complex Systems Models

One striking feature of agent-based models, and indeed of complex systems models in general, is how *simple* they are. Often, agents have only a few possible states, and only a handful of kinds of interaction. This practice has three motivations: (i) A model as detailed as the system being studied would be as hard to understand as that system. (ii) Many people working in complex systems science want to show that a certain set of mechanisms are sufficient to generate some phenomenon, like cooperation among unrelated organisms, or the formation of striped patterns. Hence using simple models, which contain only those mechanisms, makes the case. (iii) Statistical physicists, in particular, have a long tradition of using highly simplified models as caricatures of real systems.

All three motives are appropriate, in their place. (i) is completely unexceptionable; abstracting away from irrelevant detail is always worthwhile, so long as it really is irrelevant. (ii) is also fair enough, though one should be careful that the mechanisms in one's model can still generate the phenomenon when they interact with *other* effects as well. (iii) works very nicely in statistical physics itself, where there are powerful mathematical results relating to the renormalization group (139) and bifurcation theory (14), which allow one to extract certain kinds of *quantitative* results from simplified models that share certain *qualitative* characteristics with real systems. (We have seen a related principle when discussing cellular automata models above.) There is, however, little reason to

think that these universality results apply to most complex systems, let alone ones with adaptive agents!

6. EVALUATING MODELS OF COMPLEX SYSTEMS

We do not build models for their own sake; we want to see what they do, and we want to compare what they do both to reality and to other models. This kind of evaluation of models is a problem for all areas of science, and as such little useful general advice can be given. However, there are some issues that are peculiar to models of complex systems, or especially acute for them, and I will try to provide some guidance here, moving from figuring out just what your model does, to comparing your model to data, to comparing it to other models.

6.1. Simulation

The most basic way to see what your model does is to run it; to do a simulation. Even though a model is entirely a human construct, every aspect of its behavior following logically from its premises and initial conditions, the frailty of human nature is such that we generally cannot perceive those consequences, not with any accuracy. If the model involves a large number of components that interact strongly with each other—if, that is to say, it's a good model of a complex system—our powers of deduction are generally overwhelmed by the mass of relevant, interconnected detail. Computer simulation then comes to our aid, because computers have no trouble remembering large quantities of detail, nor in following instructions.

6.1.1. Direct Simulation

Direct simulation—simply starting the model and letting it go—has two main uses. One is to get a sense of the typical behavior, or of the range of behavior. The other, more quantitative, use is to determine the distribution of important quantities, including time series. If one randomizes initial conditions, and collects data over multiple runs, one can estimate the distribution of desired quantities with great accuracy. This is exploited in the time-series method of surrogate data (above), but the idea applies quite generally.

Individual simulation runs for models of complex systems can be reasonably expensive in terms of time and computing power; large numbers of runs, which are really needed to have confidence in the results, are correspondingly more costly. Few things are more dispiriting than to expend such quantities of time and care, only to end up with ambiguous results. It is almost always

worthwhile, therefore, to carefully think through what you want to measure, and why, before running anything. In particular, if you are trying to judge the merits of competing models, effort put into figuring out how and where they are *most* different will generally be well-rewarded. The theory of experimental design offers extensive guidance on how to devise informative series of experiments, both for model comparison and for other purposes, and by and large the principles apply to simulations as well as to real experiments.

6.1.2. Monte Carlo Methods

Monte Carlo is the name of a broad, slightly indistinct family for using random processes to estimate deterministic quantities, especially the properties of probability distributions. A classic example will serve to illustrate the basic idea, on which there are many, many refinements.

Consider the problem of determining the area A under an curve given by a known but irregular function $f(x)$. In principle, you could integrate f to find this area, but suppose that numerical integration is infeasible for some reason. (We will come back to this point presently.) A Monte Carlo solution to this problem is as follows: pick points at random, uniformly over the square. The probability p that a point falls in the shaded region is equal to the fraction of the square occupied by the shading: $p = A/s^2$. If we pick n points independently, and x of them fall in the shaded region, then $x/n \to p$ (by the law of large numbers), and $s^2x/n \to A$. s^2x/n provides us with a stochastic estimate of the integral. Moreover, this is a probably approximately correct (§2.1.3) estimate, and we can expect, from basic probability theory, that the standard deviation of the estimate around its true value will be proportional to $n^{-1/2}$, which is not bad.[16] However, when faced with such a claim, one should always ask what the proportionality constant is, and whether it is the best achievable. Here it is not: the equally simple, if less visual, scheme of just picking values of x uniformly and averaging the resulting values of $f(x)$ always has a smaller standard deviation (140, ch. 5).

This example, while time-honored and visually clear, does not show Monte Carlo to its best advantage; there are few one-dimensional integrals that cannot be done better by ordinary, non-stochastic numerical methods. But numerical integration becomes computationally intractable when the domain of integration has a large number of dimensions, where "large" begins somewhere between four and ten. Monte Carlo is much more indifferent to the dimensionality of the space: we could replicate our example with a 999-dimensional hypersurface in a 1000-dimensional space, and we'd still get estimates that converged like $n^{-1/2}$, so achieving an accuracy of $\pm\varepsilon$ will require evaluating the function f only $O(\varepsilon^{-2})$ times.

Our example was artificially simple in another way, in that we used a uniform distribution over the entire space. Often, what we want is to compute the

expectation of some function $f(x)$ with a nonuniform probability $p(x)$. This is just an integral, $\int f(x)p(x)dx$, so we could sample points uniformly and compute $f(x)p(x)$ for each one. But if some points have very low probability, so they only make a small contribution to the integral, spending time evaluating the function there is a bit of a waste. A better strategy would be to pick points according to the actual probability distribution. This can sometimes be done directly, especially if $p(x)$ is of a particularly nice form. A very general and clever indirect scheme is as follows (14). We want a whole sequence of points, $x_1, x_2, \ldots x_n$. We pick the first one however we like, and after that we pick successive points according to some Markov chain: that is, the distribution of x_{i+1} depends only on x_i, according to some fixed function $q(x_i, x_{i+1})$. Under some mild conditions,[17] the distribution of x_t approaches a stationary distribution $q^*(x)$ at large times t. If we could ensure that $q^*(x) = p(x)$, we would know that the Markov chain was converging to our distribution, and then, by the ergodic theorem, averaging $f(x)$ along a trajectory would give the expected value of $f(x)$. One way to ensure this is to use the "detailed balance" condition of the invariant distribution, that the total probability of going from x to y must equal the total probability of going the other way:

$$p(x)q(x,y) = p(y), \tag{39}$$

$$\frac{q(x,y)}{q(y,x)} = \frac{p(y)}{p(x)} \equiv h(x,y). \tag{40}$$

So now we just need to make sure that [40] is satisfied. One way to do this is to set $q(x,y) = \min(1,h(x,y))$; this was the original proposal of Metropolis et al. (141). Another is $q(x,y) = (h(x,y))/(1 + h(x,y))$. This method is what physicists usually mean by "Monte Carlo," but statisticians call it **Markov chain Monte Carlo**, or "MCMC." While we can now estimate the properties of basically arbitrary distributions, we no longer have independent samples, so evaluating the accuracy of our estimates is no longer a matter of *trivial* probability.[18] An immense range of refinements have been developed over the last fifty years, addressing these and other points; see the further reading section for details.

Keep in mind that Monte Carlo is a stochastic simulation method only in a special sense—it simulates the probability distribution $p(x)$, *not* the mechanism that generated that distribution. The dynamics of Markov chain Monte Carlo, in particular, often bear no resemblance whatsoever to those of the real system.[19] Since the point of Monte Carlo is to tell us about the properties of $p(x)$ (what is the expectation value of this function? what is the probability of configurations with this property? etc.), the actual trajectory of the Markov chain is of no interest. This point sometimes confuses those more used to direct simulation methods.

6.2. Analytical Techniques

Naturally enough, analytical techniques are not among the tools that first come to mind for dealing with complex systems; in fact, they often do not come to mind at all. This is unfortunate, because a lot of intelligence has been devoted to devising approximate analytical techniques for classes of models that include many of those commonly used for complex systems. A general advantage of analytical techniques is that they are often fairly insensitive to many details of the model. Since any model we construct of a complex system is almost certainly much simpler than the system itself, a great many of its details are just wrong. If we can extract nontrivial results insensitive to those details, we have less reason to worry about this.

One particularly useful, yet neglected, body of approximate analytical techniques relies on the fact that many complex systems models are Markovian. In an agent-based model, for instance, the next state of an agent generally depends only on its present state, and the present states of the agents it interacts with. If there is a fixed interaction graph, the agents form a Markov random field on that graph. There are now very powerful and computationally efficient methods for evaluating many properties of Markov chains (58,142), Markov random fields (143), and (closely related) graphical models (144) *without* simulation. The recent books of Peyton Young (145) and Sutton (146) provide nice instances of using analytical results about Markov processes to solve models of complex social systems, without impractical numerical experiments.

6.3. Comparisons with Data

6.3.1. General Issues

We can only compare particular aspects of a model of a system to particular kinds of data about that system. The most any experimental test can tell us, therefore, is how similar the model is to the system *in that respect*. One may think of an experimental comparison as a test for a *particular* kind of *error*, one of the infinite number of mistakes which we could make in building a model. A good test is one which is very likely to alert us to an error, if we have made it, but not otherwise (50).

These ought to be things every schoolchild knows about testing hypotheses. It is very easy, however, to blithely ignore these truisms when confronted with, on the one hand, a system with many strongly interdependent parts, and, on the other hand, a model that tries to mirror that complexity. We must decide which features of the model *ought* to be similar to the system, and how similar. It is important not only that our model be able to adequately reproduce those phe-

nomena, but that it not entail badly distorted or nonexistent phenomena in other respects.

6.3.2. Two Stories and Some Morals

Let me give two examples from very early in the study of complex systems, which nicely illustrate some fundamental points.

The first has to do with pattern formation in chemical oscillators (147). Certain mixtures of chemicals in aqueous solution, most famously the Belusov-Zhabotinsky reagent, can not only undergo cyclic chemical reactions, but will form rotating spiral waves, starting from an initial featureless state. This is a visually compelling example of self-organization, and much effort has been devoted to understanding it. One of the more popular early models was the "Brusselator" advanced by Prigogine and his colleagues at the Free University of Brussels; many similarly named variants developed. Brusselator-type models correctly predicted that these media would support spiral waves. They all, further, predicted that the spirals would form only when the homogeneous configuration was unstable, and that then they would form spontaneously. It proved very easy, however, to prepare the Belusov-Zhabotisnky reagent in such a way that it was "perfectly stable in its uniform quiescence," yet still able to produce spiral waves if excited (e.g., by being touched with a hot wire) (148). The Brusselator and its variants were simply unable to accommodate these phenomena, and had to be discarded in favor of other models. The fact that these were qualitative results, rather than quantitative ones, if anything made it more imperative to get rid of the Brusselator.

The second story concerns the work of Varela and Maturana on "autopoesis." In a famous paper (149), they claimed to exhibit a computational model of a simple artificial chemistry where membranes not only formed spontaneously, but a kind of metabolism self-organized to sustain the membranes. This work influenced not just complex systems science but theoretical biology, psychology, and even sociology (150). When, in the 1990s, McMullin made the first serious effort to reproduce the results, based on the description of the model in the paper, that description proved *not* to match the published simulation results. The discrepancy was only resolved by the fortuitous rediscovery of a mass of papers, including Fortran code, that Varela had left behind in Chile when forced into exile by the fascist regime. These revealed a crucial change in one particular reaction made all the difference between successful autopoesis and its absence. (For the full story, see (151,152).) Many similar stories could be told of other models in complex systems (153); this one is distinguished by McMullin's unusual tenacity in trying to replicate the results, Varela's admirable willingness to assist him, and the happy ending.

The story of autopoesis is especially rich in morals. (1) Replication is essential. (2) It is a good idea to share not just data but programs. (3) *Always* test the robustness of our model to changes in its parameters. (This is fairly common.) (4) *Always* test your model for robustness to small changes in qualitative assumptions. If your model calls for a given effect, there are usually several mechanisms that could accomplish it. If it does not matter which mechanism you actually use, the result is that much more robust. Conversely, if it does matter, the overall adequacy of the model can be tested by checking whether *that* mechanism is actually present in the system. Altogether too few people perform such tests.

6.3.3. Comparing Macro-data and Micro-models

Data are often available only about large aggregates, while models, especially agent-based models, are about individual behavior. One way of comparing such models to data is to compute the necessary aggregates, from direct simulation, Monte Carlo, etc. The problem is that many different models can give the same aggregated behavior, so this does not provide a powerful test between different models. Ideally, we'd work back from aggregate data to individual behaviors, which is known, somewhat confusingly, as **ecological inference**. In general, the ecological inference problem itself does not have a unique solution. But the aggregate data, if used intelligently, can often put fairly tight constraints on the individual behaviors, and micro-scale can be directly checked against those constraints. Much of the work here has been done by social scientists, especially American political scientists concerned with issues arising from the Voting Rights Act (154), but the methods they have developed are very general, and could profitably be applied to agent-based models in the biological sciences, though, to my knowledge, they have yet to be.

6.4. Comparison to Other Models

Are there other ways of generating the data? There generally are, at least if "the data" are some very gross, highly summarized pattern. This makes it important to look for differential signatures, places where discrepancies between different generative mechanisms give one some *leverage*. Given two mechanisms that can both account for our phenomenon, we should look for some *other* quantity whose behavior will be different under the two hypotheses. Ideally, in fact, we would look for the statistic on which the two kinds of model are *most* divergent. The literature on experimental design is relevant here again, since it considers such problems under the heading of **model discrimination**, seeking to

maximize the power of experiments (or simulations) to distinguish between different classes of models (155,156).

Perhaps no aspect of methodology is more neglected in complex systems science than this one. While it is always perfectly legitimate to announce a new mechanism as *a* way of generating a phenomenon, it is far too common for it to be called *the* way to do it, and vanishingly rare to find an examination of how it *differs* from previously proposed mechanisms. Newman and Palmer's work on extinction models (157) stands out in this regard for its painstaking examination of the ways of discriminating between the various proposals in the literature.

7. INFORMATION THEORY

Information theory began as a branch of communications engineering, quantifying the length of codes needed to represent randomly varying signals, and the rate at which data can be transmitted over noisy channels. The concepts needed to solve these problems turn out to be quite fundamental measures of the uncertainty, variability, and the interdependence of different variables. Information theory thus is an important tool for studying complex systems, and in addition is indispensable for understanding complexity measures (§8).

7.1. Basic Definitions

Our notation and terminology follows that of Cover and Thomas's standard textbook (158).

Given a random variable X taking values in a discrete set $\mathcal{A}$, the **entropy** or **information content** $H[X]$ of X is

$$H[X] \equiv -\sum_{a \in \mathcal{A}} \Pr(X = a) \log_2 \Pr(X = a) \, . \qquad [41]$$

$H[X]$ is the expectation value of $-\log_2 \Pr(X)$. It represents the uncertainty in X, interpreted as the mean number of binary distinctions (bits) needed to identify the value of X. Alternately, it is the minimum number of bits needed to encode or describe X. Note that $H[X] = 0$ if and only if X is (almost surely) constant.

The **joint entropy** $H[X,Y]$ of two variables X and Y is the entropy of their joint distribution:

$$H[X,Y] \equiv -\sum_{a \in \mathcal{A}, b \in \mathcal{B}} \Pr(X = a, Y = b) \log_2 \Pr(X = a, Y = b) \, . \qquad [42]$$

The **conditional entropy** of X given Y is

$$H[X|Y] \equiv H[X,Y] - H[Y]. \quad [43]$$

$H[X|Y]$ is the average uncertainty remaining in X, given a knowledge of Y.

The **mutual information** $I[X;Y]$ between X and Y is

$$I[X;Y] \equiv H[X] - H[X|Y]. \quad [44]$$

It gives the reduction in X's uncertainty due to knowledge of Y and is symmetric in X and Y. We can also define higher-order mutual informations, such as the third-order information $I[X;Y;Z]$,

$$I[X;Y;Z] \equiv H[X] + H[Y] + H[Z] - H[X,Y,Z], \quad [45]$$

and so on for higher orders. These functions reflect the joint dependence among the variables.

Mutual information is a special case of the **relative entropy**, also called the **Kullback-Leibler divergence** (or **distance**). Given two *distributions* (not variables), P and Q, the entropy of Q relative to P is

$$D(\mathrm{P} \parallel \mathrm{Q}) \equiv \sum_x P(x) \log \frac{\mathrm{P}(x)}{\mathrm{Q}(x)}. \quad [46]$$

D measures how far apart the two distributions are, since $D(\mathrm{P}\|\mathrm{Q}) \geq 0$, and $D(\mathrm{P}\|\mathrm{Q}) = 0$ implies the two distributions are equal almost everywhere. The divergence can be interpreted either in terms of codes (see below), or in terms of statistical tests (159). Roughly speaking, given n samples drawn from the distribution P, the probability of our accepting the false hypothesis that the distribution is Q can go down no faster than $2^{-n\mathrm{D}(\mathrm{P}\|\mathrm{Q})}$. The mutual information $I[X;Y]$ is the divergence between the joint distribution $\Pr(X,Y)$, and the product of the marginal distributions, $\Pr(X)\Pr(Y)$, and so measures the departure from independence.

Some extra information-theoretic quantities make sense for time series and stochastic processes. Supposing we have a process $\overline{X} = ...,X_{-2},X_{-1},X_0,X_1,X_2,...$, we can define its **mutual information function** by analogy with the autocovariance function (see §3.2),

$$I_{\overline{X}}(s,t) = I[X_s; X_t], \quad [47]$$

$$I_{\overline{X}}(\tau) = I[X_t; X_{t+\tau}], \quad [48]$$

where the second form is valid only for strictly stationary processes. The mutual information function measures the degree to which different parts of the series are dependent on each other.

The **entropy rate** h of a stochastic process is

$$h \equiv \lim_{L \to \infty} H[X_0 \mid X_{-L}^{-1}], \quad [49]$$

$$= H[X_0 \mid X_{-\infty}^{-1}]. \quad [50]$$

(the limit always exists for stationary processes), where h measures the process's unpredictability, in the sense that it is the uncertainty which remains in the next measurement even given complete knowledge of its past. In nonlinear dynamics, h is called the **Kolmogorov-Sinai (KS) entropy**.

For continuous variables, one can define the entropy via an integral,

$$H[X] \equiv -\int p(x) \log p(x) dx, \quad [51]$$

with the subtlety that the continuous entropy not only can be negative, but depends on the coordinate system used for x. The relative entropy also has the obvious definition,

$$D(\mathrm{P} \,\|\, \mathrm{Q}) \equiv \int p(x) \log \frac{p(x)}{q(x)} dx, \quad [52]$$

but is coordinate-independent and non-negative. So, hence, is the mutual information.

Optimal Coding. One of the basic results of information theory concerns codes, or schemes for representing random variables by bit strings. That is, we want a scheme that associates each value of a random variable X with a bit string. Clearly, if we want to keep the average length of our code-words small, we should give shorter codes to the more common values of X. It turns out that the average code-length is minimized if we use $-\log \Pr(x)$ bits to encode x, and it is always possible to come within one bit of this. Then, on average, we will use $\mathbf{E}[-\log \Pr(x)] = H[X]$ bits.

This presumes we know the true probabilities. If we think the true distribution is Q when it is really P, we will, on average, use $\mathbf{E}[-\log \mathrm{Q}(x)] \geq H[X]$. This quantity is called the **cross-entropy** or **inaccuracy**, and is equal to $H[X] + D(\mathrm{P}\|\mathrm{Q})$. Thus, finding the correct probability distribution is equivalent to minimizing the cross-entropy, or the relative entropy (160).

The Khinchin Axioms and Rényi Information. In 1953, A.I. Khinchin published a list of four reasonable-looking axioms fora measure of the information $H[X]$ associated with a random variable X (161). He then proved that the Shannon information was the unique functional satisfying the axioms, up to an overall multiplicative constant. (The choice of this constant is equivalent to the choice of the base for logarithms.) The axioms were as follows.

- The information is a functional of the probability distribution of X, and not on any of its other properties. In particular, if f is any invertible function, $H[X] = H[f(X)]$.
- The information is maximal for the uniform distribution, where all events are equally probable.
- The information is unchanged by enlarging the probability space with events of zero probability.
- If the probability space is divided into two subspaces, so that X is split into two variables Y and Z, the total information is equal to the information content of the marginal distribution of one subspace, plus the mean information of the conditional distribution of the other subspace: $H[X] = H[Y] + \mathbf{E}[H(Z|Y)]$.

A similar axiomatic treatment can be given for the mutual information and the relative entropy.

While the first three of Khinchin's axioms are all highly plausible, the fourth is somewhat awkward. It is intuitively more plausible to merely require that, if Y and Z are independent, then $H[Y,Z] = H[Y] + H[Z]$. If the fourth axiom is weakened in this way, however, there is no longer only a single functional satisfying the axioms. Instead, any of the infinite family of entropies introduced by Rényi satisfies the axioms. The **Rényi entropy** of order α, with α any non-negative real number, is

$$H_\alpha[X] \equiv \frac{1}{1-\alpha} \log \sum_{i:p_i>0} p_i^\alpha \qquad [53]$$

in the discrete case, and the corresponding integral in the continuous case. The parameter α can be thought of as gauging how strongly the entropy is biased towards low-probability events. As $\alpha \to 0$, low-probability events count more, until at $\alpha = 0$, all possible events receive equal weight. (This is sometimes called the **topological entropy**.) As $\alpha \to \infty$, only the highest-probability event contributes to the sum. One can show that, as $\alpha \to 1$, $H_\alpha[X] \to H[X]$, i.e., one recovers

the ordinary Shannon entropy in the limit. There are entropy rates corresponding to all the Rényi entropies, defined just like the ordinary entropy rate. For dynamical systems, these are related to the fractal dimensions of the attractor (162,163).

The **Rényi divergences** bear the same relation to the Rényi entropies as the Kullback-Leibler divergence does to the Shannon entropy. The defining formula is

$$D_\alpha(\mathrm{P} \parallel \mathrm{Q}) \equiv \frac{1}{\alpha - 1} \log \sum q_i \left(\frac{p_i}{q_i} \right)^\alpha , \qquad [54]$$

and similarly for the continuous case. Once again, $\lim_{\alpha \to 1} D_\alpha(\mathrm{P}\|\mathrm{Q}) = D(\mathrm{P}\|\mathrm{Q})$. For all $\alpha > 0$, $D_\alpha(\mathrm{P}\|\mathrm{Q}) \geq 0$, and is equal to zero if and only if P and Q are the same. (If $\alpha = 0$, then a vanishing Rényi divergence only means that the supports of the two distributions are the same.) The Rényi entropy $H_\alpha[X]$ is nonincreasing as α grows, whereas the Rényi divergence $D_\alpha(\mathrm{P}\|\mathrm{Q})$ is nondecreasing.

Estimation of Information-Theoretic Quantities. In applications, we will often want to estimate theoretic quantities, such as the Shannon entropy or the mutual information, from empirical or simulation data. Restricting our attention, for the moment, to the case of discrete-valued variables, the empirical distribution will generally converge on the true distribution, and so the entropy (say) of the empirical distribution ("sample entropy") will also converge on the true entropy. However, it is not the case that the sample entropy is an *unbiased* estimate of the true entropy. The Shannon (and Rényi) entropies are measures of variation, like the variance, and sampling tends to reduce variation. Just as the sample variance is a negatively biased estimate of the true variance, sample entropy is a negatively biased estimate of the true entropy, and so sample mutual information is a positively biased estimate of true information. Understanding and controlling the bias, as well as the sampling fluctuations, can be very important.

Victor (164) has given an elegant method for calculating the bias of the sample entropy; remarkably, the leading-order term depends only on the alphabet size k and the number of samples N, and is $(k-1)/2N$. Higher-order terms, however, depend on the true distribution. Recently, Kraskov et al. (165) have published an adaptive algorithm for estimating mutual information, which has very good properties in terms of both bias and variance. Finally, the estimation of entropy *rates* is a somewhat tricky matter. The best practices are to either use an algorithm of the type given by (166), or to fit a properly dynamical model. (For discrete data, variable-length Markov chains, discussed in §3.6.2 above, generally work very well, and the entropy rate can be calculated from them very simply.) Another popular approach is to run one's time series through a standard compression algorithm, such as `gzip`, dividing the size in bits of the output by

the number of symbols in the input (167). This is an absolutely horrible idea; even under the circumstances under which it gives a consistent estimate of the entropy rate, it converges much more slowly, and runs more slowly, than employing either of the two techniques just mentioned (168,169).[20]

7.2. Applications of Information Theory

Beyond its original home in communications engineering, information theory has found a multitude of applications in statistics (159,160) and learning theory (144,170). Scientifically, it is very natural to consider some biological systems as communications channels, and so analyze their information content; this has been particularly successful for biopolymer sequences (171) and especially for neural systems, where the analysis of neural codes depends vitally on information theory (172,173). However, there is nothing prohibiting the application of information theory to systems that are not designed to function as communications devices; the concepts involved require only well-defined probability distributions. For instance, in nonlinear dynamics (174,175) information-theoretic notions are very important in characterizing different kinds of dynamical system (see also §3.6). Even more closely tied to complex systems science is the literature on "physics and information" or "physics and computation," which investigates the relationships between the mechanical principles of physics and information theory, e.g., Landauer's principle, that erasing (but not storing) a bit of information at temperature T produces $k_B T \ln 2$ joules of heat, where k_B is Boltzmann's constant.

8. COMPLEXITY MEASURES

We have already given some thought to complexity, both in our initial rough definition of "complex system" and in our consideration of machine learning and Occam's Razor. In the latter, we saw that the relevant sense of"complexity" has to do with families of models: a model class is complex if it requires large amounts of data to reliably find the best model in the class. On the other hand, we initially said that a complex system is one with many highly variable, strongly interdependent parts. Here, we will consider various proposals for putting some mathematical spine into that notion of a system's complexity, as well as the relationship to the notion of complexity of learning.

Most measures of complexity for systems formalize the intuition that something is complex if it is difficult to describe adequately. The first mathematical theory based on this idea was proposed by Kolmogorov; while it is *not* good for analyzing empirical complex systems, it was very important historically, and makes a good point of entry into the field.

8.1. Algorithmic Complexity

Consider a collection of measured data-values, stored in digitized form on a computer. We would like to say that they are complex if they are hard to describe, and measure their complexity by the difficulty of describing them. The central idea of Kolmogorov complexity (proposed independently by Solomonoff (176) and Chaitin) is that one can describe the data set by writing a program which will reproduce the data. The difficulty of description is then measured by the length of the program. Anyone with much experience of other people's code will appreciate that it is always possible to write a longer, slower program to do a given job, so what we are really interested in is the shortest program that can exactly replicate the data.

To introduce some symbols, let x be the data, and $|x|$ their size in bits. The Kolmogorov or algorithmic complexity of x, $K(x)$, is the length of the shortest program that will output x and then stop.[21] Clearly, there is always some program which will output x and then stop, for instance, "`print(`x`); end`." Thus $K(x) \leq |x| + c$, where c is the length of the print and end instructions. This is what one might call a literal description of the data. If one cannot do better than this—if $K(x) \approx |x|$—then x is highly complex. Some data, however, is highly compressible. For instance, if x consists of the second quadrillion digits of π, a very short program suffices to generate it.[22]

As you may already suspect, the number of simple data sets is quite limited. Suppose we have a data set of size n bits, and we want to compress it by k bits, i.e., find a program for it which is $n - k$ bits long. There are at most 2^{n-k} programs of that length, so of all the 2^n data sets of size n, the fraction that can be compressed by k bits is at most 2^{-k}. The precise degree of compression does not matter—when we look at large data sets, almost all of them are highly complex. If we pick a large data set *at random*, then the odds are very good that it will be complex. We can state this more exactly if we think about our data as consisting of the first n measurements from some sequence, and let n grow. That is, $x = x_1^n$, and we are interested in the asymptotic behavior of $K(x_1^n)$. If the measurements x_i are independent and identically distributed (IID), then $K(x_1^n)/|x| \to 1$ almost surely; IID sequences are **incompressible**. If x is a realization of a stationary (but not necessarily IID) random process $\overline{X}$, then (177,10)

$$\lim_{n\to\infty} \mathbf{E}\left[\frac{K(X_1^n)}{n}\right] = h(\overline{X}), \qquad [55]$$

the entropy rate (§7) of $\overline{X}$. Thus, random data has high complexity, and the complexity of a random process grows at a rate that just measures its unpredictability.

This observation goes the other way: complex data look random. The heuristic idea is that if there were any regularities in the data, we could use them to shave at least a little bit off the length of the minimal program. What one can show formally is that incompressible sequences have *all* the properties of IID sequences—they obey the law of large numbers and the central limit theorem, pass all statistical tests for randomness, etc. In fact, this possibility, of defining "random" as "incompressible," is what originally motivated Kolmogorov's work (107, ch. 3).

Kolmogorov complexity is thus a very important notion for the foundations of probability theory, and it has extensive applications in theoretical computer science (177) and even some aspects of statistical physics (178). Unfortunately, it is quite useless as a measure of the complexity of natural systems. This is so for two reasons. First, as we have just seen, it is maximized by *independent* random variables; we want *strong dependence*. Second, and perhaps more fundamental, it is simply not possible to calculate Kolmogorov complexity. For deep reasons related to Gödel's Theorem, there cannot be any procedure for calculating $K(x)$, nor are there any accurate approximation procedures (177).

Many scientists are strangely in denial about the Kolmogorov complexity, in that they think they can calculate it. Apparently unaware of the mathematical results, but aware of the relationship between Kolmogorov complexity and data compression, they reason that file compression utilities should provide an estimate of the algorithmic information content. Thus one finds many papers which might be titled `gzip` as a measure of complexity,"[23] and the practice is even recommended by some otherwise-reliable sources (e.g., (73)). However, this is simply a confused idea, with absolutely nothing to be said in its defense.

8.2. Refinements of Algorithmic Complexity

We saw just now that algorithmic information is really a measure of randomness, and that it is maximized by collections of independent random variables. Since complex systems have many strongly dependent variables, it follows that the Kolmogorov notion is not the one we really want to measure. It has long been recognized that we really want something which is small both for systems which are strongly ordered (i.e., have only a small range of allowable behavior) and for those which are strongly disordered (i.e., have independent parts). Many ways of modifying the algorithmic information to achieve this have been proposed; two of them are especially noteworthy.

8.2.1. Logical Depth

Bennett (179–181) proposed the notion of the **logical depth** of data as a measure of its complexity. Roughly speaking, the logical depth $L(x)$ of x is the

number of computational steps the minimal program for x must execute. For incompressible data, the minimal program is `print`(x), so $L(x) \approx |x|$. For periodic data, the minimal program cycles over printing out one period over and over, so $L(x) \approx |x|$ again. For some compressible data, however, the minimal program must do nontrivial computations, which are time-consuming. Thus, to produce the second quadrillion digits of π, the minimal program is one that *calculates* the digits, and this takes considerably more time than reading them out of a list. Thus, π is deep, while random or periodic data are shallow.

While logical depth is a clever and appealing idea, it suffers from a number of drawbacks. First, real data are not, so far as we know, actually produced by running their minimal programs,[24] and the run-time of that program has no known *physical* significance, and that's not for lack of attempts to find one (182). Second, and perhaps more decisively, there is still no procedure for finding the minimal program.

8.2.2. Algorithmic Statistics

Perhaps the most important modification of the Kolmogorov complexity is that proposed by Gács, Tromp and Vitanyi (183), under the label of "algorithmic statistics." Observe that, when speaking of the minimal program for x, I said nothing about the inputs to the program; these are to be built in to the code. It is this which accounts for the length of the programs needed to generate random sequences: almost all of the length of `print`(x) comes from x, not `print`(). This suggests splitting the minimal program into two components, a "model" part, the program properly speaking, and a "data" part, the inputs to the program. We want to put all the regularities in x into the model, and all the arbitrary, noisy parts of x into the inputs. Just as in probability theory a "statistic" is a function of the data that summarizes the information they convey, Gács et al. regard the model part of the program as an **algorithmic statistic**, summarizing its regularities. To avoid the trivial regularity of `print`() when possible, they define a notion of a **sufficient** algorithmic statistic, based on the idea that x should be in some sense a typical output of the model (see their paper for details). They then define the complexity of x, or, as they prefer to call it, the **sophistication**, as the length of the shortest sufficient algorithmic statistic.

Like logical depth, sophistication is supposed to discount the purely random part of algorithmic complexity. Unlike logical depth, it stays within the confines of description in doing so; programs, here, are just a particular, mathematically tractable, kind of description. Unfortunately, the sophistication is still uncomputable, so there is no real way of applying algorithmic statistics.

8.3. Statistical Measures of Complexity

The basic problem with algorithmic complexity and its extensions is that they are all about finding the shortest way of exactly describing a single configuration. Even if we could compute these measures, we might suspect, on the basis of our discussion of over-fitting in §2 above, that this is not what we want. Many of the details of any particular set of data are just noise, and will not generalize to other data sets obtained from the same system. If we want to characterize the complexity of the system, it is precisely the generalizations that we want, and not the noisy particulars. Looking at the sophistication, we saw the idea of picking out, from the overall description, the part which describes the regularities of the data. This idea becomes useful and operational when we abandon the goal of *exact* description, and allow ourselves to recognize that the world is full of noise, which is easy to describe statistically; we want a statistical, and not an algorithmic, measure of complexity.

I will begin with what is undoubtedly the most widely used statistical measure of complexity, Rissanen's **stochastic complexity**, which can also be considered a method of model selection. Then I will look at three attempts to isolate the complexity of the system as such, by considering how much information would be required to predict its behavior, *if* we had an optimal statistical model of the system.

8.3.1. Stochastic Complexity and the Minimum Description Length

Suppose we have a statistical model with some parameter θ, and we observe the data x. The model assigns a certain likelihood to the data, $\Pr_\theta(X = x)$. One can make this into a loss function by taking its negative logarithm: $L(\theta,x) = -\log \Pr_\theta(X = x)$. Maximum likelihood estimation minimizes this loss function. We also learned, in §7, that if $\Pr_\theta$ is the correct probability distribution, the optimal coding scheme will use $-\log \Pr_\theta(X = x)$ bits to encode x. Thus, maximizing the likelihood can also be thought of as minimizing the encoded length of the data.

However, we do not yet have a complete description: we have an encoded version of the data, but we have not said what the encoding scheme, i.e., the model, is. Thus, the total description length has two parts:

$$C(x,\theta,\theta) = L(x,\theta) + D(\theta,\Theta), \qquad [56]$$

where $D(\theta,\Theta)$ is the number of bits we need to specify θ from among the set of all our models Θ. $L(x,\theta)$ represents the "noisy" or arbitrary part of the description, the one which will not generalize; the model represents the part which does generalize. If $D(\theta,\Theta)$ gives short codes to simple models, we have the desired kind of tradeoff, where we can reduce the part of the data that looks like noise

only by using a more elaborate model. The **minimum description length principle** (184,185) enjoins us to pick the model that minimizes the description length, and the **stochastic complexity** of the data is that minimized description-length:

$$\theta_{\mathrm{MDL}} = \arg\min_{\theta} C(x,\theta,\Theta)\,, \qquad [57]$$

$$\theta_{\mathrm{SC}} = \min_{\theta} C(x,\theta,\Theta)\,. \qquad [58]$$

Under not-too-onerous conditions on the underlying data-generating process and the model class Θ (185, ch. 3), as we provide more data θ_{MDL} will converge on the model in Θ that minimizes the generalization error, which here is just the same as minimizing the Kullback-Leibler divergence from the true distribution.[25]

Regarded as a principle of model selection, MDL has proved very successful in many applications, even when dealing with quite intricate, hierarchically layered model classes ((186) presents a nice recent application to a biomedical complex system; see §3.4 for applications to state-space reconstruction.) It is important to recognize, however, that most of this success comes from carefully tuning the model-coding term $D(\theta,\Theta)$ so that models that do not generalize well turn out to have long encodings. This is perfectly legitimate, but it relies on the tact and judgment of the scientist, and often, in dealing with a complex system, we have no idea, or at least no *good* idea, what generalizes and what does not. If we were malicious, or short-sighted, we could always ensure that the particular data we got have a stochastic complexity of just one bit.[26] The model that gives us this complexity will then have absolutely horrible generalization properties.[27]

Whatever its merits as a model selection method, stochastic complexity does not make a good measure of the complexity of natural systems. There are at least three reasons for this.

1. The dependence on the model-encoding scheme, already discussed.

2. The log-likelihood term, $L(x,\theta)$ in C_{SC} can be decomposed into two parts, one of which is related to the entropy rate of the data-generating process, and so reflects its intrinsic unpredictability. The other, however, indicates the degree to which even the most accurate model in θ is misspecified. Thus it reflects our ineptness as modelers, rather than any characteristic of the process.

3. Finally, the stochastic complexity reflects the need to specify some particular model, and to represent this specification.

> While this is necessarily a part of the modeling process for us, it seems to have no *physical* significance; the system does not need to *represent* its organization, it just *has* it.

8.3.2. Complexity via Prediction

Forecast Complexity and Predictive Information. Motivated in part by concerns such as these, Grassberger (187) suggested a new and highly satisfactory approach to system complexity: complexity is the amount of information required for optimal prediction. Let us first see why this idea is plausible, and then see how it can be implemented in practice. (My argument does not follow that of Grassberger particularly closely. Also, while I confine myself to time series, for clarity, the argument generalizes to any kind of prediction (188).)

We have seen that there is a limit on the accuracy of any prediction of a given system, set by the characteristics of the system itself (limited precision of measurement, sensitive dependence on initial conditions, etc.). Suppose we had a model that was maximally predictive, i.e., its predictions were at this limit of accuracy. Prediction, as I said, is always a matter of mapping inputs to outputs; here the inputs are the previous values of the time series. However, not all aspects of the entire past are relevant. In the extreme case of independent, identically distributed values, *no* aspects of the past are relevant. In the case of periodic sequences with period p, one only needs to know which of the p phases the sequence is in. If we ask how *much* information about the past is relevant in these two cases, the answers are clearly 0 and log p, respectively. If one is dealing with a Markov chain, only the present state is relevant, so the amount of information needed for optimal prediction is just equal to the amount of information needed to specify the current state. One thus has the nice feeling that both highly random (IID) and highly ordered (low-period deterministic) sequences are of low complexity, and more interesting cases can get high scores.

More formally, any predictor f will translate the past of the sequence x^{-} into an effective state, $s = f(\mathrm{x}^{-})$, and then make its prediction on the basis of s. (This is true whether f is formally a state-space model or not.) The amount of information required to specify the state is $H[\mathrm{S}]$. We can take this to be the complexity of f. Now, if we confine our attention to the set $\mathcal{M}$ of maximally predictive models, we can define what Grassberger called the "true measure complexity" or "forecast complexity" of the process as the minimal amount of information needed for optimal prediction:

$$C = \min_{f \in \mathcal{M}} H[f(X^{-})] \,. \qquad [59]$$

Grassberger did not provide a procedure for finding the maximally predictive models, nor for minimizing the information required among them. He did, however, make the following observation. A basic result of information theory, called the **data-processing inequality**, says that $I[A;B] \geq I[f(A);B]$, for any variables A and B—we cannot get more information out of data by processing it than was in there to begin with. Since the state of the predictor is a function of the past, it follows that $I[X^-;X^+] \geq I[f(X^-);X^+]$. Presumably, for optimal predictors, the two informations are equal—the predictor's state is just as informative as the original data. (Otherwise, the model would be missing some potential predictive power.) But another basic inequality is that $H[A] \geq I[A;B]$—no variable contains more information about another than it does about itself. So, for optimal models, $H[f(X^-)] \geq I[X^-;X^+]$. The latter quantity, which Grassberger called the **effective measure complexity**, can be estimated purely from data, without intervening models. This quantity—the mutual information between the past and the future—has been rediscovered many times, in many contexts, and called **excess entropy** (in statistical mechanics), **stored information** (189), **complexity** (190–192), or **predictive information** (193); the last name is perhaps the clearest. Since it quantifies the degree of statistical dependence between the past and the future, it is clearly appealing as a measure of complexity.

Grassberger-Crutchfield-Young Statistical Complexity. The forecasting complexity notion was made fully operational by Crutchfield and Young (102,194), who provided an effective procedure for finding the minimal maximally predictive model and its states. They began by defining the **causal states** of a process, as follows. For each history x^-, there is some conditional distribution of future observations, $\Pr(X^+|x^-)$. Two histories x_1^- and x_2^- are equivalent if $\Pr(X^+|x_1^-) = \Pr(X^+|x_2^-)$. Write the set of all histories equivalent to x^- as $[x^-]$. We now have a function ε that maps each history into its equivalence class: $\varepsilon(x^-) = [x^-]$. Clearly, $\Pr(X^+|\varepsilon(x^-)) = \Pr(X^+|x^-)$. Crutchfield and Young accordingly proposed to forget the particular history and retain only its equivalence class, which they claimed would involve no loss of predictive power; this was later proved to be correct (195, theorem 1). They called the equivalence classes the "causal states" of the process, and claimed which these were the simplest states with maximal predictive power; this is also was right (195, theorem 2). Finally, one can show that the causal states are the *unique* optimal states (195, theorem 3); any other optimal predictor is really a disguised version of the causal states. Accordingly, they defined the **statistical complexity** of a process C as the information content of its causal states.

Because the causal states are purely an objective property of the process being considered, C is too; it does not depend at all on our modeling or means of description. It is equal to the length of the shortest description of the past that is *relevant* to the actual dynamics of the system. As we argued should be the case above, for IID sequences it is exactly 0, and for periodic sequences it is $\log p$.

One can show (195, theorems 5 and 6) that the statistical complexity is always at least as large as the predictive information, and generally that it measures how far the system departs from statistical independence.

The causal states have, from a statistical point of view, quite a number of desirable properties. The maximal prediction property corresponds exactly to that of being a sufficient statistic (159); in fact they are minimal sufficient statistics (159,165). The sequence of states of the process form a Markov chain. Referring back to our discussion of filtering and state estimation (§3.5), one can design a recursive filter that will eventually estimate the causal state without any error at all; moreover, it is always clear whether the filter has "locked on" to the correct state or not.

All of these properties of the causal states and the statistical complexity extend naturally to spatially extended systems, including, but not limited to, cellular automata (196,197). Each point in space then has its own set of causal states, which form not a Markov chain but a Markov field, and the local causal state is the minimal sufficient statistic for predicting the future of that point. The recursion properties carry over, not just temporally but spatially: the state at one point, at one time, helps determine not only the state at that same point at later times, but also the state at neighboring points at the same time. The statistical complexity, in these spatial systems, becomes the amount of information needed about the past of a given point in order to optimally predict its future. Systems with a high degree of local statistical complexity are ones with intricate spatio-temporal organization, and, experimentally, increasing statistical complexity gives a precise formalization of intuitive notions of self-organization (197).

Crutchfield and Young were inspired by analogies to the theory of abstract automata, which led them to call their theory, somewhat confusingly, **computational mechanics**. Their specific initial claims for the causal states were based on a procedure for deriving the minimal automaton capable of producing a given family of sequences[28] known as Nerode equivalence classing (198). In addition to the theoretical development, the analogy to Nerode equivalence-classing led them to describe a procedure (102) for estimating the causal states and the ε-machine from empirical data, at least in the case of discrete sequences. This Crutchfield-Young algorithm has actually been successfully used to analyze empirical data, for instance, geomagnetic fluctuations (199). The algorithm has, however, been superseded by a newer algorithm that uses the known properties of the causal states to guide the model discovery process (105) (see §3.6.3 above).

Let me sum up. The Grassberger-Crutchfield-Young statistical complexity is an objective property of the system being studied. This reflects the *intrinsic* difficulty of predicting it, namely the amount of information that is actually relevant to the system's dynamics. It is low both for highly disordered and trivially ordered systems. Above all, it is calculable, and has actually been calculated for a range of natural and mathematical systems. While the initial formulation

was entirely in terms of discrete time series, the theory can be extended straightforwardly to spatially extended dynamical systems (196), where it quantifies self-organization (197), to controlled dynamical systems and transducers, and to prediction problems generally (188).

8.4. Power Law Distributions

Over the last decade or so, it has become reasonably common to see people (especially physicists) claiming that some system or other is complex, because it exhibits a power law distribution of event sizes. Despite its popularity, this is simply a fallacy. No one has demonstrated any relation between power laws and any kind of formal complexity measure. Nor is there any link tying power laws to our intuitive idea of complex systems as ones with strongly interdependent parts.

It is true that, *in equilibrium statistical mechanics*, one does not find power laws *except* near phase transitions (200), when the system *is* complex by our standard. This has encouraged physicists to equate power laws as such with complexity. Despite this, it has been known for half a century (5) that there are many, many ways of generating power laws, just as there are many mechanisms that can produce Poisson distributions, or Gaussians. Perhaps the simplest one is that recently demonstrated by Reed and Hughes (201), namely exponential growth coupled with random observation times. The observation of power laws alone thus says nothing about complexity (except in thermodynamic equilibrium!), and certainly is not a reliable sign of some specific favored mechanism, such as self-organized criticality (202,203) or highly optimized tolerance (204–206).

8.4.1. Statistical Issues Relating to Power Laws

The statistics of power laws are poorly understood within the field of complex systems, to a degree that is quite surprising considering how much attention has been paid to them. To be quite honest, there is little reason to think that many of the things claimed to be power laws actually *are* such, as opposed to some other kind of heavy-tailed distribution. This brief section will attempt to inoculate the reader against some common mistakes, most of which are related to the fact that a power law makes a straight line on a log-log plot. Since it would be impractical to cite all papers that commit these mistakes, and unfair to cite only some of them, I will omit references here; interested readers will be able to assemble collections of their own very rapidly.

Parameter Estimation. Presuming that something is a power law, a natural way of estimating its exponent is to use linear regression to find the line of best fit to the points on the log-log plot. This is actually a consistent estimator, if the data really do come from a power law. However, the loss function used in linear regression is the sum of the squared distances between the line and the points ("least squares"). In general, the line minimizing the sum of squared errors is *not* a valid probability distribution, and so this is simply not a reliable way to estimate the *distribution*.

One is much better off using maximum likelihood to estimate the parameter. With a discrete variable, the probability function is expressed as follows: $\Pr(X = x) = x^{-\alpha}/\zeta(\alpha)$, where $\zeta(\alpha) = \sum_{k=1}^{\infty} k^{-\alpha}$ is the Riemann zeta function, which ensures that the probability is normalized. Thus the maximum likelihood estimate of the exponent is obtained by minimizing the negative log-likelihood, $L(\alpha) = \Sigma_i \alpha \log x_i + \log \zeta(\alpha)$, i.e., $L(\alpha)$ is our loss function. In the continuous case, the probability density is $(\alpha - 1)c^{\alpha-1}/x^{\alpha}$, with $x \geq c > 0$.

Error Estimation. Most programs used to perform linear regression also provide an estimate of the standard error in the estimated slope, and one sometimes sees this reported as the uncertainty in the power law. This is an entirely unacceptable procedure. Those calculations of the standard error assume that measured values having Gaussian fluctuations around their true means. Here that would mean that the log of the empirical relative frequency is equal to the log of the probability plus Gaussian noise. However, by the central limit theorem, one knows that the relative frequency is equal to the probability plus Gaussian noise, so the former condition does not hold. Notice that one can obtain asymptotically reliable standard errors from maximum likelihood estimation.

Validation, R^2. Perhaps the most pernicious error is that of trying to validate the assumption of a power law distribution by either eye-balling the fit to a straight line, or evaluating it using the R^2 statistic, i.e., the fraction of the variance accounted for by the least-squares regression line. Unfortunately, while these procedures are good at confirming that something is a power law, if it really is (low Type I error, or high statistical significance), they are very bad at alerting you to things that are *not* power laws (they have a very high rate of Type II error, or low statistical power). The basic problem here is that *any* smooth curve looks like a straight line, if you confine your attention to a sufficiently small region—and for some non–power-law distributions, such "sufficiently small" regions can extend over multiple orders of magnitude.

To illustrate this last point, consider Figure 5, made by generating 10,000 random numbers according to a log-normal distribution, with a mean log of 0 and a standard deviation in the log of 3. Restricting attention to the "tail" of random numbers ≥ 1, and doing a usual least-squares fit, gives the line shown in

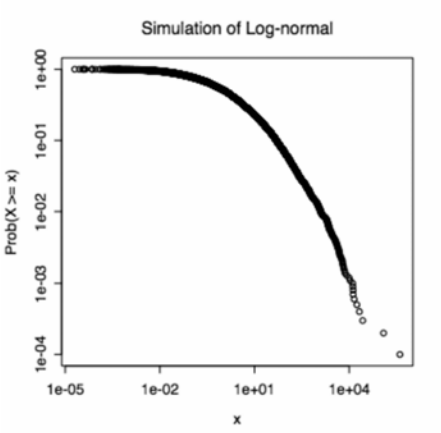


Figure 5. Distribution of 10,000 random numbers, generated according to a log-normal distribution with $\mathbf{E}[\log X] = 0$ and $\sigma(\log X) = 3$.

Figure 6. One might hope that it would be easy to tell that this data does not come from a power law, since there are a rather large number of observations (5,112), extending over a wide domain (more than four orders of magnitude). Nonetheless, R^2 is 0.962. This, in and of itself, constitutes a demonstration that getting a high R^2 is not a reliable indicator that one's data was generated by a power law.[29]

An Illustration: Blogging. An amusing empirical illustration of the difficulty of distinguishing between power laws and other heavy-tailed distributions is provided by political weblogs, or "blogs"—websites run by individuals or small groups providing links and commentary on news, political events, and the writings of other blogs. A rough indication of the prominence of a blog is provided by the number of other blogs linking to it—its **in-degree**. (For more on network terminology, see Part II, chapter 4, by Wuchty, Ravasz and Barabási, this volume.) A widely read essay by Shirky claimed that the distribution of in-degree follows a power law, and used that fact, and the literature on the growth of scale-free networks, to draw a number of conclusions about the social organization of the blogging community (207). A more recent paper by Drenzer and Farrell (208), in the course of studying the role played by blogs in general political debate, re-examined the supposed power-law distribution.[30] Using a large population of inter-connected blogs, they found a definitely heavy-tailed distribution which, on a log-log plot, was quite noticeably concave (Figure 7); nonetheless, R^2 for the conventional regression line was 0.898.

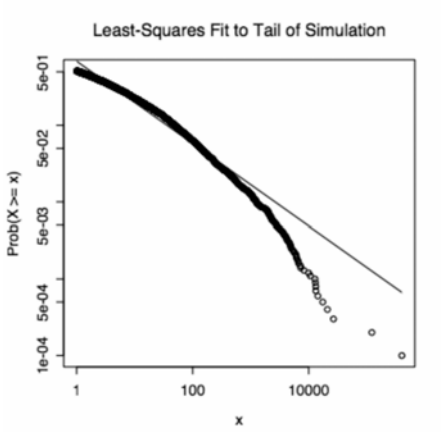


Figure 6. Inability of linear regression on log-log plots to correctly identify power law distributions. Simulation data (circles) and resulting least-squares fit (line) for the 5,112 points in Figure 5 for which $x \geq 1$. The R^2 of the regression line is 0.962.

Maximum likelihood fitting of a power law distribution gave $\alpha = -1.30 \pm 0.006$, with a negative log-likelihood of 18481.51. Similarly fitting a log-normal distribution gave $\mathbf{E}[\log X] = 2.60 \pm 0.02$ and $\sigma(\log X) = 1.48 \pm 0.02$, with a negative log-likelihood of 17,218.22. As one can see from Figure 8, the log-normal provides a very good fit to almost all of the data, whereas even the best fitting power-law distribution is not very good at all.[31]

A rigorous application of the logic of error testing (50) would now consider the probability of getting at least this good a fit to a log-normal if the data were actually generated by a power law. However, since in this case the data were $e^{18481.51-17218.22} \approx 13$ million times more likely under the log-normal distribution, any sane test would reject the power-law hypothesis.

8.5. Other Measures of Complexity

Considerations of space preclude an adequate discussion of further complexity measures. It will have to suffice to point to some of the leading ones. The **thermodynamic depth** of Lloyd and Pagels (182) measures the amount of information required to specify a trajectory leading to a final state, and is related both to departure from thermodynamic equilibrium and to retrodiction (209). Huberman and Hogg (210), and later Wolpert and Macready (211), proposed to measure complexity as the *dissimilarity* between different levels of a given system, on the grounds that self-similar structures are actually very

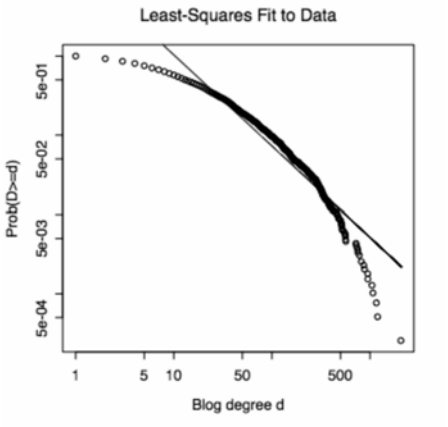


Figure 7. Empirical distribution of the in-degrees of political weblogs ("blogs"). Horizontal axis: number of incoming links d; vertical axis: fraction of all blogs with at least that many links, $\Pr(D \geq d)$; both axes are on a log-log scale. Circles show the actual distribution; the straight line is a least-squares fit to these values. This does not produce a properly normalized probability distribution but it does have an R^2 of 0.898, despite the clear concavity of the curve.

easy to describe. (Say what one level looks like, and then add that all the rest are the same!) Wolpert and Macready's measure of self-dissimilarity is, in turn, closely related to a complexity measure proposed by Sporns, Tononi, and Edelman (212–214) for biological networks, which is roughly the amount of information present in higher-order interactions between nodes which is not accounted for by the lower-order interactions. Badii and Politi (10) propose a number of further **hierarchical scaling complexities**, including one that measures how slowly predictions converge as more information about the past becomes available. Other interesting approaches include the **information fluctuation** measure of Bates and Shepard (215), and the predictability indices of the "school of Rome" (216).

8.6. Relevance of Complexity Measures

Why measure complexity at all? Suppose you are interested in the patterns of gene expressions in tumor cells and how they differ from those of normal cells. Why should you care if I analyze your data and declare that (say) healthy cells have a more complex expression pattern? Assuming you are not a numerologist, the only reason you *should* care is if you can learn something from that number—if the complexity I report tells you something about the

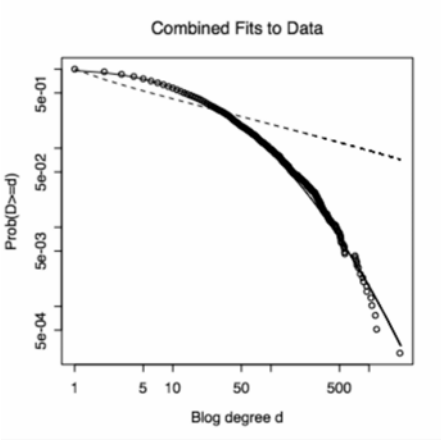


Figure 8. Maximum likelihood fits of log-normal (solid line) and power law (dashed line) distributions to the data from Figure 7 (circles); axes as in that figure. Note the extremely tight fit of the log-normal over the whole range of the curve, and the general failure of the power-law distribution.

thermodynamics of the system, how it responds to fluctuations, how easy it is to control, etc. A good complexity measure, in other words, is one which is *relevant* to many other aspects of the system measured. A bad complexity measure lacks such relevance; a really bad complexity measure would be positively misleading, lumping together things with no real connection or similarity just because they get the same score. My survey here has focused on complexity measures that have some claim to relevance, deliberately avoiding the large number of other measures which lack it (216).

9. GUIDE TO FURTHER READING

9.1. General

There is no systematic or academically detailed survey of the "patterns" of complex systems, but there are several sound informal discussions: Axelrod and Cohen (218), Flake (219), Holland (220), and Simon (221). The book by Simon, in particular, repays careful study.

On the "topics," the only books I can recommend are the ones by Boccara (222) and Flake (219). The former emphasizes topics from physics, chemistry, population ecology, and epidemiology, along with analytical methods, especially from nonlinear dynamics. Some sections will be easier to understand if one is familiar with statistical mechanics at the level of, e.g., (200), but this is

not essential. It does not, however, describe any models of adaptation, learning, evolution, etc. Many of those topics are covered in Flake's book, which however is written at a much lower level of mathematical sophistication.

On foundational issues about complexity, the best available surveys (10,195) both neglect the more biological aspects of the area, and assume advanced knowledge of statistical mechanics on the part of their readers.

9.2. Data Mining and Statistical Learning

There are now two excellent introductions to statistical learning and data mining: (223) and (31). The former is more interested in computational issues and the initial treatment of data; the latter gives more emphasis to pure statistical aspects. Both are recommended unreservedly. Baldi and Brunak (95) introduce machine learning via its applications to bioinformatics, and this may be especially suitable for readers of the present volume.

For readers seriously interested in understanding the theoretical basis of machine learning, (224) is a good starting point. The work of Vapnik (22,225,226) is fundamental; the presentation in (22) is enlivened by many strong and idiosyncratic opinions, pungently expressed. (40) describes the very useful class of models called "support vector machines," as well as giving an extremely clear exposition of key aspects of statistical learning theory. Those interested in going further will find that most of the relevant literature is still in the form of journals—*Machine Learning*, *Journal of Machine Learning Research* (free online at www.jmlr.org), *Neural Computation*—and especially annual conference proceedings—Computational Learning Theory (COLT), International Conference on Machine Learning (ICML), Uncertainty in Artificial Intelligence (UAI), Knowledge Discovery in Databases (KDD), Neural Information Processing Systems (NIPS), and the regional versions of them (EuroCOLT, Pacific KDD, etc.).

Much of what has been said about model selection could equally well have been said about what engineers call **system identification**, and in fact *is* said in good modern treatments of that area, of which (227) may be particularly recommended.

In many respects, data mining is an extension of exploratory data analysis; the classic work by Tukey (228) is still worth reading. No discussion of drawing inferences from data would be complete without mentioning the beautiful books by Tufte (229–231).

9.3. Time Series

Perhaps the best all-around references for the nonlinear dynamics approach are (60) and (232). The former, in particular, succeeds in integrating standard principles of statistical inference into the nonlinear dynamics method. (73), while less advanced than those two books, is a model of clarity, and contains an

integrated primer on chaotic dynamics besides. Ruelle's little book (16) is *much* more subtle than it looks, full of deep insights. The SFI proceedings volumes (233,234) are very worthwhile. The journals *Physica D*, *Physical Review E*, and *Chaos* often have new developments.

From the statistical wing, one of the best recent textbooks is (55); there are many, many others. That by Durbin and Koopman (60) is particularly strong on the state-space point of view. The one by (235) Azencott and Dacunha-Castelle is admirably clear on both the aims of time series analysis, and the statistical theory underlying classical methods; unfortunately it typography is less easy to read than it should be. (236) provides a comprehensive and up-to-date view of the statistical theory for modern models, including strongly nonlinear and non-Gaussian models. While many of the results are directly useful in application, the proofs rely on advanced theoretical statistics, in particular the geometric approach pioneered by the Japanese school of Amari et al. (237). This **information geometry** has itself been applied by Ay to the study of complex systems (238,239).

At the interface between the statistical and the dynamical points of view, there is an interesting conference proceedings (240) and a useful book by Tong (241). Pearson's book (242) on discrete-time models is very good on many important issues related to model selection, and exemplifies the habit of control theorists of cheerful stealing whatever seems helpful.

Filtering. Linear filters are well-described by many textbooks in control theory (e.g., (243)), signal processing, time series analysis (e.g., (55)), and stochastic dynamics (e.g., (58)).

(89) provides a readable introduction to optimal nonlinear estimation, draws interesting analogies to nonequilibrium statistical mechanics and turbulence, *and* describes a reasonable approximation scheme. (90) is an up-to-date textbook, covering both linear and nonlinear methods, and including a concise exposition of the essential parts of stochastic calculus. The website run by R.W.R. Darling, www.nonlinearfiltering.webhop.net, provides a good overview and extensive pointers to the literature.

Symbolic Dynamics and Hidden Markov Models. On symbolic dynamics, formal languages and hidden Markov models generally, see (10). (198) is a good first course on formal languages and automata theory. Charniak is a very readable introduction to grammatical inference. (244) is an advanced treatment of symbolic dynamics emphasizing applications; by contrast, (116) focuses on algebraic, pure-mathematical aspects of the subject. (163) is good on the stochastic properties of symbolic-dynamical representations. Gershenfeld (245) gives a good motivating discussion of hidden Markov models, as does Baldi and Brunak (95), while (94) describes advanced methods related to statistical signal processing. Open-source code for reconstructing causal-state models from state is available from http://bactra.org/CSSR.

9.4. Cellular Automata

General. There is unfortunately no completely satisfactory unified treatment of cellular automata above the recreational. Ilachinski (246) attempts a general survey aimed at readers in the physical sciences, and is fairly satisfactory on purely mathematical aspects, but is more out of date than its year of publication suggests. Chopard and Droz (247) has good material on models of pattern formation missing from Ilachinski, but the English is often choppy. Toffoli and Margolus (248) is inspiring and sound, though cast on a piece of hardware and a programming environment that are sadly no longer supported. Much useful material on CA modeling has appeared in conference proceedings (249–251).

CA as Self-Reproducing Machines. The evolution of CA begins in (252), continues in (253), and is brought up to the modern era in (254); the last is a beautiful, thought-provoking and modest book, sadly out of print. The modern era itself opens with (255).

Mathematical and Automata-Theoretic Aspects. Many of the papers in (256) are interesting. Ilachinski (146), as mentioned, provides a good survey. The Gutowitz volume (250) has good material on this topic, too. (257) is up-to-date.

Lattice gases. (124) is a good introduction, and (258) somewhat more advanced. The pair of proceedings edited by Doolen (259,260) describe many interesting applications, and contain useful survey and pedagogical articles. (There is little overlap between the two volumes.)

9.5. Agent-Based Modeling

There do not seem to be any useful textbooks or monographs on agent-based modeling. The *Artificial Life* conference proceedings, starting with (255), were a prime source of inspiration for agent-based modeling, along with the work of Axelrod (261). (262) is also worth reading. The journal *Artificial Life* continues to be relevant, along with the *From Animals to Animats* conference series. Epstein and Axtell's book (263) is in many ways the flagship of the "minimalist" approach to ABMs; while the arguments in its favor (e.g., (264,265)) are often framed in terms of social science, many apply with equal force to biology.[32] (266) illustrates how ABMs can be combined with extensive real-world data. Other notable publications on agent-based models include (267), spanning social science and evolutionary biology, (268) on agent-based models of morphogenesis, and (269) on biological self-organization.

(131) introduces object-oriented programming and the popular Java programming language at the same time; it also discusses the roots of object-orientation in computer simulation. There are many, many other books on object-oriented programming.

9.6. Evaluating Models of Complex Systems

Honerkamp (58) is great, but curiously almost unknown. Gershenfeld (245) is an extraordinary readable encyclopedia of applied mathematics, especially methods which can be used on real data. Gardiner (270) is also useful.

Monte Carlo. The old book by Hammersley and Handscomb (140) is concise, clear, and has no particular prerequisites beyond a working knowledge of calculus and probability. (271) and (272) are both good introductions for readers with some grasp of statistical mechanics. There are also very nice discussions in (58,31,142). Beckerman (143) makes Monte Carlo methods the starting point for a fascinating and highly unconventional exploration of statistical mechanics, Markov random fields, synchronization, and cooperative computation in neural and perceptual systems.

Experimental design. Bypass the cookbook texts on standard designs, and consult Atkinson and Donev (155) directly.

Ecological inference. (273) is at once a good introduction, and the source of important and practical new methods.

9.7. Information Theory

Information theory appeared in essentially its modern form with Shannon's classic paper (274), though there had been predecessors in both communications (275) and statistics, notably Fisher (see Kullback (159) for an exposition of these notions), and similar ideas were developed by Wiener and von Neumann, more or less independently of Shannon (56). Cover and Thomas (158) is, deservedly, the standard modern textbook and reference; it is highly suitable as an introduction, and handles almost every question most users will, in practice, want to ask. (276) is a more mathematically rigorous treatment, now free online. On neural information theory, (172) is seminal, well-written, still very valuable, and largely self-contained. On the relationship between physics and information, the best reference is still the volume edited by Zurek (12), and the thought-provoking paper by Margolus.

9.8. Complexity Measures

The best available survey of complexity measures is that by Badii and Politi (10); the volume edited by Peliti and Vulpiani (277), while dated, is still valuable. Edmonds (278) is an online bibliography, fairly comprehensive through 1997. (195) has an extensive literature review.

On Kolmogorov complexity, see Li and Vitanyi (177). While the idea of measuring complexity by the length of descriptions is usually credited to the trio of Kolmogorov, Solomonoff, and Chaitin, it is implicit in von Neumann's 1949 lectures on the "Theory and Organization of Complicated Automata" (252, Part I, especially pp. 42–56).

On MDL, see Rissanen's book (185), and Grünwald's lecture notes (270). Vapnik (22) argues that when MDL converges on the optimal model, SRM will too, but he assumes independent data.

On statistical complexity and causal states, see (195) for a self-contained treatment, and (188) for extensions of the theory.

10. ACKNOWLEDGMENTS

This work has been supported by a grant from the James S. McDonnell Foundation, and was finished while enjoying the hospitality of the Institut des Systèmes Complexes, the Laboratoire de l'Informatique du Parallèlisme and the Exystence Thematic Institute on Discrete and Computational Aspects of Complex Systems at the École Normale Supériere de Lyon. I am grateful to Michel Morvan and Cris Moore, organizers of the "Science et Gastronomie 2003" workshop, for allowing me to present some of this material there, and to my fellow participants for their critical comments. It is a pleasure to acknowledge discussions with Dave Albers, Satinder Singh Baveja, Philippe Binder, Sven Bruckener, Sandra Chapman, Markus Christen, Michael Cohen, Jim Crutchfield, Gunther Eble, Dave Feldman, Walter Fontana, Peter Grassberger, Rob Haslinger, John Holland, Herbert Jaeger, Jürgen Jost, Michael Lachmann, Alex Lancaster, Norm Margolus, Cris Moore, Mark Newman, Scott Page, Mitchell Porter, Kristina Shalizi, Carl Simon, Eric Smith, Ricard Solé, Bill Tozier, Erik-van Nimwegen and Nick Watkins. Special thanks to Prof. Cohen for permission to take a triangle he drew and add a corner to it; to Profs. Drenzer and Farrell for permission to use their data on weblogs; to Bill Tozier for suggesting that the leitmotifs of complex systems are analysis patterns, and for advice on agents; to Kara Kedi for daily validation of von Neumann's remarks on the complexity of cats; to Mathew Dafilis and Nigel Phillips for spotting misprints; to Kristina Shalizi for assistance on linear models of time series and for careful reading of a draft; and to Kristina and Rob for resisting the idea that there *is* such a thing as complex systems science. Finally, I wish to acknowledge the patience of the editors.

11. NOTES

1. Several books pretend to give a unified presentation of the topics. To date, the only one worth reading is (222), which however omits all models of adaptive systems.

2. Not all data mining is strictly for predictive models. One can also mine for purely descriptive models, which try to, say, cluster the data points so that more similar ones are closer together, or just assign an overall likelihood score. These, too, can be regarded as minimizing a cost function (e.g., the dissimilarity within clusters plus the similarity across clusters). The important point is that

good descriptions, in this sense, are implicitly predictive, either about other aspects of the data or about further data from the same source.

3. A subtle issue can arise here, in that not all errors need be equally bad for us. In scientific applications, we normally aim at accuracy as such, and so all errors *are* equally bad. But in other applications, we might care very much about otherwise small inaccuracies in some circumstances, and shrug off large inaccuracies in others. A well-designed loss function will represent these desires. This does not change the basic principles of learning, but it can matter a great deal for the final machine (280).

4. Here and throughout, I try to follow the standard notation of probability theory, so capital letters (X, Y, etc.) denote random variables, and lower-case ones particular values or realizations—so X = the role of a die, whereas $x = 5$ (say).

5. This is called the **convergence in probability** of $\hat{L}(\theta)$ to its mean value. For a practical introduction to such convergence properties, the necessary and sufficient conditions for them to obtain, and some thoughts about what one can do, statistically, when they do not, see (51).

6. The precise definition of the VC dimension is somewhat involved, and omitted here for brevity's sake. See (224,40) for clear discussions.

7. For instance, one can apply the independent-sample theory to learning feedback control systems (281).

8. Actually, the principle goes back to Aristotle at least, and while Occam used it often, he never used exactly those words (282, translator's introduction).

9. This is very close to the notion of the power of a statistical hypothesis test (283), and almost exactly the same as the severity of such a test (50).

10. One could, of course, build a filter that uses later y values as well; this is a **non-causal** or **smoothing** filter. This is clearly not suitable for estimating the state in real time, but often gives more accurate estimates when it is applicable. The discussion in the text generally applies to smoothing filters, at some cost in extra notation.

11. Equivalent terms are **future-resolving** or **right-resolving** (from nonlinear dynamics) and **deterministic** (the highly confusing contribution of automata theory).

12. Early publications on this work started with the assumption that the discrete values were obtained by dividing continuous measurements into bins of width ε, and so called the resulting models "ε-machines." This name is unfortunate: that is usually a bad way of discretizing data (§3.6.4); the quantity ε plays no role in the actual theory, and the name is more than usually impenetrable to outsiders. While I have used it extensively myself, it should probably be avoided.

13. An alternate definition (10) looks at the entropy rate (§7) of the symbol sequences: a generating partition is one that maximizes the entropy rate, which

is the same as maximizing the extra information about the initial condition x provided by each symbol of the sequence $\Phi(x)$.

14. Quantum versions of CA are an active topic of investigation, but unlikely to be of biological relevance (246).

15. In a talk at the Santa Fe Institute, summer of 2000; the formula does not seem to have been published.

16. A simple argument just invokes the central limit theorem. The number of points falling within the shaded region has a binomial distribution, with success parameter p, so asymptotically x/n has a Gaussian distribution with mean p and standard deviation $\sqrt{p(1-p)/n}$. A nonasymptotic result comes from Chernoff's inequality (281), which tells us that, for all n, we have $\Pr(|x/n - p| \geq \varepsilon) < 2e^{-2n\varepsilon^2}$.

17. The chain needs to be irreducible, meaning one can go from any point to any other point, and positive recurrent, meaning that there's a positive probability of returning to any point infinitely often.

18. Unless our choices for the transition probabilities are fairly perverse, the central limit theorem still holds, so asymptotically our estimate still has a Gaussian distribution around the true value, and still converges as $N^{-1/2}$ for large enough N, but determining what's "large enough" is trickier.

19. An important exception is the case of equilibrium statistical mechanics, where the dynamics under the Metropolis algorithm *are* like the real dynamics.

20. For a pedagogical discussion, with examples, of how compression algorithms may be misused, see http://bactra.org/notebooks/cep-gzip.html.

21. The issue of what language to write the program in is secondary; writing a program to convert from one language to another just adds on a constant to the length of the overall program, and we will shortly see why additive constants are not important here.

22. Very short programs can calculate π to arbitrary accuracy, and the length of these programs does not grow as the number of digits calculated does. So one could run one of these programs until it had produced the first two quadrillion digits, and then erase the first half of the output, and stop.

23. (167) is perhaps the most notorious; see (168) and especially (169) for critiques.

24. It is certainly legitimate to regard any dynamical process as also a computational process, (284–286,195), so one could argue that the data *are* produced by some kind of program. But even so, this computational process generally does not resemble that of the minimal Kolmogorov program at all.

25. It is important to note (185, ch. 3) that if we allowed any possible model in Θ, finding the minimum would, once again, be incomputable. This restriction to a definite, perhaps hierarchically organized, class of models is vitally important.

26. Take our favorite class of models, and add on deterministic models that produce particular fixed blocks of data with probability 1. For any of these mod-

els θ, $L(x,\theta)$ is either 0 (if x is what that model happens to generate) or ∞. Then, once we have our data, and find a θ that generates that and nothing but that, rearrange the coding scheme so that $D(\theta,\Theta) = 1$; this is always possible. Thus, $C_{SC}(x,\Theta) = 1$ bit.

27. This does not contradict the convergence result of the last paragraph; one of the not-too-onerous conditions mentioned in the previous paragraph is that the coding scheme remain fixed, and we're violating that.

28. Technically, a given regular language (§3.6).

29. If I replace the random data by the *exact* log-normal probability distribution over the same range, and do a least-squares fit to that, the R^2 actually increases, to 0.994.

30. Professors Drenzer and Farrell kindly shared their data with me, but the figures and analysis that follow are my own.

31. Note that the log-normal curve fitted to the *whole* data continues to match the data well even in the tail. For further discussion, omitted here for reasons of space, see http://bactra.org/weblog/232.html.

32. In reading this literature, it may be helpful to bear in mind that by "methodological individualism," social scientists mean roughly what biologists do by "reductionism."

12. REFERENCES

1. Gamma E, Helm R, Johnson R, Vlissides J. 1995. *Design patterns: elements of reusable object-oriented software*. Addison-Wesley J, Reading, MA 1995.
2. Anderson RW. 1988. Random-walk learning: a neurobiological correlate to trial-and-error. In *Progress in neural networks*, pp. 221–244. Ed. OM Omidvar, J Dayhoff. Academic Press, Boston.
3. Mueller S, Marchettod J, Airaghi S, Koumoutsakos P. 2002. Optimization based on bacterial chemotaxis. *IEEE Trans Evolut Comput* **6**:16–29.
4. Simon HA. 1962. The architecture of complexity: Hierarchic systems. *Proc Am Philos Soc* **106**:467–482 (reprinted as chap. 8 of [221]).
5. Simon HA. 1955. On a class of skew distribution functions. *Biometrika* **42**:425–440.
6. Turing A. 1952. The chemical basis of morphogenesis. *Philos Trans Roy Soc B* **237**:37–72.
7. Strong SP, Freedman B, Bialek W, Koberle R. 1998. Adaptation and optimal chemotactic strategy for *E. coli*. *Phys Rev E* **57**:4604–4617 (http://arxiv.org/abs/adap-org/9706001).
8. Alon U, Surette MG, Barkai N, Leibler S. Robustness in bacterial chemotaxis. *Nature* **397**:168–171.
9. Yi T-M, Huang Y, Simon MI, Doyle J. 2000. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc Natl Acad Sci USA* **97**:4649–4653.
10. Badii R, Politi A. 1997. *Complexity: hierarchical structures and scaling in physics*. Cambridge UP, Cambridge.
11. Fontana W, Buss, LW. 1994. "Arrival of the fittest": towards a theory of biological organization. *Bull Math Biol* **56**:1–64 (http://www.santafe.edu/~walter/Papers/arrival.US.ps.gz).
12. Zurek WH, ed. 1990. *Complexity, entropy, and the physics of information*. Addison-Wesley, Reading, MA.
13. Frisch U. 1995. *Turbulence: the legacy of A.N. Kolmogorov*. Cambridge, Cambridge UP

14. Cross MC, Hohenberg P. 1993. Pattern formation out of equilibrium. *Rev Mod Phys* 65:851–1112.
15. Ball P. 1999. *The self-made tapestry: pattern formation in nature*. Oxford UP, Oxford.
16. Holland JH. 1992. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. 2nd ed. MIT Press, Cambridge (1st ed. 1975, U Michigan P, Ann Arbor).
17. Mitchell M. 1996. *An introduction to genetic algorithms*. MIT Press, Cambridge.
18. Gintis H. 2000. *Game theory evolving: a problem-centered introduction to modeling strategic interaction*. Princeton UP, Princeton.
19. Hofbauer J, Sigmund K. 1988. *The theory of evolution and dynamical systems: mathematical aspects of selection*. Cambridge UP, Cambridge.
20. Fischer KH, Hertz JA. 1988. *Spin glasses*. Cambridge Studies in Magnetism. Cambridge UP, Cambridge.
21. Stein DL. 2003. Spin glasses: still complex after all these years? In *Quantum decoherence and entropy in complex systems*. Ed. T Elze. Springer, Berlin.
22. Vapnik VN. 2000. *The nature of statistical learning theory*, 2nd ed. Springer, Berlin.
23. Akaike H. 1973. Information theory and an extension of the maximum likelihood principle. In *Proceedings of the second international symposium on information theory*, pp. 267–281. Ed. BN Petrov, F Caski. Akademiai Kiado, Budapest (repr. in [287, pp. 199–213].
24. Akaike H. 1977. On entropy maximization principle. In *Applications of statistics*, pp. 27–41. Ed. PR Krishnaiah. North-Amsterdam, Holland
25. Schwarz G. 1978. Estimating the dimension of a model. *Ann Stat* **6**:461–464.
26. van de Geer S. 2000. *Empirical processes in M-estimation*. Cambridge UP, Cambridge.
27. Biggers ED. 1928. *Behind that curtain*. Grosset and Dunlap, New York.
28. Valiant LF. 1984. A theory of the learnable. *Commun Assoc Comput Machinery* **27**:1134–1142.
29. Shao X, Cherkassky V, Li W. 2000. Measuring the VC-dimension using optimized experimental design. *Neural Comput* **12**:1969–1986.
30. Meir R. 2000. Nonparametric time series prediction through adaptive model selection. *Machine Learning* **39**:5–34.
31. Hastie T, Tibshirani R, Friedman J. 2001. *The elements of statistical learning: data mining, inference, and prediction*. Springer, New York.
32. Ripley BD. 1996. *Pattern recognition and neural networks*. Cambridge UP, Cambridge.
33. Wahba G. 1990. *Spline models for observational data*. Society for Industrial and Applied Mathematics, Philadelphia.
34. Anthony M, Bartlett PL. 1999. *Neural network learning: theoretical foundations*. Cambridge UP, Cambridge.
35. Zapranis A, Refenes A-P. 1999. *Principles of neural model identification, selection and adequacy: with applications to financial econometrics*. Springer, London.
36. Engel A, Van den Broeck C. 2001. *Statistical mechanics of learning*. Cambridge UP, Cambridge.
37. Breiman L, Friedman J, Olshen R, Stone C. 1984. *Classification and regression trees*. Wadsworth, Belmont, CA.
38. Gigerenzer G, Todd PM, ABC Research Group. 1999. *Simple heuristics that make us smart*. Oxford UP, Oxford.
39. Herbrich R. 2002. *Learning kernel classifiers: theory and algorithms*. MIT Press, Cambridge.
40. Cristianini N, Shawe-Taylor J. 2000. *An introduction to support vector machines: and other kernel-based learning methods*. Cambridge UP, Cambridge.
41. Pearl J. 2000. *Causality: models, reasoning, and inference*. Cambridge UP, Cambridge.
42. Shafer G. 1996. *The art of causal conjecture*. MIT Press, Cambridge.
43. Spirtes P, Glymour C, Scheines R. 2001. *Causation, prediction, and search*, 2nd ed. MIT Press, Cambridge.

44. Dayan P, Hinton GE, Neal RM, Zemel, RS. 1995. The Helmholtz machine. *Neural Comput* **7**:889–904 (http://www.cs.utoronto.ca/~hinton/absps/helmholtz.htm).
45. Domingos P. 1999. The role of Occam's Razor in knowledge discovery. *Data Mining and Knowledge Discovery* **3**:409–425 (http://www.cs.washington.edu/home/pedrod/dmkd99.pz.gz).
46. Klein JL. 1997. *Statistical visions in time: a history of time series analysis, 1662–1938*. Cambridge UP, Cambridge.
47. Dirac PAM. 1935. *Principles of quantum mechanics*. Clarendon Press, Oxford.
48. Knight FB. 1975. A predictive view of continuous time processes. *Ann Probability* **3**:573–596.
49. Knight FB. 1992. *Foundations of the prediction process*. Oxford Studies in Probability, Vol. 1. Clarendon Press, Oxford.
50. Mayo DG. 1996. *Error and the growth of experimental knowledge*. U Chicago P, Chicago.
51. Gray RM. 1988. *Probability, random processes, and ergodic properties*. Springer, New York (http://ee-www.stanford.edu/~gray/arp.html)
52. Basawa IV, Scott DJ. 1983. *Asymptotic optimal inference for non-ergodic models*. Springer, Berlin.
53. West BJ, Deering B. 1995. *The lure of modern science: fractal thinking*. World Scientific, Singapore.
54. Press WH, Teukolsky SA, Vetterling WT, and Flannery BP. 1992. *Numerical recipes in c: the art of scientific computing*, 2nd ed. Cambridge UP, Cambridge.
55. Shumway RH, Stoffer DS. 2000. *Time series analysis and its applications*. Springer Texts in Statistics. Springer, New York.
56. Wiener N. 1961. *Cybernetics: or, control and communication in the animal and the machine*, 2nd ed. MIT Press, Cambridge (1st ed. 1948, Wiley, New York).
57. Hubbard BB. 1996. *The world according to wavelets: the story of a mathematical technique in the making*. A.K. Peters, Wellesley.
58. Honerkamp J. 1994. *Stochastic dynamical systems: concepts, numerical methods, data analysis*. Transl. Katja Lindenberg. VCH, New York.
59. Box GEP, Jenkins GM. 1970. *Time series analysis, forecasting, and control*. Holden-Day, Oakland, CA.
60. Durbin J, Koopman SJ. 2001. *Time series analysis by state space methods*. Oxford UP, Oxford.
61. Eyink GL. 1998. Linear stochastic models of nonlinear dynamical systems. *Phys Rev E* **58**:6975–6991.
62. Barndorff-Nielsen OE, Jensen JL, Sorensen M. 1990. Parametric modelling of turbulence. *Philos Trans Roy Soc A* **332**:439–455.
63. Eyink GL, Alexander FJ. 1998. Predictive turbulence modeling by variational closure. *J Stat Phys* **91**:221–283.
64. Beran J. 1994. *Statistics for long-memory processes*. Chapman and Hall, New York.
65. Embrechts P, Maejima M. 2002. *Selfsimilar processes*. Princeton UP, Princeton.
66. Bosq D. 1998. *Nonparametric statistics for stochastic processes: estimation and prediction*, 2nd ed. Springer, Berlin.
67. Algoet P. 1992. Universal schemes for prediction, gambling and portfolio selection. *Ann Probability* **20**:901–941. See also an important Correction, *Ann Probability* **23**:474–478, 1995.
68. Takens F. 1981. Detecting strange attractors in fluid turbulence. In *Symposium on dynamical systems and turbulence*, pp. 366–381. Ed. DA Rand and LS Young. Springer, Berlin.
69. Kantz H, Schreiber T. 1997. *Nonlinear time series analysis*. Cambridge UP, Cambridge.
70. Judd K, Mees A. 1998. Embedding as a modeling problem. *Physica D* **120**:273–286.
71. Small M, Tse CK. 2004. Optimal embedding parameters: a modelling paradigm. *Physica D* **194**:283–296 (http://arxiv.org/abs/physics/0308114).
72. Kennel MB, Brown R, Abarbanel HDI. 1992. Determining minimum embedding dimension using a geometric construction. *Phys Rev A* **45**:3403–3411.
73. Sprott JC. 2003. *Chaos and time-series analysis*. Oxford UP, Oxford.
74. Smith LA. 1988. Intrinsic limits on dimension calculations. *Phys Lett A* **133**:283–288.

75. Fraser AM, Swinney HL. 1986. Independent coordinates for strange attractors from mutual information. *Phys Rev A* **33**:1134–1140.
76. Cellucci CJ, Albano AM, Rapp PE. 2003. Comparative study of embedding methods. *Phys Rev E* **67**:162–210.
77. Letellier C, Aguirre LA. 2002. Investigating nonlinear dynamics from time series: the influence of symmetries and the choice of observables. *Chaos* **12**:549–558.
78. Wiener N. 1949. *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications*. Technology Press of MIT, Cambridge.
79. Kolmogorov AN. 1941. Interpolation und extrapolation von stationären zufälligen folgen. *Bull Acad Sci USSR Math* **3**:3–14 (in Russian with German summary).
80. Kalman RE. 1960. A new approach to linear filtering and prediction problems. *ASME Trans J Basic Eng* **82D**:35–50.
81. Kalman RE, Bucy RS. 1961. New results in linear filtering and prediction. *ASME Trans J Basic Eng* **83D**:95–108.
82. Bucy RS. 1994. *Lectures on discrete time filtering*. Springer, Berlin.
83. Stratonovich RL. 1968. *Conditional markov processes and their application to the theory of optimal control*, 2nd ed., revised by the author. Transl. RN and NB McDonough, with a preface by R Bellman. Elsevier, New York. (1st ed. 1966, *Uslovnyye markovskiye protessy i ikh primeneiye k teorri optimal'nogo upravleniya*, Moscow UP, Moscow).
84. Kushner HJ. 1967. Dynamical equations for optimal nonlinear filtering. *J Differ Eqs* **3**:179–190.
85. Lipster RS, Shiryaev AN. 2001. *Statistics of random processes*, 2 vols., 2nd ed. Transl. AB Aries. Springer, Berlin (first published 1974, *Statistika sluchainykh protessov*, Nauka, Moscow).
86. Tanizaki H. 1996. *Nonlinear filters: estimation and applications*, 2nd ed. Springer, Berlin.
87. Darling RWR. 1998. *Geometrically intrinsic nonlinear recursive filters I: algorithms*. Technical Report 494, Statistics Department, University of California-Berkeley (http://www.stat.berkeley.edu/tech-reports/494.abstract).
88. Darling RWR. 1998. *Geometrically intrinsic nonlinear recursive filters ii: foundations*. Technical Report 512, Statistics Department, University of California-Berkeley (http://www.stat.berkeley.edu/tech-reports/512.abstract).
89. Eyink GL. 2000. A variational formulation of optimal nonlinear estimation. *Methodology and Computing in Applied Probability*. submitted (http://arxiv.org/abs/physics/0011049).
90. Ahmed NU. 1998. *Linear and nonlinear filtering for scientists and engineers*. World Scientific, Singapore.
91. Chomsky N. 1956. Three models for the description of language. *IRE Trans Inf Theory* **2**:113–124.
92. Charniak E. 1993. *Statistical language learning*. MIT Press, Cambridge.
93. Manning CD, Schütze H. 1999. *Foundations of statistical natural language processing*. MIT Press, Cambridge.
94. Elliott RJ, Aggoun L, Moore JB. 1995. *Hidden markov models: estimation and control*. Applications of Mathematics: Stochastic Modelling and Applied Probability, Vol. 29. Springer, New York.
95. Baldi P, Brunak S. 2001. *Bioinformatics: the machine learning approach*, 2nd ed. MIT Press, Cambridge.
96. Neal RM, Hinton GE. 1998. A view of the EM algorithm that justifies incremental, sparse, and other variants. In *Learning in Graphical Models*, pp. 355–368. Ed. MI Jordan, Kluwer Academic, Dordrecht.
97. Rissanen J. 1983. A universal data compression system. *IEEE Trans Inf Theory* **29**:656–664.
98. Willems F, Shtarkov Y, Tjalkens T. 1995. The context-tree weighting method: basic properties. *IEEE Trans Inf Theory* **41**:653–664.

99. Ron D, Singer Y, Tishby N. 1996. The power of amnesia: learning probabilistic automata with variable memory length. *Machine Learning* **25**:117–149.
100. Bühlmann P, Wyner AJ. 1999. Variable length Markov chains. *Ann Stat* **27**:480–513 (http://www.stat.berkeley.edu/tech-reports/479.abstract1).
101. Kennel MB, Mees AI. 2002. Context-tree modeling of observed symbolic dynamics. *Phys Rev E* **66**:056209.
102. Crutchfield JP, Young K. 1989. Inferring statistical complexity. *Phys Rev Lett* **63**:105–108.
103. Jaeger H. 2000. Observable operator models for discrete stochastic time series. *Neural Comput* **12**:1371–1398 (http://www.faculty.iu-bremen.de/hjaeger/pubs/oom/neco00.pdf).
104. Littman ML, Sutton RS, Singh S. 2002. Predictive representations of state. In *Advances in neural information processing*, pp. 1555–1561. Ed. TG Dietterich, S Becker, Z Ghahramani, *Systems 14*. MIT Press, Cambridge (http://www.eecs.umich.edu/~baveja/Papers/psr.pdf).
105. Shalizi CR, Shalizi KL. 2004. Blind construction of optimal nonlinear recursive predictors for discrete sequences. In *Uncertainty in artificial intelligence: proceedings of the twentieth conference*, pp. 504–511. Ed. M Chickering, J Halpern. AUAI Press, Arlington, VA (http://arxiv.org/abs/cs.LG/0406011).
106. Salmon WC. 1971. *Statistical explanation and statistical relevance*. With contributions by RC Jeffrey and JG Greeno. U Pittsburgh P.
107. Salmon WC. 1984. *Scientific explanation and the causal structure of the world*. Princeton UP, Princeton.
108. Singh S, Littman, ML, Jong NK, Pardoe D, Stone P. 2003. Learning predictive state representations. In *Proceedings of the twentieth international conference on machine learning (ICML-2003)*, pp. 712–719. Ed. T Fawcett, N Mishra. AAAI Press, New York (http://www.eecs.umich.edu/~baveja/Papers/ICMLfinal.ps.gz).
109. Upper DR. 1997. *Theory and algorithms for hidden markov models and generalized hidden markov models*. PhD thesis, University of California, Berkeley (http://www.santafe.edu/projects/CompMech/ or papers/TAHMMGHMM.html).
110. Dupont P, Denis F, Esposito Y. 2004. Links between probabilistic automata and hidden Markov models: probability distributions, learning models and induction algorithms. *Pattern Recognit* Forthcoming (http://www.info.ucl.ac.be/people/pdupont/pdupont/postscript/Links_PA_HMM_preprint.ps.gz)
111. Jaeger H. 1999. *Characterizing distributions of stochastic processes by linear operators*. Technical Report 62, German National Center for Information Technology (http://www.faculty.iu-bremen.de/hjaeger/pubs/oom_distributionsTechRep.pdf).
112. Jaeger H. 2000. *Modeling and learning continuous-valued stochastic processes with OOMs*. Technical Report 102, German National Center for Information Technology (http://www.faculty.iu-bremen.de/hjaeger/pubs/jaeger.00.tr.contoom.pdf).
113. Crutchfield JP. 1992. Unreconstructible at any radius. *Phys Lett A* **171**:52–60.
114. Bollt EM, Stanford T, Lai Y-C, Zyczkowski K. 2000. Validity of threshold-crossing analysis of symbolic dynamics from chaotic time series. *Phys Rev Lett* **85**:3524–3527.
115. Bollt EM, Stanford T, Lai Y-C, Zyczkowski K. 2001. What symbolic dynamics do we get with a misplaced partition? On the validity of threshold crossing analysis of chaotic time-series. *Physica D* **154**:259–286.
116. Kitchens BP. 1998. *Symbolic dynamics: one-sided, two-sided and countable state markov shifts*. Springer, Berlin.
117. Kennel MB, Buhl M. 2003. Estimating good discrete partitions from observed data: symbolic false nearest neighbors. *Phys Rev Lett* **91**:084102 (http://arxiv.org/abs/nlin.CD/0304054).
118. Hirata Y, Judd K, Kilminster D. 2004. Estimating a generating partition from observed time series: Symbolic shadowing. *Phys Rev E* **70**:016215.
119. Crutchfield JP, Packard NH. 1983. Symbolic dynamics of noisy chaos. *Physica D* **7**:201–223.
120. Moore C. 1997. Majority-vote cellular automata, Ising dynamics, and P-completeness. *J Stat Phys* **88**:795–805 (http://arxiv.org/abs/cond-mat/9701118).

121. Moore C, Nordahl MG. 1997. Lattice gas prediction is P-complete. Electronic preprint (http://arxiv.org/abs/nlin.CG/9704001).
122. Hardy J, Pomeau Y, de Pazzis O. 1976. Molecular dynamics of a classical lattice gas: transport properties and time correlation functions. *Phys Rev A* **13**:1949–1960,.
123. Frisch U, Hasslacher B, Pomeau Y. 1986. Lattice-gas automata for the Navier-Stokes equation. *Phys Rev Lett* **56**:1505–1508.
124. Rothman DH, and Zaleski S. 1997. *Lattice-gas cellular automata: simple models of complex hydrodynamics*. Cambridge UP, Cambridge.
125. Fisch R, Gravner J, Griffeath D. 1991. Threshold-range scaling of excitable cellular automata. *Stat Comput* **1**:23–39 (http://psoup.math.wisc.edu/papers/tr.zip).
126. Nilsson M, Rasmussen S, Mayer B, Whitten D. 2003. Constructive molecular dynamics (MD) lattice gases: 3-D molecular self-assembly. In *New constructions in cellular automata*, pp. 275–290. Ed. D Griffeath, C Moore. Oxford UP, Oxford.
127. Nilsson M, Rasmussen S. 2003. Cellular automata for simulating molecular self-assembly. *Discr Math Theor Comput Sci* AB(DMCS):31–42 (http://dmtcs.loria.fr/proceedings/html/dmAB0103.abs.html).
128. Bartlett MS. 1955. *An introduction to stochastic processes, with special reference to methods and applications*. Cambridge UP, Cambridge.
129. Jacquez JA, Koopman JS, Simon CP, Longini IM. 1994. The role of the primary infection in epidemics of HIV-infection in gay cohorts. *J Acq Immune Def Synd Hum Retrovirol* **7**:1169–1184.
130. Koopman J, Jacquez J, Simon C, Foxman B, Pollock S, Barth-Jones D, Adams A, Welch G, Lange K. 1997. The role of primary HIV infection in the spread of HIV through populations. *J AIDS* **14**:249–258.
131. Budd T. 2000. *Understanding object-oriented programming with Java*, 2nd ed. Addison-Wesley, Reading, MA.
132. Resnick M. 1994. *Turtles, termites and traffic jams: explorations in massively parallel microworlds*. MIT Press, Cambridge.
133. Brown JS, Duguid P. 2000. *The social life of information*. Harvard Business School Press P, Boston.
134. Bonabeau E, Dorigo M, Theraulaz G. 1999. *Swarm intelligence: from natural to artificial systems*. Oxford UP, Oxford.
135. Lerman K. *Design and mathematical analysis of agent-based systems*. E-print, Information Sciences Institute, University of Southern California, 2000 (http://www.isi.edu/~lerman/papers/fmw00_abstract.html).
136. Ossowski S. 2000. *Co-ordination in artificial agent societies: social structure and its implications for autonomous problem-solving agents*. Springer, Berlin.
137. Wooldridge M. 2000. *Reasoning about rational agents*. MIT Press, Cambridge.
138. Jonker CM, Snoep JL, Treur J, Westerhoff HV, Wijngaards WAC. 2002. Putting intentions into cell biochemistry: An artificial intelligence perspective. *J Theor Biol* **214**:105–134.
139. Chaikin PM, Lubensky TC. 1995. *Principles of condensed matter physics*. Cambridge UP, Cambridge.
140. Hammersley JM, Handscomb DC. 1964. *Monte Carlo methods*. Chapman and Hall, London.
141. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. 1953. Equations of state calculations by fast computing machines. *J Chem Phys* **21**:1087–1092.
142. Brémaud P. 1999. *Markov chains: gibbs fields, monte carlo simulation, and queues*. Springer, Berlin.
143. Beckerman M. 1997. *Adaptive cooperative systems*. Wiley, New York.
144. Jordan MI, ed. 1998. *Learning in graphical models*. Kluwer Academic, Dordrecht.
145. Young HP. 1998. *Individual strategy and social structure: an evolutionary theory of institutions*. Princeton UP, Princeton.
146. Sutton J. 1998. *Technology and market structure: theory and history*. MIT Press, Cambridge.

147. Epstein IR, Pojman JA. 1998. *An introduction to nonlinear chemical dynamics: oscillations, waves, patterns, and chaos*. Oxford UP, Oxford.
148. Winfree AT. 1987. *When time breaks down: the three-dimensional dynamics of electrochemical waves and cardiac arrhythmias*. Princeton UP, Princeton.
149. Varela FJ, Maturana HR, Uribe R. 1974. Autopoiesis: the organization of living systems, its characterization and a model. *Biosystems* **5**:187–196.
150. Luhmann N. 1984/1995. *Social systems*. Transl. J Bednarz Jr, with D Baecker, and foreword by EM Knodt. Stanford UP, Stanford. Originally published as *Soziale systeme: grundriss einer allgemeinen theorie*. Suhrkamp-Verlag, Frankfurt am Main.
151. McMullin B. 1997. The case of the independent test. *Santa Fe Inst Bull* **12**(2).
152. McMullin B, Varela FJ. 1997. *Rediscovering computational autopoesis*. Technical Report no. 97-02-012, Santa Fe Institute (http://www.santafe.edu/research/publications/wpabstract/199702012).
153. Mitchell M, Hraber PT, Crutchfield JP. 1993. Revisiting the edge of chaos: evolving cellular automata to perform computations. *Complex Syst* **7**:89–130. (http://www.cse.ogi.edu/~mm/revedge.pdf).
154. Schuessler AA. 1999. Ecological inference. *Proc Natl Acad Sci USA* **96**:10578–10581.
155. Atkinson AC, Donev AN. 1992. *Optimum experimental designs*. Clarendon Press, Oxford.
156. Borowiak DS. 1989. *Model discrimination for nonlinear regression models*. Marcel Dekker, New York.
157. Newman MEJ, Palmer RG. 2003. *Modeling extinction*. Oxford UP, Oxford (http://arxiv.org/abs/adap-org/9908002).
158. Cover TM, Thomas JA. 1991. *Elements of information theory*. Wiley, New York.
159. Kullback S. 1968. *Information theory and statistics*, 2nd ed. Dover, New York.
160. Kulhavy R. 1996. *Recursive nonlinear estimation: a geometric approach*. Springer, Berlin.
161. Khinchin AI. 1957. *Mathematical foundations of information theory*. Transl. RA Silverman, MD Friedman. Dover, New York. Originally published as two Russian articles in *Uspekhi Matematicheskikh Nauk* (**7**:3–20, 1953; **9**:17–75, 1956).
162. Ruelle D. 1989. *Chaotic evolution and strange attractors: the statistical analysis of time series for deterministic nonlinear systems*. Cambridge UP, Cambridge. Notes prepared by Stefano Isola.
163. Beck C, Schlögl F. 1993. *Thermodynamics of chaotic systems: an introduction*. Cambridge UP, Cambridge.
164. Victor JD. 2000. Asymptotic bias in information estimates and the exponential (Bell) polynomials. *Neural Comput* **12**:2797–2804.
165. Kraskov A, Stögbauer H, Grassberger P. 2003. Estimating mutual information. *Phys Rev E*. Submitted (http://arxiv.org/abs/cond-mat/0305641).
166. Kontoyiannis I, Algoet M. Sukhov YuM, and Wyner AJ. 1998. Nonparametric entropy estimation for stationary processes and random fields, with applications to english text. *IEEE Trans Inf Theory* **44**:1310–1327 (http://www.dam.brown.edu/people/yiannis/PAPERS/suhov2.pdf).
167. Benedetto D, Caglioti E, Loreto V. 2002. Language trees and zipping. *Phys Rev Lett* **88**:048702 (http://arxiv.org/abs/cond-mat/0108530).
168. Khmelev DV, Teahan WJ. 2003. Comment on "Language trees and zipping." *Phys Rev Lett* 90:089803 (http://arxiv.org/abs/cond-mat/0205521).
169. Goodman J. 2002. Extended comment on "Language trees and zipping." Electronic pre-print (http://arxiv.org/abs/cond-mat/0202383).
170. MacKay DJC. 2003. *Information theory, inference, and learning algorithms*. Cambridge UP, Cambridge (http://www.inference.phy.cam.ac.uk/mackay/itprnn/book.html).
171. Durbin R, Eddy S, Krogh A, Mitchison G. 1998. *Biological sequence analysis: probabilistic models of proteins and nucleic acids*. Cambridge UP, Cambridge.
172. Rieke F, Warland D, van Steveninck RDR, Bialek W. 1997. *Spikes: exploring the neural code*. MIT Press, Cambridge.

173. Abbott LF, Sejnowski TJ, eds. 1998. *Neural codes and distributed representations: foundations of neural computation*. MIT Press, Cambridge.
174. Billingsley P. 1965. *Ergodic theory and information*. Wiley, New York.
175. Katok AB, Hasselblatt B. 1995. *Introduction to the modern theory of dynamical systems*. Cambridge UP, Cambridge.
176. Solomonoff RJ. 1964. A formal theory of inductive inference. *Inf Control* **7**:1–22, 224–254 (http://world.std.com/~rjs/pubs.html).
177. Li M, Vitanyi PMB. 1993. *An introduction to Kolmogorov complexity and its applications*. Springer, New York.
178. Zurek WH. 1998. Algorithmic randomness, physical entropy, measurements, and the demon of choice. E-print, arxiv.org (http://arxiv.org/abs/quant-ph/9807007).
179. Bennett CH. 1985. Dissipation, information, computational complexity and the definition of organization. In *Emerging syntheses in science*, pp. 215–234. Ed D Pines. Santa Fe Institute, Santa Fe, NM.
180. Bennett CH. 1986. On the nature and origin of complexity in discrete, homogeneous locally-interacting systems. *Found Phys* **16**:585–592.
181. Bennett CH. 1990. How to define complexity in physics, and why. In *Complexity, entropy, and the physics of information*, pp. 137–148. Ed. WH Zurek. Addison-Wesley, Reading, MA.
182. Lloyd S, Pagels H. 1988. Complexity as thermodynamic depth. *Ann Phys* **188**:186–213.
183. Gács P, Tromp JT, and Vitanyi PMB. 2001. Algorithmic statistics. *IEEE Trans Inf Theory* **47**:2443–2463 (http://arxiv.org/abs/math.PR/0006233).
184. Rissanen J. 1978. Modeling by shortest data description. *Automatica* **14**:465–471.
185. Rissanen J. 1989. *Stochastic complexity in statistical inquiry*. World Scientific, Singapore.
186. Hraber PT, Korber BT, Wolinsky S, Erlich H, Trachtenberg E. 2003. *HLA and HIV infection progression: application of the minimum description length principle to statistical genetics*. Technical Report 03-04-23, Santa Fe Institute (http://www.santafe.edu/research/publications/wpabstract/200304023).
187. Grassberger P. 1986. Toward a quantitative theory of self-generated complexity. *Int J Theor Phys* **25**:907–938.
188. Shalizi CR. 2001. *Causal architecture, complexity and self-organization in time series and cellular automata*. PhD thesis, University of Wisconsin-Madison (http://bactra.org/thesis/).
189. Shaw R. 1984. *The dripping faucet as a model chaotic system*. Aerial Press, Santa Cruz, CA.
190. Lindgren K, Nordahl MG. 1988. Complexity measures and cellular automata. *Complex Syst* **2**:409–440.
191. Li W. 1991. On the relationship between complexity and entropy for Markov chains and regular languages. *Complex Syst* **5**:381–399.
192. Arnold D. 1996. Information-theoretic analysis of phase transitions. *Complex Syst* **10**:143–155.
193. Bialek W, Nemenman I, Tishby N. 2001. Predictability, complexity and learning. *Neural Comput* **13**:2409–2463 (http://arxiv.org/abs/physics/0007070).
194. Crutchfield JP, Young K. 1990. Computation at the onset of chaos. In *Complexity, entropy, and the physics of information*, pp. 223–269. Ed. WH Zurek. Addison-Wesley, Reading, MA.
195. Shalizi CR, Crutchfield JP. 2001. Computational mechanics: Pattern and prediction, structure and simplicity. *J Stat Phys* **104**:817–879 (http://arxiv.org/abs/cond-mat/9907176).
196. Shalizi CR. 2003. Optimal nonlinear prediction of random fields on networks. *Discr Math Theor Comput Sci* AB(DMCS):11–30 (http://arxiv.org/abs/math.PR/0305160).
197. Shalizi, CR, Shalizi KL, Haslinger R. 2004. Quantifying self-organization with optimal predictors. *Phys Rev Lett* 93:118701 (http://arxiv.org/abs/nlin.AO/0409024).
198. Lewis HR, Papadimitriou CH. 1998. *Elements of the theory of computation*. Prentice-Hall, Upper Saddle River, NJ.
199. Clarke RW, Freeman MP, Watkins NW. 2003. Application of computational mechanics to the analysis of natural data: an example in geomagnetism. *Phys Rev E* 67:0126203 (http://arxiv.org/abs/cond-mat/0110228).

200. Chandler D. 1987. *Introduction to modern statistical mechanics*. Oxford UP, Oxford.
201. Reed WJ, Hughes BD. 2002. From gene families and genera to incomes and Internet file sizes: why power laws are so common in nature. *Phys Rev E* 66:067103.
202. Bak P, Tang C, Wiesenfeld K. 1987. Self-organized criticality: An explanation of 1/f noise. *Phys Rev Lett* **59**:381–384.
203. Jensen HJ. 1998. *Self-organized criticality: emergent complex behavior in physical and biological systems*. Cambridge UP, Cambridge.
204. Carlson JM, Doyle J. 1999. Highly optimized tolerance: A mechanism for power laws in designed systems. *Phys Rev E* **60**:1412–1427.
205. Carlson JM, Doyle J. 2000. Highly optimized tolerance: robustness and design in complex systems. *Phys Rev Lett* **84**:2529–2532.
206. Newman MEJ, Girvan M, Farmer JD. 2002. Optimal design, robustness, and risk aversion. *Phys Rev Lett* 89:028301 (http://arxiv.org/abs/cond-mat/0202330).
207. Shirky C. 2003. Power laws, weblogs, and inequality. In *Extreme democracy*. Ed. M Ratcliffe, J Lebkowsky. Forthcoming. First published online February 2003 (http://www.shirky.com/writings/powerlaw_weblog.html).
208. Drenzer D, Farrell H. 2004. The power and politics of blogs. *Persp Politics*. Submitted (http://www.utsc.utoronto.ca/~farrell/blogpaperfinal.pdf).
209. Crutchfield JP, Shalizi CR. 1999. Thermodynamic depth of causal states: objective complexity via minimal representations. *Phys Rev E* **59**:275–283 (http://arxiv.org/abs/cond-mat/9808147).
210. Huberman BA, Hogg T. 1986. Complexity and adaptation. *Physica D* **22**:376–384.
211. Wolpert DH, Macready WG. 2000. Self-dissimilarity: an empirically observable measure of complexity. In *Unifying themes in complex systems*. Ed. Y Bar-Yam. Perseus Books, Boston (http://www.santafe.edu/research/publications/wpabstract/199712087).
212. Sporns O, Tononi G, Edelman GM. 2000. Connectivity and complexity: the relationship between neuroanatomy and brain dynamics. *Neural Networks* **13**:909–992 (http://php.indiana.edu/~osporns/nn_connectivity.pdf).
213. Sporns O, Tononi G, Edelman GM. 2000. Theoretical neuroanatomy: Relating anatomical and functional connectivity in graphs and cortical connection matrices. *Cerebral Cortex* **10**:127–141.
214. Sporns O, Tononi G. 2002. Classes of network connectivity and dynamics. *Complexity* **7**:28–38 (http://php.indiana.edu/~osporns/complexity_2002.pdf).
215. Bates J, Shepard H. 1993. Measuring complexity using information fluctuation. *Phys Lett A* **172**:416–425 (http://physics.unh.edu/people/profiles/bates_shepard.pdf).
216. Boffetta G, Cencini M, Falcioni M, Vulpiani A. 2002. Predictability: A way to characterize complexity. *Phys Rep* **356**:367–474 (http://arxiv.org/abs/nlin.CD/0101029).
217. Feldman DP, Crutchfield JP. 1998. Measures of statistical complexity: why? *Phys Lett A* **238**:244–252 (http://hornacek.coa.edu/dave/Publications/MSCW.html).
218. Axelrod R, Cohen MD. 1999. *Harnessing complexity: organizational implications of a scientific frontier*. Free Press, New York.
219. Flake GW. 1998. *The computational beauty of nature: computer explorations of fractals, chaos, complex systems and adaptation*. MIT Press, Cambridge.
220. Holland JH. 1998. *Emergence: from chaos to order*. Addison-Wesley, Reading.
221. Simon HA. 1996. *The sciences of the artificial*, 3rd ed. MIT Press, Cambridge.
222. Boccara N. 2004. *Modeling complex systems*. Springer, Berlin.
223. Hand D, Mannila H, Smyth P. 2001. *Principles of data mining*. MIT Press, Cambridge.
224. Kearns MJ, Vazirani UV. 1994. *An introduction to computational learning theory*. MIT Press, Cambridge.
225. Vapnik VN. 1979/1982. *Estimation of dependencies based on empirical data*. Transl. S Kotz. Springer, Berlin. From *Vosstanovlyenie Zavicimostei po Empiricheckim Dannim*, Nauka, Moscow.
226. Vapnik VN. 1998. *Statistical learning theory*. Wiley, New York.

227. Nelles O. 2001. *Nonlinear system identification: from classical approaches to neural networks and fuzzy models*. Springer, Berlin.
228. Tukey JW. 1977. *Exploratory data analysis*. Addison-Wesley, Reading, MA.
229. Tufte ER. 1983. *The visual display of quantitative information*. Graphics Press, Cheshire, CT.
230. Tufte ER. 1990. *Envisioning information*. Graphics Press, Cheshire, CT.
231. Tufte ER. 1997. *Visual explanations: images and quantities, evidence and narrative*. Graphics Press, Cheshire, CT.
232. Abarbanel HDI. 1996. *Analysis of observed chaotic data*. Springer, New York.
233. Casdagli M, Eubank S, eds. 1992. *Nonlinear modeling and forecasting*. Addison-Wesley, Reading, MA.
234. Weigend AS, Gershenfeld NA, eds. 1993. *Time series prediction: forecasting the future and understanding the past*. Addison-Wesley, Reading, MA.
235. Azencott R, Dacunha-Castelle D. 1984/1986. *Series of irregular observations: forecasting and model building*. Springer, Berlin. Originally published as *Series d'observations irregulieres*, Masson, Paris.
236. Taniguchi M, Kakizawa Y. 2000. *Asymptotic theory of statistical inference for time series*. Springer, Berlin.
237. Amari S, Nagaoka H. 1993/2000. *Methods of information geometry*. American Mathematical Society, Providence, RI. Transl. D Harada from *Joho Kika no Hoho*, Iwanami Shoten Publishers, Tokyo.
238. Ay N. 2002. An information-geometric approach to a theory of pragmatic structuring. *Ann Probability* **30**:416–436 (http://www.mis.mpg.de/preprints/2000/prepr5200-abstr.html).
239. Ay N. 2001. *Information geometry on complexity and stochastic interaction*. Technical Report 95-2001, Max Planck Institute for Mathematics in the Sciences (http://www.mis.mpg.de/preprints/2001/prepr9501-abstr.html).
240. Cutler CD, Kaplan DT, eds. 1997. *Nonlinear dynamics and time series: building a bridge between the natural and statistical sciences*. American Mathematical Society, Providence, RI.
241. Tong H. 1990. *Nonlinear time series: a dynamical systems approach*. Oxford UP, Oxford.
242. Pearson RK. 1999. *Discrete-time dynamic models*. Oxford UP, New York.
243. Stengel RF. 1994. 1986. *Optimal control and estimation*. Dover, New York. Originally published as *Stochastic optimal control: theory and application*. Wiley, New York.
244. Lind D, Marcus B. 1995. *An introduction to symbolic dynamics and coding*. Cambridge UP, Cambridge.
245. Gershenfeld N. 1999. *The nature of mathematical modeling*. Cambridge UP, Cambridge.
246. Ilachinski A. 2001. *Cellular automata: a discrete universe*. World Scientific, Singapore.
247. Chopard B, Droz M. 1998. *Cellular automata modeling of physical systems*. Cambridge UP, Cambridge.
248. Toffoli T, Margolus N. 1987. *Cellular automata machines: a new environment for modeling*. MIT Press, Cambridge.
249. Farmer JD, Toffoli T, Wolfram S, eds. 1984. *Cellular automata: proceedings of an interdisciplinary workshop, Los Alamos, NM 87545, March 7–11*. North-Holland, Amsterdam. Also published in *Physica D* **10**(1–2), 1984.
250. Gutowitz H, ed. 1991. *Cellular automata: theory and experiment*. MIT Press, Cambridge. Also published in *Physica D* **45**(1–3), 1990.
251. Manneville P, Boccara N, Vichniac GY, Bidaux R, eds. 1990. *Cellular automata and modeling of complex systems: proceedings of the winter school, Les Houches, France, February 21–28, 1989*. Springer, Berlin.
252. von Neumann J. 1966. *Theory of self-reproducing automata*, Ed. and completed by AW Burks. U Illinois P, Urbana.
253. Burks AW, ed. 1970. *Essays on cellular automata*. U Illinois P, Urbana.
254. Poundstone W. 1984. *The recursive universe: cosmic complexity and the limits of scientific knowledge*. William Morrow, New York.

255. Langton CG, ed. 1988. *Artificial life*. Addison-Wesley, Reading, MA.
256. Wolfram S. 1994. *Cellular automata and complexity: collected papers*. Addison-Wesley, Reading, MA (http://www.stephenwolfram.com/publications/books/ca-reprint/).
257. Griffeath D, Moore C, eds. 2003. *New constructions in cellular automata*. Oxford UP, Oxford.
258. Rivet J, Boon J. 2001. *Lattice gas hydrodynamics*. Cambridge UP, Cambridge.
259. Doolen GD, ed. 1989. *Lattice gas methods for partial differential equations: a volume of lattice gas reprints and articles*. Addison-Wesley, Reading, MA.
260. Doolen GD, ed. 1991. *Lattice gas methods: theory, applications, and hardware*. MIT Press, Cambridge. Also published in *Physica D* **47**(1–2), 1991.
261. Axelrod R. 1984. *The evolution of cooperation*. Basic Books, New York.
262. Varela FJ, Bourgine P, eds. 1992. *Toward a practice of autonomous systems: proceedings of the first European conference on artificial life*. MIT Press, Cambridge.
263. Epstein JM, Axtell R. 1996. *Growing artificial societies: social science from the bottom up*. MIT Press, Cambridge.
264. Epstein JM. 1999. Agent-based computational models and generative social science. *Complexity* **4**(5):41–60.
265. Macy MW, Willer R. 2002. From factors to actors: Computational sociology and agent-based modeling. *Ann Rev Sociol* **28**:143–66.
266. Gimblett R, ed. 2001. *Integrating geographic information systems and agent-based modeling techniques for understanding social and ecological processes*. Oxford UP, Oxford.
267. Kohler TA, Gumerman GJ, eds. 2000. *Dynamics in human and primate societies: agent-based modeling of social and spatial processes*. Santa Fe Institute Studies in the Sciences of Complexity. Oxford UP, Oxford.
268. Bonabeau E. 1997. From classical models of morphogenesis to agent-based models of pattern formation. *Artificial Life* **3**:191–211.
269. Camazine S, Deneubourg J-L, Franks NR, Sneyd J, Theraulaz G, and Bonabeau E. 2001. *Self-organization in biological systems*. Princeton UP, Princeton.
270. Gardiner CW. 1990. *Handbook of stochastic methods for physics, chemistry, and the natural sciences*, 2nd ed. Springer, Berlin.
271. Newman MEJ, Barkema GT. 1999. *Monte Carlo methods in statistical physics*. Clarendon Press, Oxford.
272. MacKeown PK. 1997. *Stochastic simulation in physics*. Springer, Singapore.
273. King G. 1997. *A solution to the ecological inference problem: reconstructing individual behavior from aggregate data*. Princeton UP, Princeton.
274. Shannon CE. 1948. A mathematical theory of communication. *Bell Syst Tech J* **27**:379–423. Reprinted in *The mathematical theory of communication*. Ed. Shannon CE, Weaver W. U Illinois P, Urbana, 1963 (http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html).
275. Hartley RVL. 1928. Transmission of information. *Bell Syst Tech J*, pp. 535–563.
276. Gray RM. 1990. *Entropy and information theory*. Springer, New York (http://www-ee.stanford.edu/~gray/it.html).
277. Peliti L, Vulpiani AV, eds. 1988. *Measures of complexity: proceedings of a conference held in Rome, September 30–October 2, 1987*. Springer, Berlin.
278. Edmonds BH. 1997. *Hypertext bibliography of measures of complexity* (http://www.cpm.mmu.ac.uk/~bruce/combib/).
279. Grünwald P. 2005. A tutorial introduction to the minimum description length principle. In *Advances in minimum description length: theory and applications*. Ed. P Grünwald, IJ Myung, M Pitt. MIT Press, Cambridge (http://arxiv.org/abs/math.ST/0406077).
280. Skouras S. 2001. *Decisionmetrics: A decision-based approach to econometric modeling*. Technical Report 01-11-064, Santa Fe Institute (http://www.santafe.edu/research/publications/wpabstract/200111064).
281. Vidyasagar M. 1997. *A theory of learning and generalization: with applications to neural networks and control systems*. Springer, Berlin.

282. William of Ockham. 1964. *Philosophical writings: a selection*. Transl. with an introduction by Philotheus Boehner. Bobbs-Merrill, Indianapolis. Originally published in various European cities during the early 1300s.
283. Lehmann EL. 1997. *Testing statistical hypotheses*, 2nd ed. Springer Texts in Statistics. Springer, Berlin.
284. Churchland PS, Sejnowski TJ. 1992. *The computational brain*. MIT Press, Cambridge.
285. Giunti M. 1997. *Computation, dynamics, and cognition*. Oxford UP, Oxford.
286. Margolus N. 1999. Crystalline computation. In *Feynman and computation: exploring the limits of computers*, pp. 267–305. Ed. AJG Hey. Perseus Books, Reading, MA (http://arxiv.org/abs/nlin.CG/9811002).
287. Hirotugu Akaike. 1998. *Selected papers of Hirotugu Akaike*. Ed. E Parzen, K Tanabe, G Kitagawa. Springer, Berlin.
288. Shannon CE, Weaver W, eds. 1963. *The mathematical theory of communication*. U Illinois P, Urbana.

2

NONLINEAR DYNAMICAL SYSTEMS

Joshua E. S. Socolar

Physics Department, Duke University, Durham, North Carolina

The concepts and techniques developed by mathematicians, physicists, and engineers to characterize and predict the behavior of nonlinear dynamical systems are now being applied to a wide variety of biomedical problems. This chapter serves as an introduction to the central elements of the analysis of nonlinear dynamics systems. The fundamental distinctions between linear and nonlinear systems are described and the basic vocabulary used in studies of nonlinear dynamics introduced. Key concepts are illustrated with classic examples ranging from simple bistability and hysteresis in a damped, driven oscillator to spatiotemporal modes and chaos in large systems, and to multiple attractors in complex Boolean networks. The goal is to give readers less familiar with nonlinear dynamics a conceptual framework for understanding other chapters in this volume.

1. INTRODUCTION

The latter half of the twentieth century saw remarkable advances in our understanding of physical systems governed by nonlinear equations of motion. This development has changed the scientific worldview in profound ways, simultaneously supplying a dose of humility—the recognition that deterministic equations do not guarantee quantitative predictability—and a great deal of insight into the qualitative and statistical aspects of dynamical systems. One of the byproducts has been the realization that the mathematical constructs developed for modeling simple physical systems can be fruitfully applied to more complex systems, some of which are of great interest to the biomedical community. Examples range from electrical signal propagation in cardiac tissue, where one

Address correspondence to: Joshua E.S. Socolar, Physics Department, Duke University, 107 Physics Building, Durham, NC 27708; 919-660-2557 (socolar@phy.duke.edu).

might well expect physical theory to play a significant role, to the logic of neural networks or gene regulation, where the role of physical/mathematical modeling may be less obvious.

The establishment of a connection between physical theory and biomedical observations generally involves a combination of fundamental physical reasoning and *a posteriori* model validation. The field of nonlinear dynamics is of crucial importance for both purposes. It provides both the techniques for analyzing the equations of motion that emerge from the physical theory and a useful language for framing questions and guiding the process of model validation (as noted by Shalizi, Part II, chapter 1, this volume). Because the concept of a nonlinear dynamical system is rich enough to encompass an extremely broad range of processes in which the future configurations of a system are determined by its past configurations, the methods of analysis developed in the field are useful in a huge variety of contexts.

To appreciate the validity of a particular research result involving the application of nonlinear dynamical theory and properly interpret the specific conclusions, it is important to grasp the broad conceptual basis of the work. The purpose of this chapter is to explain the meaning and crucial consequences of nonlinearity so as to provide an operational understanding of the principles underlying the modeling discussed in other chapters. (Almost all of the chapters in the present volume rely on techniques and approaches whose roots lie in the development of nonlinear dynamics as a discipline. The chapters by Subramanian and Narang [Part III, chapter 2.2], Lubkin [Part III, chapter 3.1], Tabak [Part III, chapter 5.2], Solé [Part III, chapter 6.2], and Segel [Part III, chapter 4.1] all make direct reference to nonlinear dynamical models of precisely the sort discussed here.) Along the way, certain fundamental terms will be defined and illustrated with examples, but the reader interested in the details of the mathematics will have to look elsewhere. Two excellent textbooks that do not require familiarity with mathematical concepts beyond basic calculus are (28) and (3). For treatments of more advanced topics, a good place to start is (25).

Nonlinear dynamics enters the biomedical literature in at least three ways. First, there are cases in which experimental data on the temporal evolution of one or more quantities are collected and analyzed using techniques grounded in nonlinear dynamical theory, with minimal assumptions about the underlying equations governing the process that produced the data. That is, one seeks to discover correlations in the data that might guide the development of a mathematical model rather than guess the model first and compare it to the data. (See the chapter by Shalizi [Part II, chapter 1] in this volume for a discussion of time-series analysis.) Second, there are cases in which symmetry arguments and nonlinear dynamical theory can be used to argue that a certain simplified model should capture the important features of a given system, so that a phenomenological model can be constructed and studied over a broad parameter range. Often this leads to models that behave qualitatively differently in different regions

of parameter space, and one region is found to exhibit behavior quite similar to that seen in the real system. In many cases, the model behavior is rather sensitive to parameter variations, so if the model parameters can be measured in the real system the model shows realistic behavior at those values, and one can have some confidence that the model has captured the essential features of the system. Third, there are cases in which model equations are constructed based on detailed descriptions of known (bio)chemistry or biophysics. Numerical experiments can then generate information about variables inaccessible to physical experiments.

In many cases, all three approaches are applied in parallel to the same system. Consider, for example, the problem of fibrillation in cardiac tissue. At the cellular level, the physics of the propagation of an electrical signal involves complicated physicochemical processes. Models involving increasingly realistic descriptions of the interior of the cell, its membrane, and the intercellular medium are being developed in attempts to include all the features that may give rise to macroscopic properties implicated in fibrillation. (See, for example, articles in (5).) At the same time, recognizing the general phenomenon of action potential propagation as similar to chemical waves in reaction-diffusion systems allows one to construct plausible, though idealized, mathematical models in which phenomena quite similar to fibrillation can be observed and understood (10,11)). These models can then be refined using numerical simulations that incorporate more complicated features of the tissue physiology. In parallel with these theoretical efforts, experiments on fibrillation or alternans in real cardiac tissue yield time series data that must be analyzed on its own (with as little modeling bias as possible) to determine whether the proposed models really do capture the relevant physics (13).

Almost all mathematical modeling of biomedical processes involves a significant computational component. This is less a statement about the complexity of biomedical systems than a reflection of the mathematical structure of nonlinear systems in general, even simple ones. Indeed, in large measure the rise of nonlinear dynamics as a discipline can be attributed to the development of the computer as a theoretical tool. Though one can often prove theorems about general features of solutions to a set of nonlinear equations, it is rarely possible to exhibit those solutions in detail except through numerical computation. Moreover, it is often the case that the numerical simulation has to be done first in order to give some direction to theoretical studies. Though the catalogue of well-characterized, generic behaviors of deterministic nonlinear systems is large and continues to grow, there is no *a priori* method for classifying the expected behavior of a particular nonlinear dynamical system unless it can be directly mapped to a previously studied example.

Rather than attempting a review of the state of the art in time-series analysis, numerical methods, and theoretical characterization of nonlinear dynamical systems, this chapter presents some of the essential concepts using a few exam-

ples. Section 2 presents the vocabulary needed for the description of dynamical systems of any type, linear or nonlinear, large or small, continuous or discrete. Section 3 presents the fundamental ideas relevant for understanding the behavior of small systems, i.e., systems characterized by a small number of dynamical variables. It begins with a discussion of the concept of nonlinearity itself, then proceeds to build on it using two canonical examples: the damped, driven oscillator and the logistic map. In section 4 new issues that arise in large systems are introduced, again in the context of two characteristic examples: the cardiac system and the Boolean model of genetic regulatory networks. It is hoped that these discussions will provide a context that will help readers understand the import of other chapters in this book.

2. DYNAMICAL SYSTEMS IN GENERAL

The term **dynamical system** refers to any physical or abstract entity whose configuration at any given time can be specified by some set of numbers, called **system variables**, and whose configuration at a later time is uniquely determined by its present and past configurations through a set of rules for the transformation of the system variables. Two general types of transformation rules are often encountered. In **continuous-time systems** the rules are expressed as equations that specify the time derivatives of the system variables in terms of their current (and possible past) values. In such cases, the system variables are real numbers that vary continuously in time. The Newtonian equations of motion describing the trajectories of planets in the solar system represent a continuous-time dynamical system. In **discrete-time systems** the rules are expressed as equations giving new values of the system variables as functions of the current (and possibly past) values. Though classical physics tells us that all systems are continuous-time systems at their most fundamental level, it is often convenient to use descriptions that describe the system configurations only at a discrete set of times and describe the effects of the continuous evolution as discrete jumps from one configuration to another.

A set of equations describing a continuous-time dynamical system takes the form

$$\dot{x}(t) = \mathbf{f}(\mathbf{x}(t); \mathbf{p}, t). \qquad [1]$$

Here the components of the vector $\mathbf{x}$ are the system variables and the vector $\mathbf{f}$ represents a function of all of the system variables at fixed values of the parameters $\mathbf{p}$. The overdot on the left indicates a first time derivative.[1] Note that $\mathbf{f}$ can depend explicitly on time, as would be the case, for example, in a system driven by a time-varying external force. In some systems, time delays associated with

finite speeds of signal propagation cause **f** to depend also on values of **x** at times earlier than *t*.

In **spatially extended systems**, each system variable is a continuous function of spatial position as well as time and the equations of motion take the form of partial differential equations:

$$\frac{\partial x}{\partial t} = \mathbf{f}(\mathbf{x}, \nabla\mathbf{x}, \nabla\mathbf{x}^2, ...; \mathbf{p}, t) \,, \qquad [2]$$

where ∇ is the spatial gradient, ∇^2 is the Laplacian, and the dots represent higher-order derivatives of **x**. The parameters **p** may also be externally imposed functions of position, as would be the case for a system evolving in an inhomogeneous environment. (If the environment itself is affected by the system variables, however, then variables representing the environment become system variables rather than external parameters.)

A set of equations describing a discrete-time dynamical system takes the form

$$\mathbf{x}(t + 1) = \mathbf{F}(\mathbf{x}(t); \mathbf{p}t). \qquad [3]$$

Here the function **F** directly gives the new **x** at the next time step, rather than the derivative from which a new **x** can be calculated. The function **F** is often referred to as a **map** that takes the system from one time step to the next.

In all cases, the evolution of the system is described as a motion in **state space**, the space of all possible values of the vector **x**. A **trajectory** is a directed path through state space that shows the values of the system variables at successive times. The theory of dynamical systems is concerned with classifying the types of trajectories that can occur, determining whether they are robust against small variations in the system variables, categorizing the ways in which the possible trajectories change as parameters **p** are varied, and developing techniques both for simulating trajectories numerically and inferring the structure of trajectories from incomplete sets of observations of the system variables. The most basic structures arising in the classification of state space trajectories—fixed points, limit cycles, transients, basins of attraction, and stability—will be explained below as they arise in the context of some simple examples.

3. LINEAR SYSTEMS AND SOME BASIC VOCABULARY

A **linear system** is one for which any two solutions of the equations of motion can be combined through simple addition to generate a third solution, given appropriate definitions of the zeros of the variables. The system of equations can be extremely complicated, representing large numbers of variables with all sorts

of logical structures associated with the connections between them, including complicated networks of causal relationships among variables, time delays between cause and effect, arbitrarily complex spatial inhomogeneities, or even externally imposed noise. The way to recognize a linear dynamical system is that its equations of motion will involve only polynomial functions of degree one in the system variables; there will be no products of different system variables or nontrivial functions of any individual variable. (Examples of nontrivial functions include squares or square roots, threshold functions that specify discontinuous switching of parameters as the system variables change, or quantities that have simple geometric interpretations but turn out to be complicated functions of the fundamental variables.) Gradients of any order may appear, however, as well as coefficients that are nontrivial functions of spatial position and time.[2]

For all types of linear systems, the constraint that the sum of any two solutions also must be a solution has profound consequences. Simply put, the full range of behavior of a linear system is understood as soon as its behavior in an infinitesimal region of its state space is understood. In the absence of an external driving force, there is one special solution to any linear system where the variables are time-independent—everything just sits still. This is called the **fixed point**. A trivial example is the equilibrium point of a weight hanging from an ideal spring in a perfectly uniform gravitational field. Here the system variables are the position and velocity of the weight, which can both be defined to be zero at the fixed point. Another example is the surface of a liquid that may have ripples governed by surface tension (capillary waves) described by a linear theory. The system variables here are a field representing the height of the liquid at all points in space and the time derivative of that field. Again, the system variables can be defined to be zero at the fixed point corresponding to a quiescent, flat surface.

When variables are defined so as to be zero at the fixed point, linearity implies that every solution can be multiplied by an arbitrary factor to yield another solution. Thus solutions with arbitrarily large amplitudes can be multiplied by an arbitrarily small factor to yield solutions infinitesimally close to the fixed point, indicating that the nature of solutions very near the fixed point determines all of the possible solutions. The situation is further simplified by the fact that solutions in the vicinity of the fixed point come in only three types—stable, unstable, and marginal.

In a linear **stable system**, all solutions asymptotically approach the fixed point as time progresses. The typical case is that beginning from any initial point in state space, the variables decay toward the fixed point by first rapidly approaching a particular line in state space and then relaxing exponentially along that line toward the origin.[3] In an **unstable system**, all solutions that do not start exactly on the fixed point diverge from it exponentially at long times. The **marginal** case, in which the variables neither decay to zero nor diverge, occurs pri-

marily in "Hamiltonian" systems, in which conservation of energy prohibits convergence or divergence of nearby phase space trajectories. Such **dissipationless** systems are of great interest in quantum mechanics and statistical mechanics, but systems of interest in biomedicine always involve strongly **dissipative** processes, which include, for example, all processes involving frictional forces. Marginal stability then occurs only as a very special case where parameters have been carefully tuned, though there have been suggestions that marginal stability can reappear spontaneously in certain self-organizing nonlinear systems (2).

Now we are often interested in dissipative systems that are subjected to external driving of some sort, whether it be a steady input of energy or a driving with more complicated temporal structure. In such systems, the notion of a fixed point must be generalized to include steady or regularly repeating motions. For example, if the ceiling from which a weighted, damped spring is hanging were constantly oscillating up and down, the weight would not sit at a fixed point but could exhibit regular oscillations with a period that matches the oscillation of the ceiling. Such trajectories are called **limit cycles**, and, like fixed points, they may be stable or unstable.

Stable fixed points and limit cycles are called **attractors**, as trajectories in state space eventually flow toward them and then stay very close to them at long times. If we begin observing a system when it is far from its attractor and watch for a long time, we will be able to detect its motion toward the attractor for a while, but at some point it will be so close to the attractor that we can no longer resolve the difference. The portion of the trajectory over which we can observe progress toward the attractor is called a **transient**. The set of points in state space that lie on transients associated with a particular attractor is called the **basin of attraction** of the attractor. In a stable linear system, all points in state space lie in the same basin of attraction. In other words, for any initial configuration of the system variables, the ultimate fate of the system is the same fixed point or limit cycle.

4. NONLINEAR EFFECTS IN SIMPLE SYSTEMS

Linear systems are often studied in great detail. They can be solved exactly and hence make for good textbook problems; and linear equations can be used as good approximations to nonlinear ones in situations where the trajectories stay very close to a stable fixed point or limit cycle. They cannot capture, however, many of the most important qualitative features of real systems.

In a **nonlinear system**, the equations of motion include at least one term that contains the square or higher power, a product of system variables (or more complicated functions or them), or some sort of threshold function, so that the addition of two solutions does *not* yield a valid new solution, no matter how the

system variables are defined. *All physical systems describable in terms of classical equations of motion are nonlinear.* (The quantum mechanical theory of atoms and molecules is a linear theory: the connection between it and the comparatively macroscopic processes at the cellular level and larger is beyond the scope of the present discussion.) In all real systems, deviations of large enough amplitude require nonlinear terms in the relevant model. There is no such thing as a truly linear spring or a waves on a fluid obeying a perfectly linear equation of motion. This is why the study of nonlinear dynamics has such broad relevance.

The consequences of nonlinearity are profound. Most importantly, nonlinear systems may contain multiple attractors, each with its own basin of attraction. Thus the fate of a nonlinear dynamical system may depend on its initial state, and a whole new set of phenomena arises associated with the way in which basins of attraction shift as parameters are varied.

Nonlinearity can also give rise to an entirely new type of attractor. Limit cycles in nonlinear systems may be quite complicated, circling around in a bounded region of state space many times before finally closing on themselves. It is even possible (and quite common) for a trajectory to be confined to a region of state space where there are no stable limit cycles or fixed points. The system then appears to follow an irregular trajectory that is said to lie on a **strange attractor**. The trajectory comes arbitrarily close to closing on itself, but never quite does, and two identical systems that come arbitrarily close to each other in state space diverge rapidly thereafter. A strange attractor is the state space structure associated with the phenomenon known as **chaos**.

Finally, in spatially extended systems nonlinearities can give rise to **pattern formation**, the spontaneous creation of attractors with nontrivial spatial structure in a system with no externally imposed inhomogeneities. (See (33,8,20) for textbook treatments of pattern formation and spatiotemporal structures in large systems. For more technical treatments from a physics perspective, see (17,6,4)). Examples include the formation of stripes or spots in chemical reaction-diffusion systems and excitable media, which find applications in such processes as butterfly wing coloration and cardiac electrodynamics.

The essential features of nonlinear systems can be illustrated with the simplest of examples, the driven, damped oscillator. Figure 1 shows a picture of the system. We assume the spring is nonlinear: it gets stiffer under compression and softer under extension. With an appropriate definition of the zero of x, the position of the mass, the equation of motion can be written as $m\ddot{x} = -\gamma\dot{x} + k(h-x) + k'(h-x)^2$, where m, γ, k, and k' are constants and h is the deviation of the ceiling from its average height. Defining $z = \dot{x}$, the equation of motion can be written as two coupled equations in a two-variable state space: $\dot{v} = -\gamma v - k(h-x) + k'(h-x)^2$ and $\dot{x} = v$. We will consider cases in which the ceiling oscillates according to $h = \alpha \sin(\omega t)$.

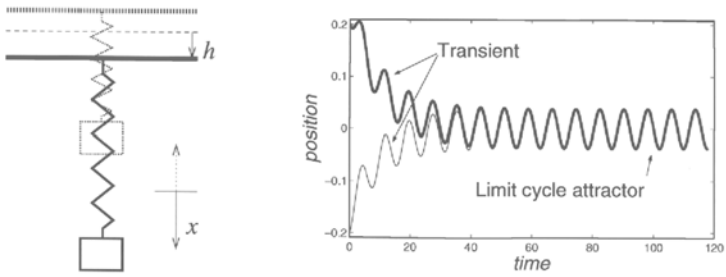


Figure 1. A simple nonlinear dynamical system. Left: A mass is attached to the ceiling by a spring. The force exerted on the mass by gravity and the spring together is $k(h - x) + k'(h - x)^2$, where h is the displacement of the ceiling from its nominal height and x is the displacement of the mass from its resting position. The solid and dotted images represent the spring and mass at different times during a cycle in which the ceiling is oscillating. Right: Two time series for the linear case $k' = 0$.

Figure 1 also shows the behavior of the mass when $k' = 0$, which makes the system linear. Two time series are shown for a particular choice of the drive frequency ω, and one sees that the long term behavior in the two cases is identical. The difference between the two curves in the early stages corresponds to transients that depend on the details of the initial configuration. The behavior that is reached in the long term is called a limit cycle attractor.

When $k' = 0$, one can see immediately from the equations that the strength of the drive, α, is not an important parameter in determining the qualitative structure of the motion. The solution for a given α can simply be rescaled by multiplication so as to correspond to a different value of α.

When k' is nonzero, so that the system is nonlinear, one often finds behavior similar to that shown in Figure 1, i.e., convergence of all transients (in the domain of initial conditions of interest) to the same solution. In the nonlinear case, however, it is possible to see quite different behavior. Figure 2 shows one simple nonlinear effect: one can have two different long term solutions for a single value of the system parameters. The differences produced by different initial conditions in this case are not limited to transient effects. This phenomenon of **bistability** is a generic feature of nonlinear dynamics, and its presence in all sorts of biomedical systems indicates that nonlinearities play a fundamental role in their function.

The presence of bistability in a system raises the question of which initial conditions will lead to which orbit, or which points in state space lie in which basin of attraction. Even for systems as simple as the damped, driven oscillator,

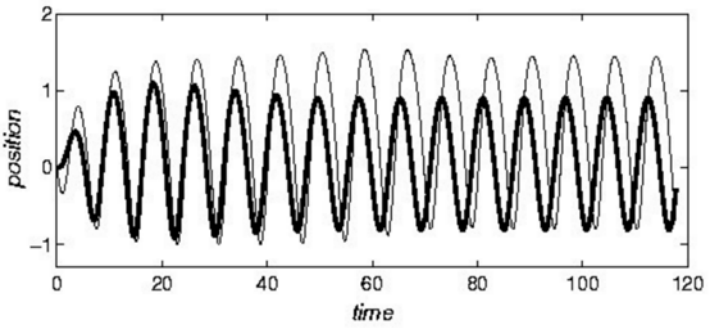


Figure 2. Two time series for the same system as in Figure 1, but with parameters set such that two different long time solutions exist. Which solution is realized depends upon the choice of initial conditions.

the boundaries in state space of the basins of attraction can be quite complex, making it extremely difficult to predict which orbit the system will eventually reach from a specified initial configuration. (See chapter 5 of (24) for details.) In studying more complicated systems, one often finds multiple attractors with basins of attraction that can vary widely in size.

A feature generically associated with bistability is **hysteresis**, the dependence of the observed solution on the direction in which a parameter is varied. For example, Figure 3 shows curves indicating the amplitude of oscillation of the mass in our simple model as the drive frequency ω is slowly ramped up and then down. For small ω and large ω there is only one attractor. In the intermediate range, however, there are two (plus an unstable periodic orbit that is not seen). During the upsweep, the system stays in the basin of one of the attractors until that attractor is destroyed, at which point it is attracted to the stable orbit of significantly different amplitude. During the downsweep, the same process happens in reverse, except that the jump occurs at a lower value of ω.

The jump to a different solution in a hysteretic system is an example of a **bifurcation**. More generally, the theory of bifurcations describes the transitions that occur between structurally different solutions as a system parameter is varied. Such transitions may correspond to the creation or destruction of fixed points or simply to changes in the stability properties of existing fixed points. In the oscillator example, one may observe a bifurcation upon variation of any of the parameters k, k', γ, α, or ω. The precise values of the parameters at which a bifurcation occurs are called a **critical point** in parameter space. The mathematical theory of how solutions can be created or destroyed as parameters are varied is well developed and full of beautiful structures (12,24,25,28).

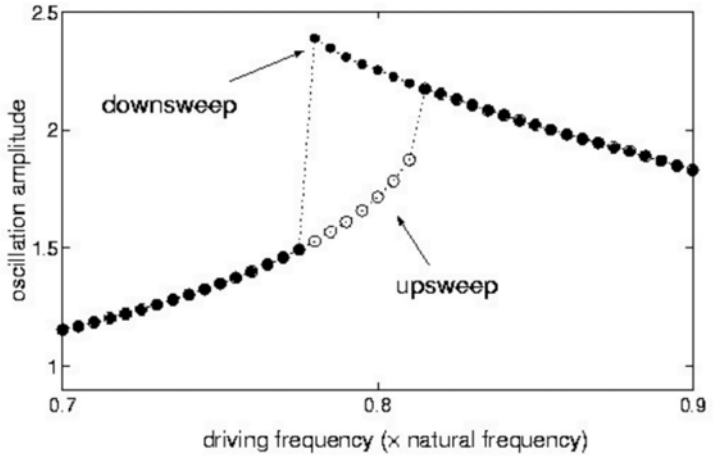


Figure 3. The limit cycle amplitudes for different values of the driving frequency in a nonlinear oscillator. The solid circles correspond to oscillations observed as the driving frequency is slowly ramped up from 0.7. The open circles correspond to oscillations observed as the driving frequency is slowly ramped down from 0.9. The dotted lines are guides to the eye. The drive frequency is measured in units of the natural frequency of the linear oscillator. The time series shown in Figure 2 correspond to the two limit cycles that coexist at a driving frequency of 0.8 for $\gamma = 0.18$, $k = 1$, $k' = 0.5$, and $\alpha = 0.3$.

As mentioned above, nonlinear dynamical systems sometimes exhibit chaos, motion that never settles into a fixed point or limit cycle. The system stays confined to a finite region of state space, but never returns precisely to one of the points it has visited before. In fact, almost all dynamical systems are chaotic for some range of parameter values and our simple driven oscillator is no exception. An example of a chaotic orbit in this system is shown in Figure 4.

Much attention has been devoted to the characterization of strange attractors. Three aspects of the theory are of particular interest for practical purposes. The first is the determination of the **dimension of the attractor**. The attractor itself is a geometric object, a set of points in state space, that has a dimension which can be non-integral, sometimes called a **fractal dimension**. Most importantly, the dimension is finite and lower than the dimension of the full state space. Its origin in a set of deterministic equations for a relatively small number of variables makes it fundamentally different from the erratic trajectories associated with random, or stochastic, processes. This raises the possibility that an experimentally observed time series suggesting erratic, unpredictable behavior actually arises from a deterministic, though nonlinear, set of equations. A beauti-

ful theorem shows that the important topological features of strange attractors can be reconstructed from a time series measurement of a single variable (30). One can construct the state space structure of the attractor using time-delayed values of that variable rather than synchronous measurements of all of the system variables. The reconstruction is said to be **embedded** in a space of dimension equal to the number of time delays used and the theorem says that as long as the embedding dimension is large enough, the topological features of the trajectory will be accurately reconstructed. (See (24) for a discussion of embedding.) This has led to the development of a number of computational tools for analyzing time series data to determine whether a system can be modeled using a small number dynamical variables or not, though prohibitive difficulties are almost always encountered if more than about 10 variables are required. (Aspects of this sort of time-series analysis are discussed in the preceding chapter by Shalizi.)

The second item of interest is the characterization of the strength of the chaos (its degree of unpredictability) via the **Lyapunov exponents**. Suppose a system is following a trajectory in state space that is a long time solution to the equations of motion. We imagine an almost exact copy of the system at time $t = 0$. The copy has exactly the same parameter values as the original—it is the same system—but the variable values at $t = 0$ differ by a tiny amount from the original. In a chaotic system, the difference between the variable values in the copy and the original will grow (on average) with time. Ignoring short-time scale fluctuations, the difference between a given variable, say x, in the two systems will grow exponentially: $\delta x = \delta x_0 \exp(\lambda t)$. The quantity λ, with dimensions of 1/time, is called the Lyapunov exponent.[4]

A large λ indicates rapid divergence of nearby trajectories, which implies that prediction of future values of the variables requires extremely precise knowledge of the present values. The consequence of exponential divergence is that accurate prediction becomes prohibitively difficult over times larger than a few times λ^{-1}. This is only a quantitative issue: chaos does *not* imply some mysterious new source of randomness of the type, say, that is found in measurements on quantum systems. Nevertheless, the mathematics of exponential growth makes a qualitative difference in practice for would-be predictors of the motion. The increase in precision of measurement required to make accurate predictions is so rapid within the desired time interval covered that useful long-time prediction is impossible.

The third item of interest is the nature of the transition to chaos as a parameter is varied, i.e., the type of bifurcation that leads to the emergence of a strange attractor. Perhaps the most celebrated result in chaos theory is the proof by Feigenbaum that all discrete maps in a broad class go through a quantitatively identical transition, dubbed the **period-doubling route to chaos** (9). In a period-doubling bifurcation, a periodic orbit undergoes a change in which only every other cycle is identical. One then still has a periodic orbit, but its period is twice

as long as the original. The deviations from the original simple orbit can grow larger as the bifurcation parameter is ramped further, eventually leading to a second period-doubling bifurcation, so that the new orbit has a period that is four times longer than the original. In fact, an entire period-doubling cascade can occur within a finite range of the bifurcation parameter, leading finally to a chaotic attractor. Feigenbaum showed that the sequence of bifurcations has a structure that is the same for a large class of discrete maps. The details of the particular map under study become irrelevant as we approach the end of the period-doubling cascade.

To see a sequence of period-doubling transitions leading to chaos, one need look no farther than our simple nonlinear oscillator. Figure 4 shows a set of solutions, with $k = 1$, $k' = 0.5$, $\gamma = 0.5$, and $\omega = 0.8$, for four different values of the drive amplitude α. The time series on the top row show that bifurcations occur as the drive amplitude is varied. At some critical point between the first and second panels ($\alpha = 0.7$ and $\alpha = 0.76$) the solution undergoes a structural change. It begins with a limit cycle with some period (the exact value is unimportant). It then changes to a limit cycle that has a period approximately twice as long as the original. In terms of the original, every other cycle looks different. In a periodically forced system such as this one, the second limit cycle is often called a "2:1 state," referring to the fact that there are two periods of the driver for every one period of the limit cycle. By the time we get to the third panel, the 2:1 state has itself undergone a period-doubling bifurcation, leading to a 4:1 state. In between the third and fourth panels, an infinite sequence of period doublings has occurred, leading, finally, to a strange attractor and its trademark erratic time series.

The bottom row of Figure 4 shows a view of the same motion that clarifies the nature of the bifurcations a bit. Each of these plots is a projection of the trajectory corresponding to the time series above it. In the present case, the state space is three dimensional, the three dimensions corresponding to the position and velocity of the mass and the position of the ceiling. The figures show the projection of a path through the 3D space onto a 2D plane. Here the differences between the four solutions are easier to see at a glance.[5]

A standard method for analyzing such a situation is through the construction of a discrete **return map** from its current position in state space to its position one drive period later. This is the theoretical equivalent of taking a movie of the motion with a strobe light that flashes in synchrony with the ceiling oscillations. In that movie, the ceiling will appear to sit still, while the weight will jump from point to point according to the map. If the weight is on a simple limit cycle as described above, it will appear fixed in the movie. In this way, we see that a periodic orbit of a continuous-time system corresponds to a fixed point of a discrete-time system.[6] The motion of the system in the strobed movie is said to occur on a **Poincaré section** of the state space. For practical and analytical

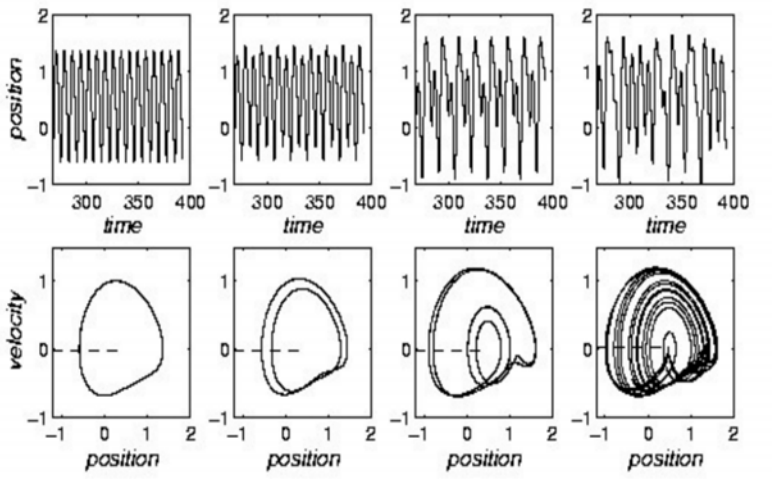


Figure 4. The period doubling route to chaos in the oscillator with $k = 1$, $k' = 0.5$, $\gamma = 0.5$, and $\omega = 0.8$ for $\alpha = 0.700$, 0.760, 0.794, and 0.810. Top: time series of the position (x vs. t). Bottom: phase space plots ($\dot{x}$ vs. x).

reasons, one often works directly with a discrete map that takes one point on the Poincaré section into the next, rather than the underlying differential equations.

Since the system is deterministic, the map that takes one point to the next is unique. In the oscillator example above, the Poincaré section may be taken to be the half plane corresponding to the points where the phase of the drive has some chosen value. The dashed line drawn on the lower set of plots in Figure 4 schematically represents the projection of this plane onto the $x - \dot{x}$ plane. As time progresses, the system keeps looping around the state space in a clockwise direction, passing through the Poincaré section once every time around. Each time the section is crossed, the position and velocity of the oscillator are observed. In the present case (and many others) it is sufficient to keep track of only one variable, say the position at each piercing of the Poincaré section. In this way we obtain a discrete sequence of x values, x_n. A return map f defined by $x_{n+1} = f(x_n;\alpha)$ can then be constructed, where we write α explicitly to indicate that the map depends on the bifurcation parameter.

For the simple limit cycle on the left, the system returns to the same point on every cycle. For this value of α, the fixed point x^* satisfying $x^* = f(x^*)$ is stable. For the second case shown in the figure, x_n will alternate between two values. In this case it is the map f^2 (two successive applications of f) that has a

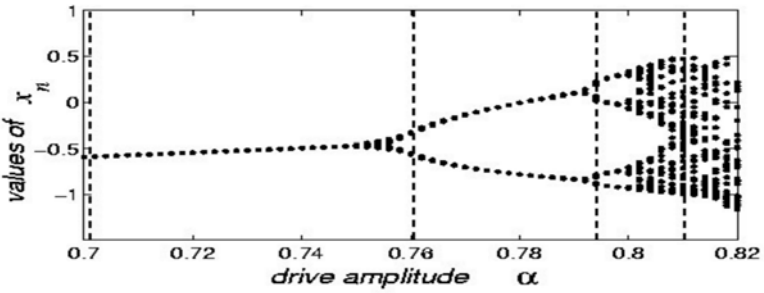


Figure 5. The bifurcation diagram for the oscillator with $k = 1$, $k' = 0.5$, $\gamma = 0.5$, and $\omega = 0.8$, in the range $\alpha = (0.7, 0.82)$. The dashed vertical lines indicate the values of α used in Figure 4.

fixed point. Though it cannot be seen from these figures, there may still be a fixed point of f in this system, but it has become unstable at this value of α. For the chaotic orbit on the right, the sequence x_n will contain an infinite number of different points (assuming we can wait long enough to collect them). Moreover, there are infinitely many values of m for which f^m has a fixed point. These fixed points cannot be literally on the attractor—if the system were ever to hit one of these points exactly, it would stay on the associated periodic orbit—but they are infinitesimally close to it. These periodic orbits are said to be **embedded** in the strange attractor, and their presence has been exploited both for control purposes and for the derivation of mathematical properties of the strange attractors. (See (24) for details.)

A useful way to exhibit the types of bifurcations that occur in a given system is to form a **bifurcation diagram** from the return map. Sets of values of x_n are collected for many different values of the bifurcation parameter and plotted on a single figure, as shown in Figure 5. Each vertical slice of the figure shows all of the x_n's observed for the corresponding value of α. The sequence of period doubling bifurcations is visible, and is a common structure in systems with only a few variables. Two other features common to experiments are visible in the figure. First, at the critical point for the first period doubling bifurcation (near $\alpha = 0.76$), the data are slightly smeared out. This is because near the transition the 1:1 limit cycle is just barely stable, which in turn implies that the transient relaxation to the limit cycle is very long. The plot was made by integrating the equations of motion up through about 50 cycles and recording data from the last 40 cycles. In the present case, the smearing could easily be reduced by waiting

longer before recording the data; the plot is presented as it is to illustrate the general point that slow relaxation near the critical point makes it more difficult to get clean data there. Second, the chaotic region above $\alpha \approx 0.81$ appears rather sparsely filled. Again, this is partly because the runs from which data were gathered only covered 40 cycles of the drive. For longer runs, the data for a given α would form rather dense bands with some visible gaps. Experimental data similar to Figure 5, however, would constitute clear evidence of chaotic behavior.

The presence of a period-doubling route to chaos in a wide variety of systems, together with the recognition that simple bifurcations can be classified into generic types, is very encouraging. It means that many of the features of nonlinear dynamical systems are **universal**, that is, they are independent of the quantitative details of a model. There are now several other routes to chaos that have been characterized, including quasiperiodic attractors that finally give way to chaos, and intermittent behavior in which long periods of nearly regular behavior are interrupted by chaotic bursts (17,23). This type of universality allows educated guesses about how to construct models that exhibit the features observed in experiments.

A basic vocabulary of bifurcations and transitions to chaos is now well developed, and one's first inclination upon observing chaos in an experiment should be to classify its onset as a particular known type. The known classification scheme is not exhaustive, however, and there continue to be cases in which theoretical understanding requires exploring the mathematics of new types of transitions. This is particularly true in systems with very many variables or systems described by partial differential equations that lead to complex patterns.

5. TWO TYPES OF COMPLEXITY: SPATIAL STRUCTURE AND NETWORK STRUCTURE

Thus far the discussion has been limited to systems with only a few degrees of freedom. The effects of nonlinearity become much more difficult to characterize or predict when many degrees of freedom interact. The complexity of the solutions can become overwhelming, in fact, and many fundamental mathematical questions about such systems remain open. Nevertheless, the language and techniques of nonlinear dynamics are helpful in formulating fruitful questions and reporting results.

There are two different ways in which a system can involve a large number of degrees of freedom, both of which are commonly encountered in biomedicine. First, a system can be spatially extended, consisting of a few variables that take on different values at different spatial points. Though such systems may be described by just a few PDEs, the solutions can involve spatial structures of exceedingly complex form. A steadily driven chemical reaction, for example, can display ever-changing patterns of activity as spiral waves are continually formed

and destroyed by propagating wavefronts. As one might expect, there is a whole zoo of observed patterns and bifurcations in such systems, obtained both from physical experiments and numerical simulations on systems as diverse as vertically vibrated layers of sand, layers of fluid heated from below, chemical reaction-diffusion systems, and optical systems involving broad laser beams in feedback loops containing nonlinear elements. Studies of such systems appear to be relevant for explaining pattern formation on butterfly wings, cardiac alternans and fibrillation, and the behavior of neuronal tissue, to name just a few examples.

Not all spatially extended systems show dynamics qualitatively different from simple systems of a few variables. Typically, there is a length scale associated with the spatial patterns one sees in a snapshot of the system. This could be, for example, the average width of stripes observed in a stationary or moving pattern or the size of a square in a checkerboard pattern. If the system is not too large compared to this characteristic length, the dynamics generally takes the forms discussed in section 3.

To analyze spatiotemporal dynamics, one often tries to define new variables that make the problem as simple as possible. These variables take the form of spatially varying functions of the natural field variables, and these function are called **modes** of the system. Choosing a useful set of modes can be difficult, though in some cases symmetry considerations make the task easier. For example, when the equations of motion are unchanged by uniform spatial shifts, it is often useful to use a Fourier decomposition, in which the modes is are simple sine waves of different wavelength. In other cases it may be natural to define modes associated with spatial structures whose amplitudes grow or shrink particularly rapidly or capture salient features of the observed patterns.

The partial differential equations of motion are then transformed into ordinary differential equations governing the amplitudes of the different modes. In the case of systems that are not too large compared to their characteristic length scale, one usually finds that all but a few of the mode amplitudes decay rapidly to zero. The long time dynamics of the system is then well represented by coupled ordinary differential equations for a few variables and the methods of section 3 can be applied even though the corresponding spatiotemporal behavior may look rather complicated.

If the system is *large* compared to the natural length scale of the spatial pattern, the situation becomes substantially more complex. Figure 6 shows a snapshot a the convection pattern in a fluid heated from below. PDEs used to model this system reproduce the observed behavior very well, and the phenomenon is now known as "spiral defect chaos" (7,19). The spiral structures in the pattern move around in erratic ways, and theoretical understanding of the motion is far from complete. It can be extremely difficult, for example, to answer one of the most fundamental questions about observed erratic behavior: Does it correspond to a strange attractor or just to an extremely long transient?

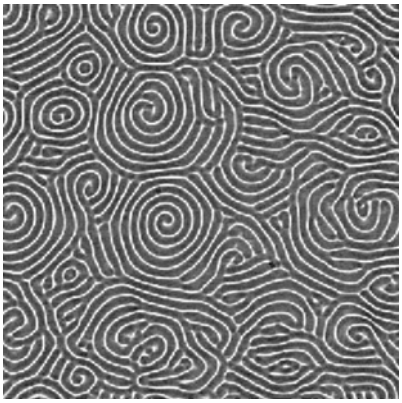


Figure 6. Spiral defect chaos: an example of spatiotemporal chaos in a system that is large compared to the characteristic length scale of the pattern (the width of the stripes). Image courtesy of G. Ahlers.

The theory of spatiotemporal pattern formation and chaos in large systems spans several loosely connected approaches. One is to study the dynamics of isolated typical structures. For example, one can study the speed with which a single pulse moves and spreads through an otherwise homogeneous medium. Another is to look carefully at critical points where universal bifurcation structures can be identified. An example of this is the onset of stripe structures in a homogeneous medium, such as the regions of sinking and upwelling in a fluid heated from below, or the development of chemical patterns in reaction-diffusion systems. Still another is to analyze the **short-time Lyapunov exponents** and associated modes of instability. These exponents do not quite have the same meaning as the true Lyapunov exponents, but are similar in spirit. The true exponents are defined as global properties of the full limit cycle or strange attractor. The short time exponents describe the local stability properties of a trajectory over a finite time interval and the modes associated with positive short-time exponents can reveal the locations in a pattern where instabilities will make prediction difficult over near term.

Still another theoretical approach to complex spatiotemporal behavior is to identify local structures that control the evolution of the pattern and try to describe their collective behavior in statistical terms. Often, the objects of interest are "defects" in an otherwise regular pattern. In spiral defect chaos, for example, it is known that there is another attractor consisting of uniform stripes, so it is tempting to think of the core of a spiral can be thought of as a defect in a stripe pattern. For topological reasons, the defect (a left-handed spiral, say) cannot be

removed except by drifting off the boundary of the system or annihilating with another defect of opposite topological sign (a right-handed, spiral). The dynamics of the system may be described as a changing pattern of defect locations. Since the behavior of dissipative nonlinear dynamical systems with many degrees of freedom resembles in several ways the behavior of systems treated by statistical mechanics, the language of statistical mechanics sometimes creeps into discussions of deterministic nonlinear dynamical systems. In particular, bifurcations are sometimes referred to as **phase transitions**.

The statistical approach is very tempting to physicists. The tools of statistical mechanics have been spectacularly successful in analyzing phase transitions in equilibrium systems, and they become more and more accurate as system size increases. They are based on the idea that the details of how a system moves in its very high-dimensional state space are unimportant for making statistical predictions about the states it is likely to be found in. Unfortunately, the fundamental assumptions of statistical mechanics are strongly violated in the driven, dissipative systems of biomedical interest. At present, there is increasing evidence that statistical mechanics can account for the behavior of a number of deterministic systems far from equilibrium. Examples include coupled discrete maps that undergo bifurcation that are quantitatively similar to equilibrium phase transitions (18), and sheared granular materials like sand in which effective temperatures can be defined for describing the wanderings of individual grains in space and time (16). These applications of statistical mechanics are as yet poorly understood, however, and in the absence of a fundamental theory of nonequilibrium pattern formation, theoretical insight comes largely from numerical simulation of model equations and analysis of the generic types of behavior on a case by case basis.

Construction of an appropriate model for a spatially extended dissipative system involves a healthy dose of intuition as well as a few constructive principles. One generally begins with the selection of the simplest PDE that incorporates the symmetries and general features of the physical system. This may be a known set of equations that is selected either because its solutions seem to match the observed behavior qualitatively or because the underlying physics of the system is expected to be in the same universality class. One then simulates the system numerically and attempts to find a regime in parameter space where the spatiotemporal dynamics is roughly reminiscent of the real system. Analysis of the model can then lead to hypotheses about the effects of varying parameters in the real system. To obtain more accurate predictions, one then adds terms to the model that alter the detailed behavior without changing the big picture. Given the complexity of types of bifurcations that can occur in large systems, however, one often discovers new and unexpected attractors, or one finds that detailed models just don't work and aspects of the physics that were thought to be irrelevant actually must be included in order to obtain reasonable representations of the true dynamical attractors. A good example of progress of this sort is

the modeling of cardiac electrodynamics, which is being approached from several directions simultaneously. The construction of PDE models based on detailed understanding of the physiology of cardiac tissue (26), the construction and simulation of PDE models reproducing realistic patterns of spatiotemporal activity (10,34), and the construction of ODE models or discrete maps reproducing bifurcations observed in small pieces of cardiac tissue (31), are all being brought to bear in a grand attempt to explain the onset and characteristics of arrhythmias from simple period doubling to fibrillation.

The second way in which complexity can enter the world of nonlinear dynamics is through the sheer abundance of distinct variables and the logical, or causal, relations among them. Even without taking into account the spatial distribution of concentrations of molecules, for example, the immune system or the metabolic network in a cell can exhibit surprising behavior due to nonlinear interactions among the concentrations of distinct types of molecules. The nature of these connections, and in particular the topology of the network representing them, has become a central theme in current research. As yet, rather little is understood about the dynamical processes that occur within such large networks containing many feedback loops. Our purpose here is just to illustrate the way in which nonlinear dynamics becomes the natural language for discussing the behavior of complex systems of this sort.

Consider the problem of modeling the regulatory network that governs gene expression in a cell. At its most basic level, the cell can be thought of as a dynamical system of interacting biomolecules produced through the mechanisms of gene expression. In this picture, the future chemistry of a cell is determined by which genes are expressed at any given moment. The products of transcription and translation of genes interact in extraordinarily complicated ways and act back on the processes of transcription and translation so as to influence which genes are expressed at a later time. To model this system the nonlinear dynamicist might begin by defining the system variables to be the levels of expression of each gene. Thus the life history of a cell becomes a trajectory through a state space of dimension equal to the number of genes in the network.

Modeling of the detailed interactions among all of the proteins and nucleic acids in the cell would make for a horrifically complicated mathematical system, from which it would be very difficult to glean any useful insights. Instead, one can hope (and perhaps expect) that many features of the state space trajectories are universal, i.e., that they do not depend on the details of the interactions. One is then led to devise models that retain the general logical structure of genetic regulatory networks but are defined by interactions simple enough to be efficiently simulated and studied analytically. One approach, pioneered in the context of genetic regulatory networks by Kauffman (14,15) is to assume that gene expression level is a Boolean variable and that the logical relations among different genes' activities are essentially random. As it turns out, the behavior of

such networks is surprisingly complex and suggestive, and is a subject of active research.

Boolean functions describing the switching of system variables between binary values can be thought of as an approximation of a map corresponding to a very complicated set of differential equations for the underlying physical processes. These functions, which must produce a binary output given some set of binary inputs, are strongly nonlinear. Since the system variables can only have two distinct values, the notion of a state corresponding to the sum of two other distinct states is not even well defined. Even in this extreme situation, however, the concepts of nonlinear dynamics provide a useful framework for discussing network behavior. Here we present a bare-bones description of this framework as an illustration of how the concepts discussed above enter the discussion.

A Boolean network is a collection of N logic gates, each having some fixed number of inputs, K_i, and one binary output, σ_i, where $i = 1, \ldots N$ indexes the gates. The inputs to a gate are a subset of size K_i of the outputs from all of the gates. Each gate is also characterized by a truth table T_i that determines σ_i as a function of the inputs. On each (discrete) time step, all of the gates apply their truth tables to their inputs and update their outputs accordingly. Each T_i is assumed to be selected randomly from a weighted distribution of all the possible truth tables with K_i inputs. To complete the definition of the model, one must specify the K_i's and the procedure for choosing which σ_i's act as inputs to a given gate. The best studied cases are networks in which all K_i are the same and the choice of which gates are inputs to any given gate is completely random. The result is a "random Boolean network" (RBN), sometimes referred to as a "Kauffman net."

The system variables in an RBN are simply the values of the outputs of the gates. The parameters of a particular model network are the choices of which outputs serve as inputs to each gate and which Boolean function is assigned to each gate. Instead of specifying all of these parameters explicitly, however, we specify a random procedure for choosing them. The number of inputs to each gate and the probabilities assigned to each of the different truth tables are taken as the parameters of the model. Note that when we discuss the behavior of the model at a certain set of parameter values, we are now talking about the average or typical behavior of a whole class of individual RBNs—those constructed according to a specified probabilistic procedure—rather than the detailed behavior of one specific dynamical system. (For more on probabilistic procedures for constructing the wiring diagrams of biological networks, see this volume, Part II, chapter 4, by Wuchty, Ravasz, and Barabási.)

In an RBN, the trajectory associated with the differential equations becomes a sequence of vertices in a state space that is a discrete set of points. If there are N gates in the network, each point in state space is an N-dimensional vector. Now because the number of distinct states is finite, the total number of possible states being 2^N, the sequence must eventually arrive at a point that has been vis-

ited before. From then on, it must cycle on the same loop forever. This means that, strictly speaking, all attractors on any Boolean network are periodic limit cycles. When N is very large, however, 2^N is astronomically huge and these cycles can become extremely long.

A surprising aspect of RBNs is the existence of two qualitatively different behaviors for different parameter regimes. A parameter q_1 can be defined that corresponds to the probability that changing the value of one randomly selected input to a randomly selected gate will result in a change in the output of that gate (27). A qualitative change in the network behavior is observed as q_1 is varied, which can be accomplished by changing K or changing the weights of the different truth tables. For small values of q_1, typical networks have only a few attractors; almost all of the gates wind up stuck on one value or the other and the duration of the attractor cycles are short. For larger values of q_1, a number of gates of order N remain active and the cycles are extremely long. The attractors in the two regimes also have markedly different stability properties. In the case of small q_1, small externally imposed perturbations, like changing the output value of a single gate for one time step, have little effect. The system quickly returns to the original attractor. For large q_1, on the other hand, small perturbations often place the system in the basin of a different attractor. The regime in which one observes short, stable cycles is called "ordered," and the region with exponentially long, attractors that are sensitive to small perturbations is called "chaotic." The latter term is meant to emphasize the erratic nature of the attractors over many times steps, but is not a rigorous description of the attractors over the tremendously long times associated with their cycle durations.

RBNs at the critical value of q_1 exhibit a unique balance of attractor stability and flexibility (15). The discovery of these special and totally unanticipated properties of critical RBNs is an indication of the power of the nonlinear dynamics conceptual framework. Even though these specific RBN models are not faithful representations of real biological processes, they reveal nonlinear dynamical structures that are likely to arise also in models that incorporate more realistic details, and therefore suggest new ways of understanding of the integrated behavior of the genome.

6. DISCUSSION AND CONCLUSIONS

This chapter is intended only to establish some of the vocabulary of nonlinear dynamics and give some indication of the rich behaviors that fall within its domain. Many important phenomena have been neglected entirely to this point. Three stand out as requiring some comment, however brief: the effects of stochastic processes; the role of boundary conditions; and the phenomena of frequency locking and synchronization.

Dynamical systems are, by definition, deterministic. They are therefore capable of exhibiting exquisitely detailed mathematical structures, and one might well ask whether these structures survive in the presence of stochastic influences, or **noise**. Conversely, dynamical systems with rather mundane behavior could respond in unexpected ways in the presence of noise. The theory of noisy dynamical systems, which in many cases is studied under the heading of nonequilibrium statistical mechanics, is a rich topic in its own right and is likely to be highly relevant for understanding some biomedical processes. It is also true, however, that the effects of noise can often be safely neglected, either because the details washed out by the noise are on such a fine scale as to be uninteresting, or because the feedback elements in the system allow it to operate reliably even when noise is a strong influence.

In spatially extended dynamical systems, boundary conditions can play a crucial role in determining the nature of the solutions to equations describing a bulk material. The same PDE can exhibit very different solutions when the boundary conditions are changed, and the realistic modeling of a system may depend just as much on getting the boundary conditions right as it does on modeling the bulk process. This often means having to understand the physics of a material or interface that was originally thought to be external to the system. Many analytical and numerical studies of PDEs are performed on domains that are artificially modeled as having no boundaries, like a torus. This is often quite useful, but care must be taken in applying intuition from these studies to the interpretation of experiments.

Phase locking is a phenomenon that occurs when two autonomous systems that oscillate at different natural frequencies are weakly coupled. While for extremely weak coupling there exist **quasiperiodic** trajectories of the coupled system that never exactly repeat but do not have the positive Lyapunov exponents associated with chaos, slightly stronger coupling tends to cause the two original systems to lock into a periodic trajectory in which the ratio of the periods of oscillation of the two original systems is a rational number. The most famous case of this is the phase locking of the moon's rotation about its axis to its orbit around the earth, which is why we on earth always see the same side of the moon. When elements are added to a system to induce phase locking, or when a large number of systems become phase locked in a 1:1 pattern, the phenomenon is sometimes called **synchronization**. In studying natural systems where synchronization is observed, it may be helpful to keep in mind the fact that it could be a straightforward consequence of nonlinear dynamics principles (29).

Finally, in an age in which the control and manipulation of biological systems is attracting so much interest and speculation, it is worth noting that there is a vast and growing literature on the **control** of dynamical systems. In this context, control means applying signals, hopefully of low power, in order to get a system to follow a desired trajectory in state space. (Two useful textbooks for basic elements of control theory are (21,22).) This may mean steering the system

from one attractor to another, keeping the system on a trajectory that is unstable in the absence of control, or combinations of the two. One example of a biomedical problem that naturally involves control theory, but also pushes its current limits, is the prevention of cardiac arrhythmias in humans (see Part III, chapter 3.3, by Glass.) Here one has a spatially extended system large enough to support complex spatiotemporal activity, though the desired behavior is a simple, regular heartbeat. There is some reason to hope that nonlinear dynamics models will provide useful descriptions of cardiac electrodynamics and new ideas for suppressing instabilities associated with certain types of arrhythmias. Recent work has focused on the onset of alternans (period doubling) in paced cardiac tissue (31) and the manipulation or destruction of spiral waves in excitable media models (1).

The world of nonlinear dynamical systems is full of complex structures and surprising behavior. There is now a well-developed language for characterizing all sorts of attractors and bifurcations as parameters are varied. The classification schemes will (probably) never be complete, however, and studies of systems as complex as living tissues and biological networks (metabolic, genetic, immunological, neuronal, ecological) are highly likely to uncover new mathematical structures. Systems with strongly stochastic elements or many interacting variables will require further connections to be made between nonlinear dynamics proper and statistical mechanics. As indicated by many of the chapters in this volume, all of these concepts can and should be brought to bear in the study of biomedical systems.

6. NOTES

1. If the equations of motion contain a second derivative of x_1, say, the above form is recovered by defining $x_2 = \dot{x}_1$ and writing $\dot{x}_2$ wherever the second derivative of x_1 appeared in the equations.

2. The coefficients are the external parameters designated by **p** above.

3. It is possible, however, for there to be an initial increase in some variables before the ultimate relaxation toward the origin occurs. This can happen when the eigenvectors associated with significantly different modes are **nonnormal** (not perpendicular to each other in state space). See (32) for a discussion of this effect and presentation of several examples.

4. Strictly speaking, this is a bit of a misnomer, as an exponent should not be a dimensionful quantity. The physically relevant quantity is the Lyapunov exponent λ multiplied by some characteristic time in the system.

5. Note that at points where the trajectory appears to cross itself it must be really separated in the third dimension since the future behavior is unique once an initial point in state space is given.

6. The situation is simple when driving is explicitly exhibited and is strictly periodic. For other systems, the strobe may have to be triggered in a slightly more subtle way. It has to flash when a system variable passes through a particular value, rather than at precisely equal time intervals, but the basic idea is the same.

7. REFERENCES

1. Aranson I, Levine H, Tsimring L. 1994. Controlling spatiotemporal chaos. *Phys Rev Lett* 72:2561.
2. Bak P. 1996. *How nature works: the science of self-organized criticality*. Copernicus, New York.
3. Baker GL, Gollub JP. 1990. *Chaotic dynamics: an introduction*. Cambridge UP, Cambridge.
4. Bohr T, Jensen MH, Paladin G, Vulpiani A. 1998. *Dynamical systems approach to turbulence*. Cambridge UP, Cambridge.
5. Cabo C, Rosenbaum DS. 2003. *Quantitative cardiac electrophysiology*. Marcel Dekker, New York.
6. Cross MC, Hohenberg PC. 1993. Pattern formation outside of equilibrium. *Rev Mod Phys* 65:851–1112.
7. Egolf DA, Melnikov IV, Bodenschatz E. 1998. Importance of local pattern properties in spiral defect chaos. *Phys Rev Lett* 80:3228–3231.
8. Epstein IR, Pojman JA. 1998. *An introduction to chemical dynamics: oscillations, waves, patterns, and chaos*. Oxford UP, New York.
9. Feigenbaum MJ. 1978. Quantitative universality for a class of nonlinear transofrmations. *J Stat Phys* 19:25.
10. Fenton F, Karma A. 1998. Fiber-rotation-induced vortex turbulence in thick myocardium. *Phys Rev Lett* 81:481.
11. Glass L. 1996. Dynamics of cardiac arrhythmias. *Phys Today* 49:40.
12. Golubitsky M, Schaeffer DG. 1988. *Singularities and groups in bifurcation theory*. Springer Verlag, New York.
13. Hall GM, Bahar S, Gauthier DJ. 1999. The prevalence of rate-dependent dynamics in cardiac tissue. *Phys Rev Lett* 82:2995–2999.
14. Kauffman SA. 1974. The large-scale structure and dynamics of gene control circuits. *J Theor Biol* 44:167.
15. Kauffman SA. 1993. *Origins of order*. Oxford UP, New York.
16. Makse H. 2002. Thermodynamics and effective temperatures in sheared granular matter and emulsions. *Eur Phys J E* 9:265–269.
17. Manneville P. 1990. *Dissipative structures and weak turbulence*. Academic Press, New York.
18. Miller J, Huse DA. 1993. Macroscopic equilibrium from microscopic irreversibility in a chaotic coupled-map lattice. *Phys Rev E* 48:2528–2535.
19. Morris SW, Bodenschatz E, Cannell DS, Ahlers G. 1993. Spiral defect chaos in large aspect ratio Rayleigh-Benard convection. *Phys Rev Lett* 71:2026–2029.
20. Murray JD. 2002. *Mathematical biology*, Vol. I, 3rd ed. Springer, Berlin.
21. Ogata K. 1987. *Discrete-time control systems*. Prentice-Hall, Englewood Cliffs, NJ.
22. Ogata K. 2001. *Modern control engineering*. Prentice-Hall, Englewood Cliffs, NJ.
23. Ostlund S, Rand D, Sethna J, Siggia E. 1983. Universal properties of the transition from quasiperdiocity to chaos in dissipative systems. *Physica D* 8:303.
24. Ott E. 2002. *Chaos in dynamical systems*, 2nd ed. Cambridge UP, Cambridge.
25. Ott E, Grebogi C, Yorke JA. 1990. Controlling chaos. *Phys. Rev. Lett*. 64:1196.

26. Roth BJ, Krassowska W. 1998. The induction of reentry in cardiac tissue, the missing link: how electric fields alter transmembrane potential. *Chaos* 8:204–220.
27. Socolar JES, Kauffman SA. 2003. Scaling in ordered and critical random Boolean networks. *Phys Rev Lett* 90:068702.
28. Strogatz SH. 1994. *Nonlinear dynamics and chaos*. Perseus Books, Reading, MA.
29. Strogatz SH. 2000. From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D* 143:1–20.
30. Takens F. 1980. Detecting strange attractors in turbulence. In *Dynamical systems and turbulence*, pp. 366–381. Ed. DA Rand, LS Young. Springer lecture notes in mathematics, Vol. 898. Springer, New York.
31. Tolkacheva EG, Romeo MM, Guerraty M, Gauthier DJ. 2004. Condition for alternans and its control in a two-dimensional mapping model of paced cardiac dynamics. *Phys Rev E* 69:031904.
32. Trefethen LN. 1997. Pseudospectra of linear operators. *SIAM Rev* 39:383–406.
33. Walgraef D. 1997. *Spatio-temporal pattern formation: with examples from physics, chemistry, and materials science*. Springer, New York.
34. Winfree AT. 1997. Heart muscle as a reaction-diffusion medium: the roles of electric potential diffusion, activation front curvature, and anisotropy. *Int J Bifurcations Chaos* 7:487–526.

3

BIOLOGICAL SCALING AND PHYSIOLOGICAL TIME: BIOMEDICAL APPLICATIONS

Van M. Savage and Geoffrey B. West

Bauer Center for Genomics Research, Harvard University, Cambridge, Massachusetts; Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico

In collaboration with A.P. Allen, J.H. Brown, B.J. Enquist, J.F. Gillooly, A.B. Herman, and W.H. Woodruff

A framework for the development of quantitative theories that capture the body size and body temperature dependence of many cellular and physiological rates and times is presented. These theories rely on basic properties of biological systems, such as the invariance of terminal units, and on fundamental constraints taken from physics and chemistry, such as energy minimization of flow through resource-distribution networks and statistics of biochemical reaction kinetics. The primary postulate of this framework is that metabolic rate—the rate at which organisms take in resources from the environment, distribute these resources throughout their bodies, and process these resources by means of biochemical reactions—is perhaps the most fundamental rate in all of biology and is a major determinant, through both direct and indirect effects, of most cellular and physiological rates. The pervasive effects of metabolic rate are due to the facts that cellular rates work in concert to produce the rates manifested at the whole-organism level, and that the power created by metabolism must be allocated to individual maintenance, ontogenetic growth, and reproduction. Here we outline the derivations of the body size and body temperature dependence of metabolic rate. Using the primacy of metabolic rate, we then describe the ongoing development of theories that connect the theory of biological scaling to several biomedical processes, including ontogenetic growth, nucleotide substitution rates, sleep, and cancer growth. Empirical data are presented that confirm the mass and temperature dependence of metabolic rate as well as predictions for lifespan, ontogenetic growth trajectories, and sleep cycle times. Insights gleaned from these theories could potentially lead to important biomedical applications, such as methods for calculating proper drug dosing or for frustrating processes related to tumor angiogenesis.

Address correspondence to: Van M. Savage, 7 Divinity Avenue, Bauer Center for Genomics Research, Harvard University, Cambridge, MA 02138.

1. INTRODUCTION

An Anacreontick

Busy, curious, thirsty fly,
Gently drink, and drink as I;
Freely welcome to my cup,
Could'st thou sip, and sip it up;
Make the most of Life you may,
Life is short and wears away.

Just alike, both mine and thine,
Hasten quick to their decline;
Thine's a Summer, mine's no more,
Though repeated to threescore;
Threescore Summers when they're gone,
Will appear as short as one.

—William Oldys

When Oldys wrote this verse in the eighteenth century, poetically expressing the similarity of life histories of different organisms and their corresponding rates of life processes, he surely did not appreciate that the difference in perspective between a fly and a human is not merely psychological, but has, in fact, a physiological basis. For example, it is now known that the total number of lifetime heartbeats is approximately the same for all mammals, even though larger mammals generally live longer than smaller ones. Indeed, lifespan varies in a systematic way with the body size and body temperature of an organism (1), and it is believed that this is a consequence of the rate at which organisms live and process energy. Almost all physiological variables and all biological rates and times, including heart and respiratory rates (2,3), gestation periods (2,4), developmental times (5), sleep times (6,7), and even cancer growth rates (8–11), scale in a systematic and interrelated way with body size and temperature. In addition, many structural properties, including the radius of the aorta (12), the density of mitochondria (13), and genome length scale in a similar fashion (14,15). All of these phenomena scale with mass as a power law with exponents that are invariably simple multiples of one quarter. In the past few years we and our colleagues have attempted to discover the mathematical form of these scaling relationships, and perhaps most importantly, to develop mechanistic, dynamical theories for their origins, based on an underlying set of general principles (1,12,16). Through the development of these theories and the gathering of empirical data, we have discovered new scaling relationships for the rates and times of many cellular (13), physiological (5), and ecological processes (17–20).

Our starting point is the assumption that at every organizational level there exist average idealized organisms, or biological systems, whose properties are

calculable. These idealizations provide a baseline for quantitatively analyzing, understanding, and assessing real biological systems, which are viewed as perturbations or variations from the idealized norms. This idea is motivated by the existence of the many scaling laws that reveal the remarkable similarity among organisms of vastly different sizes, temperatures, and structures despite the extraordinary complexity and diversity of life. Living organisms span a mass range of over 21 orders of magnitude from the smallest mycoplasma (10^{-13} g) to the largest mammals and plants (10^{8} g). Overall, the life process covers almost 30 orders of magnitude from gene structure and the terminal oxidase molecules of the respiratory complex up to ecosystems. This vast range exceeds that of the earth's mass relative to the galaxy's, which is *only* 18 orders of magnitude, and is comparable to an electron relative to a cat. By focusing on metabolic rate, which we assume places fundamental constraints on all organisms, we gain a foothold for studying this immense diversity. For our idealized organisms, fundamental properties of resource-distribution networks are the paramount evolutionary traits, and the aforementioned perturbations and variations from some average idealized norm in real biological systems are presumed to be due to local environmental niches or external conditions that select for many other, often unrelated traits. Comparing organisms over large ranges in body size and temperature effectively averages over environments and diminishes the importance of evolutionary innovation in response to specific environmental conditions. Consequently, a coarse-grained quantitative, predictive description becomes conceptually feasible, so that a generalized theory can apply over many orders of magnitude.

Allometric scaling relates biological parameters to body mass, M. The best-known of these is for basal metabolic rate, which was first shown by Kleiber (21) and Brody (22) to scale as $M^{3/4}$ for mammals and birds. (For a recent and extensive compilation and analyses of metabolic rate data for mammals, see Savage et al. (3).) This observation was extended by Hemmingsen (23) to ectotherms and unicellular organisms and later by other researchers to many other taxa, including plants (12,24,25). More recently, it was extended to the respiratory complex within mitochondria down to the terminal oxidase molecules (the universal respiratory machinery responsible for the production of ATP, the basic currency of aerobic metabolism), thereby covering an astonishing 27 orders of magnitude (Figure 1) (13). A synthesis of the enormous amount of data encoded in allometric scaling was summarized in the early 1980s in four books that convincingly showed the predominance of quarter-power scaling across all scales and almost all forms of life (2,26–28).

After body size, the biggest determinant of biological rates and times is body temperature (29,30). Basal metabolic rate for hibernating mammals, birds in torpor, amphibians, reptiles, plants, and unicellular organisms have all been shown to scale as a Boltzmann factor, $e^{-E/kT}$, where E is the activation energy for biochemical reactions, k is Boltzmann's constant, and T is absolute temperature

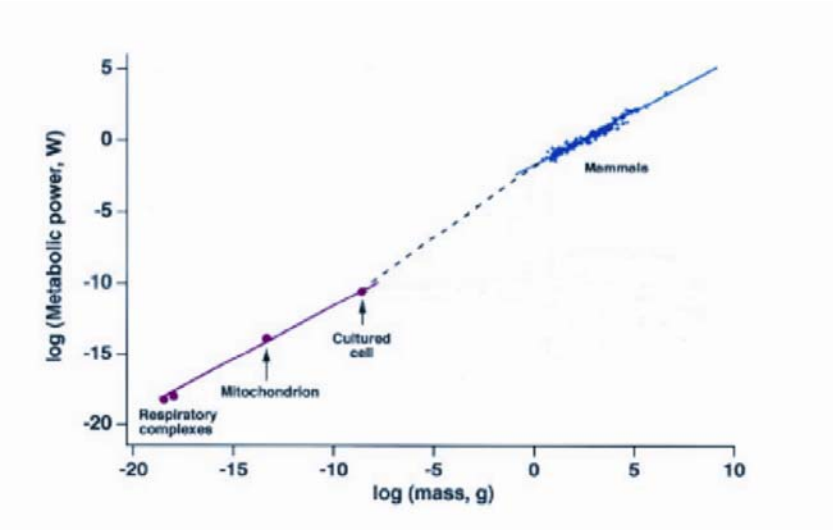


Figure 1. A logarithmic plot of metabolic rate as a function of mass. The entire range is shown, covering 27 orders of magnitude, from a cytochrome oxidase molecule and respiratory complex through a mitochondrion and single cell in vitro (red dots) up to whole mammals (blue dots). The solid red and blue lines through the corresponding dots represent $M^{3/4}$ fits. The dashed line is the predicted linear extrapolation from the mass for the smallest mammals to an isolated mammalian cell. This figure is reproduced with permission from West et al. (2002) (13).

(Figure 2) (1). Typically, E has a value in the range of 0.6–0.7 eV, reflecting a common biochemistry underlying most of life (1,18). Ontogenetic growth rates, heart rates, and even rates of conflict between beetles also scale with a similar Boltzmann factor (18). Lifespan (1), time to first reproduction (5), and the intrinsic rate of increase for a population (17,18) all scale as an inverse Boltzmann factor.

What is remarkable is that body size (as expressed in quarter-power allometric scaling) and temperature (as expressed by the Boltzmann factor) explain the dominant variation among biological rates: for example, correcting metabolic rate for mass and temperature reduces the variation from fifteen orders of magnitude variation to approximately one (1). Therefore, these two variables, along with just two numbers, E and 1/4, provide a surprisingly robust baseline for biological phenomena.

An intriguing consequence of these laws is the emergence of approximately invariant quantities, something physicists recognize as signatures of fundamental underlying constraints. For example, lifespan increases as $M^{1/4}e^{E/kT}$ (Figure 3),

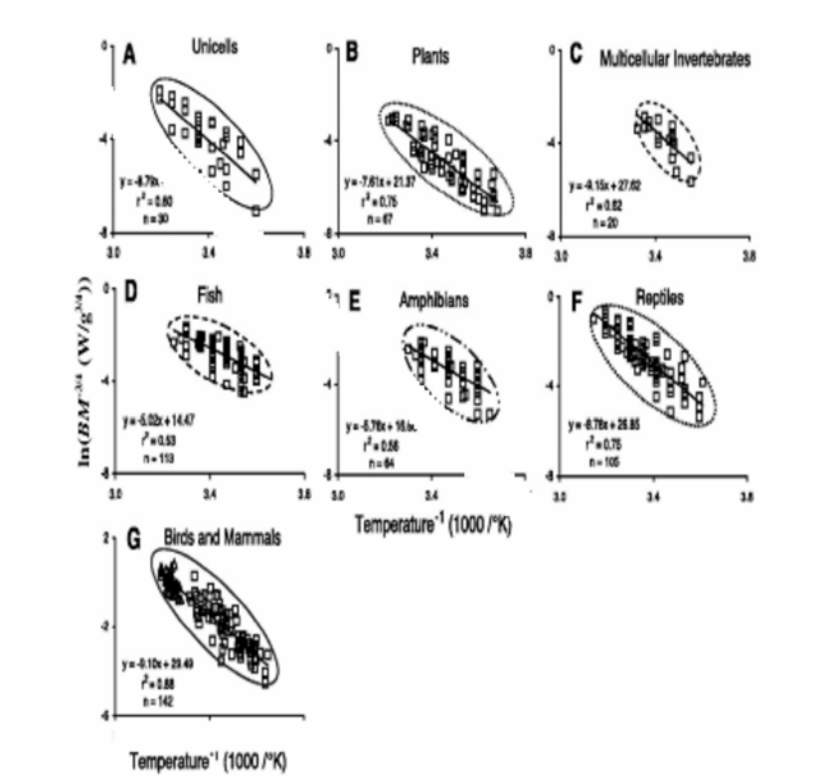


Figure 2. Effects of temperature (1000/°K) on mass-corrected resting metabolic rate, $\ln(BM^{-3/4}$ $(W/g^{3/4}))$, for unicells (**A**), plants (**B**), multicellular invertebrates (**C**), fish (**D**), amphibians (**E**), reptiles (**F**), and birds and mammals (**G**). Birds (black symbols) and mammals (open symbols) are shown at normal body temperature (triangles) and during hibernation or torpor (squares). This figure is reproduced with permission from Gillooly et al. (2000) (1).

whereas heart rate decreases as $M^{-1/4}e^{-E/kT}$, so the number of heartbeats per lifetime is approximately invariant, independent of size and temperature (lifetime heartbeats ~1.5 x 10^9 for mammals). Hearts are not fundamental but the molecular machinery of aerobic metabolism is, and it also has an analogous invariant, namely, the number of lifetime turnovers of the respiratory complex (~10^{16}) (13).

Until recently most theoretical investigations focused on the metabolic rate of a specific taxonomic class. The broader challenge is to understand the overwhelming ubiquity and universality of quarter-power scaling with mass and the exponential variation with temperature, and to connect the existence of these scaling relationships to unifying principles that determine how life is structured and the constraints under which it has evolved. In the second section of this

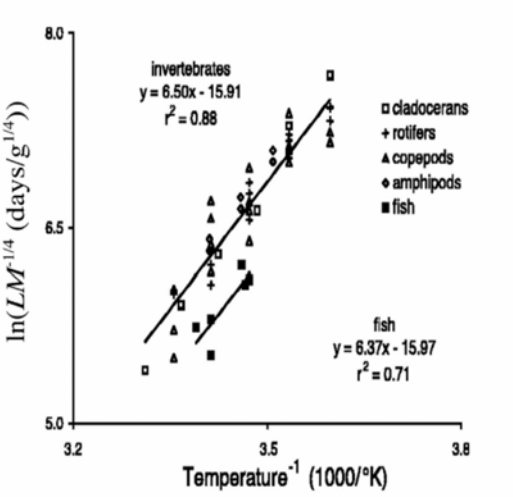


Figure 3. Effects of temperature, T^{-1} (1000/°K), on mass-corrected lifespan, $\ln(LM^{-1/4}$ (days/$g^{1/4}$)). Data are for fish and aquatic invertebrates held at different constant temperatures in the laboratory. This figure is reproduced with permission from Gillooly et al. (2000) (1).

chapter, we provide a general theory for the origin of these scaling relationships based on previously published work (1,12). In the third section we focus on the development of subsequent theories that apply biological scaling to biomedical problems. In the final section we summarize the findings presented in this chapter and speculate about future work in these areas.

2. MODEL DESCRIPTION: THEORY FOR THE ORIGIN OF SCALING RELATIONSHIPS

Metabolic rate sets the pace of life. By modeling the dynamics of the cardiovascular system based on some general assumptions, which are independent of the detailed dynamics and design, the body size dependence for metabolic rate, $B \propto M^{3/4}$, can be derived. The temperature dependence is given by a Boltzmann factor, $B \propto e^{-E/kT}$, and is a direct consequence of the kinetics of the underlying biochemical reactions responsible for the production of ATP. As mentioned above, these two variables alone explain a surprising amount of the variation in metabolic rate and set a baseline for analyzing biological organisms. It is likely that the residual variation points to interesting biological differences between organisms and gives clues as to what these differences are.

2.1. Body Size Dependence

> So, naturalists observe, a flea has smaller fleas that on him prey; and these have smaller still to bite 'em, And so proceed ad infinitum. Thus, every poet in his kind is bit by him that comes behind.
>
> —*Jonathan Swift*

We posit that the effect of body size on all physiological variables is determined by the scaling properties of hierarchical resource-distribution networks, such as the cardiovascular and respiratory systems, which deliver essential nutrients and metabolites to cells. There are three main assumptions, all presumed to be derivative from the processes of evolution and natural selection, that define the theory and are postulated to characterize the resource-distribution networks (12,16): (i) they are space filling in order to service all cells through the supply of nutrients and the removal of wastes, (ii) the energy to deliver resources is minimized, and (iii) their terminal units (e.g., capillaries) are invariant. We now review how these three assumptions are used to derive the scaling of metabolic rate with body size.

In order to describe the network we need to determine how the radii, r_k, and lengths, l_k, of the tubes change throughout the network; k denotes the level of the branching, beginning with the aorta at $k = 0$ and terminating at the capillaries where $k = N$. The number of branches per node (the branching ratio), n, is assumed to be constant throughout the network. To characterize the branching we introduce scale factors via the ratios $\beta_k = r_{k+1}/r_k$ and $\gamma_k = l_{k+1}/l_k$. Since capillaries are invariant units, these scale factors completely determine the network except for the number of levels, which is a function of body size.

The first assumption (i), that networks are space-filling (31), ensures that all tissues are supplied by capillaries. The organism is composed of many groups of cells, referred to here as "service volumes," v_N, which are supplied by a single capillary. The total volume to be filled, or serviced, is given by $V = N_N v_N$, where N_N is the number of capillaries. For a network with many levels, N, complete space-filling implies that this same volume, V, is filled at all scales by an analogous volume, v_k, defined by branches at each level k. Since $r_k << l_k$, $v_k \propto l_k^3$, so space-filling constrains only branch lengths, l_k. Thus, $V \approx N_k v_k \propto N_k l_k^3$, and since V is independent of k, we have $\gamma \approx n^{-1/3}$. We assume this relation is valid throughout the network, although it becomes less realistic for small values of k.

A more explicit statement of assumption (ii) is that the continuous feedback implicit in evolutionary adaptation has lead to resource-distribution networks that, on average, minimize the energy required to support flow through the system. There are two independent contributions to energy loss: energy dissipated by viscous forces, which is only important in smaller vessels, and energy reflected at branch points, which is entirely eliminated by impedance matching. In large vessels (e.g., arteries) pulse waves suffer little attenuation or dissipation,

and as a result, the branching is area preserving, which leads to a constant blood velocity. In small vessels (capillaries, arterioles) the pulse is strongly damped because Poiseuille flow dominates and significant energy is dissipated. This leads to area-increasing branching, so blood flow slows down, almost ceasing in the capillaries.

A detailed treatment of pulsatile flow is complicated. Here, we present a condensed version that contains the features pertinent to the scaling problem. In contrast to non-pulsatile, Poiseuille flow, blood vessels cannot be considered rigid for pulsatile flow because vessels expand and contract as the pulse wave generated by the contraction of the heart propagates along them. The classic Poiseuille resistance of the rigid tube, relating the fluid volume flow rate to the driving pressure gradient, is thereby generalized to a complex impedance, signifying attenuated wave propagation (32–34).

The equation of motion governing fluid flow is the Navier-Stokes equation (35). Neglecting nonlinear terms responsible for turbulence, this is:

$$\rho \frac{\partial \mathbf{v}}{\partial t} = \mu \nabla^2 \mathbf{v} - \nabla \mathbf{p}. \qquad [1]$$

Here, the vector $\mathbf{v}$ is the local fluid velocity at some time t, $\mathbf{p}$ is the local pressure, and ρ is the fluid density. If the fluid is incompressible, then local conservation of fluid requires $\nabla \bullet \mathbf{v} = 0$. When combined with Eq. [1], this gives the subsidiary condition

$$\nabla^2 \mathbf{p} = 0. \qquad [2]$$

The analogous equation governing the elastic motion of the tube is the Navier equation. Neglecting nonlinear terms, this is given by:

$$\rho_w \frac{\partial^2 \zeta}{\partial t^2} = E \nabla^2 \xi - \nabla \mathbf{p}\,, \qquad [3]$$

where the vector ζ is the local displacement of the tube wall, ρ_w is its density, and E is its modulus of elasticity. These three coupled equations, [1], [2], and [3], must be solved subject to boundary conditions that require the continuity of velocity and force at the tube wall interfaces.

In the approximation where the vessel wall thickness, h, is small compared to the static equilibrium value of the vessel radius, r, i.e., $h << r$, the problem can be solved analytically, as first shown by Womersley (32), to give

$$\left(\frac{c}{c_0}\right)^2 \sim \frac{J_2(i^{3/2}\alpha)}{J_0(i^{3/2}\alpha)} \quad \text{and} \quad Z \sim \frac{c_0^2 \rho}{\pi r^2 c}\,, \qquad [4]$$

where ω is the angular frequency of the wave, $\alpha \equiv (\omega\rho/\mu)^{1/2}r$ is a dimensionless parameter known as the Womersley number, and $c_0 \equiv (Eh/2\rho r)^{1/2}$ is the classic Korteweg-Moens velocity (33,34). The wave velocity, c, and therefore, Z, are both complex functions of ω, so the wave is attenuated and dispersed as it propagates. The character of the wave depends critically on whether $|\alpha|$ is less than or greater than 1. This can be seen explicitly in Eq. [4], where the behavior of the Bessel functions changes from a power-series expansion for small $|\alpha|$ to an expansion with oscillatory behavior when $|\alpha|$ is large. In humans, α has a value of around 15 in the aorta, 5 in the arteries, 0.04 in the arterioles, and 0.005 in the capillaries. When α is large (>1), Eq. [4] gives $c \sim c_0$, which is a purely real quantity, so the wave suffers neither attenuation nor dispersion, demonstrating that viscosity plays almost no role in these large vessels. In this region (large vessels), Eq. [4] also gives $Z \sim \rho c_0/\pi r^2$, and the minimization of energy loss is attained through impedance matching, which eliminates the reflection of pulse waves at junctions, leading to area-preserving for the vessels, $\pi r_k^2 = \mathrm{n}\pi r_{k+1}^2$, so that $\beta = n^{-1/2}$.

For small vessels where $|\alpha| < 1$, the role of viscosity becomes increasingly important until it eventually dominates the flow. Eq. [4] gives $c \sim (1/4)i^{1/2}\alpha c_0 \rightarrow 0$, in quantitative agreement with observation (33,34). Because c now has a significant imaginary part, the traveling wave is heavily damped, leaving an almost steady oscillatory flow whose impedance is, from Eq. [4], given by the classic Poiseuille formula, $Z_k = 8\mu l_k/\pi r_k^4$. Unlike energy loss due to possible reflections at branch points, energy loss due to viscous dissipative forces cannot be entirely eliminated. However, it can be minimized using the classic method of Lagrange multipliers to enforce the appropriate constraints (12,36,37). To sustain a given metabolic rate in an organism of fixed mass with a given volume of blood, V_b, the cardiac output must be minimized subject to a space-filling geometry. This leads to $\beta = n^{-1/3}$. Thus, for small vessels area-preserving branching is replaced by area-increasing branching, and blood slows down, allowing efficient diffusion of oxygen from the capillaries to the cells. Branching, therefore, changes continuously through the network, so that β is *not* independent of k but changes continuously from $n^{-1/2}$ at the aorta to $n^{-1/3}$ at the capillaries. Consequently, the network is not strictly self-similar, but within the two different regions (pulsatile and Poiseuille) self-similarity is a reasonable approximation that is well supported by empirical data (33,34,38–41).

To derive allometric relations we need to connect the scaling of vessel size within an organism to its body mass, M. A natural vehicle for this is the total volume of blood in the network, V_b, which can be shown to depend linearly on M if cardiac output is minimized (12,37). V_b is given by

$$V_b = \sum_{k=0}^{N} N_k V_k = \sum_{k=0}^{N} n^k \pi r^2 l_k = \left[n^N V_N \frac{(n\gamma\beta_<^2)^{-N+\bar{k}-1} - 1}{(n\gamma\beta_<^2)^{-1} - 1} + \frac{(n\gamma\beta_<^2)^{-\bar{k}} - 1}{(n\gamma\beta_<^2)^{-1} - 1} \right], \qquad [5]$$

where we have divided the sum into regions according to the different values of β; $\beta_<$ is for the area-preserving branching that minimizes cardiac output for pulsatile flow in large vessels, and $\beta_>$ is for the area-increasing branching that minimizes the energy loss for Poiseuille flow in small vessels. Further, V_N is the volume of a single capillary, and $\bar{k}$ is the number of levels from the transition level between the two regions to the capillaries, which is the same for all mammals, i.e., it does not scale with organism size (12). Substituting the scaling relationships for γ, $\beta_<$, and $\beta_>$ into Eq. [5], we find that the first term in the square brackets scales as $n^{N/3}$ and that the second term is independent of N. Thus, for $N >> 1$ the first term dominates, and the leading-order behavior for the blood scales as $n^{4N/3}V_N$. By assumption (iii), capillaries are invariant units, so V_N is independent of M, which implies $V_b \propto n^{4N/3}$. Also by assumption (iii), the metabolic rate per capillary, B_N, is invariant, which implies $B = n^N B_N \propto n^N \propto V_b^{3/4}$. Since energy minimization requires $V_b \propto M$, this implies $B \propto M^{3/4}$. This matches the well-known empirical result.

Many other relations follow from this theory, including the determination of the scaling of the radii and lengths of all vessels and the blood flow and pulse rate in each of them. Furthermore, it shows that the 3/4 exponent is only an approximation, and that deviations can be expected for small mammals where the number of vessels that can support a pulse is small compared to that of a large mammal (13). Such deviations are indeed suggested by the data (3,42).

Metabolic energy is conserved as it flows through cells and mitochondria, which may possess hierarchical networks of cellular transport and chemical reactions respectively. Surprisingly little is known about intracellular transport networks, and an important challenge, both theoretical and experimental, is the construction of realistic models. The continuity of flow of metabolic energy through this series of sequential networks, a "hierarchy of hierarchies," imposes boundary conditions between each level that lead to constraints on densities of mitochondria and respiratory complexes (both considered invariant terminal units at their appropriate level) (13). This explains why there are typically a few hundred mitochondria per cell in vivo, but several thousand in vitro, and why there are several thousand respiratory complexes per mitochondrion (and not ten or ten million).

2.2. Body Temperature Dependence

Rates of chemical reactions, R, depend crucially upon temperature (29,30), as first demonstrated by Arrhenius (43) with his famous equation:

$$\frac{d \ln R}{dT} = -\frac{E}{kT^2} . \qquad [6]$$

Arrhenius divided molecules in a chemical system into two types: normal and activated. Activated molecules have an energy equal to E, the activation energy, and participate in chemical reactions, while normal molecules are dormant. He then postulated that the normal and activated molecules are in thermodynamic equilibrium (43,44). Although the essence of Arrhenius' argument is correct, it is now well established that Eq. [6] is actually probabilistic, derivable via statistical mechanical arguments. Modern kinetic theory holds that reactions can take place only if molecules collide with an energy greater than or equal to some minimum, E. Treating molecules as billiard balls, the kinetic energy of a collision increases linearly with the masses of the molecules and as the square of the velocity of the molecules relative to one another. Using Maxwell-Boltzmann theory, a velocity distribution function for the molecules in a system can be derived, and the probability of a collision occurring with an energy greater than or equal to E can be determined (45–47). The leading-order solution for the probability is proportional to the Boltzmann factor, $e^{-E/kT}$, which can be thought of as proportional to the fraction of molecules with energy, E, at a given temperature, T. Hence, the increase in the average energy per collision, and therefore, the increase in chemical reaction rates as a function of T, is proportional to the Boltzmann factor, which is precisely the solution to the Arrhenius equation, Eq. [6]: $R \propto e^{-E/kT}$. It is important to note that temperature also increases the frequency of collisions. However, the effect of this on reaction rates is a pre-factor that scales slowly as a power of T, and, therefore, is subdominant to the exponential behavior of the Boltzmann factor.

Biochemical reactions necessary for metabolism within organisms are similar to reactions in a chemistry laboratory, except enzymes catalyze many metabolic reactions and the medium for the reactions is the mitochondrial membrane immersed in water. For every species of organism, there is a minimum temperature below which metabolic rate ceases, an optimal temperature at which metabolic rate is maximum, and a very narrow temperature range above the optimum where metabolic rate rapidly decreases. We are primarily interested in the "biologically relevant" temperature range defined to be between the minimum and the optimum temperatures, where effects such as the freezing of water at low temperatures or the denaturing of proteins at high temperatures are negligible (1,17,18). If temperature is in a "biologically relevant" regime and the effects of temperature on enzyme functions are sub-exponential, the temperature dependence of metabolic reaction rates is the same as that of non-biological reactions, so that metabolic rate scales with a Boltzmann factor, $B \propto e^{-E/kT}$, where E now represents an activation energy for metabolic reactions (1,46). When combined with body size scaling, this implies that the characteristic power per gram, and thus, typical biological rates (1,5) scale as

$$R_{BIO} \propto M^{-1/4} E^{-E/kT}, \qquad [7]$$

whereas typical times (5) scale as

$$t_{BIO} \propto M^{1/4} e^{E/kT}. \qquad [8]$$

This exponential dependence on temperature is closely tied to Q_{10} factors, which are frequently used in biology. A Q_{10} factor is defined as the percentage change in a biological rate or time when there is a 10°C change in temperature. Since the temperature in degrees Celsius is given by $T_c = T - 273$, it follows that to leading order, $e^{-E/kT} \propto e^{\alpha T_c}$, where $\alpha = E/k(273)^2$. A value of $E \approx 0.6$ eV is typical for metabolic reactions, so $\alpha \approx 0.1°C^{-1}$. Thus, a typical value for Q_{10}, which is the ratio of rates, is $Q_{10} \approx e \approx 2.72$. Since Q_{10} factors are only approximations to the exact formula, they are probably best used as a guideline, and Boltzmann factors should be used for precise tests or calculations.

The activation energy, E, can be interpreted in a couple of ways, depending on whether the core reactions for metabolism occur in series or in parallel. If the reactions occur in series, the rate of supply of reactants to a reaction is set by the preceding reaction in the chain, and thus, the overall rate of the chain of reactions is set by its slowest, rate-limiting step. In this case, E is simply the activation energy of this rate-limiting reaction. However, if there are N core reactions that occur in parallel and are supplied with molecules from different pools, the overall rate of metabolism is given by the average rate, which to leading order is proportional to

$$\frac{1}{N}\sum_{i=0}^{N} e^{-E_i/kT}. \qquad [9]$$

In the special case where the activation energies for all of the core reactions are near some average value, $\bar{E}$ ($E_i = \bar{E}$ or $(E_i - \bar{E})/kT << 1$ for all i), the overall reaction rate for metabolism is approximately $e^{-\bar{E}/kT}$. Similarly, if there is a subset of reactions that have activation energies near some average value, $\bar{E}$, and this average is significantly smaller than activation energies for all other reactions, the rate will be given by $e^{-\bar{E}/kT}$. Real biochemical networks consist of a combination of reactions in series and reactions in parallel. Since many empirical results agree with the temperature dependence in Eqs. [7] and [8] (see the §1), it is likely that one of the scenarios outlined above, or some combination thereof, is correct for biological systems.

For a wide range of rates and times across a broad assortment of taxa, measured values of E are typically in the range of 0.6–0.7 eV, reflecting some shared biochemistry (1,5,17,18). Indeed, these values for E bracket the average value for biochemical reactions (48,49) and are very close to the activation energy for the oxidation of NaDH, which is common to most of life and may be a rate-limiting step!

3. BIOMEDICAL APPLICATIONS

Many medical studies are performed on animals other than humans, especially on mice and rats. Understanding how to interpret data from animal studies and how to compare them with similar conditions in humans requires the use of scaling relationships. In this section we briefly describe several areas in which progress has been made in applying the theory of biological scaling to biomedical problems.

3.1. Ontogenetic Growth and Developmental Times

The theory of biological scaling naturally leads to a general growth equation applicable to all animals (50,51),

$$B = N_c B_c + E_c \frac{dN_c}{dt} . \qquad [10]$$

Here, N_c is the total number of cells in the organism, B_c is the metabolic rate given to maintenance per cell, and E_c is the energy needed to create a fully grown, new cell. Metabolic energy is transported through the network to cells where it is used either for maintenance (the first term on the right side of Eq. [10]), including replacement of cells, or for growth of new cells (the second term on the right side of Eq. [10]). Substituting $N_c = m/m_c$ into Eq. [10], where m is the ontogenetic mass at time t after birth and m_c is the average cell mass, gives an equation for the rate of growth of an organism:

$$\frac{dm}{dt} = \left(\frac{B_0 m_c}{E_c}\right) m^{3/4} - \left(\frac{B_c}{E_c}\right) m , \qquad [11]$$

where B_0 is a taxon-dependent normalization constant for the scaling of metabolic rate. The parameters of the resulting growth equation are determined solely by fundamental cellular properties, such as their metabolic rate and the energy required to create them. The model gives a natural explanation for why animals stop growing: the number of cells supplied ($N_c \propto m$) scales faster than the number of supply units (since $B \propto N_N \propto m^{3/4}$) and leads to formulae for the asymptotic mass of the organism: $M = (B_0 m_c / B_c)^4$. Eq. [11] can be solved analytically to determine $m(t)$, leading to a classic sigmoidal curve. From the ensuing equations, a universal scaling curve for growth is derived that is well fit by data from many different organisms (Figure 4). Ontogenetic development is therefore a universal phenomenon determined by basic cellular properties. Furthermore, this theory

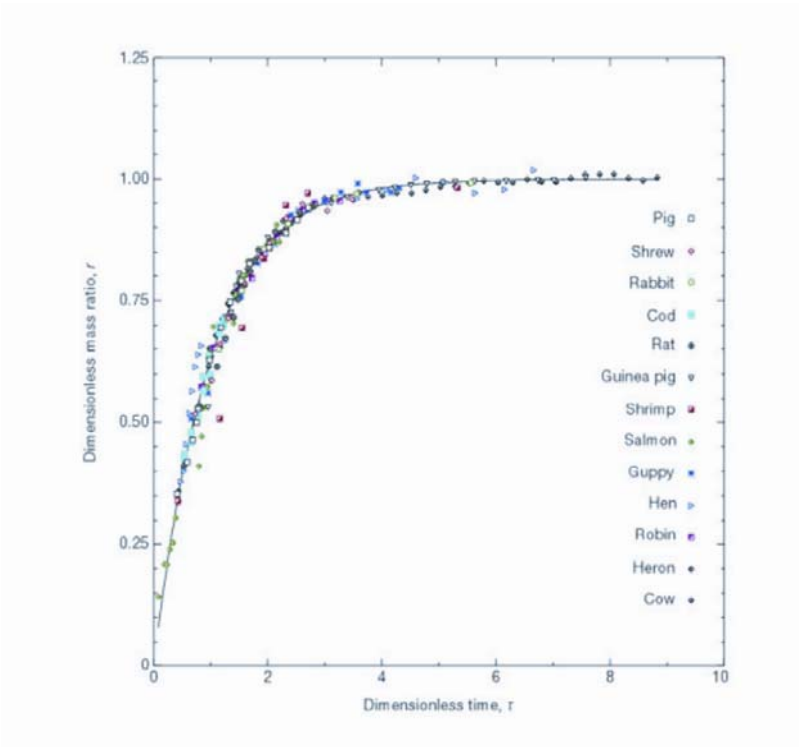


Figure 4. Universal growth curve. A plot of the dimensionless mass ratio, $r = 1 - R \equiv (m/M)^{1/4}$, versus the dimensionless time variable, $\tau = (at/4M^{1/4}) - \ln[1 - (m_0/M)^{1/4}]$, where $a = B_0 m_c/E_c$ and m_0 is the mass at birth. Data are for a wide variety of species with determinate and indeterminate growth. When plotted in this way, our model predicts that growth curves for all organisms should fall on the same universal parameterless curve $1 - e^{-\tau}$ (shown as a solid line). The model identifies r as the proportion of total lifetime metabolic power used for maintenance and other activities. This figure is reproduced with permission from West et al. (2001) (50).

for ontogenetic growth leads to scaling laws for other growth characteristics, such as doubling times and the relative energy devoted to maintenance.

3.2. Nucleotide Substitution Rates, Cellular Energetics, and Genome Size

One of the most interesting consequences of these ideas is that nucleotide substitution rates in DNA and rates of molecular evolution (52,53) can be characterized by combining the theory of biological scaling with Kimura's classic neutral theory of molecular evolution (54). It is assumed that point mutations, and therefore substitutions, occur at a rate proportional to metabolic rate. This is because most mutations are due to processes such as free radical production or

replication error during cell division, which are consequences of metabolism. The predictions of this theory are supported by molecular divergence data from mitochondrial and nuclear genomes (55). By accounting for the effects of body size and temperature on metabolic rate, a single molecular clock explains heterogeneity in rates of nucleotide substitution across genes, taxa, and thermal environments. This clock, however, "ticks" at a constant substitution rate per mass-specific metabolic rate rather than per unit time. For taxonomic groups with DNA repair rates and genome lengths that are roughly constant, this predicts that the number of substitutions per lifetime is invariant because lifetime scales inversely with metabolic rate (Figure 2).

In collaboration with our colleagues, we are currently analyzing other cellular level properties, including scaling laws for cell size, coding and non-coding genome lengths, and RNA abundances. Unicellular eukaryotes vary in mass and volume by several orders of magnitude ($\approx 10^{-10}$ to 10^{-6} g), and their metabolic rates scale as cell mass to the 3/4-power (Figure 1). This raises interesting questions about how external exchange surfaces and internal structures and rate processes within cells scale (13). Moreover, recent data for multicellular organisms suggest that the mass of certain cell types scale with body size whereas other types remain roughly constant. The scaling of cell size is therefore not just a question of the size of an organism or of endothermy versus ectothermy, but is also related to questions of functionality. Recent progress on the scaling of cell size and genome length suggests that these problems may well be interrelated.

By understanding how much variation in genome length is due just to changes in the size and temperature of the organism and identifying the residuals of these more dominant patterns, it may be possible to provide a better definition of biological or genomic complexity and possibly help resolve the C-value paradox (56–60), i.e., why is genome length seemingly uncorrelated with organismal complexity. A few groups have begun to study connections between the theory of biological scaling and genome length (14,15). Furthermore, it is well known that in eukaryotes total genome length scales linearly with cellular mass. Our colleagues and we believe that we have begun to understand this relationship, and we have discovered similar relationships for coding genome length. Further progress on the scaling of genome length may lead to very simple methods for estimating the gene number of an organism based on the cellular mass, and this could potentially facilitate gene searches.

3.3. Drug Dosing: Scaling from Rats to Humans

Tests to determine the appropriate level of drug dosage are often done on mice or rats and then extrapolated to humans (61,62). Guiot et al. (9) have discussed the use of the growth model described in §3.1, as applied to tumor growth, for determining levels of therapeutic tumor dosages. In mammals the extrapolation of drug dosage for the treatment of a given condition is, for simplicity, sometimes calculated assuming a linear dependence with body mass.

However, the global ability to process drugs should scale as the metabolic rate, $M^{3/4}$, and residence times should scale as the inverse of mass-specific metabolic rate, $M^{1/4}$. These 3/4 extrapolations can provide guidelines for drug dosing and are more appropriate than linear extrapolations. However, factors particular to the biochemistry of a given species are also important and must be used in tandem with allometric estimates.

To illustrate the problems associated with extrapolations of drug dosing, we consider the elephant, Tusko. West et al. (63) wanted to determine whether LSD had the same effect on elephants as a naturally occurring state in male elephants known as "on musth," which has associated biochemical changes (63). They cited studies that gave appropriate drug dosing of LSD for cats, rhesus macaques, and humans. Although appropriate drug dosing varies considerably among these species, their estimate for appropriate drug dosing of an elephant was a nearly linear extrapolation of the appropriate dosing for cats. Since an elephant weighs approximately 1000 times as much as a cat, this means Tusko received approximately 1000 times more LSD than a cat. In contrast, if drug dosing is scaled according to metabolic rate, the LSD dosing for an elephant is approximately 180 times that of a cat. While this calculation does not account for other, possibly important, factors, it naively suggests that Tusko was given five times more LSD than was appropriate. Tusko died about two hours after receiving the injection.

3.4. Spread of Infectious Diseases

Epidemiological equations are very similar to those for population growth, and it is likely that our theory for population growth and mortality rates (17) could be extended to study the spread of infectious diseases. Certainly, new factors particular to specific diseases must be considered. For instance, water abundance and biting rates of mosquitoes play an important role in the transmission rates of malaria (64). Our work on populations demonstrates that the biologically mediated processing of energy and materials is a direct consequence of the metabolic rates of the constituent organisms, and therefore, that describing higher-level phenomena in terms of individual-level interactions is fruitful. A similar theory could aid studies of the spread of infectious diseases, especially if combined with previous theoretical models that have already proved useful in intervention strategies (65,66).

3.5. Cancer Growth

Recently, Deisboeck and his collaborators have published exciting work on the application of the above growth model (§3.1) to tumor growth (8–10). Taking the ontogenetic growth model described above as a point of departure, they analyzed growth curves of tumors, both in vivo and in vitro, for rats and humans. Their results closely matched the form of the ontogenetic growth equa-

tion, but the parameters were determined by fitting the data and not from first principles. Although the basic ideas behind our growth model should apply to tumors, the theory does not include the presence of necrotic tissue. As such, the derived parameters might be misleading.

Concurrently, Alex Herman and we have been developing a theory for tumor growth that includes necrosis, with the initial goal of predicting the effects of tumor size and host size on the tumor's metabolic rate. Using the network theory for the cardiovascular system, we intend to make predictions that apply to both avascular and vascular tumors. For the latter, we are combining the mechanism of angiogenesis with the network theory (§2.1) to describe, or possibly to predict, the structure and dynamics of the tumor vascular system. Predictions about which properties of cancer vasculature allow cancer cells to grow quickly and at the expense of the rest of the body will be made. Preliminary results suggest that a substantial amount of tumor data can be parsimoniously explained and that disparate empirical findings for tumors can be interrelated in a novel way. We have a manuscript in preparation in which we provide a theory for the allometry of asymptotic tumor sizes and doubling times and derive tumor growth trajectories in a mechanistic fashion (11).

3.6. Sleep

Sleep is one of the most noticeable and widespread phenomena in multicellular animals, occurring in mammals, birds, amphibians, reptiles, and insects (67). Recent neurobiological studies have uncovered a great deal of information about the mechanisms involved in sleep, but a convincing demonstration of the function of sleep is considered one of the most important, unsolved problems in science. Some of the most-studied and best-known hypotheses for the function of sleep are related to metabolic rate. These include rest for the body or brain (6,68), cortical reorganization associated with memory and learning (69–73), and cellular repair in the body or brain (74–79). However, there is a remarkable absence of quantitative theories to elucidate or distinguish between these metabolically based theories for sleep.

We have developed a quantitative theory for mammalian sleep that relates fundamental parameters of sleep to metabolic rate and thus, to body size. This theory is based on the hypothesis that processes related to metabolism or metabolic damage, most notably cortical reorganization and cellular repair, occur during sleep. For example, sleep cycle time—the time between the endings of periods of REM sleep, i.e., the amount of REM sleep plus non-REM sleep in a single cycle—increases with body size in a systematic way (see Figure 5). Using this theory we are also able to derive previously unknown relationships between sleep time, awake time, and body size, and these relationships are supported by available data for mammals (7). These findings suggest that a metabolic theory for sleep is well founded and is possibly the dominant explanation for why animals sleep the amounts that they do.

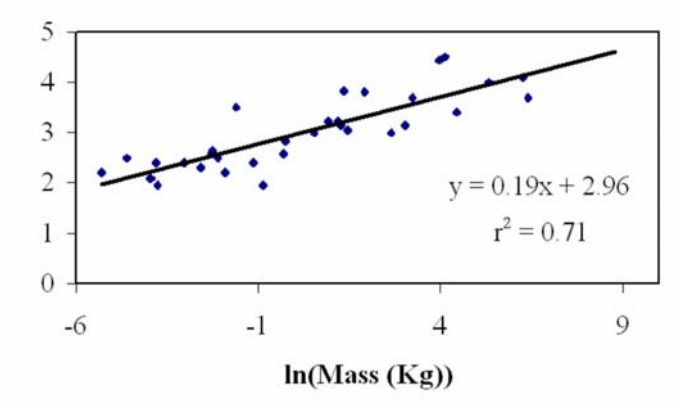


Figure 5. Plot of the logarithm of sleep cycle time, the period between REM and non-REM sleep, in minutes versus the logarithm of body size in Kg, ln(Mass(Kg)). The slope computed using OLS regression is 0.19 ($p < 0.0001$, $n = 32$, 95% CI: 0.14, 0.23). Note that the confidence intervals exclude the naive expectation of 1/4. This is because the scaling of sleep cycle time should scale as brain mass to the 1/4, and since brain mass scales as body mass to the 3/4 (93), the predicted exponent is actually 3/16 = 0.1875, which is in close agreement with the measured exponent. This figure will appear in Savage and West GB (to be published) (7).

3.7. Lifespan

The lifespan, L, of animals scales as $L = L_0 M^{1/4} e^{E/kT}$, reflecting the fact that larger, colder animals generally live longer than smaller, warmer ones (1,2). The scaling exponent is the same for all species, but among different taxa of mammals L_0 varies by more than a factor of two. Furthermore, birds live a factor of ten longer than mammals of similar mass. There is evidence that these variations may be due to differences in the rates of radical production by mitochondria and rates of DNA repair of different species (80–82). Consequently, combining theories of aging and empirical findings (83–87) with the theory of biological scaling may give new insights into the process of aging.

4. DISCUSSION AND CONCLUSIONS

The paradigm and principles developed here suggest novel ways of using quantitative analytic thinking to attack many fundamental problems in the biomedical sciences. This work has enormous potential at all scales and in a variety of different contexts, ranging from the highly practical (such as pharmacology, cancer growth, and aging as well as immunology; see this volume, Part III,

chapter 4.3, by Perelson, Bragg, and Wiegel) to the conceptual (such as formulating quantitative physical theories that capture the properties of intracellular network structures, illuminate the function of sleep, and are complementary to the algorithmic principles of the genetic code).

The success of the theory of biological scaling should be viewed as a mere beginning. By its very nature, the model should be thought of as a zeroth-order description of "average idealized organisms" that embody the essential features of biological systems. In the future, more detailed analyses should be performed that use the theory of biological scaling as a point of departure. Indeed, a major conceptual challenge presented by our research is, "Why does it work so well?" Is there some fixed point, or deep basin of attraction that operates within the general dynamical structure and ensures that, in spite of enormous complexity, the general features of biological systems are robust against significant perturbations (88)?

In conclusion, our theory offers a comprehensive, quantitative, integrated explanation for the scaling of many biological rates and times with body size and temperature. It explains why body size and temperature have such a powerful influence on biological structure and function at all levels of organization. As discussed in §3, the theory of biological scaling is already being applied in exciting ways to several biomedical systems, and we are hopeful that it will lead to important contributions to the biomedical field.

Methods:

Data for sleep cycle times are from (89), and methods are detailed therein. Data for body mass are not given in (89), so most values for body mass come from (90), which is an earlier compilation that (89) draws from heavily. When masses given in (90) did not correspond to sleep measurements listed in (89), the values given in (91) were taken, and when the mass could not be found there, we used the average of the range of values given in (92). Some of the original sources given in (89) were consulted to determine which species were used. In a few cases, the logarithmic averages of body masses were calculated for groups of species (e.g., four species of *Microtus* and five species of *Peromyscus*). This was done in order to be consistent with the original sleep data in (89).

5. ACKNOWLEDGMENTS

It is important to recognize that we carried out this work in a close, intense collaboration between biologists, biochemists, and physicists, and that the research represents an excellent example of successful interdisciplinary science. We would especially like to thank our collaborators listed on the title page of

this chapter and who comprise the "scaling group" at the Santa Fe Institute, the University of New Mexico, Los Alamos National Laboratory, and the University of Arizona. V.M.S. and G.B.W. are grateful for the support of the Thaw Charitable Trust, a Packard Interdisciplinary Science Grant, an NSF Biocomplexity Grant, and LANL/LDRD Grant 20030050DR. G.B.W. is also grateful for NSF Award PHY-0202180.

6. REFERENCES

1. Gillooly JF, Brown JH, West GB, Savage VM, Charnov EL. 2001. Effects of size and temperature on metabolic rate. *Science* **293**:2248–2251.
2. Peters RH. 1983. *The ecological implications of body size*. Cambridge UP, Cambridge.
3. Savage VM, Gillooly JF, Woodruff WH, West GB, Allen AP, Enquist BJ, Brown JH. 2004. The predominance of quarter-power scaling in biology. *Funct Ecol* **18**:257–282.
4. Blueweiss L, Fox H, Kudzma V, Nakashima D, Peters R, Sams S. 1978. Relationships between body and size and some life history parameters. *Oecologia* **37**:257–272.
5. Gillooly JF, Charnov EL, West GB, Savage VM, Brown JH. 2002. Effects of size and temperature on developmental time. *Nature* **417**:70–73.
6. Zepelin H, Rechtschaffen A. 1974. Mammalian sleep, longevity, and energy metabolism. *Brain Behav Evol* **10**:425–470.
7. Savage VM, West GB. Paper to be submitted to *Science*. In progress.
8. Delsanto PP, Guiot C, Degiorgis PG, Condat CA, Mansury Y, Deisboeck TS. A growth model for multicellular tumor spheroids. 2004. arXiv:biological-physics/0307136.
9. Guiot C, Degiorgis PG, Delsanto PP, Gabriele P, Deisboeck TS. 2003. Does tumor growth follow a "universal law"? *J Theor Biol* **225**:147–151.
10. Deisboeck TS, Mansury Y, Guiot C, Degiorgis PG, Delsanto PP. 2004. Insights from a novel tumor model: indications for a quantitative link between tumor growth and invasion. arXiv:biological-physics/0309096.
11. Herman AB, Savage VM, West GB. Paper to be submitted to *Nature*. In progress.
12. West GB, Brown JH, Enquist BJ. 1997. A general model for the origin of allometric scaling laws in biology. *Science* **276**:122–126.
13. West GB, Woodruff WH, Brown JH. 2002. Allometric scaling of metabolism from molecules and mitochondria to cells and mammals. *Proc Natl Acad Sci USA* **99**:2473–2478.
14. Kozlowski J, Konarzewski M, Gawelczyk AT. 2003. Cell size as a link between noncoding DNA and metabolic rate scaling. *Proc Natl Acad Sci USA* **100**:14080–14085.
15. Lynch M, Conery JS. 2003. The origins of genome complexity. *Science* **302**:1401–1404.
16. West GB, Brown JH, Enquist BJ. 1999. The fourth dimension of life: fractal geometry and allometric scaling of organisms. *Science* **284**:1677–1679.
17. Savage VM, Gillooly JF, Brown JH, West GB, Charnov EL. 2004. Effects of body size and temperature on population growth. *Am Naturalist* **163**:429–441.
18. Brown JH, Gillooly JF, Allen AP, Savage VM, West GB. 2004. Towards a metabolic theory of ecology. *Ecology* **85**:1771–1789.
19. Allen AP, Brown JH, Gillooly JF. 2002. Global biodiversity, biochemical kinetics, and the energetic-equivalence rule. *Science* **297**:1545–1548.
20. Ernest SKM, Enquist BJ, Brown JH, Charnov EL, Gillooly JF, Savage VM, White EP, Smith FA, Hadly EA, Haskell JP, Lyons SK, Maurer BA, Niklas KJ, Tiffney B. 2003. Statistical mechanics of complex ecological aggregates. *Ecol Lett* **6**:990–995.
21. Kleiber M. 1932. Body size and metabolism. *Hilgardia* **6**:315–353.
22. Brody, S. 1945. *Bioenergetics and growth*. Reinhold, New York.

23. Hemmingsen AM. 1960. Energy metabolism as related to body size and respiratory surfaces, and its evolution. *Rep Steno Memorial Hospital Nordisk Insulin Laboratorium (Copenhagen)*, **9**:6–110.
24. Enquist BJ, West GB, Charnov EL, Brown JH. 1999. Allometric scaling of production and life history variation in vascular plants. *Nature* **401**:907–911.
25. West GB, Brown JH, Enquist BJ. 1999. A general model for the structure, and allometry of plant vascular systems. *Nature* **400**:664–667.
26. Schmidt-Nielsen K. 1984. *Scaling: why is animal size so important?* Cambridge UP, Cambridge.
27. McMahon TA, Bonner JT. 1983. *On size and life*. Scientific American Library, New York.
28. Calder WA. 1984. *Size, function, and life history*. Harvard UP, Cambridge/
29. Cossins AH, Bowler K. 1987. *Temperature biology of animals*. Chapman and Hall, London.
30. Somero GS. 1997. Temperature relationships: from molecules to biogeography. In *Handbook of physiology*, Vol. 13, pp. 1391–1444. Ed. WH Dantzler. Oxford UP, New York.
31. Mandelbrot BB. 1982. *The fractal geometry of nature*. Freeman, San Francisco.
32. Womersley JR. 1955. Oscillatory motion of a viscous liquid in a thin-walled elastic tube, I: the linear approximation for long waves. *Philos Mag* **46**:199–221.
33. Caro CG, Pedley TJ, Schroter RC, Seed WA. 1978. *The mechanics of circulation*. Oxford UP, Oxford.
34. Fung YC. 1984. *Biodynamics*. Springer, New York.
35. Landau LD, Lifshitz EM. 1978. *Fluid mechanics*. Pergamon Press, Oxford.
36. Marion JB, Thornton ST. 1988. *Classical dynamics of particles and systems*. Harcourt Brace Jovanovich, San Diego.
37. West GB, Brown JH, Enquist BJ. 2000. The origin of universal scaling laws in biology. In *Scaling in biology*, pp. 87–112. Ed. JH Brown, GB West. Oxford UP, Oxford.
38. Iberall AS. 1967. Anatomy and steady flow characteristics of the arterial system with an introduction to its pulsatile characteristics. *Math Biosci* **1**:375–395.
39. Sherman TF 1981. On connecting large vessels to small: the meaning of Murray's law. *J Gen Physiol* **78**:431–453.
40. Zamir M, Sinclair P, Wonnacott TH. 1992. Relation between diameter and flow in major branches of the arch of the aorta. *J Biomech* **25**:1303–1310.
41. Zamir M. 1999. On fractal properties of arterial trees. *J Theor Biol* **197**:517–526.
42. Dodds PS, Rothman DH, Weitz JS. 2001. Re-examination of the "3/4-law" of metabolism. *J Theor Biol* **209**:9–27.
43. Arrhenius S. 1889. Über die Reaktionsgeschwindigkeit bei der Inversion von Rohzucker durch Sauren. *Z Physik Chem* **4**:226–248.
44. Glasstone S. 1940. *Textbook of physical chemistry*. Van Nostrand, New York.
45. Landau LD, Lifshitz EM. 1980. *Statistical physics*. Pergamon Press, Oxford.
46. Baierlein R. 1999. *Thermal physics*. Cambridge UP, Cambridge.
47. Westerhoff HV, Dam KV. 1987. *Thermodynamics and control of biological free-energy transduction*. Elsevier, Amsterdam.
48. Raven JA, Geider RJ. 1988. Temperature and algal growth. *New Phytologist* **110**:441–461.
49. Vetter RAH. 1995. Ecophysiological studies on citrate-synthase, I: enzyme regulation of selected crustaceans with regard to temperature. *J Comp Physiol B* **165**:46–55.
50. West GB, Brown JH, Enquist BJ. 2001. A general model for ontogenetic growth. *Nature* **413**:628–631.
51. West GB, Enquist BJ, Brown JH. 2002. Ontogenetic growth: modeling universality and scaling [brief communication]. *Nature* **420**:626–627.
52. Zuckerkandl E, Pauling L. 1965. Evolutionary divergence and convergence in proteins. In *Evolving genes and proteins*, pp. 97–166. Ed. V Bryson, HJ Vogel. Academic Press, New York.
53. Li WH. 1997. *Molecular evolution*. Sinauer Associates, Sunderland, MA.

54. Kimura M. 1983. *The neutral theory of molecular evolution*. Cambridge UP, Cambridge.
55. Gillooly JF, Allen AP, West GB, Brown JH. 2005. The rate of DNA evolution: effects of body size and temperature on the molecular clock. *Proc Natl Acad Sci USA* **102**:140–145.
56. Beaton MJ, Cavalier-Smith T. 1999. Eukaryotic non-coding DNA is functional: evidence from the differential scaling of cryptomonad genomes. *Proc Roy Soc London B—Biol Sci* **266**:2053–2059.
57. Cavalier-Smith T. 1980. r- and K-tactics in the evolution of protist developmental systems: cell and genome size, phenotype diversifying selection, and cell cycle patterns. *Biosystems* **12**(1–2):43–59.
58. Gregory TR. 2001. The bigger the C-value, the larger the cell: genome size and red blood cell size in vertebrates [review]. *Blood Cells Mol Dis* **27**:830–843.
59. Gregory TR. 2001. Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma [review]. *Biol Rev Camb Philos Soc* **76**:65–101.
60. Olmo E, Capriglione T, Odierna G. 1989. Genome size evolution in vertebrates: trends and constraints [review]. *Comp Biochem Physiol B* **92**:447–453.
61. Chandler M, Spain M. 2003. Pitfalls to using small animals in preclinical testing are being eliminated. *Preclinica* **1**:1–2.
62. Jain RK. 1996. Delivery of molecular medicine to solid tumors. *Science* **271**:1079–1080.
63. West LJ, Pierce CM. Thomas WD. 1962. Lysergic acid diethylamide: its effect on a male Asiatic elephant. *Science* **138**:1100–1103.
64. MacDonald G. 1957. *The epidemiology and control of malaria*. Oxford UP, London.
65. Ancel LW, Levin BR, Richardson AR, Stojilkovic I. 2001. *Two-tiered evolution of Neiserria meningitis: how within-host ecology and between-host epidemiology expedite phase shifting*. Santa Fe Institute Working Paper 01-12-079.
66. Ancel LW, Newman MEJ, Martin M, Schrag S. 2002. *Applying network theory to epidemic intervention: modelling the spread and control of Mycoplasma pneumoniae*. Santa Fe Institute Working Paper 01-12-078.
67. Campbell SS, Tobler I. 1984. Animal sleep: a review of sleep duration across phylogeny. *Neurosci Biobehav Rev* **8**:269–300.
68. Mamelak M. 1997. Neurodegeneration, sleep, and cerebral energy metabolism: a testable hypothesis. *J Geriat Psychiatry Neurol* **10**:29–32.
69. Maquet P. 2001. The role of sleep in learning and memory. *Science* **294**:1048–1052.
70. Frank MG, Issa NP, Stryker MP. 2001. Sleep enhances plasticity in the developing visual cortex. *Neuron* **30**:275–287.
71. Hoffman KL, McNaughton BL. 2002. Sleep on it: cortical reorganization after-the-fact. *Trends Neurosci* **25**:1–2.
72. Stickgold R, Hobson JA, Fosse R, Fosse M. 2001. Sleep, learning, and dreams: off-line memory reprocessing. *Science* **294**:1052–1057.
73. Tononi G, Cirelli C. 2003. Sleep and synaptic homeostasis: a hypothesis. *Brain Res Bull* **62**:143–150.
74. Adam K, Oswald I. 1977. Sleep is for tissue restoration. *J Roy Coll Phys London* **11**:376–388.
75. Inoue S, Honda K, Komoda Y. 1995. Sleep as neuronal detoxification and restitution. *Behav Brain Res* **69**:91–96.
76. Moruzzi G. 1966. The functional significance of sleep with particular regard to the brain mechanisms underlying consciousness. In *Brain and conscious experience*, pp. 345–379. Ed. JC Eccles, Springer, Berlin.
77. Reimund E. 1991. Sleep deprivation-induced neuronal damage may be due to nicotinic acid depletion. *J Med Hypo* **34**:275–277.
78. Reimund E. 1994. The free radical flux theory of sleep. *J Med Hypo* **43**:231–233.
79. Siegel JM. 2003. Why we sleep. *Sci Am* **289**:92–97.
80. Cortopassi GA, Wang E. 1996. There is substantial agreement among interspecies estimates of DNA repair activity. *Mech Ageing Dev* **91**:211–218.

81. Perez-Campo R, Lopez-Torres M, Cadenas S, Rojas C, Barja G. 1998. The rate of free radical production as a determinant of the rate of aging: evidence from the comparative approach. *J Comp Physiol B—Biochem Syst Environ Physiol* **168**:149–158.
82. Yu BP. 1993. *Free radicals in aging*. CRC Press, Boca Raton, FL.
83. Thorbecke GJ. 1975. *FASEB Monog*, ed. KF Heumann. Plenum Press, New York.
84. Arantes-Oliveira N, Berman JR, Kenyon C. 2003. Healthy animals with extreme longevity. *Science* **302**:611.
85. Sohal R, Weindruch R. 1996. Oxidative stress, caloric restriction, and aging. *Science* **273**:59–63.
86. Weindruch R. 1996. Caloric restriction and aging. *Sci Am* **274**:46–52.
87. Weindruch R, Walford RL. 1988. *The retardation of aging and disease by dietary restriction*. Charles C. Thomas, Springfield, IL.
88. Mauroy B, Filoche M, Weibel ER, Sapoval B. 2004. An optimal bronchial tree may be dangerous. *Nature* **427**:633–636.
89. Zepelin H. 1989. Mammalian sleep. In *Principles and practice of sleep medicine*, pp. 30–49. Ed. MH Kryger, T Roth, WC Dement. Saunders, Philadelphia.
90. Meddis R. 1983. The evolution of sleep. In *Sleep mechanisms and functions in humans and animals*, pp. 57–106. Ed. A Mayes. Van Nostrand Reinhold, Cambridge.
91. Smith FA, Lyons SK, Ernest SKM, Jones KE, Kaufman DM, Dayan T, Marquet PA, Haskell JP. 2003. The body mass of late Quaternary mammals. *Ecology* **84**:3403, E084.
92. Nowak RN. 1991. *Walker's mammals of the world*. Johns Hopkins UP, Baltimore.
93. Pagel MD, Harvey PH. 1991. The taxon-level problem in the evolution of mammalian brain size: fact and artifacts. *Am Naturalist* **132**:344–359.

4

THE ARCHITECTURE OF BIOLOGICAL NETWORKS

Stefan Wuchty, Erszébet Ravasz, and Albert-László Barabási

Department of Physics, University of Notre Dame, Notre Dame, Indiana

An ambitious goal of contemporary biological research is the elucidation of the structure and functions of networks that constitute cells and organisms. In biological systems, networks appear in many different disguises, ranging from protein interactions to metabolic networks. The emergence of these networks is driven by self-organizing processes that are governed by simple but generic laws. While unraveling the complex and interwoven systems of different interacting units, it has become clear that the topology of networks of different biological origin share the same characteristics on the large scale. In this chapter, we survey the most prominent characteristics of biological networks, focusing on the emergence of the scale-free architecture and the hierarchical arrangement of modules.

1. INTRODUCTION

Understanding complex systems often requires a bottom–up approach, breaking the system into small and elementary constituents and mapping out the interactions between these components. In many cases, the myriad components and interactions are best characterized as networks. For example, the society is a network of people connected by various links, including friendships (36), collaborationships (29,59), sexual contacts (33), or scientific coauthorships (437,8). Electronic communication relies on two very different networks: the physical network wiring the routers together (Internet) (1,14) and the web of homepages linked by URLs (World Wide Web) (2,9,32). Airline, cell-phone, power-grid, or

Address correspondence to: Albert-László Barabási, Department of Physics, 225 Nieuwland Science Hall, University of Notre Dame, Notre Dame, IN 46556 USA; tel 574-631-5767, fax 574-631-5259 (alb@nd.edu).

business networks represent further examples of complex networks of technological, scientific or economic interest.

In biological systems networks emerge in many disguises, from food webs in ecology to various biochemical nets in molecular biology. In particular, the wide range of interactions between genes, proteins, and metabolites in a cell are best represented by various complex networks. During the last decade, genomics has produced an incredible quantity of molecular interaction data, contributing to maps of specific cellular networks. The emerging fields of transcriptomics and proteomics have the potential to join the already extensive data sources provided by the genome-wide analysis of gene expression at the mRNA and protein levels (10,11,40). Indeed, extensive protein–protein interaction maps have been generated for a variety of organisms including viruses (16,34), prokaryotes, like *H. pylori* (45), and eukaryotes, like *S. cerevisiae* (17,22–24,49,52) and *C. elegans* (58). Beyond the current focus on uncovering the structure of the genomes, proteomes, and interactomes of various organisms, some of the most extensive data sets are the metabolic maps (28,39), catalyzing an increasing number of studies, focusing on the architecture of the metabolism (15,26,57).

Networks offer us a new way to categorize systems of very different origin under a single framework. This approach has uncovered unexpected similarities between the organization of various complex systems, indicating that the networks describing them are governed by generic organization principles and mechanisms. Understanding the driving forces that invest different networks with similar topological features enables systems biology to combine the numerous details about molecular interactions into a single framework, offering means to address the structure of the cell as a whole.

2. BASIC NETWORK FEATURES

A node's degree (or connectivity), giving the number of links k the node has, is the most elementary network measure. For example, in Figure 1 node i has exactly three links ($k_i = 3$). The overall graph, however, is characterized by the average degree, $\langle k \rangle$, which has the value $\langle k \rangle = 2.6$ for this example. Yet, the average degree does not capture the potential degree variations present in the network. This is better characterized by the degree distribution, $P(k)$, which gives the number of nodes with exactly k links (see Table 1).

Planing a trip from Anchorage, Alaska, to Alice Springs in the outbacks of Australia requires finding the shortest paths through a particular airline's transportation network. As in most networks, there are multiple paths between any two nodes i and j; a useful distance measure is the length of the shortest path, l_{ij} (see Figure 1). In a network of N nodes, the mean path length is defined as

$$\langle l \rangle = \frac{2}{N(N-1)} \sum_{j \neq l} \sum_{i=1}^{N} l_{ij}, \qquad [1]$$

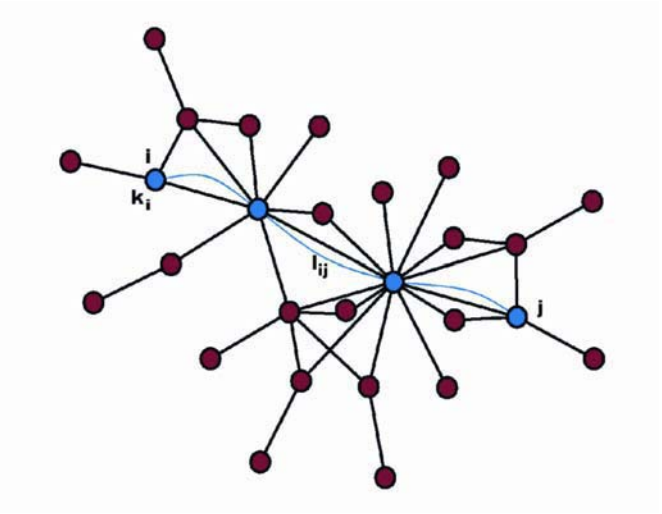


Figure 1. Characterizing a simple network: in the figure, both nodes, *i* and *j*, have three links (*k* = 3). The shortest path between these nodes, indicated in blue, has length $l_{ij} = 3$.

Table 1

Parameter	Definition
Total number of nodes	N
Degree of node *i*	k_i
Shortest path length	l_{ij}
Mean path length	$\langle l \rangle = \frac{2}{N(N-1)} \sum_{k \neq i} \sum_{i=1}^{N} l_{ij}$
Clustering coefficent	$C_i = \frac{2n_i}{k_i(k_i - 1)}$
Mean clustering coefficient	$\langle C \rangle = \frac{1}{N} \sum_{i=1}^{N} C_i$

offering a measure of the network's navigability. A network that can be "crossed" by a relatively small number of steps is often referred to as displaying the "small world" property, first illustrated on social networks, indicating that two randomly chosen individuals can be connected by only six intermediate acquaintances (36).

Nodes in many real systems exhibit a tendency to cluster, which can be quantified using the clustering coefficient (60), a measure of the degree to which

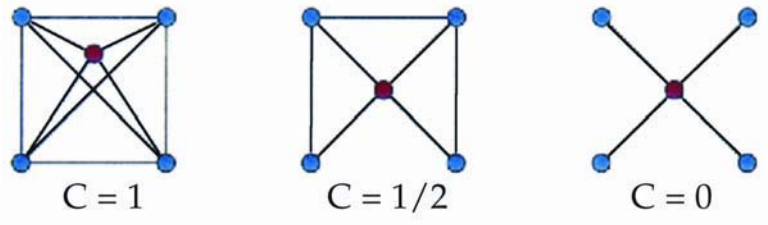


Figure 2. The clustering coefficient *C* offers a measure of the degree of interconnectivity in the neighborhood of a node. For example, the red node whose neighbors are all connected to each other has $C = 1$ (left), whereas the red node with no links between its neighbors has $C = 0$ (right).

the neighbors of a particular node are connected to each other (Figure 2). For example, in a friendship network *C* reflects the degree to which friends of a particular person are friends with each other as well. Formally, the clustering coefficient of node *i* is defined as

$$C_i = \frac{2n_i}{k_i(k_i - 1)}, \qquad [2]$$

where n_i denotes the number of links connecting the k_i neighbors of node *i* to each other. Accordingly, we can define the average clustering coefficient as

$$\langle C \rangle = \frac{1}{N}\sum_{i=1}^{N} C_i . \qquad [3]$$

An additional important measure of the network's structure is the function $C(k)$, defined as the average clustering coefficient of all nodes with *k* links. If $C(k)$ is independent of *k*, the network is either homogeneous or it is dominated by numerous small tightly linked clusters. In contrast, if $C(k)$ follows $C(k) \sim k^{-1}$, the network has a hierarchical architecture, meaning that sparsely connected nodes are part of highly clustered areas (12,27,46,47). In such hierarchical networks communication between the different highly clustered neighborhoods is maintained by a few hubs.

As we will see below, the degree distribution $P(k)$ and the *k* dependence of $C(k)$ can have generic features, allowing us to classify various networks. Parameters such as the average degree $\langle k \rangle$, average path length $\langle l \rangle$, and average clustering coefficient $\langle C \rangle$ characterize the unique properties of the particular network under consideration, and therefore are less generic.

3. NETWORKS MODELS

The main role of network models is to explain the emergence and behavior of some of the most important network characteristics. As they play a crucial role in shaping our understanding of complex networks, we need to pay attention to some of the more important models.

3.1. Random Networks

While graph theory initially focused on regular graphs, since the 1950s large networks with no apparent design principles, described as random graphs (8), were proposed as the simplest and most straightforward realization of a complex network. According to the Erdös-Rényi (ER) model of random graphs (13), we start with N nodes and connect each pair of nodes with probability p, creating a graph with approximately $[pN(N-1)]/2$ randomly distributed links (first column in Figure 3). The ER graph has an exponential degree distribution and exhibits the small-world property. Indeed, in the ER network most nodes have approximately the same number of links, $k \approx \langle k \rangle$ (first column in Figure 4), and the mean path length is proportional to the network size, $\langle l \rangle \sim \log N$.

The growing interest in complex systems prompted many scientists to ask a simple question: are real networks behind diverse complex systems, like the cell, fundamentally random?

3.2. Scale-free Networks

A highly nontrivial development in our understanding of complex networks was the discovery that for most large networks, including the metabolic and protein interaction networks (24,26), the degree distribution follows a power-law:

$$P(k) \sim k^{-\gamma}. \qquad [4]$$

These networks are called scale-free, as a power law does not support the existence of a characteristic scale. Two mechanisms, absent from the classical random network model, are responsible for the emergence of this power-law degree distribution (5,6). First, most networks grow through the addition of new nodes, which link to nodes already present in the system. Second, in most real networks there is a higher probability to link to a node with a large number of connections, a property called preferential attachment. The scale-free model introduced by Barabási and Albert (the BA model; second columns in Figures 3 and 4) incorporates these features (5). Starting from a small graph, at each time step a

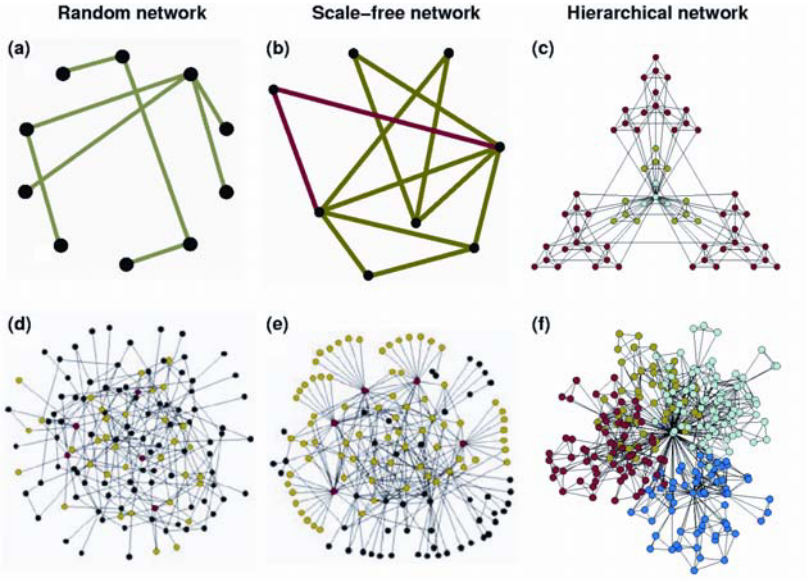


Figure 3. (a) The random network model is constructed by laying down N nodes and connecting each pair of nodes with probability p. The figure shows a particular realization of such a network for $N = 10$ and $p = 0.2$. (b) The scale-free model assumes that the network constantly grows by the addition of new nodes. The figure depicts the network at time t (nodes connected by green links) and after addition of a new node at time $t + 1$ (red links). With the introduction of new nodes, already highly connected ones are more favored to be connected to the new one than less connected ones. This procedure is called *preferential attachment*. (c) The iterative construction of a hierarchical network starts from a fully connected cluster of four nodes (blue), which is replicated three times. Subsequently, the peripheral nodes of each replica (green) are connected to the central node of the original module. Repeating replication and the connection step with the 16-node module (red) leads to a 64-node network that provides scale-free topology and is built by nested modules. (d) The random network is rather homogeneous, i.e., most nodes have about the same number of links. (e) In contrast, a scale-free network is extremely inhomogeneous: while the majority of nodes have one or two links, a few nodes have a large number of links, preserving system integrity. To show this, five nodes with the highest number of links are colored red and their first neighbors green. While in the random network only 27% of the nodes are reached by the five most-connected nodes, in the scale-free network more than 60% are, demonstrating the key role hubs play in a scale-free network. Note that both networks contain the same number of nodes and links. (f) A hierarchical network still preserves a scale-free organization and displays the inherent modularity of nodes. The node's affiliation with a certain module is indicated by different colors. However, the underlying network structure clearly indicates blurred boundaries of its modules.

node with m links is added to the network, connecting to a previously present node i with probability

$$\Pi_i = \frac{k_i}{\sum_j k_j}, \qquad [5]$$

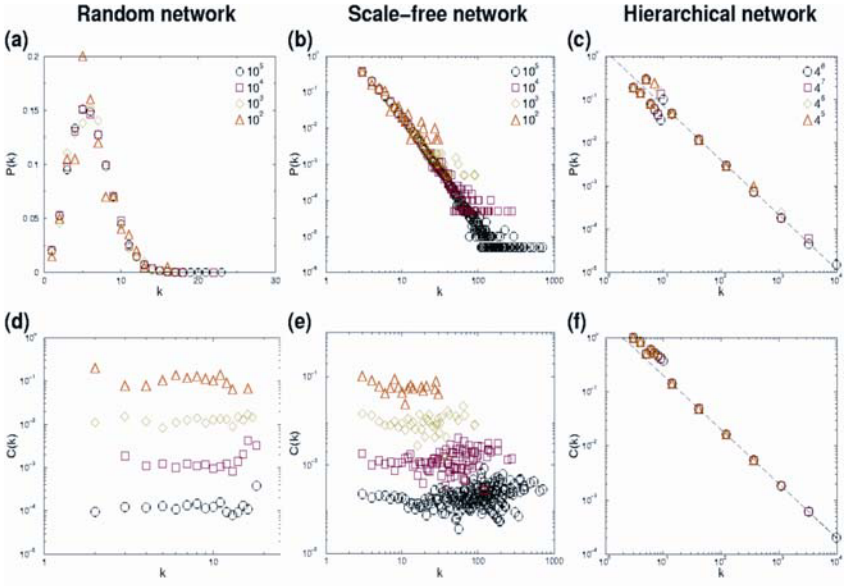


Figure 4. (**a**) For the random graph, the degree distribution, $P(k)$, which gives the probability that a randomly selected node has exactly k edges, follows a Poisson distribution that is strongly peaked at the average degree $<k>$ and decays exponentially for large k. (**b,c**) The $P(k)$ values of a scale-free and a hierarchical network do not have a peak and decay as a power-law, $P(k) \sim k^{\gamma}$. (**d,e**) For both a random and scale-free network, the $C(k)$ function, which denotes the mean clustering coefficient for nodes with exactly k links, is independent of k. (**f**) In contrast, $C(k)$ of a hierarchical network depends on k, decaying as $C(k) \sim k^{-1}$. Insets correspond to the number of underlying networks.

where k_i is the degree of node i. The network generated by this growth process will be scale-free with degree exponent $\gamma = 3$. In a scale-free network the probability that a node is highly connected ($k >> <k>$) is statistically more significant than in a random graph. Thus, the network's properties are often determined by a relatively small number of highly connected nodes or hubs. An important consequence of the hubs is that scale-free networks exhibit high tolerance to random perturbations but are sensitive to targeted attack upon the highly connected nodes (3). Accordingly, failure of randomly selected nodes cannot destroy the network's integrity. However, systematic removal of the hubs will rapidly fragment the network. This feature is of particular importance for biological systems, since it reflects a biochemical network's resilience against random mutations. Therefore, highly connected nodes in biochemical networks might be potential candidates for drug targets.

The presence of hubs in a scale-free network has a fundamental impact on virus spreading as well. Classical epidemiological models predict that infectious diseases with transmission probability under an epidemic threshold will inevita-

bly die out. However, in scale-free networks the epidemic threshold is reduced to zero (43). Thus, as some social and sexual networks are known to exhibit a scale-free topology (33), even extremely weakly infectious viruses can spread and prevail, making random immunization ineffective.

3.3. Hierarchical Networks

Many real networks are expected to be fundamentally modular, meaning that the network can be seamlessly partitioned into a collection of modules. Each module is expected to perform an identifiable task, separable from the function of other modules (19,31,50,61). Therefore, we must reconcile the scale-free property with the network's potential modularity. Numerical simulations indicate that neither the random nor the scale-free network model are modular.

In order to account for the coexistence of modularity, local clustering, and scale-free topology in real systems, we have to assume that clusters combine in an iterative manner, generating a hierarchical network (7,46). Such networks emerge from an iterative duplication and integration of clustered nodes, a process that in principle can be repeated indefinitely. Our starting point is a small cluster of four densely linked nodes. Next we generate three replicas of this hypothetical module and connect the three external nodes of the replicated clusters to the central node of the old cluster, obtaining a large 16-node module. Subsequently, we again generate three replicas of this 16-node module and connect the 16 peripheral nodes to the central node of the old module, obtaining a new module of 64 nodes (third column of Figure 3).

The hierarchical network model seamlessly integrates a scale-free topology with an inherent modular structure by generating a network that has a power-law degree distribution with degree exponent $\gamma = 1 + \ln 4/\ln 3 = 2.26$. Yet the most important signature of this hierarchical modularity is the fact that the clustering coefficient, $C(k)$, scales as k^{-1} (third column of Figure 4). Note that for a network generated by the ER and BA models $C(k)$ is independent of k.

Modularity does not, however, imply clear-cut subnetworks that are linked in well-defined ways. In fact, the boundaries of modules are often considerably blurred, triggered by highly connected nodes that interconnect modules.

4. BIOLOGICAL NETWORKS

4.1. Metabolic Networks

The structure of metabolic networks was addressed by two independent studies by Fell and Wagner and Jeong et al. Fell and Wagner assembled a list of stoichiometric equations that represent the central routes of the energy metabo-

lism and small-molecule building block synthesis in *E. coli* (15,57). A substrate graph was defined by the nodes representing all metabolites, two substrates being considered linked if they occurred in the same reaction. They found the substrate graph to be scale-free, with **glutamate**, **coenzyme A**, **2-oxoglutarate**, **pyruvate**, and **glutamine** having the highest degree, which were viewed as an evolutionary core of the *E. coli*.

At the same time, Jeong et al. analyzed the metabolic networks of 43 organisms representing all three domains of life (26), finding that the power-law degree distribution for both incoming and outgoing edges holds for organisms of all kingdoms. Furthermore, the average separation between nodes has the same value for all organisms under consideration, regardless of the number of substrates found in the given species. Interestingly, the ranking of the most connected substrates is largely identical for all organisms. A recent study comparing the system-level properties of metabolic networks in various organisms indicates that the structural features of these networks are more conserved than the components themselves (44,61).

4.2. Protein Interaction Networks

Protein interactions offer another opportunity to study cellular networks, considering proteins as nodes and physical interactions (binding) as links. It has been shown that interaction networks of *S. cerevisiae* and *H. pylori* proteins exhibit distinct scale-free behavior (24,56; see also this volume, Part III, chapter 1.3, by Wagner). Although protein interaction data are derived from different sources and retrieved by different methods, the emergence of the scale-free property appears to be a robust feature. As previously discussed, scale-free networks are vulnerable to targeted attack on their highly connected nodes. Therefore, mutations of highly interacting proteins are expected to be lethal for the cell. This prediction is supported by explicit measurements (25). Figure 5 represents the yeast protein interaction network, illustrating the basic feature that hubs keep many sparsely connected nodes together.

4.3. Protein Domain Networks

The domain architecture of proteins was studied by considering protein domains as nodes and their co-occurrence in proteins as links (4,62,63), documenting again the emergence of a scale-free architecture. Although the methods and sources of domain information were different, the scale-free features of the networks were found to be robust. Domains that appear in cellular functions crucial for the maintenance of multicellular organisms, such as signal transduction and cell–cell contacts, were found to be the most connected. Thus, domains like

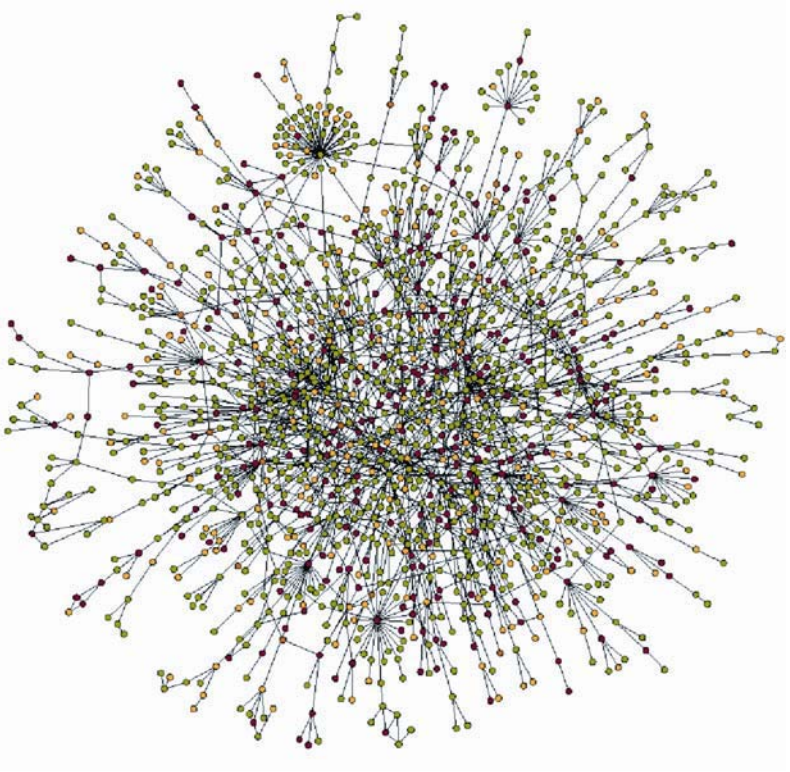


Figure 5. Map of the protein–protein interaction network of *S. cerevisiae* (24). The color code of nodes refers to the phenotypic effect the deletion of the respective protein has on the organism (red: lethal; green: viable; orange: slowed growth; yellow: unknown).

kinases, immunoglobulins, and zinc-fingers played an important role. Interestingly, the increasing complexity of an organism's domain architecture was found to decrease the slope of the degree distribution, and highly connected domains constantly accumulated links due to the organismic complexity. Similarly, interactions of domain families generated from sequence and structural data (41,63) revealed that highly connected domains on sequence level appear to be the most frequently interacting as well.

4.4. Hierarchies in Biological Networks

The clustering coefficient of metabolic networks varies with the inverse degree, $C(k) \sim k^{-1}$, indicating the presence of a hierarchical modularity. In order to discern the discrete modules, we can define a topological overlap, which

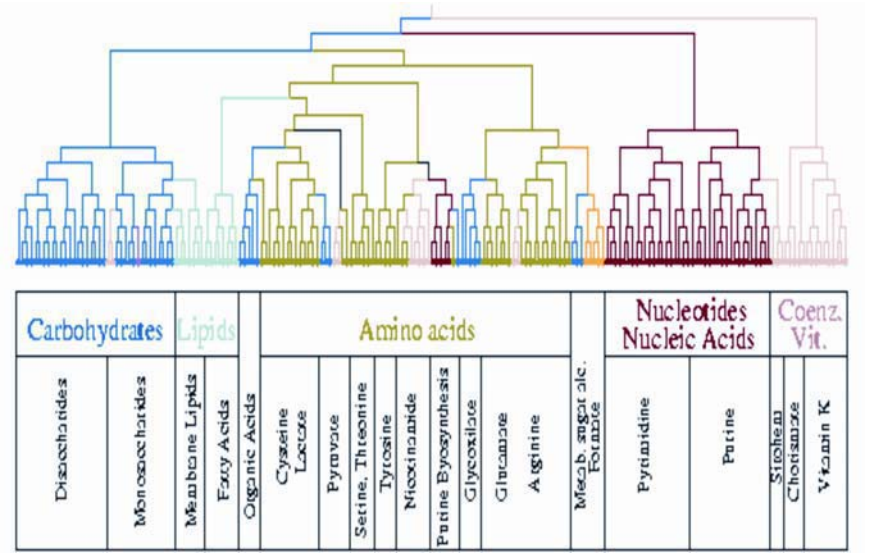


Figure 6. Hierarchies of topological modules in *E. coli* metabolism. The branches of the tree obtained by average-linkage clustering of the topological overlap of metabolites (47) are color-coded to reflect the predominant biochemical classification of their substrates. The biochemical classes represent carbohydrate metabolism (blue); nucleotide and nucleic acids metabolism (red); protein, peptide, and amino acid metabolism (green); lipid metabolism (cyan); aromatic compound metabolism (dark pink); monocarbon metabolism (yellow), and coenzym metabolism (light orange) (39).

scales from 0 to 1, reflecting the degree to which two metabolites *i* and *j* interact with the same substrates. Substrates that are part of larger metabolic modules appear to have a high topological overlap with their neighbors. The application of average-linkage clustering to the obtained overlap matrix has been used to uncover the topological modules present in the metabolism (Figure 6). The clustering identified a hierarchy of nested topological modules of increasing size and decreasing interconnectedness. The hierarchical tree offers a breakdown of the metabolism into several large modules that are further partitioned into smaller but more integrated submodules, reflecting a certain degree of inherent self-similarity. Some of these modules have been found to be in excellent agreement with the known functional classification of metabolites. Other approaches to discern modules in metabolic networks focused on the appearance of edges in mutual shortest paths within the network (18,21). The most frequent edges were identified and removed in an iterative manner, uncovering again the underlying functional modules.

Finally, modularity is not an exclusive property of the metabolism. Indeed, the protein interaction network of *S. cerevisiae* (66), based on four independent

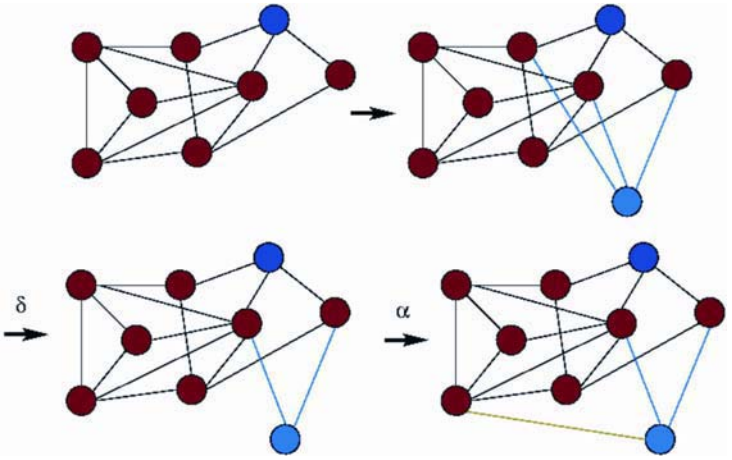


Figure 7. Mechanism of gene duplication and divergence model. At each time step a gene is randomly duplicated, retaining all of its links (blue nodes and edges). Subsequently, interactions of the duplicated gene are deleted or newly added with probabilities δ and α, respectively (green edge).

databases (22,35,52,65), and the conformational spaces of RNA (64), also reflect a modular architecture.

4.5. Mechanisms of Proteome Evolution

The origin of scale-free behavior in biological networks continues to offer some unresolved questions. Recently, however, it has been shown that a simple model based on gene duplication can lead to the experimentally observed scale-free topology of protein–protein interaction networks (42,51,53,56). In the model, at each time step a gene is randomly chosen and duplicated. The copied gene retains all interactions of the original gene. To mimic the potential loss or gain of interactions due to random mutations, interactions of the duplicated genes are deleted or newly added with probabilities δ and α, respectively (Figure 7). The emerging network can be shown analytically to have a power-law degree distribution, a high clustering coefficient, and a visual structure similar to the protein–protein interaction network shown in Figure 5.

5. CONCLUSIONS

A power-law degree distribution, the quantitative signature of a scale-free network, has emerged as one of the few universal laws characterizing cellular networks. Of even greater immediate importance is the intriguing possibility of using the insights provided by scale-free models as a framework to facilitate

analysis of biological networks at a higher level of abstraction. Such approaches could reveal salient features of biological phenomena missed by non-network-based approaches.

The appearance of hierarchical modularity in biological networks supports the assumption that evolution acts on many levels. The accumulation of local changes, affecting the small highly integrated modules, slowly impacts the larger, less integrated modules as well. Thus, evolution might act in a self-similar fashion, copying and reusing existing modules to further increase the organism's complexity. Especially in the face of eukaryotic evolution, this network-based framework might be suitable to describe the explosion of complexity in the development of the single-celled *S. cerevisiae* toward the multicellular *H. sapiens*.

It is widely accepted that different cellular functions—such as information storage, processing, and execution—are carried out by the genome, transcriptome, proteome, and metabolome. Although the functional distinction between these organizational levels is not always clear cut since, e.g., the proteome is crucial for short-term information storage, all cellular functions can be described by networks of various heterogeneous components. One way to visualize the complex relationships between these components is to organize them into a simple complexity pyramid (38) in which various molecular components—genes, RNAs, proteins and metabolites—organize themselves into recurrent patterns such as metabolic pathways and genetic regulatory motifs. In turn, motifs and pathways are seamlessly integrated to form functional modules that are responsible for distinct cellular functions (19). These modules are nested in a hierarchical fashion and define the cell's large-scale organization (Figure 8).

Our present knowledge about the architecture of biological networks emphasizes two major aspects: (1) discrete cellular functions are mediated with the aid of distinct albeit often-blurred modules; (2) network integrity is assured by a handful of highly connected nodes, making networks robust against random failures but exceedingly vulnerable to targeted attack. These features explain the observation that many mutations have little or no phenotypic effect (55), which appears to be consistent with the presence of genes that either cannot propagate their failure or whose function can be replaced by other components of the network. The presence of genes that integrate multiple signals and can trigger widespread changes upon their failure proves the crucial role of highly connected genes.

For example, the tumor suppressor gene **p53** has been identified as such a highly connected and thus crucial node, which, once mutated, severely jeopardizes genome stability and integration of signals related to control of the cell-cycle and apoptosis (30,54). Emphasizing its crucial role, disfunctional p53 proteins are involved in more than half of all human cancer phenotypes. From a biomedical point of view, highly connected proteins in general and proteins that maintain the integrity of modules can be perceived as disease factors and thus potential drug targets. With the increasing ability to identify and collect protein–

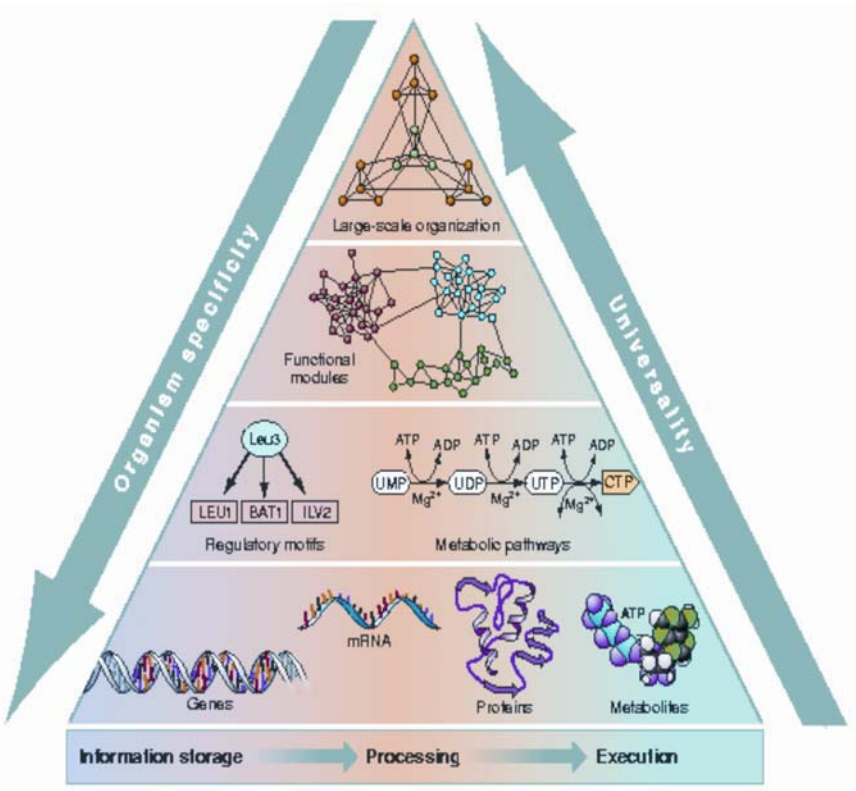


Figure 8. From the *particular* to the *universal*. The bottom (level 1) of the pyramid shows schematic representations of the cell's functional organization: genome, transcriptome, proteome, and metabolome. Insights into the cell's organization can be obtained if we consider the components to be linked by functional relationships, such as regulatory motifs and metabolic pathways (level 2). In turn, they are the building blocks of operational modules (level 3), which are nested and considerably blurred, generating a scale-free hierarchical architecture (level 4). Although the individual components are unique, the topological properties of biological networks share astounding similarities. This suggests that universal organizing principles apply to all kinds of complex networks (38).

protein interactions, the determination of modules and highly connected proteins will become a major issue in the fast and effective identification of potential drug targets.

The recent progress in biological networks has successively uncovered the skeleton and organization of networks, offering important insights about the assembly and functionality of components and subnetworks. In the future we will need to go several steps further, addressing the dynamic aspects of various cellular networks (see also this volume, Part III, chapter 2.1, by Huang, Sultan, and Ingber). The analysis of fluxes and fluctuations along the links in metabolic

and regulatory pathways will play an especially major role, significantly influencing potential biotechnological applications.

6. REFERENCES

1. Vázquez A, Pastor-Satorras R, Vespignani A. 2002. Large-scale topological and dynamical properties of the internet. *Phys Rev E* **65**:066130.
2. Albert R, Jeong H., Barabási A-L. 1999. Diameter of the world-wide web. *Nature* **401**:130–131.
3. Albert R, Jeong H, Barabási A-L. 2000. Attack and error tolerance of complex networks. *Nature* **406**:378.
4. Apic G, Gough J, Teichmann S. 2001. Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *J Mol Biol* **310**:311–325.
5. Barabási A-L, Albert R. 1999. Emergence of scaling in random networks. *Science* **286**:509–512.
6. Barabási A-L, Albert R, Jeong H. 1999. Mean-field theory for scale-free random networks. *Physica A* **272**:173–187.
7. Barabási A-L, Ravasz E., Vicsek T. 2001. Deterministic scale-free networks. *Physica A* **299**:559–564.
8. Bollobás B. 1985. *Random graphs*. Academic Press, London.
9. Broder A, Kumar R, Maghoul F, Raghavan P, Rajalopagan S, Stata R, Tomkins A, Wiener J. 2000. Graph structure in the web. *Comput Netw* **33**:309–320.
10. Burge C. 2001. Chipping away at the transcriptome. *Nature Genet* **27**:232–234.
11. Caron H, van Schaik B, van der Mee M, Baas F, Riggins G, van Sluis P, Hermus M-C, van Asperen R, Boon K, Voute PA, Heisterkamp S. 2001. The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science* **291**:1289–1292.
12. Dorogovtsev SN, Goltsev AV, Mendes JFF. 2002. Pseudofractal scale-free web. *Phys Rev E* **65**:066122.
13. Erdös P, Rényi A. 1960. On the evolution of random graphs. *Publ Math Inst Hung Acad Sci* **5**:17–61.
14. Faloutsos M, Faloutsos P, Faloutsos C. 1999. On power-law relationships of the internet topology. *Comput Commun Rev* **29**:251–262.
15. Fell D, Wagner A. 2000. The small world of metabolism. *Nature Biotech* **189**:1121–1122.
16. Flajolet M, Rotondo G, Daviet L, Bergametti F, Inchauspe G, Tiollais P, Transy C, Legrain P. 2000. A genomic approach to the hepatitis C virus. *Gene* **242**:369–379.
17. Gavin A, Bösche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick J, Michon A-M. 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**:141–147.
18. Girvan M, Newman M. 2002. Community structure in social and biological networks. *Proc Natl Acad Sci USA* **99**:7821–7826.
19. Hartwell LH, Hopfield JJ, Leibler S, Murray AW. 1999. From molecular to modular cell biology. *Nature* **402**:C47–C52.
20. Ho Y, Gruhler A, Heilbut A, Bader G, Moore L, Adams S-L, Millar A, Taylor P, Bennett K, Boutillier K. 2002. Systematic identification of protein complexes in saccharomyces cerevisiae by mass spectrometry. *Nature* **415**:180–183.
21. Holme P, Huss M, Jeong H. 2003. Subnetwork hierarchies in biochemical pathways. *Bioinformatics* **19**:532–538.
22. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci USA* **98**:4569–4574.

23. Ito T, Tashiro K, Muta S, Ozawa R, Chiba T, Nishizawa M, Yamamoto K, Kuhara S, Sakaki Y. 2000. Towards a protein–protein interaction map of the budding yeast: a comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proc Natl Acad Sci USA* **97**:1143–1147.
24. Jeong H, Mason S, Barabási A-L, Oltvai ZN. 2001. Lethality and centrality in protein networks. *Nature* **411**:41–42.
25. Jeong H, Oltvai ZN, Barabási A-L. 2003. Prediction of protein essentiality based on genomic data. *ComPlexUs* **1**:19–28.
26. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási A-L. 2000. The large-scale organization of metabolic networks. *Nature* **407**:651–654.
27. Jung S, Kim S, Kahng B. 2002. A geometric fractal growth model for scale free networks. *Phys Rev E* **65**:056101.
28. Karp PD, Riley M, Saier M, Paulsen I, Paley S, Pellegrini-Toole A. 2000. The EcoCyc and MetaCyc databases. *Nucleic Acids Res* **28**:56–59.
29. Kochen M, ed. 1989. *The small world*. Ablex, Norwood, NJ.
30. Kohn K. 1999. Molecular interaction map of mammalian cell-cycle control and dna repair systems. *Mol Biol Cell* **10**:2703–2734.
31. Lauffenburger D. 2000. Cell signaling pathways as control modules: complexity for simplicity. *Proc Natl Acad Sci USA* **97**:5031–5033.
32. Lawrence S, Giles CL. 1999. Accessibility of information on the web. *Nature* **400**:107–109.
33. Liljeros F, Edling C, Amaral L, Aberg Y. 2001. The web of human sexual contacts. *Nature* **411**:907–908.
34. McGraith S, Holtzman T, Moss B, Fields S. 2000. Genome-wide analysis of vaccinia virus protein–protein interactions. *Proc Natl Acad Sci USA* **97**:4879–4884.
35. Mewes H, Frishman D, Gruber C, Geier B, Haase D, Kaps A, Lemcke K, Mannhaupt G, Pfeiffer F. 2000. MIPS: a database for genomes and protein sequences. *Nucleic Acids Res* **28**:37–40.
36. Milgram S. 1967. The small-world problem. *Psych Today* **2**:60–67.
37. Newman M. 2001. The structure of scientific collaboration networks. *Proc Natl Acad Sci USA* **98**:404–409.
38. Oltvai ZN, Barabási A-L. 2002. Life's complexity pyramid. *Science* **298**:763–764.
39. Overbeek R, Larsen N, Pusch G, D'Souza M, Selkov Jr E, Kyrpides N, Fonstein M, Maltsev N, Selkov E. 2000. WIT: integrated system for high-throughput genome sequence analysis and metabolic reconstruction. *Nucleic Acids Res* **28**:123–125.
40. Pandey A, Mann M. 2000. Protemics to study genes and genomes. *Nature* **405**:837–846.
41. Park J, Lappe M, Teichmann A. 2001. Mapping protein family interations: intramolecular and intermolecular protein family interaction repertoires in the pdb and yeast. *J Mol Biol* **307**:929–938.
42. Pastor-Satorras R, Smith E, Solé R. 2002. *Evolving protein interaction networks through gene duplication*. Santa Fe Working Paper 02-02-008.
43. Pastor-Satorras R, Vespignani A. 2001. Epidemic spreading in scale-free networks. *Phys Rev Lett* **86**:3200–3203.
44. Podani J, Oltvai ZN, Jeong H, Tombor B, Barabási A-L, Szathmary E. 2001. Comparable system-level organization of archae and eukaryotes. *Nature Genet* **29**:54–56.
45. Rain J-C, Selig L, DeReuse H, Battaglia V, Reverdy C, Simon S, Lenzen G, Petel F, Wojcik J. 2001. The protein–protein interaction map of *Helicobacter pylori*. *Nature* **409**:211–215.
46. Ravasz E, Barabási A-L. 2002. Hierarchical organization in complex networks. *Phys Rev E* **67**:026122.
47. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási A-L. 2002. Hierarchical organization of modularity in metabolic networks. *Science* **297**:1551–1555.
48. Redner S. 1998. How popular is your paper? An empirical study of the citation distribution. *Eur Phys J* **B4**:131135.

49. Schwikowski B, Uetz P, Fields S. 2000. A network of protein–protein interactions in yeast. *Nature Biotech* **18**:1257–1261.
50. Shen-Orr S, Milo R, Mangan S, Alon U. 2002. Network motifs in the transcriptional regulation network of *E. coli*. *Nature Genet* **31**:64–68.
51. Solé R, Pastor-Satorras R, Smith E, Kepler T. 2002. A model of large-scale proteome evolution. *Adv Compl Sys* **5**:43–54.
52. Uetz P, Giot L, Cagney G, Mansfield T, Judson R, Knight J, Lockshorn D, Narayan V, Srinivasan M, Pochart P. 2000. A comprehensive analysis of protein–protein interactions of *Saccharomyces cerevisiae*. *Nature* **403**:623–627.
53. Vazquez A, Flammini A, Maritan A, Vespignani A. 2003. Modeling of protein interaction networks. *ComPlexUs* **1**:38–44.
54. Vogelstein B, Lane D, Levine A. 2000. Surfing the p53 network. *Nature* **408**:307–310.
55. Wagner A. 2000. Mutational robustness in genetic networks of yeast. *Nature Genet* **24**:355–361.
56. Wagner A. 2001. The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. *Mol Biol Evol* **18**:1283–1292.
57. Wagner A, Fell DA. 2001. The small world inside large metabolic networks. *Proc Roy Soc London Ser B* **268**:1803–1810.
58. Walhout A, Sordella R, Lu X, Hartley J, Temple G, Brasch M, Thierry-Mieg N, Vidal M. 2000. Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* **287**:116–122.
59. Wasserman S, Faust K. 1994. *Social network analysis: methods and applications*. Cambridge UP, Cambridge.
60. Watts DJ, Strogatz SH. 1998. Collective dynamics of small-world networks. *Nature* **393**:440–442.
61. Wolf Y, Karev G, Koonin E. 2002. Scale-free networks in biology: new insights into the fundamentals of evolution? *Bioessays* **24**:105–109.
62. Wuchty S. 2001. Scale-free behavior in protein domain networks. *Mol Biol Evol* **18**:1694–1702.
63. Wuchty S. 2002. Interaction and domain networks of yeast. *Proteomics* **2**:1715–1723.
64. Wuchty S. 2003. Small-worlds in RNA. *Nucleic Acids Res* **31**:1108–1117.
65. Xenarios I, Fernandez E, Salwinski L, Duan X, Thompson M, Marcotte E, Eisenberg D. 2001. DIP: the database of interacting proteins: 2001 update. *Nucleic Acids Res* **29**:239–241.
66. Yook SH, Oltvai ZN, Barabási A-L. 2003. Functional and topological characterization of protein–protein interaction networks. Submitted.

5

ROBUSTNESS IN BIOLOGICAL SYSTEMS: A PROVISIONAL TAXONOMY

David C. Krakauer

Santa Fe Institute, Hyde Park Road, Santa Fe, New Mexico

Biology is a domain of tension: on the one hand, biology is concerned with transformation and the generation of diversity; on the other, biology is concerned with the persistence of improbable structural regularities. The historical sciences in biology, principally evolution, have focused on change. The mechanistic sciences in biology, principally medicine, have focused on stability. Robustness, as a research program, aims to uncover those evolved mechanisms promoting the persistence of regularities. Here I organize mechanisms of robustness into a phenomenological taxonomy, grouping biological mechanisms into principles of robust organization. These include: Redundancy, Purging, Feedback, Modularity, Spatial Compartmentalization, Distributed Processing, and the Extended Phenotype. I present case studies in which mechanisms representative of each principle are described. These case studies serve to illustrate the ubiquity of specialized robustness mechanisms in all complex biosystems.

1. A FUNDAMENTAL BIOLOGICAL DICHOTOMY: ROBUSTNESS AND EVOLVABILITY

Biologists have been motivated by two fundamental sets of questions. One set is associated with the generation and maintenance of genotypic, phenotypic, and functional diversity. The second set is associated with genotypic, phenotypic, and functional invariance. Evolutionary theory, following Darwin (9,10),

Address correspondence to: David C. Krakauer, Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501 (krakauer@santafe.edu).

has been concerned largely with transformation—from one species into another with coincident suites of modified adaptive complexes. Mechanistic biology—to include physiology and cell biology—has focused on the mechanisms underlying robustness of genotype and phenotype. Thus not only are robustness and evolvability obverse trends in biological system mechanics, they are also represented by two largely independent research traditions: the historical sciences relying on comparative data and theory, and the ahistorical sciences relying on laboratory data and description.

This caricature of our predicament suggests that two quite distinct problems need to be overcome in order to develop unified theories of biosystems. One is to establish the utility of evolutionary thinking in mechanistic science, and the other is to impress the importance of robustness upon evolutionary theory. Such a project would go some way towards reintroducing the phenotype into evolutionary theory.

Much has been written on the subject of transformation. Population genetics is concerned with the study of changing gene frequencies through time (26). Quantitative genetics is concerned with the change in the mean and variance of phenotypes across generations (14). In neither case has it been possible to explicitly incorporate detailed mechanistic components of the phenotye into these models. A recent movement in this direction involves work on the *genotype-to-phenotype map* and the *representation problem* (62). The genotype–phenotype map describes the process of development required to decode a genome into a viable phenotype. The representation problem is concerned with the way in which the variational properties of the genome are dependent upon the precise manner in which phenotypes are encoded in genotypes. To put it another way, are all phenotypes equally accessible from a given genotype configuration, and if not, does this depend upon the way in which phenotypes are represented in genetic data structures? Assuming a fixed representation, are there some phenotypes that are unlikely to ever be realized even in the face of overwhelming selective advantage? If this is so, then these impediments to isotropic adaptive transformation are likely to be associated with just those mechanisms ensuring unity of type, stability of genomes across generations, and homeostatic stability of the phenotype.

One path through the labyrinth of biological robustness is to keep hold of two of Ariadne's threads: one connected to limits to evolvability and associated mechanisms limiting variation, and the other to mechanistic inquiries into homeostasis and the regulation of cellular and individual phenotype (17). Robustness thus relates to two critical properties of complex biosystems: the long-term limits to evolutionary change and the short-term persistence of system function. Put differently, robustness mechanisms are one of the bridges connecting the dynamics of ontogeny with the dynamics of phylogeny by limiting phenotypic variation and also providing some means of exploring alternative genotypes without compromising the phenotype.

2. GENOTYPIC VERSUS ENVIRONMENTAL VERSUS FUNCTIONAL ROBUSTNESS

When speaking of robustness it is worth bearing in mind the plethora of definitions the word attracts. For an extensive list see [www.discusss.santa.edu/robustness]. These are to some degree domain-specific. In ecology, stability or robustness is a measure of the preservation of species diversity upon species removal (38), or the permanence of a configuration when perturbing some variable of ecological interest. In medicine, robustness is associated with healing and compensation, neither of which imply a return to the original phenotype but rather a restoration of wild-type function (56). In linguistics, robustness relates to competence and comprehensibility despite incomplete information and ambiguity (27). Thus structural transformation is acceptable subject to information remaining decodable. In paleontology, robustness relates to the continuity of lineages across geological eras (13) and the persistence of lineages during mass extinction events. In metabolism, robustness relates to limited phenotypic variation across large changes in kinetic parameters (21,63). In cell biology, robustness can describe how cell fate decisions remain constant when transcription regulation is stochastic (25), or how conserved RNA secondary structures can remain resistant to point mutations (15).

In each of these cases robustness relates to either (1) non-detectable or minor modification in phenotype following a large perturbation to the genotype, (2) non-detectable or minor modification in phenotype following a large perturbation to the phenotype from the environment, or (3) non-detectable or minor modification in function following a large perturbation to the genotype or phenotype with or without a correlated change in the phenotype. The important distinction between genotypic and environmental robustness is that in the first case perturbations are inherited, whereas in the second case they are not. Functional robustness can be achieved through phenotypic invariance or phenotypic plasticity. In one case the phenotype resists perturbations, and in the second case the phenotype tracks perturbations. Genotypic and environmental robustness can be measured through the environmental (V_e) or mutational variance (V_m) of a trait, whereas functional robustness can be measured as the variance in geometric mean fitness. It is often the case that a single mechanism leads to all three forms of robustness, in which case we observe *congruence* (4) between the genotype and phenotype.

3. PRINCIPLES AND PARAMETERS OF ROBUST ORGANIZATION

In (31,32), Krakauer and Plotkin describe three *principles* that have arisen in an effort to understand the evolutionary response to mutations: the principle of canalization, the principle of neutrality, and the principle of redundancy. We

contrast these with the *parameters* of robustness—those mechanisms by which these principles are realized. The principles and parameters metaphor is derived from linguistics (7), where the principles are defined as the invariant properties of universal grammar and the parameters the local rules and practices of language. Here we extend these principles to include feedback, modularity, spatial compartmentalization, distributed processing, and the extended phenotype. Another way of thinking about the principles is as higher grades in a theoretical taxonomy of robustness. All mechanisms employing some form of redundancy are classed together, as are those employing modularity and so on. As we work down the classificatory tree of robustness, we eventually hit the unique mechanical instantiation that gives rise to robustness. Our classification is more Linnean than Darwinian, as we have no external principle with which to organize mechanism.

We give a brief introduction to each of these principles below, and subsequently go on to discuss in more detail a few models developed to address specific robustness mechanisms in biology.

3.1. Redundancy

A common means of identifying the function of a gene is to perform a knockout experiment, removing or silencing a gene early in development. By assaying the resultant phenotype, the putative function of the absent gene can be inferred. In many such experiments there is no scoreable phenotype: the knockout leaves the phenotype in the wild-type condition. Biologists refer to a gene x on a background y as functionally redundant (57). This is taken to mean that the target gene is one of at least two or more genes contributing to the phenotype epistatically (27). Removal of a redundant gene x leads to compensation by remaining members of a redundant set $\mathbf{y}$. Let $f(g)$ be the fitness of gene or genome g; then redundancy implies that $f(x,\mathbf{y}) = f(\mathbf{y})$. When $\mathbf{y}$ has a cardinality of one and $\mathbf{y} = x$, then functional redundancy reduces to the special case of a redundant copy of x. Redundancy as a principle is more general, and describes any case in which the mechanism of robustness is only operative upon perturbation. Hence redundancy is a variational property, not contributing to fitness directly, but indirectly operating at the population level. Individuals with a redundancy property are not fitter than those without, but those without will on occasion suffer the consequences.

True redundancy might be rarer than "artefactual" redundancy, or experimental neutrality, in which the effect of perturbation remains below an experimental detection limit (47). Assuming that we are able to detect small changes, the degree of redundancy describes the degree of correlation among genes contributing to a single function. Models of redundancy in biology tend to focus on the evolutionary preservation of redundant components, and hence employ

population genetics approaches. More recently differential equation-based models for the dynamics of regulatory systems following structural perturbation have also been explored (61).

3.2. Feedback Control

Elementary feedback control systems have three components: a *plant* (the system under control), a *sensor* (measuring the output of the plant) and a *controller* (generating the plants input) (12). A measure of performance is often the degree to which the output of a plant approximates some function of the input to the controller. In biology a plant could be RNA or protein concentration, protein kinase activation, immune effector cell abundance, or species abundance. Inputs in each of these cases would be transcription factors, protease concentrations, chemical agonists bound to receptors, antigen concentrations, and death rates. The controllers are more often than not aggregates of several mechanisms. Feedback is a mechanism of robustness as it enables plants to operate efficiently over a range of exogenous input values. The question remains as to whether the controller is robust to variations in the plant—does it provide *robust stability*? For example, in biology, can a single feedback controller regulate the concentrations of several different proteins?

The theoretical literature in linear feedback control is very well developed in engineering. Biology has borrowed extensively from this literature. Nonlinear feedback control is another issue, and there are few canonical models (1).

3.3. Modularity

Repeated representations of functionally distinct character complexes capable of recombination or shuffling is an example of modularity. Modularity aims to capture structures that balance autonomy and integration. Within a complex there is strong integration, whereas populations of complexes are only weakly coupled (55). This has also been called near decomposability (54). In genetics, modularity involves a minimum of pleiotropy, in which sets of genes contributing to one complex or trait (for example, organ system) make little contribution to other complexes or traits (16,49,50). These modular genetic systems are found in different genomic contexts performing a similar function. Of course, modularity can be defined at levels of organization above that of the gene (64)—e.g., the extent to which organs operate independently during homeostasis. The dissociability of modules provides one means of damage limitation through encapsulation.

There are no collectively agreed upon models for analyzing modularity in biosystems. To date quantitative genetics models have been used to explore the

limits to the evolution of modularity, and neural network models have been used to explore how modularity can lead to more efficient task management (6). For a comprehensive review of the recent literature in the area of modularity see (53).

3.4. Purging—Antiredundancy

Whereas redundancy buffers the effect of perturbation, purging acts in the opposite fashion—amplifying the effects of perturbation—so as to ensure the purity of a population (29,31). A gene x on a genetic background $\mathbf{y} + \mathbf{z}$ is functionally antiredundant when the target gene is one of at least two or more genes ($x + \mathbf{y}$) contributing to the phenotype epistatically, and, when removal of gene x leads to a greater perturbation in the presence of $\mathbf{y}$ than in the absence of $\mathbf{y}$: $f(x,\mathbf{y} + \mathbf{z}) \geq \mathbf{f(z)} >> \mathbf{f(y + z)}$.

Purging is only effective when individual replication rates are sufficiently large to tolerate the effects of removal of defective components. Thus apoptosis—programmed cell death—is a common strategy for eliminating cells upon damage to their genomes or upon infection, provided these cell types are capable of regeneration. Nerve cells and germ cells produce factors that strongly inhibit apoptosis (37), as removal in these cases has deleterious consequences. In case of severe infection it can make sense to purge nerve cells (28).

Recent models dealing with purging-type phenomena have involved stochastic models assuming finite populations. The key insight from the study of antiredundancy is the ability of particulate, hierarchical systems to exploit cellular turnover to eliminate and replace deleterious components from populations.

3.5. Spatial Compartmentalization

Compartmental systems are those comprised of a finite number of macroscopic subsystems called compartments, each of which is well mixed. Compartments interact through the exchange of material (22). The spatial compartmentalization of reactions leads to robustness by minimizing covariance among reaction components participating in functionally unrelated processes. Thus spatial de-correlation through compartmentalization substitutes for temporal correlation in biological functions. Robustness is achieved in at least two ways: (1) minimizing interference—chemical, epistatic, or physiological, and (2) minimizing mutual dependencies and thereby attenuating the propagation of error through a system. The study of spatial compartmentalization is particularly rich in theoretical ecology and epidemiology (34), where it has been used to explore the maintenance of antigenic diversity, restrictions on pathogen virulence, and seasonal forcing, and more recently in molecular biology, where proteins have been found to be compartmentalized (48).

From a modeling perspective, compartmentalization is often approached from the perspective of metapopulation dynamics or coupled oscillators, in which space is assumed to be discrete (implicit space) and non-local (19). An alternative approach is based on continuous space (explicit space) with local interactions and employs partial differential equations to study diffusion and advection of components (42). A third approach assumes discrete space with local interactions employing coupled map lattices and cellular automata. A fourth approach analyzes the statistical connectivity properties of undirected graphs and their response to node or edge elimination (2).

3.6. Distributed Processing

Distributed processing describes those cases in which an integrated set of functions are carried out by multiple, semiautonomous units (20,40). The most obvious example is that of nerve cells comprising the nervous system. Distributed processing, or connectionism, might be assumed to be a combination of modularity and spatial compartmentalization, but differs in that a single function is emergent from the collective activities of units, and correlated activity is thereby a desired outcome.

The robustness properties of connectionist models are the ability to (1) identify incomplete patterns, (2) generalize from a subset of learned patterns, and (3) degrade gracefully upon removal of individual nodes.

Connectionist models range from a simple application of linear algebra, dynamical systems, and Hamiltonian representations of steady states, through to the use of statistical mechanics models of frustrated systems such as spin glasses.

3.7. Extended Phenotypes

The extended phenotype concept was introduced by Dawkins (11) as a means of emancipating the gene from the discrete vehicle (often taken to be the individual organism). Thus while the gene's most proximal effect is to encode proteins, more distally, and as a byproduct, these participate in cells, tissues, organs, individuals, behaviors, mental states, and on through to cultures. There is no implication of determinism or strong causality in this statement. The extended phenotype notion merely recognizes that the boundary of physical embodiment need not represent the boundary of genic action.

In nonhuman biosystems the importance of the extended phenotype to robustness is not contested: from animal artifacts (ant nests, termite mounds, bird nests, and spider webs) and from animal behavior (policing, reconciliation, and dominance). In human society the issue is more controversial and the evidence

correspondingly weaker. However, it remains a fascinating question to pose: to what extent do human institutions represent instances of mechanisms for biological robustness? In the non-reductive (gene-independent) example of medical care and hospitals the case is obvious. There are indications, though, that behavioral rules, such as reciprocity and sharing, are to some extent causally related to the actions of our genes (8).

Modeling in this area tends to be either game theoretical (39) or some variant of population genetics to allow for both vertical and horizontal transmission. This is a nascent field for theory.

4. CASE STUDIES OF ROBUST PRINCIPLES

In the remainder of this chapter I have chosen case studies to illustrate the application of theory in the study of biological robustness. I have done so because as of yet there is no unified theory of biological robustness, only collections of illustrative models. These models vary in the degree to which they deal with robustness explicitly, and yet all them bear on the question in some fundamental way.

4.1. Redundancy in Genetic Networks

Wagner (60) has studied dynamical models for evolution of transcription regulation circuits. Gene duplication is thought of as a mutational event necessary to establish the genetic diversity for subsequent diversity in spatiotemporal patterning during development. Wagner poses this question: what is the average proportion of genes likely to be involved in a duplication event, such that the initial effect on the phenotype of duplication is minimized? In other words, what fraction of genes is capable of performing redundantly? This question can be inverted by asking how many genes from a portion of genome made up from duplicate sets can be deleted and made to preserve the same phenotype. In the first case the perturbation involves adding genes, and in the second eliminating genes. Wagner models the gene expression dynamics in much the same way connectionist modelers describe neural networks. The activity of gene i is denoted by S_i. The magnitude of transcriptional activation between gene i and gene j is given by weight matrix entry w_{ij}. The dynamics of gene expression in discrete time are

$$S_i(t+\tau) = \sigma\left[\sum_i^N w_{ij} S_j(t)\right] = \sigma[h_i(t)]\,. \qquad [1]$$

The function $\sigma[.]$ is the sign function. The output of interest is the steady-state levels of gene expression in the network $\bar{\mathbf{S}}$ as a function of the initial conditions of gene expression $\bar{\mathbf{S}}(0)$ and network connectivity. Whereas duplication (duplication function π) of one or more genes (k) creates a network in a higher dimensional state space, deletion (deletion function δ) creates a network in a lower dimensional state space:

$$\pi\colon \{-1,1\}^{N} \rightarrow \{-1,1\}^{N+k}, \qquad [2]$$

and hence

$$(S_1, ..., S_k, S_{k+1}, ..., S_N) \rightarrow (S_1, S_1, ..., S_k, S_k, S_{k+1}, ..., S_N) \qquad [3]$$

and for deletions

$$\delta\colon \{-1,1\}^{N+k} \rightarrow \{-1,1\}^{N}, \qquad [4]$$

and hence

$$(S_1, S_1, ..., S_k, S_k, S_{k+1}, ..., S_N) \rightarrow (S_1, ..., S_k, S_{k+1}, ..., S_N). \qquad [5]$$

Wagner compares the wild-type equilibrium states ($\bar{\mathbf{S}}$) and the state following duplication ($\pi\colon \bar{\mathbf{S}}$) using the Hamming distance between ($\bar{\mathbf{S}}$) and ($\pi\colon \bar{\mathbf{S}}$) as the robustness metric. It is observed that small duplications and large duplications have the least impact on phenotypic change. And hence small and large deletions are likely to have the least impact on phenotype. Intermediate sized duplications (around 40% of genes) have the greatest impact on phenotype. In a region of the genome made up from sets of duplicate genes, perturbations involving deletions of just under half of the genome are expected to have the greatest effect on the phenotype, whereas genotypes are expected to be robust against perturbations involving a few or almost all genes.

Redundancy in this model does not refer to the duplicate genes, but the phenotypic invariance relating to epistasis in the transcriptional network. The explanation for this result is fairly obvious. Duplicating all the genes leaves the network effectively unchanged. Small numbers of duplications proportionately influence a small number of connected pairs. Intermediate sized duplications are likely to be most disruptive.

4.2. Modularity in Genetic Regulatory Networks

In *Drosophila* the anterior–posterior body axis is segmented. Segmentation is initiated by maternal factors at the embryo stage. Those factors initiating seg-

mentation are expressed transiently, and it is left to a segment polarity network to maintain the definition of segment boundaries. Segment polarity networks abound in insect orders, whereas the patterns of stable segmentation are variable. Von Dassow et al. (58,59) suggest that the segment polarity network is a robust evolutionary module, recruited by different insect species, and provided with different inputs to produce diverse patterns of segmentation. In order for this to be the case, parametric variation in reaction coefficients should leave the patterning ability of the network intact.

In order to model the network, Von Dassow simulated large systems of coupled first-order differential equations. For example, the rate of transcription of mRNA M_i from gene E_i, assuming a concentration of binding transcription factor x_i, a maximum rate of transcription $T_{\max}$, and a rate of decay dm_i, is given by

$$\dot{m}_i = T_{\max}\left[\frac{x_i^c}{k^c + x^c}\right] - dm_i\,, \qquad [6]$$

where the parameter k determines the value at which the transcription factor X_i has half maximum effect on the rate of translation of the gene E_i. The subsequent translation of M_i into a protein P_i with a maximum rate of translations $r_{\max}$ and a rate of decay $d_P p_i$ is of the form

$$\dot{p}_i = r_{\max}\left[\frac{m_i}{m_i + k}\right] - d_P p_i\,. \qquad [7]$$

These proteins are then free to bind to other proteins forming complexes with novel transcription activity (e.g., a p_i might bind to a p_j to induce x_k, etc).

Equations of this form assume saturation of enzymes and substrates. As a consequence, over large variations in parameter values steady-state concentrations of protein products and complexes remain unchanged. Saturation is the assumption behind the derivation of the familiar Michelis-Menten rate law: the concentration of substrate is in large excess over the concentration of enzyme (66). In the limiting case of very high values of the constant c, coupled differential equations can be effectively replaced by Boolean networks. In this case, only the topology of the network and the initial conditions, not the kinetic constants, have an influence on steady states. Thus stable variation of segmentation in insect orders might be achieved through variation in initial conditions with disregard for variation in kinetic parameters. Species diversity would derive from feeding different initial conditions through the same network without regard for species-specific variation in rate constants. If saturation is not justified this robust modularity disappears. The empirical validity of saturation in developmental networks remains to be established.

4.3. Feedback Control in Immune Regulation and Signal Transduction

4.3.1. Segel and Bar-Or's Adaptive Control Model for Immune Effector Action

The immune system is configured so as to maximize damage to pathogens and minimize damage to *self*. These, however, are not orthogonal goals, and hence the regulation of infection by the immune system requires feedback control in order to prevent an overenthusiastic immune response from destroying healthy tissues.

Segel and Bar-Or (51,52) approach the problem as follows (see also this volume, Part III, chapter 4, by Segel). Assume a population of immune effector cells E, a population of pathogens P and a noxious chemical N. E are able to kill P, as is N. However, N can also damage the host and thereby compromise production of E. It is assumed that the immune system seeks to minimize damage to the host by maximizing the efficiency of the immune response. Damage to the host δ is calculated as the time averaged abundance of P and N, where damage from P occurs at a rate h_pP and damage from N at a rate h_NN. Thus

$$\delta = \frac{1}{T}\int_0^T [h_p P(t) + h_N N(t)]dt\,. \qquad [8]$$

Assuming the dynamical system:

$$\dot{N} = Se - g_N N, \qquad [9]$$

$$\dot{P} = rP - aEPN, \qquad [10]$$

$$\dot{E} = E[\mu_p P(1 - E/E_{\max}) - g_E]. \qquad [11]$$

where the crucial parameter, s, the secretion coefficient of noxious chemicals, in response to immune activation, is assumed to be under constitutive control by the host. The function $\delta(s)$ has a unique minimum for any given value of the pathogen proliferation coefficient r, moreover $d(\delta(s))/dr > 0$.

The problem for feedback control is to determine the optimal value of s for a variety of pathogens with different proliferation rates. Segel and Bar-Or suggest one way, which requires that the host employ two performance measures: a *kill indicator* chemical, K, produced in response to immune activity NPE, and a *harm indicator* chemical produced in response to instantaneous damage—$h_pP + h_NN$. Include these two chemicals in the dynamical system:

$$\dot{K} = c_k(aEPN) - g_k K, \qquad [12]$$

$$\dot{H} = c_h(h_p P + h_N N) - g_h H. \quad [13]$$

Now harm from the pathogen (H_p) is not the same as harm inflicted indirectly through the immune response (H_I). Estimate $H_p = H/(1 + k_p N)$ and assuming that $H = H_I + H_p$, then an adaptive s coefficient might change according to the Michaelean rate law:

$$s = s_1 + \frac{s_2 K H_p}{1 + s_3 H_1 + s_4 K H_p}. \quad [14]$$

An immune response making use of multiple sources of feedback information can operate effectively over a far greater range of parameter values and variable values than one without. This form of robustness through feedback control is typical of biological systems.

One caveat to be observed at this point regards the arbitrary nature of the functional response curves assumed in this model and in others like it. In other words, constant non-saturating rates of immune effector proliferation and pathogen replication. To what extent is feedback destabilized by increasing nonlinearities in response functions? The purpose of these models is often concerned with "proof of principle," establishing the plausibility of intuitive notions of control, rather than empirical fitting of experimental data.

4.3.2. Barkai and Leibler's Chemotaxis Network

Feedback is no less important in regulating reactions within a cell as among populations of cells. As with variation in pathogen parameters in populations, there can as easily be variation in inputs to a cell. This means that fine-tuning parameters in advance (through evolution) to maximize a function for fixed parameters is likely to be far from robust.

Chemotaxis in bacteria describes the purposeful motion of bacteria swimming towards increasing concentrations of nutritive chemicals. Bacteria swim in alternating bouts of *smooth runs*, during which they move along a single vector, and *tumbling*, during which they randomly reorient to a new vector. An observed property of bacterial chemotaxis is *adaptation*, whereby the steady-state tumbling frequency in a homogeneous chemical environment is independent of the concentration of chemical. This is a means of ensuring constant responsiveness (3). Barkai and Leibler (5) ask whether feedback circuits in the putative chemotactic network are responsible for this adaptive property.

A nutritive chemical, or *ligand L*, binds to an enzymatic receptor *E*. The receptor transitions between a modified and unmodified state at a rate proportional to the concentration of *L*, denoted *l*. *L* represents the input to a cellular

signal transduction system, and the concentration of active enzyme (A) interfacing with the propulsive flagellum is the system output. An adaptive system has the characteristic that the steady-state concentration of $A(\overline{a})$ is independent of l.

The key to the robust adaptive property is to make the modification and unmodification transformation of E dependent only on the concentration of A. Yi et al. (65) point out that this adaptive property of the network is a consequence of *integral feedback control*. In mathematical terms:

$$\dot{x} = a, \qquad [15]$$

$$a = a_1 - \overline{a} = k(l - x) - \overline{a}. \qquad [16]$$

Here the time integral of the system error (x), the difference between the actual output (a_1) and the desired equilibrium output $(\overline{a})$, is fed back into the system. The parameter k is the gain of the system. In this way one obtains robust asymptotic tracking of variations in input l.

4.4. Antiredundancy through Apoptosis in Neoplastic Lineages

Tumorigenesis marks the onset of unregulated cell proliferation. In most long-lived mammals, progress towards tumorigenesis involves the cumulative loss of important regulatory genes monitoring the genetic state of defective cells. An important class of regulatory genes are the *tumor suppressor genes* (35,36), which respond to mutations by inducing programmed cell death (apoptosis) or repairing damaged DNA. Apoptosis represents a strategy of antiredundancy or purging, in which defective cells are removed and subsequently replaced by the descendants of healthy cells in the surrounding tissue. Purging as a mechanism of robustness thus depends crucially on population sizes large enough to allow for the replacement of eliminated cells.

Plotkin and Nowak (46) have modeled the waiting time for dividing cells undergoing mutation and mutation-induced apoptosis to reach a tumorigenic state. Assume that L genes in the genome of dividing cells regulate healthy cell cycle function. For each cell, count the number of mutations in L and call it k. When the value of $k = n$, the cell is tumorigenic. During each cell division a cell with k mutations can divide and remain in the same state with a probability q_k or mutate with a probability $p_k = 1 - q_k$. Any cell with $k \geq 1$ mutations is under the surveillance of tumor suppressor genes and can be induced into apoptosis with a probability α_k. Apoptosis will fail with a probability $\beta_k = 1 - \alpha_k$.

These probabilities can be used to construct a Markovian model of cancer progression, with three important assumptions: (1) there are no population dynamics—cell populations are of a large fixed size with no fixation of mutant lineages, (2) there are symmetric mutations such that only the total number of

mutations k and not the position of these mutations in a string of length L is significant, (3) the cell with n mutations is an absorbing state. With these assumptions one can write down an $(n + 1)$ x $(n + 1)$ transition matrix:

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \\ n-1 \\ n \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & \cdots & n-1 & n \end{array} \\ \begin{pmatrix} q_0 & p_0 & 0 & \cdots & \cdots & 0 \\ \alpha_1 & \beta_1 q_1 & \beta_1 p_1 & \cdots & & 0 \\ \alpha_2 & 0 & \beta_2 q_2 & \beta_2 p_2 & & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ \alpha_{n-1} & 0 & \cdots & 0 & \beta_{n-1} q_{n-1} & \beta_n q_n \\ 0 & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix} \end{array}$$

This is a flexible formulation, as it allows for either *genomic instability*, in which $\alpha_0 > \alpha_1 > ... > \alpha_{n-1}$, which describes how the incidence of mutations reduces the efficacy of the apoptotic response, or when $\alpha_0 < \alpha_1 < ... < \alpha_{n-1}$, which reflects an increasing probability of cells with more mutations undergoing effective surveillance. I will only discuss the case in which $q = q_1 = q_2 = ... = q_{n-1}$ and $\alpha = \alpha_1 = \alpha_2 = ... = \alpha_{n-1}$.

The effects of apoptotic purging can be demonstrated by comparing the waiting time for $k = n$ of a non-apoptotic cell, assuming thereby that $\alpha_i = 0$ for all i, and the alternative case with apoptosis as described above in which $\alpha_i > 0$ for all i.

The waiting without apoptosis for one cell in a tissue of N cells to obtain n mutations is given by

$$T = \frac{1}{\log(1/q)(n-1)!^N} \int_0^\infty \Gamma(n,a)^N \, da \,, \qquad [17]$$

where $\Gamma(.,.)$ is the incomplete Gamma function. The waiting time for a single cell with apoptosis to obtain n mutations is given by

$$T = \frac{p\beta(\alpha + p_o)}{\alpha p_0 (\alpha q + p)(1 - \alpha/(p + \alpha q))^n} - \frac{1}{\alpha} \,. \qquad [18]$$

In the case without apoptosis, the waiting time depends inversely on the logarithm of replication fidelity q. With apoptosis the waiting time grows exponentially with n. Thus purging of damaged cells prolongs the waiting time to tumorigenesis, and thereby increases the latency of cancer.

4.5. Spatial Compartmentalization of Predators and Prey: Infectious Disease

Theoretical immunology is in large part based on reinterpretation of the immune system as an interaction between predators and prey. Whereas in ecol-

ogy these might be carnivores and herbivores, in immunology these might be cytotoxic T cells and virus-infected cells. Immune effector cells proliferate in response to antigen presented by infected cells, in which the rate of proliferation is likely to be proportional to the number of infected cells presenting antigen. The destruction of infected cells brings about a concomitant reduction in effector cell proliferation. We therefore expect oscillatory dynamics. In ecology one of the principal measures of population stability is the variance in species abundance. Large-amplitude oscillations are thought to make populations vulnerable, whereas low-amplitude oscillations are a sign of robustness (38). In ecology species extinction is at stake, in immunology a loss of effector cells and a loss of regulatory control are at stake.

Jansen and de Roos (23) studied the following two-compartment model. Consider two populations of predators, P_1 and P_2, and prey populations, N_1 and N_2. Predators are able to migrate from one compartment to another with probability $m/2$:

$$\dot{n}_1 = rn_1 - n_1 p_1 , \qquad [19]$$

$$\dot{p}_1 = n_1 p_1 - \mu p_1 + \frac{m}{2}(p_2 - p_1) , \qquad [20]$$

$$\dot{n} = rn_2 - n_2 p_2 , \qquad [21]$$

$$\dot{p}_1 = n_2 p_2 - \mu p_2 + \frac{m}{2}(p_1 - p_2) . \qquad [22]$$

The rate of predator proliferation is given by rp_i and the death rate μp_i. From an immunological perspective we might think of two strains of infecting virus and their corresponding T cell receptors.

Assuming equal densities of predators ($p_1 = p_2$) and prey ($n_1 = n_2$), the model reduces to the non-spatial Lotka-Volterra model in which densities oscillate permanently at an amplitude determined by the initial conditions. However, if small differences in density are allowed between compartments, these transiently increase with a correlated reduction in the amplitude of oscillations in the average densities. This is because in the compartmental model large-amplitude oscillations are diffusively unstable (statistical stabilization), whereas in the single-population model oscillations of any amplitude can be maintained. Thus, establishing compartments in which pathogens will be attacked (such as lymph nodes), rather than fostering the likeness of a single population, should allow organisms to limit variation in pathogen densities.

4.6. Distributed Processing in the Nervous System

The connectionist modeling paradigm has become the dominant theoretical framework for thinking about information processing by the nervous system (20,41). While the mapping from neural network to neural systems is highly approximate, the objective in connectionist models is to explore the properties and limits of a "*gedankenexperiment*" in which information is distributed over a population of homogeneous, computationally trivial units. Out of this research have arisen the following robustness observations: (1) pattern recognition of corrupted inputs, (2) categorization or generalization of noisy inputs, and (3) graceful degradation in response to graded perturbations in network input or network structure. There is some sense that network models are intrinsically fault-tolerant as a result of the distributed nature of the information representation. The aforementioned principles of redundancy and modularity are likely to participate in connectionist robustness but do not exactly capture the distributed nature of the information in a neural network model.

The canonical representation of a feedforward neural network is

$$S_i = f\left(\sum_j w_{ij} S_j - \theta_i\right), \qquad [23]$$

where S_i is the output of unit i, w_{ij} are the weights from unit j to unit i, and θ_i is the activation threshold of unit i. The function $f(.)$ is most often of the form of a nonlinear squashing function or a step-function. The robustness of a network can be assessed as the deviation of the actual output vector (**S**) from an desired output vector (**O**). A common metric is the RMS error:

$$\varepsilon = \sqrt{\frac{1}{N}\sum_i^N (S_i - O_i)^2}\ . \qquad [24]$$

Perturbations in S_j or w_{ij} can then be assessed quantitatively. An alternative error function for binary or "bipolar" units is to use the Hamming distance between **S** and **O**.

4.6.1. Joanisse and Seidenberg on Verb Morphology

There has been some debate on whether brain-injured patients have a greater difficulty in constructing the irregular past tense of familiar verbs or the regular past tense of nonsense (nonce) words. The impairment has been used to discriminate between damage to rule-following (regular) versus damage to associative memories (irregular). The construction of the past tense has become a

paradigmatic linguistic system for studying the difference between look-up tables for exceptions and rules for common verbs (64). Joanisse and Seidenberg (24) constructed a simple neural network model in which output units represent a sequence of phonological features—ordered lists of vowels (V) and consonants (C). Thus, each word can be aligned with a basic template: CCVVCCC–VC. Tasted might read "C0V0CC0VC," in which "0" denotes wildcards or empty slots not filled by a given verb.

Each verb is represented by a unique hidden unit in the network. In addition, the network contains semantic units to render verb meanings. Input units encode basic phonology as with the output units. Thus, inputs connect to hidden units that connect to output units. Semantic units also connect to hidden units recurrently. One of the tasks of the network is to take a phonological input and a tense marker and generate an identical output (autoassociative mapping), while another required a semantic input to be mapped onto an appropriate phonological unit.

Perturbations to the network involved severing a proportion of connections or by adding Gaussian noise to semantic units or phonological units. "Lesions" to 5% or less of the connections had almost no effect on performance (as measured by proportion of correct outputs given a target vector—the Hamming metric). Perturbations of over 5% and higher led to a roughly linear reduction in system performance. Perturbations to the phonological units tended to produce "irregularization" errors, whereas damage to the semantic units tended to produce regularization of irregular verbs.

Thus, this network was able to preserve its basic function over a small range of perturbations, above which it degraded gracefully. This linear reduction in system performance is a result of the distributed nature of the computation. Moreover, the way in which the model lost robustness reflected, in some way, the pattern of language deficit observed in Alzheimer's or Parkinsonian patients.

4.7. The Extended Phenotype of Human Culture

The derivation of human culture from genetic processes remains a controversial and often poorly posed enterprise. However, it is possible to ask whether there are universal tendencies among human populations to institutionalize rules that minimize the impact of perturbations. In other words, are there rules, norms, and procedures that serve to make human populations more robust? The mathematical study of the stability of human culture to social perturbations is in the domain of game theory.

One area in which the human species has been stated to be unique is in the possession of arbitrary symbols combined with a combinatorial grammar. An essential early step in the evolution of language has been the evolution of phonological rules, in which phonemes are combined into words. Why should this transition take place? Why use compositional signals rather than expanding the number of phonemes? Nowak and Krakauer (27) as well as Nowak, Krakauer,

and Dress (43) present a simple formalism of this problem and demonstrate that one important selection pressure in favor of compositional signals is a need to become robust against errors in signal perception (18). As a result of space limitations, I shall only demonstrate the nature of the signalling problem, and omit the full solution.

Assume that a language L employs n signals to communicate about n objects. When two individuals communicate, they obtain a payoff:

$$F = \sum_{i=1}^{n} a_i \,. \qquad [25]$$

If all objects have the same value, then the total payoff is simply $F = kn$. In reality, communication is error prone. Denote the probability of mistaking a signal i for a signal j u_{ij}. The error matrix $\mathbf{U}$ is a row stochastic error matrix. The diagonal values u_{ii} give the probability of correct communication. Hence,

$$F = \sum_{i=1}^{n} a_i u_{ii} \,. \qquad [26]$$

The error matrix can be defined in terms of similarity between any two signals i and j: s_{ij}. Similarity is a value between 0 and 1, and hence $u_{ij} = s_{ij} / \sum_{k=1}^{n} s_{ik}$. This enables us to write the payoff in terms of signal similarity:

$$F = \sum_{i=1}^{n} \left(\frac{a_i}{\sum_{j=1}^{n} s_{ij}} \right). \qquad [27]$$

Signals are embedded in some metric space X and d_{ij} denotes the distance between i and j. Assume that similarity is a monotonically decreasing function of distance, $s_{ij} = f(d_{ij})$. One choice of function is $s_{ij} = \exp(-\alpha d_{ij})$, where the parameter α is a measure of the resolution of perception.

For a given number of objects we wish to find the optimum configuration of sounds $x_1, ..., x_n$ in a sound continuum that maximize the payoff function:

$$F = \sum_{i=1}^{n} \left(\frac{1}{\sum_{j=1}^{n} \exp(-\alpha \mid x_i - x_j \mid)} \right). \qquad [28]$$

It can be proved that the maximum value of F, as n tends to infinity, converges to

$$F_{\max} = 1 + \alpha/2. \quad [29]$$

For any given value of perceptual accuracy α, the payoff converges to a maximum as a result of perceptual error. Increasing the number of signals increases the number of objects that can be communicated about, but at the cost of increased ambiguity. We have called this the *linguistic error limit*. It is our hypothesis that phonology, word formation, and simple grammar evolved through a need for greater robustness in response to inevitable errors of communication. The key to understanding how this works is to think in terms of composite words, in which a word W_{ij} consists of phonemes i and j. The similarity between words W_{ij} and W_{kl} is given by $s_{ik}s_{jl}$. The payoff to a language that contains n^2 words to describe n^2 objects is

$$F = \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\frac{1}{\sum_{j=1}^{n} \sum_{l=1}^{n} s_{ik} s_{jl}} \right)$$

$$= \left[\sum_{i=1}^{n} \left(\frac{1}{\sum_{j=1}^{n} s_{ij}} \right) \right]^2, \quad [30]$$

and for words of length L

$$= \left[\sum_{i=1}^{n} \left(\frac{1}{\sum_{j=1}^{n} s_{ij}} \right) \right]^L. \quad [31]$$

Hence, the total payoff can now grow exponentially with the length of words. Words, according to this formulation, are a cultural robustness mechanism.

5. AWAITING A SYNTHESIS OF ROBUSTNESS IN BIOLOGICAL SYSTEMS

I have presented a superficial overview of various research projects aimed at understanding robustness in biological systems. I have tried to organize this work into a number of principles of robustness—a theoretical taxonomy—so that common patterns and mechanisms might become apparent to the reader. It is unfortunate that there does not exist a single theory of biological robustness that might be applied to these several different problems. The historical, and to

some extent contingent, nature of biological organization is in large part responsible for this theoretical deficit.

There are, however, glimpses of intersection among principles—redundancy, modularity, spatial compartmentalization, and distributed processing share the use of a multiplicity of self-contained units discretely connected, to ensure a degree of autonomy of processing. The feedback control, the developmental module, and the connectionist model all exploit saturation effects to damp the consequences of nonlinearity. Almost all the models assume some form of sparse connectivity, whether it be among neurons, classes of mutation, modules, signaling molecules, or immune effectors. From this perspective it can be observed that robustness is a property of a large class of complex systems, and that a general theory might be expressed in some abstract terms that transcend system particulars. With a strong general theory of robustness we might start exporting insights from evolved systems to engineered systems, where robustness is frequently minimal.

In a work in progress, we (33) have been developing a general theory of robustness for energetic and informational flows over biological networks. This work stresses the mechanisms by which networks, once perturbed, reconfigure to alternative, redundant input sources so as to continue operating. This work has also highlighted the fact that a satisfactory definition of robustness requires some means of excluding inert or linear aggregates (systems of noninteracting components) as robust. In other words, robustness theories need to include some measure of contributions from both network flows (the system is in some sense functional) and from invariance upon elimination of a subset of flows (the function can be retained upon perturbation). Robustness is a variational property of evolved/engineered systems, and if we do not assume this dual measure approach (flows and invariance), rocks and dead organisms strike us as rather too robust upon perturbation!

6. REFERENCES

1. Aeyels D, Lamnabhi-Lagarrigue F, Schaft AJvd, Nonlinear Control Network. 1999. *Stability and stabilization of nonlinear systems*. Springer, London.
2. Albert R, Jeong H, Barabasi AL. 2000. Error and attack tolerance of complex networks. *Nature* **406**:378–382.
3. Alon U, Surette MG, Barkai N, Leibler S. 1999. Robustness in bacterial chemotaxis. *Nature* **397**:168–171.
4. Ancel LW, Fontana W. 2000. Plasticity, evolvability, and modularity in RNA. *J Exp Zool* **288**:242–283.
5. Barkai N, Leibler S. 1997. Robustness in simple biochemical networks. *Nature* **387**:913–917.
6. Calabretta R, Nolfi S, Parisi D, Wagner GP. 1998. Emergence of functional modularity in robots. In *From animals to animats*, pp. 497–504. Ed. R Pfeifer, B Blumberg, J-A Meyer, SW Wilson. MIT Press, Cambridge.

7. Chomsky N. 1981. Principles and parameters in syntactic theory. In *Explanations in linguistics*, pp. 32–75. Ed. N Hornstein, D Lightfoot. Longman, London.
8. Constantino JN, Todd RD. 2000. Genetic structure of reciprocal social behavior. *Am J Psychiatry* **157**:2043–2045.
9. Darwin C. 1859. *On the origin of the species by means of natural selection, or, the preservation of favoured races in the struggle for life*. J. Murray, London.
10. Darwin C. 1874. *The descent of man, and selection in relation to sex*. A.L. Burt, New York.
11. Dawkins R. 1982. *The extended phenotype: the gene as the unit of selection*. Freeman, San Francisco.
12. Emanuel P, Leff E. 1979. *Introduction to feedback control systems*. McGraw-Hill, New York.
13. Erwin DH. 2001. Lessons from the past: biotic recoveries from mass extinctions. *Proc Natl Acad Sci USA* **98**:5399–5403.
14. Falconer DS, Mackay TFC. 1996. *Introduction to quantitative genetics*. Essex, England, Longman.
15. Fontana W. 2002. Modelling "evo-devo" with RNA. *Bioessays* **24**:1164–1177.
16. Goldberg E. 1995. Rise and fall of modular orthodoxy. *J Clin Exp Neuropsychol* **17**:193–208.
17. Gould SJ. 2002. *The structure of evolutionary theory*. Harvard UP, Cambridge.
18. Grassly NC, Von Haeseler A, Krakauer DC. 2000. Error, population structure and the origin of diverse sign systems. *J Theor Biol* **206**:369–378.
19. Hanski I. 2001. Spatially realistic theory of metapopulation ecology. *Naturwissenschaften* **88**:372–381.
20. Hertz J, Krogh A, Palmer RG. 1991. *Introduction to the theory of neural computation*. Addison-Wesley, Redwood City, CA.
21. Hurst LD, Randerson JP. 2000. Dosage, deletions and dominance: simple models of the evolution of gene expression. *J Theor Biol* **205**:641–647.
22. Jacquez JA. 1985. *Compartmental analysis in biology and medicine*. U Michigan P, Ann Arbor.
23. Jansen VAA, de Roos A. 2000. The role of space in reducing predator–prey cycles. In *The geometry of ecological interactions*, pp. 183–200. Ed. U Dieckmann, R Law, JAJ Metz. Cambridge UP, Cambridge.
24. Joanisse MF, Seidenberg MS. 1999. Impairments in verb morphology after brain injury: a connectionist model. *Proc Natl Acad Sci USA* **96**:7592–7597.
25. Kepler TB, Elston TC. 2001. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys J* **81**:3116–3136.
26. Kimura M, Ohta T, Aoki K. 1985. *Population genetics and molecular evolution: papers marking the sixtieth birthday of Motoo Kimura*. Japan Scientific Societies Press, Tokyo; Springer, New York.
27. Krakauer DC, Nowak MA. 1999. Evolutionary preservation of redundant duplicated genes. *Semin Cell Dev Biol* **10**:555–559.
28. Krakauer DC. 2000. Evolving cell death in the virus-infected nervous system. *Trends Neurosci* **23**:611–612.
29. Krakauer DC, Sasaki A. 2002. Noisy clues to the origin of life. *Proc R Soc Lond B Biol Sci* **269**:2423–2428.
30. Krakauer DC, Page KM, Sealfon S. 2002. Module dynamics of the GnRH signal transduction network. *J Theor Biol* **218**:457–470.
31. Krakauer DC, Plotkin JB. 2002. Redundancy, antiredundancy, and the robustness of genomes. *Proc Natl Acad Sci USA* **99**:1405–1409.
32. Krakauer DC, Plotkin JB. 2004. Principles and parameters of molecular robustness. In *Robust design: a repertoire for biology, ecology and engineering*, pp. 71–103. Ed. E Jen. Oxford UP, Oxford.
33. Ay N, Krakauer DC. 2005. *Information geometry of robust biological networks*. SFI Preprint.

34. Levin SA, Grenfell B, Hastings A, Perelson AS. 1997. Mathematical and computational challenges in population biology and ecosystems science. *Science* **275**:334–343.
35. Levine AJ. 1993. The tumor suppressor genes. *Annu Rev Biochem* **62**:623–651.
36. Levine AJ. 1997. p53, the cellular gatekeeper for growth and division. *Cell* **88**:323–331.
37. Matsumoto K, Nakayama T, Sakai H, Tanemura K, Osuga H, Sato E, Ikeda JE. 1999. Neuronal apoptosis inhibitory protein (NAIP) may enhance the survival of granulosa cells thus indirectly affecting oocyte survival. *Mol Reprod Dev* **54**:103–111.
38. May RM. 1973. Stability and complexity in model ecosystems. *Monogr Popul Biol* **6**:1–235.
39. Maynard Smith J. 1982. *Evolution and the theory of games*. Cambridge UP, Cambridge.
40. McClelland JL, Artificial Intelligence and Psychology Project. 1988. *Parallel distributed processing: implications for cognition and development*. Departments of Computer Science and Psychology, Carnegie Mellon University, Learning Research and Development Center, University of Pittsburgh.
41. McClelland JL, Rumelhart DE. 1988. *Explorations in parallel distributed processing: a handbook of models, programs, and exercises*. MIT Press, Cambridge.
42. Murray JD, Stanley EA, Brown DL. 1986. On the spatial spread of rabies among foxes. *Proc R Soc Lond B Biol Sci* **229**:111–150.
43. Nowak MA, Krakauer DC, Dress A. 1999. An error limit for the evolution of language. *Proc R Soc Lond B Biol Sci* **266**:2131–2136.
44. Nowak MA, Plotkin JB, Krakauer DC. 1999. The evolutionary language game. *J Theor Biol* **200**:147–162.
45. Nowak MA, Krakauer DC. 1999. The evolution of language. *Proc Natl Acad Sci USA* **96**:8028–8033.
46. Plotkin JB, Nowak MA. 2002. The different effects of apoptosis and DNA repair on tumorigenesis. *J Theor Biol* **214**:453–467.
47. Ponte E, Bracco E, Faix J, Bozzaro S. 1998. Detection of subtle phenotypes: the case of the cell adhesion molecule csA in *Dictyostelium*. *Proc Natl Acad Sci USA* 95:9360–9365.
48. Prior IA, Hancock JF. 2001. Compartmentalization of Ras proteins. *J Cell Sci* **114**:1603–1608.
49. Raff RA, Sly BJ. 2000. Modularity and dissociation in the evolution of gene expression territories in development. *Evol Dev* **2**:102–113.
50. Raff EC, Raff RA. 2000. Dissociability, modularity, evolvability. *Evol Dev* **2**:235–237.
51. Segel LA, Bar-Or RL. 1999. On the role of feedback in promoting conflicting goals of the adaptive immune system. *J Immunol* **163**:1342–1349.
52. Segel LA. 2001. Controlling the immune system: diffuse feedback via a diffuse informational network. *Novartis Found Symp* **239**:31–40, discussion 40–51.
53. Schlosser G, Wagner GP. 2004. *Modularity in development and evolution*. Chicago UP, Chicago.
54. Simon H. 1962 The architecture of complexity. *Proc Am Phil Soc* **106**:467–482.
55. Slotine JJ, Lohmiller W. 2001. Modularity, evolution, and the binding problem: a view from stability theory. *Neural Netw* **14**:137–145.
56. Stearns S. 1998. *Evolution in health and disease*. Oxford UP, Oxford.
57. Tautz D. 1992. Redundancies, development and the flow of information. *Bioessays* **14**:263–266.
58. Von Dassow G, Meir E, Munro EM, Odell GM. 2000. The segment polarity network is a robust developmental module. *Nature* **406**:188–192.
59. Von Dassow G, Odell GM. 2002. Design and constraints of the Drosophila segment polarity module: robust spatial patterning emerges from intertwined cell state switches. *J Exp Zool* **294**:179–215.
60. Wagner A. 1994. Evolution of gene networks by gene duplications: a mathematical model and its implications on genome organization. *Proc Natl Acad Sci USA* **91**:4387–4391.
61. Wagner A. 1996. Genetic redundancy caused by gene duplications and its evolution in networks of transcriptional regulators. *Biol Cybern* **74**:557–567.

62. Wagner GP, Altenberg L. 1996. Complex adapatations and the evolution of evolvability. *Evolution* **50**:967–976.
63. Westerhoff HV, Groen AK, Wanders RJ. 1984. Modern theories of metabolic control and their applications (review). *Biosci Rep* **4**:1–22.
64. Winther RG. 2001. Varieties of modules: kinds, levels, origins, and behaviors. *J Exp Zool* **291**:116–129.
65. Yi TM, Huang Y, Simon MI, Doyle J. 2000. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc Natl Acad Sci USA* **97**:4649–4653.
66. Jordan PC. 1979. *Chemical kinetics and transport*. Plenum Press, New York.
67. Pinker S. 1999. *Words and rules*. Perseus Publishing, Reading, MA.

Part III

COMPLEX ADAPTIVE BIOSYSTEMS: A MULTI-SCALED APPROACH

1.1

NOISE IN GENE REGULATORY NETWORKS

Juan M. Pedraza and Alexander van Oudenaarden

Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts

Gene expression is based on biochemical processes that are inherently stochastic. The resulting fluctuations in mRNA and protein levels can sometimes be exploited but generally need to be controlled for reliable function of regulatory networks. From models of these biochemical processes it is possible to obtain analytical expressions for the stochastic properties of the resulting distributions of expression levels. We present a review of the two main analytical techniques for modeling stochastic gene expression.

1. INTRODUCTION

Noise is often perceived as being undesirable and unpredictable; however, living systems are inherently noisy, and are optimized to function in the presence of fluctuations. Biochemical and genetic pathways are sensitive to noise: some organisms can exploit fluctuations to introduce diversity into a population, as occurs with the lysis-lysogeny bifurcation in phage λ (1,2) or phase variation in bacteria (3). In contrast, stability against fluctuations is essential for a gene regulatory cascade controlling cell differentiation in a developing embryo (4). Robustness to noise can be part of the function of a given network architecture, since structures like feedback loops can be used to reduce noise (5) (see also preceding chapter 5, Part II, by Krakauer). Stochastic fluctuations in gene

Address correspondence to: Juan M. Pedraza, van Oudenaarden Biophysics Laboratory, Department of Physics, Room 13-2042, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307 (juan@mit.edu).

expression lead to non-genetic individuality: even in the case when two individuals are genetically identical, protein concentrations between the two individuals can vary significantly because of the stochastic nature of protein synthesis. Pioneering experiments by Ko et al. (6) demonstrated that gene expression levels can vary significantly from cell to cell in an isogenic population. It is thought that the main source of the noise in gene expression arises from statistical fluctuations in the concentrations of mRNAs, transcriptional and translational machinery, and regulatory proteins (see also related chapters 1.2 (by Hofacker and Stadler) and 1.3 (by Wagner), Part III, this volume). In a single cell these concentrations can fluctuate widely since many of the molecules are present at low numbers. For example the lactose repressor protein in *Escherichia coli* is present at only 30 copies on average. Several stochastic models (7–13) have been proposed that recently have been tested experimentally (14–17) (see (18) for an excellent review). The goal of this chapter is to review two powerful analytical modeling techniques that can be used to determine the noise characteristics of genetic networks: (i) the master equation approach and (ii) the Langevin approach. These approaches will be applied to calculate the statistical properties of noisy genetic circuits. A more general treatment of master equations and the Langevin technique can be found in van Kampen (19).

2. THE MASTER EQUATION APPROACH

The master equation corresponds to the statement that the probability of being in a given state changes depending on the probabilities of transition to and from any other state in the system. It provides the full probability distribution when it can be directly solved. Unfortunately, this is not often the case, so we must settle for some of the moments of the distribution. These are easily obtained from the generating function, so we will work with the master equation in a form in which it depends on the generating function rather than the distribution.

The genetic network is defined by N state variables $n_1 \ldots n_N$ and M rate constants $k_1 \ldots k_M$. The variables denote the number of copies of a certain chemical species such as mRNAs or proteins. Before applying the master equation approach to determine the noise properties of a genetic network we will start by obtaining the master equation in the generating function form for some elementary chemical equations.

2.1. Synthesis from a Template

In numerous genetic reactions, such as transcription and translation, mRNAs and proteins are synthesized from a template (DNA and mRNA, respec-

tively). After synthesis the number of templates is not changed. The corresponding reaction is therefore:

$$A \xrightarrow{k} A + B\,. \qquad [1]$$

Molecule A produces molecule B at rate k (in units of (concentration $\times$ time)$^{-1}$). The master equation describes how the probability to be in state $[n_1, n_2]$ (n_1 A molecules, n_2 B molecules) at time t changes in time. For the reaction above:

$$\dot{p}(n_1,n_2,t) = -kn_1p(n_1,n_2,t) + kn_1p(n_1,n_2-1,t)\,. \qquad [2]$$

The first term reflects a transition from state $[n_1, n_2]$ to state $[n_1, n_2 + 1]$ and therefore leads to a decrease in $p(n_1n_2, t)$. The second term denotes the transition $[n_1, n_2 - 1] \rightarrow [n_1, n_2]$ and leads to increased $p(n_1, n_2, t)$. The master equation above is linear and can be solved for the moments by constructing the moment-generating function. In general for N system variables the moment-generating function is given by:

$$F(z_1,z_2,...,z_N,t) = \sum_{n_1,n_2,...,n_N} z_1^{n_1} z_2^{n_2} ... z_N^{n_N}\, p(n_1,n_2,...,n_N,t)\,, \qquad [3]$$

where the sum runs over all possible states for each n_i (in this case, from 0 to ∞). This function has the following useful properties:

$$F\big|_1 = 1, \quad \left.\frac{\partial F}{\partial z_i}\right|_1 = \langle n_i \rangle, \quad \left.\frac{\partial^2 F}{\partial z_i^2}\right|_1 = \langle n_i(n_i-1)\rangle, \quad \left.\frac{\partial^2 F}{\partial z_i \partial z_j}\right|_1 = \langle n_i n_j \rangle\,, \qquad [4]$$

where $|_1$ means that the function is evaluated at $z_j = 1$ for all j. These expressions justify the name "moment generating": we can obtain the moments of the probability distribution by evaluating the partial derivatives of the function.

Multiplying the master equation above by $z_1^{n_1} z_2^{n_2}$ on both sides gives

$$\sum_{n_1,n_2} z_1^{n_1} z_2^{n_2}\, \dot{p}(n_1,n_2,t) = -k \sum_{n_1,n_2} z_1^{n_1} z_2^{n_2} n_1 p(n_1,n_2,t) + k \sum_{n_1,n_2} z_1^{n_1} z_2^{n_2} n_1 p(n_1,n_2-1,t)\,. \qquad [5]$$

This equation can be simplified significantly by realizing that

$$\frac{\partial F}{\partial z_1} = \sum_{n_1} \left(n_1 z_1^{n_1-1}\right) p(n_1,n_2,t) \quad \Rightarrow \quad z_1 \frac{\partial F}{\partial z_1} = \sum_{n_1} n_1 z_1^{n_1} p(n_1,n_2,t)$$

$$\sum_{n_1=0,n_2=0} n_1 z_1^{n_1} z_2^{n_2}\, p(n_1,n_2-1,t) = z_1 \sum_{n_1=0,n_2=-1} \left(n_1 z_1^{n_1-1}\right) z_2^{n_2}\, p(n_1,n_2-1,t) \qquad [6]$$

$$= z_1 \sum_{n_1=0, n'_2=0} \left(n_1 z_1^{n_1-1}\right) z_2^{n'_2+1} p(n_1, n'_2, t) = z_1 z_2 \frac{\partial F}{\partial z_1},$$

where the change in the lower limit of the sum for n_2 is allowed since $p(n_1, -1, t) = 0$. This leads to

$$\dot{F}(z_1, z_2, t) = k z_1 (z_2 - 1) \frac{\partial F}{\partial z_1}. \quad [7]$$

In the special case of synthesis from a fixed number of templates ($n_1 = n$), the equation for the moment-generating function reduces to

$$\dot{F}(z_2, t) = kn(z_2 - 1)F. \quad [8]$$

This equation can be explicitly solved, but in itself it does not represent the full process. We therefore will obtain the expressions for the other terms before combining them to model a real situation.

2.2. Degradation

Now consider the degradation reaction

$$B \xrightarrow{\gamma} 0. \quad [9]$$

This reaction can represent two different processes: degradation, where molecule B is converted into a species that is not part of the subset of interest, and dilution, where it is physically separated from the volume of interest. In the latter context γ is the degradation rate and $\ln(2)/\gamma$ the half-life of the molecule. The master equation for this reaction is

$$\dot{p}(n_1, t) = -\gamma n_1 p(n_1, t) + \gamma(n_1 + 1) p(n_1 + 1, t). \quad [10]$$

Using the same strategy as above the time evolution of the moment-generation function yields:

$$\dot{F}(z_1, t) = -\gamma(z_1 - 1) \frac{\partial F}{\partial z_1}. \quad [11]$$

2.3. Forward Reaction, Conservation of Total Number of Molecules

Now consider the reaction

$$A \xrightarrow{k} B\,, \qquad [12]$$

where $n_0 + n_1 = n = \text{const}$. Since the total number n is conserved, the system is defined by only one variable. We will use n_2 as the single variable to define this system. For the reaction above:

$$\dot{p}(n_2,t) = -k(n-n_2)p(n_2,t) + k(n-n_2+1)p(n_2-1,t)\,. \qquad [13]$$

This leads[1] to:

$$\begin{aligned}\dot{F}(z_2,t) &= -knF + kz_2\frac{\partial F}{\partial z_2} + k(n+1)z_2F - kz_2\left(z_2\frac{\partial F}{\partial z_2} + F\right)\\ &= kn(z_2-1)F - kz_2(z_2-1)\frac{\partial F}{\partial z_1}\,. \end{aligned} \qquad [14]$$

Based on these elementary reactions, larger chemical networks can be built up. The results above are summarized in Table 1.

Table 1

	Reaction type	$\dot{F} =$
I	$A \xrightarrow{k} A + B$	$kz_1(z_2-1)\dfrac{\partial F}{\partial z_1}$
II	$B \xrightarrow{\gamma} 0$	$-\gamma(z_1-1)\dfrac{\partial F}{\partial z_1}$
III	$A \xrightarrow{k} B$ ($n_1 + n_2 = n = \text{const.}$, in terms of n_2)	$kn(z_2-1)F - kz_2(z_2-1)\dfrac{\partial F}{\partial z_2}$
IV	$A \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} B$	$k_1(z_2-z_1)\dfrac{\partial F}{\partial z_1} + k_{-1}(z_1-z_2)\dfrac{\partial F}{\partial z_2}$

2.4. Noise Properties of a Constitutively Expressed Gene

Based on the results for these elementary reactions the equation for the moment-generating functions of more complex networks can be easily deduced. First let us consider a constitutive expressed gene in a single copy in the chromosome of a bacterium. In this case the state of this system at any time is defined by the number of mRNA molecules r and number of proteins p for that gene. mRNA molecules are synthesized off the template DNA strand at a rate k_R and are translated at a rate k_P. The mRNA and protein degradation are described by the destruction rates γ_R and γ_P, respectively (Figure 1).

Based on the results in Table 1, the moment-generating function can be deduced directly:

$$\dot{F}(z_1,z_2,t) = k_R(z_1-1)F + k_P z_1(z_2-1)\frac{\partial F}{\partial z_1} - \gamma_R(z_1-1)\frac{\partial F}{\partial z_1} - \gamma_P(z_2-1)\frac{\partial F}{\partial z_2}. \quad [15]$$

The first two terms are the transcription and translation reactions (Table 1, type I) and the last two terms model degradation of mRNA and proteins, respectively (Table 1, type II). Below the equation will be solved for the moments in the steady state $(\dot{F}=0)$. In this case:

$$k_R(1-z_1)F = k_P z_1(z_2-1)\frac{\partial F}{\partial z_1} - \gamma_R(z_1-1)\frac{\partial F}{\partial z_1} - \gamma_P(z_2-1)\frac{\partial F}{\partial z_2}. \quad [16]$$

The mean mRNA level $\langle r \rangle$ and protein level $\langle p \rangle$ are found by taking the derivative with respect to z_1 and z_2, respectively:

$$k_R(1-z_1)\frac{\partial F}{\partial z_1} - k_R F = \left[k_P z_1(z_2-1) - \gamma_R(z_1-1)\right]\frac{\partial^2 F}{\partial z_1^2} + \left[k_P(z_2-1) - \gamma_R\right]\frac{\partial F}{\partial z_1} - \gamma_P(z_2-1)\frac{\partial^2 F}{\partial z_2 \partial z_1}. \quad [17]$$

Evaluating both expressions at $z_1 = z_2 = 1$ gives

$$\langle r \rangle = \frac{k_R}{\gamma_R},$$
$$\langle p \rangle = \frac{k_R k_P}{\gamma_R \gamma_P}. \quad [18]$$

These results are consistent with the equivalent deterministic system:

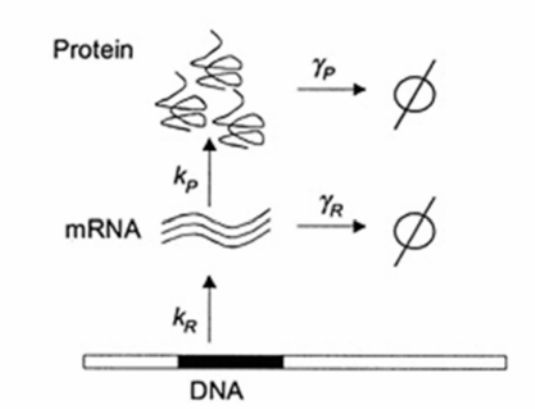


Figure 1. Basic model for constitutive expression of a single gene. Only four individual reactions are considered: creation of mRNA from a DNA template, creation of proteins from individual mRNA molecules, and the degradation/dilution of both species.

$$\begin{aligned} \langle \dot{r} \rangle &= k_R - \gamma_R \langle r \rangle , \\ \langle \dot{p} \rangle &= k_P \langle r \rangle - \gamma_P \langle p \rangle . \end{aligned} \qquad [19]$$

The fluctuations in mRNA and proteins level are found by differentiating the above equations again with respect to z_1 and z_2 and evaluating at $z_1 = z_2 = 1$:

$$\begin{aligned} \langle r^2 \rangle - \langle r \rangle^2 &= \langle r \rangle , \\ \langle rp \rangle - \langle r \rangle \langle p \rangle &= \frac{\langle p \rangle}{1 + \gamma_R / \gamma_P} , \\ \langle p^2 \rangle - \langle p \rangle^2 &= \langle p \rangle \left(\frac{k_P / \gamma_R}{1 + \gamma_P / \gamma_R} + 1 \right) . \end{aligned} \qquad [20]$$

Further moments can be obtained sequentially in this manner. Note that for a random variable with Poissonian distribution all moments are equal, so the variance over the mean equals one, as is the case for the mRNA in this model. The protein number fluctuates with a higher than Poissonian noise, the correction determined primarily by the term k_P / γ_R (the "burst size") that corresponds to the average number of proteins produced per mRNA (9).

In simple cases like this, the moments can also be obtained as a function of time. For a single gene, the noise out of equilibrium can be 40% larger than its

steady-state value in the limit of short mRNA lifetimes (9). A more detailed modeling of this process could include more intermediate processes, such as the random steps that a ribosome takes along an mRNA, but most turn out to have little effect when compared in simulations. However, when a repressor or activator is present, its binding and unbinding might have to be included in the model, for this can be a major source of noise. It is in this context that the terms shown in Table 1, type III, are needed. Furthermore, the repressor concentration itself might be fluctuating, in which case we have to consider the entire system of genes.

2.5. Linearized Matrix Formulation

The method above can also be used for interacting systems of genes, but solving it is not straightforward unless the connections are linear. Alternatively, if the system is at a stable point in steady state, the interaction can be linearized around the steady state value. A practical way of writing this out is in matrix form (9). The transition probabilities for species x_i are given by $f_i(x_1, x_2, ..., x_n)$ for creation and λ_i for destruction, and A and Γ are the matrices defined by $A_{ij} = \left.\frac{\partial f_i}{\partial x_j}\right|_{\langle x_1\rangle,\langle x_2\rangle,\ldots}$ and $\Gamma_{ij} = \gamma_i \delta_{ij}$. Letting x be the vector of chemical species, the linearized macroscopic equations are then given by $\langle \dot{x} \rangle = (A - \Gamma)\langle x \rangle$.

Note that in many cases the macroscopic equations include constant creation terms. If the system is linear it might be necessary to include an additional variable that is non-fluctuating and allows the inclusion of the constant terms in compact matrix form. As an illustration of this, the matrices for the single gene case are

$$A = \begin{bmatrix} 0 & 0 & 0 \\ k_R & 0 & 0 \\ 0 & k_P & 0 \end{bmatrix}, \quad \Gamma = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \gamma_R & 0 \\ 0 & 0 & \gamma_P \end{bmatrix}, \qquad [21]$$

where the state vector is where $x^T = (d, r, p)$ is the gene copy number. This constant state coordinate needs not to represent an actual chemical; for a system where many species have a constant creation rate, these rates can all be placed in the first column of A (setting $d = 1$ and $\Gamma_{1j} = \Gamma_{j1} = 0$). An example of this is the matrix for the case of two interacting genes, linearized around steady state, with fixed gene copy numbers d_1 and d_2, respectively, and where the first gene (r_1, p_1) represses the second (r_2, p_2) with transfer function $f(p_1)$:

$$A=\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ d_1 k_{R1} & 0 & 0 & 0 & 0 \\ 0 & k_{P1} & 0 & 0 & 0 \\ d_2 k_{R2} & 0 & d_2 \left.\dfrac{\partial f}{\partial p_1}\right|_{\langle p_1\rangle} & 0 & 0 \\ 0 & 0 & 0 & k_{P2} & 0 \end{bmatrix}, \qquad \Gamma=\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \gamma_{R1} & 0 & 0 & 0 \\ 0 & 0 & \gamma_{P1} & 0 & 0 \\ 0 & 0 & 0 & \gamma_{R2} & 0 \\ 0 & 0 & 0 & 0 & \gamma_{P2} \end{bmatrix}, \qquad [22]$$

where $k_{R2}=f\left(\langle p_1\rangle\right)-\left.\dfrac{\partial f}{\partial p_1}\right|_{\langle p_1\rangle}\langle p_1\rangle$.

Written in terms of these matrices, the master equation in generating function form would be

$$\dot{F}=\sum_i (1-z_i)\left(\Gamma_{ii}\frac{\partial F}{\partial z_i}-\sum_j A_{ij}z_j\frac{\partial F}{\partial z_j}\right). \qquad [23]$$

At steady state, $\dot{F}=0$, and taking the derivative with respect to z_l we obtain

$$0=\sum_i (1-z_i)\left(\Gamma_{ii}\frac{\partial^2 F}{\partial z_i \partial z_l}-\sum_j A_{ij}z_j\frac{\partial^2 F}{\partial z_j \partial z_l}-A_{il}\frac{\partial F}{\partial z_l}\right)-\left(\Gamma_{ll}\frac{\partial F}{\partial z_l}-\sum_j A_{lj}z_j\frac{\partial F}{\partial z_j}\right). \qquad [24]$$

Setting all $z_i = 1$, we have for each i

$$0=\Gamma_{ii}\left.\frac{\partial F}{\partial z_i}\right|_1-\sum_j A_{ij}z_j\left.\frac{\partial F}{\partial z_j}\right|_1 \Rightarrow 0=(\Gamma-A)\nabla F|_1=(\Gamma-A)\langle x_1\rangle, \qquad [25]$$

corresponding to the macroscopic result. Similarly, differentiating again and evaluating at $z_i = 1$,

$$\begin{aligned} 0&=\left(\Gamma_{ii}\left.\frac{\partial^2 F}{\partial z_i \partial z_l}\right|_1-\sum_j A_{ij}z_j\left.\frac{\partial^2 F}{\partial z_j \partial z_l}\right|_1-A_{il}\left.\frac{\partial F}{\partial z_l}\right|_1\right)+\left(\Gamma_{ll}\left.\frac{\partial^2 F}{\partial z_l \partial z_i}\right|_1-\sum_j A_{lj}z_j\left.\frac{\partial^2 F}{\partial z_j \partial z_i}\right|_1-A_{li}\left.\frac{\partial F}{\partial z_i}\right|_1\right), \\ &=\left((\Gamma-A)\nabla\nabla^T F|_1-A\Theta F|_1\right)+\left((\Gamma-A)\nabla\nabla^T F|_1-A\Theta F|_1\right)^T, \end{aligned} \qquad [26]$$

where $\Theta_{ij}=\delta_{ij}\dfrac{\partial}{\partial z_i}$. These linear equations can be solved for the means, variances, and correlations.

This approach is very general and the resulting matrix equations can be solved directly. However, even for the case of just two interacting genes this

requires a 5 x 5 matrix system as shown, so it gets cumbersome for larger systems even though most entries are zero. Using symbolic matrix manipulation software it is straightforward to obtain the desired expressions, so for known parameters this is a good method for obtaining values without further approximations.

3. THE LANGEVIN APPROACH

An alternate approach that allows for a more straightforward interpretation and scales easily to different levels of detail is to use a Langevin equation. The Langevin approach consists essentially of adding a noise term to the deterministic equations. This noise term can represent the effect of the intrinsic fluctuations (20) or the external inputs of the system (21).

For x, the concentration of some chemical species,

$$\dot{x} = f(x) \rightarrow \dot{x} = f(x) + q(x)\varepsilon(t)\,, \qquad [27]$$

where the random variable $\varepsilon(t)$ is determined by its statistical properties. Formally, this can be any random process, but in practice we assume white-noise statistics, which will give approximate values for the first two moments. The conditions for white noise are

$$\left\langle \varepsilon(t) \right\rangle = 0, \qquad \left\langle \varepsilon(t)\varepsilon(t+\tau) \right\rangle = \delta(\tau),\,, \qquad [28]$$

where $\langle\ \rangle$ denotes an ensemble average. Since we are interested in the steady state fluctuations, we will assume the coefficient of the noise term to be constant,[3] i.e., evaluated at $\left\langle x \right\rangle_{ss}$.

For the case of our basic model of the single gene, we have two macroscopic equations representing mRNA and protein creation, respectively:

$$\begin{aligned} \dot{r} &= k_R - \gamma_R r \rightarrow \dot{r} = k_R - \gamma_R r + q_r \varepsilon_r\,, \\ \dot{p} &= k_P r - \gamma_P p \rightarrow \dot{p} = k_P r - \gamma_P p + q_p \varepsilon_p\,, \end{aligned} \qquad [29]$$

where the coefficients of the noise terms are to be determined. Clearly, $\langle r \rangle = k_R/\gamma_R$ and $\langle p \rangle = k_P \langle r \rangle/\gamma_P$ from the condition of zero mean for the noise term.

The difference with the steady state $\delta r = r - \left\langle r \right\rangle$ follows the equation

$$\delta\dot{r} + \gamma_R \delta r = q_r \varepsilon_r\,. \qquad [30]$$

Fourier transforming, we obtain

$$\delta\hat{r}(\omega) = \frac{q_r\hat{\varepsilon}_r}{i\omega + \gamma_R}, \qquad [31]$$

so after multiplying by the complex conjugate and taking the average,

$$\left\langle \left|\delta\hat{r}(\omega)\right|^2 \right\rangle = \frac{{q_r}^2}{\omega^2 + \gamma_R^2}, \qquad [32]$$

The steady-state fluctuations are given by the inverse Fourier transform with t = 0:[4]

$$\delta r^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{{q_r}^2}{\omega^2 + \gamma_R^2} d\omega = \frac{{q_r}^2}{2\pi\gamma_R}\int_{-\infty}^{\infty} \frac{dx}{x^2 + 1} = \frac{{q_r}^2}{2\gamma_R}. \qquad [33]$$

But since the production of mRNA is in this model a single-step independent random process, it has a Poisson distribution, so the variance equals the mean, which implies

$$\frac{{q_r}^2}{2\gamma_R} = \frac{k_R}{\gamma_R} \Rightarrow {q_r}^2 = 2k_R. \qquad [34]$$

For the number of proteins, we have

$$\begin{aligned} \delta\dot{p} + \gamma_P\delta p &= \delta r + q_p\varepsilon_p, \\ \delta\hat{p}(\omega) &= \frac{\delta\hat{r}(w) + q_p\hat{\varepsilon}_p}{i\omega + \gamma_P}, \end{aligned} \qquad [35]$$

but in this case we also need to notice that $\left\langle \delta\hat{r}(w)\hat{\varepsilon}_p^* \right\rangle = \left\langle \delta\hat{r}(w) \right\rangle \left\langle \hat{\varepsilon}_p^* \right\rangle = 0$, since these are two independent random processes with zero mean. So in this case,

$$\left\langle \left|\delta\hat{p}(\omega)\right|^2 \right\rangle = \frac{\left\langle \left|\delta\hat{r}(\omega)\right|^2 \right\rangle + {q_p}^2}{\omega^2 + \gamma_P^2} = \frac{{q_r}^2}{\left(\omega^2 + \gamma_R^2\right)\left(\omega^2 + \gamma_P^2\right)} + \frac{{q_p}^2}{\omega^2 + \gamma_P^2}. \qquad [36]$$

We use ${q_r}^2 = 2k_R$ and ${q_p}^2 = 2k_p\frac{k_R}{\gamma_R}$ (since this represents the internal noise and for a fixed number of mRNAs the production of proteins is also a Poissonian process). Performing the inverse transform,[5]

$$\delta p^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{q_r^{\,2}}{\left(\omega^2+\gamma_R^2\right)\left(\omega^2+\gamma_P^2\right)} + \frac{q_p^{\,2}}{\omega^2+\gamma_P^2}\, d\omega = \frac{k_P^{\,2} k_R}{\left(\gamma_R^{\,2}-\gamma_P^{\,2}\right)}\left[\frac{1}{\gamma_P}-\frac{1}{\gamma_R}\right] + \frac{k_P k_R}{\gamma_P \gamma_R}. \quad [37]$$

For comparison with the previous result, note that $\langle p \rangle = k_P k_R / \gamma_P \gamma_R$, so this can be rewritten as

$$\delta p^2 = \langle p \rangle \left(\frac{k_P / \gamma_R}{1 + \gamma_P / \gamma_R} + 1 \right). \quad [38]$$

This is identical to the result obtained by the master equation. This method can be readily generalized for many interacting genes when the system is fluctuating around a steady state. As an example, we will analyze the case where one gene represses a second gene.

Let y_0, y_1 be the protein numbers of each gene, and let $f(y_0)$ be the rate of creation of the second protein as a function of the first. This means that the equations describing this system are

$$\begin{aligned} \dot{y}_0 &= k_0 - \gamma_0 y_0 , \\ \dot{y}_1 &= f\left(y_0\right) - \gamma_1 y_1 . \end{aligned} \quad [39]$$

Note that the equations include the entire process of producing a protein, so mRNA levels are no longer explicitly calculated. Including the Langevin noise term and looking at the fluctuations from steady state,

$$\begin{aligned} \delta \dot{y}_0 &= -\gamma_0 \delta y_0 + q_0 \varepsilon_0 , \\ \delta \dot{y}_1 &= f(y_0) - f(\langle y_0 \rangle) - \gamma_1 \delta y_1 + q_1 \varepsilon_1 \approx c_0 \delta y_0 - \gamma_1 \delta y_1 + q_1 \varepsilon_1 , \end{aligned} \quad [40]$$

where $c_0 = \left.\dfrac{df}{dy_0}\right|_{\langle y_0 \rangle}$, and each noise term has the same conditions as before. This linearization is valid at each stable point, but not for transitions between different stable points or for limit cycles. For very small numbers n of chemicals this also breaks down, because since these processes are mostly Poissonian, the fluctuations are of order $\sqrt{n}$, so a Taylor expansion might not be valid. Fourier transforming and taking the square and the average as before, we get

$$\left\langle \left| \delta \hat{y}_0(\omega) \right|^2 \right\rangle = \frac{q_0^{\,2}}{\omega^2 + \gamma_0^2} ,$$

$$\left\langle \left|\delta\hat{y}_1(\omega)\right|^2 \right\rangle = \frac{c_0^2 \left\langle \left|\delta\hat{y}_0(\omega)\right|^2 \right\rangle + {q_1}^2}{\omega^2+\gamma_1^2} = \frac{c_0^2 {q_0}^2}{\left(\omega^2+\gamma_1^2\right)\left(\omega^2+\gamma_0^2\right)} + \frac{{q_1}^2}{\omega^2+\gamma_1^2}. \quad [41]$$

The correlation between the genes can also be calculated from

$$\left\langle \delta\hat{y}_1 \delta\hat{y}_0^* \right\rangle = \left\langle \left(\frac{c_0 \delta\hat{y}_0(w) + q_1\hat{\varepsilon}_1}{i\omega+\gamma_1} \right) \delta\hat{y}_0^*(w) \right\rangle = \frac{c_0 \left\langle \left|\delta\hat{y}_0(\omega)\right|^2 \right\rangle}{i\omega+\gamma_1} = \frac{c_0 {q_0}^2}{\left(i\omega+\gamma_1\right)\left(\omega^2+\gamma_0^2\right)}, \quad [42]$$

where the term $\left\langle \hat{\varepsilon}_1 \delta\hat{y}_0^* \right\rangle$ vanishes because the fluctuations in the first gene are independent from the internal fluctuations in the second gene. In many cases, the decay time will be determined primarily by the dilution time, so it will be the same for all genes. This assumption simplifies the expressions that are obtained upon transforming back:

$$\left\langle \delta y_0^2 \right\rangle = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{{q_0}^2}{\omega^2+\gamma^2} d\omega = \frac{{q_0}^2}{2\gamma}, \quad [43]$$

$$\left\langle \delta y_1^2 \right\rangle = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{c_0^2 {q_0}^2}{\left(\omega^2+\gamma^2\right)^2} + \frac{{q_1}^2}{\omega^2+\gamma^2} d\omega = \frac{c_0^2 {q_0}^2}{4\gamma^3} + \frac{{q_1}^2}{2\gamma}, \quad [44]$$

$$\left\langle \delta y_1 \delta y_0 \right\rangle = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{c_0 {q_0}^2}{\left(i\omega+\gamma\right)\left(\omega^2+\gamma^2\right)} d\omega = \frac{c_0 {q_0}^2}{2\pi} \int_{-\infty}^{\infty} \frac{\left(i\omega-\gamma\right)}{\left(\omega^2+\gamma^2\right)^2} d\omega = \frac{c_0 {q_0}^2}{4\gamma^2}, \quad [45]$$

where the irrational part of the integral vanishes because of parity. From our previous results we know that for a single gene,

$$\left\langle \delta y_0^2 \right\rangle = \frac{{q_0}^2}{2\gamma} = \left\langle y_0 \right\rangle \left(\frac{k_{Po}/\gamma_{Ro}}{1+\gamma/\gamma_{Ro}} + 1 \right) \approx \left\langle y_0 \right\rangle \left(b_0+1\right) \Rightarrow {q_0}^2 \approx 2\gamma \left\langle y_0 \right\rangle \left(b_0+1\right), \quad [46]$$

where b_i is the burst size for gene i. For basic Hill-type repression,

$$f\left(y_0\right) = \frac{k_1}{1+\left(y_0/Y_{1/2}\right)^h} + B_1 \Rightarrow c_0 = \frac{-k_1}{\left(1+\left(y_0/Y_{1/2}\right)^h\right)^2} \left(\frac{\left\langle y_0 \right\rangle}{Y_{1/2}} \right)^{h-1} \frac{h}{Y_{1/2}},$$

$$= \left\langle y_1 \right\rangle^2 \left(\frac{\left\langle y_0 \right\rangle}{Y_{1/2}} \right)^{h-1} \frac{-h\gamma^2}{k_1 Y_{1/2}}, \quad [47]$$

where $k_1 + B_1$ is the maximum creation rate, $Y_{1/2}$ is the half induction point, h is the Hill coefficient, and B_1 is the basal transcription level. Assuming that the internal noise for the second gene alone has the same form,

$$q_1^{\,2} \approx 2\gamma \langle y_1 \rangle (b_1 + 1), \qquad [48]$$

the variance and correlation can be explicitly written as

$$\langle \delta y_1^2 \rangle = \left(\frac{\langle y_0 \rangle}{Y_{1/2}} \right)^{2h} \left(\frac{h\gamma}{2k_1} \right)^2 \frac{\langle y_1 \rangle^2}{\langle y_0 \rangle} 2(b_0 + 1) + \langle y_1 \rangle (b_1 + 1),$$

$$\langle \delta y_1 \delta y_0 \rangle = \langle y_1 \rangle^2 \left(\frac{\langle y_0 \rangle}{Y_{1/2}} \right)^h \frac{-h\gamma}{2k_1} (b_0 + 1). \qquad [49]$$

Note that we need the parameters of the macroscopic equations plus an "internal" parameter for each gene, $b_i = k_{Pi}/\gamma_{Ri}$, which depends on the parameters of the macroscopic equations for each gene.

4. DISCUSSION AND CONCLUSIONS

We thus have a versatile toolbox of modeling approaches at our disposal, each suitable for different situations and with different levels of approximation and scalability. A direct, full master equation approach provides every detail of the distribution in the few cases where it can be analytically solved. The properties of the noise as encoded by the variance and correlations can be obtained explicitly in a wide range of cases from the equations obtained from the generating-function version of the master equation, in some cases as a function of time. If we are only interested in the steady-state noise, the linearized matrix formulation of these equations provides a compact way of treating more complex systems. For a few variables this can be done easily, but increasing the size of the system can lead to cumbersome algebra. This matrix approach can be easily solved using numerical methods or matrix manipulation software, but then some of the insight might be lost. The Langevin approach provides an alternate, straightforward way of obtaining the noise characteristics that easily incorporate the effects of larger systems and other sources of noise.

An additional tool that we have not covered but is worth keeping in mind is the possibility of performing detailed Monte Carlo simulations of the system. Methods of varying degrees of approximation and efficiency have been developed recently (24,25), based on Gillespie's stochastic simulation algorithm (26). This algorithm is essentially exact, but depending on the complexity of the system and the computing power available, a suitable level of detail can be chosen

by using one of the preceding algorithms and following only processes that occur on the time scale of interest. The ability to determine the stochastic properties of gene expression will allow a more quantitative determination of the range of possible cellular responses to a stimulus or treatment, which will be increasingly important as treatments become more targeted and personalized. This should be especially useful in gene therapy, where the variability of the gene product could be controlled. It can also provide insight into the possible causes of disease, as in tumor formation from variations in phenotype due to haploinsufficiency (27). On the other hand, the ability to determine the stochastic properties of a circuit will permit increasingly complex metabolic engineering, allowing for higher-order traits like robustness or stability to be included in the design rather than painstakingly evolved (28).

5. NOTES

1. In this case, the sums only go up to n, instead of ∞. However, the extra terms that appear when applying the change of variables cancel with each other.

2. This can be summarized in a very practical way (13) in terms of the logarithmic gains to obtain an equation that reflects the resulting components of the noise.

3. For the case where $q(x)$ is not constant, the stochastic differential equation will be understood to follow the Stratonovich interpretation (19,22). This allows a general Fokker-Planck equation to be written in this form, but will not be necessary in the cases of interest.

4. From the Wiener-Khinchin theorem; see (23).

5. $\frac{1}{2\pi}\int \frac{d\omega}{\left(\omega^2+\gamma^2\right)^n} = \frac{1}{2\sqrt{\pi}\gamma^{2n-1}}\frac{\Gamma\left(n-1/2\right)}{\Gamma\left(n\right)}$, where $\Gamma(n)=\Gamma(n-1)(n-1)$, $\Gamma(1)=1$, and $\Gamma(1/2)=\sqrt{\pi}$.

6. REFERENCES

1. Ptashne M. 1992. *A genetic switch: phage lambda and higher organisms*. Cell Press, Cambridge.
2. Arkin A, Ross J, McAdams HH. 1998. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics* **149**:1633–1648.
3. van de Putte P, Goosen N. 1992. DNA inversions in phages and bacteria. *Trends Genet* **8**:457–462.
4. von Dassow G, Meir E, Munro EM, Odell GM. 2000. The segment polarity network is a robust developmental module. *Nature* **406**:188–192.
5. Becskei A, Serrano L. 2000. Engineering stability in gene networks by autoregulation. *Nature* **405**:590–593.

6. Ko MS, Nakauchi H, Takahashi N. 1990. The dose dependence of glucocorticoid-inducible gene expression results from changes in the number of transcriptionally active templates. *EMBO J* **9**:2835–2842.
7. Berg OG. 1978. A model for statistical fluctuations of protein number in a microbial population. *J Theor Biol* **71**:587–603.
8. McAdams HH, Arkin A. 19 1977. Stochastic mechanisms in gene expression. *Proc Natl Acad Sci USA* **94**:814–819.
9. Thattai M, van Oudenaarden A. 2001. Intrinsic noise in gene regulatory networks. *Proc Natl Acad Sci USA* **98:**8614–8619.
10. Kierzek AM, Zaim J, Zielenkiewicz P. 2001. The effect of transcription and translation initiation frequencies on the stochastic fluctuations in prokaryotic gene expression. *J Biol Chem* **276:**8165–8172.
11. Kepler TB, Elston TC. 2001. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys J* **81:**3116–3136.
12. Swain PS, Elowitz MB, Siggia ED. 2002. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci USA* **99**:12795–12800.
13. Paulsson J. 2004. Summing up the noise in gene networks. *Nature* **427**:415–418.
14. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A. 2002. Regulation of noise in the expression of a single gene. *Nature Genet* **31:**69–73.
15. Elowitz MB, Levine AJ, Siggia ED, Swain PS. 2002. Stochastic gene expression in a single cell. *Science* **297:**1183–1186.
16. Blake WJ, Kærn M, Cantor CR, Collins JJ. 2003. Noise in eukaryotic gene expression. *Nature* **422**:633–637.
17. Raser J, O'Shea, EK. 2004. Control of stochasticity in eukaryotic gene expression. *Science* **304**:1811–1814.
18. Rao CV, Wolf DM, Arkin AP. 2002. Control, exploitation and tolerance of intracellular noise. *Nature* **420**:231–237.
19. van Kampen NG. 1992. *Stochastic processes in physics and chemistry*. North-Holland, Amsterdam.
20. Thattai M, van Oudenaarden A. 2002. Attenuation of noise in ultrasensitive signaling cascades. *Biophys J* **82**:2943–2950.
21. Hasty J, Pradines J, Dolnik M, Collins JJ. 2000. Noise-based switches and amplifiers for gene expression. *Proc Natl Acad Sci USA* **97**:2075–2080.
22. Øksendal BK. 1998. *Stochastic differential equations: an introduction with applications.* Springer, New York.
23. Rowe HE. 1965. *Signals and noise in communication systems*. van Nostrand, Princeton.
24. Gillespie DT. 2001. Approximate accelerated stochastic simulation of chemically reacting systems. *J Chem Phys* **115**:1716–1733.
25. Gibson MA, Bruck J. 2000. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J Phys Chem A* **104**:1876–1889.
26. Gillespie DT. 1977. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* **81**:2340–2361.
27. Magee JA, Abdulkadir SA, Milbrandt J. 2003. Haploinsufficiency at the Nkx3.1 locus: a paradigm for stochastic, dosage-sensitive gene regulation during tumor initiation. *Cancer Cell* **3**:273–283.
28. McAdams HH, Arkin A. 2000. Towards a circuit engineering discipline. *Curr Biol* **10**:R318–R320.

1.2

MODELING RNA FOLDING

Ivo L. Hofacker

Institute for Theoretical Chemistry and Structural Biology, University of Vienna

Peter F. Stadler

Bioinformatics, Department of Computer Science, University of Leipzig, Leipzig, Germany

In recent years it has become evident that functional RNAs in living organisms are not just curious remnants from a primordial RNA world but a ubiquitous phenomenon complementing protein enzyme based activity. Functional RNAs, just like proteins, depend in many cases upon their well-defined and evolutionarily conserved three-dimensional structure. In contrast to protein folds, however, RNA molecules have a biophysically important coarse-grained representation: their secondary structure. At this level of resolution at least, RNA structures can be efficiently predicted given only the sequence information. As a consequence, computational studies of RNA routinely incorporate structural information explicitly. RNA secondary structure prediction has proven useful in diverse fields, ranging from theoretical models of sequence evolution and biopolymer folding, to genome analysis, and even the design of biotechnologically or pharmaceutically useful molecules. Properties such as the existence of neutral networks or shape space covering are emergent properties determined by the complex, highly nonlinear relationship between RNA sequences and their structures.

1. INTRODUCTION

It is not hard to argue that **RNomics**, i.e., an understanding of functional RNAs (both ncRNA genes and functional motifs in protein-coding RNAs) and

Address correspondence to: Peter F. Stadler, Department of Computer Science, University of Leipzig, Härtelstrasse 16-18, Leipzig, D-04107 Germany (peter.stadler@bioinf. uni-leipzig.de)

their interactions at a genomic level, is of utmost practical and theoretical importance in modern life sciences. A comprehensive understanding of the biology of a cell obviously requires the knowledge of the identity of *all* encoded RNAs, the molecules with which they interact, and the molecular structures of these complexes (18).

Structural genomics, the systematic determination of all macromolecular structures represented in a genome, until very recently has been focused almost exclusively on proteins. Although it is commonplace to speak of "*genes and their encoded protein products*," thousands of human genes produce transcripts that exert their function without ever producing proteins. The list of functional non-coding RNAs (ncRNAs) includes key players in the biochemistry of the cell. Many of them have characteristic secondary structures that are highly conserved in evolution. Databases (referenced below) collect the most important classes:

- **tRNA**. Transfer RNAs are the adapters that translate the triplet nucleic acid code of RNA into the amino acid sequence of proteins (128).
- **rRNA**. Ribosomal RNAs are central to the translational machinery. Recent results strongly indicate that peptide bond formation is catalyzed by rRNA (46,89,103,132,150).
- **snoRNA**. Small nucleolar RNAs are required for rRNA processing and base modification in the eukaryotic nucleolus (102,116).
- **snRNA**. Small nuclear RNAs are critical components of spliceosomes, the large ribonucleoprotein complexes that splice introns out of pre-mRNAs in the nucleus (161).
- **tmRNA**. The bacterial tmRNA (also known as 10Sa RNA or SsrA) was named for its dual tRNA-like and mRNA-like nature. tmRNA engages in a translation process, adding in trans a C-terminal peptide tag to the unfinished protein at a stalled ribosome. The tmRNA-directed tag targets the unfinished protein for proteolysis (146,160).
- **RNase P**. Ribonuclease P is responsible for the 5'-maturation of tRNA precursors. Ribonuclease P is a ribonucleoprotein, and in bacteria (and some Archaea) the RNA subunit alone is catalytically active in vitro, i.e., it is a ribozyme (8). RNase MRP, which shares structural similarities with RNase P RNA, cleaves at a specific site in the precursor-rRNA transcript to initiate processing of the 5S rRNA.
- **telRNA**. Telomerase RNA. Telomeres are the specialized DNA protein structures at the ends of eukaryotic chromosomes. Telomerase is a ribonucleoprotein reverse transcriptase that synthesizes telomeric DNA (6).
- **SRPRNA**. The signal recognition particle is a universally conserved ribonucleoprotein. It is involved in the co-translational targeting of proteins to membranes (37).

- **miRNA**. Micro-RNAs (75,77,80) regulate gene expression by regulating mRNA expression by a mechanism closely linked to RNA interference by small double-stranded RNAs (see, e.g. (7,94)). They are cleaved from their precursors, the small temporal RNAs (**stRNAs**), by the enzyme *Dicer*.

In addition, there is a diverse list of ncRNAs with sometimes enigmatic function. We give just a few examples (see also the `Rfam` database (41)): the 17-kb *Xist* RNA of humans and the smaller roX RNAs of *Drosophila* play a key role in dosage compensation and X chromosome inactivation (2,31). Several large ncRNAs are expressed from imprinted regions. Many of these are cis-antisense RNAs that overlap coding genes on the other genomic strand (see e.g. (22)). An RNA (meiRNA) regulates the onset of meiosis in fission yeast (100). Human vaults are intracellular ribonucleoprotein particles believed to be involved in multi-drug resistance. The complex contains several small untranslated RNA molecules (152). No precise function was known by the summer of 2003 for the human H19 transcript, the hrsω transcript induced by heat shock in *Drosophila*, or the *E. coli* 6S RNA, see e.g. (23).

Even though the sequence of the human DNA is known by now, the contents of about half of it remains unknown. The diversity of sequences, sizes, structures, and functions of the known ncRNAs strongly suggests that we have seen only a small fraction of the functional RNAs. Most of the ncRNAs are small, do not have translated ORFs, and are not polyadenylated. Unlike protein coding genes, ncRNA gene sequences do not seem to exhibit a strong common statistical signal, hence a reliable general purpose computational genefinder for non-coding RNA genes has been elusive. It is quite likely therefore that a large class of genes has gone relatively undetected so far because they do not make proteins (20).

Another level of RNA function is presented by functional motifs within protein-coding RNAs. A few of the best-understood examples of structurally conserved RNA motifs in viral RNAs include:

- An IRES (internal ribosomal entry site) region is used instead of a CAP to initialize translation by Picornaviridae, some Flaviviridae including Hepatitis C virus, and a small number of mRNAs, see, e.g. (62,105,115).
- The TAR hairpin structure in HIV and related Retroviruses is the target for viral transactivation.
- The RRE structure of Retroviruses serves as a binding site for the Rev protein and is essential for viral replication. The RRE is a characteristic five-fingered structural motif, see, e.g. (16).
- The CRE hairpin (148) in Picornaviridae is vital for replication.

Genes in eukaryotes are often interrupted by intervening sequences, *introns*, that must be removed during gene expression. Similarly, rRNAs are produced from a pre-rRNA that contains so-called internal and external transcribed spacers. These contain regions with characteristic secondary structures (17). RNA splicing is the process by which these parts are precisely removed from the pre-mRNA and the flanking, functional exons are joined together (40). Regulated mechanisms of alternative splicing allow multiple different proteins to be translated from a single RNA transcript. Mutations can affect splicing of certain introns, leading to abnormal conditions. For example, a form of thalassemia, a blood disorder, is due to a mutation causing splicing failure of an intron in a globin transcript, which then becomes untranslatable; see, e.g. (130). The splicing of most nuclear genes is performed by the spliceosome; however, in many cases the splicing reaction is self-contained, that is, the intron—with the help of associated proteins—splices itself out of the precursor RNA; see e.g. (93) for a review.

A textbook example of a functional RNA secondary structure is the *Rho*-independent termination in *E. coli*. The newly synthesized mRNA forms a hairpin in the 3'NTR that interacts with the RNA polymerase, causing a change in conformation and the subsequent dissociation of the Enzyme–DNA–RNA complex. For a computational analysis of the *Rho*-independent transcription terminators we refer to (162).

Only part of the mature mRNA is translated into a protein. At the beginning of the mRNA, just behind the cap, is a non-coding sequence, the so-called leader sequence (10–200 nt), which may be followed by another non-coding sequence of up to 600 nt. An increasing number of functional features in the untranslated regions of eukaryotic mRNA have been reported in recent years (67,105).

An extreme example are the Early Noduline genes. Enod40, which is expressed in the nodule primordium developing in the root cortex of leguminous plants after infection by symbiotic bacteria (127), codes for an RNA of about 700 nt that gives rise to two short peptides, 13 and 27 amino acids, respectively. The RNA structure itself exhibits significant conservation of secondary structure motifs (55), and might take part in localization of mRNA translation (101), as in the case of the bicoid gene *bcd* of *Drosophila* (87).

2. RNA SECONDARY STRUCTURES AND THEIR PREDICTION

As with all biomolecules, the function of RNAs is intimately connected to their structure. It does not come as a surprise, therefore, that most of the classes of functional RNAs listed in the introduction have, like the well-known *cloverleaf* structure of tRNAs, distinctive structural characteristics. While successful predictions of RNA tertiary structure remain exceptional feats, RNA secondary

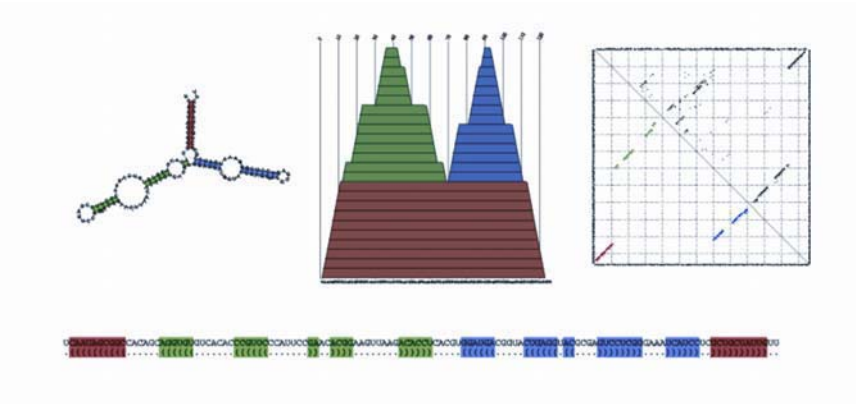


Figure 1. RNA secondary structure of a 5S ribosomal RNA. Secondary structure graph (left), mountain representations (middle), dot plot (right), and bracket notation (bottom). In the "mountain representation" each base pair (i,j) is represented by a bar from i to j. In the upper right half of the dot plot, every possible base pair (i,j) is represented by a square in row i and column j, with area proportional to its probability in thermodynamic equilibrium p_{ij}; the lower half of the plot only shows those pairs that are part of the optimal structure. In the bracket notation a secondary structure is encoded by a string of dots and brackets, where dots represent unpaired bases and matching brackets represent base pairs. In all representations base pairs of the three arms of the structure are color coded for easier comparison.

structures can be predicted with reasonable accuracy, and have proven to be a biologically useful description.

A secondary structure of a given RNA sequence is the list of (Watson-Crick and wobble) base pairs satisfying two constraints: (1) each nucleotide takes part in at most one base pair, and (2) base pairs do not cross, i.e., there are no knots or pseudo-knots. While pseudo-knots are important in many natural RNAs (145), they can be considered part of the tertiary structure for our purposes. Secondary structure can be represented in various equivalent ways (see Figure 1).

The restriction to knot-free structures is necessary for efficient computation by means of dynamic programming algorithms (55,56,95,98,119,140,149,153–155). The memory and CPU requirements of these algorithms scale with sequence length n as $\mathcal{O}(n^2)$ and $\mathcal{O}(n^3)$, respectively, making structure prediction feasible even for large RNAs of about 10000 nucleotides, such as the genomes of RNA viruses (57,64,148). There are two implementations of various variants of these dynamic programming algorithms: the `mfold` package by Michal Zuker, and the `Vienna RNA Package` by the present authors and their collaborators. The latter is freely available from http://www.tbi.univie.ac.at/.

These thermodynamic folding algorithms are based on an energy model that considers additive contributions from stacked base pairs and various types of loops; see e.g. (92,137). Two widely used methods for determining nucleic acid

thermodynamics are absorbance melting curves and microcalorimetry; see (118) for a review.

Recently, algorithms have been described that are able to deal with certain classes of pseudo-knotted structures, however, at considerable computational cost (1,36,86,114). Alternatively, heuristics such as genetic algorithms can be used (81). A common problem of all these approaches is the still very limited information about the energetics of pseudo-knots (44,66).

3. NEUTRAL NETWORKS IN THE SEQUENCE SPACE

A more detailed analysis of functional classes of RNAs shows that their structures are very well conserved while at the same time there may be little similarity at the sequence level, indicating that the structure has actual importance for the function of the molecule. In order to understand the evolution of functional RNAs one therefore has to understand the relation between sequence (genotype) and structure (phenotype). Although qualitatively there is ample evidence for neutrality in natural evolution as well as in experiments under controlled conditions in the lab, very little is known about regularities in general genotype–phenotype relations. In the RNA case, however, the phenotype can be approximated by the minimum free energy structure of RNA; see e.g. (121) for a recent review. This results in a complex, highly nonlinear genotype–phenotype map, which, however, is still computable. This simplifying assumption is met indeed by RNA evolution experiments in vitro (5) as well as by the design of RNA molecules through artificial selection (147).

There is ample evidence for redundancy in genotype–phenotype maps f in the sense that many genotypes cannot be distinguished by an evolutionarily relevant coarse-grained notion of phenotypes, which, in turn, gives rise to fitness values that cannot be faithfully separated through selection. Regarding the folding algorithms as a map f that assigns a structure $s = f(x)$ to each sequence x, we can phrase our question more precisely: we need to know how the set of sequences $f^{-1}(s)$ that folds into a given structure s is embedded in the sequence space (where the genotypes are interpreted as nodes and all Hamming distance-one neighbors are connected by an edge). The subgraphs of the sequence space that are defined by the sets $f^{-1}(s)$ are called *neutral networks* (122).

The most important global characterization of neutral networks is its average fraction of neutral neighbor $\bar{\lambda}$, usually called the (degree of) neutrality. Neglecting the influence of the distribution of neutral sequences over the sequence space, the degree of neutrality will increase with size of the pre-image. Generic properties of neutral networks (108) are readily derived by means of a random graph model. Theory predicts a phase transition-like change in the appearance of neutral networks with increasing degree of neutrality at a critical value:

$$\lambda_{cr} = 1 - \kappa^{-\frac{1}{\kappa - 1}}, \qquad [1]$$

where κ is the size of the genetic alphabet. For example, $\kappa = 4$ for the canonical genetic alphabet {A,U,G,C}. If $\bar{\lambda} < \lambda_{cr}$ then the network consists of many isolated parts with one dominating giant component. On the other hand, the network is generically connected if $\bar{\lambda} > \lambda_{cr}$. The critical value λ_{cr} is the connectivity threshold. This property of neutral networks reminds one of percolation phenomena known from different areas of physics, although the high symmetry of the sequence space, with all points being equivalent, introduces a difference in the two concepts.

A series of computational studies (27–30,42,43,63,122) has in the last decade drawn a rather detailed picture of the genotype–phenotype map of RNA (see also Figure 2).

(i) **More sequences than structures**. For sequence spaces of chain lengths $n \geq 10$ there are orders of magnitude more sequences than structures, and hence the map is many-to-one.

(ii) **Few common and many rare structures**. Relatively few common structures are opposed by a relatively large number of rare structures, some of which are formed by a single sequence only ("relatively" points at the fact that the numbers of both common and rare structures increase exponentially with n, but the exponent for the common structures is smaller than that for the rare ones).

(iii) **Shape space covering**. The distribution of neutral genotypes, which are sequences that fold into the same structure, is approximately random in the sequence space. As a result it is possible to define a spherical ball with diameter d_{cov} being much smaller than the diameter n of the sequence space, which contains on average for every common structure at least one sequence that folds into it.

(iv) **Existence and connectivity of neutral networks**. Neutral networks, being pre-images of phenotypes or structures in the sequence space, of common structures are connected unless specific and readily recognizable special features of RNA structures require specific non-random distribution in the {A,U,G,C} sequence space, $\mathcal{Q}^{(\mathrm{AUGC})}$. (For structures formed from sequences over a {G,C} alphabet the connectivity threshold is higher, whereas at the same time the mean number of neutral neighbors is smaller).

Shape space covering, item (iii) above, is a consequence of the high susceptibility of RNA secondary structures to randomly placed point mutations. Computer simulations (28,122) have shown that a small number of point mutations is very likely to cause large changes in the secondary structures: mutations in 10% of the sequence positions already lead almost surely to unrelated structures if the mutated positions are chosen randomly. The genotype–phenotype map of RNA

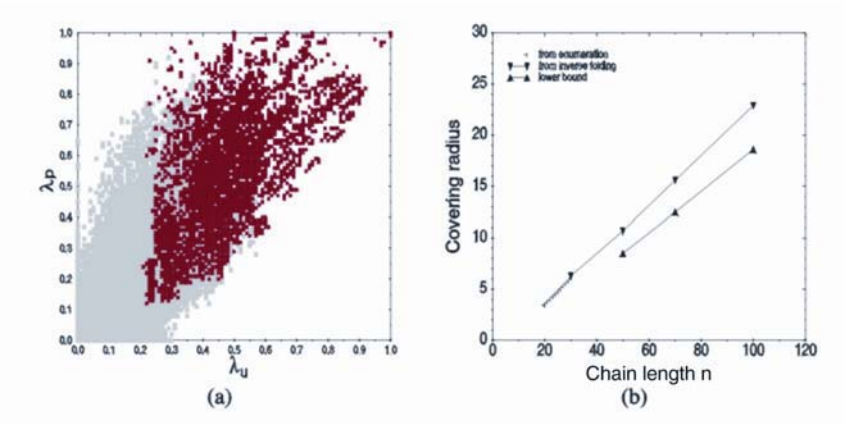


Figure 2. *Neutral Networks and Shape Space Covering*. (**a**) Neutral networks in an exhaustive survey of the *GC* sequence space with length $n = 30$ (43) are fragmented (light grey) if the fractions λ_u and λ_p of neutral mutations in the unpaired and paired parts of the sequence are below a threshold value. Above the threshold the neutral networks consist of one to four connected components. The fragmentation of the single connected component into a small number of (barely) separated subsets can be explained by the details of an energy-based folding model; see (124). (**b**) The shape space covering radius d_{cov} scales linearly with the chain length n with a slope $\varphi \approx 1/4$. Data are taken from (43).

thus exhibits a remarkable combination of robustness and fragility, typical for many complex systems (see also Part II, chapter 5, by Krakauer, this volume). Neutral networks and shape space covering are emergent properties in this setting.

The set of nodes of the neutral network $f^{-1}(s)$ is embedded in a compatible set $C(s)$ that includes all sequences that can form the structure s as suboptimal *or* minimum free energy conformation $f^{-1}(S) \subseteq C(s)$. Sequences at the intersection $C(s') \cap C(s'')$ of the compatible sets of two neutral networks in the same sequence space are of actual interest because these sequences can simultaneously carry properties of the different RNA folds. For example, they can exhibit catalytic activities of two different ribozymes at the same time (120). The intersection theorem (108) states that for all pairs of structures s' and s'' the intersection $C(s') \cap C(s'')$ is always non-empty. In other words, for each arbitrarily chosen pair of structures there will be at least one sequence that can form both. If s' and s'' are both common structures, bistable molecules that have equal preference for both structures are easy to design (25,53). A particularly interesting experimental case is described in (120).

At least features (i), (ii), and (iv) of the neutral networks of RNA seem to hold for the more complicated protein spaces as well (3,4) (see, e.g. (71) for experimental data).

The impact of these features on evolutionary dynamics is discussed in detail in (63,123): a population explores the sequence space in a diffusion-like manner along the neutral network of a viable structure. Along the fringes of the population, novel structures are produced by mutation at a constant rate (65). Fast diffusion together with perpetual innovation makes these landscapes ideal for evolutionary adaptation (30) and sets the stage for the evolutionary biotechnology of RNA (123).

4. CONSERVED RNA STRUCTURES

As we have seen, even a small number of randomly placed point mutations very likely leads to a complete disruption of the RNA structure. Secondary structure elements that are consistently present in a group of sequences with less than, say, 95% average pairwise identity are therefore almost certainly the result of stabilizing selection, not a consequence of the high degree of sequence conservation. If selection acts to preserve structure, then this structure must have some function. It is of considerable practical interest therefore to efficiently compute the consensus structure of a collection of such RNA molecules.

A promising approach to this goal is the combination of the "phylogenetic" information that is contained in the sequence co-variations and the information on the (local) thermodynamic stability of the molecules. Such methods for predicting RNA conserved and consensus secondary structure fall into two broad groups: those starting from a multiple sequence alignment and algorithms that attempt to solve the alignment problem and the folding problem simultaneously. The main disadvantage of the latter class of methods (38,39,117,133) is their high computational cost, which makes them unsuitable for long sequences such as 16S or 23S RNAs. Most of the alignment based methods start from thermodynamics-based folding and use the analysis of sequence covariations or mutual information for post-processing; see, e.g. (55,70,79,84,85). The converse approach is taken in (50), where ambiguities in the phylogenetic analysis are resolved based on thermodynamic considerations.

It is important to clearly distinguish the consensus structure of a set of RNA sequences from the collection of structural features that are conserved among these sequences. Whenever there are reasons to assume that the structure of the whole molecule is conserved, one may attempt to compute a consensus structure. On the other hand, consensus structures are unsuitable when a significant part of the molecule has no conserved structures. RNA virus genomes, for instance, contain only local structural patterns (such as the IRES in Picorna viruses or the TAR hairpin in HIV). Such features can be identified with a related approach that is implemented in the `alidot` algorithm (59,60). This program ranks base pairs using both the thermodynamic information contained in the base pairing probability matrix and the information on compensatory, consistent,

and inconsistent mutations contained in the multiple sequence alignment. The approach is different from other efforts because it does not assume that the sequences have a single common structure. In this sense `alidot` combines structure prediction and motif search (15). An implementation of this algorithm is available from http://www.tbi.univie.ac.at/. This approach to surveying functional structures goes beyond search software such as `RNAmot` (34) in that it does not require any *a priori* knowledge of the functional structure motifs and it goes beyond searches for regions that are especially thermodynamically stable or well-defined (68) in that it returns a specific prediction for a structure if *and only if* there is sufficient evidence for structural conservation.

Of course, it is not possible to determine the function of a conserved structure or structural element without additional experimental input. Nevertheless, knowledge about their location can be used to guide, for instance, deletion studies (90). Knowledge of both protein coding regions and functional RNA structures in the viral genome is needed, e.g., to rationally design attenuated mutants for vaccine development.

Structure predictions of a set of sequences are conveniently summarized in the form of Hogeweg-style mountain plots (61) (see Figure 3).

The computation of consensus and conserved RNA structures has been used to compile an *Atlas* of potentially functional RNA motifs in RNA virus genomes. Detailed data are available at present for Picornaviridae (148), Hepatitis B virus (73,129), and Flaviviridae (136).

5. DISCUSSION

Structural genomics, the systematic determination of all macromolecular structures represented in a genome is at present focused almost exclusively on proteins. Over the past two decades it has become clear, however, that a variety of RNA molecules have important, and sometimes essential, biological functions beyond their roles as rRNAs, tRNAs, or mRNAs. Given a handful of related RNA sequences, reliable methods now exist to predict conserved functional RNA structures within these RNAs. Because of their small size and fast evolution, the genomes of RNA viruses supply fertile ground for such approaches, and databases of functional viral RNA structures are being built. These functional RNA motifs in the viral genome are just as essential as the encoded proteins, and thus just as promising targets for development of drugs and vaccines (90,151).

The importance of regulatory functions mediated by RNA has only now found more attention through recent studies on the phenomenon of RNA interference (11,45,49). A recent study (138) showed, furthermore, that non-coding RNA motives may act as potent "danger motifs" that trigger an adaptive immune

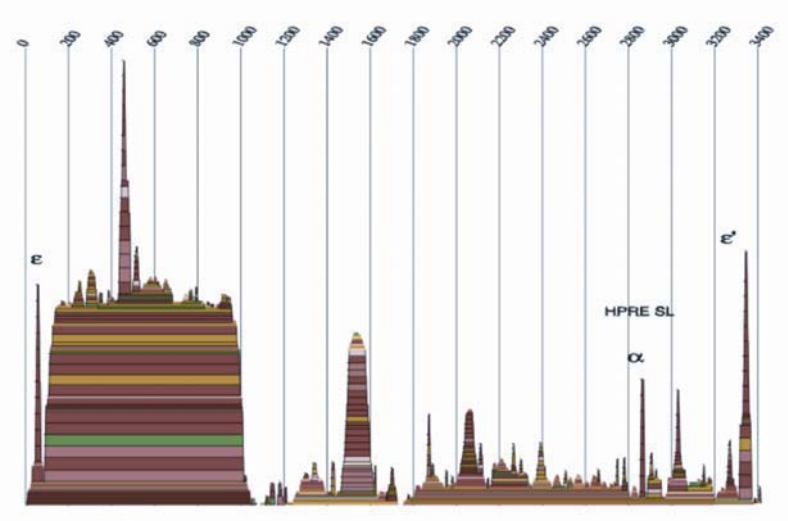


Figure 3. Predicted functional RNA structures in the genome of Hepatitis B virus. The function of the ε, ε', and α elements of the HPRE region have been determined experimentally. The prediction suggests several new conserved structures with unknown function. In the "mountain representation" each base pair (i,j) is represented by a bar from i to j. The thickness of the bar indicates its probability or the reliability of the prediction. A color scheme can be used to indicate sequence covariations. Hue encodes the number of compensatory and consistent mutations, while reduced saturation indicates that a small number of sequences is inconsistent with the structure.

system via innate immune receptors. RNA structure thus receives increased attention in molecular medicine.

A comprehensive understanding of the biology of a cell will ultimately require the knowledge of all encoded RNAs, the molecules with which they interact, and the molecular structures of these complexes (18). Various approaches to surveying genomic sequences for putative RNA genes have been devised in the last few years.

Structure-based searches use the known secondary structure of the major classes of functional RNAs. Programs such as `RNAmot` (34), `tRNAscan` (83), `HyPa` (163), `RNAMotif` (88), `bruce` (76), and many others exploit this avenue. An interesting variant that makes use of evolutionary computation is described by (165). Nevertheless, all these approaches are restricted to searching for new members of the few well-established families. The web-based resource `RNAGENiE` uses a neural network that has been trained on a wide variety of functional RNAs (166). It is capable of detecting a wider variety of functional RNAs.

Some non-coding RNAs can be found by searching for likely transcripts that do not contain an open reading frame. A survey of the *E. coli* genome for DNA regions that contain a σ70 promotor within a short distance of a *Rho*-independent terminator, for instance, resulted in 144 novel possible ncRNAs (164). This approach is limited, however, to functional RNAs that are transcribed in the "usual" manner.

Comparative approaches such as the QRNA program (113) can detect novel structural RNA genes in a pair of aligned homologous sequences by deciding whether the substitution pattern fits better with (a) synonymous substitutions, which are expected in protein-coding regions, (b) the compensatory mutations consistent with some base-paired secondary structure, or (c) uncorrelated mutations.

Another approach tries to determine functional RNAs by means of structure prediction. The basic assumption is that functional and hence conserved structures will be thermodynamically more stable (64,78). While such procedures are capable of detecting some particularly stable features, a recent study (110) concludes that "although a distinct, stable secondary structure is undoubtedly important in most non-coding RNAs, the stability of most noncoding RNA secondary structures is not sufficiently different from the predicted stability of a random sequence to be useful as a general genefinding approach." Nevertheless, in some special cases such as hyperthermophilic organisms, GC-content (and hence thermodynamic stability) proved sufficient (74).

Since most classes of functional RNAs are relatively well conserved while their sequences show little similarities, both comparative procedures and search in single sequences have to rely on structural information. While the prediction of RNA tertiary structures faces much the same problems as protein structure prediction, efficient algorithms exist for handling RNA secondary structure. As we have seen, these methods provide powerful tools for computational studies of RNA structure.

6. ACKNOWLEDGMENTS

This work is supported by the Austrian *Fonds zur Föderung der Wissenschaftlichen Forschung*, Project Nos. P-13545-MAT and P-15893, and the *DFG* Bioinformatics Initiative.

7. REFERENCES

1. Akutsu, T. 2001. Dynamic programming algorithms for RNA secondary structure prediction with pseudoknots. *Discr Appl Math* **104**:45–62.
2. Avner P, Heard E. 2001. X-chromosome inactivation: counting, choice, and initiation. *Nature Rev Genet* **2**:59–67.

3. Babajide A, Farber R, Hofacker IL, Inman J, Lapedes AS, Stadler PF, 2001. Exploring protein sequence space using knowledge-based potentials. *J Theor Biol* **212**:35–46.
4. Babajide A, Hofacker IL, Sippl MJ, Stadler PF. 1997. Neutral networks in protein space: a computational study based on knowledge-based potentials of mean force. *Folding Des* **2**:261–269.
5. Biebricher CK, Gardiner WC. 1997. Molecular evolution of RNA *in vitro*. *Biophys Chem* **66**:179–192.
6. Blackburn E. 1999. Telomerase. In *The RNA world*, pp. 609–635. Ed. R Gesteland, T Cech, J Atkins. Cold Spring Harbor Laboratory Press, New York.
7. Bosher JM, Labouesse M. 2000. RNA interference: genetic wand and genetic watchdog. *Nature Cell Biol* **2**:E31–E36.
8. Brown JW, The ribonuclease P database. 1999. *Nucleic Acids Res* **27**:314–314.
9. Bruccoleri RE, Heinrich G. 1988. An improved algorithm for nucleic acid secondary structure display. *Comput Appl Biosci* **4**:167–173.
10. Chetouani F, Monestie P, Thebault P, Gaspin C, Michot B. 1997. ESSA: an integrated and inteactive computer tool for analysing RNA structure. *Nucleic Acids Res* **25**:3514–3522.
11. Cogoni C., Macino G. 2000. Post-transcriptional gene silencing across kingdoms. *Genes Dev* **10**:638–643.
12. ten Dam EB, Pleij CW, Bosch L. 1990. RNA pseudoknots: translational frameshifting and readthrough on viral RNAs. *Virus Genes* **4**:121–135.
13. ten Dam EB, Pleij K, Draper D. 1992. Structural and functional aspects of RNA pseudoknots. *Biochemistry* **31**:11665–11676.
14. ten Dam EB, Verlaan PW, Pleij CW. 1995. Analysis of the role of the pseudoknot component in the SRV-1 gag-pro ribosomal frameshift signal: loop lengths and stability of the stem regions. *RNA* **1**:146–154.
15. Dandekar T, Hentze MW. Finding the hairpin in the haystack: searching for RNA motifs. *Trends Genet* **11**:45–50.
16. Dayton ET, Konings DAM, Powell DM, Shapiro BA, and LB, Maizel JV, Dayton AI. 1992. Extensive sequence-specific information throughout the CAR/RRE the target sequence of the human immunodeficiency virus type 1 Rev protein. *J Virol* **66**:1139–1151.
17. Denduangboripant J, Cronk QCB. 2001. Evolution and alignment of the hypervariable arm 1 of *Aeschynanthus* (Gesneriaceae) ITS2 nuclear ribosomal DNA. *Mol Phylog Evol* **20**:163–172.
18. Doudna JA. 2000. Structural genomics of RNA. *Nature Struct Biol* **7**:954–956.
19. Du Z, Holland JA, Hansen MR, Giedroc DP, Hoffman DW. 1997. Base-pairings within the RNA pseudoknot associated with the simian retrovirus-1 gag-pro frameshift site. *J Mol Biol* **270**:464–470.
20. Eddy SR. 2001. Non-coding RNA genes and the modern RNA world. *Nature Genet* 2:919–929.
21. Edmonds J. 1965. Maximum matching and a polyhedron with 0,1-vertices. *J Res Natl Bur Stand* **69B**:125–130.
22. Erdmann VA, Barciszewska MZ, Hochberg A, de Groot N, Barciszewski J. 2001. Regulatory RNAs. *Cell Mol Life Sci* **58**:960–977.
23. Erdmann VA, Szymanski M, Hochberg A. de Groot N, Barciszewski J. Collection of mRNA-like non-coding RNAs. *Nucleic Acids Res* **27**:192–195.
24. Felciano RM, Chen RO, Altmann RB. 1997. RNA Secondary strucutre as a reusable interface to biological information resources. *Gene* **190**:GC59–GC70.
25. Flamm C, Fontana W, and IH, Schuster P. 2000. RNA folding kinetics at elementary step resolution. *RNA* **6**:325–338.
26. Flamm C, Hofacker IL, Stadler PF, Wolfinger MT. 2002. Barrier Trees of Degenerate Landscapes. *Z Phys Chem* **216**:155–173.

27. Fontana W, Stadler PF, Bornberg-Bauer EG, Griesmacher T, Hofacker IL, Tacker M, Tarazona P, Weinberger ED, Schuster P. 1993. RNA folding landscapes and combinatory landscapes. *Phys Rev E* **47**:2083–2099.
28. Fontana W, Konings DAM, Stadler PF, Schuster P. 1993. Statistics of RNA secondary structures. *Biopolymers* **33**:1389–1404.
29. Fontana W, Schuster P. 1998. Continuity in evolution: on the nature of transitions. *Science* **280**:1451–1455.
30. Fontana W, Schuster P. 1998. Shaping space: the possible and the attainable in RNA genotype-phenotype mapping. *J Theor Biol* **194**:491–515.
31. Franke A, Baker BS. 2000. Dosage compensation rox! *Curr Opin Cell Biol* **12**:351–354.
32. Freier SM, Kierzek R, Jaeger JA, Sugimoto N, Caruthers MH, Neilson T, Turner DH. 1986. Improved free-energy parameters for prediction of RNA duplex stability. *Proc Natl Acad Sci USA* **83**:9373–9377.
33. Gabow HN. 1973. *Implementations of algorithms for maximum matching on nonbipartite graphs*. Stanford University, Department of Computer Science.
34. Gautheret D, Major F, Cedergren R. 1990. Pattern searching/alignment with RNA primary and secondary structures: an effective descriptor for tRNA. *Comput Appl Biosci* **6**:325–331.
35. Gavrilets S. 1997. Evolution and speciation on holey adaptive landscapes. *Trends Ecol Evol* **12**:307–312.
36. Giegerich, R, Reeder, J. 2003. *From RNA folding to thermodynamic matching including pseudoknots*. Report 2003-03, Universität Bielefeld, Germany.
37. Gorodkin J, Knudsen B, Zwieb C, Samuelsson T. 2001. SRPDB (signal recognition particle database) *Nucleic Acids Res* **29**:169–170.
38. Gorodkin J, Heyer LJ, Stormo GD. 1997. Finding common sequences and structure motifs in a set of RNA molecules. In *Proceedings of the ISMB97*, pp. 120–123. Ed. T Gaasterland, P Karp, K Karplus, Ch Ouzounis, Ch Sander, A Valencia. AAAI Press, Menlo Park, CA.
39. Gorodkin J, Heyer LJ, Stormo GD. 1997. Finding the most significant common sequence and structure motifs in a set of RNA sequences. *Nucleic Acids Res* **25**:3724–3732.
40. Green MR. 1991. Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. *Annu Rev Cell Biol* **7**:559–599.
41. Griffiths-Jones S, Bateman A, Mhairi M, Khanna A, Eddy SR. 2003. Rfam: an RNA family database. *Nucleic Acids Res* **31**:439–441.
42. Grüner W, Giegerich R, Strothmann D, Reidys CM, Weber J, Hofacker IL, Stadler PF, Schuster P. 1996. Analysis of RNA sequence structure maps by exhaustive enumeration. I. Neutral networks. *Monatsh Chem* **127**:355–374.
43. Grüner W, Giegerich R, Strothmann D, Reidys CM, Weber J, Hofacker IL, Stadler PF, Schuster P. 1996. Analysis of RNA sequence structure maps by exhaustive enumeration. II. Structures of neutral networks and shape space covering. *Monatsh Chem* **127**:375–389.
44. Gultyaev AP, van Batenburg FHD, Pleij CWA. 1999. An approximation of loop free energy values of RNA H-pseudoknots. *RNA* **5**:609–617.
45. Guru T. 2000. A silence that speaks volumes. *Nature* **404**:804–808.
46. Gutell RR, Cannone JJ, Shang Z, Du Y, Serra MJ. 2000. A story: unpaired adenosine bases in ribosomal RNA. *J Mol Biol* **304**:335–354.
47. Gutell RR. 1993. Evolutionary characteristics of RNA: inferring higher-order structure from patterns of sequence variation. *Curr Opin Struct Biol* **3**:313–322.
48. The Genome Sequencing Consortium. 2001. Gene content of the human genome. *Nature* **409**:860–921.
49. Hammond SM, Caudy AA, Hannon GJ. 2001. Post-transcriptional gene silencing by double-stranded RNA. *Nature Rev Genet* **2**:110–119.
50. Han K, Kim HJ. 1993. Prediction of common folding structures of homologous RNAs. *Nucleic Acids Res* **21**:1251–1257.

51. Han K, Kim D, Kim HJ. 1999. A vector-based method for drawing RNA secondary structure. *Bioinformatics* **15**:286–297.
52. Haslinger C. 2001. *Prediction algorithms for restricted RNA pseudoknots*. PhD dissertation, University of Vienna, Faculty of Sciences.
53. Höbartner C, Micura R. 2003. Bistable secondary structures of small RNAs and their structural probing by comparative imino proton NMR spectroscopy. *J Mol Biol* **325**:421–431.
54. Höchsmann M. 2001. Tree and forrest alignments: an algebraic dynamic programming approach for aligning trees and forests. MSc thesis. University of Bielefeld, Germany.
55. Hofacker IL, Fekete M, Stadler PF. 2002. Secondary structure prediction for aligned RNA sequences. *J Mol Biol* **319**:1059–1066.
56. Hofacker IL, Fontana W, Stadler PF, Bonhoeffer S, Tacker M, Schuster P. 1994. Fast folding and comparison of RNA secondary structures. *Monatsh Chem* **125**:167–188.
57. Hofacker IL, Huynen MA, Stadler PF, Stolorz PE. 1996. Knowledge discovery in RNA sequence families of HIV using scalable computers. In *Proceedings of the 2nd international conference on knowledge discovery and data mining, Portland, OR*, pp. 20–25. Ed. E Simoudis, J Han, U Fayyad. AAAI Press, Menlo Park, CA.
58. Hofacker IL, and PS, Stadler PF. 1998. Combinatorics of RNA secondary structures. *Discr Appl Math* **89**:177–207.
59. Hofacker IL, Fekete M, Flamm C, Huynen MA, Rauscher S, Stolorz PE, Stadler PF. 1998. Automatic detection of conserved RNA structure elements in complete RNA virus genomes. *Nucleic Acids Res* **26**:3825–3836.
60. Hofacker IL, Stadler PF. 1999. Automatic detection of conserved base pairing patterns in RNA virus genomes. *Comput Chem* **23**:401–414.
61. Hogeweg P, Hesper B. 1984. Energy directed folding of RNA sequences. *Nucleic Acids Res* **12**:67–74.
62. Huez I, Créancier L, Audigier S, Gensac M-C, Prats A-C, Prats H. 1998. Two independent internal ribosome entry sites are involved in translation initiation of vascular endothelial growth factor mRNA. *Mol Cell Biol* **18**:6178–6190.
63. Huynen MA, Stadler PF, Fontana W. 1996. Smoothness within ruggedness: the role of neutrality in adaptation. *Proc Natl Acad Sci USA* **93**:397–401.
64. Huynen MA, Perelson AS, Viera WA, Stadler PF. 1996. Base pairing probabilities in a complete HIV-1 RNA. *J Comput Biol* 3:253–274.
65. Huynen MA. 1996. Exploring phenotype space through neutral evolution. *J Mol Evol* **43**:165–169.
66. Isambert H, Siggia ED. 2000. Modeling RNA folding paths with pseudoknots: application to hepatitis delta virus ribozyme. *Proc Natl Acad Sci USA* **97**:6515–6520.
67. Jacobs GH, Rackham O, Stockwell PA, Tate W, and Brown CM. 2002. Transterm: a database of mRNAs and translational control elements. *Nucleic Acids Res* **30**:310–311.
68. Jacobson AB, Zuker M. 1993. Structural analysis by energy dot plot of large mRNA. *J Mol Biol* **233**:261–269.
69. Jaeger JA, Turner DH, Zuker M. 1989. Improved predictions of secondary structures for RNA. *Proc Natl Acad Sci USA* **86**:7706–7710.
70. Juan V, Wilson C. 1999. RNA secondary structure prediction based on free energy and phylogenetic analysis. *J Mol Biol* **289**:935–947.
71. Keefe AD, Szostak JW. 2001. Functional proteins from a random-sequence library. *Nature* **410**:715–718.
72. Keiler KC, Shapiro L, Williams KP. 2000. tmRNAs that encode proteolysis-inducing tags are found in all known bacterial genomes: A two-piece tmRNA functions in *Caulobacter*. *Proc Natl Acad Sci USA* **97**:7778–7783.
73. Kidd-Ljunggren K, Zuker M, Hofacker IL, Kidd AH. 2000. The hepatitis B virus pregenome: prediction of RNA structure and implications for the emergence of deletions. *Intervirology* **43**:154–164.

74. Klein RJ, Misulovin Z, Eddy SR. 2002. Noncoding RNA genes identified in AT-rich hyperthermophiles. *Proc Natl Acad Sci USA* **99**:7542–7547.
75. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. 2001. Identification of novel genes coding for small expressed RNAs. *Science* **294**:853–857.
76. Laslett D, Canback B, Andersson S. 2002. BRUCE: a program for the detection of transfer-messenger RNA genes in nucleotide sequences. *Nucleic Acids Res* **30**:3449–3453.
77. Lau NC, Lim LP, Weinstein EG, Bartel DP. 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294**:858–862.
78. Le S-Y, Chen J-H, Currey KM, Maizel JV. 1988. A program for predicting significant RNA secondary structures. *CABIOS* **4**:153–159.
79. Le SY, Zuker M. 1991. Predicting common foldings of homologous RNAs. *J Biomol Struct Dyn* **8**:1027–1044.
80. Lee RC, Ambros V. 2001. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* **294**:862–864.
81. Lee D, Han K. 2002. Prediction of RNA pseudoknots: comparative study of genetic algorithms. *Genome Informatics* **13**:414–415.
82. Leydold J, Stadler PF. 1998. Minimal cycle basis, outerplanar graphs. *Elec J Comb* **5**:R16 (see http://www.combinatorics.org).
83. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**:955–964.
84. Lück R, Steger G, Riesner D. 1996. Thermodynamic prediction of conserved secondary structure: application to the RRE element of HIV, the tRNA-like element of CMV, and the mRNA of prion protein. *J Mol Biol* **258**:813–826.
85. Lück R, Gräf S, Steger G. 1999. ConStruct: A tool for thermodynamic controlled prediction of conserved secondary structure. *Nucleic Acids Res* **27**:4208–4217.
86. Lyngsö RB, Pedersen CNS. 2000. RNA pseudoknot prediction in energy-based models. *J Comput Biol* **7**:409–427.
87. Macdonald PM. 1990. Bicoid mRNA localization signal: phylogenetic conservation of function and RNA secondary structure. *Development* **110**:161–171.
88. Macke TJ, Ecker DJ, Gutell RR, Gautheret D, Case DA, Sampath R. 2001. RNAMotif, an RNA secondary structure definition and search algorithm. *Nucleic Acids Res* **29**:4724–4735.
89. Maidak BL, Cole JR, Lilburn TG, Parker Jr CT, Saxman PR, Farris RJ, Garrity GM, Olsen GJ, Schmidt TM, Tiedje JM. 2001. The RDP-II (ribosomal database project). *Nucleic Acids Res* **29**:173–174.
90. Mandl CW, Holzmann H, Meixner T, Rauscher S, Stadler PF, Allison SL, Heinz FX. 1998. Spontaneous and engineered deletions in the 3'-noncoding region of tick-borne encephalitis virus: construction of highly attenuated mutants of Flavivirus. *J Virol* **72**:2132–2140.
91. Mandl CW, Aberle JH, Aberle SW, Holzmann H, Allison SL, Heinz FX. 1998. In vitro-synthesized infectious RNA as an attenuated live vaccine in a flavivirus model. *Nature Med* **4**:1438–1440.
92. Mathews DH, Sabina J, Zuker M, Turner H. 1999. Expanded sequence dependence of thermodynamic parameters provides robust prediction of RNA secondary structure. *J Mol Biol* **288**:911–940.
93. Mattick JS. 1994. Introns: evolution and function. *Curr Opin Genet Dev* **4**:823–831.
94. Matzke M, Matzke, AJM, Kooter JM. 2001. RNA: guiding gene silencing. *Science* **293**:1080–1083.
95. McCaskill JS. 1990. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers* **29**:1105–1119.
96. Meyer C, Giegerich R. 2002. Matching and significance evaluation of combined sequence/structure motifs in RNA. *Z Phys Chem* **216**:193–216.
97. Muller G, Gaspin C, Etienne A, Westhof E. 1993. Automatic display of RNA secondary structures. *Comput Appl Biosci* **9**:551–561.

98. Nussinov R, Piecznik G, Griggs JR, Kleitman DJ. 1978. Algorithms for loop matching. *SIAM J Appl Math* **35**:68–82.
99. Nussinov R, Jacobson AB. 1980. Fast algorithm for predicting the secondary structure of single-stranded RNA. *Proc Natl Acad Sci USA* **77**:6309–6313.
100. Ohno M, Mattaj IW. 1999. Meiosis: MeiRNA hits the spot. *Curr Biol* **28**:R66–R69.
101. Oleynikov Y, Singer RH. 1998. RNA localization: different zipcodes, same postman? *Trends Cell Biol* **8**:381–383.
102. Omer AD, Lowe TM, Russel AG, Ebhardt H, Eddy SR, Dennis P. 2000. Homologs of small nucleolar RNAs in Archaea. *Science* **288**:517–522.
103. Van de Peer Y, De Rijk P, Wuyts J, Winkelmans T, De Wachter R. 2000. The European small subunit ribosomal RNA database. *Nucleic Acids Res* **28**:175–176.
104. Perochon-Dorisse J, Chetouani F, Aurel S, Iscolo N, Michot B. 1995. RNA-d2: a computer programm for editing and display of RNA secondary structures. *Bioinformatics* **11**:101–109.
105. Pesole G, Mignone F, Gissi C, Grillo Ga, Licciulli F, Sabino L. 2001. Structural and functional features of eukaryotic mRNA untranslated regions. *Gene* **276**:73–81.
106. Pleij CW. 1995. Structure and function of RNA pseudoknots. *Genet Eng* **17**:67–80.
107. Rauscher S, Flamm C, Mandl C, Heinz FX, Stadler PF. 1997. Secondary structure of the 3'-non-coding region of Flavivirus genomes: comparative analysis of base pairing probabilities. *RNA* **3**:779–791 (Santa Fe Institute Preprint 97-02-010).
108. Reidys C, Stadler PF, Schuster P. 1997. Generic properties of combinatory maps: neutral networks of RNA secondary Structure. *Bull Math Biol* **59**:339–397.
109. De Rijk P, De Wachter R. 1997. RnaViz, a program for the visualization of RNA secondary structure. *Nucleic Acids Res* **25**:4679–4684.
110. Rivas E, Eddy SR. 2000. Secondary structure alone is generally not statistically significant for the detection of noncoding RNAs. *Bioinformatics* **16**:583–605.
111. Rivas E, Eddy SR. 2000. The language of RNA: a formal grammar that includes pseudoknots. *Bioinformatics* **16**:334–340.
112. Rivas E, Klein RJ, Jones TA, Eddy SR. 2001. Computational identification of noncoding RNAs in *E. coli* by comparative genomics. *Curr Biol* **11**:1369–1373.
113. Rivas E, Eddy SR. 2001. Noncoding RNA gene detection using comparative sequence analysis. *BMC Bioinformatics* **2**(8):19 pages.
114. Rivas E, Eddy SR. 1999. A dynamic programming algorithm for RNA structure prediction including pseudoknots. *J Mol Biol* **285**:2053–2068.
115. Rueckert RR. 1996. Picornaviridae: the viruses and their replication. In *Virology*, 3rd ed., pp. 609–654. Ed. NR Fields, DM Knipe, PM Howley. Lippincott-Raven, Philadelphia.
116. Samarsky DA, Fournier MJ. 1999. A comprehensive database for the small nucleolar RNAs from Saccharomyces cerevisiae. *Nucleic Acids Res* **27**:161–164.
117. Sankoff D. 1985. Simultaneous solution of the RNA folding alignment, and proto-sequence problems. *SIAM J Appl Math* **45**:810–825.
118. SantaLucia Jr. J, Turner DH. 1997. Measuring the thermodynamics of RNA secondary structure formation. *Biopolymers* **44**:309–319.
119. Schmitz M, Steger G. 1992. Base-pair probability profiles of RNA secondary structures. *Comput Appl Biosci* **8**:389–399.
120. Schultes EA, Bartel DP. 2000. One sequence, two ribozymes: implications for the emergence of new ribozyme folds. *Science* **289**:448–452.
121. Schuster P. 2001. Evolution *in silico* and *in vitro*: The RNA model. *Biol Chem* **382**:1301–1314.
122. Schuster P, Fontana W, Stadler PF, Hofacker IL. 1994. From sequences to shapes and back: a case study in RNA secondary structures. *Proc Roy Soc Lond. B* **255**:279–284.
123. Schuster P. 1995. How to search for RNA structures: theoretical concepts in evolutionary biotechnology. *J Biotechnol* **41**:239–257.

124. Schuster P, Stadler PF. 1998. Sequence redundancy in biopolymers: a study on RNA and protein structures. In *Viral regulatory structures*, pp. 163–186. Ed. G. Myers. Santa Fe Institute Studies in the Sciences of Complexity, Vol. 28. Addison-Wesley, Reading MA.
125. Shapiro BS. 1988. An algorithm for comparing multiple RNA secondary structures. *CABIOS* **4**:387–393.
126. Shapiro B, Zhang K. 1990. Comparing multiple RNA secondary structures using tree comparisons. *CABIOS* **6**:309–318.
127. Sousa C, Johansson C, Charon C, Manyani H, Sautter C, Kondorosi A, Crespi M. 2001. Translational and structural requirements of the early nodulin gene enod40, a short-open reading frame-containing RNA, for elicitation of a cell-specific growth response in the alfalfa root cortex. *Mol Cell Biol* **21**:354–366.
128. Sprinzl M, Horn C, Brown M, Ioudovitch A, Steinberg S. 1998. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res* **26**:148–153.
129. Stocsits R, Hofacker IL, Stadler PF. 1999. Conserved secondary structures in hepatitis B virus RNA. *Computer science in biology*, pp. 73–79. Proceedings of the GCB'99, Hannover, D. University of Bielefeld, Germany.
130. Stoss O, Stoilov P, Daoud R, Hartmann AM, Olbrich M, Stamm S. 2000. Misregulation of premRNA splicing that causes human diseases: concepts and therapeutic strategies. *Gene Therapy Mol Biol* **5**:9–30.
131. Sung D, Kang H. 1998. Mutational analysis of the RNA pseudoknot involved in efficient ribosomal frameshifting in simian retrovirus-1. *Nucleic Acids Res* **26**:1369–1372.
132. Szymanski M, Barciszewska MZ, Barciszewski J, Erdmann VA. 2000. 5S ribosomal RNA database Y2K. *Nucleic Acids Res* **28**:166–167.
133. Tabaska JE, Stormo GD. 1997. Automated alignment of RNA sequences to pseudoknotted structures. In Proceedings of the ISMB-97, pp. 311–318. Ed. T Gaasterland, P Karp, K Karplus, Ch Ouzounis, Ch Sander, A Valencia. AAAI Press, Menlo Park, CA.
134. Tabaska JE, Cary RB, Gabow HN, Stormo GD. 1998. An RNA folding method capable of identifying pseudoknots and base triples. *Bioinformatics* **14**:691–699.
135. Thompson JD, Higgs DG, Gibson TJ. 1994. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties, and weight matrix choice. *Nucleic Acids Res* **22**:4673–4680.
136. Thurner C, Witwer C, Hofacker I, Stadler PF. 2004. Conserved RNA secondary structures in Flaviviridae genomes. *J Gen Virol* **85**:1113–1124.
137. Walter AE, Turner DH, Kim J, Lyttle MH, Müller P, Mathews DH, Zuker M. 1994. Co-axial stacking of helixes enhances binding of oligoribonucleotides and improves predicions of RNA folding. *Proc Natl Acad Sci USA* **91**:9218–9222.
138. Wang L, Smith D, Bot S, Dellamary L, Bloom A, Bot A. 2002. Noncoding RNA danger motifs bridge innate and adaptive immunity and are potent adjuvants for vaccination. *J Clin Invest* **110**:1175–1184.
139. Wang L, Jiang T, Zhang K. 1995. Alignment of trees: an alternative to tree edit. *Theor Comput Sci* **143**:137–148.
140. Waterman MS. 1978. Secondary structure of single-stranded nucleic acids. In *Studies on foundations and combinatorics*, pp. 167–212. Advances in mathematics supplementary studies. Academic Press, New York.
141. Waterman MS, Smith TF. 1978. RNA secondary structure: A complete mathematical analysis. *Math Biosci* **42**:257–266.
142. Waterman MS. 1978. Combinatorics of RNA hairpins and cloverleaves. *Stud Appl Math* **60**:91–96.
143. Waterman MS, Gordon L, Arratia R. 1987. Phase transition in sequence matches and nucleic acid structure. *Proc Natl Acad Sci USA* **84**:1239–1243.
144. Waterman MS. 1995. *Introduction to computational biology: maps sequences, and genomes.* Chapman & Hall, London.

145. Westhof E, Jaeger L. 1992. RNA pseudoknots. *Curr Opin Struct Biol* **2**:327–333.
146. Williams KP. 2002. The tmRNA Website: invasion by an intron. *Nucleic Acids Res* **30**:179–182.
147. Wilson DS, Szostak JW. 1999. *In vitro* selection of fuctional nucleic acids. *Annu Rev Biochem* **68**:611–647.
148. Witwer C, Rauscher S, Hofacker IL, Stadler PF. 2001. Conserved RNA secondary structures in Picornaviridae genomes. *Nucleic Acids Res* **29**:5079–5089.
149. Wuchty S, Fontana W, Hofacker IL, Schuster P. 1999. Complete Suboptimal folding of RNA and the stability of secondary structures. *Biopolymers* **49**:145–165.
150. Wuyts J, De Rijk P, Van de Peer Y, Winkelmans T, De Wachter R. 2001. The European large subunit ribosomal RNA database. *Nucleic Acids Res* **29**:175–177.
151. Ying H, Zaks TZ, Wang R-F, Irvine KR, Kammula US, Marincola FM, Leitner WW, Restifo NP. 1999. Cancer therapy using a self-replicating RNA vaccine. *Nature Med* **5**:823–827.
152. van Zon A, Mossink MH, Schoester M, Scheffer GL, Scheper RJ, Sonneveld P, Wiemer EA. 2001. Multiple human vault RNAs: expression and association with the vault complex. *J Biol Chem* **276**:37715–37721.
153. Zuker M, Stiegler P. 1981. Optimal computer folding of larger RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res* **9**:133–148.
154. Zuker M, Sankoff D. 1984. RNA secondary structures and their prediction. *Bull Math Biol* **46**:591–621.
155. Zuker M. 1989. On finding all suboptimal foldings of an RNA molecule. *Science* **244**:48–52.
156. Zuker M, Somorjai RL. 1989. The alignment of protein structures in three dimensions. *Bull Math Biol* **51**:55–78.
157. Zuker M. 1989. The use of dynamic programming algorithms in RNA secondary structure prediction. In *Mathematical methods for DNA sequences*, pp. 159–184. Ed. MS Waterman. CRC Press, Boca Raton, FL.
158. Zuker M, Jaeger JA, Turner DH. 1991. A comparison of optimal and suboptimal RNA secondary structures predicted by free energy minimization with structures determined by phylogenetic comparison. *Nucleic Acids Res* **19**:2707–2714.
159. Zuker M. 1991. Suboptimal sequence alignment in molecular biology. *J Mol Biol* **221**:403–420.
160. Zwieb C, Wower J. 2000. tmRDB (tmRNA database). *Nucleic Acids Res* **28**:169–170.
161. Zwieb C. 1996. The uRNA database. *Nucleic Acids Res* **24**:76–79.
162. d'Aubenton Carafa Y, Brody E, Thermes C. 1990. Prediction of rho-independent *Escherichia coli* transcription terminators: a statistical analysis of their RNA stem-loop structures. *J Mol Biol* **216**:835–858.
163. Gräf S, Strothmann D, Kurtz S, Steger G. 2001. HyPaLib: a database of RNAs and RNA structural elements defined by hybrid patterns. *Nucleic Acids Res* **29**:196–198.
164. Chen S, Lesnik EA, Hall TA, Sampath R, Griffey RH, Ecker DJ, Blyn LB. 2002. A bioinformatics based approach to discover small RNA genes in the *Escherichia coli* genome. *BioSystems* **65**:157–177.
165. Fogel GB, Porto VW, Weekes DG, Fogel DB, Griffey RH, McNeil JA, Lesnik E, Ecker DJ, Sampath R. 2002. Discovery of RNA structural elements using evolutionary computation. *Nucleic Acids Res* **30**:5310–5317.
166. Carter RJ, Dubchak I, Holbrook SR. 2001. A computational approach ot indentify genes for functional RNAs in genomic sequences. *Nucleic Acids Res* **29**:3928–3938.

1.3

PROTEIN NETWORKS

Andreas Wagner

University of New Mexico and The Santa Fe Institute,
Albuquerque, New Mexico

This introductory survey to protein networks and their biomedical applications has three parts. First, I introduce some of the most promising and widely used experimental approaches to study protein expression on a genome-wide scale. Second, I introduce experimental approaches to identify physical protein interactions on a genome-wide scale. Finally, I present some examples of how such network information might be profitably used in medical applications.

1. INTRODUCTION

Genome-wide approaches to study the living are poised to revolutionize medicine in at least three areas: diagnosis, toxicology, and drug development. The example of drug development illustrates vividly the important role that proteins will play in this revolution. Not only are the functional products of most genes proteins, the targets for the vast majority of known pharmaceuticals are also proteins. To develop new drugs thus is to identify agents that can either block or mimic the action of key proteins in an organism's or a cell's life cycle, be it a pathogen, a cancerous cell, a cell necessary for an immune response or a cell important in some aspect of regeneration. However, efficiency is only one aspect of a good pharmaceutical, specificity being equally necessary. Ideally, a drug should interact with its intended target (protein), and only with this target.

Address correspondence to: Andreas Wagner, Department of Biology, University of New Mexico, 167A Castetter Hall, Albuquerque, NM 817131-1091 (wagnera@unm.edu).

The less specifically a drug acts, the more wide-ranging will its side effects be. Most assays aiming at characterizing the specificity of a drug rely either on organismal-level observations or on the activity of individual macromolecules. Few assays are available to explore the entire spectrum of molecular interactions a candidate pharmaceutical might engage in. Genomics, and specifically the analysis of protein networks, holds the promise to change that.

The word "network" evokes the image of objects (organisms, cells, proteins, genes, etc.) that interact with each other, usually in some complex fashion. However, browsing through the experimental literature in functional genomics, one quickly realizes that researchers in the field seem to have adopted a much broader sense of the word. Many studies in this area bear the word "network" in the title, but the studies do not identify interacting objects, be they genes or proteins. Instead, many of these studies provide circumstantial evidence—mostly through gene expression—for gene products that might be part of a network. That is, they identify genes that are expressed—either as mRNA or as protein—at similar levels during a particular part of an organism's life cycle, in a particular environmental condition, in a mutant strain, during development, or in a diseased organ. Co-occurrence of gene products under such conditions is then taken as an indication that the gene products may be part of a network that is "active" under the particular condition studied. I will use the notion of a protein network here in both the stronger sense above and in this weaker sense. The development of techniques to identify networks in the weak sense is more advanced, as are their medical applications.

This introductory survey of protein networks and their biomedical applications has three parts. First, I introduce some of the most promising approaches to study protein expression on a very large scale, that is, to identify protein networks in the weak sense. Second, I introduce experimental approaches to identify direct protein interactions on a genome-wide scale, that is, to identify protein networks in the strong sense. Finally, I present some examples of how network information can be profitably used in medical applications. At this point in time, most pertinent and publicly available experiments aim to provide a proof of principles rather than fully mature clinical applications. The literature I cite is by no means exhaustive, but intended to provide guidance for further reading.

2. LARGE-SCALE APPROACHES TO IDENTIFY PROTEIN EXPRESSION

The approach with the longest track record of identifying expressed proteins on a genome-wide large scale is two-dimensional protein electrophoresis. A second approach, mass-spectroscopic analysis of complex protein mixtures, holds great promise for the future. Both approaches identify individual proteins

using mass spectroscopy, whose pertinent features I will briefly review, especially because these features are also important in experimental techniques to identify protein interactions. For more thorough surveys see (1,24).

The principle of mass spectroscopy is that ions of different mass and charge travel at different velocity and with different trajectories through an electromagnetic field. Mass spectroscopy serves to identify the chemical composition of an unknown sample containing protein or other molecules. A mass spectrometer consists of three parts: an ionization source, an analyzer, and a detector. The ionization source is responsible for ionizing the sample, and the analyzer separates the ions in an electromagnetic field according to their mass-to-charge (m/z) ratio. (Ions of the same m/z ratio cannot be distinguished.) The detector collects the ions and records a spectrum displaying the abundance of ions of a particular m/z ratio. There are multiple approaches to implementing ionization, analysis, and detection, as well as to combining them into a spectrometer. I will mention two approaches that are particular prominent in analyzing proteins.

MALDI-TOF mass spectrometers combine Matrix-Assisted Laser Desorption Ionization (MALDI) with a time-of-flight (TOF) analyzer. MALDI is a "soft" ionization method that does not cause fragmentation of molecules in the sample. In MALDI, a sample is mixed with a matrix compound such as sinapinic acid. The mixture of sample and matrix is then exposed to laser light, which the matrix compound converts into excitation energy that ionizes the sample, such that it can enter the analyzer. The time of flight analyzer first accelerates ions in a well-defined electric field and then measures their velocity, which is characteristic for ions of given m/z. In the typical MALDI-TOF application of identifying a protein in an organism with a fully sequenced genome, one would first digest the protein enzymatically into short peptides, and then obtain a mass spectrum of the resulting mixture of peptide fragments, a so-called peptide map or mass map. The amino-acid sequence of a short peptide can often be uniquely determined from its mass. Together with the masses of other peptides from the same protein, and with the known location of genomic DNA sequences that have the nucleotide sequence necessary to encode the peptide(s), one can determine not only the complete amino-acid sequence of the protein, but also the location of its coding gene.

A second commonly used type of mass spectrometry is tandem mass spectrometry (MS-MS). MS-MS uses two analyzers. One method of combining these two analyzers, daughter ion scanning, is of particular importance for determining the sequence of peptides in complex protein mixtures. In this approach, ions pass through a first analyzer, upon which the user (or a computer) selects ions of a particular molecular weight of interest. These ions are then fragmented by bombardment with a gas in a process called collision-induced dissociation. (Collision-induced dissociation can be used to determine the amino-acid sequence of previously selected peptides, where one takes advantage of the fact that peptides fragment preferentially at their peptide bonds.) The second analyzer then gener-

ates a spectrum of the fragments that pass through the detector. Tandem mass spectroscopy allows dentification of individual compounds such as proteins in complex mixtures.

After this digression into some methodological background, I now return to the first of two methods to analyze large-scale protein expression, two-dimension protein gel electrophoresis (for in-depth reviews, see (6,25)). In this technique, soluble proteins in a cell are isolated and separated electrophoretically in two dimensions (hence the name). In the first dimension, proteins are separated according to their mass. Then the separated proteins are fractionated along a second dimension according to their charge or, more precisely, their isolectric point, in a technique called isolectric focusing. After staining proteins using agents such as silver or Coomassie blue, one is left with a complex pattern of thousands of spots, each representing one (or more) proteins that are expressed in the analyzed cells (see Figure 1 for an example). Individual proteins can then be manually excised from the gel and enzymatically digested into smaller peptides. (The most commonly used approach here is digestion with the enzyme trypsin, which cleaves proteins C-terminally of every arginine or lysine). Cleavage products can then be identified through MALDI-TOF mass spectrometry, either from large protein databases or from the completely sequenced genome, if available.

This approach suffers from limitations that can partially be overcome by automatization. They include the necessity to extract individual protein spots manually for analysis, and the limited reproducibility of the electrophoretic separation, especially across different laboratories. Another limitation, less easily overcome, is the approach's limited dynamic range. It can only detect proteins of moderate to high abundance and resolve "only" on the order of 1,000 proteins per gel. (Compare this with the complexity of the human protein complement of several hundred thousand proteins, if one includes the hundreds of alternatively spliced variants of many of the 30,000 human genes.) A second limitation, shared with many other approaches, is that the technique is not amenable to analyze insoluble membrane proteins, which may account for a large fraction all proteins in a cell.

A second prominent approach to study protein expression on a large scale circumvents two-dimensional gel electrophoresis. It identifies soluble proteins directly from complex protein mixtures extracted from living cells and typically involves three steps (25). The first of them is protein extraction from a cell and chemical or enzymatic digestion of these proteins into peptides. In a second step, the chemical complexity of the resulting mixture is reduced, for example, through separation by capillary electrophoresis. The resulting peptide fractions can then directly enter the third step: the analysis of identified peptides through tandem (MS-MS) mass spectrometry. Such direct mass spectrometry of complex protein mixtures can potentially identify thousands of proteins and, once automated, is a rapid analysis tool. Identification rates of 10^4 peptide sequences per

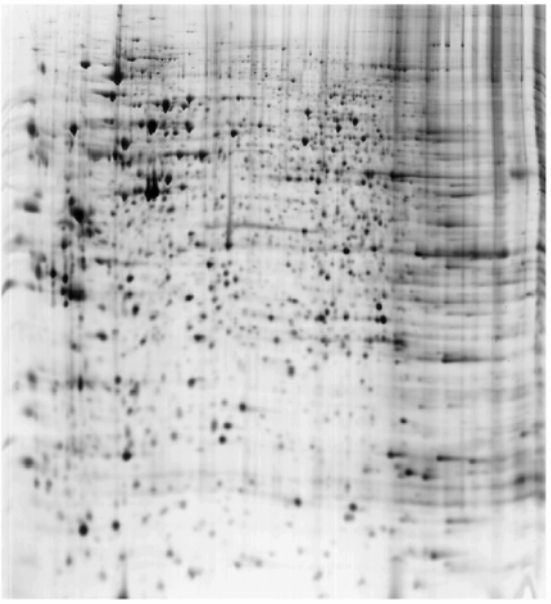


Figure 1. Image of a 2D protein electrophoresis of a whole cell lysate of colorectal cancer cells. Obtained from the SWISS-2DPAGE database (17) at http://us.expasy.org/ch2d/cours2d/gels24.html. Reprinted with permission from the Swiss Institute of Bioinformatics, Geneva, Switzerland.

hour have been achieved. However, as in the case of two-dimensional gel electrophoresis, insoluble proteins and proteins of very low abundance are still a problem. (Unfortunately, many key regulators of cellular processes fall in the latter category.)

In typical applications of either 2D electrophoresis or direct mass-spectrometry, the primary goal is not identification of a large number of proteins expressed in one cell or tissue, but a comparison of protein expression. For example, one may be interested in how particular environmental conditions, such as excessive heat or drought, affect protein expression, in order to identify protein networks that might contribute to heat or drought tolerance. To do so, one needs to compare protein expression between an experimentally manipulated environmental condition and a control (unmanipulated) condition. Another application is the identification of gene functions through the effects that mutations in a gene of interest have on the expression of other genes. Such mutant studies are pursued to obtain information about potential genetic and physical interactions of the gene's products. Comparative analysis of protein expression also has many uses in biomedicine, especially in disease diagnosis. Diseased tissue can

often be distinguished from healthy tissue on the basis of molecular markers, even in the absence of obvious histopathological abnormalities. The most comprehensive of such markers is the entire spectrum of expressed proteins.

In 2D protein electrophoresis, two principal approaches are pursued to compare expression states between different experimental conditions. The first is to carry out two electrophoretic protein separations, one for each of the desired conditions. The resulting spot patterns then need to be matched across gels, a formidable task given the limited reproducibility of individual gels (26). In a second approach, proteins are chemically labeled before separation to distinguish proteins that are expressed under different experimental conditions (27). (The very same approach is also used in direct mass-spectrometric identification of proteins in complex mixtures.) One commonly used labeling technique is that of in vivo labeling, where one population of cells is grown in a standard medium, and the other population in a medium enriched in a stable isotope such as heavy nitrogen (^{15}N). The two cell populations are mixed before protein extraction. Such differential labeling results in slightly different masses of the same gene product in two cell samples, which can be used to quantify to what extent its expression has changed. One of the disadvantages of this technique is that tissue samples, e.g., as obtained through biopsies, cannot be differentially labeled. In situations like these, the technique of post-extraction labeling is useful, as exemplified by isotope-coded affinity tagging (ITAG). Here, proteins extracted from two cell types are chemically modified through the addition of particular chemical moieties such as alkyl groups to specific amino acids such as cysteine. The modifying agent has a different isotope composition for the two samples. After pooling the modified proteins, they can again be analyzed jointly, because their shifted masses distinguish them (1,27).

The data resulting from these approaches is complex and represents relative abundances of the thousands of proteins in a cell. Fortunately, the problem of analyzing data of similar complexity has arisen earlier in large-scale measurements of mRNA gene expression through microarrays. Thus, a plethora of computational tools are available to analyze such data (3,5,8,29).

I will briefly comment on the identification of protein networks through the analysis of large-scale mRNA expression data. Most large-scale analysis of mRNA gene expression has goals similar to that of quantifying protein expression, namely, to provide hints about interacting gene products that form part of a network controlling a biological process. The key difference is that mRNA is only an intermediate in the process of gene expression. From this shortcoming arise a number of disadvantages to reconstructing protein networks from mRNA expression (15). For example, the expression of many proteins is translationally (and not transcriptionally) regulated. For such proteins, mRNA gene expression is a poor indicator of protein abundance. Second, and similarly, differential protein degradation is an important factor in determining protein concentration. It cannot be captured by mRNA gene expression analysis. Finally, the activity of

gene products is often regulated through post-translational modifications such as glycosylation and phosphorylation. mRNA expression, by definition, cannot resolve such differences. In contrast, protein expression studies can identify some of them. For example, different glycosylation states of the same protein manifest themselves in 2D electrophoresis as a closely spaced series of spots. Similarly, protein phosphorylation has specific effects on the mass/charge signature of proteins that can be detected with electrophoretic and mass spectroscopic approaches. However, post-translational modifications do pose enormous challenges to protein expression analysis, especially in light of the fact that hundreds of chemical modifications are possible. In particular, often only a small fraction of the same protein molecules are modified per cell, necessitating enormously sensitive approaches to detect them. In addition, while a single peptide sequence (obtained through tandem mass spectrometry) can be sufficient to identify a protein uniquely, this holds no longer for the identification of post-translational modifications (1).

3. IDENTIFYING PROTEIN INTERACTIONS

Protein networks in the strong sense are identified through approaches that detect interactions of proteins with each other. I will briefly discuss three approaches to detect such protein interactions on a large scale, the yeast two-hybrid assay, a second technique to identify protein complexes, and protein arrays or protein "chips" (reviewed in (23)).

3.1. The Yeast Two-Hybrid Assay

As originally proposed (10), the yeast two-hybrid assay is a technique to identify interactions between two specific proteins A and B (not necessarily from yeast). It takes advantage of the fact that many eukaryotic transcriptional activators consist of two functionally different and separable parts: a DNA binding domain responsible for binding of the transcription factor to DNA, and a transcriptional activation domain responsible for interacting with RNA polymerase. Even when separated (e.g., attached to different polypeptides), these domains can still exert their roles of DNA binding and transcriptional initiation.

To detect interactions between protein A and protein B, the two-hybrid assay first uses recombinant DNA techniques to generate two hybrid proteins. One of them is a hybrid where protein A is fused to the transcriptional activation domain of a yeast transcription factor, such as the gene product of *GAL4*. This hybrid protein is also called the bait protein. The second hybrid protein is a protein where the transcriptional activation domain of the transcription factor is fused to protein B. This hybrid protein is also called the prey (see also Figure 2).

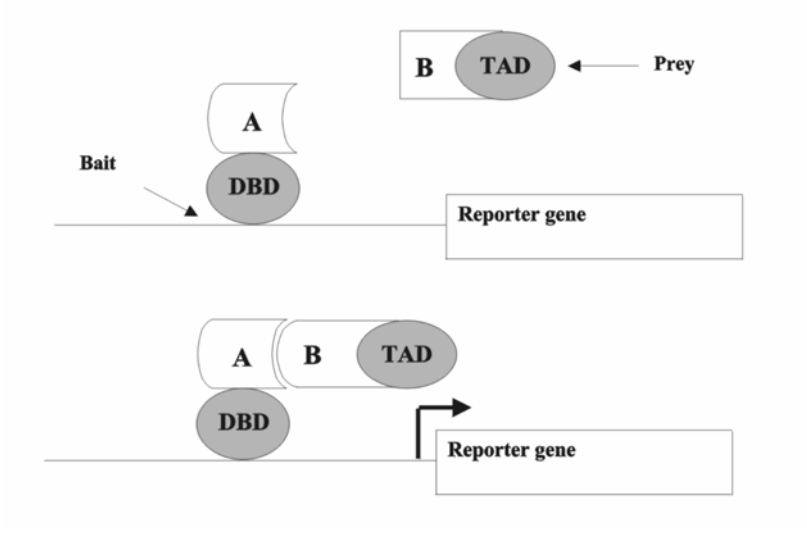


Figure 2. Schematic overview of the yeast two-hybrid assay. See text for details. DBD = DNA binding domain; TAD = transcriptional activation domain.

The assay is carried out in yeast cells that harbor an engineered yeast gene, called a reporter gene, whose upstream region contains a short DNA sequence to which the DNA binding domain of the transcription factor can bind. (Reporter genes are genes whose expression can be easily monitored.) If the two hybrid proteins are expressed together in the same cell, the reporter genes will be expressed if and only if proteins A and B physically interact, and thus build a bridge between the transcriptional activation and the DNA binding domain of the transcription factor (Figure 2).

The two-hybrid approach has been applied to detect not only interactions among two proteins but also on a much larger scale to detect interactions of most proteins in a genome (4,11,12,18,19,28,30). This is a formidable task, considering that even for a small genome like that of the yeast *Saccharomyces cerevisiae*, with some 6,000 genes, 18 million pairwise interactions need to be screened. One approach in carrying out such massive screens takes advantage of the fact that yeast cells reproduce facultatively sexually. They come in two haploid mating types called *a* and α, which can fuse to form a diploid cell.

To screen for all pairwise protein interactions in a genome, one can generate two libraries of molecules, including a bait library containing fusions of all the genes encoding proteins of interest with the coding region for the DNA binding domain of a yeast transcription factor. After yeast cells of mating type *a* are transformed with this library, each transformed cell will express a fusion protein

corresponding to one member of the library. Exactly analogously, a prey library is generated by fusing the genes for the proteins of interest to the coding region for the transcriptional activation domain of the yeast transcription factor. Yeast cells of the opposite mating type α are transformed to give a population of cells, each of which expresses one of the prey fusion proteins.

All yeast cells used in this genome-scale assay bear a special kind of reporter gene that allows them to survive on a selective medium. For example, they may be defective in the gene *HIS3* necessary to synthesize the amino acid histidine, but they contain a reporter gene where the *HIS3* gene is fused to a regulatory DNA region containing the DNA sequence necessary for binding of the transcription factor used in the assay. Yeast cells in which this reporter gene is expressed can grow and divide on a medium lacking histidine.

In the final step of the assay, the previously transformed yeast cells of opposite mating type are allowed to mate and fuse. The resulting diploid cells are exposed to a medium lacking histidine. Only cells where prey and bait protein interact physically will express *HIS3* and will thus survive and form colonies on this medium. The proteins whose interaction allowed them to survive can be identified through DNA sequencing of their encoding genes.

The first genome-wide protein interaction screens to which the two-hybrid assay was applied were carried out in the yeast proteome itself. They yielded maps of protein interactions involving some 1000 proteins (18,30). Variations of the approach have been applied successfully to analyze protein interactions in other microbes, such as the bacterium *Helicobacter pylori* (28), and protein interactions between viral and cellular proteins (4, 11).

The yeast two-hybrid approach has several commonly recognized shortcomings. One of them is the use of fusion proteins, which can lead to bait or prey misfolding. Another problem is that the assay forces coexpression of proteins in the same compartment of a cell or an organism, although the proteins may not co-localize in vivo. These shortcomings lead to potentially high false positive and false negative error rates, i.e., to detection of spurious interactions, and to a failure to detect actual interactions. These error rates may well exceed 50% (7,31).

3.2. Large-Scale Identification of Protein Complexes

Another class of techniques, which I will illustrate with one prototypical example, is designed to identify the proteins that are part of a multiprotein complex (13,16). The departure point of a typical experiment (Figure 3) is some protein A of interest, which is attached to a solid support via a chemical tag that forms part of the protein. A frequently used tag is glutathione S-transferase (GST), an enzyme that binds the tripeptide glutathione with high affinity. To attach this tag to a protein A, a fusion gene containing the coding region of GST

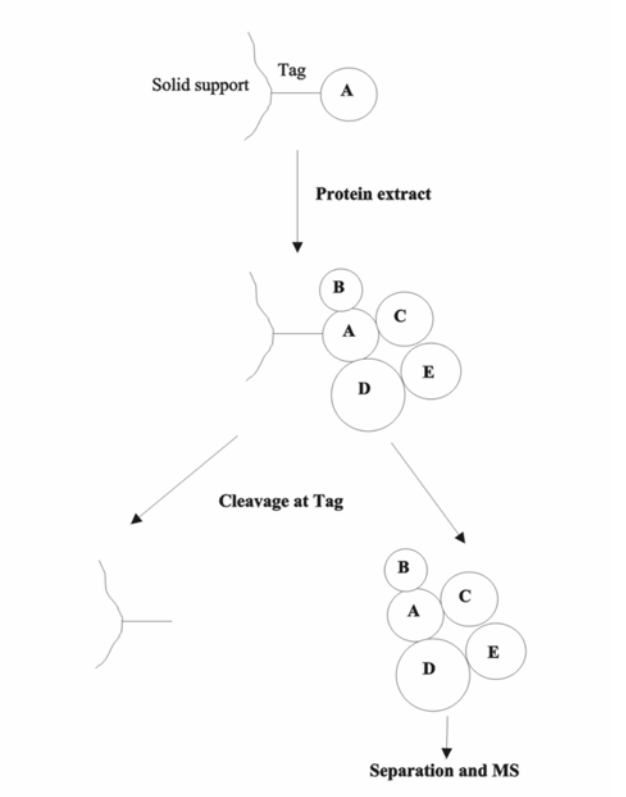


Figure 3. A prototypical strategy to isolate protein complexes. Letters **A–E** indicate proteins. The bait protein A is reversibly bound through a chemical tag to a solid substrate that forms part of an affinity chromatography column. Upon exposure to a protein extract from a cell, proteins that bind to A, either directly or indirectly, are retained on the column. This complex of proteins can be released from the column, and analyzed further. See text for details.

and that of protein A is generated. The fusion protein expressed from this gene can be bound onto a solid support of glutathione-coated agarose beads that form part of an affinity chromatography column. An important feature of this binding is its reversibility, that is, the protein A–GST fusion (and anything else attached to it) can be eluted from the solid support using an excess of glutathione.

When the chromatography column is exposed to a protein extract from a cell, all proteins capable of binding (directly or indirectly) to protein A will be retained on the column. After eluting protein A and its attached proteins, the components of the protein complex are separated through gel electrophoresis. Individual proteins in the complex can then be identified through mass spec-

trometry. Many variations of this approach exist, most notably among them one that uses co-immunoprecipitation to isolate a complex of interest. Here, the tag is a peptide epitope that can be recognized by an antibody, a feature that can be used to render a protein complex insoluble.

To identify numerous protein complexes, it is necessary to repeat an experiment of this kind with many different bait proteins. The largest-scale approaches so far have identified more than 500 protein complexes in the yeast *Saccharomyces cerevisiae* (13,16).

3.3. Protein Arrays (Protein "Chips")

A protein array consists of hundreds or thousands of well-defined locations on a small surface, each of which contains many copies of one protein. Protein arrays are useful not only to study protein–protein interactions, but also a variety of other aspects of protein function. There are again several approaches (reviewed in (35)), of which I will present only one representative example, along with one application (34).

In one approach to creating a protein array, a collection of many different genes that encode proteins of interest is established. Each gene in the collection is fused to the coding region of GST (see above) and to a stretch of DNA encoding many histidine residues (PolyHis). Upon expression of these chimeric genes, the fusion proteins can be attached to a glass slide via their PolyHis tail, each protein in a separate and well-defined spot. (The glass slide constitutes the protein chip.)

Zhu and collaborators (34) used such protein arrays to ask which proteins in the yeast *Saccharomyces cerevisiae* interact with calmodulin. Calmodulin is an important protein in many organisms and mediates the action of calcium ions on cells. It is, for example, involved in the secretion of proteins and in the motility of vesicles inside cells. To identify which yeast proteins interact with calmodulin thus means to identify which proteins contribute to regulating the effects of Ca^{2+} on yeast cells. The employed assay uses biotinylated calmodulin, that is, calmodulin to which biotin has been attached. Biotin is a compound that is very tightly and specifically bound by the protein streptavidin, its sole purpose in this assay. Another important ingredient of the assay is streptavidin that is chemically modified by attaching a green fluorescent dye, Cy3, to it. In the assay, the array is exposed to a solution of biotinylated calmodulin and Cy3-labeled streptavidin. All proteins that interact with calmodulin will have biotinylated calmodulin and Cy3-labeled streptavidin bound to them. The corresponding spots on the array will appear green when exposed to light of the appropriate wavelength.

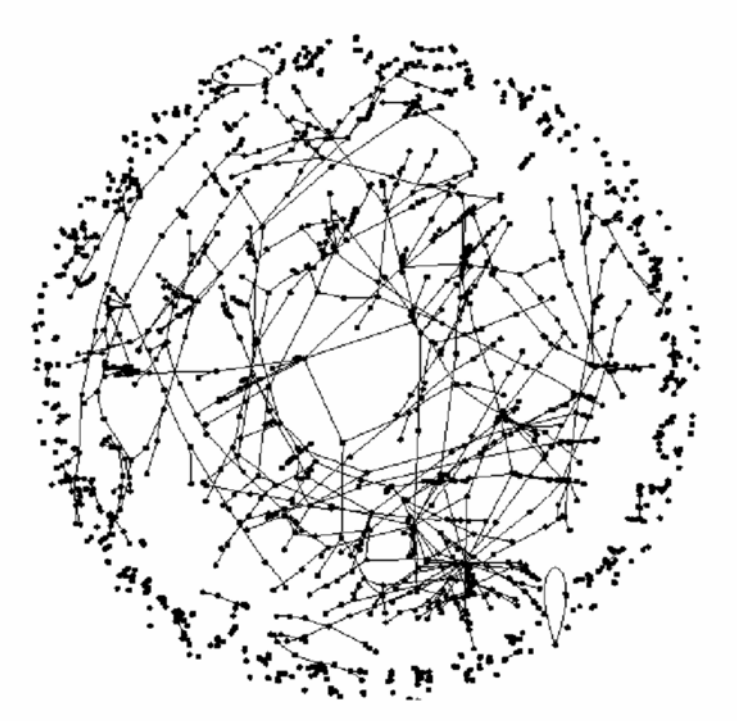


Figure 4. A graph-based representation of the protein interaction network of the yeast *Saccharomyces cerevisiae*. Nodes in the graph correspond to proteins. Black lines connect protein pairs for which at least one of three sources of empirical data (18,22,30) indicates that the proteins physically interact.

3.4. The Complexity of Protein Interaction Networks

It comes as no surprise that the above techniques to identify protein interactions and protein complexes, especially those techniques that can already be applied on a genome-wide scale, reveal protein interaction networks of great complexity. Figure 4 shows an example of such a network in the yeast *Saccharomyces cerevisiae*, as defined by yeast two-hybrid data (18,30) and manually curated protein interaction data (22). This network comprises 1893 proteins and 2364 pairwise interactions. Similar networks in higher organisms are even larger in size. For example, a recently published protein interaction map of the fruit fly *Drosophila melanogaster* involves 7048 proteins and 20,405 interactions (14). (These numbers do not represent accurate estimates of the number of protein interactions in vivo, because of some shortcomings of the technique to identify protein interactions cited above.)

Since the first such protein interaction maps became available, their structural and functional characterization have received great attention. Such networks reveal some simple structural features that they share with other biological networks (20,32). In particular, it has been noted that the number d of interactions per protein often resembles a power law, $P(d) \sim d^{-\gamma}$, where γ is a constant characteristic of the network. In addition, such networks show a high degree of clustering of interactions within small groups of nodes. Such clustering is often measured through a clustering coefficient C (33). To define the clustering coefficient $C(v)$ of a node (protein) v in a graph, consider all k_v nodes adjacent to a node v, and count the number m of edges (protein interactions) that exist among these k_v nodes (not including edges connecting them to v). The maximally possible m is $k_v(k_v-1)/2$, in which case all m nodes are connected to each other. Let $C(v) := m/(k_v(k_v - 1)/2)$. $C(v)$ measures the "cliquishness" of the neighborhood of v, i.e., what fraction of the nodes adjacent to v are also adjacent to each other. In extension, the clustering coefficient C of the graph is defined as the average of $C(v)$ over all v. It can be orders of magnitude larger in biological networks than in random networks of similar size and connectivity distribution. Despite some intriguing propositions (2), the functional and biological significance of such functional features is still unclear. In addition, because of their size and complexity, these networks may harbor biologically important structural features that remain completely unexplored.

4. MEDICAL APPLICATIONS

A hopefully temporary shortcoming of all presently available approaches to identify protein interactions on a genome-wide scale is their high rate of detecting spurious interactions and missing actually occurring interactions (7). Nevertheless, a number of commercial ventures (e.g., the German company Cellzome [www.cellzome.com] or the French company Hybrigenics [www.hybrigenics.com]) are already dedicating themselves to providing information on pairwise protein interactions and protein complexes (protein networks in the strong sense) to the pharmaceutical industry, for the purpose of drug discovery. Such commercial interest is a strong indicator of the promise protein network information holds for the future of drug discovery. However, the vast majority of studies published to date used network information in the weak sense, as defined above, in proof-of-principle biomedical applications. I will now provide a few examples of such applications.

4.1. Mechanisms of Drug Action

Analysis of protein expression can be useful in comparing the effects of newly identified antibiotics to those of existing antibiotics. This serves the pur-

pose of obtaining information about the antibiotic's mechanism of action, potential toxicity, and possible mechanisms of bacterial resistance. An application is described by Evers and Gray (9), who identified the novel antimicrobial compound Ro-64-1874. The effects of this compound on protein expression in the bacterium *Haemophilus influenzae* was compared to the effects of a series of well-characterized other antibiotics that affect cellular processes such as transcription, translation, or tRNA synthesis. It was found that Ro-64-1874 causes a protein expression profile (obtained through 2D electrophoresis) very distinct from most characterized antibiotics but similar to that of the antibiotic trimethoprim, which is an inhibitor of the enzyme dihydrofolate reductase (DHFR). This assay thus suggests that Ro-64-1874 may also inhibit DHFR activity.

4.2. Drug Target Validation

The ideal inhibitory drug is highly specific, i.e., it inhibits a single protein or other gene product so completely as if the encoding gene itself was absent. Marton and collaborators (21) developed a test for drug specificity that is based on this observation. For their proof-of-principle experiments, they chose the drug target calcineurin in the budding yeast *Saccharomyces cerevisiae*. Calcineurin is a protein phosphatase implicated in T-cell activation, apoptosis, and cardiac hypertrophy in higher organisms. In yeast, it is involved in the regulation of the cell cycle and in osmotic homeostasis. Marton and collaborators tested two drugs—FK506 and Cyclosporin A—both immunosupressants that can inhibit calcineurin activity in higher organisms and in yeast. How specific is the action of these drugs? They asked this question by comparing gene expression profiles between yeast mutants from which the calcineurin gene had been deleted, and strains where calcineurin had been pharmacologically inactivated by application of either drug. If the drugs act with high specificity, one would expect that both the genetic and pharmacological inactivation yield very similar effects on gene expression. This is what the authors observed. In contrast, deletions in genes unrelated to the calcineurin pathway yielded effects on gene expression uncorrelated with those of pharmacological inhibition. This approach is also suited to identify the exact nature of a drug's off-target effects, by studying how gene expression changes in pharmacologically versus genetically manipulated organisms (21).

4.3. Toxicology

Cyclosporin A is an immunosuppressant with a number of side effects, most importantly among them nephrotoxicity. What causes this side effect? Evers and

Gray (9) synthesized results from a series of studies that identify a likely cause using a protein expression analysis approach. Two-dimensional protein electrophoresis in rats showed that application of cyclosporin A inhibits the expression of the protein calbindin-D in rat kidneys. This response is associated with increased calcium excretion and calcium accumulation in the kidney. Calbindin 28 has been implicated in calcium transport and calcium buffering, thus making it a prime candidate for mediating this side effect.

4.4. Diagnosis

One among a growing number of proof-of-principle studies for diagnostic applications is that by Gordon and collaborators (cited in (21)), which describes the use of gene expression analysis to distinguish two types of lung tumors: malignant pleural mesothelioma (MPM) and adenocarcinoma (ADCA). Accurate differential diagnosis is critical in case of these two tumors, as treatment strategies differ greatly for them. However, diagnosis is difficult using conventional methods. Using large-scale gene expression analysis, the authors identified eight genes that changed their expression level most drastically in a "training" set of 16 ADCA and 16 MPM tissue samples. Using these biomarker genes as diagnostic indicators, the authors were able to diagnose 99% of presented tumor samples as being either ADCA or MPM (21).

The above examples are hand selected from a growing literature in this area. They illustrate the great potential of protein network-based approaches for medical applications. With continued improvements in experimental techniques to characterize such networks, they will without doubt play an important role in the future of biomedicine.

5. ACKNOWLEDGMENTS

I would like to thank the NIH for its support through grant GM63882-02.

6. REFERENCES

1. Aebersold R, Goodlett DR. 2001. Mass spectrometry in proteomics. *Chem Rev* **101**:269–295.
2. Albert R, Jeong H, Barabási AL. 2000. Error and attack tolerance of complex networks. *Nature* **406**:378–382.
3. Alon U, Barkai N, Notterman DA, Gish K, Ybarra S, Mack D, Levine AJ. 1999. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc Natl Acad Sci USA* **96**:6745–6750.
4. Bartel PL, Roecklein JA, SenGupta D, Fields S. 1996. A protein linkage map of *Escherichia coli* bacteriophage T7. *Nature Genet* **12**:72–77.

5. BenDor A, Shamir R, Yakhini Z. 1999. Clustering gene expression patterns. *J Comput Biol* **6**:281–297.
6. Dunn MJ, Gorg A. 2001. Two-dimensional polyacrylamide gel electrophoresis for proteome analysis. In *Proteomics*, pp. 43–60. Ed SR Pennington, MJ Dunn. Springer, New York.
7. Edwards AM, Kus B, Jansen R, Greenbaum D, Greenblatt J, Gerstein M. 2002. Bridging structural biology and genomics: assessing protein interaction data with known complexes. *Trends Genet* **18**:529–536.
8. Eisen MB, Spellman PT, Brown PO, Botstein D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* **95**:14863–14868.
9. Evers S, Gray CP. 2001. Application of proteome analysis to drug development and toxicology. In *Proteomics*, pp. 225–236. Ed SR Pennington, MJ Dunn. Springer, New York.
10. Fields S, Song OK. 1989. A novel genetic system to detect protein protein interactions. *Nature* **340**:245–246.
11. Flajolet M, Rotondo G, Daviet L, Bergametti F, Inchauspe G, Tiollais P, Transy C, Legrain P. 2000. A genomic approach of the hepatitis C virus generates a protein interaction map. Gene **242**:369–379.
12. Fromont-Racine M, Rain JC, Legrain P. 1997. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nature Genet* **16**:277–282.
13. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Leutwein C, Bouwmeester T, Kuster B, Superti-Furga G. 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. *FASEB J* **16**:A523–A523.
14. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E, Vijayadamodar G, Pochart P, Machineni H, Welsh M, Kong Y, Zerhusen B, Malcolm R, Varrone Z, Collis A, Minto M, Burgess S, McDaniel L, Stimpson E, Spriggs F, Williams J, Neurath K, Ioime N, Agee M, Voss E, Furtak K, Renzulli R, Aanensen N, Carrolla S, Bickelhaupt E, Lazovatsky Y, DaSilva A, Zhong J, Stanyon CA, Finley RL Jr, White KP, Braverman M, Jarvie T, Gold S, Leach M, Knight J, Shimkets RA, McKenna MP, Chant J, Rothberg JM. 2003. A protein interaction map of *Drosophila melanogaster*. *Science* **302**:1727–1736.
15. Gygi SP, Rochon Y, Franza BR, Aebersold R. 1999. Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* **19**:1720–1730.
16. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P, Bennett K, Boutilier K, Yang L, Wolting C, Donaldson I, Schandorff S, Shewnarane J, Vo M, Taggart J, Goudreault M, Muskat B, Alfarano C, Dewar D, Lin Z, Michalickova K, Willems AR, Sassi H, Nielsen PA, Rasmussen KJ, Andersen JR, Johansen LE, Hansen LH, Jespersen H, Podtelejnikov A, Nielsen E, Crawford J, Poulsen V, Sorensen BD, Matthiesen J, Hendrickson RC, Gleeson F, Pawson T, Moran MF, Durocher D, Mann M, Hogue CW, Figeys D, Tyers M. 2002. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* **415**:180–183.
17. Hoogland C, Anchez J-C, Tonella L, Binz P-A, Bairoch A, Hochstrasser DF, Appel RD. 2000. The 1999 SWISS-2DPAGE database update. *Nucleic Acids Res* **28**:286–288.
18. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci USA* **98**:4569–4574.
19. Ito T, Tashiro K, Muta S, Ozawa R, Chiba T, Nishizawa M, Yamamoto K, Kuhara S, Sakaki Y. 2000. Toward a protein–protein interaction map of the budding yeast: a comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proc Natl Acad Sci USA* **97**:1143–1147.
20. Jeong H, Mason SP, Barabási A-L, Oltvai ZN. 2001. Lethality and centrality in protein networks. *Nature* **411**:41–42.
21. Marton MJ, Derisi JL, Bennett HA, Iyer VR, Meyer MR, Roberts CJ, Stoughton R, Burchard J, Slade D, Dai HY, Bassett DE Jr, Hartwell LH, Brown PO, Friend SH. 1998. Drug target vali-

dation and identification of secondary drug target effects using DNA microarrays. *Nature Med* **4**:1293–1301.
22. Mewes HW, Heumann K, Kaps A, Mayer K, Pfeiffer F, Stocker S, Frishman D. 1999. MIPS: a database for genomes and protein sequences. *Nucleic Acids Res* **27**:44–48.
23. Pandey A, Mann MP. 2001. Proteomics to study genes and genomes. *Nature* **405**:837–846.
24. Patterson SD, Aebersold R, Goodlett DR. 2001. Mass spectrometry-based methods for protein identification and phosphorylation site analysis. In *Proteomics*, pp. 87–124. Ed. SR Pennington, MJ Dunn. Springer, New York.
25. Patton WF. 2001. Detecting proteins in polyacrylamide gels and on electroblot membranes. In *Proteomics*, pp. 65–81. Ed. SR Pennington, MJ Dunn. Springer, New York.
26. Pleissner K-P, Oswald H, Wegner S. 2001. Image analysis of two-dimensional gels. In *Proteomics*, pp. 131–146. Ed. SR Pennington, MJ Dunn. Springer, New York.
27. Quadroni M, James P. 2001. Enhancing high-throughput proteome analysis: the impact of stable isotope labeling. In *Proteomics* pp. 151–169. Ed. SR Pennington, MJ Dunn. Springer, New York.
28. Rain JC, Selig L, De Reuse H, Battaglia V, Reverdy C, Simon S, Lenzen G, Petel F, Wojcik J, Schachter V, Chemama Y, Labigne A, Legrain P. 2001. The protein–protein interaction map of *Helicobacter pylori*. *Nature* **409**:211–215.
29. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM. 1999. Systematic determination of genetic network architecture. *Nature Genet* **22**:281–285.
30. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamodar G, Yang M, Johnston M, Fields S, Rothberg JM. 2000. A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**:623–627.
31. von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, Bork P. 2002. Comparative assessment of large-scale data sets of protein–protein interactions. *Nature* **417**:399–403.
32. Wagner A. 2001. The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. *Mol Biol Evol* **18**:1283–1292.
33. Watts DJ, Strogatz SH. 1998. Collective dynamics of small-world networks. *Nature* **393**:440–442.
34. Zhu H, Bilgin M, Bangham R, Hall D, Casamayor A, Bertone P, Lan N, Jansen R, Bidlingmaier S, Houfek T, Mitchell T, Miller P, Dean RA, Gerstein M, Snyder M. 2001. Global analysis of protein activities using proteome chips. *Science* **293**:2101–2105.
35. Zhu H, Snyder M. 2001. Protein arrays and microarrays. *Curr Opin Chem Biol* **5**:40–45.

1.4

ELECTRONIC CELL ENVIRONMENTS: COMBINING GENE, PROTEIN, AND METABOLIC NETWORKS

P. Dhar

Bioinformatics Institute, Singapore

M. Tomita

Institute for Advanced Biosciences, Keio University, Fujisawa, Kanagawa, Japan

Whole-cell modeling is an emerging field of science that takes a systems view of the cell. The grand challenge is to understand, model, and simulate cellular complexity, and by extension, an organism. In this chapter we have tried to outline the raw material for modeling a cell, the advantages and limitations of various modeling strategies, the currently available tools, the biomedical applications, and the prospects for future growth in this field.

1. INTRODUCTION

The classical method of studying biology has been dominated by a ubiquitous reductionist approach, that is, dividing a problem into well-defined components and studying each one in isolation. Though this strategy has largely succeeded in enhancing our knowledge of life, it has also resulted in creation of rules that come with *exceptions*! This is primarily because knowledge emerged

Address correspondence to: Masaru Tomita, Institute for Advanced Biosciences, Keio University, 5322 Endo, Fujisawa, Kanagawa, 252-8520 Japan (mt@sfc.keio.ac.jp).

from investigation of the modular description of components. "Rules" emerged from intra-modular knowledge and "exceptions" from unidentified inter-modular interactions, an unchartered biological territory—until recently. Biology has undergone a remarkable transition: from single investigator-driven to multi-investigator-driven research. Nowadays it is fashionable to talk of a special biological metamorphosis called "Systems Biology." Systems biology is an *emergent* phenomenon that arose from a need to combine biology with mathematics, physics, chemistry, and computer science.

Analysis of the cell in its entirety assumes importance in view of the fuzzy boundaries that exist among pathways. Crosstalk among cellular pathways exists among functionally specialized components, e.g., DNA polymerase, which both acts as a catalyst for synthesizing new strands and communicates with the cell repair machinery. The DNA polymerase "interface" is indicative of the fact that a given cellular component may be connected both upstream and downstream into a mesh of transactions. These transactions bring about dynamic interaction among otherwise "static" genes and proteins. The whole cell network is extensive, demonstrates nonlinearity, and is difficult to describe in terms of concentrations alone. The origin of nonlinearity lies in the feedback loops, rate constants, and the inherent randomness and noise in gene expression, in addition to coupled vector and scalar processes. Considering cellular events as entirely modular processes is therefore an intrinsically error-prone assumption. Systems biology deals with these issues and offers an in-silico view of the genetic and metabolic pathways, thus providing a clear mechanistic basis as well as a practical view of a biological process, especially the emergent phenomena.

Ever since in-silico modeling paved the way for creation of the first hypothetical virtual cell (34), researchers have been aggressively pushing the case for creating more complicated virtual systems. The challenge is not only to simulate a reaction, but to simulate it *accurately* in the presence of diverse physiological conditions and feedback loops. Figure 1 shows a schematic representation of quantitative modeling of cellular processes. In the following sections we review some of the background work, modeling, and simulation tools—especially E-Cell—and the progress we expect to see in this field.

2. BIOMEDICAL BACKGROUND

The idea of systems biology is not new. In 1948 Norbert Wiener took a systems approach in search of general biological laws and posited the principles of cybernetics (38). Though his attempt was new and farsighted, his timing was not right due to the nonavailability of biological data. Throughout the 1960s and 1970s researchers from the fields of mathematics and engineering enthusiastically pursued the idea of transferring knowledge from physics to

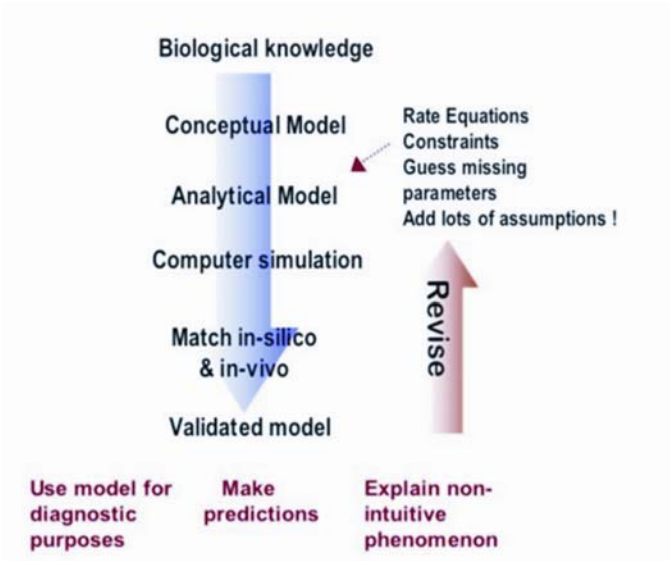


Figure 1. Animating cellular processes *in silico*.

biology (1,2,22), but they realized the limitations of the classical systems approach. The underlying assumption was that physical laws were as applicable to biological systems as to engineering systems. The dynamic and nonlinear nature of living systems was not well understood, and researchers were comfortable in viewing organisms as deterministic systems. Furthermore, the ease with which measurements could be taken from physical systems was in stark contradiction to biological systems, which posed major data-gathering challenges, leading to problems in building precise analytical models. As a result, initial modeling efforts were successful only to the extent of simulating cellular events, not in explaining fundamental principles. Following this, attempts were made to construct toy models of biological systems under an assumption of steady-state conditions. The most prominent of these were biochemical systems theory and metabolic control theory. Another aspect that received more and more attention was the appearance of patterns at different levels of biological complexity (24).

The last two decades have witnessed remarkable advancements in the field of molecular and cellular biology. This has led to a better understanding of biological complexity, from molecules to organisms. The traditional and immensely successful reductionist approach has resulted in the creation of huge databases. However, to understand biology at its default (i.e., systems) level, data needs to be woven into a system that not only portrays the known interactions but also

helps uncover new relationships. Early attempts at building biological models focused primarily on metabolic pathways because of the availability of a rich inventory of qualitative and quantitative data. With the emergence of new information, researchers moved from a data-modeling to a process-modeling approach (36) using kinetic equations and rate laws. Some of the earliest papers in biochemical pathway modeling were published during the 1960s and 1970s (8,14,25–27). It was quite challenging to solve algebraic equations manually, thus restricting the "model bandwidth" to only a few equations. During the 1980s and 1990s, the availability of massive computational power tilted the balance in favor of more powerful modeling strategies (9,10). However, despite the impressive strides in biological computer modeling, a few constraints restricted its full-scale growth. Some of the constraints were (a) a lack of high throughput and high-quality data, (b) understanding of complexity, and (c) difficulty in experimentally validating computer models. In 1997 a major milestone in biological modeling was reached when a virtual *Mycoplasma genitalium* with 127 genes was created (34), thereby signaling the arrival of an era of credible in-silico modeling. Over the last few years, using sound theoretical and experimental data, biochemical and gene expression models have been published, pushing the field of in-silico modeling to new levels. Excellent reviews have appeared on this subject recently (3,12,16,17,23,35,39).

3. MODELING AND SIMULATION

Biochemical systems come with a variety of features: forward reactions, reverse reactions, feedback loops, redundancy, stability, and modularity. The challenge is to quantitatively represent each of these features and integrate them into the model. A number of parallels between biological systems and engineering systems have prompted researchers to adopt reverse-engineering approaches, using well-established concepts from physics. Some of the traits that biological and engineering systems share include: rapid communication and response, accurate error detection and correction, fuzzy control, amplification, adaptation, and robustness. However, biological systems are open (i.e., interact with the environment, thereby providing an unlimited supply of building blocks for nucleic acids and proteins), nonlinear (a well-defined input does not always lead to the predicted output) (see this volume, Part II, chapter 2, by Socolar), and exhibit an emergent property (system behavior cannot be explained by individual components).

3.1. Why Do We Need Modeling?

1. Every aspect of biological phenomena cannot be easily captured by experiments alone. To answer complex "what if" questions, one needs novel tools and strategies to supplement the wet bench approach.

2. Modeling helps identify essential components, i.e., the major players of a given system, and to filter out redundant (i.e., nonessential) elements, which may represent evolutionary "debris."

3. Modeling facilitates rejection of false hypotheses and enables a more precise understanding of the non-intuitive behavior of a system.

4. A good model can accurately predict the future state of the system in the presence/absence of a perturbation.

5. We can easily knock-in or knock-out components from the system and study their upstream/downstream effects. A model is an inexpensive alternative to wet lab tests.

3.2. Limitations of Modeling

1. Modeling may result in duplication of experimental results.

2. The incompleteness of knowledge may result in limited predictive power.

3. Models are often constructed to answer very specific questions, without considering the big picture. Though somewhat unavoidable, the model must grow with time.

4. Even a good model may sometimes yield incorrect predictions. For example, the model of T7 phage hinted at a connection between genomic rearrangement and its growth rate, a prediction that proved incorrect (7).

3.3. Mathematical Basis

Translating biochemistry into mathematics is what eventually drives computer simulation, but this is not always a straightforward process unless the data are clean. Modeling with differential equations enables extrapolation to future states. Generally, analytical models are built with ordinary differential equations and/or stochastic equations, the accuracy of which depends on the assumed parameters, rate laws, and concentrations (see Table 1). In order to find missing parameters, various mathematical approaches are used: for example, simulated annealing, genetic algorithms.

Advantages: Precision, does not permit vague/fuzzy statements, combines explanation with predictions, connects various levels of biological organization

Limitations: nonlinearity (exhibited by biological systems) is very hard to solve analytically.

3.4. Traits of a Good Model

An ideal model is experimentally validated, analyzable, and open for manipulation and optimization. The quality of a model is directly proportional to the quality of the data. Good data are taken under uniform experimental conditions, are based on time-series measurements, and are obtained by standard experimental protocols. The rule of the thumb includes (a) keep it as simple as possible and as complex as necessary, (b) strive for realistic goals, (c) do not

Table 1. Formalisms commonly used in modeling cellular processes

Cell process	Modeling formalism
Gene expression	Boolean, rule-based, stochastic
Signaling pathway	Boolean, rule-based, stochastic
Metabolic pathway	Ordinary differential equations, S-systems
Membrane transport, cytoplasmic streaming	Reaction diffusion, deterministic partial differential equations, spatial stochastic equations

aim at deepest level of understanding at the very beginning, (d) ignore noncontributory factors while studying a given phenomena, and (e) incorporate reliable data and assumptions.

3.5. Validity

No model can completely fulfill all the items on a researchers' wish list. The question of validity is not absolute. Validity depends on the purpose of model analysis. A good data fit alone is not a reliable criterion for model validity. What is important must be differentiated from what can be ignored. The standard process is to check if the simulation results match experimental observations, and reject, modify, or develop new hypotheses. Computer simulations may yield a number of possible answers to a problem, which must be confirmed through active experimental backup. The model validates or rejects only those hypotheses that lie within its scope—depending on whether it is qualitative or quantitative. A validated model is stable under a range of physiological conditions in addition to the defined in-silico conditions. The potential impacting factors that are usually ignored are: the cell cycle phase, circadian rhythm, cell age, temperature, pH, cell morphology, ion concentration gradient, and a nonuniform mix of nutrients, proteins, and ions in the cytoplasm of the cell.

3.6. Problems in Building Good Models

1. Biology is currently undergoing a *phase transition* from low-volume output (single investigator-driven) to high-throughput (multi-investigator-driven) research. Currently, we do not have enough accurate quantitative data to develop reliable theoretical models that could accurately explain and predict system be-

havior. In comparison, models in physics and chemistry are robust and reliable because the physical laws are very well known and hold under a wide range of conditions. By its nature, mathematical modeling requires prior knowledge of all the factors and conditions that affect a reaction or, by extension, a system. For small, well-isolated systems some/most of the information can be obtained in vitro, but it is extremely hard to imagine all possible factors that impact a given phenomenon. Added to this is the computational constraint, especially if the model is based on stochastic methods. Due to a lack of adequate data, there will always be a problem of unknown components contributing to non-intuitive results.

2. The cell is more like a gel than a soupy bag of enzymes and substrates. Thus the in-vitro rate constants obtained from aqueous solutions do not perfectly match the description of real intracellular environments. Furthermore, gradients, compartments, and inhomogeneous mixtures of substances arise in the cytoplasm. Models commonly assume well-mixed environments to avoid mathematical and computational complexities.

3. The rate constants are mostly assumed and/or extrapolated from experimental results. Even when data are available, there is always an inherent variation regarding the organism and the experimental protocols.

4. A mechanistic model is purely data driven and incorporates a large number of parameters, the values of which are not always possible to obtain with accuracy. To overcome this limitation, parameter-finding algorithms are used that try to find a parameter value closest to its in-vivo counterpart. However, if the search space is large the accuracy of such predictions can be limited.

5. In many cases the reaction kinetics is completely unknown.

6. Temporal inactivation or degradation of enzymes is not generally considered in kinetic models.

7. Metabolic channeling is the movement of substrates between several active sites in a multi-enzyme complex (37) within a co-localized environment. Since it is a special case of metabolism where anatomical separation of a pathway is important, the global/local impact of such molecular crowding is presently unclear.

8. The time scales of intracellular transactions vary from 10^{-6} seconds (diffusion) to many hours (gene regulation). Thus, it is difficult to choose an ideal time step for concurrent simulation of gene expression, diffusion, metabolism, and signal transduction.

9. Complex systems show a collective behavior of individual components, more commonly known as emergent phenomenon, i.e., the whole is greater than the sum of its parts. The real behavior of a cell cannot be guessed merely by looking at individual components. At the moment it is unclear at what step we should stop adding complexity to the system beyond which it becomes redundant.

3.7. Modeling Cellular Pathways

3.7.1. Gene Regulatory Network

Gene regulation is the central theme of the dynamic genome. The cell regulatory system not only determines the Boolean (on/off) state of a gene but also how much and how long it expresses. This is achieved by means of (a) anatomical placement of the components reflecting a typical topology of the gene network, and (b) an active interaction among DNA, RNA, protein, and ions. Thus, the first task in building a gene network model is to focus on its qualitative topological features before moving to quantitative descriptions. In view of the large amounts of coexpression data gathered from microarray experiments, it now seems feasible to "reverse engineer" genetic networks. Based on high-throughput data, the relationships among interacting elements can be traced, leading to construction of a virtual regulatory gene network. The accuracy of this network depends on the accuracy of the data in terms of concentrations, rate constants, expression levels, and copy numbers. The development of techniques like cDNA microarray and oligo-based chips permits study of spatiotemporal gene expression in a cell (4,19). Of the various methods of modeling gene regulation, the stochastic approach proposed by McAdams and Arkin (20) seems to be most accurate, as recently observed from experimental data (6), since it captures noise in gene expression (see this volume, Part III, chapter 1.1, by Pedraza and Oudenaarden). The biological basis of gene expression modeling (12,20,21), its mathematical basis (5), and the grand challenges awaiting the modeling community (32) have been excellently reviewed in the past.

3.7.2. Metabolic Pathways

The fundamental feature of metabolic pathways is the presence of an intermediate between two enzyme-catalyzed reactions. Intermediates coexist as linear chains (one incoming reaction and one outgoing reaction), branched chains (one incoming reaction, >1 outgoing reactions), loops (metabolites common to one incoming and at least two outgoing reactions), or cycles (a series of linear and branched reactions that roll over to the starting point). A step is the fundamental unit of a pathway and describes a single reaction event. A sequentially joined array of steps with one entry and many exit points forms a pathway. A collection of pathways having many entry and exit points forms a network. What follows are a few prominent concepts in biochemical pathway analysis.

1. A **stoichiometric network** describes a series of spatially homogenous transformations governed by standard laws to remove or produce substances in fixed proportion. A stoichiometric matrix indicates the number of molecules consumed or produced in each reaction. The stoichiometric network has a well-

defined topology and has strong mathematical foundations in graph theory and convex set theory. In stoichiometric network analysis (SNA), the enzyme kinetics, rate constants, activators, or inhibitors are not important. The only input in the matrix is the number of substances participating in a given reaction and the number of reactions in the system. **Advantages**: SNA does not depend on the nature of the reaction, that is, can be deterministic or stochastic, discrete or continuous, has direct applications in metabolic flux analysis, allows ease in generation of a matrix, and the balance equations for substance concentrations can be written precisely. **Limitations**: cannot represent the evolutionary growth of a network or its regulatory properties, leading to limited predictive power. Energy metabolites must be balanced and bidirectional steps are sometimes hard to resolve. Furthermore, distinguishing parallel pathways from the stoichiometric matrix is very challenging.

2. **Metabolic flux analysis** (MFA) is one of the most important tools in metabolic engineering. MFA makes use of experimental data to study flux distribution in a system. A complete dynamic description of metabolic pathways requires detailed kinetic and regulatory information, which is rarely available from experimental data. An alternative approach is to abstract the dynamic nature of the pathways as a (more easily manageable) steady-state system, i.e., instead of guessing kinetic parameters. Systemic properties are determined mathematically under a quasi-steady state assumption. A very useful concept in MFA is the presence of elementary modes. Elementary modes are the simplest flux distribution reaction steps that cannot be further decomposed (28,29). They represent idealized situations and do not incorporate regulatory feedback loops. The number of elementary modes may indicate the number of alternative routes available to accomplish a certain biochemical task. Whether some or all are actually used by the cells can only be determined under stringent experimental conditions. In addition, with an increase in the number of reaction steps, the number of elementary modes also increases, leading to a combinatorial explosion. This problem can be partly overcome by identifying and grouping isoenzymes. Despite this limitation, elementary flux analysis provides a strong mathematical tool to satisfy both stoichiometry and thermodynamic requirements. MFA concepts were recently used to predict gene knockout phenotypes in *E. coli* (40). MPA (metabolic pathway analysis), an offshoot of MFA, investigates entire flux distributions, whereas MFA focuses on unitary flux distributions.

3. **Metabolic control analysis** (MCA) is a theoretical approach used to analyze the relative control of fluxes and intermediate concentrations of metabolites. In contrast to the traditional single rate-limiting step concept, the current belief centers around distributed control of enzymes driving the reactions. MCA has a strong theoretical basis and is easy to understand. However, it is quite challenging to implement and test MCA concepts experimentally. The possibility of modifying flux using MCA principles through specific biochemical path-

ways has immense pharmaceutical applications. For example, in a human pathogenic condition caused by a special metabolic pathway of a pathogenic organism, the enzyme with the highest control coefficient would be a natural choice for inhibition.

4. **Some useful definitions**: *Control coefficient* describes the dependency of flux or metabolite concentration on the property of the enzyme that mediates the reaction. *Elasticity coefficient* or *elasticity* is a measure of the catalytic property of an enzyme (in terms of reaction rate) in response to varying concentrations of the metabolites participating or influencing the reaction. *Pathway flux* is the rate at which input substances are processed into their subsequent forms. *Metabolomics* is a quantitative estimation of all known and unknown metabolites in a system under specific physiological conditions. *Metabolite target analysis* describes the correlation between a genetic mutation and its corresponding enzyme. *Metabolic profiling* accounts for the role of known metabolites in selected biochemical pathways. *Metabolic fingerprinting* is the classification of metabolites based on their origin and biological weight.

To build a metabolic model, the concentration of substances, the rate constant and reaction mechanism, as well as the strength of such impacting factors as pH, temperature, co-factors, and ions are considered. The steady overall state in the model can be considered as a composite of smaller steady-state conditions (quasi-steady state) embedded in a larger dynamic system that is open and well connected to the environment. Application of the steady-state principle helps uncover unstable states that result in oscillations. For maintenance of the steady state, the product inhibition loop may be employed in a wide variety of physiological conditions.

Biochemical reactions can be described in two ways, depending on the concentration of participating substances. At lower concentrations they are represented as stochastic events that occur when particles collide at certain probabilities. A heuristic solution to randomness is provided by mathematically sound approaches that trace the evolution of an event over a period of time. At higher concentrations biochemical reactions are represented as deterministic events best described by ordinary differential equations. If the spatial element needs to be used during modeling, partial differential equations are employed.

3.8. Software Tools

A mathematically accurate and computationally robust tool is required to translate life into numbers. Systems Biology Markup Language is a modeling and simulation language driven by a group of researchers from Caltech (http://www.sbml.org). The aim is to enable a diverse set of tools to talk to each other by building cross-platform compatible models. Some of the tools include:

1. **A-Cell** (http://www.fujixerox.co.jp/crc/cng/A-Cell/) is a tool for modeling biochemical reactions that provides backward compatibility with previously constructed models.

2. **BioUML** (http://www.biouml.org) is a Java framework for systems biology ranging from databases with experimental data, tools for formalized description of biological systems structure, and functioning, as well as tools for visualization and simulation.

3. **CellDesigner** (http://www.systems-biology.org) is a biochemical network modeling tool with a graphical user interface. It is designed to be SBW (Systems Biology Workbench) compliant and to support the SBML (Systems Biology Markup Language) format.

4. **Cellware** (http://www.cellware.org) is the first grid-enabled quantitative tool for modeling and simulating cellular processes. It is capable of multi-algorithmic simulation from a few reactions to the whole-cell level.

5. **Dbsolve** (http://websites.ntl.com/~igor.goryanin/) is an integrated development environment based on ordinary differential equations. It is especially useful for calculation of steady-state, fitting, and optimization procedures.

6. The **Discrete Metabolic Simulation System** (DMSS) (http://www.bio.cam.ac.uk/~mw263/ftp/doc/ISMB99.ps) models reactions based on competing metabolite concentrations or metabolite affinities to enzymes, including metabolite and enzyme concentrations.

7. **Gepasi** (http://www.gepasi.org/) is also based on ordinary differential equations and simulates the steady-state and time-course behavior of reactions based on stoichiometry and reaction kinetics values.

8. **Jarnac** (http://www.cds.caltech.edu/~hsauro/Jarnac.htm) is a cell modeling language employed to describe metabolic, signal transduction, and gene networks. Jarnac interacts with the user through the Jdesigner frontend.

9. **Netbuilder** (http://strc.herts.ac.uk/bio/maria/NetBuilder/) is an interactive graphical tool for representing and simulating genetic regulatory networks in multicellular organisms, using electronic engineering principles for pathway layout.

10. **SigPath** (http://icb.med.cornell.edu/crt/SigPath/index.xml) is an information system designed to support quantitative studies on the signaling pathways and networks of the cell. It helps manage information on protein–protein interactions, protein–small molecule interactions, elementary chemical reactions, and enzymatic reactions.

11. **StochSim** (http://info.anat.cam.ac.uk/groups/comp-cell/StochSim.html) is a program that provides a stochastic simulator in which individual biomolecules or molecular complexes are represented as individual software objects.

12. **Trelis** (http://sourceforge.net/projects/trelis) is a graphical Monte Carlo simulation tool for modeling the time evolution of chemical reaction systems

involving small numbers of reactant molecules such as occur in subcellular biological processes like genetic regulatory networks

13. **Virtual Cell** (http://www.nrcam.uchc.edu/) integrates biochemical and electrophysiological data with microscopic images and has a web-based interface.

14. **VLX Biological Modeler** (http://teranode.com/products/vlxbiological.php) allows scientists to create and annotate visual models of complex biological systems, perform simulations for exploring and predicting dynamic system behaviors, analyze experimental data with models for hypothesis testing, and effectively communicate their findings.

15. **E Cell** (http://www.e-cell.org) is a modeling and simulation environment from Keio University, Japan. The primary objective of the E-Cell group is to create a dynamic quantitative model of metabolic and gene regulatory networks in lower and higher model organisms. However, the model building itself involves a fair number of assumptions and rate constants. For example, to model a random bi–bi reaction may incorporate 10 parameters for a single reaction, without taking into consideration activation/inhibition by other molecules. Its basic features are listed below.

E Cell (short for Electronic Cell) is a generic object-oriented environment for modeling and simulating molecular processes of user-definable models, equipped with graphical interfaces that allow observation and interaction. The E Cell approach enables modeling of gene expression, signaling, and metabolism. The first publicly released version of the E Cell simulation system was introduced in 2001 under open source. E Cell is currently in its third version. Versions 1.0 and 3.0 are Linux-based, while version 2 is Windows-based (31).

E Cell 3.0 has been developed with an aim to providing a generic tool that runs in a high-performance software environment. It has a geometry information interface and a capability to integrate different simulation algorithms, including approaches based on a variable-process model, differential equations, diffusion reactions, and particle dynamics. One of the main features of E Cell is its ability to integrate subsystems at different time scales. The modeling architecture is tailored to individual modeling needs and allows users to add components to the system. In addition, it is an integrated modeling environment and can simulate both deterministic and stochastic models. Version 3 has been customized for biologists with little or no programming knowledge.

The first application of E Cell was the creation of a virtual cell with 127 genes in silico (34). The *Mycoplasma* genomic raw material was used to construct a virtual cell with "minimum cellular metabolism." This hypothetical cell consumed glucose from the culture medium, generated ATP, and exported lactate. Transcription and translation steps were modeled to biosynthesize proteins within the cell. The cell also consumed glycerol and fatty acid, producing phosphatidyl glycerol for membrane structure. The cell was "self-supporting," but incapable of proliferation. *Mycoplasma genitalium* was chosen for this virtual

experiment because of its small genome size (580 kb). The information on the kinetic properties of genes and proteins was mostly obtained from the KEGG and BioCyc (previously called EcoCyc) databases.

4. FUTURE WORK AND ITS RELEVANCE TO BIOMEDICINE

In-silico biological modeling has finally arrived and is here to stay. At one end of the spectrum, scientists add components and pathways to the system to enhance its properties, while at the other end cell parts and pathways are removed to rid the system of undesirable components. An optimum way to test the result of such combinations is to perform computer simulations. The computer simulations are reasonably accurate, low-cost, fast, and scalable. Of all the modeling approaches, analytical modeling is the most powerful approach, as it makes it possible to understand the exact regulatory topology of a biochemical pathway. The best estimate about kinetic parameters can be made either directly using standard experimental protocols or by inference from the literature (41).

Living systems represent a continuous and nonlinear interaction of substances that are not only large numerically but divergent in variety. Due to the inherently dynamic nature of a biological system, the traditional modular approach does not hold good in all situations. What is needed is a systems-based approach that not only models the current state accurately but also predicts all possible future states in the presence of varying environmental conditions and perturbations. The systems approach offers a possibility of addressing such questions as follows: Given a certain gene mutation, what would be its downstream impact on the immediate and/or related regulatory and metabolic pathways? How many parameters (and in what combinations) can be tweaked to produce a continuously dividing malignant cell? Given a good quantitative model of a parasitic metabolic pathway, which are the most important and highly connected nodes that can be perturbed to produce large-scale effects? Assuming that a drug binds specific proteins or genes, how many pathways will be affected, and in what sequence, both upstream and downstream? Thus, it is even possible to determine the "virtual side effects" of a drug by conducting such computer-based experiments. However, to make such virtual experiments more accurate, noise-free high-throughput data coupled with a reliable in-vivo validation system is required. The future will see the significant growth of integrative models that not only consider different cellular processes in parallel (metabolism, gene expression, and signal transduction) but also combine diverse modeling strategies (deterministic and stochastic) (30,33).

One of the problems with porting in-vitro data to computer models is the gross dissimilarity between in-vitro and in-vivo systems. While the former are buffer based, the latter represent gel-based environments. Thus, we need better assay systems that provide conditions similar to in-vivo situations. In the future,

simulations must also allow transitions between various granularities of models. The modeling environment should allow scaling up of a rough coarse-grained model (based on qualitative data) to a fine-grained model (based on quantitative data) without having to discard the previous one.

One would also need better data-mining tools and techniques and smarter algorithms to find a proper genomic syntax that can be fed into the model with a fair degree of accuracy. As a result of intensive research in metabolomics, it is hoped that metabolite analysis will provide a clue to novel gene functions. We still do not completely understand how cells maintain robustness and stability in environments fluctuating in terms of ion concentrations, nutrients, pH, and temperature. Modelers now assume an ideal situation that does not consider all these issues, but as more knowledge accrues models will need to be further constrained.

An important question that merits answer is: Do networks exhibit symbiosis and epistasis? If yes, what are the features that promote such crosstalk? Is this interaction physiology- or environment-driven? What is the role of redundancy in the evolution of networks? Given that stochasticity in gene expression is determined by extrinsic and intrinsic factors, how does noise evolve over a period of time? Does noise have any role in pushing gene expression toward more heuristic solutions? Often we curve-fit the data without considering mechanistic models that might provide real control parameters for the system. But to reach that state we need a thoroughly validated model that has failed many times over. In the future we will see more and more forward-looking modeling approaches, i.e., fitting the biological system to the model, as opposed to reverse engineering approaches (fitting the model to the biological system). Other areas that merit attention are: development of a modeling markup language (15); using a common theoretical framework for representing biological knowledge; obtaining validated and time-series high-throughput data; and developing tools capable of integrating large and complex networks. Despite all this, we still do not know if mathematics is the right tool for representing biological systems? If not, what is the best way to model the dynamic cell? Is there a "law of biological complexity" that has roots in physics or chemistry?

5. REFERENCES

1. Ashby WR. 1957. *An introduction to cybernetics*. Chapman and Hall, London.
2. Bertalanffy L. 1973. *General systems theory*. Penguin, Harmondsworth.
3. Bower JM, Bolouri H. 2001. *Computational modeling of genetic and biochemical networks*. MIT Press, Cambridge.
4. Brown PA, Botstein D, 1999. Exploring the new world of the genome with DNA microarrays. *Nature Genet* **21**(suppl.):33–37.
5. De Jong H. 2002. Modeling and simulation of genetic regulatory systems: a literature review. *J Comp Biol* **9**(1):67–103.

6. Elowitz MB, Levine AJ, Siggia ED, Swain PS. 2002. Stochastic gene expression in a single cell. *Science* **297**:1183–1186.
7. Endy D, Yu L, Yin J, Molineaux IJ. 2000. Computation, prediction, and experimental tests of fitness for bacteriophage T7 mutants with permuted genomes. *Proc Natl Acad Sci USA* **97**:5375–5380.
8. Garfinkel D. 1968. The role of computer simulation in biochemistry. *Comp Biomed Res* **2**(1):31.
9. Garfinkel D. 1980. Computer modeling complex biological systems and their simplifications. *Am J Phys* **239**(1):R1.
10. Garfinkel D. 1985. Computer-based modeling of biological systems which are inherently complex: problems, strategies and methods. *Biomed Biochim Acta* **44**(6):823.
11. Gatenby RA, Maini PK. 2003. Cancer summed up. *Nature* **421**:321.
12. Hasty J, McMillen D, Collins JJ. 2002. Engineered gene circuits. *Nature* 420:224–230.
13. Hasty J, McMillen D, Isaacs F, Collins JJ. 2001. Computational studies of gene regulatory networks: *in numero* molecular biology. *Nature Rev Genet* **2**:268–279.
14. Heinrich R, Rapoport SM. 1977. Metabolic regulation and mathematical models. *Prog Biophys Mol Biol* **32**:1.
15. Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, Cuellar AA, Dronov S, Gilles ED, Ginkel M, Igor V, Goryanin I, Hedley WJ, Hodgman TC, Hofmeyr J-H, Hunter PJ, Juty NS, Kasberger JL, Kremling A, Kummer U, Le Novere N, Loew LM, Lucio D, Mendes P, Mjolsness ED, Nakayama Y, Nelson MR, Nielsen PF, Sakurada T, Schaff JC, Shapiro BE, Shimizu TS, Spence HD, Stelling J, Takahashi K, Tomita M, Wagner J, Wang J. 2003. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19**(4):524–531.
16. Ideker T, Galitski T, Hood L. 2001. A new approach to decoding life: systems biology. *Annu Rev Genomics Hum Genetics* **2**:343–372.
17. Kitano H. 2002. Computational systems biology. *Nature* **420**:206–210.
18. Kitano H. 2002. Systems biology: a brief overview. *Science* **295**:1662–1664.
19. Lockhart DJ, Winzeler EA. 2000. Genomics, gene expression and DNA arrays. *Nature* **405**:827–836.
20. McAdams HM, Arkin A. 1998. Simulation of prokaryotic genetic circuits. *Annu Rev Biophys Biomol Struct* **27**:199–224.
21. McAdams HM, Arkin A. 1999. It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet* **15**(2):65–69.
22. Mesarovic MD. 1968. Systems theory and biology: view of a theoretician. In *Systems theory in biology*, pp. 59–87. Ed. MD Mesarovic. Springer, New York.
23. Rao CV, Wolf DM, Arkin AP. 2002. Control, exploitation and tolerance of intracellular noise. *Nature* **420**:231–237.
24. Rosen R. 1978. *Fundamentals of measurement and representation of natural systems*. North-Holland, New York.
25. Savageau MA. 1969a. Biochemical system analysis. I. Some mathematical properties of the rate law for the component enzymatic reactions. *J Theor Biol* **25**:365.
26. Savageau MA. 1969b. Biochemical system analysis. II. The steady-state solutions for an *n*-pool system using a power law approximation. *J Theor Biol* **25**:370.
27. Savageau MA. 1976. *Biochemical systems analysis*. A study of function and design in molecular biology. Addison-Wesley, Reading, MA.
28. Schuster S, Dandekar T, Fell DA. 1999. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol* **18**:326–332.

29. Schuster S, Fell D, Dandekar T. 2000. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnol* **18**:326–332.
30. Takahashi K, Yugi K, Hashimoto K, Yamada Y, Pickett C, Tomita M. 2002. Computational challenges in cell simulation: a software engineering approach. *IEEE Intelligent Syst* **17**(5):64–71.
31. Takahashi K, Ishikawa N, Sadamoto Y, Sasamoto H, Ohta S, Shiozawa A, Miyoshi M, Naito Y, Nakayama Y, Tomita M. In press. E-Cell 2: multi-platform E-Cell simulation system. *Bioinformatics*.
32. Tomita M. 2001. Whole cell simulation: a grand challenge of the 21st century. *Trends Biotechnol* **19**:205–210.
33. Tomita M. 2002. Towards computer aided design (CAD) of useful microorganisms. *Bioinformatics* **17**:1091–1092.
34. Tomita M, Hashimoto K, Takahashi K, Shimizu T, Matsuzaki Y, Miyoshi F, Saito K, Tanida S, Yugi K, Venter JC, Hutchinson CA. 1999. E-Cell: software environment for whole cell simulation. *Bioinformatics* **15**(1):72–84.
35. Voit EO. 2000. *Computational analysis of biochemical systems: a practical guide for biochemists and molecular biologists*. Cambridge University, Cambridge.
36. Voit EO. 2002. Models-of-data and models-of-processes in the post-genomic era. *Math Biosci* **180**:263–274.
37. Welch GR, Easterby JS. 1994. Metabolic channeling versus free diffusion: transition-time analysis. *Trends Biochem Sci* **19**:193–197.
38. Wiener N. 1948 *Cybernetics: control and communication in animal and the machine*. Wiley & Sons, New York.
39. Wolkenhauer O. 2001. Systems biology: the reincarnation of systems theory applied in biology? *Briefings in Bioinformatics* **2**(3):258–270.
40. Edwards JS, Palsson BØ. 2000. The *Escherichia coli* MG1655 in silico; metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci USA* **97**:5528–5533.
41. Selkov E, Basmanova S, Gaasterland T, Goryanin I, Gretchkin Y, Maltsev N, Nenashev V, Overbeek R, Panyushkina E, Pronevitch L, Selkov Jr E, Yunus I. 1996. The metabolic pathway collection from EMP: the enzymes and metabolic pathways database. *Nucleic Acids Res* **24**(1):26–28.

2.1

TENSEGRITY, DYNAMIC NETWORKS, AND COMPLEX SYSTEMS BIOLOGY: EMERGENCE IN STRUCTURAL AND INFORMATION NETWORKS WITHIN LIVING CELLS

Sui Huang, Cornel Sultan, and Donald E. Ingber

Vascular Biology Program, Departments of Surgery and Pathology, Children's Hospital and Harvard Medical School, Boston, Massachusetts

The genomic revolution has led to the systematic characterization of all the genes of the genome and the proteins they encode. But we still do not fully understand how many cell behaviors are controlled, because many important biological properties of cells emerge at the *whole-system* level from the collective action of thousands of molecular components, which is orchestrated through specific regulatory interactions. In this chapter we present two distinct approaches based on the concept of molecular networks to understand two fundamental *system properties* of living cells: their ability to maintain their shape and mechanical stability, and their ability to express stable, discrete cell phenotypes and switch between them. We first describe how structural networks built using the principles of tensegrity architecture and computational models that incorporate these features can predict many of the complex mechanical behaviors that are exhibited by living mammalian cells. We then discuss how genome-wide biochemical signaling networks produce "attractor" states that may represent the stable cell phenotypes, such as growth, differentiation, and apoptosis, and which explain how cells can make discrete cell fate decisions in the presence of multiple conflicting signals. These network-based concepts help to bridge the apparent gap between emergent system features characteristic of living cells and the underlying molecular processes.

Address correspondence to: Donald E. Ingber, Department of Pathology, Children's Hospital, Karp Family Research Laboratories, Room 11.127, 300 Longwood Avenue, Boston, MA 02115 (donald.ingber@childrens.harvard.edu).

1. INTRODUCTION: MOLECULAR BIOLOGY AND COMPLEX SYSTEM SCIENCES

A major goal of the study of complex systems is to formally describe and understand how a large number of different *parts* interact and "self-organize" into a *whole system* that exhibits properties that cannot be understood by studying the components in isolation (2). This goal transcends various levels of description and is of particular importance in biomedical research because higher organisms extend over multiple levels of organization. They are hierarchical systems that integrate their smallest constituent parts—individual molecules including DNA, proteins, and lipids—across multiple levels of organization, to organelles, cells, tissues, organs, and the organism (Figure 1). Thus, in order to understand the whole living system at the "macro" level in terms of molecular parts of the "micro" level, it is necessary to traverse multiple levels of description and size scales through many iterations of integration.

The advent of recombinant DNA technology, almost four decades ago, and the concomitant progress in protein biochemistry have led to great advances in our understanding of the lowest level of organization, the genes and molecular parts that comprise living systems. Analysis of how these components interact has led to elucidation of the fundamental principles of living cells, such as the genetic code, the transcription of genes into mRNA, and the translation of mRNA into proteins. Since then the unfathomable complexity of other molecular processes of living systems, such as the cell's growth cycle and regulation of its behavior by external signals, has attracted most attention in biology. Most molecular biologists now almost entirely focus their efforts on the identification of new genes and proteins and the characterization of their role in these processes.

However, it has recently become clear that, rather than studying individual proteins and pathways separately, an integrative approach is necessary. This is reflected in the burgeoning area of "Systems Biology," which seeks not only to systematically characterize and categorize all the molecular parts of living organisms using massively parallel analytic techniques, but also to understand the functional interactions between the molecules using computational approaches. However, despite these efforts, most researchers use the new high-throughput technologies of genomics simply to accelerate, and expand to the genome-scale, the discovery of new molecular pathways. In contrast, the importance of vertical integration across different levels of organization is still largely neglected.

Moreover, in molecular as well as systems biology, it is still often assumed that the ability to describe mechanistic details through experimentation or the use of mathematical models is equivalent to "understanding" the behavior of a complex system. For instance, a cell is typically thought to enter cell division because a growth factor binds to a cell surface receptor, activates a biochemical cascade (e.g., ras–raf–MEK-Erk), and triggers expression of the protein

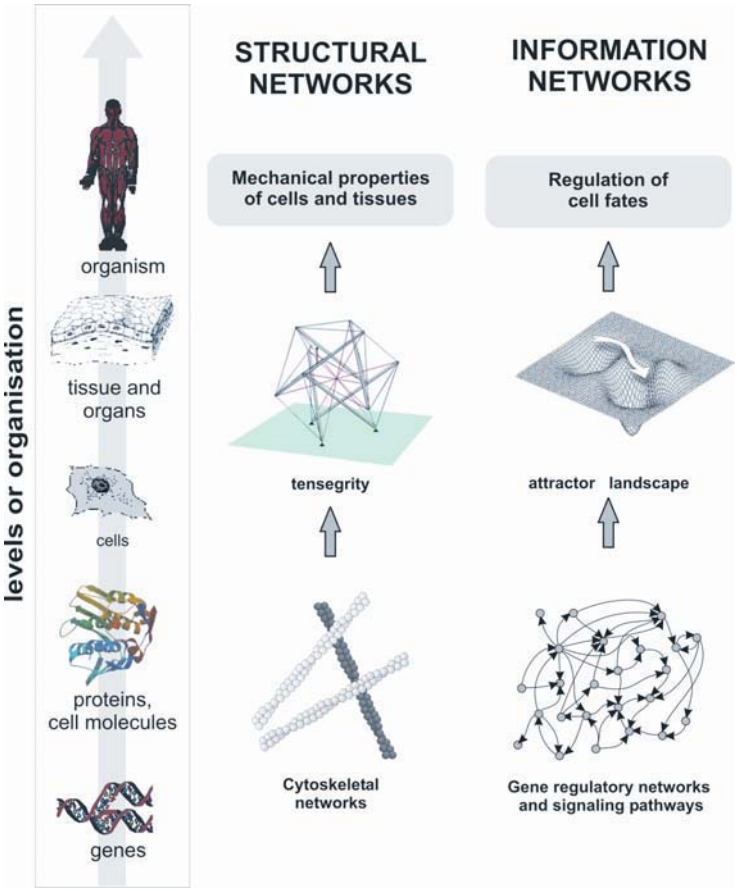


Figure 1. **Hierarchical levels of organization in complex living organisms**. At each level of the hierarchy, new entities and level-specific rules that govern their behavior emerge from the interactions of the entities of the lower level. Thus, there is no "scale-invariance" as in fractals. Here we focus on the "emergence" at the cellular and tissue levels of the characteristic mechanical properties (left) and features of cell fate behavior (right).

cyclin D; this then would lead to the phosphorylation and inactivation of the cell-cycle inhibitor protein Rb, which, in turn, would result in the induction of proteins involved in DNA replication (56). Although we are used to present this type of Rube Goldberg machine-like mechanism of a molecular process in a living cell as if it were explanatory, the reality is that such mechanistic representations are essentially descriptive. All we do is describe a chain of events at a lower (molecular) level than the one used to make our initial observation (cell or

tissue level). Even formal and quantitative approaches like system dynamics-based modeling of particular molecular pathways are descriptive because they use an established set of rules (e.g., formal reaction kinetics and diffusion) to describe and predict in detail the time evolution of a *particular instance* of a class of system whose generic behavior is known given the set of equations and conditions (see this volume, Part II, chapter 2, by Socolar). While useful to predict the behavior of a particular system in a "bottom–up" approach once its component parts are sufficiently characterized, none of these approaches directly address the challenge of integration by transcending various levels organization and elucidating the basic rules involved.

In view of the rapid rise of molecular biology and genomics, some biologists did voice caution about the limitations of this descriptive and reductionist stance (12,33,37,45,49,55,62,53), essentially calling attention to Aristotle's insight that the "whole is different from the sum of its parts." However, it is only now, at the threshold of post-genomic and systems biology, that life scientists are beginning to realize that an accurate description of all the parts that comprise a living cell is not equal to understanding how it functions (29). Sometimes capturing the impression of the whole picture with a glance can give deeper insights and yield information not obtained by reproducing it pixel by pixel. Although biologists have yet to adopt the approach of "coarse-graining" to gain insight into fundamental, system-wide properties, this method is often used by physicists (25,71). Only by adjusting our focus plane to various levels of organization, and "zooming" in and out on the magnification, can we reveal the fundamental principles that govern what makes the whole (the organism) different from the sum of its parts (the molecules).

To do so, biologists must free themselves from their "divide and conquer" mentality and their adherence to molecular description, be it qualitative or quantitative, one at a time or in massively parallel fashion, as the only mode of explanation. Instead, they must join physicists in their willingness to embrace abstraction and generalization. Conceptual and formal tools are also needed that go beyond descriptive mathematical modeling of particular molecular pathways. In fact, at the same time that the molecular biologist now faces this new challenge, the science of "complex systems" appears to have matured into a discipline in its own right. Even if a rigorous scientific underpinning remains to be established (if it is possible at all), it has encouraged scientists from various fields, such as physics, biology, engineering, business, and the social sciences, to join forces and to a take more formal and general approach to understand principles of complex systems rather than simply creating models that only reenact them in silico with all the details.

A specific formalism that is particularly useful here is based on the idea that a complex system can be treated as a *network* of interacting parts in a most general sense. A network can be a physical or mechanical structure, as well as an abstract representation of how information flows between interacting elements

within a system. Thus, in this chapter, we focus on the principles of organization that govern how cells control both their physical structure and biochemical function as a result of interactions within underlying networks of interacting of proteins and genes. In the process, we will describe how simple, rule-governed behaviors, such as cell shape stability and cell fates, represent "emergent properties" of the underlying molecular networks. Finally, we raise the possibility that the interface between these two networks, one structural and the other informational, is at the core of the evolution and functioning of complex living systems, such as cells and whole organisms, that operate at size scales much higher than that of molecular reactions and flows.

2. COMPLEXITY IN LIVING SYSTEMS

How does the information encoded within DNA and biochemical reactions map into the observable properties of living cells that comprise all organisms? This old riddle of the genome–phenome relationship can be split into two more specific questions: (1) How do interactions between biochemical components lead to the production of a physical object with distinct structural and mechanical properties characteristic of a living cell? and (2) How do interactions among genes and proteins lead to the development of a coherently functioning machinery for information processing that governs how living cells will behave and adapt to their surroundings (see Figure 1)? Thus, we need to address the question of how higher-level behaviors emerge, in the context of both the hardware (structure) and the software (information-processing programs) of the cell.

In contrast to simpler, widely discussed emergent phenomena, such as pattern formation in physicochemical systems or patterns of animal flocks, the emergence of new properties in biology has two particular characteristics which are absent in most non-living complex systems:

(i) ***Hierarchy of multiple levels of emergence***. New properties emerge at multiple hierarchical levels that cover many size scales in living organisms. Genes and amino-acid sequences determine protein structure, i.e., their three-dimensional (3D) shape and mechanical properties. Proteins and other macromolecules (e.g., lipids, sugars, nucleic acids) self-assemble to create functional multimolecular complexes and intracellular organelles, such as the ribosomes, mitochondria, the nucleus, the cytoskeleton, and the plasma membrane, as well as the extracellular matrix (ECM). These organelles together form the cells. Cells interlink with each other and with the extracellular matrix to form tissues. Multiple tissues combine to form organs that, in turn, are linked together to form the organism.

As the elements at each level (molecules, organelles, cells, tissues, organs) interact, they give rise to "emergent properties" that are characteristic of the next higher level. For example, individual proteins may exhibit a low level of cata-

lytic activity and move free in solution. However, when multiple proteins with different enzymatic activities assemble together, they can form a higher-order enzyme complex that exhibits stable 3D form as well as novel functions based on coupled metabolic processing activities. For example, the pyruvate dehydrogenase enzyme complex has a mass approaching 10 million Daltons in mammals and it exhibits a highly organized, pentagonal dodecahedral shape (69). Similarly, individual cells of the pancreas can secrete digestive enzymes in a polarized manner (i.e., from the apical pole of the cell); however, disease (pancreatitis) results if these cells dissociate from each other and their orienting extracellular matrix scaffold, and lose their higher-order tissue architecture. Thus, each level has its specific rules of interaction that involve structural as well as dynamic constraints since the nature of the parts and interactions of each level are different. Therefore, unlike fractals, we have discrete layers of patterns governed by distinct, level-specific rules, and there is no general "scale-invariance," although some principles apply to various scales, as we will see.

(ii) ***Heterogeneity of interacting elements***. The entities in systems with emergent (e.g., individual molecules with characteristic 3D structure and function, multimolecular complexes with novel enzyme activities, organelles with specialized metabolic functions, living cells that move and grow) do not form a uniform population, as is the case for the molecules in self-organizing patterns of chemicals or of individuals in schools of fish. Instead, these entities are unique individuals, or belong to classes of entities with similar properties that can be clearly distinguished from each other. For instance, cells that arise from molecular self-assembly and give rise to tissues can be classified into hundreds of qualitatively different classes or types (e.g., liver, muscle, nerve, skin).

The complexity of each biological network leads to individuality of the emergent entities (e.g., a cell type), even if their component parts are identical (e.g., the genes). This variety of individual entities, which serve as building blocks, enables combinatorial diversity at the next level of integration (e.g., tissues). This introduces new types of interaction rules at each size scale, thus adding a unique layer of complexity characteristic of living organisms. As mentioned earlier, it is the heterogeneity and size of the population of the molecular parts that has necessitated both the massively parallel analytical methods and the detailed modeling approaches of systems biology. However, systems biology does not currently embrace the concepts of hierarchy and heterogeneity in complex systems.

3. MODEL: NETWORKS AS THE GENERAL CONCEPTUAL FRAMEWORK

Given that the cells are the most basic building unit of life, which itself is a complex system, we will first focus our discussion on how cell shape and func-

tion emerge from interactions among thousands of interacting molecules and genes. Our goal is to uncover principles that govern how many heterogeneous, interacting molecular components can self-assemble and self-organize to produce higher-level features characteristic of whole living cells. However, as we will show, the same design principle may also govern how emergence occurs at higher levels of organization (e.g., tissues, organs, organism), even though the higher-level networks are composed of different players with distinct rules of interaction.

In chemistry, aggregate variables may be used to represent an average property of a homogenous population of parts. Unfortunately, because of the hierarchy of emergence and the heterogeneity of the parts in a whole organism, this approach is not well suited to describe living systems. In fact, this is the major limitation in most past studies that attempt to explain cell structure and mechanics using conventional engineering approaches (e.g., continuum mechanics), as well as cell function using laws of mass action for molecular interactions.

In contrast, networks provide a simple general formalism for abstraction in order to study how the collective action of interacting parts gives rise to emergent properties, and thus, a means to handle hierarchical complexity. Because the essential ingredients that make the whole different from the sum of its parts are the *interactions* between the heterogeneous components, a biological system can be formalized as a large *network* that consists of the component elements (the molecules) and their links (their interactions), which need not be identical. The major point here is that network models can be applied to both *structural* systems (i.e., physical scaffolds that lend mechanical stability to the network) and *information-processing* systems (i.e., the abstract diagrams that represent how elements of the network influence each other's activities and the behavior of the whole). Applied to mammalian cells, the structural network is the "cytoskeleton" that determines how the building blocks (proteins) are physically attached to each other to give the cell its physical shape and mechanical stability. The information-processing network is the regulatory network that determines how the state of interacting elements (genes and proteins) influence each other, and thereby process the information that is encoded in the genome or received from the external milieu to generate a distinct cell behavior.

The common basic property of both networks in living cells, one structural and the other informational, is that the collective action of their constituent molecular elements gives rise to a system with emergent properties. However, there is a fundamental formal difference. Structural network models describe a concrete object while information networks are an abstraction. Nevertheless, while one may think of the structural network as a physical scaffold in which every individual building block (including all members of the same class of elements) has to be depicted in the model, like in an architect's blueprint, the model of the cell that we discuss below also offers some abstraction. Specifically, models that contain a few elements (with their prototypic mechanical properties) mimic

characteristic global properties of whole living cells that contain millions of such elements. On the other hand, information-processing networks, i.e., signal transduction and gene regulatory networks, are full abstractions in that every individual interacting element that occurs only once in the model actually represents hundreds to billions of copies of that particular type of molecular species. Another difference between the structural and information networks is that the former takes into account position and physicality, whereas interactions in the latter can be represented mathematically as a graph because, in a first approximation, space does not play a role (see also this volume, Part II, chapter 4, by Wuchty, Ravasz, and Barabási).

At first glance, one might think that the structural cytoskeletal network maintains the cell's shape, whereas the biochemical information-processing network determines the behavioral state of the cell. However, as will be discussed below, evolution has led to assembly of cells in which structure and information-processing functions are tightly coupled—a fundamental property of living systems at all size scales. Another central property of biological systems, such as a cell, is that they need to be stable, yet flexible. Cells are continuously challenged by chemical and physical stimuli: not only do cells have to resist random perturbations and maintain their structure and behavioral program, they also have to be flexible enough to respond appropriately to specific external signals that require distinct changes in both cell mechanical and biochemical behaviors, such as during cell migration and differentiation. Interestingly, death of both cells and whole organisms is characterized by a rapid increase in rigidity (*rigor mortis)*, with a complete loss of the flexibility that dominates the living state. Thus, this unification of *robustness* with *flexibility*, both in terms of cell structure and behavior, is a hallmark of complex living systems.

4. RESULTS

4.1. Structural Networks in Cells

4.1.1. From Molecular Biochemistry to Cellular Mechanochemistry: The Cytoskeleton

Cells are comprised of thousands of molecules that are arranged and connected in specific ways so as to produce distinct structures and biochemical functions; they are not membranes filled with viscous colloidal solution. In particular, mammalian cells contain an internal molecular framework or "cytoskeleton" that provides shape stability to the cell, and orients much of the cell's metabolic and signal-transducing machinery (35,36). The cytoskeleton is an interconnected 3D network or lattice comprised of three major classes of filamentous protein polymers—microfilaments, microtubules, and intermediate filaments. A subset of the microfilaments that contain myosin as well as actin

are contractile; these filaments actively generate mechanical tension through a filament–filament sliding mechanism similar to that used in muscle (35,39). Thus, the entire cytoskeleton and cell exists in a state of isometric tension. In essence, by organizing this multimolecular network, the cell translates a structureless chemistry into a physical entity with well-defined mechanical properties. For example, this tensed intracellular scaffold is largely responsible for the viscoelastic properties of the cell. It also generates the tractional forces that drive cell movement as well as changes in cell shape.

4.1.2. Cellular Tensegrity

Past work on cell shape and mechanics ignored the cytoskeleton and assumed that the cell is essentially an elastic membrane surrounding a viscous or viscoelastic cytosol (16,17,22). In contrast, over the past twenty years, we and others have been able to show that the cytoskeleton is the major determinant of cell shape and mechanics, and that cells may use a particular form of architecture known as "tensegrity" to organize and stabilize this molecular network (reviewed in (39)).

Tensegrity was defined by Buckminster Fuller as a building principle in which structural shape of a network of structural members is guaranteed by continuous tensional behaviors of the system and not by local compressional member behaviors (21). The purest representation of the tensegrity principle is found in the creations of the sculptor, Kenneth Snelson, which are composed of a continuous network of high-tension cables and a discontinuous (isolated) set of compression struts (Figure 2). However, the tensegrity principle also applies to

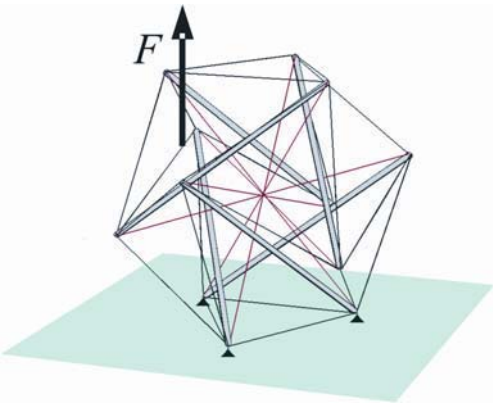


Figure 2. **Tensegrity model**. A prestressed tensegrity structure composed of 6 compression-resistant struts (white struts) interconnected by 24 tension cables (black lines) on its periphery; this model also contains radial cables connecting the ends of the struts to the cell center (red lines). The theoretical tensegrity model of the cell is based on this architecture. In the cell model, the black lines correspond to viscoelastic actin cables, the red lines to viscoelastic intermediate filaments (of different time constants), and the white struts to rigid microtubules.

all geodesic structures (37,39). Tensegrity networks have the property that they are self-stabilizing in the sense that they yield equilibrium configurations with all cables in tension only due to the internal interactions between its components, i.e., without the need for external forces. Specifically, the tension members pull against the resisting compression members and thereby create an internal tensile stress or "prestress" (isometric tension) that stabilizes the entire system (the same prestress may be generated by the compression members pushing out against a surrounding resistance network). Moreover, both multimodular and hierarchical tensegrity networks can be created that are governed by the same rules and that exhibit integrated system-wide behaviors when exposed to external stress (39).

In the cellular tensegrity model, the whole cell is a prestressed tensegrity structure; however, geodesic structures are also found in the cell at smaller size scales (37,39). In the model, tensional forces are borne by cytoskeletal microfilaments and intermediate filaments, and these forces are balanced by interconnected structural elements that resist compression. These latter elements include microtubule struts within the cytoskeleton and cell surface adhesions to the surrounding extracellular matrix. However, biological systems are dynamic and highly complex in that individual filaments can have dual functions and hence bear either tension or compression in different structural contexts or at different size scales. The tensional prestress that stabilizes the whole cell is generated actively by the actomyosin apparatus within contractile microfilaments. Additional passive contributions to this prestress come from cell distension through adhesions to the ECM and other cells, osmotic forces acting on the cell membrane, and forces exerted by filament polymerization. Intermediate filaments that interconnect at many points along microtubules, microfilaments, and the nuclear surface provide mechanical stiffness to the cell based on their material properties and on their ability to act as suspensory cables that interconnect and tensionally stiffen the entire cytoskeleton and nuclear lattice. In addition, the internal cytoskeleton interconnects at the cell periphery with a highly elastic, cortical cytoskeletal network directly beneath the plasma membrane. The entire integrated cytoskeleton is then permeated by a viscous cytosol and enclosed by a differentially permeable surface membrane.

Unlike the isotropic viscous cytoplasm that dominated past models of cell mechanics, the tensegrity-stabilized cytoskeletal network optimizes structural efficiency (strength/mass ratio) by relying on internal tension, rather than on continuous compression when exposed to an external force. Tensegrity systems also can easily change shape with minimal energy consumption, for example, as compared to classical truss structures that require an excessive amount of energy even for minor shape modification. Most importantly, as with all complex networks composed of multiple interacting components, the macroscopic properties of tensegrity networks (e.g., their mechanical stability, ability to grow and rear-

range, structural efficiency, viscoelastic behavior) are emergent properties that arise from the particular architecture used to stabilize the 3D network.

4.1.3. Computational Tensegrity Models Predict Complex Cell Behaviors

If the complex mechanical behaviors of cells, including their global mechanical stability, flexibility, ability to remodel, and optimal strength/mass ratio, represent emergent properties of cell structural networks, then we should be able to get insight into how this takes place by studying and modeling cytoskeletal mechanics (60). Existing paradigms assume that the static and dynamic mechanical behaviors of living cells respectively originate from two distinct compartments—the elastic cortical membrane and the viscous cytoplasm. Recent work, however, has revealed that cell dynamic behavior reflects a generic system property of the cell at some higher level of molecular interaction as it is characterized by a wide spectrum of time constants (18). Dimitrije Stamenovic, working with our group (58), and with others (72,73,65,66), have shown that a theoretical formulation of the cellular tensegrity model based on first mechanistic principles can predict various static mechanical properties of living mammalian cells. More recently, we found that the tensegrity model also can explain dynamic cell mechanical behaviors, as described below.

The theoretical tensegrity model of the cell is a deterministic physics-based model which assumes that contractile microfilaments and intermediate filaments carry a stabilizing tensile prestress in the cytoskeleton that is balanced by internal microtubule struts and extracellular adhesions. The cytoskeleton and substrate together were assumed to form a self-equilibrated, stable mechanical system; the prestress carried by the cables is balanced by the compression of the struts. The simplified tensegrity network used in the computational model is composed of 24 tensed, linearly viscoelastic (Kelvin-Voigt), "microfilament" cables and 6 rigid "microtubule" struts; 12 additional tensed Kelvin-Voigt "intermediate filament" cables extend from the surface of the structure to the cell center and the basal ends of 3 struts are fixed to mimic cell substrate adhesion (Figure 2). Importantly, work on variously shaped models has revealed that even the simplest prestressed tensegrity network embodies the key mechanical properties of all prestressed tensegrities (this is the degree of abstraction mentioned above). In this computational model, the material properties of the tensile filaments can be varied independently. The equilibrium solution around which the linear mathematical model was derived for frequency response calculations is a prestressable configuration (63). The prestress is a measure of the tension in the cables. The input was a vertical, sinusoidally varying force applied at the center of a strut; the output was its corresponding vertical displacement (Figure 2).

Analysis of the variations of the dynamic elastic modulus G' and dynamic frictional modulus G'' with the level of prestress for various frequencies in this

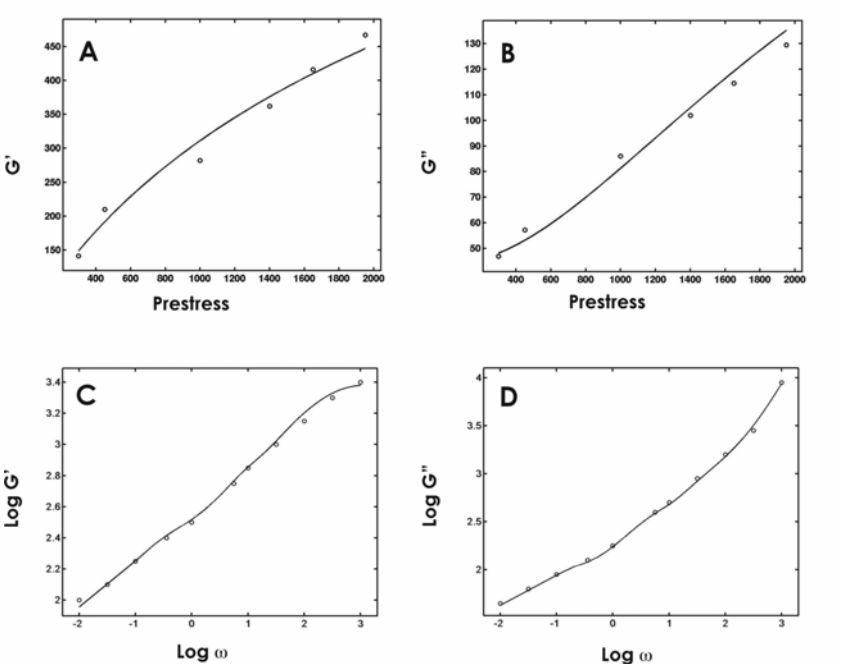


Figure 3. **Emergent mechanical properties of the tensegrity model: simulation versus experimental data**. **A**. Cellular elastic (*G*') and frictional (*G*") moduli predicted by a computational tensegrity model (solid lines) versus data obtained from experiments with living cells (circles) (64). Data in **A** and **B** that show the dependencies of moduli on prestress are re-plotted from Stamenovic et al. (59); data in **C** and **D** that show frequency dependencies are re-plotted from Fabry et al. (18). The frequency (ω) is given in Hz, whereas prestress and elastic and frictional moduli are in Pa.

computational tensegrity model revealed that these dependencies increased approximately linearly over a wide range of prestress (64). These results nicely mimic experimental observations that demonstrated the same behavior in living cells (59) (Figure 3) and confirm similar results obtained with a slightly different tensegrity structure (5). Deviations from the experimental results were only observed at very low prestress, where the cables are almost slack. Adherent living cells actively generate tension within their contractile microfilaments and, thus, their cytoskeleton is always prestressed. Importantly, analysis of the frequency dependencies of *G*' and *G*" of the tensegrity structure also revealed a wide distribution of time constants that closely mimicked behavior previously observed in living cells (18) (Figure 3). Similar results were obtained for other types of loading and for tensegrity structures of higher complexity (64). However, better re-

sults were obtained with heterogeneous models in which different filaments exhibited different levels of stiffness.

Thus, a key feature of the cellular tensegrity network—the level of cytoskeletal prestress—is critical for control of both static and dynamic mechanical behavior in whole cells. As predicted by the model, the global system architecture and inhomogeneity of time constants between individual elements also significantly contributes to the emergent properties of the system: the whole network behaves differently than an individual Kelvin-Voigt cable. This finding that both elastic and frictional behaviors of living cells naturally fall out from the tensegrity model indicates that the viscous properties of mammalian cells are not due to fluid behavior of the cytosol. Rather, these complex mechanical properties of cells emerge from collective mechanical interactions among the distinct molecular filaments that comprise the cytoskeletal network. These results emphasize the importance of the tensionally prestressed cytoskeleton for cell mechanical behavior and add further support for the universality of the cellular tensegrity model (37,39).

4.1.4. Biological Implications of Tensegrity Beyond the Cytoskeleton

In a more encompassing biological interpretation, a mechanical design principle that uses networks composed of discrete elements rather than a single mechanical continuum allows molecules (e.g., the proteins that form the filaments) to bridge the gap between microscopic structureless biochemistry and macroscopic mechanics and pattern in just one step of self-assembly. However, as mentioned in the introduction, living systems harbor a hierarchy of many levels of emergence over many size scales. Of interest thus is that the principle of tensegrity is scalable, and in fact operates at various size scales, from molecule to organism (9,37,39). For example, tensegrity may govern how individual molecules, such as proteins, and multimolecular structures (e.g., lipid micelles) gain their mechanical stability and 3D form (19,37,39,75). Geodesic forms also are dominant in molecular systems including viruses, the simplest example of a "living" system; interestingly, tensegrity was used to explain the geodesic structure of viral capsids (7).

At a larger size scale in living tissues, cells are attached to anchoring scaffolds that are also 3D structural networks composed of fibrillar extracellular matrix molecules. Because cells apply cytoskeletally generated tractional forces on their adhesions, these extracellular matrix networks are also prestressed and hence stabilized through tensegrity. Local increases in tissue tension are sensed by individual adjacent cells that respond by switching into a proliferative state, thereby increasing cell mass to match increases in applied macroscopic forces. In this manner, tension-dependent changes in cell growth allow higher-order tissue and organ structures, such as glandular buds and brain gyri, to be sculpted

by organ-level mechanical forces during morphogenesis. Finally, provision of stability and flexibility within a biological network through a tensegrity force-balance is most obvious at the highest level of organization in the hierarchy of life. The musculoskeletal system that allows humans to walk and hold our bodies in various positions gains its stability through a balance between continuous tension (muscles, tendons, ligaments) and local compression (bones) that generates a tensile prestress (tone).

Thus, tensegrity appears to represent a fundamental design principle that is used to stabilize biological networks at all size scales in the hierarchy of life, as well as throughout evolution (37). The flexibility and stability provided by use of tensegrity also may have contributed significantly to the process of hierarchical self-assembly and environmental selection that first led to the origin of cellular life (38), as well as to the development of multicellular organisms comprised of interconnected networks of cells, tissues and organs (39).

Importantly, the complex mechanical behaviors of a tensegrity system represent emergent properties of the whole network, and not properties of the individual structural members. For these reasons, tensegrity may provide a means to incorporate "physicality" and spatial constraints into models of complex network systems that commonly are only thought of in terms of information flow. Interestingly, most biochemical reactions proceed in a "solid-state" in living cells, i.e., many of the enzymes, substrates, and reactants are physically immobilized on insoluble cytoskeletal scaffolds (6,36,50). Thus, mechanical properties of structural networks, and hence tensegrity principles, may also directly impact information flow in biological systems, as will be discussed below.

4.2. Information Networks

On the hardware side, the mechanical properties of the cells are the obvious properties that emerge from interactions between structural proteins. In contrast, on the software side, the emergence of some simple, fundamental, higher-level features from interactions among regulatory genes and proteins is not immediately apparent. Here we show that, despite the complexity of molecular pathways within a cell, global cell behaviors associated with a change of phenotype exhibit simple rule-governed properties that emerge from interactions in the regulatory network of the cell.

4.2.1. Cell Fates as Emergent Properties

The global behavior of a cell within a tissue in a multicellular organism can be reduced to a few behavioral modes or phenotypes, the so-called "cell fates": the proliferative state, the migratory state, differentiation, senescence or the state

of commitment to cell death (apoptosis), etc. (30). During proliferation, cells are in a biochemical state in which they can replicate DNA and divide to increase tissue cell mass. During differentiation, cells undergo a phenotypic change from an immature precursor cell to a distinct cell type, such as a red blood cell, liver cell or nerve cell, that carries out tissue-specific tasks. In apoptosis, cells respond to particular signals by switching on a suicide program and undergo cell death. Each cell fate is characterized by a distinct profile of activation of the 30,000 or so genes in the human genome.

Cell fates are stable, typically mutually exclusive cellular states (27,30,67). The conditional selection of these cell fates within the population of cells in a tissue gives rise to the next level of emergence: the tissue and organs that consist of distinct spatial patterns of cells that exhibit different fates, including various specialized cell types. The tissue is a cellular society that requires social behavior of its members in order to maintain its global structural and functional stability. Thus, the balance between division, differentiation, and death of individual cells needs to be tightly regulated within different tissue microenvironments so that the whole tissue optimally responds to all environmental signals.

Cell fate switching is governed by a molecular network of genes, proteins, and other cellular components that give rise to the emergent property we recognize as cell fate. For simplicity, let us here focus on the gene regulatory network, and ask the more general question: How can the mutual regulation of ~30,000 genes in the genome give rise to stable, mutually exclusive cellular states (fates), each characterized by a distinct gene activation profile? For example, why does a differentiated liver cell not drift away to become a nerve cell if the difference is just in the pattern of gene activation? As described above, another important property for development is that cells unite stability (maintenance of identity in response to perturbations) with flexibility (ability to change identity in response to critical stimuli); in fact, it is this property that allows development to take place in the first place. This and other qualities are fundamental, emergent properties that, as we will see, arise as a consequence of how information is processed by the underlying gene regulatory network and the architecture of that network.

4.2.2. Network Dynamics Leads to Stable States: Attractors in Gene Regulatory Networks

Let us examine how gene regulatory interactions can collectively give rise to a global network behavior that satisfies the requirements for development of a specialized cell phenotype, and eventually, a whole living organism. Without distinct regulatory interactions between the genes, any combination of gene activities across the genome would be possible. This would result in an unstructured continuum of gene activation profiles, but no directed developmental

processes, no robust expression profiles, and hence no differentiated cell types. In reality, genes interact with each other through the regulatory proteins they encode. Each gene (or its encoded protein) has a very specific set of interaction partners based on its molecular structure, and each interaction has a distinct mode (e.g., stimulatory, inhibitory). Thus, the genome contains a hard-wired interaction network. Interactions between genes therefore introduce constraints in the whole network, such that many gene activation profiles become unstable and are never realized. It is this collapse of the vast space of theoretically possible configurations of gene activity combinations that leads to distinct dynamics and the robustness of a limited number of cell phenotypes (43).

In technical terms, constraint of the dynamics by these molecular interactions means that the *high-dimensional state space* of gene activation is highly structured. One can define a state space as the *N*-dimensional space in which every point represents a different network state defined by a distinct gene activation profile, where *N* is the number of genes. Now assume that gene *A* unconditionally inhibits the expression of gene *B*; then all network states in which both *A* and *B* are active will be unstable, thus forcing the network to "move" in state space until it hits a stable state. The network may also cycle between a few states. Thus, taking all the interactions into account, it can be shown that the network can change its activity profile in only a few directions (following stable trajectories) until it reaches a stable state, the so-called "*attractor state*," which can be a fixed-point or cycling attractor (42,43). The existence of unstable regions and of multiple stable attractors impose a substructure to the state space, which might be imagined as an "*attractor landscape*," as shown in Figure 4. Accordingly, the state of the network (and hence of the cell) can be viewed as a marble on that landscape: it is forced to roll along valleys (trajectories) into the pits (attractors) (Figure 4). This attractor landscape therefore captures the constrained, global dynamics of cell fate switching (i.e., phenotypic control). In fact, Waddington, Delbrück, Monod, Jacob, Kauffman, and others have all proposed (in various forms) that the distinct, phenotypic differentiation states that we observe in living systems correspond to attractors in the state space defined by the molecular activities of the underlying network (15,42,48,68). Thus, attractors in the state space map into stable phenotypic states (differentiation to distinct cell types, cell proliferation, programmed cell death, etc.), and the trajectories represent directed developmental processes.

In the landscape of a real gene regulatory network, the attractors would represent cell states that are stable to many random perturbations. At the same time, the network would allow the cell to switch to other attractors given the appropriate sets of conditions, such as the presence of external regulatory signals that promote a particular cell fate (30). This highly structured landscape with latent, "preexisting" possibilities creates the stage on which the developmental program is played out. Interestingly, Waddington similarly proposed an "epigenetic landscape," with a marble whose position represents developmental state rolling

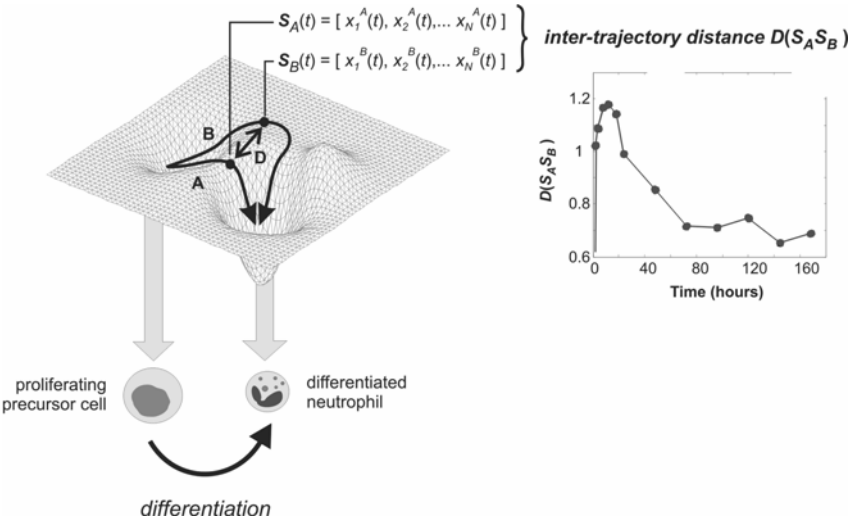


Figure 4. **Cell fates as attractors**. The structure of the *N*-dimensional state space (*N* = number of interacting genes in the network, e.g., *N* = 10,000) is schematically shown as a three-dimensional topographic "attractor landscape" that is conceptually equivalent to Waddington's "epigenetic landscape" (Figure 5). Each point in the landscape represents a cell state *S*, defined by the profile of the activation state *x* (measured as mRNA level) of all the *N* genes: $S = [x_1(t), x_2(t), ... x_N(t)]$. The pits in the landscape are the attractor states that represent the stable cell fates, in this case the precursor cell and the differentiated neutrophil. Transition into the differentiated state can be triggered by two pharmacologically distinct differentiation-inducing agents that perturb the state of the precursor cell in different ways such that the cells takes two different trajectories, *A* and *B*, respectively, to reach the neutrophil state. Monitoring the change of $S(t)$ along these two high-dimensional trajectories, $S_A(t)$ and $S_B(t)$, respectively, as the change of gene expression profile using DNA microarrays would allow calculation of the inter-trajectory distance *D*. The inset on top illustrates a case for the time course of *D* for a set of 2600 genes in a differentiation process. In this case, the course of *D* shows initial, rapid divergence of the trajectory, followed by terminal convergence in more than 50% of the state dimensions as the cell reaches the differentiated state, indicating the approach to a high-dimensional attractor state. Reprinted with permission from Huang (76).

down valleys (Figure 5) based on his observation that cells "switch between distinct, well recognizable types" during development, and that intermediates are rare and unstable (67,68). This picture captures the basic rules governing cell fate dynamics, and we can now argue that, although proposed as an intuitive representation, Waddington's epigenetic landscape is in principle the state space of the molecular network that controls cell fates. Thus, the attractor landscape represents the emergent properties of the interaction network.

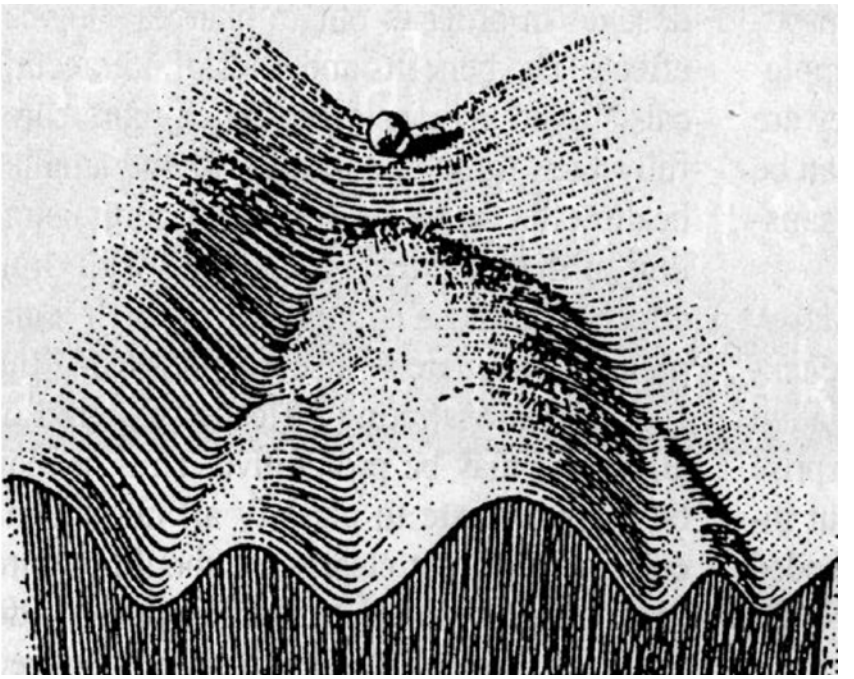


Figure 5. Waddington's idea of "epigenetic landscape." Although the model first proposed by Waddington in 1940 to explain a cell's decision between distinct, "discrete" developmental fates was merely an intuitive metaphor, it may be regarded as the structure of the state space representing the dynamics of gene regulatory networks, the "attractor landscape" in Figure 4. (Reprinted with permission from Waddington (1956) (67).

Much as the architecture of the cytoskeletal network affects the emerging macroscopic mechanical properties of cells and tissues, the architecture of the gene regulatory network (the wiring diagram of the gene–gene interaction) determines the specific "topography" of the attractor landscape, and hence cell behavior. However, not all network architectures give rise to a "reasonable" state space structure. For instance, using simulations of random continuous or discrete network models, it has been demonstrated that a fully connected network (in which every gene affects every other) would be unstable, that is, devoid of fixed-point attractor states. In contrast, sparsely connected networks are more likely to produce a dynamics with stable states (26,43). Apparently, the network architecture has been shaped through evolution by continuous growth (increase in gene number) and rewiring such that it gives rise to "biologically reasonable" dynamics with multiple stationary, stable attractor states. These attractors are just stable enough to resist random perturbations, but at the same time they allow the existence of multiple cell fates and switching between them in response to distinct perturbations during embryonic development.

The recent availability of information about large protein and gene networks (containing thousands of components) in baker's yeast (*S. cerevisiae*) made possible by new large-scale, high-throughput biochemical methods has stimulated investigations into the natural architecture of biological regulatory networks (see also this volume, Part II, chapter 4, by Wuchty, Ravasz, and Barabási). These studies revealed that the almost genome-wide protein-interaction network and gene regulatory network are indeed sparsely connected in these cells. Moreover, several interesting features of the network architecture, such as a near power-law distribution of connectivity (number of interaction partners per molecule), a propensity to modularity, and use of hierarchical structure, were found to be present (41,46,70). Interestingly, a power-law architecture appears to have beneficial consequences for system-wide dynamics (1,20). Specifically, the regime in the space of possible network architectures for "biologically reasonable" networks (i.e., those that exhibit ordered behavior with small attractors) is larger because the networks tolerate higher connectivity without becoming chaotic.

4.2.3. Biological Implications of Attractor States

As in the case of structural networks and the tensegrity model that allows prediction of some macroscopic mechanical properties of the cell based on emergent features of the model, the generic global behavior of the cell is predicted by the model of an attractor landscape. In fact, the existence of distinct stable cell fates (proliferation, apoptosis, quiescence, etc.) and of different differentiated cell types (liver, skin, neuron, etc.) that are robust to many perturbations, yet can switch between distinct states under restricted conditions, is itself a strong indication that they are attractors of an underlying molecular network. Similarly, robust developmental trajectories, corresponding to long valleys leading to lowest points in the landscape, can be explained as emergent properties of the genome-wide network of genetic interactions. However, the dynamic networks approach and the attractor landscape idea may also provide new insight into other cell biological phenomena that have previously resisted straightforward explanation by the conventional paradigm, which emphasizes the role of individual signal-transduction pathways.

Cell fate regulation in tissue homeostasis. As predicted by the dynamic network model, cell fates represent discrete, mutually exclusive, stable states that require specific signals to transition to each other, when such a transition is possible. For instance, differentiation and proliferation are well known to be mutually exclusive and robust (27,67); in many cell systems just quitting the proliferation state by overexpressing the cell cycle inhibitor protein p21 forces the cell to automatically enter the differentiation program (14,52,61). That cell fates are robust and can be realized just by "placing" cells in the corresponding

"basin of attraction," from which they will reach the attractor state, is best recapitulated by the observation that many nonspecific pharmacological stimuli that activate multiple proteins across several signaling pathways often trigger expression of the same set of cellular phenotypes. For instance, differentiation of many cell types can be turned on by a large variety of nonspecific agents, including DMSO or ethanol (3,44,47,57,74). Specifically, the differentiation of a promyelocyte cell line into mature neutrophils (the major white blood cells involved in innate immune response) can be elicited not only by DMSO, but also by treatment with retinoic acid, hypoxanthin, actinomycine D, flavone, etc. (13). In these cases, it appears that simultaneous perturbation of multiple targets in different pathways results in channeling of the biochemical effects into common end-programs, and hence the same "default" cell fate.

Perhaps the most striking cellular manifestation of the idea that cell fates represent attractor states comes from experiments in which cell shape was varied as an independent control parameter using microfabricated geometric islands of extracellular matrix proteins to which mammalian cells normally adhere (10,32). The traditional mechanistic, pathway-centered explanation of cell fate switching assumes that a specific, "instructive signal," i.e., a messenger molecule that interacts with its cognate cell surface receptor, tells the cell which particular genes to activate in order to establish a new cell phenotype. However, when these instructive signals (e.g., soluble growth factors and insoluble extracellular matrix molecules) were held constant, cell shape distortion alone was able to switch endothelial cells between proliferation, apoptosis, and differentiation (31). Thus, variation in one continuous control parameter (cell shape) that is devoid of the molecular specificity normally assumed to carry "instructive" information led to switching between multiple, mutually exclusive cell fates, and produced effects reminiscent of a biological "phase transition." Essentially, cell distortion triggered the cell to "select" between different preexisting attractor states.

Integration of structural and information networks. Importantly, because cell shape is governed by changes in cytoskeletal shape and mechanics, phenotypic control by cell distortion is a clear example of how structural networks can impact information-processing networks in living cells. From a mechanistic point of view one can then ask, how can a "nonspecific" parameter, such as cell shape, elicit the detailed molecular changes associated with cell growth, differentiation, and apoptosis? If cell fates are attractors, then a large variety of molecular signals will push the cells into the few available behavioral modes that the cell can adopt: again, regulation corresponds to selection among a limited number of preexisting fates, rather than instruction of how to behave.

Changes in cell shape imposed by the microfabricated constraints lead to massive rearrangements of the cytoskeleton that maintains shape stability in response to external influences according to the tensegrity rules that govern these structural networks. Visualization of the actin cytoskeleton in cells grown on micropatterns, for example, revealed that the actin bundles of the cell reorient

depending on the shape of the microfabricated adhesive island and map out tension field lines within the cell (8,51).

But how does cell mechanics affect cell fate? Cell anchoring to extracellular matrix substrates, such as these islands, is mediated by cell surface integrin receptor molecules that cluster within small anchoring sites known as "focal adhesions." Actin filaments insert at these focal sites of attachment between integrins and the extracellular matrix, and apply traction forces to these adhesions much like the tension in a tent membrane is transmitted through ropes to the pegs that anchor it into the ground. These cellular anchoring structures at the cell membrane are also the nucleation sites for the formation of large, multimolecular complexes of proteins that are involved in signal transduction, and hence mediate cellular information processing (23). Such complexes at adhesion sites facilitate the interactions between signaling proteins. For instance, the activation of many of the signal-transduction molecules, such as the aforementioned ras–raf–MEK–Erk mitogenic pathway, depends on the configuration of the actin-cytoskeleton (28), whereas assembly of signaling protein complexes at the focal site depends on the tension in the actin bundles (11). Thus, focal adhesions also represent sites of mechanotransduction (24,34): they sense the mechanical tension of the cell that is modulated in response to the geometry of the environment (51).

As described above, the 3D shape of molecules dictates their mechanical and biochemical behavior—another example of emergence from the level of their component parts (e.g., from amino acids to catalytic enzymes). Importantly, altering molecular shape through chemical modification or mechanical distortion alters biochemistry by changing thermodynamic and kinetic parameters (37,40). The biochemical information-processing network of the cell is therefore governed by physical interactions that depend on the 3D shape and mechanical properties of the individual molecules and hence on the state of the cytoskeletal network that they comprise. Thus, structural networks and information networks integrate as a result of mechanochemistry. Specifically, mammalian cells contain structures that link cytoskeleton with signaling pathways, thereby allowing mechanical forces to feed back to regulate cellular information processing.

The biochemical details as to the precise molecules that transduce the mechanical forces into biochemical signaling are still not fully understood, although strong experimental evidence now supports the implication of several specific signal-transducing proteins (23). However, given the fact that researchers commonly strive to uncover all of the "instructive pathways" by which cell fates are regulated, it appears that, instead of being carried along linear molecular pathways, information is processed in a distributed manner over the network of interacting regulatory molecules. Many of these molecules physically associate with the load-bearing elements of the structural cytoskeletal network that stabilizes cell shape. If the activities of associated regulatory molecules were to

change in response to mechanical distortion, this integrated structural and information-processing network would be perfectly designed to sense the diffuse signals that emanate from a concerted rearrangement of the cytoskeleton in response to mechanical stress or physical changes in cell shape. In fact, both molecules that physically associate with the cytoskeleton in the focal adhesion site and at other locations throughout the cell have been shown to change their activity in response to applied mechanical stress or cell distortion (23).

Because the wiring of the signaling networks produce attractors that correspond to only a limited number of distinct cell fates, the cell may naturally and reliably sense a broad spectrum of signals and simultaneously orchestrate multiple molecular responses to produce coherent behavioral programs. In other words, the existence of attractors representing distinct cell behaviors facilitates the evolution of a form of regulation that connects signals devoid of molecular specificity like mechanical forces to the internal regulatory machinery that governs specific cell fates.

4.2.4. Experimental Evidence for Attractors in Gene Regulatory Networks

The characteristic dynamics of cell fate control, the mutual exclusivity of different phenotypes, and their robustness in living cells all suggest that distinct cell fates represent attractors that emerge in the dynamic network of gene regulatory interactions. But can we directly view the structure of the attractor landscape network at the molecular level without knowledge of the precise wiring diagram of the underlying genome-wide regulatory? To map out this state space, it would be necessary to simultaneously measure the activation state of the genome-wide set of molecular activities that are responsible for cell fate switching. The arrival of technologies for the massively parallel monitoring of genes now opens this possibility to follow trajectories of cell states in a high-dimensional state space of the regulatory network. Gene expression profiling using DNA microarrays allows the parallel measurement of the level of >10,000 mRNAs in cells and tissues; this represents a surrogate measure for genome-wide gene activation profiles, and hence for cell states.

One way to uncover the existence of a high-dimensional attractor in real cells, where unlike in computer simulations we cannot systematically sample the states in state space, is to approach it from different directions of the state space and demonstrate the convergence of the high-dimensional trajectories (Figure 4). For instance, expression profiling to probe such trajectories in state space could be used to monitor cells that are induced by two different (biochemically distinct) stimuli to undergo the very same cell fate switch, e.g., from a proliferative state to a differentiated state. As noted above, such scenarios are common, and as such already suggest that the differentiated state is a stable attractor. If the two trajectories in gene expression space first diverge but then converge with

respect to a large portion, if not all, of the "gene dimensions" of the state space, this would be strongly suggestive of an entry into an attractor state. Initial studies for in vitro neutrophil differentiation suggest that this is in fact the case (Figure 4).

Such network perturbation experiments would, when performed systematically and in many cell types, allow us to obtain a first glance at the structure of the "attractor landscape" of the genome without knowledge of all the details of the "wiring diagram" of the genomic regulatory network. Unfortunately, the manipulation of the activation state of individual genes in living cells is still cumbersome compared to the situation in computer-simulated networks, such that systematic network perturbations that may reveal more detailed information about the structure of the attractor are still limited. Nevertheless, such experiments are a first small step toward the molecular characterization of Waddington's "epigenetic landscape" (Figure 5) and an essential intermediate step toward our understanding of how the genome maps into the phenome.

4.2.5. Hierarchical Considerations: Signaling Networks Beyond the Cell

Similar to the structural networks discussed above, information networks will extend beyond the limits of intracellular regulation. Cells in various states (attractors) signal to each other via physical cell–cell contacts, soluble cytokines, and insoluble matrix scaffolds, thus forming an extracellular communication network. The dynamics of such "cellular networks" can also be viewed in a framework of state space concepts, with stable behavioral modes that involve many cell types and their secreted products representing a coherent, robust physiological program of the tissue, such as inflammation, immune response, regeneration, development, and toxicity. These distinct "tissue fates" also exhibit properties of state space trajectories and attractors. For instance, immune system decisions between mutually exclusive, robust responses are common place, as in the Th1/Th2 dichotomy in the T-cell immune response (54; see also Part III, chapter 4.1, by Segel, this volume). Moreover, it now appears that cancer is not just a cell-autonomous disease in which mutant cells evolve to proliferate in an unconstrained fashion (see this volume, Part III, chapters 6.1, by Pienta, and 6.2, by Solé, Gonzales Garcia, and Costa), but also involves a dysregulation at the tissue level in which non-mutant cells of the tumor bed (stroma) and its blood vessels play a central role—thus, the cancerous tumor itself may be an unfortunately stable "tissue fate" (see also this volume, Part III, chapter 6.3, by Mansury and Deisboeck).

Biomedical research is only at the beginning of appreciating these higher-level interactions as formal networks, because most of leading edge "systems biology" research is still carried out on single-cell model organisms focusing on individual molecular pathways. But experiments in the near future that system-

atically elucidate these cell–cell interaction networks will enable the next step in understanding biological regulation: marching up to a higher level of organization in the vertical hierarchy of integration that characterizes complex living organisms.

5. CONCLUSION

Tensegrity is a principle that ensures structural stability within networks comprised of multiple structural components, and hence governs their self-assembly. Tensegrity is used at all size scales in the hierarchy of life, and it may have played an important role in the mechanism by which hierarchical self-assembly of inorganic components and small organic molecules led to formation of living cells (38). Use of the tensegrity principle by cells also provides an energy-efficient way to build macroscopic hierarchical structures using tiers of interconnected molecular networks (9).

On the other hand, the emergence of attractor landscapes within sparsely connected information-processing biochemical networks provides a mechanism for establishment of a limited number of stable network states that may have enabled evolution to harness a wide variety of environmental signals, including mechanical perturbation, for the regulation of cell fates. Thus, from the perspective of organismal biology, linking tensegrity-based structural networks and physical constraints to cell fate regulation is a central requirement for the evolution of organisms of increasing size that cannot rely solely on chemical interactions with their environment for control of their behavior. Living cells and tissues must deal with macroscopic physical phenomena such as mechanical forces, including tension, compression, shear, surface tension, and osmotic stresses. These physical signals can regulate specific modes of cell behavior controlled by molecular networks because of the link between structural networks and biochemical reactions (mechanochemistry on the cytoskeleton), and because of the existence of information-processing networks that produce an attractor landscape with stable states.

In both complex cellular structural networks and information networks, simple properties emerge through the collective action of the parts that includes mechanical and biochemical interactions. Hence the study of network properties helps to bridge the gap between microscopic biochemistry and macroscopic structure and behavior. Thus, elucidation of how simple, rule-governed behaviors (e.g., mechanical properties of cells and their behavioral control) emerge at higher levels of organization may eventually lead to a fuller understanding of the inner working of the living organism across many size scales. A key challenge for a conceptual understanding of the fundamental principles of complex living systems will be to learn when the myriad details can be abstracted away, and when they matter. Although our work represents only a first step toward an un-

derstanding of universal principles rather than specific details, hopefully, it opens up a new avenue of investigation in cell and molecular biology.

6. ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health, the Air Force Office of Scientific Research, and the National Aeronautics and Space Administration.

7. REFERENCES

1. Aldana M, Cluzel P. 2003. A natural class of robust networks. *Proc Natl Acad Sci USA* **100**:8710–8714.
2. Bar-Yam Y. 1997. *Dynamics in complex systems*. Studies in Nonlinearity. Perseus Publishing, Reading, MA.
3. Bogomolova EV. 2001. Isopropyl alcohol induced mycelium formation in lithobiontic black yeasts. *Mikologiya i Fitopatologiya* **35**(4):24–28.
4. Burridge K, Chrzanowska-Wodnicka M. 1996. Focal adhesions, contractility, and signaling. *Annu Rev Cell Dev Biol* **12**:463–518.
5. Cañadas P, Laurent VM, Oddou C, Isabey D, Wendling S. 2002. A cellular tensegrity model to analyse the structural viscoelasticity of the cytoskeleton. *J Theor Biol* **218**:155–173.
6. Carpenter CL. 2000. Actin cytoskeleton and cell signaling. *Crit Care Med* **28**(4 suppl):N94–N99.
7. Caspar DLD. 1980. Movement and self-control in protein assemblies. *Biophys J* **32**:103–138.
8. Chen CS, Alonso JL, Ostuni E, Whitesides GM, Ingber DE. 2003. Cell shape provides global control of focal adhesion assembly. *Biochem Biophys Res Commun* **307**:355–361.
9. Chen CS, Ingber DE. 1999. Tensegrity and mechanoregulation: from skeleton to cytoskeleton. *Osteoarthritis Cartilage* **7**:81–94.
10. Chen CS, Mrksich M, Huang S, Whitesides GM, Ingber DE. 1997. Geometric control of cell life and death. *Science* **276**:1425–1428.
11. Chrzanowska-Wodnicka M, Burridge K. 1996. Rho-stimulated contractility drives the formation of stress fibers and focal adhesions. *J Cell Biol* **133**:1403–1415.
12. Coffey DS. 1998. Self-organization, complexity and chaos: the new biology for medicine. *Nature Med* **4**:882–885.
13. Collins SJ. 1987. The HL-60 promyelocytic leukemia cell line: proliferation, differentiation, and cellular oncogene expression. *Blood* **70**:1233–1244.
14. Das D, Pintucci G, Stern A. 2000. MAPK-dependent expression of p21(WAF) and p27(kip1) in PMA-induced differentiation of HL60 cells. *FEBS Lett* **472**:50–52.
15. Delbrück M. 1949. Colloques Internationaux du CNRS. In *Unités biologiques douées de continuité génétique*. CNRS, Paris.
16. Dong C, Skalak R, Sung KL. 1991. Cytoplasmic rheology of passive neutrophils. *Biorheol* **28**:557–567.
17. Evans E, Yeung A. 1989. Apparent viscosity and cortical tension of blood granulocytes determined by micropipet aspiration. *Biophys J* 56:151–160.
18. Fabry B, Maksym GN, Butler JP, Glogauer M, Navajas D, Fredberg JF. 2001. Scaling the microrheology of living cells. *Phys Rev Lett* **87**:148102-1–148102-4.

19. Farrell HM Jr, Qi PX, Brown EM, Cooke PH, Tunick MH, Wickham ED, Unruh JJ. 2002. Molten globule structures in milk proteins: implications for potential new structure–function relationships. *J Dairy Sci* 85:459–471.
20. Fox JJ, Hill CC. 2001. From topology to dynamics in biochemical networks. *Chaos* **11**:809–815.
21. Fuller B. 1961. Tensegrity. *Portfolio Artnews Annu* **4**:112–127.
22. Fung YC, Liu SQ. 1993. Elementary mechanics of the endothelium of blood vessels. *ASME J Biomech Eng* **115**:1–12.
23. Geiger B, Bershadsky A. 2001. Assembly and mechanosensory function of focal contacts. *Curr Opin Cell Biol* **13**:584–592.
24. Geiger B, Bershadsky A. 2002. Exploring the neighborhood: adhesion-coupled cell mechanosensors. *Cell* **110**:139–142.
25. Gell-Mann M. 1994. *The quark and the jaguar: adventures in the simple and the complex*. Freeman, San Francisco.
26. Glass L, Hill C. 1998. Ordered and disordered dynamics in random networks. *Europhys Lett* **41**:599–604.
27. Goss RJ. 1967. The strategy of growth. In *Control of cellular growth in the adult organism*, pp. 3–27. Ed. H Teir, T Rytömaa T. Academic Press, London.
28. Howe AK, Aplin AE, Juliano RL. 2002. Anchorage-dependent ERK signaling—mechanisms and consequences. *Curr Opin Genet Dev* **12**:30–35.
29. Huang S. 2000. The practical problems of post-genomic biology. *Nature Biotechnol* **18**:471–472.
30. Huang S. 2002. Regulation of cellular states in mammalian cells from a genome-wide view. In *Gene regulation and metabolism: post-genomic computational approach*, pp. 181–220. Ed. Collado-Vides J, Hofestädt R. MIT Press, Cambridge.
31. Huang S, Ingber DE. 2000. Shape-dependent control of cell growth, differentiation and apoptosis: switching between attractors in cell regulatory networks. *Exp Cell Res* **261**:91–103.
32. Huang S, Chen SC, GM, Ingber DE. 1998. Cell-shape-dependent control of p27Kip and cell cycle progression in human capillary endothelial cells. *Mol Biol Cell* **9**:3179–3193.
33. Ingber DE, Jamieson JD. 1985. Cells as tensegrity strutures: architectural regulation of histodifferentiation by physical forces transduced over basement membrane. In *Gene expression during normal and malignant differentiation*, pp. 13–32. Ed. LC Anderson, CG Gahmberg, P Ekblom. Academic Press, Orlando, FL.
34. Ingber DE. 1991. Integrins as mechanochemical transducers. *Curr Opin Cell Biol* **3**:841–848.
35. Ingber DE. 1993. Cellular tensegrity: defining new rules of biological design that govern the cytoskeleton. *J Cell Sci* **104**:613–627.
36. Ingber DE. 1993. The riddle of morphogenesis: a question of solution chemistry or molecular cell engineering? *Cell* **75**:1249–1252.
37. Ingber DE. 1998. The architecture of life. *Sci Am* **278**:48–57.
38. Ingber DE. 2000. The origin of cellular life. *Bioessays* **22**(12):1160–1170.
39. Ingber DE. 2003. Tensegrity. I. Cell structure and hierarchical systems biology. *J Cell Sci* **116**(Pt 7):1157–1173.
40. Ingber DE. 2003. Tensegrity. II. How structural networks influence cellular information-processing networks. *J Cell Sci* **116**(Pt 8):1397–1408.
41. Jeong H, Mason SP, Barabasi AL, Oltvai ZN. 2001. Lethality and centrality in protein networks. *Nature* **411**:41–42.
42. Kauffman SA. 1969. Metabolic stability and epigenesis in randomly constructed genetic nets. *J Theor Biol* **22**:437–467.
43. Kauffman SA. 1993. *The origins of order*. Oxford UP, New York.
44. Kulyk WM, Hoffman LM. 1996. Ethanol exposure stimulates cartilage differentiation by embryonic limb mesenchyme cells. *Exp Cell Res* **223**(2):290–300.

45. Lewontin RC. 2000. *The triple helix : gene, organism, and environment*. Harvard UP, Cambridge.
46. Maslov S, Sneppen K. 2002. Specificity and stability in topology of protein networks. **Science 296**:910–913.
47. Messing RO. 1993. Ethanol as an enhancer of neural differentiation. *Alcohol Alcoholism* **2**(suppl.):289–293.
48. Monod J, Jacob F. 1961. General conclusions: teleonomic mechanisms in cellular metabolism, growth and differentiation. *Cold Spring Harbor Symp Quant Biol* **26**:389–401.
49. Morange M. 2001. *The misunderstood gene*. Harvard UP, Cambridge.
50. Ovadi J, Srere PA. 2000. Macromolecular compartmentation and channeling. *Intl Rev Cytol* **92**:255–280.
51. Parker KK, Brock AL, Brangwynne C, Mannix RJ, Wang N, Ostuni E, Geisse NA, Adams JC, Whitesides GM, Ingber DE. 2002. Directional control of lamellipodia extension by constraining cell shape and orienting cell tractional forces. *FASEB J* **16**(10):1195–1204.
52. Parker SB, Eichele G, Zhang P, Rawls A, Sands AT, Bradley A, Olson EN, Harper JW, Elledge SJ. 1995. p53-independent expression of p21Cip1 in muscle and other terminally differentiating cells. *Science* **267**:1024–1027.
53. Rose SPR. 1998. *Lifelines: biology beyond determinism*. Oxford UP, Oxford.
54. Santana MA, Rosenstein Y. 2003. What it takes to become an effector T cell: the process, the cells involved, and the mechanisms. *J Cell Physiol* **195**(3):392–401.
55. Schultz SG. 1996. Homeostasis, Humpty Dumpty, and integrative biology. *News Physiol Sci* **11**:238–246.
56. Sherr CJ. 1994. G1 phase progression: cycling on. *Cell* **79**:561–555.
57. Spremulli EN, Dexter DL. 1984. Polar solvents: a novel class of antineoplastic agents. *J Clin Oncol* **2**(3):227–241.
58. Stamenovic D, Fredberg JJ, Wang N, Butler JP, Ingber DE. 1996. A microstructural approach to cytoskeletal mechanics based on tensegrity. *J Theor Biol* **181**:125–136.
59. Stamenovic D, Liang Z, Chen J, Wang N. 2002. Effect of the cytoskeletal prestress on the mechanical impedance of cultured airway smooth muscle cells. *J Appl Physiol* **92**:1443–1450.
60. Stamenovic D, Ingber DE. 2002. Models of cytoskeletal mechanics and adherent cells. *Biomech Modeling Mechanobiol* **1**:95–108.
61. Steinman RA, Hoffman B, Iro A, Guillouf C, Liebermann DA, el-Houseini ME. 1994. Induction of p21 (WAF-1/CIP1) during differentiation. *Oncogene* **9**:3389–3396.
62. Strohman RC. 1997. The coming Kuhnian revolution in biology. *Nature Biotechnol* **15**:194–200.
63. Sultan C, Corless M, Skelton RE. 2001. The prestressability problem of tensegrity structures: some analytical solutions. *Intl J Solids Struct* **38**:5223–5252.
64. Sultan C, Stamenovic D, Ingber DE. 2004. A computational tensegrity model predicts dynamic rheological behaviors in living cells. *Annu Biomed Eng* **32**:520–530.
65. Volokh KY, Vilnay O, Belsky M. 2000. Tensegrity architecture explains linear stiffening and predicts softening of living cells. *J Biomech* **33**:1543–1549.
66. Volokh KY, Vilnay O, Belsky M. 2002. Cell cytoskeleton and tensegrity. *Biorheol* **39**:63–67.
67. Waddington CH. 1956. *Principles of embryology*. Allen & Unwin Ltd, London.
68. Waddington CH. 1940. *Organisers and genes*. Cambridge UP, Cambridge.
69. Wagenknecht T, Grassucci R, Radke GA, and Roche TE. 1991. Cryoelectron microscopy of mammalian pyruvate dehydrogenase complex. *J Biol Chem* **266**:24650–24656.
70. Wagner A. 2002. Estimating coarse gene network structure from large-scale gene perturbation data. *Genome Res* **12**(2):309–315.
71. Weiss JN, Qu Z, Garfinkel A. 2003. Understanding biological complexity: lessons from the past. *FASEB J* **17**:1–6.

72. Wendling S, Cañadas P, Oddou C, Meunier A. 2002. Interrelations between elastic energy and strain in a tensegrity model: contribution to the analysis of the mechanical response in living cells. *Comput Meth Biomech Biomed Engin* **5**(1):1–6.
73. Wendling S, Oddou C, Isabey D. 1999. Stiffening response of a cellular tensegrity model. *J Theor Biol* **96**(3):309–325.
74. Yu ZW, Quinn PJ. 1994. Dimethyl sulphoxide: a review of its applications in cell biology. *Biosci Rep* **14**:259–281.
75. Zanotti G, Guerra C. 2003. Is tensegrity a unifying concept of protein folds? *FEBS Lett* **534**:7–10.
76. Huang S, Eichler G, Bar-Yam Y, Ingber DE. 2005. Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys Rev Lett* **94**:128701.

2.2

SPATIOTEMPORAL DYNAMICS OF EUKARYOTIC GRADIENT SENSING

K.K. Subramanian and Atul Narang

Department of Chemical Engineering,
University of Florida, Gainesville

The crawling movement of eukaryotic cells in response to a chemical gradient is a complex process involving the orchestration of several subcellular activities. Although a complete description of the mechanisms underlying cell movement remains elusive, the very first step of *gradient sensing*, enabling the cell to perceive the imposed gradient, is becoming more transparent. The increased understanding of this step has been driven by the discovery that within 5–10 seconds of applying a weak chemoattractant gradient, membrane phosphoinositides, such as PIP_3, localize at the front end of the cell, where they activate a process of intense actin polymerization and trigger the extension of a protrusion. This train of events implies that the key to gradient sensing is a mechanistic understanding of the phosphoinositide localization. Since the phosphoinositide distribution is highly localized compared to the shallow chemoattractant gradient, it has been suggested that the cell merely amplifies the chemoattractant gradient. However, this cannot be true since the phosphoinositide localization can display a bewildering array of spatial distributions that bear no resemblance to the external chemoattractant profile. For instance, a single phosphoinositide localization is produced in the face of multiple chemoattractant sources. More surprisingly, the localization forms at a random location even if the chemoattractant concentration is uniform. Here we show that all these seemingly complex dynamics are consistent with the so-called activator-inhibitor class of models. To this end, we formulate and simulate an activator-inhibitor model of gradient sensing based on the phosphoinositide signaling pathways. Specifically, membrane-resident phosphoinositides play the role of activator, and cytosolic inositol phosphates act as inhibitor. The remarkable agreement between the simulated and observed dynamics of phosphoinositide localization supports our conjecture that gradient sensing is a manifestation of an activator-inhibitor mechanism. However, the molecular identity of the activator and inhibitor remains unresolved. We discuss several competing hypotheses in the literature regarding the identity of these molecules.

Address correspondence to: Atul Narang, Department of Chemical Engineering, University of Florida, Room 237 CHE, Gainesville, FL 32608 (ksubrama@che.ufl.edu).

1. INTRODUCTION

Cell migration plays a crucial role in our birth, survival, and death. We are conceived as amorphous fertilized eggs. It is cell migration, among other processes, that sculpts a richly structured embryo from a fertilized egg. While we live, cell migration heals our wounds (1) and protects us from surrounding pathogens (2). But when we age, cell migration can accelerate death. In some instances, cancer metastasis is caused by directed migration of tumor cells from the primary to the preferred sites of metastasis (3). Evidently, a better understanding of the mechanism of cell migration will have profound biomedical consequences.

Most eukaryotic cells move by crawling on a surface. The crawling movement occurs in response to an external stimulus, which is frequently a chemical concentration gradient. The resultant motion propels the cells forward along the direction of highest increase in concentration. The chemical that induces the movement is called a *chemoattractant* and the movement itself is called *chemotaxis*. Eukaryotic chemotaxis is cyclic, and each cycle consists of four phases (Figure 1): (1) extension of a protrusion, (2) adhesion of the protrusion to the surface, (3) contraction of the cell body, and (4) retraction of the tail. Each phase of the cycle is a complex process involving the coordinated action of a large constellation of molecules (4). In this work, we confine our attention to *gradient sensing*, the mechanism that enables the cell to read the external gradient and extend a protrusion precisely at the *leading edge*, the region exposed to the highest chemoattractant concentration.

Figure 1. The four phases of a chemotactic cycle.

The extension of the protrusion involves localized actin polymerization at the leading edge. Soon after the cells are exposed to a chemoattractant gradient, the leading edge develops fingerlike actin-based structures called *filopodia*. The space between the filopods then fills up with an actin mesh to form a wide, sheetlike *lamellipod*. The localized polymerization of actin at the leading edge implies that the gradient sensing machinery amplifies the external signal. Indeed, the chemoattractant gradients imposed in the extracellular space are often quite small (1–2% concentration change over the length of the cell) (5), but actin polymers synthesized in response to the gradient are found exclusively at the leading edge (6). The key problem in the study of gradient sensing is elucidation

of the mechanism that mediates the formation of a highly polarized distribution of actin polymers in response to a relatively mild chemoattractant gradient.

The chemoattractant gradient is transmitted to the actin polymerization machinery by a signal transduction pathway that starts with receptors on the cell surface and terminates in proteins that catalyze actin polymerization. It is therefore conceivable that actin polymers inherit their polarized distribution from some molecule that is upstream of the polymers in the pathway. Recent experiments have shown that certain membrane-resident phosphoinositides—namely, phosphatidylinositol 3,4,5-phosphate (PIP_3)[1] and phosphatidylinositol 4,5-phosphate (PIP_2)—are among the earliest polarized components of the pathway. Moreover, these phosphoinositides activate the enzymes that catalyze actin polymerization, thus generating the force that pushes the membrane forward. In other words, the morphological polarity corresponding to the extension of a protrusion is driven by the chemical polarity corresponding to the phosphoinositide localization. Thus, the key question in the study of gradient sensing becomes: What is the mechanism of phosphoinositide localization?

Our attempts to address this question exemplify the modeling methodology enunciated in this volume (see Part II, chapter 2, by Socolar). We observed that the seemingly complex patterns of phosphoinositide localization were strongly reminiscent of the spatiotemporal dynamics associated with the activator-inhibitor class of models. Encouraged by this analogy, we screened the phosphoinositide signaling pathways for activators and inhibitors, and selected the pair that seemed most consistent with the prevailing experimental literature. We then formulated and simulated an activator-inhibitor model in which membrane-resident phosphoinositides and cytosolic inositol phosphates played the roles of activator and inhibitor, respectively. As we show in §2, the simulations are in remarkable agreement with all the spatiotemporal dynamics observed in the literature. These findings strongly suggest that the dynamics of the gradient sensing mechanism are consistent with the activator-inhibitor class of models. However, there are several hypotheses in the literature regarding the molecular identity of the activator and the inhibitor. These open questions, which can only be resolved by experiments, are discussed in §3.

We begin by giving a brief description of the phosphoinositide signaling pathways and the spatiotemporal dynamics of the phosphoinositide localization.

1.1. Signaling Pathways

The signaling steps that follow receptor activation are the subjects of ongoing research. The model systems studied most intensively are *Dictyostelium discoideum* and neutrophils. In these systems, receptor-ligand binding activates heterotrimeric G-proteins, which stimulate PI3K, the enzyme that catalyzes the synthesis of PIP_3 (Figure 2a). In neutrophils, it has been shown that PIP_3

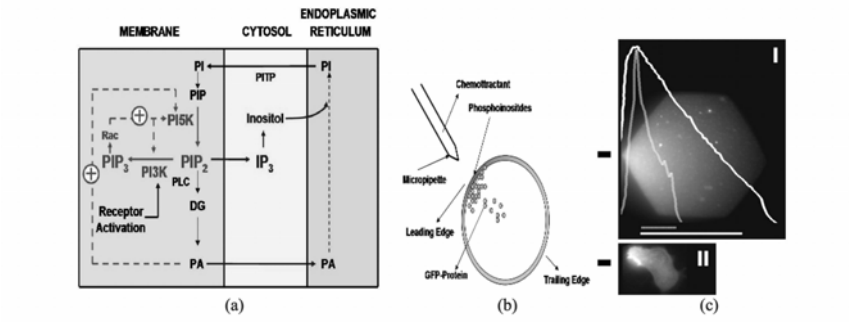


Figure 2. (**a**) The phosphoinositide cycle. The abbreviations used are: PI = phosphatidylinositol; PIP = phosphatidylinositol 4-phosphate; PIP_2 = phosphatidylinositol 4,5-phosphate; PIP_3 = phosphatidylinositol 3,4,5-phosphate; DG = diacylglycerol; IP_3 = inositol 1,4,5-triphosphate; PI3K = phosphatidylinositol 3-kinase; PI5K = phosphatidylinositol 4-phosphate 5-kinase; PITP = phosphatidylinositol transport protein. (**b**) The principle of fluorescent imaging experiments. (**c**) Polarization of PIP_3 in response to a steady chemoattractant gradient (14). The curve with a shallow gradient represents the chemoattractant distribution within a hexagonal field of view. The curve with a sharp gradient shows the resulting distribution of the marker for PIP_3 within the cell shown in the lower figure.

activates the Rac-GEF, P-Rex1, which in turn activates the small GTPase, Rac (7). There is strong evidence that Rac activates PI5K (8). It has also been proposed that Rac activates PI3K (9). Both of these steps create a positive feedback loop in which synthesis of phosphoinositides stimulates the synthesis of even more phosphoinositides. There is yet another positive feedback loop because activation of PI5K increases the rate of synthesis of PIP_2 and its downstream product, phosphatidic acid (PA), which is a potent activator of PI5K (10). Consequently, upon receptor activation, the synthesis rates of PIP_2 and PIP_3 can rapidly accelerate to high levels. Such high synthesis rates can be sustained for no more than a second because the concentration of phosphatidylinositol (PI) in the plasma membrane is quite small (11). Depletion of PI in the plasma membrane is prevented by the cytosolic PI transport protein (PITP), which transfers readily available PI from the endoplasmic reticulum to the plasma membrane (12). The PIP_2 formed by successive phosphorylation of PI is hydrolyzed by phospholipase C (PLC) to diacylglycerol (DG) and cytosolic inositol 1,4,5-triphosphate (IP_3). Diacylglycerol is converted to PA and transferred to the endoplasmic reticulum for regeneration of PI. The inositol produced by rapid dephosphorylation of IP_3 via multiple pathways (13), also participates in PI regeneration.

Recent experiments have established a causal link between localized phosphoinositide formation and lamellipod extension. Specifically, it has been shown that

1. PIP_2, in conjunction with GTP-bound Cdc42, is a strong activator of N-WASP, which in turn activates Arp2/3 (15).

2. Activated Arp2/3 mediates actin polymerization by nucleating the sides of preexisting actin filaments. This promotes the formation of the branched filament network found in lamellipods (16).

3. Actin polymerization by Arp2/3 can drive lamellipod protrusion (17).

4. Ruffles form at the very same time and locations as PIP_2 localization (18).

Taken together, these facts suggest that the localization of phosphoinositides at the leading edge plays a crucial role in lamellipod extension.

1.2. Dynamics

The principle of the experiments used to study the spatiotemporal dynamics in response to various chemoattractant gradients is illustrated in Figure 2b. Motile cells are transfected with chimeric proteins made by fusing a fluorescent protein either to the molecule of interest, or to a "marker" molecule that binds specifically to the molecule of interest and thus "reports" on it. The marker molecules commonly used for reporting on PIP_2 and PIP_3 are the pleckstrin homology (PH) domains derived from various proteins (19). The transfected cells are then exposed to various chemoattractant concentration profiles and the movement of the fluorescent chimeric proteins is visualized using confocal microscopy. The chemoattractant profiles imposed include *steady or time-varying gradients*, obtained by appropriate manipulation of the chemoattractant flow rate through a micropipette, and *steady uniform profiles*, obtained by immersing the cell in chemoattractant. Each chemoattractant profile reveals particular aspects of the dynamics associated with gradient sensing.

Steady and time-varying chemoattractant gradients show the existence of *amplification* and help identify the first amplified component. In response to such chemoattractant profiles, it has been observed that

- When G-proteins are absent (20) or inactive (14), there is no polarization. Evidently, the chemoattractant profile is transmitted to the cell through the receptors and G-proteins.
- The receptors remain uniformly distributed in both neutrophils (21) and *Dictyostelium* (22). Furthermore, receptor occupancy (23) and G-protein activity (24,25) are not significantly polarized. It follows that receptors and G-proteins are required for transmitting the extracellular signal, but they are not the source of the amplification.

- Membrane-resident phosphoinositides, PIP_3 (14,26–28) and PIP_2 (29,18), are strongly polarized. In neutrophils, the gradient of the marker for PIP_3 is six times the chemoattractant gradient (14) (Figure 2c).

At first sight, the strong polarization of phosphoinositides seems explicable in terms of a simple amplification model. It suffices to postulate that phosphoinositide synthesis responds to receptor activation in a highly cooperative manner (Hill-type kinetics). In this case, the phosphoinositide distribution will be similar in shape, but steeper in slope, when compared to the chemoattractant concentration profile. The following observation precludes this simple model. If a cell that is already polarized in a certain direction is exposed to a *modest* chemoattractant gradient along a different direction, a new phosphoinositide localization and pseudopod does not develop at the point with the highest chemoattractant concentration. Instead, the existing phosphoinositide localization and pseudopod turn and reorient themselves along the new gradient (14,25,30). It is as if the preexisting phosphoinositide localization and leading edge is more sensitive to chemotactic signals than all other regions of the cell. This phenomenon, called *polarized sensitivity*, suggests that during the course of phosphoinositide localization at the leading edge a diffusible substance is formed that rapidly migrates away from the leading edge and somehow inhibits the generation of a new localization in other regions of the cell. Interestingly, if the new chemoattractant gradient is relatively large and localized, the existing localization dissolves and a new one grows at the maximum of the new gradient (31,32,27).

Steady uniform chemoattractant concentrations reveal two properties of motile cells, namely, *spontaneous polarization* and *adaptation*.

When cells are exposed to such chemoattractant concentration profiles, phosphoinositides accumulate uniformly along the entire plasma membrane within 5–10 seconds. However, this uniform pattern does not persist for long. Within a few minutes, the phosphoinositide distribution polarizes at a random location. This phenomenon is called *spontaneous* polarization (33,34) to emphasize the fact that the cells polarize even though the chemoattractant concentration is macroscopically uniform. The random location of the polarization suggests that this phenomenon is stochastic in nature.

In *Dictyostelium*, the phosphoinositides return to their *pre-stimulus* uniform distribution within 100–200 seconds (27), but the cells polarize eventually. It is significant that phosphoinositides return to the pre-stimulus level even though the chemoattractant concentration is several orders of magnitude higher than the pre-stimulus level. This remarkable phenomenon is a manifestation of *adaptation* (35). It follows that in *Dictyostelium*, the time scale of activation is 5–10 seconds, whereas the time scale of adaptation is 100–200 seconds. In neutrophils, the cells polarize spontaneously before there is any perceptible reduction

in phosphoinositide levels at the plasma membrane (33). This suggests that in neutrophils the time scale of adaptation is so slow that the cells polarize spontaneously before there is significant adaptation.

In short, the phosphoinositide distributions display four distinct types of spatiotemporal dynamics in response to various chemoattractant gradients: 1) amplification and threshold, 2) polarized sensitivity, 3) spontaneous polarization, and 4) adaptation. Several mathematical models have been developed to capture one or more of these dynamics. Iglesias and coworkers have formulated a series of models that focus on adaptation and amplification (36). Although these models display perfect adaptation, they capture neither polarized sensitivity nor spontaneous polarization, and amplification is quite weak compared to experiments. The model in (37) is concerned primarily with amplification. The remaining models are formally similar insofar as they contain a *short-range activator* that is synthesized autocatalytically, and a *long-range inhibitor* that inhibits the synthesis of the activator (38–41). These models differ only with respect to the postulated mechanisms of activation and inhibition. In the next section, we present our model and show that it captures the dynamics of amplification, threshold, polarized sensitivity and spontaneous polarization. Although the other two activator-inhibitor models developed by Meinhardt and Postma et al. have not been studied in detail, the fact that they are formally similar to our model suggests that they possess similar dynamic properties. The discussion of adaptation is deferred to §3.

2. MODEL AND SIMULATION

2.1. Model

The model is an abstraction of the phosphoinositide cycle (Figure 2a). It contains three variables corresponding to three "lumped" pools (Figure 3a)—namely, membrane phosphoinositides (P), cytosolic and its phosphates (I), and phosphoinositides in the endoplasmic reticulum (P_s). The concentrations of these variables are denoted by p, i, and p_s, respectively. It is assumed that

1. The cell is two-dimensional and disk-shaped. Thus, p and p_s are based on the length of the plasma membrane and i is based on the area of the cytosol.

2. Radial gradients of the cytosolic inositol phosphate pool are negligibly small. This is reasonable because inositol phosphates diffuse rapidly.[2] It follows from assumptions 1 and 2 that the angle, θ is the only spatial variable.

3. In the absence of receptor activation, there is *basal* synthesis and degradation of P and I. The basal synthesis rates of P and I, denoted c_p and c_i, follow zero-order kinetics. The basal degradation rates, denoted $r_{p,d}$ and $r_{i,d}$, obey first-order kinetics with rate constants k_p and k_i, respectively, i.e.,

$$r_{p,d} \equiv k_p p, \quad r_{i,d} \equiv k_i i$$

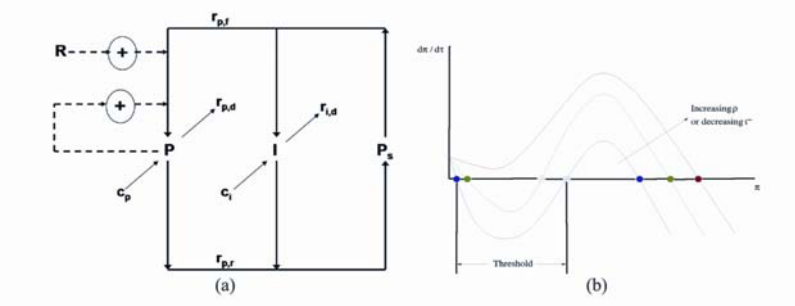


Figure 3. (**a**) The model scheme is an abstraction of the PI cycle shown in Figure 2. (**b**) Initial dynamics of the membrane phosphoinositides at any point of the cell membrane. The initial concentrations of active receptors, ρ, and inositol phosphates, ι^-, are treated as control parameters. When ρ is small or ι^- is large, there is a threshold (defined as the distance between the lower and intermediate steady states). As ρ increases or ι^- decreases, the threshold becomes progressively smaller. At sufficiently large values of ρ or sufficiently small values of ι^-, the threshold disappears completely.

4. The receptors instantly inherit the chemoattractant profile imposed on the cells, and the receptor-mediated rate of formation of membrane phosphoinositides per unit length of membrane is given by

$$r_{p,f} \equiv k_f r(t,\theta) p^2 p_s.$$

Here, k_f denotes the rate constant and $r(t,\theta)$ denotes the concentration of active receptors. The dependence on p^2 represents the autocatalytic and cooperative kinetics with respect to membrane phosphoinositides. This is an idealization of the positive feedback loops shown in Figure 3a. Thus, the membrane phosphoinositides will play the role of *local activator* in the model.

5. The inositol phosphate pool (I) stimulates transfer of phosphoinositides from the plasma membrane to the endoplasmic reticulum. Thus, the rate of removal of membrane phosphoinositides per unit length of the membrane is

$$r_{p,r} \equiv k_r pi,$$

where k_r denotes the rate constant. The rationale for this assumption is as follows. Inositol participates in the regeneration of PI in the endoplasmic reticulum (Figure 3a). An increase in inositol concentration will, therefore, drive the transport of PA from the plasma membrane to the endoplasmic reticulum and its subsequent conversion to PI. Thus, the inositol phosphates act as a *global inhibitor*.

To simulate the experiments, it is assumed that before the cell is subjected to a chemoattractant perturbation ($t < 0$), it is at a homogeneous steady

state, $(\bar{p},\bar{p_s},\bar{i})$, corresponding to a uniform chemoattractant concentration; hence, $r = \bar{r}$ = constant. At time $t = 0$, the cell is perturbed by exposing it to a chemoattractant profile that is instantly mirrored by the active receptor profile, $r(t,\theta)$. The dynamics of P, P_s, and I are then governed by the equations

$$\frac{\partial p}{\partial t} = k_f r p^2 p_s - k_r p i + c_p - k_p p + \frac{D_p \partial^2 p}{R^2 \partial \theta^2}, \quad [1]$$

$$\frac{\partial p_s}{\partial t} = -\left(k_f r p^2 p_s - k_r p i + c_p - k_p p\right) + \frac{D_{p_s}}{R^2}\frac{\partial^2 p_s}{\partial \theta^2}, \quad [2]$$

$$\frac{\partial i}{\partial t} = s\left(k_f r p^2 p_s - k_r p i\right) + c_i - k_i i + \frac{D_i}{R^2}\frac{\partial^2 i}{\partial \theta^2}. \quad [3]$$

Here, D_p, D_{p_s}, D_i denote the lateral diffusivities of P, P_s, and I, respectively, and R denotes the cell radius. The factor s, denoting membrane length per unit cell area, is required since synthesis and removal rates of P are based on the length of the plasma membrane. Since the cell is circular, all concentrations and fluxes must be equal at $\theta = 0$ and $\theta = 2\pi$. Thus, we impose the *periodic* boundary conditions

$$x(0,t) = x(2\pi,t), \frac{\partial x(0,t)}{\partial \theta} = \frac{\partial x(2\pi,t)}{\partial \theta}, t > 0, \quad [4]$$

where $x \equiv p, p_s, i$. The initial conditions for the reduced equations are

$$p(0,\theta) = \bar{p},\ p_s(0,\theta) = p_t - \bar{p},\quad i(0,\theta) = \bar{i},\quad 0 < \theta < 2\pi. \quad [5]$$

Here, the initial condition for p_s reflects the assumption that the total amount of phosphoinositide in the membrane and the endoplasmic reticulum is conserved, so that the average phosphoinositide concentration, denoted p_t, is constant (39).

It is convenient to define the dimensionless variables

$$\pi \equiv \frac{p}{p_t}, \pi_s \equiv \frac{p_s}{p_t}, \iota \equiv \frac{i}{s p_t}, \xi \equiv \frac{\theta}{2 \times 3.1416}, \tau \equiv \frac{t}{1/\left(k_r s p_t\right)},$$

and dimensionless parameters

$$\rho \equiv \frac{r}{r_t} \quad \kappa_f \equiv \frac{k_f r_t p_t^2}{k_r s p_t} \quad \psi_p \equiv \frac{c_p / p_t}{k_r s p_t} \quad \kappa_p \equiv \frac{k_p}{k_r s p_t} \quad \delta_{p_s} \equiv \frac{D_{p_s} / C^2}{k_r s p_t},$$

$$\psi_i \equiv \frac{c_i/(sp_t)}{k_r sp_t} \quad \kappa_i \equiv \frac{k_i}{k_r sp_t} \quad \delta_i \equiv \frac{D_i/C^2}{k_r sp_t} \quad \rho^- \equiv \frac{r^-}{r_t} \quad \pi \equiv \frac{p}{p_t} \quad \pi_s^- \equiv \frac{p_s^-}{p_t} \quad \iota^- \equiv \frac{i^-}{sp_t},$$

where C denotes the circumference of the cell and r_t denotes the total number of receptors. Thus, we arrive at the dimensionless equations

$$\frac{\partial \pi}{\partial \tau} = \kappa_f \rho \pi^2 \pi_s - \pi \iota + \psi_p - \kappa_p \pi + \delta_p \frac{\partial^2 \pi}{\partial \xi^2}, \qquad [6]$$

$$\frac{\partial \pi_s}{\partial \tau} = -\left(\kappa_f \rho \pi^2 \pi_s - \pi \iota + \psi_p - \kappa_p \pi\right) + \delta_{p_s} \frac{\partial^2 \pi_s}{\partial \xi^2}, \qquad [7]$$

$$\frac{\partial \iota}{\partial \tau} = \kappa_f \rho \pi^2 \pi_s - \pi \iota + \psi_i - \kappa_i \iota + \delta_i \frac{\partial^2 \iota}{\partial \xi^2}, \qquad [8]$$

with initial conditions

$$\pi(0,\xi) = \pi^-, \pi_s(0,\xi) = 1 - \pi^-, \iota(0,\xi) = \iota^-, 0 \le \xi < 1 \qquad [9]$$

and periodic boundary conditions

$$x(0,\tau) = x(1,\tau), \frac{\partial x(0,\tau)}{\partial \xi} = \frac{\partial x(2\pi,\tau)}{\partial \xi}, \tau > 0. \qquad [10]$$

where $x \equiv \pi, \pi_s, \iota$.

2.2. Simulations

Equations [6]–[10], along with various choices of $\rho(\tau,\xi)$ described below, were simulated using the NAG subroutine D03PHF (43). In this subroutine, the spatial coordinate is discretized by finite differences, the PDEs are reduced to ODEs by the method of lines, and the resulting system of ODEs is integrated by using a backward differentiation formula. The parameter values used in the simulations are shown in Table 1. The rationale for the choice of parameter values can be found in (39). To facilitate comparison of the simulations with experimentally observed dynamics, it is useful to note that $k_r sp_t \sim 1$ 1/sec (39). Hence, each unit of dimensionless time, τ, corresponds roughly to 1 sec.

Table 1. Parameter values used in the simulations (from (39) and (42))

$\psi_i = 0.01$	dimensionless rate of basal synthesis of inositol phosphates
$\psi_p = 0.01$	dimensionless rate of basal synthesis of phosphoinositides
$\delta_i = 1$	dimensionless angular diffusivity of inositol phosphates
$\delta_p = 0.001$	dimensionless angular diffusivity of phosphoinositides in plasma membrane
$\delta_{p_s} = 0.001$	dimensionless angular diffusivity of phosphoinositides in endoplasmic reticulum
$\kappa_f = 2.4$	dimensionless rate constant for receptor mediated phosphoinositide formation
$\kappa_p = 0.1$	dimensionless rate constant for basal degradation of phosphoinositides
$\kappa_i = 0.1$	dimensionless rate constant for basal degradation of inositol phosphates
$\kappa^- = 6.7 \times 10^{-3}$	dimensionless rate constant for receptor-ligand dissociation
$r_t = 50{,}000$	total number of receptors

2.2.1. Amplification and Threshold

In the following simulation, the cell is assumed to be at a uniform steady state ($\bar{\pi}$, $\bar{\pi}_s$, $\bar{\iota}$) corresponding to the uniform active receptor profile $\rho = \bar{\rho}$. At time $t \geq 0$, the cell is exposed to a steady chemoattractant gradient and the active receptor distribution instantly inherits the chemoattractant concentration profile (Figure 4a). Despite the mild gradient of ρ, a pronounced phosphoinositide peak ultimately develops at the leading edge, $\xi = 1/2$ (Figure 4b). Compared to the polarized distribution of membrane phosphoinositides, the concentration profile of inositol phosphates is virtually flat (Figure 4c).

In terms of the model, formation of the phosphoinositide peak can be explained as follows. Because of their autocatalytic and cooperative kinetics, membrane phosphoinositides (P) are strongly amplified beyond a certain threshold. To see this, observe that immediately after receptor activation, $\pi_s \approx \bar{\pi}_s = 1 - \bar{\pi}$, $\iota \approx \bar{\iota}$, and diffusion is negligibly small compared to the reaction. Hence, the *initial* dynamics of the membrane phosphoinositides at any point of the plasma membrane is approximated by the equation

$$\frac{\partial \pi}{\partial \tau} = \kappa_f \rho \pi^2 (1 - \pi) - \pi \bar{\iota} + \psi_p - \kappa_p \pi . \qquad [11]$$

Figure 3b shows that if ρ is small at a point, the membrane phosphoinositides display *bistable* dynamics at that point, i.e., there are two stable steady states separated by an unstable steady state, which acts as a threshold because π moves

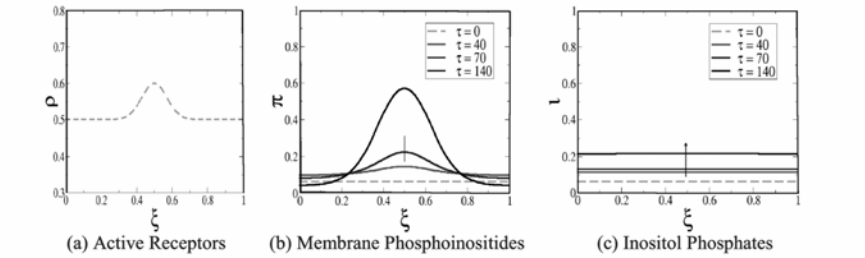


Figure 4. Amplification of a small chemoattractant gradient: (**a**) distribution of active receptors at $t \geq 0$, (**b**) development of a pronounced membrane phosphoinositide peak at the leading edge of the cell at $\xi = 1/2$, (**c**) growth of the uniformly distributed inositol phosphate pool.

to the upper steady state only if it crosses the unstable steady state. If ρ is large at a point, the threshold disappears at that point, and every π, no matter how small, tends to move toward the upper steady state. Now, at time $t < 0$, $\rho = \bar{\rho}$ everywhere and there is a threshold at every point. Upon activation of receptors at $t = 0$, ρ becomes significantly large in a neighborhood of the leading edge. In this region, the threshold vanishes and the concentration of P starts growing. Although P diffuses slowly, it does have a tendency to spread throughout the cell membrane. This tendency is contained by the formation of cytosolic inositol phosphates (I), which rapidly diffuse away from the leading edge, and acquire a relatively flat profile. The rapid diffusion of I has a twofold effect. In the neighborhood of the leading edge, the inhibitory effect of I is diminished, so that localized growth of the phosphoinositide peak increases. Outside this neighborhood, the higher concentration of I promotes transfer of membrane phosphoinositides from the plasma membrane to the endoplasmic reticulum. The net effect of this transfer is to deplete the plasma membrane of its phosphoinositides, thus preventing the peak from spreading beyond the leading edge. Hence, within the leading edge, the steady-state concentration of P is higher than the initial basal level. Outside the leading edge, it is lower than the initial basal level.

2.2.2. Polarized Sensitivity

The model has also been used to explore the phenomenon of polarized sensitivity. To this end, the model cell is exposed to a chemoattractant gradient and allowed to form a steady-state phosphoinositide peak consistent with this gradient. After the steady state has been reached, the polarized cell is subjected to a new chemoattractant gradient that is different from the previous gradient. The

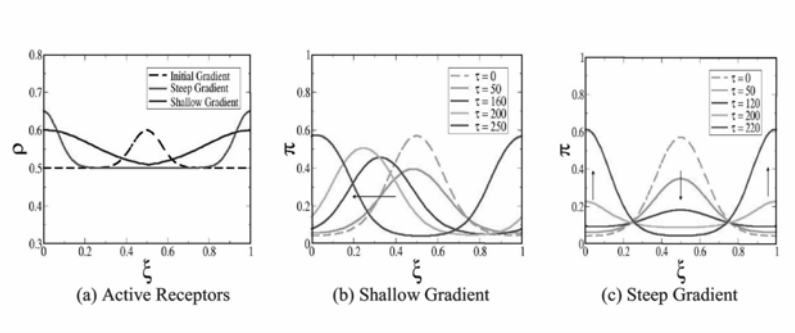


Figure 5. Response of a preexisting phosphoinositide peak to a new chemoattractant gradient (polarized sensitivity): (**a**) the active receptor concentration; (**b**) when the new gradient is shallow, the phosphoinositide peak moves as a traveling wave towards the new steady state; (**c**) when the new gradient is sharp, the original phosphoinositide peak collapses and a new peak is created at the maximum of the new gradient.

simulations show that if the new chemoattractant gradient declines gradually from its maximum in such a way that the receptors in the neighborhood of the preexisting phosphoinositide peak sense the influence of the new gradient, the preexisting peak moves like a traveling wave to the point at which the new chemogradient gradient has a maximum (Figure 5b). On the other hand, if the new chemoattractant gradient is highly localized some distance away from the preexisting phosphoinositide peak such that the receptors in the neighborhood of the preexisting peak do not sense the influence of the new chemoattractant gradient, there is no wave motion. The preexisting peak retracts, and a new peak grows at the maximum of the new gradient (Figure 5c). These results are consistent with the phosphoinositide dynamics observed in experiments.

To explain the wavelike motion of the peak (Figure 5b), it is useful to observe that the steady-state membrane phosphoinositide peak formed in response to the first gradient is "inert" everywhere except in the two thin "transition layers" surrounding the peak within which there is a sharp change in the gradient of membrane phosphoinositides. By "inert," we mean that outside these transition layers nothing is happening at steady state—there is neither diffusion nor synthesis of membrane phosphoinositides. The transition layers, on the other hand, are sites of intense activity even at steady state. In the upper half of a transition layer, there is rapid synthesis of membrane phosphoinositides, which then diffuse into the lower half of the transition layer, from where they are promptly removed. The steady state is maintained by this precarious balance between synthesis of membrane phosphoinositides in the upper half of the transition layers and their removal in the lower half of the transition layers. If the balance is disturbed by imposing a shallow chemoattractant gradient that increases the rate of

phosphoinositide synthesis relative to its rate of diffusion, the transition layer moves in a wavelike fashion at a velocity that is proportional to the net rate of phosphoinositide accumulation within the transition layer.

The response to sharp chemoattractant gradients (Figure 5c) can be explained as follows. After the steady state has developed, the concentration of the inhibitor, *I*, is high throughout the cell. This tends to increase the threshold at all points of the plasma membrane. When the gradient is switched, the active receptor concentration decreases at the previous "front" and increases at the current "front." It follows from the earlier discussion regarding thresholds (see Figure 3b) that at the current "front" the tendency of the threshold to increase due to elevated inhibitor concentrations is mitigated by the higher active receptor concentration. However, at the "previous" front the tendency of the threshold to increase due to elevated inhibitor concentrations is further exacerbated by the lower active receptor concentration. The thresholds at the "previous" front become so large that, despite the large concentrations of membrane phosphoinositides, they fall short of the threshold, and the preexisting peak collapses.

2.2.3. Spontaneous Polarization

The random location of the polarity in *spontaneous* polarization suggests that some variable that is upstream of the phosphoinositides in the signal transduction pathway undergoes stochastic fluctuations. The most upstream source of the stochastic fluctuations is receptor-ligand binding (44). Following (5), we construct a stochastic model of receptor-ligand binding.

To this end, we partition a cell containing r_t receptors into n equal sections. We assume that this cell is exposed to some uniform chemoattractant concentration l. If ligand binds instantaneously to the receptors, the mean number of active receptors in each section, denoted r_m, is $(r_t/n)l/(k^-/k^+ + l)$, where k^+ and k^- are the rate constants for receptor-ligand association and dissociation, respectively. The number of active receptors in each section, denoted, r_i, is given by the stochastic differential equation

$$dr_i = \left[\left(k^+ l + k^-\right)\left(r_m - r_i\right)\right]dt + \sqrt{k^+ l\left(r_t/n - r_m\right) + k^- r_m}\, dW_i, i = 1, 2...n\,. \qquad [12]$$

where the first term on the right denotes the *deterministic* part of receptor ligand binding, which has the effect bringing r_i back to its mean value, r_m. The second term on the right denotes the *stochastic* part of the binding process. Here, dW_i denotes the Wiener process, which is a Gaussian random number generator with zero mean and standard deviation $(dt)^{1/2}$. It should be noted that the standard deviation of the random process, δr_i, is proportional to $(r_t)^{1/2}$, but the relative spread, defined as the ratio, $\delta r/r_m$, is inversely proportional to $(r_t)^{1/2}$. Thus, the

smaller the number of total receptors in a cell, the more pronounced the stochastic effects. If we define the dimensionless variables

$$\tau \equiv \frac{t}{1/(k_r sp_t)}, \rho_i = \frac{r_i}{r_t/n}, \rho_m = \frac{r_m}{r_t/n}, \omega_i = W_i \sqrt{\frac{k_r sp_t}{r_t/n}},$$

and the dimensionless parameters,

$$\lambda = \frac{l}{k^-/k^+}, \kappa^- = \frac{k^-}{k_r sp_t},$$

and equation [12] becomes

$$d\rho_i = \left[\kappa^-(\lambda+1)(\rho_m - \rho_i)\right]d\tau + \sqrt{\kappa^-\left(\lambda(1-\rho_m)+\rho_m\right)}d\omega_i, i = 1,2...n\,. \qquad [13]$$

To simulate the experiments showing spontaneous polarization, it is assumed that at $\tau = 0$ the cell is exposed to a small and uniform chemoattractant concentration, $\lambda = \lambda^-$. At $\tau \geq 0$, the cell is immersed in a uniform concentration of chemoattractant $\lambda^+ > \lambda^-$. Euler integration of equation [13] yields a noisy active receptor distribution, a typical snapshot of which is shown in Figure 6a. The response to stochastic receptor-ligand binding is then simulated by choosing this noisy active receptor distribution as the function $\rho(\tau,\xi)$ in model equations [6]–[8]. The simulation shows that the sudden increase from λ^- to λ^+ causes an initial uniform increase in membrane phosphoinositides, but this is followed by the formation of a phosphoinositide peak at a random location (Figure 6b).

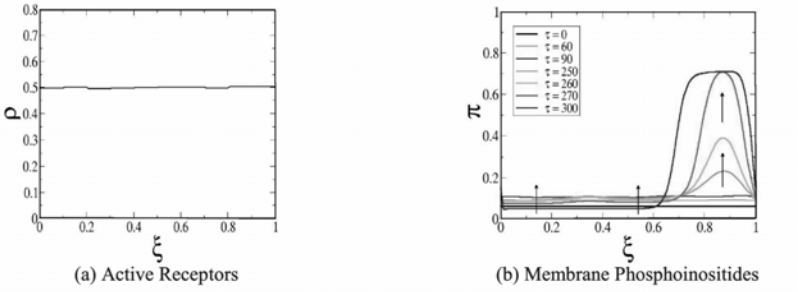


Figure 6. Spontaneous polarization in response to a uniform but noisy chemoattractant profile: (**a**) a snapshot of the noisy active receptor distribution generated by the Tranquillo-Lauffenburger model, and (**b**) development of a phosphoinositide peak in response to this noisy signal.

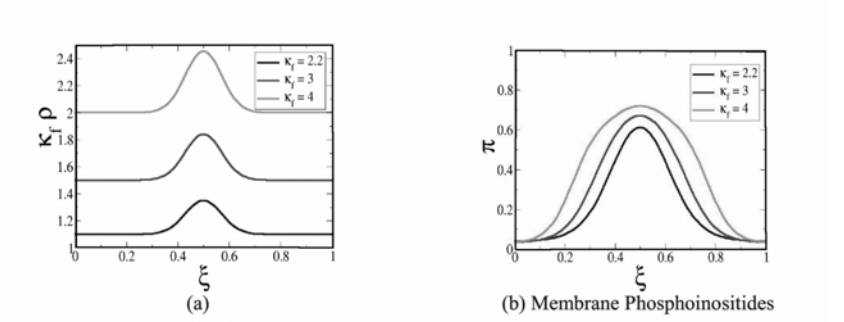


Figure 7. Variation of the steady-state phosphoinositide peaks with respect to different distributions of $\kappa_f\rho$: (**a**) distributions of $\kappa_f\rho$, (**b**) corresponding phosphoinositide peaks.

Recently, we have shown that the phosphoinositide peak forms even though the environment is macroscopically uniform because there is a certain range of chemoattractant concentrations within which the homogeneous steady state of the model is Turing unstable, i.e., unstable with respect to all nonhomogeneous perturbations (40). Hence, any kind of noise drives the cell away from the homogeneous steady state and results in the formation of a phosphoinositide peak. This property is characteristic of activator-inhibitor models (45).

2.2.4. Variation of the Phosphoinositide Peak With Respect to Kinetic Parameters and External Signal

Figure 7 shows the variation of the steady-state phosphoinositide peak with respect to three different distributions of the function $\kappa_f\rho$. We assume that these distributions are achieved by fixing the active receptor distribution, ρ, and varying the parameter, κ_f, i.e., cells with progressively higher levels of PI3K or PI5K are subjected to the same chemoattractant gradient. The simulations show that the development of the phosphoinositide peak occurs only within a certain range of κ_f. Even within this range of existence

1. The chemoattractant gradient required to provoke peak formation decreases with κ_f, and becomes zero for sufficiently large κ_f.

2. The width of the peak increases with κ_f (Figure 7).

When $\kappa_f > 5$, the peak disappears completely. At such large values of κ_f, no initial perturbation can provoke the formation of a stable phosphoinositide peak. Similarly, decreasing κ_f narrows the peak until at a sufficiently small value ($\kappa_f <$ 2.2) peak formation cannot be induced.

These results have the following physical interpretation. The parameter κ_f is the ratio of the characteristic velocities of phosphoinositide synthesis and removal. If κ_f is small, phosphoinositide synthesis is rapidly opposed by inhibitory

action of inositol phosphates, resulting in high thresholds and narrow peaks. When $\kappa_f < 2.2$, even the maximum possible increment of ρ cannot abolish the threshold at the leading edge, thus making it impossible to provoke a peak. Conversely, if κ_f is large, the inositol phosphate pool responds slowly, so that the peak is wider and the threshold lower. At very large values of κ_f, the threshold is zero and membrane phosphoinositide synthesis is much faster than its removal. Hence, the minutest chemoattractant gradient results in propagation of phosphoinositides throughout the membrane before inositol phosphates can exert the inhibitory effect required for formation of the polarized steady state. Thus, the phosphoinositides ultimately increase to a near uniform steady state.

Now, we could just as well assume that the three different distributions of $\kappa_f\rho$ shown in Figure 7 were obtained by keeping the parameter κ_f fixed, and varying the active receptor distribution ρ. In this case, the simulations would imply that the geometry of the phosphoinositide peak varies with the shape of the external chemoattractant.

3. FUTURE WORK

We have shown above that the model captures all the dynamics except adaptation. To be sure, we modeled adaptation in our earlier work by assuming slow desensitization and resensitization of the receptors (39). In response to uniform increases of the chemoattractant concentration, the model displayed perfect adaptation over several orders of magnitude of the chemoattractant concentrations. However, the mechanism proposed in the model was inconsistent with experiments, since adaptation occurs at some level below the receptors (46) and G-proteins (24). This deficiency was resolved in more recent work, wherein we have taken due account of the fact that adaptation acts below the receptors (40). Simulations of this modified model are in good agreement with experiments.

Although our local-activator-global-inhibitor model captures the spatiotemporal dynamics observed in experiments, the identity of the local activator and global inhibitor are subjects of considerable debate. Specifically,

1. Experimental data concerning the polarization of PIP_2 is not definitive. It is difficult to discern spatial variations of the PIP_2 distribution because its concentration in the plasma is relatively high. It has been argued that the appearance of PIP_2 polarization observed in earlier work (18) does not reflect a localized increase in the concentration of PIP_2—it is a consequence of the high surface area created at the leading edge by the formation of membrane folds (47). On the other hand, it has been shown that PI5K, the enzyme that synthesizes PIP_2, localizes to the ruffles immediately after chemoattractant stimulation (29,48).

2. The role of PIP_3 remains elusive. It has been shown, for instance, that PI3K, the enzyme that catalyzes the synthesis of PIP_3, is sufficient, but not necessary for polarization. Delivery of exogenous PIP_3 to neutrophils provokes polarity development (9). However, in *Dictyostelium* cells lacking both PI3K1 and PI3K2 show only partial defects in chemotaxis (49).

3. The influence on polarization of molecules that are downstream of PIP_3 is controversial. In *Dictyostelium*, phosphoinositides are polarized even if actin polymerization is inhibited by latrunculin (27). This implies that phosphoinositides are polarized exclusively by reactions upstream of actin polymerization. In neutrophils, on the other hand, the polarization is almost completely abolished when actin polymerization is inhibited by latrunculin or promoted by jasplakinolide (33).

4. Several potential global inhibitors have been hypothesized, but the evidence supporting their inhibitory role is either lacking or inconclusive. The enzyme PTEN, which catalyzes dephosphorylation of PIP_3 to PIP_2, migrates in a manner that strongly suggests that it may be a global inhibitor. While PIP_3 localizes to the "front" of the cell, PTEN translocates to the "back" of the cell (50,51). However, the mechanism of this translocation is unknown. We have proposed that the cytosolic pool of inositol phosphates plays the role of global inhibitor (39). This hypothesis is supported by experiments which show that blocking the synthesis of inositol phosphates stimulates chemotaxis (53). However, the inhibitory mechanism appears to be different from the one hypothesized by us. Indeed, Luo et al. (52) have shown that in *Dictyostelium* the inositol phosphate InsP7 competitively inhibits the binding of various PH-domain-containing effector enzymes to PIP_3, thus retarding their recruitment to the membrane. Finally, it has also been suggested that cGMP fulfills the role of inhibitor (37), but this hypotheses has not been rigorously tested.

The resolution of these outstanding issues will play a crucial role in fostering further model development. Indeed, all the existing gradient sensing models are "lumped" because our ignorance regarding the identity of the activator and the inhibitor argues against the development of more detailed models. As noted in this volume (see Part III, chapter 1.4, by Dhar and Tomita), a model is only as good as the data it purports to explain. At present, we have reliable data on the spatiotemporal dynamics of the phosphoinositide localization. By establishing that these nontrivial dynamics are consistent with the activator-inhibitor model, we hope that we have sharpened the focus of the experimental search for the variables. To the extent that the activator-inhibitor model is a valid representation of the gradient sensing mechanism, the experimentalist can now focus attention on variables possessing the dynamic properties of the activator and inhibitor. Further refinement of the model must await the identification of these variables.

The theory of pattern formation was originally inspired by a desire to explain tissue differentiation (54). The theory showed that in a reaction-diffusion system chemical inhomogeneities could arise spontaneously, and that these inhomogeneities could confer distinct attributes to different parts of a tissue. Yet, the link between this theory and experiments remains tenuous because of the difficulties encountered in identifying the variables of the theory, namely, the activators and the inhibitors. In recent years, it has become evident that there are chemical inhomogeneities even within a single cell. The chemical and morphological polarization observed in the eukaryotic gradient sensing mechanism is a paradigm of such subcellular pattern formation. Given our deep understanding of cellular and molecular biology, it would not be surprising if pattern formation theory faces its first rigorous experimental tests at the level of cells rather than tissues.

4. ACKNOWLEDGMENTS

We would like to thank Professor D.A. Lauffenburger (MIT) for introducing us to the problem and for highlighting the important questions. We are also grateful to Dr. Orion Weiner (Harvard) and Professor Richard Dickinson (University of Florida) for their thoughtful comments.

5. NOTES

1. Abbreviations used: PI = phosphatidylinositol; PIP = phosphatidylinositol 4-phosphate; PIP_2 = phosphatidylinositol 4,5-phosphate; PIP_3 = phosphatidylinositol 3,4,5-phosphate; DG = diacylglycerol; IP_3 = inositol 1,4,5-triphosphate; PI3K = phosphatidylinositol 3-kinase; PI5K = phosphatidylinositol 4-phosphate 5-kinase; PITP = phosphatidylinositol transport protein.

2. To be sure, this implies that the angular gradients are also negligibly small, and this will be borne out by the simulations below. We retain the angular variation only because it simplifies the numerical simulations. Neglecting the angular variation yields an inconvenient "mixed" system of equations containing an ODE describing the evolution of the inositol phosphates along with the other PDEs. At any rate, we suffer no loss of generality by retaining the angular variation.

6. REFERENCES

1. Martin P. 1997. Wound healing—aiming for perfect skin regeneration. *Science*, **276**:75–81.
2. Jones G. 2000. Cellular signaling macrophage migration and chemotaxis. *J Leuk Biol* **68**:593–602.

3. Moore M. 2001. The role of chemoattraction in cancer metastases. *Bioessays* **23**:674–676.
4. Lauffenburger DA, Horwitz AF. 1996. Cell migration: a physically integrated molecular process. *Cell* **84**:359–369.
5. Tranquillo RT, Lauffenburger DA, Zigmond SH. 1988. A stochastic model for leucocyte random motility and chemotaxis based on receptor-binding fluctuations. *J Cell Biol* **106**:303–309.
6. Coates T, Watts R, Hartman R, Howard T. 1992. Relationship of F-actin distribution to development of polar shape in human polymorphonuclear neutrophils. *J Cell Biol* **117**:765–774.
7. Welch HCE, Coadwell WJ, Ellson CD, Ferguson GJ, Andrews SR, Erdjument-Bromage H, Tempst P, Hawkins PT, Stephens LR. 2002. P-Rex1, a PtdIns(3,4,5)P_3- and Gβγ-regulated guanine-nucleotide exchange factor for Rac. *Cell* **108**:809–821.
8. Tolias KF, Hartwig JH, Ishihara H, Shibasaki Y, Cantley LC, Carpenter CL. 2000. Type Iα phosphatidyl-4-phosphate 5-kinase mediates Rac-dependent assembly. *Curr Biol* **10**:153–156.
9. Weiner O, Neilsen P, Prestwich G, Kirschner M, Cantley L, Bourne H. 2002. A PtdInsP(3)- and Rho GTPase-mediated positive feedback loop regulates neutrophil polarity. *Nature Cell Biol* **4**:509–513.
10. Ishihara H, Shibasaki Y, Kizuki N, Wada T, Yazaki Y, Asano T, Oka Y. 1998. Type I phosphatidylinositol 4-phosphate 5-kinases. *J Biol Chem* **273**:8741–8748.
11. Willars GB, Nahorski SR, Challiss RAJ. 1998. Differential regulation of muscarinic acid acetylcholine receptor-sensitive polyphosphoinositide pools and consequences for signaling in human neuroblastoma cells. *J Biol Chem* **273**:5037–5046.
12. Cockcroft S. 1999. Mammalian phosphatidylinositol transfer proteins: emerging roles in signal transduction and vesicular traffic. *Chem Phys Lipids* **98**:23–33.
13. Berridge MJ, Irvine RF. 1989. Inositol phosphates and cell signaling. *Nature* **34**:197–205.
14. Servant G, Weiner OD, Herzmark P, Balla T, Sedat JW, Bourne HR. 2000. Polarization of chemoattractant receptor signaling during neutrophil chemotaxis. *Science* **287**:1037–1040.
15. Zigmond SH. 2000. How WASP regulates actin polymerization. *J Cell Biol* **150**:F117–F119.
16. Mullins RD. 2000. How WASP-family proteins and the Arp2/3 complex convert intracellular signals into cytoskeletal structures. *Curr Opin Cell Biol* **12**:91–96.
17. Borisy GG, Svitkina TM. 2000. Actin machinery: pushing the envelope. *Curr Opin Cell Biol* **12**:104–112.
18. Tall EG, Spector I, Pentyala SN, Bitter I, Rebecchi MJ. 2000. Dynamics of phosphatidylinositol 4,5-biphosphate in actin-rich structures. *Curr Biol* **10**:743–746.
19. Balla T, Bondeva T, Varnai P. 2000. How accurately can we image inositol lipids in living cells? *Trends Pharmacol Sci* **21**:238–241.
20. Lilly PJ, Devreotes PN. 1995. Chemoattractant and GTPγs-medicated stimulation of adenylyl cyclase in *Dictyostelium* requires translocation of CRAC to membranes. *J Cell Biol* **129**:1659–1665.
21. Servant G, Weiner OD, Neptune ER, Sedat JW, Bourne HR. 1999. Dynamics of a chemoattractant receptor in living neutrophils during chemotaxis. *Mol Cell Biol* **10**:1163–1178.
22. Xiao Z, Zhang N, Murphy DB, Devreotes PN. 1997. Dynamic distribution of chemoattractant receptors in living cells during chemotaxis and persistent stimulation. *J Cell Biol* **139**:365–374.
23. Ueda M, Sako Y, Tanaka T, Devreotes P, Yanagida T. 2001. Single-molecule analysis of chemotactic signaling in *Dictyostelium* cells. *Science* **294**:864–867.
24. Janetopoulos C, Jin T, Devreotes P. 2001. Receptor-mediated activation of heterotrimeric G-proteins in living cells. *Science* **291**:2408–2411.
25. Jin T, Zhang N, Long Y, Parent CA, Devreotes PN. 2000. Localization of the G protein βγ complex in living cells during chemotaxis. *Science* **287**:1034–1036.
26. Meili R, Ellsworth C, Lee S, Reddy TBK, Ma H, Firtel RA. 1999. Chemoattractant-mediated transient activation and membrane localization of Akt/PKB is required for efficient chemotaxis to cAMP in *Dictyostelium*. *EMBO J* **18**:2092–2105.
27. Parent CA, Blacklock BJ, Froehlich WM, Murphy DB, Devreotes PN. 1998. G protein signaling events are activated at the leading edge of cells. *Cell* **95**:81–91.

28. Haugh JM, Codazzi F, Teruel M, Meyer T. 2000. Spatial sensing in fibroblasts mediated by 3' phosphoinositides. *J Cell Biol* **151**:1269–1279.
29. Honda A, Nogami M, Yokozeki T, Yamazaki M, Nakamura H, Watanabe H, Kawamoto K, Nakayama K, Morris AJ, Frohman MA, Kanaho W. 1999. Phosphatidylinositol 4-phosphate 5-kinase α is a downstream effector of the small G protein AFR6 in membrane ruffle formation. *Cell* **99**:521–532.
30. Zigmond SH, Levitsky HI, Kreel BJ. 1981. Cell polarity: an examination of its behavioral expression and its consequences for polymorphonuclear leucocyte chemotaxis. *J Cell Biol* **89**:585–592.
31. Chung CY., Funamoto S, Firtel RA. 2001. Signaling pathways controlling cell polarity and chemotaxis. *Trends Biochem Sci* **26**:557–566.
32. Firtel RA, Chung CY. 2000. The molecular genetics of chemotaxis: sensing and responding to chemoattractant gradients. *Bioessays* **22**:603–615.
33. Wang F, Herzmark P, Weiner O, Srinivasan S, Servant G, Bourne H. 2002. Lipid products of PI(3)Ks maintain persistent cell polarity directed motility in neutrophils. *Nature Cell Biol* **4**:513–518.
34. Wedlich-Soldner R, Li R. 2003. Spontaneous cell polarization: undermining determinism. *Nature Cell Biol* **5**:267–270.
35. Othmer H, Schaap P. 1998. Oscillatory camp signaling in the development of *Dictyostelium discoideum*. *Comments Theor Biol* **5**:175–282.
36. Levchenko A, Iglesias P. 2002. Models of eukaryotic gradient sensing: application to chemotaxis of amoebae and neutrophils. *Biophys J* **82**:50–63.
37. Rappel W-J, Thomas P, Levine H, Loomis W. 2002. Establishing directing during chemotaxis in eukaryotic cells. *Biophys J* **83**:1361–1367.
38. Meinhardt H. 1999. Orientation of chemotactic cells and growth cones: models and mechanisms. *J Cell Sci* **112**:2867–2874.
39. Narang A, Subramanian KK, Lauffenburger DA. 2001. A mathematical model for chemoattractant gradient sensing based on receptor-regulated membrane phospholipid signaling dynamics. *Annu Biomed Eng* **29**:677–691.
40. Subramanian KK, Narang A. 2004. A mechanistic model for eucaryotic gradient sensing: spontaneous and induced phosphoinositide polarization. *J Theor Biol* **231**:49–67.
41. Postma M, Van Haasert P. 2001. A diffusion-translocation model for gradient sensing by chemotactic cells. *Biophys J* **81**:1314–1323.
42. Zigmond SH, Sullivan SJ, Lauffenburger DA. 1982. Kinetic analysis of chemotactic peptide receptor modulation. *J Cell Biol* **92**:34–43.
43. *NAG fortran library introductory guide, mark 18*. 1999. Numerical Algorithms Group, Oxford.
44. Lauffenburger DA, Linderman JJ. 1993. *Receptors: model for binding, trafficking and signaling*, 1st ed. Oxford UP, New York.
45. Kerner BS, Osipov VV. 1994. *Autosolitons: a new approach to problems of self-organization and turbulence*, 1st ed. Kluwer Academic Publishers, Dordrecht.
46. Kim J, Soede R, Schaap P, Valkema R, Borleis J, Haastert P, Devreotes P, Hereld D. 1997. Phosphorylation of chemoattractant receptors is not essential for chemotaxis or termination of G-protein-mediated chemotaxis or termination of G-protein-mediated responses. *J Biol Chem* **272**:27313–27318.
47. van Rheenan J, Jalink K. 2002. Agonist-induced PIP2 hydrolysis actin dynamics: regulation at a global but not at a micrometer scale. *Mol Biol Cell* **13**:3257–3267.
48. Doughman R, Firestone A, Wojtasiak M, Bunce M, Anderson R. 2003. Membrane ruffling requires coordination between type Iα phosphatidylinositol phosphate kinase and Rac signaling. *J Biol Chem* **278**:23036–23045.
49. Iijima M, Devreotes P. 2002. Tumor suppressor PTEN mediates sensing of chemoattractant gradients. *Cell* **109**:599–610.

50. Funamoto S, Meili R, Lee S, Parry L, Firtel R. 2002. Spatial and temporal regulation of 3-phosphoinositides by PI 3-kinase and PTEN mediates chemotaxis. *Cell* **109**:611–623.
51. Iijima M, Huang Y, Devreotes P. 2002. Temporal and spatial regulation of chemotaxis. *Dev Cell* **3**:469–478.
52. Luo HR, Huang YH, Chen JC, Saiardi A, Iijima M, Ye K, Huang Y, Nagata E, Devreotes P, Snyder SH. 2003. Inositol pyrophosphate mediate chemotaxis in *Dictyostelium* via pleckstrin homology domain-PtdIns(3,4,5)P_3 interactions. *Cell* **114**:559–572.
53. Li Z, Jiang H, Xie W, Zhang Z, Smrcka A, Wu D. 2000. Roles of PLC-β2 and –β3 and PI3Kγ in chemoattractant-mediated signal transduction. *Science* **287**:1046–1049.
54. Turing A. 1952. The chemical basis of morphogenesis. *Phil Trans Roy Soc Lond B* **237**:37–72.

2.3

PATTERNING BY EGF RECEPTOR: MODELS FROM *DROSOPHILA* DEVELOPMENT

Lea A. Goentoro and Stanislav Y. Shvartsman

Lewis-Sigler Institute for Integrative Genomics,
Princeton University, Princeton, New Jersey

The epidermal growth factor receptor (EGFR) belongs to a large class of receptor tyrosine kinases. Abnormal EGFR signaling is associated with severe developmental defects and many types of cancers. Many individual molecules mediating the EGFR-induced responses became drug targets in oncology and other areas of medicine. However, neither the contribution of EGFR to tissue morphogenesis in development nor the exact role of deregulated EGFR signaling in diseases is understood at this time. The key challenge is to integrate the existing molecular and cellular information into a systems-level description of the EGFR network in tissues. Systems-level descriptions are impossible without quantitative models. Even the simplest models of EGFR signaling in tissues must simultaneously account for ligand transport, binding, signal transduction, and gene expression. Given this complexity, such tissue-level models are difficult to test; therefore, they require appropriate experimental paradigms for their validation. We suggest that model organisms of developmental genetics, such as the fruit fly *Drosophila melanogaster*, can be used as experimental systems for the development and validation of computational descriptions of EGFR signaling in tissues.

1. INTRODUCTION

The epidermal growth factor receptor (EGFR) is an evolutionarily conserved regulator of epithelial tissues. The first identified receptor tyrosine kinase

Stanislav Y. Shvartsman, Lewis-Sigler Institute for Integrative Genomics, Princeton University, Carl Icahn Laboratory, Washington Road, Princeton, NJ 08544; (609) 258-7071; stas@princeton.edu.

and a founding member of the ErbB receptor family, EGFR has been implicated in countless physiological and pathological contexts (1,2). Most commonly, EGFR is activated by extracellular ligands. Ligand binding induces dimerization of the receptor and activates the kinase in its cytoplasmic domain. By recruiting and phosphorylating the cytoplasmic targets, the activated receptor couples to signal transduction pathways and controls cellular responses (see also the preceding chapter 2.2, by Subramanian and Narang). While experiments in cell culture keep providing invaluable insights into the structure and function of the EGFR network, more complex experimental systems are required to study EGFR signaling in tissues. Co-culture models and cultured tissues can be used to probe EGFR signaling in multicellular systems (3,4). Finally, analysis of the organism-level effects of EGFR signaling requires studies in vivo.

EGFR activation in vivo is mediated by autocrine and paracrine signals. Secreted ligands usually bind to receptors on the ligand-producing cells or their neighbors. Receptor activation depends on the rates of ligand release, receptors levels, and tissue architecture. Typically, ligand/receptor levels and activation of downstream pathways are assessed using in situ hybridization or immunohistochemistry. Since these techniques are nontrivial to quantitate, even the "simplest" parameters of autocrine and paracrine networks, such as ligand concentrations, cannot be measured directly. In contrast, in the studies conducted in vitro, one can both control the exogenous ligand concentration and measure receptor levels using a number of quantitative assays.

In theory, modeling and computations can bridge the apparent gap between the in-vitro and in-vivo studies of EGFR biology (5). Again in theory, the biochemical parameters measured in vitro can provide inputs to the tissue-level models. These models can estimate the parameters that are either impossible or difficult to measure directly. For example, the information about receptor dynamics generated in cell culture can be combined with the microscopically derived information about the tissue architecture in order to compute the spatial distribution of autocrine and paracrine signals (6). In this way, cellular and biochemical studies can drive the development of mechanistic models of cell communication in tissues.

The experimental validation of tissue-level models requires a flexible experimental system. With its advanced experimental genetics, the fruit fly *Drosophila melanogaster* serves as an excellent testing ground for validation of models of EGFR signaling in tissues (7). Further, the high evolutionary conservation makes the *Drosophila* EGFR network an excellent model for the more complex mammalian EGFR systems (8). In this chapter we first describe two examples of EGFR-mediated patterning in fruit fly development. These examples serve to illustrate how EGFR signaling is exquisitely tuned to produce the appropriate patterns of gene expression during development. Next we describe some of our initial work in the mechanistic modeling of these systems. Our emphasis is on the spatial range of autocrine and paracrine signals and the dynam-

ics of feedback loops in the EGFR network. And finally, we conclude by describing some of the current challenges involved in modeling EGFR at the tissue level.

2. TWO EXAMPLES OF EGFR SIGNALING IN FRUIT FLY DEVELOPMENT

EGFR is recurrently used in fruit fly development (see Shilo (7) for a recent review). The pleiotropic nature of EGFR signaling was realized when mutations affecting different stages of development were mapped to the same location, the *Egfr* gene.[1] Within several years, the *Drosophila* EGFR ligands, all secreted molecules, were identified: Gurken, Spitz, and Keren, all three being homologous to the human Transforming Growth Factor alpha (TGFα); and Vein, a homologue of the human Neuregulin. Several inhibitors have also been identified. Most relevant to our discussion here is Argos, a secreted inhibitor acting through sequestering the activating ligands (9). At this time, no mammalian counterpart of Argos has been identified.

The presence of multiple ligands and inhibitors of EGFR signaling does not necessarily imply redundancy in the system. Gurken, Spitz, and Keren are produced as transmembrane precursors and activated through a proteolytic cleavage. In contrast, Vein and Argos are expressed in their active secreted forms. Furthermore, Spitz, Vein, and Argos, when acting in a feedback fashion, require a different threshold of signaling activity for their induction. The different modes of induction reflect the different modules of regulations used to limit the availability of the molecules. The different regulations are translated into different characteristics of the molecules (e.g., time delay upon induction, steep versus gradual response to induction) that may govern the choice of using a specific ligand/inhibitor at a given place and time. And finally, the molecules have different transport and kinetic properties (e.g., diffusion coefficient, binding rate constants) that would influence the way they regulate EGFR activity.

In the following two sections, we review two well-studied examples of EGFR-mediated tissue patterning during the egg and embryonic development in the fruit fly *Drosophila melanogaster*. During embryogenesis, an initial gradient of EGFR activity created by a locally secreted activator is fine-tuned by a secreted inhibitor (10,13,14). Spitz acts as the activator that induces the expression of Argos in a feedback fashion. During egg development, the initial domain of EGFR activity is first expanded by a secreted activator and later refined by a secreted inhibitor (11). In this case, Gurken acts as the initial activator that induces the expression of Spitz and subsequently Argos. Thus, Gurken, Spitz, and Argos define a spatially distributed network controlling the EGFR activation. These two examples illustrate the patterning versatility of the autocrine and paracrine EGFR networks.

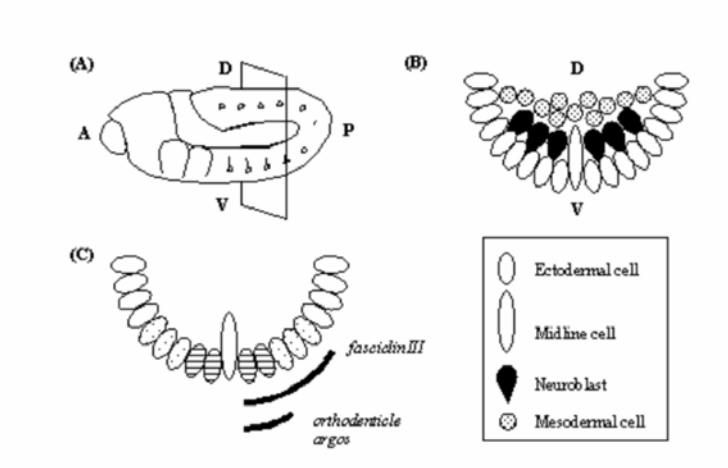


Figure 1. (**A**) A mid-stage fly embryo (D = dorsal, V = ventral, A = anterior, P = posterior). (**B**) Cross-section of the embryo at the indicated plane in (Figure 1A). (**C**) The ventral ectodermal cells are patterned by Egfr into two subgroups. The ventral-most cells (striped) express all the genes listed, whereas the ventrolateral cells (dotted) express the *fasciclinIII* gene only (12).

2.1. EGFR Signaling in Embryogenesis: Ventral Ectodermal Patterning

A fly embryo halfway through embryogenesis is shown in Figure 1A. Three layers of cells are present at this stage (Figure 1B): the ectoderm, which will form the larval epidermis; the neuroblast, which will give rise to the nervous system; and the mesoderm, which will develop into muscle and connective tissues. The midline cells, which divide the ectoderm along the dorsoventral axis, are distinct from their neighboring ectodermal cells. These midline cells will later delaminate from the ectoderm and give rise to specific neurons and midline glial cells. EGFR induces two different fates in the ventral ectoderm: the ventral-most and ventrolateral fates (12). The two fates are identifiable by the expression of different marker genes (Figure 1C).

The following mechanism, summarized in Figure 2, was proposed by Golembo et al. (10,13,14) to explain the EGFR-mediated patterning of ventral ectoderm. The patterning process starts when the *single-minded* gene, expressed exclusively in the midline cells, induces the expression of *rhomboid* (15). Rhomboid is a protease that cleaves the Spitz transmembrane precursor into the active secreted form (16). As a result, the midline acts as a local source of secreted Spitz, establishing a gradient of EGFR activation in the neighboring ectodermal cells.

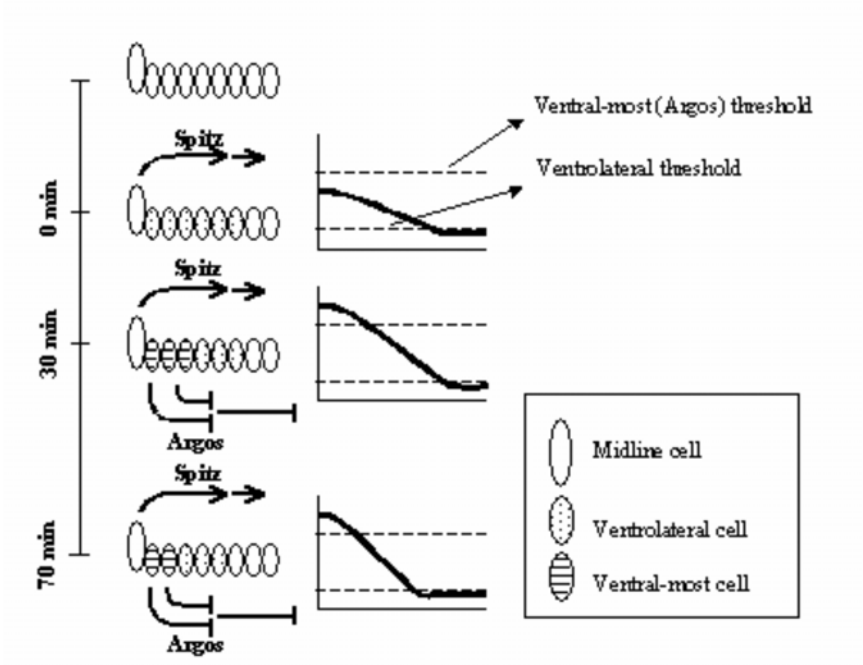


Figure 2. Ventral ectodermal patterning. In the graphs, the *y*-axis is the Egfr activity and the *x*-axis the distance away from the midline. All cells are initially uniform. Patterning is initiated by secretion of Spitz from the midline cells. This creates a graded activation of EGFR and induction of the low-threshold ventrolateral fate (dotted) in nearby ectodermal cells. In time, the signaling activity increases such that the cells nearest to the midline (striped) reach the high threshold of the ventral-most fate and start secreting Argos. Inhibition by Argos modifies the gradient of activity, restricting the domain of each ventral fate. This pattern persists for *at least* 3 more hours. (Note: The timeline is approximated from various published data.)

Subsequently, the cells nearest to the midline, hence the ones exposed to the highest level of Spitz, start secreting Argos. This establishes a negative feedback that is thought to refine the gradient of Spitz-induced EGFR activity and regulate the relative number of cells adopting each of the ventral fates. As a result of the interplay between Spitz and Argos, five rows of cells flanking the midline are patterned. The two rows nearest to the midline receive a high level of EGFR activation and adopt the ventral-most fate. The next three rows of cells receive a moderate level of EGFR activation and adopt the ventrolateral fates. In summary, the initial gradient of EGFR activity induced by a localized secretion of an activator (Spitz) is refined by a secreted inhibitor (Argos), creating two distinct fates.

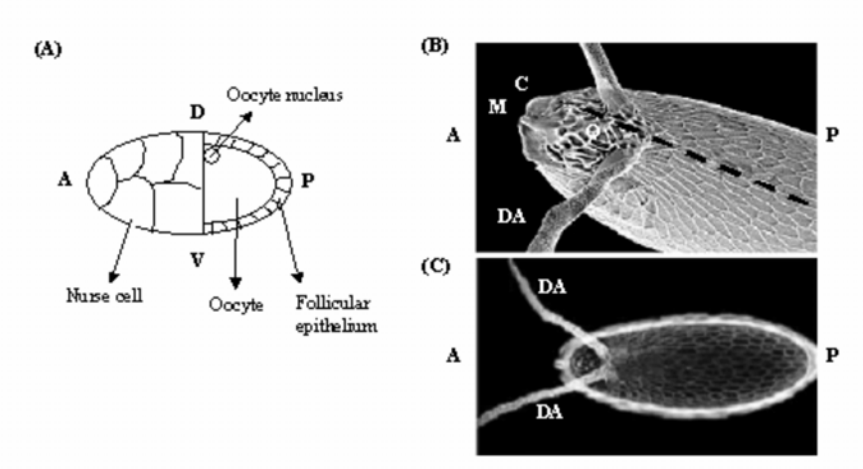


Figure 3. (**A**) A mid-stage egg chamber. At this stage, the nucleus has just migrated from the posterior to the future dorsal position (A = anterior, P = posterior, D = dorsal, V = ventral). (**B**) A scanning electron micrograph of the dorsal-anterior section of a mature egg chamber (19). (Reproduced from (19) with permission "Development," Company of Biologists.) Specialized regions can be identified: the micropyle (M), the entry point of the sperm; the operculum (O), a distinct region between the dorsal appendages (DA) that serves as the larval exit door; and the collar region (C), which delineates the operculum. The broken line indicates the midline (i.e., the dorsal-most aspect of the egg). (**C**) A dark micrograph of a mature fly egg chamber with the dorsal appendages extending out from the dorsal-anterior side.

2.2. EGFR Signaling in Oogenesis: Eggshell Patterning

A fly egg is composed of three types of cells (Figure 3A): the oocyte, which later develops into the embryo; the nurse cells, which supply nutrients to the oocyte; and the follicle cells, which form an epithelium enveloping the oocyte. The EGFR-mediated patterning of the follicle cells is highly regulated in space and time (17,18). The net result of this patterning process is division of the initially equivalent follicle cells into distinct populations. Each population of follicle cells gives rise to a specialized structure of the eggshell. EGFR signaling first divides the follicle cells into dorsal and ventral cells. Afterward, EGFR signaling further subdivides the dorsal cells into several subpopulations that give rise to different dorsal structures (Figure 3B) (18). Here we focus on the respiratory appendages extending from the dorsal-anterior side of the egg, whose positioning along the dorsoventral axis requires EGFR signaling (Figure 3C) (20).

Gurken is localized around the oocyte nucleus throughout egg development (21). The relevant patterning process starts during mid-oogenesis, at the time when the oocyte nucleus has just migrated from its previous posterior position to a random point along the oocyte anterior circumference (22,23). Gurken

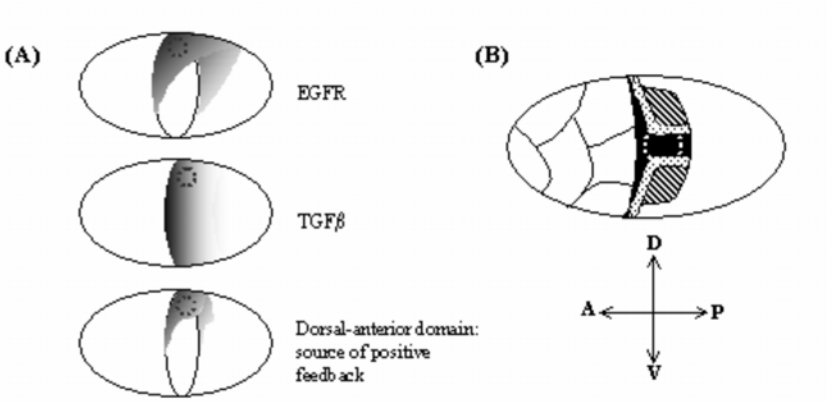


Figure 4. In all figures, the broken circle indicates the position of the oocyte nucleus. (**A**) The overlap between the dorsal-ventral gradient of Gurken-mediated EGFR signaling and the anterior-posterior gradient of TGFβ signaling defines the dorsal-anterior region. (B) The dorsal-anterior is patterned further by EGFR signaling into three different fates: the operculum-producing dorsal midline cells (black), the dorsolateral cells that position the future dorsal appendages (dotted), and the appendage progenitors (striped). The arrows indicate the orientation in all figures (A = anterior, P = posterior, D = dorsal, V = ventral).

released locally from the oocyte induces a gradient of EGFR signaling in the overlying follicle cells. The follicle cells closest to the oocyte nucleus, hence receiving a high level of EGFR signaling, adopt the dorsal fate; the cells far from the nucleus receive a low level of EGFR signaling and express the ventral marker genes.

A dorsal-anterior domain in the follicular epithelium is defined by the overlap between the gradient of the Gurken-mediated high EGFR activity and a transverse gradient of activity of the TGFβ pathway (Figure 4A) (24). Over time, the EGFR signaling further divides the dorsal-anterior domain into three distinct groups of cells: the dorsal midline cells, which will contribute to the production of operculum; the dorsolateral cells, which will specify the position of the dorsal appendages; and the dorsal appendage anlagen, which will the secrete the appendage materials (Figure 4B).

The mechanism by which EGFR signaling mediates the subsequent refinements in the dorsal cells is not well understood. The currently accepted mechanism (Figure 5), proposed by Wasserman and Freeman (11), is by no means the complete story. Expression of *rhomboid* is induced through a crosstalk between the EGFR and TGFβ signaling in the dorsal-anterior domain (Figure 4A) (24). Rhomboid cleaves the Spitz transmembrane precursors, leading to localized secretion of Spitz. The released Spitz acts as a positive feedback that amplifies the initial EGFR activity induced by Gurken.

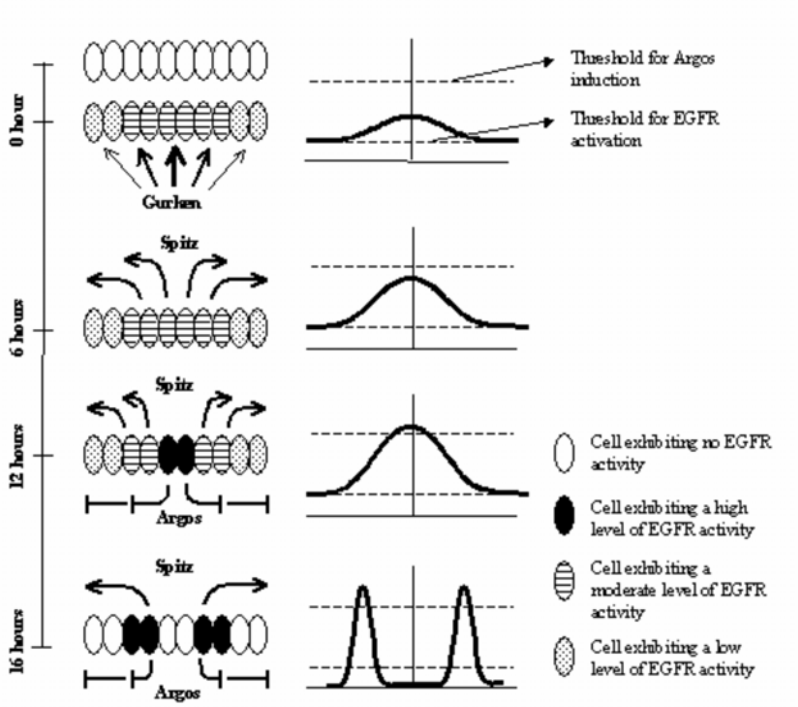


Figure 5. Eggshell patterning. In the graphs, the *y*-axis is the EGFR activity and the *x*-axis the distance away from the midline. All cells are initially uniform. Patterning is initiated by secretion of Gurken from the oocyte. This creates a graded activation of EGFR in the cells nearest to the oocyte nucleus. Although omitted from the subsequent figures, Gurken secretion is constant for *at least* 12 hours. The AP gradient of the TGFβ ligand is not formed until ~6 hours later; following that, the cells start secreting Spitz. The following mechanism is proposed by Wasserman and Freeman (11). Due to amplification by Spitz, the cells nearest the midline reach the high threshold and start secreting Argos. Inhibition by Argos splits the domain of EGFR activity into two and restricts expansion to the lateral side. As a consequence of its own inhibitory action, the domain of Argos expression also splits into two stripes, coinciding with the high-EGFR activity domains. These Argos-expressing cells become the dorsolateral cells; the cells in which the high Egfr activity is extinguished by Argos become the dorsal midline cells; the more lateral cells in which the moderate Egfr activity is turned off by Argos become the appendage progenitor cells. This pattern persists, throughout the morphogenetic movements during the appendage formation, until the end of oogenesis (~6 more hours). (Note: The timeline is approximated from various published data.)

Above a certain high threshold of EGFR activity, Argos is induced. Argos negative feedback is thought to split the initial single domain of high EGFR activity and at the same time restrict further expansion of signaling activation into the lateral side. Consequently, the domain of high EGFR activity is gradually

reduced into two L-shaped stripes positioned symmetrically around the midline (Figure 4B). The cells in which the EGFR activity is initially very high and then abruptly extinguished by Argos adopt the dorsal midline fate; the cells in which the EGFR activity is maintained high adopt the dorsolateral fate; the cells in which the EGFR activity is moderate and then quenched by Argos become the dorsal appendage anlagen. In short, the initial gradient of EGFR activity induced by a localized input (Gurken) is first amplified by a positive feedback (Spitz) and afterward refined by a negative feedback (Argos), thereby creating three distinct fates.

In both patterning systems, EGFR signaling mediates the patterning of gene expression in the epithelial cells. Just as analyzing the similarities between the two systems enables us to derive common regulatory principles in EGFR signaling, examining the differences between the two systems may reveal further mechanistic insights. In the ovary, Argos inhibits EGFR signaling in the Argos-producing cells and their neighbors (11). In the embryo, on the other hand, the Argos-producing cells appear to be refractory to its inhibitory action (10,14). Understanding this observed difference may lead to a better grasp of the way Argos functions. Another difference between the two systems is the absence of positive feedback in the ventral ectoderm.[2] The ventral ectodermal patterning has been shown to be highly robust. Halving or doubling the level of input from the midline does not alter the pattern (10). In contrast, varying the level of input from the oocyte generates diverse eggshell morphology phenotypes (20). In particular, defects in the positive feedback lead to eggshells with fused dorsal appendages (11). Hence, it appears that the positive feedback is necessary to transform a pattern mediated by a simple inductive signal (Gurken) into a more complex spatial pattern.

Long-range inhibition by secreted Argos plays a central role in both mechanisms. Most likely, these mechanisms will have to be revised to account for the recent discovery of the fact that Argos directly interacts with the EGFR ligand Spitz (9), and not with the receptor itself.

3. MODELING AND COMPUTATIONAL ANALYSIS OF AUTOCRINE AND PARACRINE NETWORKS

Here we describe our initial steps toward building mechanistic models of autocrine and paracrine EGFR signaling in epithelial layers. In each case we briefly describe the model and illustrate the representative questions that the model can help to address. The details of the derivations and computational analysis can be found elsewhere (25–30). All the parameters used from this section onward are listed in Table 1.

Table 1. Parameters in the model (25–30,39,40)*

Param.	Description	Value
D	Ligand diffusion coefficient	10^{-10}–10^{-6} cm^2/s (37,38)
R_{tot}	Number of receptors per cell	10^3–10^5 rec/cell
k_{on}	Receptor-ligand forward binding rate constant	1.667 x 10^9 cm^3/mole-s
k_{off}	Complex dissociation rate constant	0.02 s^{-1}
k_e	Complex endocytosis rate constant	0.02 s^{-1}
R_{cell}	Receptor cell surface density (cell area ~25 μm^2)	40–4000 rec/μm^2
h	Height of the extracellular medium	0.5 μm
C_T	Threshold complex concentration for Rho induction	
τ	Time scale for Rho degradation	~20 min (49)
$I_{\|i-m\|,\|j-n\|}$	Cell–cell coupling coefficient	

*Unless indicated otherwise, the references for the parameters can be found in these papers and the references therein.

3.1. Models of Ligand Transport and Binding

The mechanisms of EGFR-mediated patterning in *Drosophila* development depend on the spatial ranges of EGFR ligands. Spitz was identified as a short-ranged ligand acting over 3–4 cell diameters (31–33), while Gurken was proposed to act as a long-range morphogen that can act over more than 10 cell diameters (34–36). The spatial ranges of Gurken and Spitz were derived from observing their effects on the expression of EGFR-target genes. At this time, the mechanisms governing the differences in the apparent ranges of the ligand are not well understood. Since both molecules are secreted, their spatial range can be tuned by the rates of extracellular transport and ligand-receptor interaction. Given a large and rapidly growing amount of information about each of these processes in the EGFR system, it is reasonable to ask if the experimentally derived estimates of ligand range can be interpreted in terms of the elementary processes, such as binding and receptor-mediated endocytosis. The ability to predict and manipulate the spatial ranges of secreted growth factors can be used to develop computational models of patterning networks and design new experiments for evaluating proposed mechanisms. In the following, we use a simplified geometry of cell–cell communication to illustrate a mechanistic model of ligand transport (Figure 6A).

In the model, the ligand diffuses between the receptor-covered and reflecting surfaces. This geometry approximates the one in egg development where EGFR ligands diffuse in the thin gap between the oocyte and the follicular epithelium. The motion of the secreted ligand is modeled by free diffusion with an

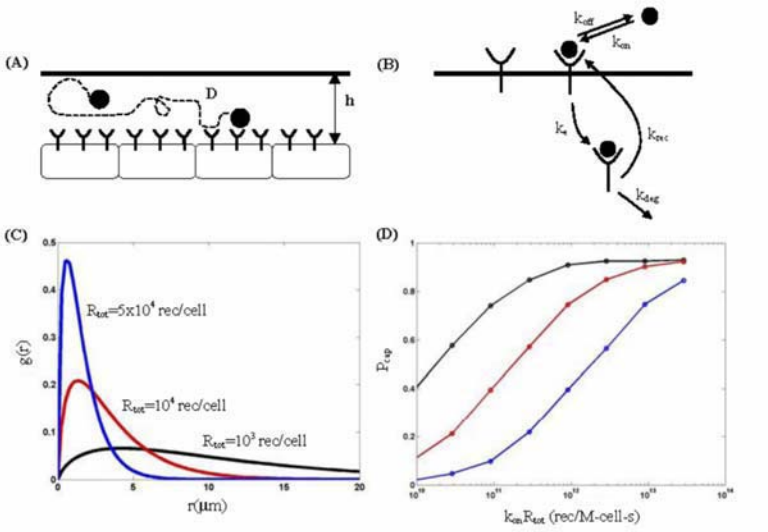


Figure 6. (**A**) A simplified model of ligand transport, binding, and trafficking. Ligand diffuses in the gap between the reflective and receptor-covered surfaces. Receptor density is uniform across the surface of the epithelial layer. (**B**) Ligand-receptor interactions (see Table 1 for definition of parameters). (**C**) Probability density function for the lateral distances traveled by secreted ligands in the time between the binding events. $g(r)dr$ is equal to the probability that a ligand will be bound between r and $r + dr$ (see Eq. [2]). All computations are performed on a hexagonal cell with an area of 25 μm^2. (**D**) Fraction of the ligands that are recaptured by the ligand-releasing cell plotted as a function of the cell surface receptor number (R_{tot}), ligand-receptor affinity (k_{on}), and extracellular ligand diffusivity (D). The curves, from top to bottom, correspond to $D = 10^{-9}$ cm^2/s, $D = 10^{-8}$ cm^2/s, and $D = 10^{-7}$ cm^2/s.

effective diffusion coefficient D. The diffusion coefficient can vary between the low values of growth factor diffusion in extracellular matrices (10^{-10} cm^2/s) (37) and the typical values for protein diffusion in an aqueous solution (10^{-6} cm^2/s) (38). We assume that the number of receptors per cell, R_{tot}, is constant. As illustrated in Figure 6B, ligand-receptor interactions are characterized by kinetic rate constants k_{on} and k_{off}; the endocytosis of receptor-bound ligands is modeled as a first-order process with rate constant k_e; we assume that internalized ligand is not recycled. The last assumption is based on the observation that, for the mammalian TGFα, recycling is negligible (39). In the absence of measurements in the *Drosophila* EGFR system, the rate constants are approximated by their counterparts measured in mammalian systems (39,40). In the following, we show how this model can be used to quantify the distance traveled by a secreted ligand.

Analysis of the distance traveled by a ligand between the subsequent binding events requires solving the problem of ligand transport in the gap above the

receptor-covered plane. The statistical properties of the time and distance to the first binding event depend on the forward binding rate constant, k_{on}, the cell surface density of receptors, R_{cell}, the ligand diffusion coefficient in the extracellular matrix, D, and the height of the extracellular medium, h. For the relevant ranges of these parameters, the distribution of time to capture is given by an exponentially distributed random variable: The probability that a ligand is bound for the first time after time t (P_{first}) is given by

$$P_{\text{first}}(t) = \exp\left(-\frac{k_{\text{on}}R_{\text{cell}}}{h}t\right). \quad [1]$$

This expression leads to the probability density function for the lateral distance traveled until the first binding event (Figure 6C): The probability that a ligand is bound in the ring between radii r and $r + dr$ is given by

$$g(r)dr = \left(\frac{k_{\text{on}}R_{\text{cell}}}{Dh}\right)K_0\left(r\sqrt{\frac{k_{\text{on}}R_{\text{cell}}}{Dh}}\right)rdr\,, \quad [2]$$

where K_0 is the modified Bessel function (41). We see that for reasonable ranges of receptor densities and binding rate constants, a majority of the secreted ligands is bound for the first time after traveling a very short distance. As an immediate consequence, the ligand-producing cell can recapture a significant fraction of the secreted ligands (Figure 6D).

Once secreted, a ligand undergoes several cycles of binding, dissociation, and extracellular diffusion before it is removed from the extracellular medium by receptor-mediated endocytosis (Figure 6B). The rate constants for dissociation and endocytosis of ligand-receptor complexes, k_{off} and k_e, determine the number of binding events until the first endocytosis event. In the simplest model, the number of binding events is a geometrically distributed random variable with the mean equal to $1 + (k_{off}/k_e)$ (25). Based on the measurements for the human EGFR-TGFα interactions ($k_{off} \cong k_e \approx 0.1\ \text{min}^{-1}$) (40), the ligand will be internalized after ~2 binding events. Thus, this simple model predicts that inhibition of receptor-mediated endocytosis can extend the range of secreted ligands. This prediction can be tested in mutants with defects in the genes mediating receptor-mediated endocytosis (42,43). (In other systems, where ligand is efficiently recycled to the cell surface, the opposite can be true. In fact, endocytosis might actually be the main mechanism for the spatial propagation of secreted signals (44).)

According to this simple model, the relative rates between binding interactions, endocytosis, and diffusion influence the spatial range of secreted ligands. This might explain the apparent differences in the spatial ranges of Gurken and Spitz. An experimental test of this explanation requires measurement of the relevant rate constants of Gurken and Spitz.

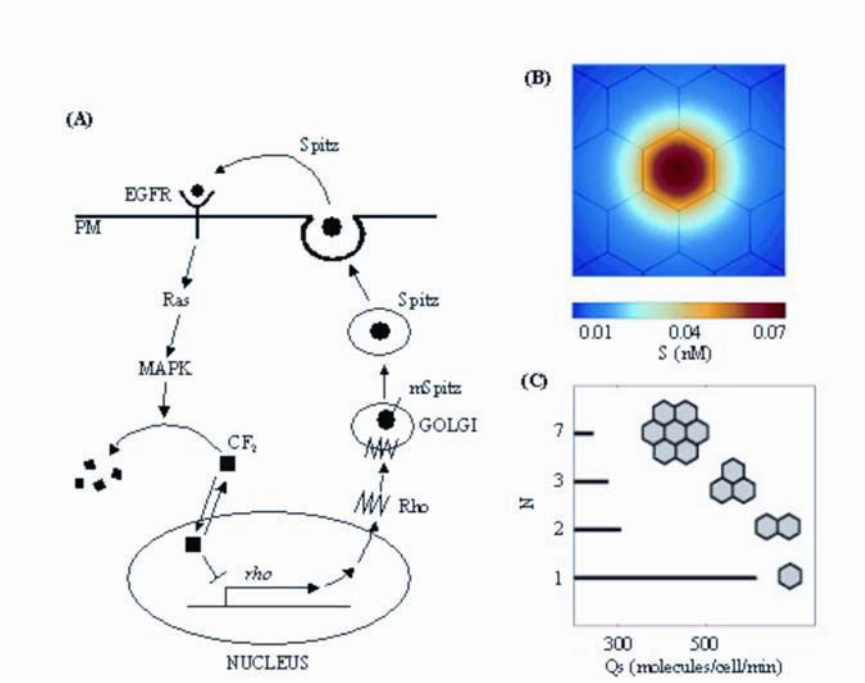


Figure 7. (**A**) A tentative structure of a positive feedback loop in the Rhomboid/Spitz system. Ligand binding stimulates ligand release. Receptor activation leads to activation of the canonical Ras/MAPK signal transduction pathway. The MAPK activity leads to degradation of CF_2, a transcription factor that inhibits transcription of the ligand-releasing protease, *rhomboid* (*rho*). In the absence of CF_2 inhibition, *rhomboid* is synthesized. The mature Rhomboid (Rho) protein cleaves the transmembrane Spitz (mSpitz) in the Golgi into its secreted form (PM = plasma membrane). (**B**) The steady-state ligand field due to a single ligand-releasing cell. Parameters: $h = 0.5$ μm, $k_c = 0.1$ min^{-1}, $k_{off} = 0.1$ min^{-1}, $R_{cell} = 1 \times 10^4$ receptors/cell surface, $D = 1 \times 10^{-7}$ cm^2/s, $k_{on} = 0.1$ Nm^{-1}min^{-1}, maximal rate of ligand release $Q_s = 100$ molecules/cell-min, cell area = 25 μm^2. (**C**) A small cluster of cells with constitutively active Rhomboid expression can generate an expanding wave of Rhomboid induction. The critical value of ligand release necessary to generate the wave is plotted as a function of the number of cells within the cluster (N). The generation function for Rhomboid was approximated by a Heaviside function, such that Rhomboid expression can be only in two states, "on" and "off." See Pribyl et al. (25) for the detailed definition of model parameters and its computational analysis.

3.2. Positive Feedback by Rhomboid and Spitz

The Rhomboid/Spitz module amplifies the oocyte-derived Gurken signal in eggshell patterning (11,45,46). In the emerging picture, EGFR-activated Ras/MAPK pathway relieves the transcriptional repression of *rhomboid* (47,48). Rhomboid then stimulates the secretion of Spitz that binds to EGFR on the ligand-producing cells and their neighbors (Figure 7A). This information about signaling in a single cell can be combined with the transport model from the

previous section to analyze the operation of the Rhomboid/Spitz feedback in a multicellular system such as an epithelial layer. The resulting description is useful in analyzing the effects of exogenous signals presented to the epithelial layer. For example, in the case of oogenesis, it is important to characterize the domain affected by Gurken and Gurken-induced EGFR ligands (4,11,45).

To understand the operation of the Rhomboid/Spitz circuit we started to develop models of autocrine signaling in epithelial layers (25). In addition to ligand transport, binding, and internalization, these models account for Rhomboid induction and Rhomboid-mediated Spitz release. Rhomboid induction was modeled as a threshold-like function, σ, of the total number of ligand-receptor complexes on the cell surface. The balance for the level of Rhomboid in the cell (i,j), $P_{i,j}$, takes the following form:

$$\frac{dP_{i,j}}{dt} = -\frac{P_{i,j}}{\tau} + \sigma\left(C_{i,j}^{\text{tot}} - C_T\right), \qquad [3]$$

where $C_{i,j}^{\text{tot}}$ is the total number of occupied EGF receptors in the cell (i,j) and C_T is the threshold-value for Rhomboid induction. The time-scale for Rhomboid degradation, $\tau \sim 20$ minutes, can be estimated from the experiments in the embryo (49).

Receptor occupancy on any given cell within the epithelial layer depends on the pattern of ligand release, and hence the pattern of Rhomboid expression in the entire layer. Our analysis suggests that ligand binding and transport rapidly adjust to the much slower dynamics of Rhomboid expression. In other words, the equations for ligand binding and transport reach the steady state dictated by the pattern of Rhomboid across the epithelial layer. In the ligand-limited regime, receptor occupancy for a given cell is computed from the linear superposition of ligand fields due to individual cells (Figure 7B). As a result, the dynamics of cells coupled by secreted signals can be described entirely in terms of the intracellular variables:

$$\frac{dP_{i,j}}{dt} = -\frac{P_{i,j}}{\tau} + \sigma\left(\sum_{m,n} I_{|i-m|,|j-n|} P_{m,n} - C_T\right), \qquad [4]$$

where $I_{|i-m|,|j-n|}$ is the cell–cell coupling coefficient that quantifies the strength and the spatial range of autocrine and paracrine signals. Importantly, these coefficients were derived as a function of the biophysical parameters of the problem, such as the diffusion and binding rates, as well as the rates and levels of ligand release by single cells within the layer. We found that the coupling coefficients decay rapidly as a function of cell–cell distance. This suggests that only a small number of cell–cell interactions must be taken into account in calculating recep-

tor occupancy on any given cell, a fact that is very useful in solving the problem numerically.

This biophysical framework can be used to predict the possible effects of localized perturbations of epithelial layers. For example, Peri et al. (35) constitutively activated Rhomboid in a small group of cells within the follicular epithelium. The effect of this perturbation was localized to its neighborhood: the EGFR-target genes were affected a few cell diameters from the cluster with the constitutively active ligand release. What is the outcome of such perturbations in general? Under what conditions will they remain localized or, alternatively, generate a propagating wave where secreted Spitz will be inducing Rhomboid expression and further Spitz release from the neighboring cells? This question can be easily addressed with the described model. For example, Figure 7C shows how the transition between the stationary and propagating patterns is affected by the size of the perturbation and the rate of ligand release. Clearly, a high rate of ligand release and a large size of perturbation promote generation of traveling waves. Because of its potential for "runaway" behavior, a positive feedback is tightly regulated. Genetic studies in the ovary indicate that the domain of the positive feedback is restricted in space, presumably to prevent propagation of traveling waves (24).

3.3. Pattern Formation by Interacting Feedback Loops

Dorsal appendage morphogenesis provides a genetically tractable system for studying the mechanisms by which simple inductive cues are converted into more elaborate spatial patterns. The components of the mechanism proposed by Wasserman and Freeman are well established. But, is the proposed mechanism actually correct? Specifically, does it account for the phenotypes that are induced by various genetic manipulations of the DER network and can it make testable predictions? These questions led us to develop our initial phenomenological model of EGFR-mediated patterning in *Drosophila* oogenesis (26,30). The model accounts for the interactions between the spatially nonuniform input by Gurken and the feedback loops by Spitz and Argos (Figure 8A). We formulated the model in one spatial dimension and assumed that the characteristic size of the pattern greatly exceeds the size of a single cell. This led to a system of nonlinear reaction-diffusion equations that was analyzed by time integration and numerical bifurcation analysis.

Our main goal was to test whether the mechanism could account for the various eggshell morphology phenotypes. We were particularly interested in the phenotypes generated by manipulations of the dose and the spatial distribution of the oocyte-derived signal (see Nilson and Schupbach (20) for a comprehensive review). It is known that a systematic decrease in the level of Gurken signal

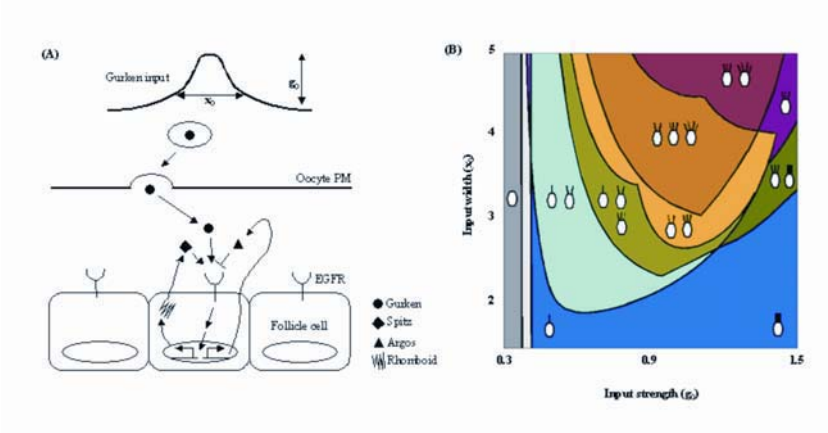


Figure 8. (**A**) Input and feedback loops in the model of pattern formation by peak splitting. (**B**) Summary of the results of computational analysis of the one-dimensional model of pattern formation by *Drosophila* EGFR autocrine feedback loops. The regions of existence of different stationary patterns as a function of the width (x_0) and the amplitude (g_0) of the input (Gurken) signal. Patterns with different numbers of peaks are associated with the eggshells with different numbers of dorsal appendages (shown by insets). See Shvartsman et al. (30) for a detailed definition of model parameters and its computational analysis.

can generate eggshells with one or zero dorsal appendages. At the same time, an increase in the dose leads to eggs with increased inter-appendage distance or one broad dorsal appendage. These observations provide important constraints on the modeling.

Analysis of the phenomenological model shows that the peak-splitting mechanism can be realized in one spatial dimension (Figure 8B). This means that a single-peaked input in the model, mimicking the oocyte-derived Gurken, can generate a stable pattern with two large-amplitude peaks in the spatial distribution of Rhomboid. The two-peaked pattern emerges as a result of the *instability* of the one-peaked solution that is realized at lower inputs. At a critical input level, this single-peaked solution splits, giving rise to the blueprint for formation of two dorsal appendages. Thus, patterning leading to formation of dorsal appendages can be viewed as a transition between the two kinds of solutions in the model (i.e., one- and two-peaked).

The variations in the level and the spatial distribution of Gurken input can induce transitions between different classes of patterns that are characterized by the different number of large-amplitude peaks in the spatial distribution of Rhomboid. We correlate these patterns with the dorsal appendage phenotypes in mutants with either lower Gurken doses or with defects in EGFR signal transduction (Figure 8B) (20). Predicted transitions between the zero-, one-, and two-peaked patterns in the model correspond to the experimentally observed transi-

tions between eggshells with zero, one, and two dorsal appendages. Finally, stable patterns with three or four peaks emerge in the model when the two-peaked pattern is destabilized by the same mechanisms that generate the two-peaked pattern itself. Since the number of peaks in the pattern corresponds to the number of dorsal appendages, this finding provides the mechanistic basis for explaining complex morphologies in mutants of *Drosophila melanogaster* (50,51). In addition, this versatility in patterning may account for more complex eggshell morphologies in related fly species (52,53).

This phenomenological model did not explicitly account for the details of EGFR interaction with its ligands. Our main goal was to examine the pattern formation capability of the localized input (i.e., Gurken) modulated by a network of spatially distributed feedback loops (i.e., Argos and Spitz). As with any phenomenological modeling, our model tested the sufficiency of the proposed mechanism, but did not prove it at the genetic or biochemical level. A mechanistic approach to modeling of EGFR-mediated signaling in oogenesis is now possible, based on the biochemical analysis of EGFR/Argos/Spitz interactions (9).

4. CONCLUSIONS AND OUTLOOK

At this time, only a few dozen out of ~30,000 EGFR-related PubMed entries are dedicated to modeling and computational analysis of EGFR signaling. Most of the existing models are formulated at the molecular and cellular levels (5). However, to understand how this system operates in vivo we need modeling at the level of tissues. Even the simplest models of EGFR signaling in multicellular systems must simultaneously account for ligand release, transport, binding, intracellular signaling, and gene expression. Given this complexity, the integrated models are nontrivial to test experimentally. We believe that a combined modeling-experimental approach is possible in *Drosophila*, where a number of genetic tools are available for implementing the model-directed manipulations in vivo.

We have described two systems from *Drosophila* development where modeling seems both feasible and necessary. In both cases, a large amount of data was summarized in the form of a complex patterning mechanism. The feasibility of these mechanisms depends on the quantitative parameters, such as the spatial ranges of secreted signals and the strengths of the feedback loops. Modeling can be used to elucidate the quantitative constraints on the proposed patterning mechanisms and to dissect the relative contributions of multiple cellular processes. We are just beginning to develop and test mechanistic models of EGFR signaling in tissues. Currently, we rely on the large amount of biochemical and cellular experiments in mammalian systems. In the future, direct biophysical characterization of the *Drosophila* EGFR network will be required in order to develop truly mechanistic models of EGFR signaling in fruit fly development.

We hope that, in addition to being applicable to the mechanisms in *Drosophila* development, our models will be useful for the quantitative description of EGFR signaling in higher organisms. Indeed, the molecular components and the feedback loops in EGFR signaling are conserved across species. For example, the positive feedback, similar to the one in the Rhomboid/Spitz system, was identified in radiation responses of human autocrine carcinoma cells (54). There, a pulse of ionizing radiation induces a primary wave of EGFR activation that was then amplified by the positive feedback, which is based on the MAPK-mediated TGFα release and recapture by the cell. Central to this feedback is activation of the ligand-releasing protease (TACE) that serves as the rate-limiting component that controls ligand availability and, hence, receptor activation.

The negative feedback loop by Argos does not have a direct counterpart in mammalian EGFR systems. The closest mode of regulation, discovered by Maihle and colleagues (55), relies on a secreted form of EGFR. Secreted receptors compete with the ones on the cell surface for the extracellular ligands, and in this way control the level of cellular EGFR activation. This mode of regulation has been described for both the ErbB1 and ErbB3 receptors, indicating that it is a general mechanism in the ErbB receptor family. Several lines of evidence from the same study support the physiological significance of this negative mode of control. For example, the levels of secreted receptors can be used as diagnostic markers in ovarian epithelial cancer (56). In the future, it will be important to investigate whether this mode of regulation also contributes in the developmental context.

5. ACKNOWLEDGMENTS

The authors are indebted to Trudi Schupbach and Lazaros Batsilas for their critical reading of the manuscript. SYS thanks Sasha Berezhkovskii, Doug Lauffenburger, Mark Lemmon, Cyrill Muratov, Gregory Reeves, Michal Pribyl, and Steve Wiley for numerous helpful discussions. This work was supported by the grants from the NSF and the Searle Foundation.

6. NOTES

1. In accordance to the convention, the names of genes are italicized (i.e., *gene*) while the names of proteins are written with the first letters capitalized (i.e., Protein).

2. Vein has been shown to form a positive feedback loop during the patterning of ventral ectoderm (10). However, its contribution is redundant and only important when the level of Spitz is reduced.

7. REFERENCES

1. Jorissen RN, Walker F, Pouliot N, Garrett TP, Ward CW, Burgess AW. 2003. Epidermal growth factor receptor: mechanisms of activation and signaling. *Exp Cell Res* **284**(4):31–53.
2. Yarden Y, Sliwkowski MX. 2001. Untangling the ErbB signalling network. *Nature Rev Mol Cell Biol* **2**:127–137.
3. Pierce KL, Tohgo A, Ahn S, Field ME, Luttrell LM, Lefkowitz RJ. 2001. Epidermal growth factor (EGF) receptor-dependent ERK activation by G protein-coupled receptors: a co-culture system for identifying intermediates upstream and downstream of heparin-binding EGF shedding. *J Biol Chem* **276**(25):23155–23160.
4. Vermeer PD, Einwalter LA, Moninger TO, Rokhlina T, Kern JA, Zabner J, Welsh MJ. 2003. Segregation of receptor and ligand regulates activation of epithelial growth factor receptor. *Nature* **422**(6929):322–326.
5. Wiley HS, Shvartsman SY, Lauffenburger DA. 2003. Computational modeling of the EGF receptor system: a paradigm for systems biology. *Trends Cell Biol* **13**(1):43–50.
6. Lander AD, Nie W, Wan FY. 2002. Do morphogen gradients arise by diffusion? *Dev Cell* **2**(6):785–796.
7. Shilo BZ. 2003. Signaling by the *Drosophila* epidermal growth factor receptor pathway during development. *Exp Cell Res* **284**(1):140–149.
8. Casci T, Freeman M. 1999. Control of EGF receptor signalling: lessons from fruitflies. *Cancer Metastasis Rev* **18**(2):181–201.
9. Klein DE, Nappi VM, Reeves GT, Shvartsman SY, Lemmon MA. 2004. Argos inhibits epidermal growth factor receptor signaling by ligand sequestration. *Nature* **430**(730):1040–1044.
10. Golembo M, Yarnitzky T, Volk T, Shilo BZ. 1999. Vein expression is induced by the EGF receptor pathway to provide a positive feedback loop in patterning the *Drosophila* embryonic ventral ectoderm. *Genes Dev* **13**(2):158–162.
11. Wasserman JD, Freeman M. 1998. An autoregulatory cascade of EGF receptor signaling patterns the *Drosophila* egg. *Cell* **95**(3):355–364.
12. Schweitzer R, Shilo BZ. 1997. A thousand and one roles for the *Drosophila* EGF receptor. *Trends Genet* **13**(5):191–196.
13. Golembo M, Raz E, Shilo BZ. 1997. The *Drosophila* embryonic midline is the site of Spitz processing, and induces activation of the EGF receptor in the ventral ectoderm. *Development* **122**(11):3363–3370.
14. Golembo M, Schweitzer R, Freeman M, Shilo BZ. 1996. Argos transcription is induced by the *Drosophila* EGF receptor pathway to form an inhibitory feedback loop. *Development* **122**(1):223–230.
15. Nambu JR, Franks RG, Hu S, Crews ST. 1990. The single-minded gene of *Drosophila* is required for the expression of genes important for the development of CNS midline cells. *Cell* **63**(1):63–57.
16. Lee JR, Urban S, Garvey CF, Freeman M. 2001. Regulated intracellular ligand transport and proteolysis control EGF signal activation in *Drosophila*. *Cell* **107**(2):161–171.
17. Dobens LL, Raftery LA. 2000. Integration of epithelial patterning and morphogenesis in *Drosophila* oogenesis. *Dev Dyn* **218**(1):80–93.
18. Spradling, AC. 1993. Developmental genetics of oogenesis. In *The development of drosophila melanogaster*, pp. 1–70. Ed. M Bate, AM Arias. Cold Spring Harbor Laboratory Press, Plainview, New York.
19. Queenan AM, Ghabrial A, Schupbach T. 1997. Ectopic activation of torpedo/Egfr, a *Drosophila* receptor tyrosine kinase, dorsalizes both the eggshell and the embryo. *Development* **124**(19):3871–3880.

20. Nilson LA, Schupbach T. 1999. EGF receptor signaling in *Drosophila* oogenesis. *Curr Top Dev Biol* **44**:203–243.
21. Neuman-Silberberg FS, Schupbach T. 1993. The *Drosophila* dorsoventral patterning gene *gurken* produces a dorsally localized RNA and encodes a TGF alpha-like protein. *Cell* **75**(1):165–174.
22. Montell DJ, Keshishian H, Spradling AC. 1993. Laser ablation studies of the role of the *Drosophila* oocyte nucleus in pattern formation. *Science* **254**(5029):290–293.
23. Guichet A, Peri F, Roth S. 2001. Stable anterior anchoring of the oocyte nucleus is required to establish dorsoventral polarity of the *Drosophila* egg. *Dev Biol* **237**(1):93–106.
24. Peri F, Roth S. 2000. Combined activities of Gurken and Decapentaplegic specify dorsal chorion structures of the *Drosophila* egg. *Development* **127**(4):841–850.
25. Pribyl M, Muratov CB, Shvartsman SY. 2003. Discrete models of autocrine signaling in epithelial layers. *Biophys J* **83**(6):3659–65.
26. Pribyl M, Muratov CB, Shvartsman SY. 2003. Transitions in the model of epithelial patterning. *Dev Dyn* **226**(1):155–159.
27. Pribyl M, Muratov CB, Shvartsman SY. 2003. Long-range signal transmission in autocrine relays. *Biophys J* **84**(2):883–896.
28. Shvartsman SY, Wiley HS, Deen WM, Lauffenburger DA. 2001. Spatial range of autocrine signaling: modeling and computational analysis. *Biophys J* **81**(4):1854–1867.
29. Shvartsman SY, Hagan MP, Yacoub A, Dent P, Wiley HS, Lauffenburger DA. 2002. Autocrine loops with positive feedback enable context-dependent cell signaling. *Am J Physiol Cell Physiol* **282**(3):C545–C559.
30. Shvartsman SY, Muratov CB, Lauffenburger DA. 2002. Modeling and computational analysis of EGF receptor-mediated cell communication in *Drosophila* oogenesis. *Development* **129**(11):2577–2589.
31. Bergmann A, Tugentman M, Shilo H. 2002. Regulation of cell number by MAPK-dependent control of apoptosis: a mechanism for trophic survival signaling. *Dev Cell* **2**(2):159–170.
32. Schweitzer R, Shaharabany M, Seger R, Shilo BZ. 1995. Secreted Spitz triggers the DER signaling pathway and is a limiting component in embryonic ventral ectoderm determination. *Genes Dev* **9**(12):1518–1529.
33. Freeman M. 1997. Cell determination strategies in the *Drosophila* eye. *Development* **124**(2):261–270.
34. Pai LM, Barcelo G, Schupbach T. 2000. D-cbl, a negative regulator of the Egfr pathway, is required for dorsoventral patterning in Drosophila oogenesis. *Cell* **103**(1):51–61.
35. Peri F, Technau M, Roth S. 2002. Mechanisms of Gurken-dependent *pipe* regulation and the robustness of dorsoventral patterning in *Drosophila*. *Development* **129**(12):2965–2975.
36. James KE, Dorman JB, Berg CA. 2002. Mosaic analyses reveal the function of *Drosophila* Ras in embryonic dorsoventral patterning and dorsal follicle cell morphogenesis. *Development* **129**(9):209–222.
37. Dowd CJ, Cooney CL, Nugent MA. 1999. Heparan sulfate mediates bFGF transport through basement membrane by diffusion with rapid reversible binding. *J Bio Chem* **274**(8):5236–5244.
38. Berg HC. 1983. *Random walks in biology*. Princeton University Press, Princeton.
39. Wiley HS. 2003. Trafficking of the ErbB receptors and its influence on signaling. *Exp Cell Res* **284**(1):78–88.
40. Wiley HS, Burke PM. 2001. Regulation of receptor tyrosine kinase signaling by endocytic trafficking. *Traffic* **2**(1):12–18.
41. Abramowitz M, Stegun IA. 1964. *Handbook of mathematical tables with formulas, graphs, and mathematical tables*. Government Printing Office, Washington, DC.
42. Seto ES, Bellen HJ, Lloyd TE. 2002. When cell biology meets development: endocytic regulation of signaling pathways. *Genes Dev*. **16**(11):1314–1336.

43. Lloyd TE, Atkinson R, Wu MN, Zhou Y, Pennetta G, Bellen HJ. 2002. Hrs regulates endosome membrane invagination and tyrosine kinase receptor signaling in *Drosophila*. *Cell* **108**(2):261–269.
44. Entchev EV, Gonzalez-Gaitan MA. 2002. Morphogen gradient formation and vesicular trafficking. *Traffic* **3**(2):98–109.
45. Peri F, Bokel C, Roth S. 1999. Local *gurken* signaling and dynamic MAPK activation during *Drosophila* oogenesis. *Mech Dev* **81**(1–2):75–88.
46. Sapir A, Schweitzer R, Shilo BZ. 1998. Sequential activation of the EGF receptor pathway during *Drosophila* oogenesis establishes the dorsoventral axis. *Development* **125**(2):191–200.
47. Hsu T, McRackan D, Vincent TS, Gert de Couet H. 2001. *Drosophila* Pin1 prolyl isomerase Dodo is a MAP kinase signal responder during oogenesis. *Nature Cell Biol* **3**(6):538–543.
48. Mantrova EY, Hsu T. 1998. Down-regulation of transcription factor CF2 by Drosophila Ras/MAP kinase signaling in oogenesis: cytoplasmic retention and degradation. *Genes Dev* **12**(8):1166–1175.
49. Sturtevant MA, Roark M, O'Neill JW, Biehs B, Colley N, Bier E. 1996. The *Drosophila* rhomboid protein is concentrated in patches at the apical cell surface. *Dev Biol* **174**(2):298–309.
50. Reich A, Sapir A, Shilo BZ. 1999. Sprouty, a general inhibitor of receptor tyrosine kinase signaling. *Development* **126**(18):413–447.
51. Deng WM, Bownes M. 1997. Two signalling pathways specify localised expression of the Broad-Complex in *Drosophila* eggshell patterning and morphogenesis. *Development* **124**(22):4639–4647.
52. Perrimon N, Duffy, JD. 1998. Developmental biology: sending all the right signals. *Nature* **396**(6706):18–19.
53. Hinton HE. 1981, *Biology of insect eggs*, Vol. 1. Pergamon Press, New York.
54. Dent P, Reardon D, Park J, Bowers G, Logsdon C, Valerie K, Schmidt-Ullrich R. 1999. Radiation-induced release of transforming growth factor-α activates the epidermal growth factor receptor and mitogen-activated protein kinase pathway in carcinoma cells, leading to increased proliferation and protection from radiation-induced cell death. *Mol Biol Cell* **10**:2493–2506.
55. Lee H, R. Akita W, Sliwkowski MX, Maihle NJ. 2001. A naturally occurring secreted human ErbB3 receptor isoform inhibits heregulin-stimulated activation of ErbB2, ErbB3, and ErbB4. *Cancer Res* **61**(11):4467–4473.
56. Baron AT, Cora EM, Lafky JM, Boardman CH, Buenafe MC, Rademaker A, Liu D, Fishman DA, Podratz KC, Maihle NJ. 2003. Soluble epidermal growth factor receptor (sEGFR/sErbB1) as a potential risk, screening, and diagnostic serum biomarker of epithelial ovarian cancer. *Cancer Epidemiol Biomarkers Prev* **12**(2):103–13.

3.1

DEVELOPMENTAL BIOLOGY: BRANCHING MORPHOGENESIS

S.R. Lubkin

Department of Mathematics,
North Carolina State University, Raleigh

Branching morphogenesis is a ubiquitous system in the developmental biology of macroscopic organisms. Many of the details are known, yet a unified understanding remains out of reach. Many of the relevant facts about branching morphogenesis become clearer if we include a mechanical interpretation of the interactions between tissues.

1. INTRODUCTION

Lewis Wolpert (63) has often said that "gastrulation is the most important event in your life," but it is not the most frequent phenomenon in the development of an organism. Gastrulation occurs just once, but branching morphogenesis happens early and often. The formation of branched tubular structures–glands–occurs throughout an organism, in many different tissues, and is essential to the existence of virtually all organisms which need to transport fluids more than about a millimeter.

If you are macroscopic, you need ducts. How do they form?

Branching morphogenesis is widespread in animal development, generating the form of such diverse organs as the lung, pancreas, mammary gland, salivary

Address correspondence to: S.R. Lubkin, Department of Mathematics, Box 8203, North Carolina State University, Raleigh, NC 27695-8205 (lubkin@eos.ncsu.edu); North Carolina State University, Raleigh, NC 27695-8205.

gland, and kidney. Although nearly every technique of cellular and developmental biology has been applied to it (see reviews (3,11)), and a vast amount has been learned, there is still some uncertainty as to the mechanism of branching.

When I pose the question "What makes an airplane fly?" most people answer with some variation on the theme of lift, i.e., the net force generated by the balance of flows of air around the object. Rarely does anyone answer that it is the pilot manipulating the vast number of controls in the cockpit who makes the airplane fly—though that is an equally correct answer.

When I pose the question "What makes a developing organ branch?" most biologists answer with a discussion of the switches, not of the forces. Yet just as we could not understand the airplane without understanding the physical forces that it generates, and its physical interaction with its environment, we cannot understand branching morphogenesis without understanding the physical forces that its tissues generate, and the organ's physical interactions with its environment.

The next logical step in morphogenesis research is study of the biophysical and biomechanical aspects, which are what create and modify form (53). This chapter compares the biomechanical aspects of the currently competing theories of branching morphogenesis, and suggests new experiments and new interpretations of old experiments.

It is assumed, though not proven, that the mechanism of morphogenesis is the same in all the branched organ systems, but that it is differences in gene expression and protein/polysaccharide/proteoglycan production that cause the morphological differences. Although the developmental biology of all of these organs has been widely studied, because of the assumed unity of mechanism, one organ, the rodent submandibular (salivary) gland, has been studied in the greatest detail.

The general features of branching morphogenesis in vivo are as follows. Glandular organs are constructed initially as a disorganized mass of mesenchymal tissue surrounding a finger of epithelium, which has a simple unbranched shape. Then the finger flattens slightly, and is split into two or more lobules by the formation of one or more clefts. These new lobules grow as the extracellular matrix (ECM) around the clefts gets denser and mesenchymal cells condense around the clefts and the stalks of the lobules (Figure 1). When the young lobules have grown sufficiently larger, there is another branching, in the same fashion, followed by extension and condensation, cleft-grow-cleft-grow, until a highly branched structure has formed.

The morphological differences between mature glands are substantial, but differences can be seen at the stage of clefting. For example, the lung branches only dichotomously, with rounded clefts, whereas the submandibular gland forms multiple very sharp clefts in each branching lobule. The lung also forms its dichotomous clefts in nearly orthogonal directions in successive branching generations. A mature gland is highly structured (62), and can be thought of as

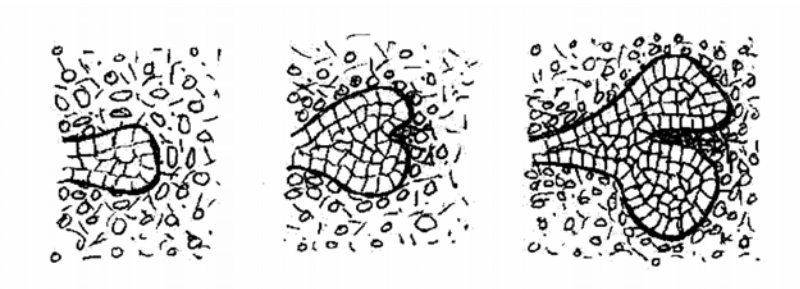


Figure 1. Sketch of the general sequence of events in a single branching event. As cleft deepens, mesenchyme in and near it becomes denser in cells and ECM materials. New branches grow, flatten, and also branch. Reprinted with permission from Lubkin and Li (2002) (37).

having fractal scaling (5). There are many interesting models that have been written of the formation of an entire branched tree (27,38). When modeling the formation of an entire branched structure, the major concern is elucidating the mechanism while correctly reproducing the fractal scaling (compare with Part II, chapter 3, by Savage and West, this volume).

For this chapter, let us focus on the single step of clefting and growth of one lobule, under the assumption that repetition with eventual halting leads to a branched structure. Our major concern is elucidating the mechanism while correctly reproducing the length and time scales and geometry observed in the real system. The fractal scaling does not matter at the scale of a single clefting. However, the focus can be made even tighter than that. To create a whole branched tree, growth is of course essential for the volume contribution alone. But while branching requires cell growth, clefting does not (42,54). Therefore in the model presented in this chapter, in the interests of simplicity and robustness, we omit growth and focus solely on clefting.

Individual glandular rudiments grow and branch normally in vitro, making mechanical, chemical, and radiative manipulation relatively easy. It has been known for decades that glandular epithelia separated from their mesenchymes form normal organs when recombined with mesenchyme of the same organ type and age, but branch abnormally or not at all when recombined with mesenchyme from other organs (1,16,17,31–33,55). Whether the directive role of the mesenchyme is mechanical, chemical, or both has been the subject of a great deal of investigation. Most dramatically, lung and salivary epithelia have been cultured without any mesenchyme at all, embedded in gels with appropriate growth factors, and the epithelia form branched tubules without mechanical or chemical support from mesenchyme (46,48).

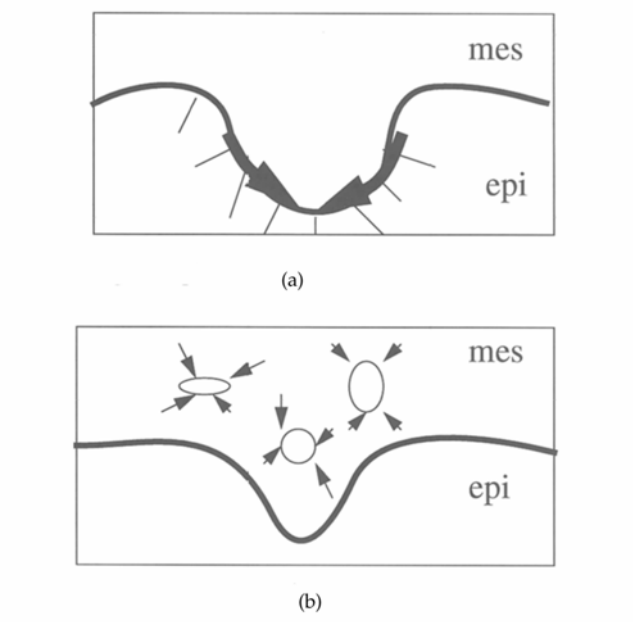


Figure 2. (a) Epithelial theory; (b) mesenchymal theory.

2. PREVIOUS WORK

There are two major theories of branching morphogenesis, each supposing that the force is from a different tissue.

1. The *epithelial theory* (7,21,56) hypothesizes a morphogenetic force at the basal end of the epithelium (Figure 2a), where it is supposed that microfilaments contract in the cleft region, in response to an external signal. The external signal is believed to be linked to the observed high rate of turnover of the basal lamina (6) in the proto-lobules. Other similar epithelial phenomena may be governed by the epithelial ECM (30).

2. The *mesenchymal theory* (26,41,43,44) hypothesizes a morphogenetic force in the contractile behavior of fibroblasts in the mesenchyme (47), condensing the mesenchyme near the epithelium (Figure 2b). The stresses created by the cellular traction forces are believed to align collagen fibrils into thick cords, which when pulled taut by further traction, push deeply into an epithelial lobule, creating a cleft.

The strongest evidence in favor of the epithelial theory is that epithelia need no mechanical link to mesenchyme in order to form tubules (18,20,46,48,59,

64,65). The strongest evidence in favor of the mesenchymal theory is that a wide variety of cell types, including fibroblasts, generate traction forces on ECM, resulting in deformation and sometimes in pattern formation (22,23,47,57, 60,61). Also, collagenases inhibit clefting, while collagenase inhibitors enhance clefting (44). Both theories are plausible, yet it would seem that if complete removal of the mesenchyme does not prevent clefting, then the mesenchyme cannot be causing the clefting. I believe that, as in most biological systems, the reality here is more complicated and subtle than our preconceived notions. Can a model help us understand the subtleties better?

We developed a mathematical model of the mechanical forces and deformations of the tissues involved in morphogenesis (37). We can use it to answer some questions.

3. MODEL

3.1. Hypothesis

It is clear that epithelia can generate morphogenetic forces in the absence of a mechanical input from mesenchyme. These forces can generate clefts even if the mesenchyme is not around the epithelium. But can the same forces generate clefts if the epithelium is embedded in mesenchyme? *We hypothesize that the branching morphogenesis observed in the mesenchyme-free experiments is not mechanically equivalent to the branching morphogenesis observed in mechanically intact rudiments or in vivo*. In this chapter we show numerical experiments to illustrate the forces and deformations in the two experimental situations—with and without mesenchyme.

To understand the relationship between forces and deformations in a morphogenetic system, we must formulate a model of the mechanics of the tissues and their interactions. But what is the constitutive law for an embryonic tissue? If I pull on my skin and let go, it bounces back. A material like skin that responds to short-term forces with reversible deformations is exhibiting short-term elasticity. If I wear braces on my teeth for two years, then remove them, the teeth do not bounce back. A material like the jawbone, which responds to long-term forces with irreversible deformations, behaves in the long term as a viscous fluid.

Most biomaterials are actually somewhere between an elastic solid and a viscous fluid. A material with short-term elasticity and long-term viscosity is in the Maxwell class of viscoelastic fluids. Embryonic tissues will bounce back from a brief deformation, but the changes associated with development are permanent. If we are only interested in permanent deformations of branching rudiments, we may ignore the short-term elastic component and focus on long-term behavior. We therefore model the embryonic epithelium and mesenchyme as

Stokes fluids. There is substantial precedent for considering the morphogenetic behavior of embryonic tissues in terms of a fluid ((38,49,50) deal with a small subset of these, while (12,15,25) consider single cells behaving as fluids). There is excellent experimental calibration of the fluid model of embryonic mechanics from Malcolm Steinberg and his colleagues (13,14).

Our desire is to keep all aspects of the model "as simple as possible, but not moreso" (A. Einstein). For example, the epithelium may or may not initially be hollow. In the salivary gland, a lumen is created as the epithelium matures, but is not present when the branches are created. For simplicity, we ignore lumens. We model a branching rudiment as a uniform epithelium inside either a uniform mesenchyme or a uniform acellular fluid (of the consistency of water or a collagen gel). It has been shown that growth-suppressed salivary gland rudiments can still branch once, though they do not grow enough to generate subsequent branches (42). Again, in the interest of simplicity, we focus on the single step of the creation of one cleft in a branching rudiment, so we choose to model the tissue as not growing.

Finally, also in the interest of simplicity, we do not try to explicitly model the force that causes the deformation. Instead, we simply apply a localized surface force at several points on the epithelial surface, and/or modify the local surface tension. Since our goal is just to understand the relationship between force and deformation in branching morphogenesis, this artificial force will serve our purposes fine.

The geometry of our model is simple. We have an epithelium-shaped region of fluid surrounded by a second fluid representing either mesenchyme or culture medium. The Stokes equations apply in each fluid:

$$\nabla p = \mu \nabla^2 \mathbf{u}\,, \qquad [1]$$

$$\nabla \cdot \mathbf{u} = \mathbf{0}\,, \qquad [2]$$

where p is the pressure, $\mathbf{u}$ is the velocity vector, and μ is the viscosity, which we shall assume to be constant within each fluid.

On the interface between the fluids, there is a surface tension γ, which can in principle vary in space, and which will be the only agent driving shape changes in this model. We can write the boundary condition between the fluids in terms of the jumps $[^\circ]$ of two quantities across the interface:

$$[\,p - 2\mu(\nabla \mathbf{u} \cdot \mathbf{n}) \cdot \mathbf{n}] = -\gamma\kappa\,, \qquad [3]$$

$$[\mu(\nabla \mathbf{u} \cdot \mathbf{n})] \cdot \tau = -\frac{\partial \gamma}{\partial \tau}\,, \qquad [4]$$

where **n** and τ are the normal and tangent vectors at each point on the interface, and κ is the local mean curvature of the interface (e.g., (34, ch. 5). Equation [3] is known as the Laplace-Young condition, and eq. [4] represents Marangoni stresses.

The clefting force is modeled as a point force applied at selected points along the interface. Since surface tension can be considered to produce a net force in the normal direction, we represent the inward-directed point forces as localized reductions of the surface tension:

$$\gamma = \gamma_0 - f \sum_i \delta(s_i), \qquad [5]$$

where γ_0 is the uniform surface tension everywhere on the epithelial surface, f is the magnitude of the local point force density, and δ is the delta function localizing the force at points s_i. In the interest of simplicity, we keep the point forces pointing in the same directions as initially, regardless of the motion of the interface.

Because the forces of one rudiment on another in vivo or in vitro can generally be neglected if the rudiments are not very close, the outside fluid is modeled as rectangular, with periodic boundary conditions.

3.2. Nondimensionalization

If we write the internal and external viscosities as μ^- and μ^+, respectively, and characteristic length and time scales as L and T, the characteristic surface tension as γ_0, and the characteristic pressure as μ^-/T, we can define three nondimensional parameters:

$$\alpha \equiv \frac{\mu^+}{\mu^-}, \qquad [6]$$

$$\beta \equiv \frac{\gamma_0 T}{\mu^- L}, \qquad [7]$$

$$\varphi \equiv \frac{f}{\gamma_0 L}. \qquad [8]$$

We call α the viscosity ratio, β the nondimensional surface tension, and φ the nondimensional clefting force. These three nondimensional parameters are all that governs the behavior of the model. There are thus only a small number of numerical experiments needed to explore all the model behavior.

3.3. Parameter Estimates

All of the parameters in the model can be estimated from published data in the literature. For example, salivary lobules are of the order of 100 μm in diameter, and branching occurs approximately every 8 hours (7). These provide estimates of the length L and time T scales for nondimensionalization. Contractility is readily estimated or measured in cell-populated collagen gels, and can give us an estimate for the clefting force f. The quality of these estimates is variable, since most applicable experiments have not been performed with this model in mind. In one prominent counterexample, contractility of various dissociated mesenchymal cells, including those of lung and submandibular gland, was measured on artificial collagen gels (47), specifically to test the mesenchymal theory of branching morphogenesis. One important factor in estimating the clefting force is that the younger the tissue, the stronger the contractility (52). Since we are concerned in this chapter with embryonic tissues, we rely on the higher published estimates of contractility.

The size of a branching rudiment and time scale of branching morphogenesis are well established, as is the viscosity of water (Table 1). We assume that if the outer "fluid" is mesenchyme, then its viscosity is within an order of magnitude of the viscosity of the epithelium inside. If the outer fluid is a collagen gel rather than a tissue, its viscosity may have a very wide range, depending on the exact composition, including pH and most importantly water content. Unfortunately, no one has reported direct measurements of the viscosity of the embedding clots in the mechanically mesenchyme-free culture systems. In some cases, even the composition of the embedding medium is not specified well enough for us to estimate its viscosity. Hence depending on the particular mesenchyme-free experiment we are modeling, the appropriate external viscosity, μ^+, could range from that of water to that of mesenchyme (or even more viscous). However, in general, we assume that the viscosity of the collagen gel in the mesenchyme-free experiments is much smaller than the viscosity of mesenchyme.

Foty and colleagues (14) and Forgacs et al. (13) measured the viscosity and surface tension of several types of embryonic tissues, and found each tissue to be mechanically very similar, within an order of magnitude of the others. As far as we know, there are no measurements reported of the tissue viscosity or surface tension of the specific components of any embryonic branching tissue, but we will assume that their values lie within the narrow range of the five tissue types measured by Foty et al.

The basal lamina of a clefting epithelium affects the epithelio-mesenchymal interactions (2) and more rapidly turns over on the lobules than in the clefts (6). This implies that there may be a reduction in the surface tension of the epithelium in the lobules, caused by faster dissipation of stresses in the lobular lamina than in already-formed clefts. However the magnitude of any surface tension from the basal lamina has not been measured, so we choose to ignore the basal lamina mechanically.

Table 1. Expected Ranges of Parameter Values

Parameter	Symbol	Units	Range	Reference(s)
Time scale	T	s	10^4–10^5	Bernfield et al. 1984 (7)
Rudiment size	L	m	10^{-4}	Bernfield et al. 1984 (7)
Epithelial viscosity	μ^-	poise	10^4–10^6	Foty et al. 1994 (14), Forgacs et al. 1998 (13), Phillips et al. 1977, 1978 (49,50)
Mesenchymal viscosity	μ^+	poise	10^4–10^6	Forgacs et al. 1998 (13), Phillips et al. 1977, 1978 (49,50)
Viscosity of embedding gel	μ^+	poise	10^0–10^6	Nogawa & Takahashi 1991 (48), Barocas et al. 1995 (4)
Viscosity of water	μ^+	poise	10^{-3}	—
Epithelial surface tension	γ_0	N/m	10^{-3}–10^{-2}	Foty et al. 1994 (14), Forgacs et al. 1998 (13)
Clefting force	f	N	10^{-7}–10^{-6}	Rappaport 1977 (51), Kolodney & Wysolmerski 1992 (28)

The clefting force is difficult to estimate. We can take estimates for single cells in very different contexts (e.g., Rappaport's (51) cleaving oocytes or fibroblasts on collagen gel (28), and estimate that an epithelial region about to cleft could generate 10–100 times the force of a single oocyte or fibroblast. This gives an estimate for f of 10^{-7} to 10^{-5} N.

Hence the possible range of α is from 10^{-9} to 10^{1}, of β from 10^{-1} to 10^{3}, and of ϕ from 10^{-1} to 10^{2}.

3.4. Methods

The fluid equations were solved by a finite-difference method (35) specifically designed to simulate two fluids of different viscosities separated by an interface. Because our major interest is in examining the effects of the viscosity of the outside fluid (mesenchyme or growth medium), the equations were solved in two space dimensions.

We performed the following numerical experiments:

1. To test the effect of the viscosity ratio α on the time course of deformations, we deformed identical 3-lobed "rudiments" at the centers of the far ends of their lobes, with the same normal force, until each lobe was nearly cleft in two. One rudiment was embedded in a material (gel or mesenchyme) of viscos-

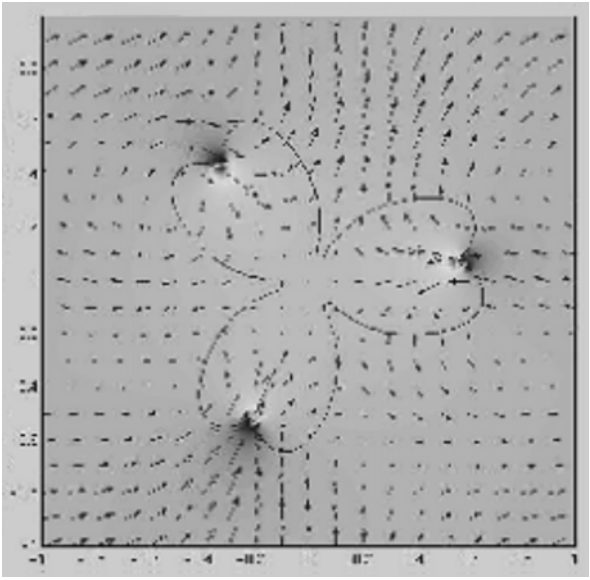


Figure 3. Pressure and flow in a typical simulation (equal viscosities). Velocity arrows are localized at the tail of each arrow. Note that pressure gradients are extremely localized. Reprinted with permission from Lubkin and Li (2002) (37).

ity equal to the epithelial viscosity; the other was embedded in a material (gel or liquid medium) of negligible viscosity; a third was embedded in a material more viscous than the epithelium.

2. To test the roles of clefting force and surface tension, initially 3-lobed "rudiments," embedded in mesenchyme or ECM of the same viscosity as the epithelium, were clefted with three identical inward-directed forces, for different values of α and ϕ.

3.5. Results

If no external force is applied to our branched rudiment, the preexisting clefts disappear. The time scale of this relaxation depends strongly on the surface tension and the viscosity ratio (35). High surface tension leads to faster cleft loss, and high viscosity ratio leads to slower cleft loss.

In all simulations, the pressure distribution was extremely uniform, except for a strong dipole near each point force (Figure 3).

In all our clefting experiments, our applied point forces were able to deform the 3-lobed rudiment into a 6-lobed rudiment, if the force was applied for long

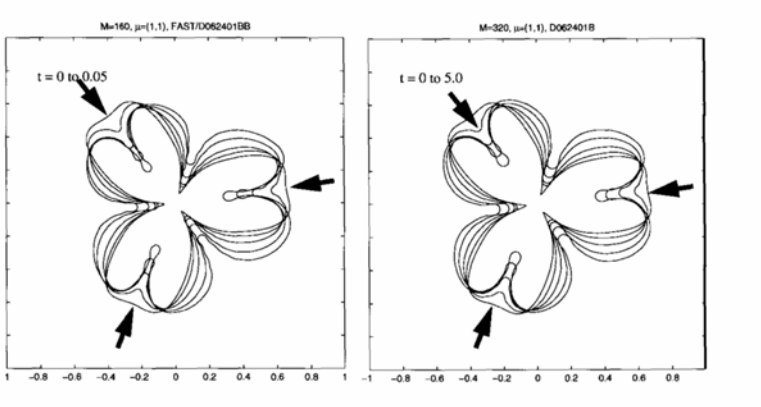


Figure 4. Effect of ϕ, relative force strength, on clefting of epithelial rudiment embedded in mesenchyme or gel of the same viscosity. Arrows indicate the directions of imposed forces, and also the direction of time sequence. Nondimensional surface tension $\beta = 0.005$, viscosity ratio $\alpha = 1$. (**a**) $\phi = 160$, $t/T = 0$ to 0.05; (**b**) $\phi = 50$, $t/T = 0$ to 5. A tripling of the relative clefting force ϕ divides the nondimensional clefting time by a factor of approximately 100. The weaker force leaves wider clefts (cf. Figure 6). Reprinted with permission from Lubkin and Li (2002) (37).

enough (Figures 4–6). The nondimensional surface tension β significantly affected the evolving shape. If β was very large, the preexisting clefts retreated before the new cleft was fully formed, regardless of the relative clefting force ϕ. For fixed surface tension β, the shape depended only subtly on the viscosity ratio and force parameters α and ϕ.

The most significant differences between simulations were in the time scales. When the nondimensional clefting force ϕ was small, it took significantly longer for clefts to form than at larger ϕ values. In particular, decreasing the clefting force by a factor of 3 increases the clefting time by a factor of about 100 (Figure 4).

Our most significant finding was that when the viscosity ratio α was high, it took significantly longer for clefts to form than at lower α values. In particular, increasing the viscosity of the mesenchyme/gel by a factor of 10 typically tripled the time to form a cleft of a characteristic depth (Figure 4). This relates directly to the question of what is going on in the mesenchyme-free experiments. Is the branching that occurs in a salivary epithelium the same when its mesenchyme is removed and replaced by a material which is much less viscous? We performed numerical experiments where for a fixed surface tension β, force ϕ, viscosity ratio α and experiment length t_{final}, a cleft formed; the same rudiment under identical conditions but whose mesenchyme/ECM was 10 times as viscous (α multiplied by 10) *failed* to form a visible cleft in the same time period (Figure 5).

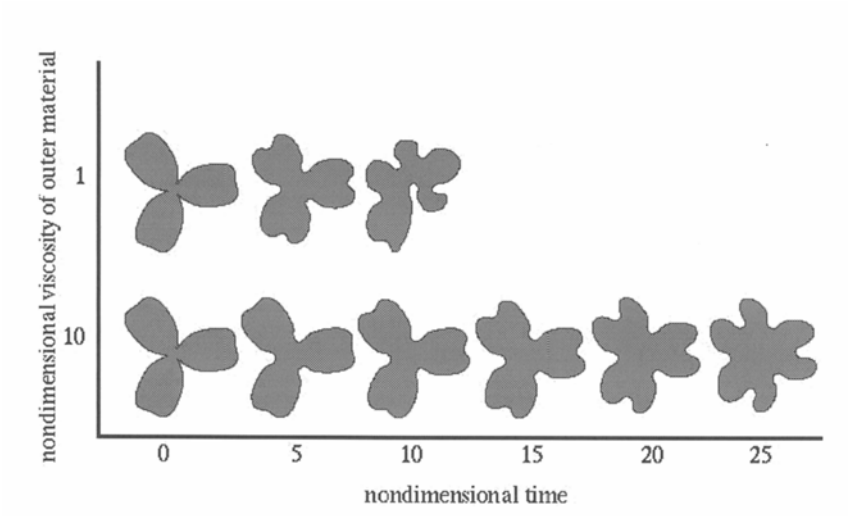


Figure 5. Effect of viscosity ratio α on evolution of the clefting epithelium. Nondimensional surface tension $\beta = 0.01$, nondimensional clefting force $\phi = 2.5$. Forces are in the same fixed directions. Upper: viscosity ratio $\alpha = 1$; lower: viscosity ratio $\alpha = 10$. Epithelium embedded in a material more viscous than itself takes longer to form the same depth cleft than if it were embedded in a material of the same viscosity. Reprinted witih permission from Lubkin and Li (2002) (37).

Viscosity ratio α had a significant effect not just on the time course of branching but, surprisingly, also on the width of the clefts that formed. Larger α (more viscous mesenchyme) led to narrower clefts (Figure 6).

4. DISCUSSION AND CONCLUSIONS

Because of the way the model [6]–[8] is scaled, the dynamics depend only on the three nondimensional ratios: $\alpha \equiv \mu^+ / \mu^-$, $\beta \equiv \gamma_0 T / \mu^- L$, and $\varphi \equiv f / \gamma_0 L$. Therefore, we can compare predicted deformations among different tissue types simply by noticing which ratios are preserved. For example, epithelial rudiments of different tissue viscosities but equal surface tensions and equal sizes, grown in a mesenchyme-free medium, should take different times to form the same size cleft from the same size force; the time should be proportional to the tissue viscosity μ^-. If they take the same time, it must be that the force or surface tension are also different. At an even simpler level, mesenchyme-free rudiments grown in gels identical in all respects except for the amount of collagen (hence viscosity and/or stiffness) should form clefts at dif-

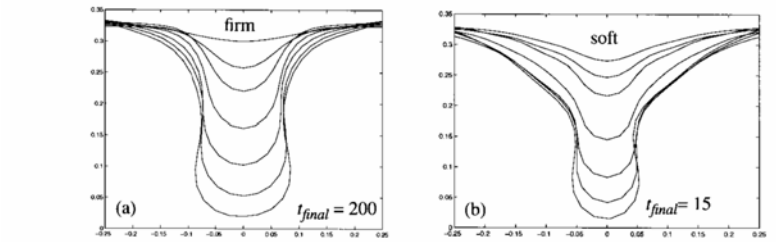


Figure 6. Magnification of a deepening cleft over time. Surface tension $\beta = 0.01$, nondimensional clefting force $\phi = 1.25$. Same initial conditions, same epithelial viscosity, same clefting force, same final depth of cleft. (**a**) Relatively firm mesenchyme. Viscosity ratio $\alpha = 100$, nondimensional time $t/T = 0$ to 200. (**b**) Relatively soft mesenchyme or ECM. Viscosity ratio $\alpha = 1$, nondimensional time $t/T = 0$ to 15. The epithelium embedded in a lower-viscosity material has taken less time to form its cleft, and the cleft is narrower than the epithelium embedded in a more viscous material. Reprinted with permission from Lubkin and Li (2002) (37).

ferent rates, and with slightly different morphology. No one has done that experiment.

Our simple model is able to make many of these predictions, and they are qualitatively robust. However, because the model is so simple, quantitative predictions are probably at best approximations. For instance, we assumed that each tissue was viscous, not viscoelastic, and had a spatially and temporally uniform viscosity, contrary to what is already known. In particular, we know that preexisting clefts have abundant collagen in them naturally making the clefts more resistant to deformation, and mesenchyme condenses around the branches, which should also make the lobe region stiffer and/or more viscous than the surrounding mesenchyme. We ignored the mechanical role of the basal lamina, for want of adequate information about its thickness, mechanics, and spatial and temporal features. The basal lamina is likely to be most important mechanically as a barrier to expansion (growth), rather than as a regulator of clefting. Since we focused in this study on clefting only, separate from growth, we expect that the omission of a term for the basal lamina is reasonable.

The most serious limitation of our model is restricting ourselves to two dimensions due to computational constraints. A two-dimensional model requires us to artificially model the clefting force as a point force, and it probably affects the quantitative observations by at least a factor of two. Other recent mechanical models of morphogenesis have either been confined to two dimensions (9,58) or, if in three dimensions, have been on computational domains much smaller than ours (8), or axisymmetric (25,36), or using a mechanical model that was easier to solve accurately (10). Methods to accurately solve multi-fluid flow problems in 3D are still being developed.

In our tests of the effect of the viscosity ratio, we found that a given epithelially generated force could cause very different amounts of deformation in a given time period, depending on how viscous was the material in which it was embedded. The major developmental conclusion we can draw from our results is this: any experiment in which the mesenchyme of a clefting rudiment is replaced by a gel of a different viscosity is *not* mechanically equivalent to an experiment with epithelium embedded in intact mesenchyme. Most crucially, the time course of a thin-ECM clefting will be substantially faster than a viscous-ECM clefting. By analogy, a hand can very easily move in water, but very slowly if embedded in clay, and hardly at all in plaster. In particular, we cannot conclude from mesenchyme-free experiments (40,46,48) that the clefting force of branching morphogenesis comes solely from the epithelium *unless* the gel is of the same viscosity as the tissue it replaces *and* the time course of clefting is the same in mesenchyme-free rudiments and intact rudiments. In short, until mechanical measurements are made of the tissues and gels involved, we cannot conclude from mesenchyme-free experiments that branching morphogenesis is driven solely by forces of epithelial origin.

The model described in this chapter has been very useful in answering a few questions, but others cannot be answered with these modeling tools. For example, clefts are known to be filled with collagen fibers (19,43,44), and the mesenchyme close to a branched epithelium is denser than that far away, yet the epithelial and mesenchymal theories could both explain these phenomena. As we showed in the case of tumor encapsulation, a dense layer of tissue can be formed equally well by contractility from outside or by suction from inside (36). The right model can clarify the implications of a hypothesis while suggesting refinements to a theory and also suggesting more rigorous experimental frameworks.

There is ample work still to be done to understand the mechanical aspects of branching morphogenesis. The model described in this chapter is extremely simple, and focused on understanding a single aspect of branching—the mechanical implications of being surrounded by mesenchyme or surrounded by a collagen gel. The conclusion is clear: pushing against a soft material is easier than pushing against a firm material. Mesenchyme provides more resistance than a typical collagen gel. To move the mesenchyme requires its cooperation.

What the model presented in this chapter does not address is whether or not the mesenchyme can generate enough force to create a cleft. It also assumes that the viscosities of the epithelium and mesenchyme are constant. An alternative hypothesis of the mechanical aspects of branching morphogenesis would assume that the mesenchyme is remaining passive, but is not resisting deformation nearly as much as we suppose, because the epithelium is in a sense melting it away with collagenases as it expands. These collagenases would have to be localized in the proliferating tips, because it has been shown (24) that collagenases added to the culture medium inhibit branching by preventing cleft formation.

Different branched organs (lung, salivary gland, mammary gland, kidney, etc.) have very different morphologies. It is possible that the different shapes are due primarily to mechanical differences in the tissues involved during branching morphogenesis (39). For example, we found that the viscosity of the surrounding mesenchyme affected the cleft shape, with relatively soft mesenchyme leading to sharper clefts than firmer mesenchyme. It is possible that the wide clefts of the embryonic lung form because lung mesenchyme is firmer relative to lung epithelium than salivary mesenchyme is to salivary epithelium. But we do not know, because no one has measured the viscosities of these specific tissues.

The recombination experiments of Lawson (31–33), Spooner and Wessells (55), Grobstein (16), Ball (1), and Kratochwil (29) indicate that combinations of epithelium and mesenchyme from different organs generally lead to morphology characteristic of the mesenchyme's organ of origin, but that in some cases, branching does not occur. While differences in growth factors may offer part of the explanation, it may be that the rest of the explanation lies in simple mechanical differences between the mesenchymes. How strong are the epithelia? How viscous are the mesenchymes?

Measurements of the tissue viscosities and surface tensions of the epithelia and mesenchymes involved could provide a simple and fascinating key to interpreting the large number of facts we have accumulated about branching morphogenesis. If we really want to understand developmental mechanisms, we will take mechanical measurements.

5. ACKNOWLEDGMENTS

This work has been supported by the National Science Foundation and the National Institutes of Health (DMS 9805611, 0201094).

6. REFERENCES

1. Ball WD 1974. Development of the rat salivary glands, III: mesenchymal specificity in the morphogenesis of the embryonic submaxillary and sublingual glands of the rat. *J Exp Zool* **188**:277–288.
2. Banerjee SD, Cohn RH, Bernfield MR. 1977. Basal lamina of embryonic salivary epithelium. *J Cell Biol* **73**:445–463.
3. Bard JBL. 1990. *Morphogenesis: the cellular and molecular processes of developmental anatomy*. Cambridge UP, Cambridge.
4. Barocas VH, Moon AG, Tranquillo RT. 1995. The fibroblast-populated collagen microsphere assay of cell traction force 2: measurement of the cell traction parameter. *J Biomech Eng* **117**(2):161–170.
5. Bassingthwaighte JB, Leibovich LS, West BJ. 1994. *Fractal physiology*. Oxford UP, New York.

6. Bernfield MR, Banerjee SD. 1982. The turnover of basal lamina glycosaminoglycan correlates with epithelial morphogenesis. *Dev Biol* **90**:291–305.
7. Bernfield MR, Banerjee SD, Koda JE, Rapraeger AC. 1984. Remodelling of the basement membrane as a mechanism of morphogenetic tissue interaction. In *The role of extracellular matrix in development*, pp. 545–572. Ed. RL Trelstad. Liss, New York.
8. Brodland GW, Clausi DA. 1994. Embryonic tissue morphogenesis modeled by FEM. *J Biomech Eng* **116**(2):146–155.
9. Chen HH, Brodland GW. 2000. Cell-level finite element studies of viscous cells in planar aggregates. *J Biomech Eng* **122**(4):394–401.
10. Davidson LA, Koehl MAR, Keller R, Oster GF. 1995. How do sea urchins invaginate? using biomechanics to distinguish between mechanisms of primary invagination. *Development* **121**(7):2005–2018.
11. Davies JA. 2002. Do different branching epithelia use a conserved developmental mechanism? *BioEssays* **24**:937–948.
12. Drury JL, Dembo M 1999. Hydrodynamics of micropipette aspiration. *Biophys J* **76**(1, pt. 1):110–128.
13. Forgacs G, Foty RA, Shafrir Y, Steinberg MS. 1998. Viscoelastic properties of living embryonic tissues: a quantitative study. *Biophys J* **74**:2227–2234.
14. Foty RA, Forgacs G, Pfleger CM, Steinberg MS. 1994. Liquid properties of embryonic tissues: measurement of interfacial tensions. *Phys Rev Lett* **72**(14):2298–2301.
15. Greenspan HP. 1977. On the dynamics of cell cleavage. *J Theor Biol* **65**:79–99.
16. Grobstein C. 1953. Epithelio-mesenchymal specificity in the morphogenesis of mouse submandibular rudiments *in vitro*. *J Exp Zool* **124**:383–414.
17. Grobstein C. 1953. Analysis *in vitro* of the early organization of the rudiments of the mouse submandibular gland. *J Morph* **93**:19–44.
18. Grobstein C. 1953. Morphogenetic interaction between embryonic mouse tissues separated by a membrane filter. *Nature* **172**:869–871.
19. Grobstein C, Cohen J. 1965. Collagenase: effect on the morphogenesis of embryonic salivary epithelium in vitro. *Science* **150**:626–628.
20. Hamamoto S, Imagawa W, Yang J, Nandi S. 1988. Morphogenesis of mouse mammary epithelial cells growing within collagen gels: ultrastructural and immunocytochemical characterization. *Cell Diff* **22(3)**:191–201.
21. Hardman P, Spooner BS. 1992. Salivary epithelium branching morphogenesis. In *Epithelial organization and development*, pp. 353–375. Ed. TP Fleming. Chapman and Hall, London.
22. Harris AK, Stopak D, Wild P. 1980. Fibroblast traction as a mechanism for collagen morphogenesis. *Nature* **290**:249–251.
23. Harris AK, Stopak D, Warner P. 1984. Generation of spatially periodic patterns by a mechanical instability: a mechanical alternative to the Turing model. *J Embryol Exp Morphol* **80**:1–20.
24. Hayakawa T, Kishi J, Nakanishi Y. 1992. Salivary gland morphogenesis: possible involvement of collagenase. *Matrix Suppl* **1**:344–351.
25. He X, Dembo M. 1997. A dynamical model of cell division. In *Dynamics of cell and tissue motion*, pp. 55–66. Ed. W Alt, A Deutsch, G Dunn. Birkhäuser, Basel.
26. Hieda Y, Nakanishi Y. 1997. Epithelial morphogenesis in mouse embryonic submandibular gland: its relationships to the tissue organization of epithelium and mesenchyme. *Dev Growth Diff* **39**:1–8.
27. Kitaoka H, Takaki R, Suki B. 1999. A three-dimensional model of the human airway tree. *J Appl Physiol* **87**(6):2207–2217.
28. Kolodney MS, Wysolmerski RB. 1992. Isometric contraction by fibroblasts and endothelial cells in tissue culture: a quantitative study. *J Cell Biol* **117**:73–82.
29. Kratochwil, K. 1969. Organ specificity in mesenchymal induction demonstrated in the embryonic development of the mammary gland of the mouse. *Dev Biol* **20**:46–71.

30. Lane MC, Koehl MAR, and Keller R. 1993. A role for regulated secretion of apical extracellular matrix during epithelial invagination in the sea urchin. *Development* **117**(3):1049–1060.
31. Lawson FW, 1972. The role of mesenchyme in the morphogenesis and functional differentiation of rat salivary epithelium. *J Embryol Exp Morphol* **27**:497–513.
32. Lawson KA. 1974. Mesenchyme specificity in rodent salivary gland development: the response of salivary epithelium to lung mesenchyme in vitro. *J Embryol Exp Morphol* **32**:469–493.
33. Lawson KA. 1983. Stage specificity in the mesenchyme requirement of rodent lung epithelium in vitro: a matter of growth control? *J Embryol Exp Morphol* **74**:183–206.
34. Leal LG. 1992. *Laminar flow and convective processes: scaling principles and asumptotic analysis*. Butterworth-Heinemann, Boston
35. Li Z, Lubkin SR. 2001. Numerical analysis of interfacial stokes flow with discontinuous viscosity and nonlinear surface tension. *Int J Numer Meth Fluids* **37**:525–540.
36. Lubkin SR, Jackson T. 2002. Multiphase mechanics of capsule formation in tumors. *J Biomech Eng* **124**:237–243.
37. Lubkin SR, Li Z. 2002. Force and deformation on branching rudiments: cleaving between hypotheses. *Biomech Modeling Mechanobiol* **1**(1):5–16.
38. Lubkin SR, Murray JD. 1995. A mechanism for early branching in lung morphogenesis. *J Math Biol* **34**:77–94.
39. Matsui R, Thurlbeck WM, Shehata EI, Sekhon HS. 1996. Two different patterns of airway branching regulated by different components of the extracellular matrix in vitro. *Exp Lung Res* **22**(6):593–611.
40. Miura T, Shiota K. 2000. Time-lapse observation of branching morphogenesis of the lung bud epithelium in mesenchyme-free culture and its relationship with the localization of actin filaments. *Int J Dev Biol* **44**(8):899–902.
41. Nakanishi Y, Ishii T. 1989. Epithelial shape change in mouse embryonic submandibular gland modulation by extracellular matrix components. *Bioessays* **11**(6):163–167.
42. Nakanishi Y, Morita T, Nogawa H. 1987. Cell proliferation is not required for the initiation of early cleft formation in mouse submandibular epithelium in vitro. *Development* **99**:429–437.
43. Nakanishi Y, Sugiura F, Kishi J-I, T. H. 1986. Collagenase inhibitor stimulates cleft formation during early morphogenesis of mouse salivary gland. *Dev Biol* **113**:201–206.
44. Nakanishi Y, Sugiura F, Kishi J, Hayakawa T. 1986. Scanning electron microscopic observation of mouse embryonic submandibular glands during initial branching: preferential localization of fibrillar structures at the mesenchymal ridges participating in cleft formation. *J Embryol Exp Morphol* **96**:65–77.
45. Nakanishi Y, Ishii T. 1989. Epithelial shape change in mouse embryonic submandibular gland: modulation by extracellular matrix components. *BioEssays* **11**:163–167.
46. Nogawa H, Ito T. 1995. Branching morphogenesis if embryonic mouse lung epithelium in mesenchyme-free culture. *Development* **121**:1015–1022.
47. Nogawa H, Nakanishi Y. 1987. Mechanical aspects of the mesenchymal influence on epithelial branching morphogenesis of mouse salivary gland. *Development* **101**:491–500.
48. Nogawa H, Takahashi Y. 1991. Substitution for mesenchyme by basement-membrane-like substratum and epidermal growth factor in inducing branching morphogenesis of mouse salivary epithelium. *Development* **112**:855–861.
49. Phillips HM, Steinberg MS, Lipton BH. 1977. Embryonic tissues as elasticoviscous liquids, II: direct evidence for cell slippage in centrifuged aggregates. *Dev Biol* **59**:124–134.
50. Phillips HM, Steinberg MS. 1978. Embryonic tissues as elasticoviscous liquids, I: rapid and slow shape changes in centrifuged cell aggregates. *J Cell Sci* **30**:1–20.
51. Rappaport R. 1977. Tensiometric studies of cytokinesis in cleaving sand dollar eggs. *J Exp Zool* **201**:375–78.
52. Reed J, Vernon RB, Abrass IB, Sage EH. 1994. TGF-B1 induces the expression of type 1 collagen and SPARC, and enhances contraction of collagen gels, by fibroblasts from young and aged donors. *J Cell Physiol* **158**:169–179.

53. Roux W 1895. The problems, methods, and scope of developmental mechanics. *Archiv für Entwicklungsmechanik der Organismen* 1:1
54. Spooner BS, Bassett KE, Spooner Jr. BS. 1989. Embryonic salivary gland epithelial branching activity is experimentally independent of epithelial expansion activity. *Dev Biol* **133**:569–575.
55. Spooner BS, Wessells NK. 1972. An analysis of salivary gland morphogenesis: role of cytoplasmic microfilaments and microtubules. *Dev Biol* **27**:38–54.
56. Spooner BS, Thompson-Pletscher HA, Stokes B, Bassett KE. 1986. Extracellular matrix involvement in epithelial branching morphogenesis. In *Developmental biology*, Vol. 3: *Cell surface development and cancer*, pp. 225–260. Ed MS Steinberg. Plenum, New York.
57. Stopak D, Harris AK. 1982. Connective tissue morphogenesis by fibroblast traction. *Dev Biol* **90**:383–398.
58. Taber LA. 2000. Pattern formation in a nonlinear membrane model for epithelial morphogenesis. *Acta Biotheoret* **48**(1):47–63.
59. Takahashi Y, Nogawa H. 1991. Branching morphogenesis of mouse salivary epithelium in basement-membrane-like substratum separated from mesenchyme by the membrane filter. *Development* **111**:327–335.
60. Vernon RB, Angello JC, Iruela-Arispe ML, Lane TF, Sage EH. 1992. Reorganization of basement membrane matrices by cellular traction promotes the formation of cellular networks in vitro. *Lab Invest* **66**(5):536–547.
61. Vernon RB, Lara SL, Drake CJ, Iruela-Arispe ML, Angello JC, Little CD, Wight TN, Sage EH. 1995. Organized type I collagen influences endothelial patterns during "spontaneous angiogenesis in vitro": planar cultures as models of vascular development. *In Vitro Cell Dev Biol* **31**(2):120–131.
62. Weibel ER. 1963. *Morphometry of the human lung*. Springer, New York.
63. Wolpert L. 1991. *The triumph of the embryo*. Oxford UP, Oxford.
64. Yang J, Larson L, Nandi S. 1982. Three-dimensional growth and morphogenesis of mouse submandibular epithelial cells in serum-free primary culture. *Exp Cell Res* **137**(2):481–85.
65. Yang J, Balakrishnan A, Hamamoto S, Elias JJ, Rosenau W, CW Beattie, Das Gupta TK, Wellings SR, Nandi S. 1987. Human breast epithelial cells in serum-free collagen gel primary culture: growth, morphological, and immunocytochemical analysis. *J Cell Physiol* **133**(2):228–234, 254–255.

3.2

MODELING CARDIAC FUNCTION

Raimond L. Winslow

The Center for Cardiovascular Bioinformatics and Modeling,
The Johns Hopkins University School of Medicine and
Whiting School of Engineering, Baltimore, Maryland

This chapter reviews the current state of integrative modeling of the heart, focusing on three topics. First, we review integration of experimental data into the most commonly used class of ventricular myocyte models—common pool models. We critically assess both the successes and failures of these models. Second, we review the formulation of a new class of myocyte models known as local-control models. While these models are more computationally intensive than are common pool models, they are able to capture critically important aspects of single channel behaviors that have a profound impact on myocyte function, and which cannot be described using common pool models. Finally, we review how cellular models may be integrated with imaging data on heart geometry and micro-anatomic structure to formulate computational models of cardiac ventricular electrical conduction.

1. INTRODUCTION

There is growing recognition that the identification of genetic and molecular building blocks from which biological systems are composed, while being critically important, is not by itself sufficient for understanding the functional properties of these systems. Rather, it is clear that the emergent, integrative behaviors of biological systems result from complex interactions between system components, and that development/analysis of computational models based directly on experimental data provides a powerful tool for understanding relationships between gene/protein expression and biological function at the level of cell and tissue in both health and disease.

Address correspondence to: Raimond L. Winslow, PhD, Room 201B, Clark Hall, The Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218 (rwinslow@bme.jhu.edu).

Cardiac electrophysiology is a field with an extensive history of integrative modeling that has been coupled closely with both the design and interpretation of experiments. The first models of the cardiac action potential (AP) were developed shortly after the Hodgkin-Huxley model of the squid AP (1,2), and were formulated in order to explain the experimental observation that, unlike the neuronal AP, cardiac APs exhibit a long-duration plateau phase. It was not long after the formulation of these early myocyte models that initial models of electrical conduction in cardiac tissue were formulated and applied to yield clinically useful insights into mechanisms of arrhythmia (3). This close interplay between experiment and integrative modeling continues today, with new model components and applications being developed in close coordination with the emergence of new subcellular, cellular, and whole-heart data describing cardiac function in health and disease.

This chapter will review the current state of integrative modeling of the heart, focusing on three topics. First, we will review the integration of experimental data into the most commonly used class of ventricular myocyte models—common pool models. These models take the form of coupled systems of ordinary differential-algebraic equations. We will examine both the successes and failures of these common pool models. Second, we will review the formulation of a new class of myocyte models known as local-control models. These models take the form of coupled systems of stochastic differential equations, whose properties are evolved in time using a combination of Monte Carlo simulation and numerical integration. While these models are more computationally intensive than common pool models, they are able to capture critically important aspects of single channel behaviors that have a profound impact on myocyte function, and which cannot be described using common pool models. Finally, we will review how cellular models may be integrated with imaging data on heart geometry and micro-anatomic structure to formulate computational models of cardiac ventricular electrical conduction.

2. CELLULAR MODELS

2.1. The Cardiac Action Potential

In order to understand the properties of modern computational models of the cardiac myocyte, it is necessary to review the ionic mechanisms giving rise to the cardiac AP. In this and all other sections of this chapter, we will focus exclusively on the description and models of the properties of cardiac ventricular myocytes, as the properties of these cells figure so importantly in the genesis of heart disease.

Figure 1A shows a schematic illustration of the large mammalian cardiac AP. The currents mediating the AP upstroke (Phase 0) are the fast inward sodium (Na^+) current (I_{Na}, for review see (4)), and to a lesser extent the L-Type

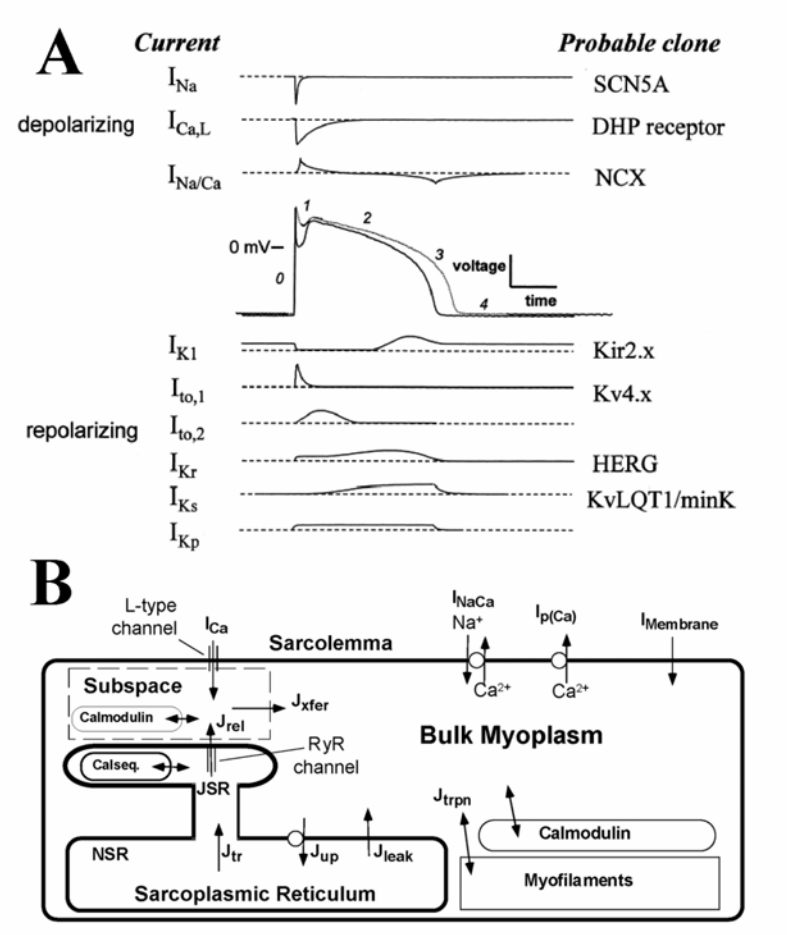


Figure 1. (**A**) Schematic illustration of the large mammalian cardiac ventricular myocyte action potential (membrane potential in mV as a function of time) illustrating depolarizing and repolarizing current (left) and alias gene names (right) encoding each of these currents. (Reprinted with permission from Tomaselli and Marban (1999) (75).) (**B**) Schematic illustration of the structure of common pool ventricular myocyte models.

Ca^{2+} current ($I_{Ca,L}$, for review see (5)). The Phase 1 notch, which is apparent in ventricular myocytes isolated from epi- and mid-myocardial regions, but largely absent in those isolated from the endocardium, is produced by activation of the voltage-dependent transient outward potassium (K^+) current ($I_{to,1}$). In the canine, a transient voltage-independent Ca^{2+}-modulated Cl^- current contributes to the Phase 1 notch ($I_{to,2}$), however, this current is not known to be expressed in the human. The Phase 2 plateau is a time during which membrane conductance is

very low, with potential being determined by a delicate balance between small inward and outward currents. The major inward plateau current is $I_{Ca,L}$, and major outward plateau currents are generated by the rapid and slow-activating delayed outward rectifier K^+ currents I_{Kr} and I_{Ks}, respectively, and the plateau K^+ current I_{Kp}. Finally, repolarization Phase 3 is produced by the hyperpolarizing activated inward rectifier K^+ current I_{K1}.

Three major ion transporters and exchangers play a critically important role in shaping properties of the cardiac AP, Ca^{2+} transient, and in long-term regulation of intracellular ion concentrations. These are the sarcolemmal Na^+–K^+ pump, the sarcolemmal Na^+–Ca^{2+} exchanger, and the SR Ca^{2+}–ATPase. The sarcolemmal Na^+–K^+ pump, present in virtually all mammalian cell membranes, extrudes 3 Na^+ ions while importing 2 K^+ ions on each cycle. This pump functions to keep intracellular Na^+ low, thereby maintaining the external versus internal gradient of Na^+, by extruding Na^+ that enters during each AP. Cycling of this pump requires hydrolysis of 1 ATP molecule, and generates a net outward movement of 1 positive charge, thus contributing to outward membrane current and influencing resting membrane potential.

The sarcolemmal Na^+–Ca^{2+} exchanger imports three Na^+ ions for every Ca^{2+} ion extruded, yielding a net charge movement. It is driven by both transmembrane voltage and intra- and extracellular Na^+ and Ca^{2+} ion concentrations. It functions in forward mode during diastole, in which case it extrudes Ca^{2+} and imports Na^+, thus generating a net inward current. It is the principal means by which Ca^{2+} is extruded from the myocyte following each AP, particularly during the diastolic interval. Due to the voltage- and Ca^{2+}-sensitivity of the exchanger, experimental evidence indicates that it can function in reverse mode during the plateau phase of the AP, in which case it extrudes Na^+ and imports Ca^{2+}, thus generating a net outward current.

A second major cytoplasmic Ca^{2+} extrusion mechanism is the SR Ca^{2+}–ATPase. This ATPase pumps Ca^{2+} from the cytosol into the NSR. The SR Ca^{2+}–ATPase has both forward and reverse components (7), with the reverse component serving to prevent overloading of the SR with Ca^{2+} at rest. An additional Ca^{2+} extrusion mechanism is the sarcolemmal Ca^{2+}–ATPase. This Ca^{2+} pump hydrolyzes ATP to transport Ca^{2+} out of the cell. However, it contributes a sarcolemmal current that is small relative to that of the Na^+–Ca^{2+} exchanger, with estimates indicating perhaps that as little as 3% of Ca^{2+} extrusion from the myocyte is mediated by this pump.

2.2. The Structure of Myocyte Models

Development of myocyte models began in the early 1960s with publication of Purkinje fiber AP models. Subsequent elaboration of these models led to development of the first biophysically based cell model describing interactions

between voltage-gated membrane currents, membrane pumps, and exchangers that regulate Ca^{2+}, Na^{+} and K^{+} levels, and additional intracellular Ca^{2+} cycling processes in the cardiac myocyte—the DiFrancesco-Noble model of the Purkinje fiber (8). This important model established the conceptual framework from which all subsequent models of the myocyte have been derived (ventricular myocytes (9–12); SA node cells (13–16), and atrial myocytes (17,18)). These models have proven reproductive and predictive properties and have been applied to advance our understanding of myocyte function in both health and disease.

Each of the integrative models of the myocyte cited above are of a type known as "common pool" models (19), the structure of which is shown in Figure 1B. In such models, Ca^{2+} flux through both L-type Ca^{2+} channels (LCCs) and ryanodine-sensitive Ca^{2+} release channels (RyRs) in the JSR membrane is directed into a single common Ca^{2+} compartment referred to as the subspace. The subspace represents the total volume of the ~5,000 diadic spaces present in the ventricular myocyte. Stern demonstrated that common pool models are structurally unstable, exhibiting all-or-none Ca^{2+} release except (possibly) over some narrow range of model parameters (20). This instability occurs because Ca^{2+} release from JSR produces a large, rapid increase of Ca^{2+} concentration in the subspace. This in turn results in a very strong positive feedback effect in which increased binding of Ca^{2+} to RyR induces further RyR channel opening and release of Ca^{2+}. Despite this inability to reproduce experimentally measured properties of graded JSR Ca^{2+} release, common pool models have been very successful in reproducing and predicting a range of myocyte behaviors. This includes properties of interval–force relationships that depend heavily on proper dynamic modeling of intracellular Ca^{2+} uptake and release mechanisms (21). In the following sections we describe the components from which common pool models are formed.

2.3. Model Components: Ion Channels and Currents

For many years Hodgkin-Huxley models have been the standard for describing membrane current kinetics (22). However, data obtained using new experimental approaches, in particular those for producing recombinant channels by coexpression of genes encoding pore-forming and accessory channel subunits in host cells, have shown these models to have significant limitations. First, while these models can be expanded to an equivalent Markov chain representation having multiple closed and inactivated states (24), many single channel behaviors such as mean open time, first latency, and a broad range of other kinetic behaviors cannot be described using these equivalent Markov models (6,25). Second, where it has been studied in detail, as is the case for cardiac Na channels, Hodgkin-Huxley models are insufficient for reproducing behaviors

that may be critically state-dependent, such as how ionic channels interact with drugs and toxins (26,27). Accordingly, much recent effort in modeling of cardiac ionic currents has focused on development of biophysically detailed Markov chain models of channel gating. We will therefore illustrate the generic concepts involved in modeling of ion channel function and membrane currents by reviewing our recent efforts to model the cardiac sodium (Na^+) channel (28). This model is able to reproduce and predict a wide range of single channel and whole-cell current properties (28), and the ways in which this model is formulated and constrained is illustrative of modern approaches to ion channel and current modeling.

The structure of the model is shown in Figure 2A. The channel can occupy any of 13 states. The top row of states corresponds to zero to four voltage sensors being activated (C_0 through C_4) plus an additional conformational change required for opening ($C_4 \rightarrow O_1$ and $C_4 \rightarrow O_2$). The bottom row of states corresponds to channel inactivation. Affinity of the inactivation particle binding site is hypothesized to increase by a scaling factor (a) as the channel activates and to decrease by the same factor as the channel deactivates. Closed–closed and closed–open (horizontal) transitions are voltage dependent and closed–inactivated (vertical0 transitions are voltage independent. Transition rates are of a form given by Eyring rate theory (24), and include explicit temperature dependence:

$$\lambda = \frac{kT}{h} \exp\left(\frac{-\Delta H_\lambda}{RT} + \frac{\Delta S_\lambda}{R} + \frac{z_\lambda FV}{RT}\right), \qquad [1]$$

where k is the Boltzmann constant, T is the absolute temperature, h is the Planck constant, R is the gas constant, F is Faraday's constant, ΔH_λ is the change in enthalpy, ΔS_λ is the change in entropy, z_λ is the effective valence (i.e., the charge moved times the fractional distance the charge is moved through the membrane), and V is the membrane potential in volts.

The probability of occupying any particular channel state is described by a set of ordinary differential equations, written in matrix notation as

$$\frac{\partial \mathbf{P}(\mathrm{t})}{\partial \mathrm{t}} = \mathbf{W}\mathbf{P}(\mathrm{t}), \qquad [2]$$

where $\mathbf{P}(t)$ is a vector state occupancy probabilities, and $\mathbf{W}$ is the state transition matrix. $\mathbf{W}$ is in general a function of voltage and time. For the voltage-clamp conditions generally used to constrain ion current models, $\mathbf{W}$ is piecewise time-independent, thus Eq. [2] has the analytic solution

$$\mathbf{P}(t) = \exp(\mathbf{W}t)\mathbf{P}(0). \qquad [3]$$

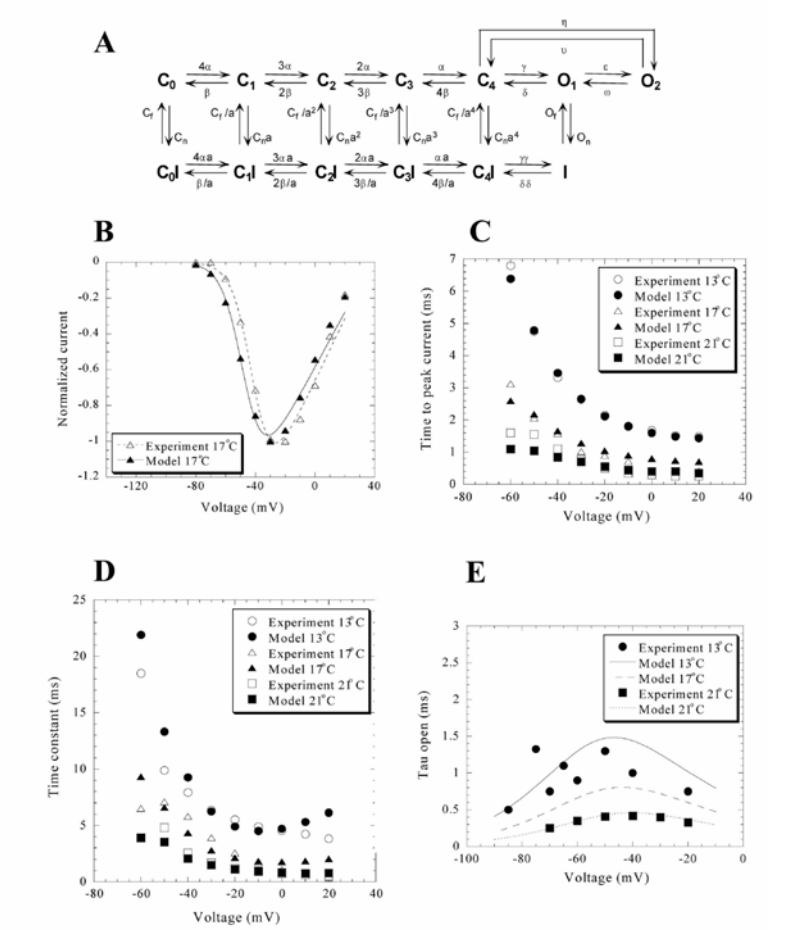


Figure 2. (**A**) Markov chain model of the human cardiac Na channel. States $C_{0\text{-}4}$ are closed states, states $O_{1,2}$ are open, conducting states, and states $C_{1\text{-}4}$ and I are inactivated states. (**B**) Normalized peak Na current (ordinate) as a function of membrane potential (mV, abscissa). Open and filled symbols are experimental and model data at 17°C, respectively. (**C**) Time to peak Na current (msec, ordinate) as a function of membrane potential (mV, abscissa). Experimental and model data are compared at 13, 17, and 21°C. Data at 13 and 21°C are model fits, and data at 17°C constitute a model prediction. (**D**) Comparison of the time constant of Na current inactivation (ms, ordinate) as a function of clamp potential (mV, abscissa) predicted by the model versus those measured experimentally at 13, 17, and 21°C. (**E**) Model predictions (solid, dashed, and dotted lines) at 13, 17, and 21°C, respectively, of single-channel open time (msec, ordinate) as a function of membrane potential (mV, abscissa). Filled symbols are experimental data.

Current through an ensemble of Na channels is calculated as

$$I_{\mathrm{Na}} = NG_{\mathrm{Na}}P_{\mathrm{open}}(t)(V(t) - E_{\mathrm{Na}}(t)), \qquad [4]$$

where $I_{Na}(t)$ is the Na current, N is the number of Na channels, G_{Na} is single-channel conductance, $P_{open}(t)$ is the probability of occupying the open states (O_1 + O_2), $V(t)$ is membrane potential, and $E_{Na}(t)$ is the reversal potential for Na given by the Nernst equation.

The number of coupled differential equations, and hence the number of parameters that need to be constrained for the model, may be reduced through application of two fundamental principles. First, the state occupancy probabilities for a Markov chain model must sum to one. Second, there are several loops in the model that must satisfy the principle of microscopic reversibility. Microscopic reversibility is derived from the law of conservation of energy and states that the product of rate constants when traversing a loop clockwise must be equal to the product when traversing the same loop counterclockwise (24). For the closed–closed-inactivated loops, satisfying microscopic reversibility requires that the transitions among the closed-inactivated states be scaled by a, the same factor used to scale the transitions between rows. Microscopic reversibility is preserved around the closed–open-inactivated loop by isolating the ΔH, ΔS, and z terms in the product and satisfying each term separately using the following equations:

$$\Delta H_{\gamma\gamma} = \Delta H_{\gamma} + \Delta H_{on} + \Delta H_{\delta\delta} + \Delta H_{cf} + 8RT \ln a - \Delta H_{\delta} - \Delta H_{cn} - \Delta H_{of}, \quad [5]$$

$$\Delta S_{\gamma\gamma} = \Delta S_{\gamma} + \Delta S_{on} + \Delta S_{\delta\delta} + \Delta S_{cf} - \Delta S_{\delta} - \Delta S_{cn} - \Delta S_{of}, \quad [6]$$

$$z_{\gamma\gamma} = z_{\gamma} + z_{on} + z_{\delta} + z_{of} - z_{\delta\delta}. \quad [7]$$

Similarly, microscopic reversibility is preserved around the closed-open-open loop using the following equations for ΔH_{η}, ΔS_{η}, and z_{η}:

$$\Delta H_{\eta} = \Delta H_{\gamma} + \Delta H_{\varepsilon} + \Delta H_{\nu} - \Delta H_{\delta} - \Delta H_{\omega}, \quad [8]$$

$$\Delta S_{\eta} = \Delta S_{\gamma} + \Delta S_{\varepsilon} + \Delta S_{\nu} - \Delta S_{\delta} - \Delta S_{\omega}, \quad [9]$$

$$z_{\eta} = z_{\gamma} + z_{\delta} - z_{\nu}. \quad [10]$$

These microscopic reversibility constraints thus reduce the dimension of the parameter estimation problem, as transition rates $\gamma\gamma$ and η are fully constrained.

The model of Figure 2A may also be viewed as a Markov chain description of single channel behavior. Single-channel gating may be simulated using the method of Clay and DeFelice (29). In this method, the length of time a channel stays in its current state (i.e., its dwell time, denoted as T_j) is calculated according to the formula

$$T_j = -(\ln r) \Big/ \sum_{k=1}^{x} \lambda_{jk} \,, \qquad [11]$$

where r is a random variable drawn from a uniform distribution on the interval [0,1] and λ_{jk} is the transition rate from state j to state k. The sum is over the x pathways out of state j. The resulting dwell time T_j is an exponentially distributed random variable with parameter $\lambda = \sum_{k=1}^{x} \lambda_{jk}$. At the end of the dwell time, the new state of the channel is determined by assigning random numbers to a portion of the interval [0,1] based on the probabilities of changing to neighboring states. These probabilities are equal to the rate constant for a particular transition divided by the sum of the rate constants for all possible transitions. Once the new state is determined, another random number is used to calculate the dwell time in the new state. At an instantaneous voltage step, channels remain in their current state, but the dwell times are recalculated.

Extensive experimental data were required to fully determine the model parameters. The majority of data were taken from human SCN5A-encoded Na channels. Experimental data obtained at temperatures of 13° and 21°C were used to constrain the model, and the ability of the model to predict data collected at 17°C was tested. Constraining data included: (a) ionic currents in response to voltage-clamp; (b) gating charge accumulation; (c) steady-state inactivation curve; (d) rate of tail current relaxation; (e) time course of recovery from inactivation; and (f) single-channel open times. A cost function defined as the squared error between simulated and experimental data (including both whole-cell current and single-channel data) was minimized to determine an optimal model parameter set. A simulated annealing algorithm (30) was needed to perform this minimization, as the cost function exhibited many local minima. The resulting model was able to reproduce a broad range of membrane current data (28), and Figures 2B–E demonstrates the ability of the model to predict channel/current properties at 17°C as well as single-channel data not included in the fitting process. A similar methodology has been used to develop quantitative models of other myocyte membrane currents, most notably I_{Kr}, I_{Ks}, I_{CaL}, and I_{to1} (12,31–33).

2.4. Model Components: Intracellular Ion Concentration Changes

We illustrate the process of modeling time-varying changes of intracellular ion concentration with reference to the common pool model architecture shown in Figure 1B. In the common pool model formulation, there are four distinct Ca^{2+} compartments (the cytosol, subspace, NSR, and JSR) and one Na^+ and potassium (K^+) compartment (the cytosol). Note that in present common pool myocyte models, the cytosolic concentrations of both Na^+,K^+ and Ca^{2+} are assumed to be

uniform. The time rate of change of concentration C_i of the ith ionic species in a given compartment is given by

$$\frac{dC_i(t)}{dt} = \frac{I_i(t)}{z_i FV}, \qquad [12]$$

where $C_i(t)$ is the concentration (typically mM) of species i; t is time (typically in msec), $I_i(t)$ is net current into the compartment carried by species i (typically in pA); z_i is the valence of the ith species, F is Faraday's constant; and V is the compartment volume (typically in units of pL). One such equation may be defined for the concentration of each ionic species in each model compartment. Ion flux between compartments, related to the term $I_i(t)$ in Eq. [12], is produced either by: (a) diffusion due to differences in ion species concentration between adjacent compartments (as is the case for the flux term J_{xfer} in Figure 1B representing Ca^{2+} diffusion from the subspace to the cytosol); (b) gating of ion channels in the sarcolemmal or JSR membrane (as is the case for Ca^{2+} flux J_{rel} in Figure 1B from the JSR into the subspace through RyR channels); or (c) the action of membrane transporters and exchangers (for example, Ca^{2+} flux through the SR Ca^{2+}–ATPase, labeled J_{up} in Figure 1B). The form of the algebraic equations describing the function of membrane transporters and exchangers, including their concentration, voltage, and in some instances ATP dependence, may be found in the published equations for a number of myocyte models. In addition, buffering of Ca^{2+} by negatively charged phospholipid head groups in the sarcolemmal and JSR subspace membrane, by cytosolic myofilaments (troponin) and by calsequestrin in the JSR is modeled. Buffering due to mechanisms other than myofilaments is described using the rapid buffer approximation of Wagner and Keizer (34).

2.5. Composite Equations for Common Pool Models

Common pool models of the cardiac myocyte consist of systems of nonlinear ordinary differential-algebraic equations describing the time evolution of model state variables. These state variables are: (a) probability of occupancy of ion channel states (Eq. [2]) and current flux through open channels (Eq. [4]); (b) concentrations of ion species in model compartments (Eq. [12]); and (c) time evolution of membrane potential. Currently, all biophysically detailed models of the myocyte assume that since these cells are spatially compact they are isopotential, with time-rate-of-change of membrane potential given by

$$\frac{dv(t)}{dt} = -\left\{\sum_i I_i^{ion}[v(t)] + \sum_i I_i^{pump}[v(t), c(t)]\right\}, \qquad [13]$$

where $v(t)$ is membrane potential, $I_i^{ion}[v(t)]$ is the current carried by the ith membrane current, and $I_i^{pump}[v(t),c(t)]$ is the current through the ith membrane pump/exchanger, which can depend on both membrane potential $v(t)$ and the relevant ion concentration $c(t)$. Figure 3 shows examples of simulated normal APs (solid line, Figure 3C) and Ca^{2+} transients (solid line, Figure 3D) compared with those measured from isolated canine ventricular myocytes (AP = solid line, Figure 3A; Ca^{2+} transient = solid line, Figure 3B). These data demonstrate that common pool models have been quite successful in reconstruction of the AP and in reconstructing some aspects (the time-varying waveform) of the Ca^{2+} transient. In the following section, we illustrate how such models may be used to gain insight into cardiovascular disease mechanisms.

2.6. Application: Modeling the Molecular Basis of Heart Failure

Heart failure (HF), the most common cardiovascular disorder, is characterized by ventricular dilatation, and decreased myocardial contractility and cardiac output. Prevalence in the general population is over 4.5 million, and increases with age to levels as high as 10%. New cases number approximately 400,000 per year. Patient prognosis is poor, with mortality roughly 15% at one year, increasing to 80% at six years subsequent to diagnosis. It is now the leading cause of sudden cardiac death (SCD) in the United States. An increased understanding of the molecular basis of this disease therefore offers the possibility of improved treatments that can reduce the risk of SCD.

Experimental studies have now identified two major features of the cellular phenotype of heart failure. First, ventricular myocytes isolated from failing human (35) and canine (36,37) hearts exhibit significant AP prolongation. An example of this AP prolongation recorded from normal versus failing canine ventricular myocytes is shown in Figure 3A (normal and failing APs shown in solid and dashed lines, respectively). Duration of the failing AP (~660 msec) is roughly twice that of the normal (~330 msec). AP duration is controlled by the balance between inward and outward membrane currents during the plateau phase of the AP. Possible explanations for this prolongation are therefore HF-induced upregulation of inward currents, and/or downregulation of outward currents. Second, failing ventricular myocytes exhibit altered Ca^{2+} transients. An example of normal and failing Ca^{2+} transients obtained simultaneously with the AP recordings of Figure 3A is shown in Figure 3B. Differences between normal (solid line) and failing (dotted line) Ca^{2+} transients include: (a) reduced amplitude; and (b) reduced rate of decline of the Ca^{2+} transient subsequent to repolarization of the AP.

There is little evidence to support the idea that upregulation of inward currents is responsible for prolongation of AP duration in HF, as the majority of measurements of whole-cell Na^+ and Ca^{2+} current density show no change in the

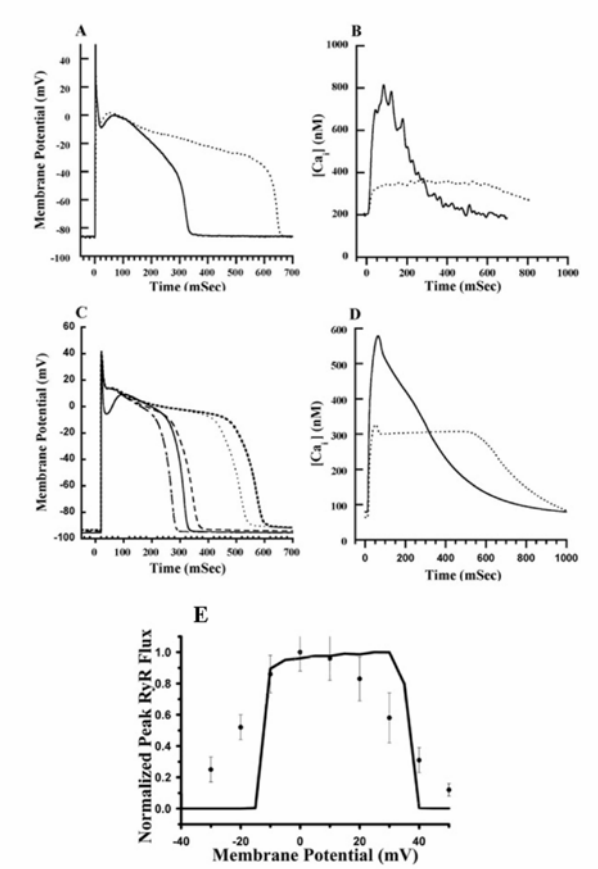


Figure 3. Model versus experimental action potentials and Ca^{2+} transients. Each action potential and Ca^{2+} transient is in response to a 1 Hz pulse train, with responses measured in the steady state. (**A**) Experimentally measured membrane potential (mV, ordinate) as a function of time (msec, abscissa) in normal (solid) and failing (dotted) canine myocytes. (**B**) Experimentally measured cytosolic Ca^{2+} concentration (nmol/L, ordinate) as a function of time (msec, abscissa) for normal (solid) and failing (dotted) canine ventricular myocytes. (**C**) Membrane potential (mV, ordinate) as a function of time (msec, abscissa) simulated using the normal canine myocyte model (solid), and with the successive downregulation of I_{to1} (dot-dashed, 66% downregulation), I_{K1} (long-dashed, downregulation by 32%), SERCA2 (rightmost short-dashed, downregulation by 62%) and NCX1 (dotted, upregulation by 75%). (**D**) Cytosolic Ca^{2+} concentration (nmol/L, ordinate) as a function of time (msec, abscissa) simulated using the normal (solid) and heart failure (dotted) model. (**E**) Normalized peak RyR Ca^{2+} release flux (ordinate) as a function of membrane potential (mV, abscissa) measured experimentally (filled circles) and predicted using the common pool model (solid line). Reprinted with permission of the Royal Society of London from Winslow et al. (2001) (23).

density of these currents (37). However, downregulation of voltage-gated K currents is known to occur in HF. Measurements of whole-cell inward rectifier current I_{K1} show that current density at hyperpolarized membrane potentials is reduced in HF by ~50% in human (38) and by ~40% in dog (36). Measurements

of I_{to1} (39) show that in end-stage HF, current density is reduced by up to 70% in human and canine tachycardia pacing-induced HF. Human and canine Ca^{2+}-independent transient outward current, I_{to1}, is a combination of currents encoded by the KCND3 and KCNA4 genes (39,40), and KCND3 expression has been shown to be reduced in HF (41). There appears to be no change in expression of the HERG or KCNQ1 gene encoding α-subunits of the I_{Kr} and I_{Ks} channels, respectively, in HF.

Expression of diverse proteins involved in the processes of EC coupling have also been measured in normal and failing myoyctes. These proteins include: (a) the SR Ca^{2+}–ATPase encoded by the SERCA2 gene; (b) the phospholamban protein encoded by the PLN gene; and (c) the sodium–calcium (Na^{+}–Ca^{2+}) exchanger protein encoded by the NCX1 gene. Measurements indicate there is an approximately 50% reduction of SERCA2 mRNA (37,42,43), expressed SR Ca^{2+}–ATPase protein level, and direct SR Ca^{2+}–ATPase uptake rate (37) during HF. There is a 55% increase in NCX1 mRNA levels, and an approximate factor of two increase in Na^{+}–Ca^{2+} exchange activity in human (43–45) and canine HF (37). There is uncertainty as to whether mRNA and expressed protein level of phospholamban is decreased (46) or unchanged (47) in human HF, and evidence that expressed protein level is decreased by a percentage equal to that of the SR Ca^{2+}–ATPase in the failing canine heart (37).

It is therefore critically important to understand how these changes in gene expression, protein levels, and current densities measured experimentally impact on the morphology of the AP and Ca^{2+} transient. In particular, it is key to know which of these changes have the greatest functional effect. To answer this question, we developed a computational model of the failing ventricular myocyte (28). The above data suggest the following *minimal* model of altered repolarization and Ca^{2+} handling in ventricular cells from the failing canine heart: (a) reduced expression of I_{K1} and I_{to1}; (b) downregulation of the SR Ca^{2+}–ATPase; and (c) upregulation of the electrogenic Na^{+}–Ca^{2+} exchanger. Since the density, but not the kinetic behavior, of each of the four transporters and ion currents comprising the minimal model appears altered in HF, we incorporated information on this altered gene and protein expression in the canine cell model by varying the density of these four membrane transporters (I_{to1}, I_{K1}, SR Ca^{2+}–ATPase, and Na^{+}–Ca^{2+} exchanger) within experimentally derived limits (37).

The model has been used to test the hypothesis that this minimal set of heart failure-induced changes can account for prolongation of AP duration, as well as decreased peak amplitude and decay rate of the Ca^{2+} transient observed in failing myocytes. Figures 3C and 3D demonstrate the ability of the model to reconstruct APs and Ca^{2+} transients measured in both normal and failing canine midmyocardial ventricular myocytes. Figure 3C shows a normal model AP (solid line), and model APs corresponding to the additive effects of sequential downregulation of I_{to1} (by 62%; dot-dashed line), I_{K1} (by 32%; long–dashed line), and SR Ca^{2+}-ATPase (by 62%; rightmost short-dashed bold line), followed by

upregulation of Na^+–Ca^{2+} exchanger (by 75%; dotted line). Changes of transporter amplitude were based on average values derived from experiments using mid-myocardial failing canine ventricular myocytes. Model simulations indicate that downregulation of I_{to1} produces a modest *shortening*, not lengthening, of AP duration. On first consideration, this seems an anomalous effect since I_{to1} is an outward K current, but is one which agrees with the experimental results of Zygmunt et al. (48) in canine myocytes (see their Fig. 2). The mechanism of this AP duration shortening has been investigated in detail using computational models (33), and results show that reduction of the Phase 1 notch depth through downregulation of I_{to1} reduces the electrical driving force on inward Ca^{2+} current and hence shortens AP duration. The additional downregulation of I_{K1} (long-dashed line) produces modest AP prolongation, consistent with the fact that outward current through I_{K1} is activated primarily at potentials which are hyperpolarized relative to the plateau potential. The most striking result is shown by the short-dashed line in Figure 3C—significant AP prolongation occurs following downregulation of SR Ca^{2+}–ATPase. This downregulation results in a near doubling of AP duration that is similar to that observed experimentally (Figure 3A). Finally, the model predicts that upregulation of Na^+–Ca^{2+} exchanger, when superimposed on these other changes, contributes to modest APD *shortening* due to reverse mode Na^+–Ca^{2+} exchange and generation of a net outward current during the plateau phase of the AP.

This modeling has provided important insights into the mechanism of AP prolongation and altered Ca^{2+} transients in heart failure. Prior to this work, the consensus was that downregulation of the genes encoding the I_{to1} and I_{K1} outward K currents was responsible for AP prolongation—a very intuitive and reasonable hypothesis. The model indicates that this is not likely to be the case. Rather, the main contributor to AP prolongation involves downregulation of the gene encoding the SR Ca^{2+}–ATPase. Subsequent model simulations have shown that downregulation of this transport process alone has a severe effect on prolongation of the AP, a prediction confirmed by experiments in which cyclopiazonic acid is used to block SR Ca^{2+}–ATPase transport (49). This modeling illustrates the value of using quantitative models to interpret the consequences of changes in gene and protein expression on cell function. It also points out how prediction of a cellular phenotype using knowledge of underlying molecular changes *must* be based on interpretations derived from quantitative experimentally based models.

2.7. A New Class of Myocyte Models

While common pool models are able to reconstruct APs with high fidelity, they are unable to reproduce a very fundamental behavior of cardiac myocytes—SR Ca^{2+} release that is smoothly and continuously graded with influx of

trigger Ca^{2+} through sarcolemmal LCCs. This failure is demonstrated in Figure 3E. The figure shows normalized peak Ca^{2+} flux through RyR channels (ordinate) as a function of membrane potential (mV; abscissa). Filled circles are experimental measurements from the work of Wier et al. (50), showing that release flux increases smoothly to a maximum flux at about 0 mV, and then decreases to near zero at more depolarized potentials. Release flux increases from –40 to 0 mV since over this potential range the open probability of LCCs increases very steeply reaching a maximum value. Release flux decreases over the potential range greater than 0 mV because the electrical driving force on Ca^{2+} decreases monotonically. The solid line shows release flux for the Jafri-Rice-Winslow guinea pig ventricular myocyte model. Release is all-or-none, with regenerative release initiated at a membrane potential causing opening of a sufficient number of LCCs (~–15 mV), and release terminating at the potential for which electrical the driving force is reduced to a critical level (~+40 mV).

This all-or-none behavior of Ca^{2+} release in common pool models has very important implications for common pool model dynamics. LCCs not only undergo voltage- but also Ca^{2+}-dependent inactivation (51,52). Inactivation depends on local subspace Ca^{2+} concentration, and occurs as Ca^{2+} binding to calmodulin (52), which is tethered to the LCC, induces the channel to switch from a normal mode of gating to a mode in which transitions to open states are extremely rare. Recent experimental data have demonstrated that voltage-dependent inactivation of LCCs is a slow and weak process, whereas Ca^{2+}-dependent inactivation is relatively fast and strong (52,53) (see Figure 4C). This implies in turn that there is a very strong coupling between Ca^{2+} release from JSR into the local subspace, and regulation of inactivation of the LCC. When this newly revealed balance between voltage- and Ca^{2+}-dependent inactivation is incorporated into common pool models, the models become unstable, exhibiting alternating short and long duration APs (31,54) (see Figure 4D). The reason for this is intuitively clear—since JSR Ca^{2+} release is all or none in these models, Ca^{2+}-dependent inactivation of LCCs is all-or-none, depending on whether release has or has not occurred. Since L-type Ca^{2+} current is a major contributor to inward current during the plateau phase of the AP, its biphasic inactivation leads to instability of AP duration. This, unfortunately, constitutes a fatal weakness of common pool models.

The fundamental failure of common pool models described above suggests that more biophysically based models of excitation–contraction coupling must be developed and investigated. Understanding of the mechanisms by which Ca^{2+} influx via LCCs triggers Ca^{2+} release from the JSR has advanced tremendously with the development of experimental techniques for simultaneous measurement of LCC currents and Ca^{2+} transients and detection of local Ca^{2+} transients, and this has given rise to the local control hypothesis of EC coupling (19,50,55,56). As illustrated schematically in Figure 4A, this hypothesis asserts that opening of an *individual* LCC in the T-tubular membrane triggers Ca^{2+} release from the

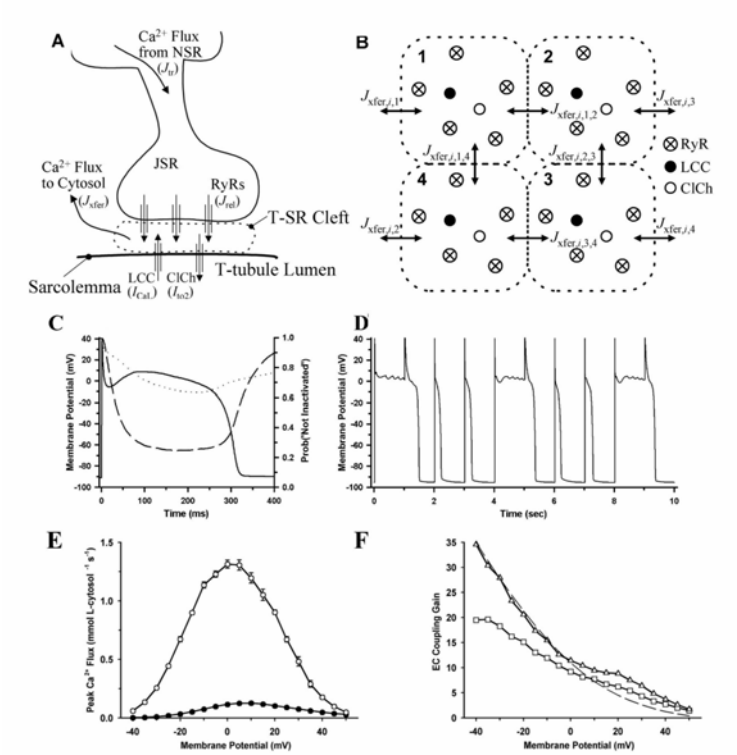


Figure 4. (**A**) Structure of the LCC-RyR complex, denoted as the functional unit (FU). A single LCC in the sarcolemmal membrane is associated with 5 RyR in the closely apposed JSR membrane. ClCh denotes a single Ca^{2+}-modulated Cl^- channel that is thought to be co-located in the dyadic space. (**B**) Structure of the Ca^{2+} release unit (CaRU). Each CaRU consists of 4 FUs, with Ca^{2+} diffusion between adjacent FUs and into the surrounding cytosolic space. (**C**) Solid line is an action potential (membrane potential in mV, left ordinate; time in msec, abscissa) predicted by the local-control myocyte model. Dotted line is the fraction of channels (right ordinate) not voltage inactivated, and the dashed line is the fraction not Ca^{2+}-inactivated during the action potential shown by the solid line. (**D**) Behavior of the common pool myocyte model when the balance between voltage- and Ca^{2+}-inactivation is as shown in panel **C**. Note the instability of action potentials. (**E**) Peak Ca^{2+} flux (ordinate) through RyRs (open symbols) and LCCs (filled symbols) as a function of membrane potential (mV, abscissa). (**F**) EC coupling gain (ordinate, ratio of peak RyR to LCC flux) as a function of membrane potential (mV, abscissa).

small cluster of RyRs located in the closely apposed (~12 nm) JSR membrane. Thus, the local control hypothesis asserts that release is all-or-none at the level of these individual groupings of LCCs and RyRs (referred to as the functional unit). However, LCC:RyR clusters are physically separated at the ends of the sarcomeres (57). These clusters therefore function in an approximately independent fashion. The local control hypothesis asserts that graded control of SR

Ca^{2+} release, in which Ca^{2+} release from JSR is a smooth continuous function of Ca^{2+} influx, is achieved by statistical recruitment of elementary Ca^{2+} release events in these independent diadic spaces. Thus, central to the local-control hypothesis is the assertion that the co-localization of LCCs and RyRs is a structural component that is fundamental to the property of graded Ca^{2+} release and force generation at the level of the cell. This concept of channel co-localization contributing in fundamental ways to cell behavior is a general theme of biophysical signal processing in excitable cells.

We have recently implemented a local-control model of myocyte function (54). As a compromise between structural and biophysical detail versus tractability, a "minimal model" of local control of Ca^{2+} release, referred to as the Ca^{2+} release unit (CaRU) model, was developed. Figure 4B shows a schematic of the CaRU model. This model is intended to mimic the properties of Ca^{2+} sparks in the T-tubule/SR (T-SR) junction (Ca^{2+} sparks are elementary SR Ca^{2+} release events arising from opening of a cluster of RyRs (58)). Figure 4B shows a cross-section of the model T-SR cleft, which is divided into four individual diadic subspace compartments arranged on a 2×2 grid. Each subspace (SS) compartment contains a single LCC and 5 RyRs in its JSR and sarcolemmal membranes, respectively. All 20 RyRs in the CaRU communicate with a single local JSR volume. The 5:1 RyR to LCC stoichiometry is chosen to be consistent with recent estimates indicating that a single LCC typically triggers the opening of 4–6 RyRs (59). Each subspace is treated as a single compartment in which Ca^{2+} concentration is uniform; however, Ca^{2+} may diffuse passively to neighboring subspaces within the same CaRU. The division of the CaRU into four subunits allows for the possibility that an LCC may trigger Ca^{2+} release in adjacent subspaces (i.e., RyR recruitment) under conditions where unitary LCC currents are large. Since LCC:RyR clusters are physically separated (57), each model CaRU is assumed to function independently of other CaRUs. Upon activation of RyRs, subspace Ca^{2+} concentration will become elevated. This Ca^{2+} will freely diffuse to either adjacent subspace compartments (J_{iss}), or into the cytosol (J_{xfer}) along its concentration gradient. The local JSR compartment is refilled via passive diffusion of Ca^{2+} from the network SR (NSR) compartment (J_{tr}).

The local control simulation algorithm is described in detail in the appendix of Greenstein and Winslow (54). Simulation of the dynamics of each CaRU requires both numerical integration of the ordinary differential equations describing local subspace and JSR Ca^{2+} balance, as well as Monte Carlo simulation of LCC and RyR channel gating in the approximately ~12,500 CaRUs in the cell (there are ~50,000 LCCs per ventricular myocyte). The state of each channel is described by a set of discrete valued random variables that evolve in time as described by Markov processes. Time steps for CaRU simulations are adaptive and are chosen to be sufficiently small based on channel transition rates. The CaRU simulations occur within the (larger) time step used for the numerical integration of the system of ordinary differential equations describing the time-

evolution of global state variables. As a result of the embedded Monte Carlo simulation, all model state variables and ionic currents/fluxes will contain a component of stochastic noise. These fluctuations introduce a degree of variability to simulation output.

Figures 4C–F show macroscopic properties of APs and SR Ca^{2+} release in this hybrid stochastic/ODE model. Figure 4C shows the relative balance between the fraction of LCCs *not* voltage-inactivated (dotted line) and *not* Ca^{2+}-inactivated (dashed line) during the AP. These fractions were designed to fit the experimental data of Linz and Meyer (53). The solid line shows a local-control model AP. This AP should be contrasted with those produced by the common pool model when the same relationship between LCC voltage- and Ca^{2+}-dependent inactivation as shown in Figure 4C is used. Clearly, the local-control model exhibits stable APs whereas the common pool model does not. Figure 4E shows the voltage dependence of peak LCC Ca^{2+} influx ($F_{LCC(max)}$ = filled circles, ordinate) and peak RyR Ca^{2+} release flux ($F_{RyR(max)}$ = open circles, ordinate) in response to voltage-clamp steps to the indicated potentials (mV, abscissa). Ca^{2+} release flux is a smooth and continuous function of membrane potential, and hence triggers Ca^{2+}, as shown by the experimental data in Figure 4D. EC coupling gain may be defined as by Wier et al. (50), as the ratio $F_{RyR(max)}/F_{LCC(max)}$, and is plotted as a function of voltage in Figure 4F (triangles). EC coupling gain is monotonically decreasing with increasing membrane potential, and agrees with corresponding experimental measurements made by Wier (50). The role of inter-subspace coupling on gain is demonstrated in Figure 4F, by comparison of control simulations (triangles) to those in the absence of inter-subspace coupling (squares). With inter-subspace coupling intact, EC coupling gain is greater at all potentials, but the increase in gain is most dramatic at more negative potentials. In this negative voltage range, LCC open probability is submaximal, leading to sparse LCC openings. However, unitary current magnitude is relatively high, so that in the presence of Ca^{2+} diffusion within the CaRU the rise in local Ca^{2+} due to the triggering action of a single LCC can recruit and activate RyRs in adjacent subspace compartments within the same T-SR junction. The net effect of inter-subspace coupling is therefore to increase the magnitude and slope of the gain function preferentially in the negative voltage range. These simulations therefore offer an intriguing glimpse of how the co-localization and stochastic gating of individual channel complexes can have a profound effect on the overall integrative behavior of the cell.

3. MODELS OF THE CARDIAC VENTRICLES

Computational models of the cardiac myocyte have contributed greatly to our understanding of myocyte function. This is in large part due to a rich interplay between experiment and modeling—an interplay in which experiments

inform modeling and modeling suggests new experiments. However, modeling of cardiac ventricular conduction has to a large extent lacked this interplay. While it is now possible to measure electrical activation of the epicardium at relatively high spatial resolution, the difficulty of measuring the geometry and fiber structure of hearts that have been electrically mapped has limited our ability to relate ventricular structure to conduction via quantitative models. As described in the following sections, we are approaching this problem by: (a) mapping ventricular activation using high-density epicardial electrode arrays; (b) measuring and modeling ventricular geometry and fiber orientation at high spatial resolution using diffusion tensor magnetic resonance imaging (DTMRI); (c) constructing computational models of the imaged hearts; and (d) comparing simulated conduction properties with those measured experimentally in the same heart. This is one approach to "closing the loop" between experiment and modeling at the whole-heart level.

3.1. Mapping of Epicardial Conduction in Canine Hearts

We have recently performed electrical mapping studies in which epicardial conduction in response to various current stimuli has been measured using electrode arrays consisting of a nylon mesh with 256 electrodes and electrode spacing of ~5 mm sewn around its surface. Bipolar epicardial twisted-pair pacing electrodes were sewn onto the right atrium (RA) and the right-ventricular (RV) free-wall. Four to ten glass beads filled with gadolinium-DTPA (~5 mM) were attached to the sock as localization markers, and responses to different pacing protocols we recorded. Figure 5A shows an example of measurement of activation time (color bar, in msec) measured in response to an RV stimulus pulse applied at the epicardial locations marked in red. After all electrical recordings are obtained, the animal is euthanized with a bolus of potassium chloride, and the heart is then scanned with high-resolution T1-weighted imaging in order to locate the gadolinium-DTPA filled beads in scanner coordinates. The heart is then excised, sock electrode locations are determined using a 3D digitizer (MicroScribe 3DLX), and the heart is formalin-fixed in preparation for DTMRI.

3.2. Measuring the Fiber Structure of the Cardiac Ventricles Using DTMRI

DTMRI is based on the principle that proton diffusion in the presence of a magnetic field gradient causes signal attenuation, and that measurement of this attenuation in several different directions can be used to estimate a diffusion tensor at each image voxel (60,61). Several studies have now confirmed that the principal eigenvector of the diffusion tensor is locally aligned with the long axis of cardiac fibers (62–64).

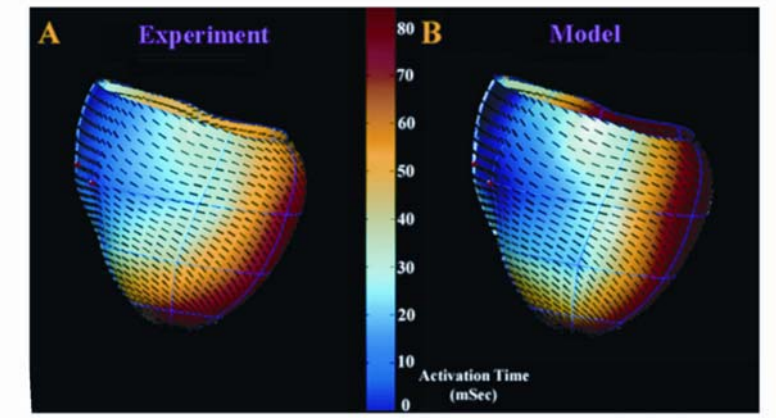


Figure 5: (**A**) Electrical activation times (indicated by color bar) in response to right RV pacing as recorded using electrode arrays. Data was obtained from a normal canine heart that was subsequently reconstructed using DTMRI. Activation times are displayed on the epicardial surface of a finite-element model fit to the DTMRI reconstruction data. Fiber orientation on the epicardial surface, as fit to the DTMRI data by the FEM model, is shown by the short line segments. (**B**) Activation times predicted using a computational model of the heart mapped in **A**.

Use of DTMRI for reconstruction of cardiac fiber orientation provides several advantages over traditional histological methods. First, DTMRI yields estimates of the absolute orientation of cardiac fibers, whereas histological methods yield estimates of only fiber inclination angle. Second, DTMRI performed using formalin-fixed tissue: (a) yields high-resolution images of the cardiac boundaries, thus enabling precise reconstruction of ventricular geometry using image segmentation software; and (b) eliminates flow artifacts present in perfused heart, enabling longer imaging times, increased signal-to-noise ratio, and improved spatial resolution. Third, DTMRI provides estimates of fiber orientation at more than one order of magnitude more points than is possible with histological methods. Fourth, reconstruction time is greatly reduced (~60 hours versus weeks to months) relative to that for histological methods.

DTMRI data acquisition and analysis for ventricular reconstruction has been semi-automated. Once image data are acquired, software written in the MatLab programming language is used to estimate epicardial and endocardial boundaries in each short-axis section of the image volume using either the method of region growing or the method of parametric active contours (65). Diffusion tensor eigenvalues and eigenvectors are computed from the DTMRI data sets at those image voxels corresponding to myocardial points, and fiber orientation at each image voxel is computed as the primary eigenvector of the diffusion tensor.

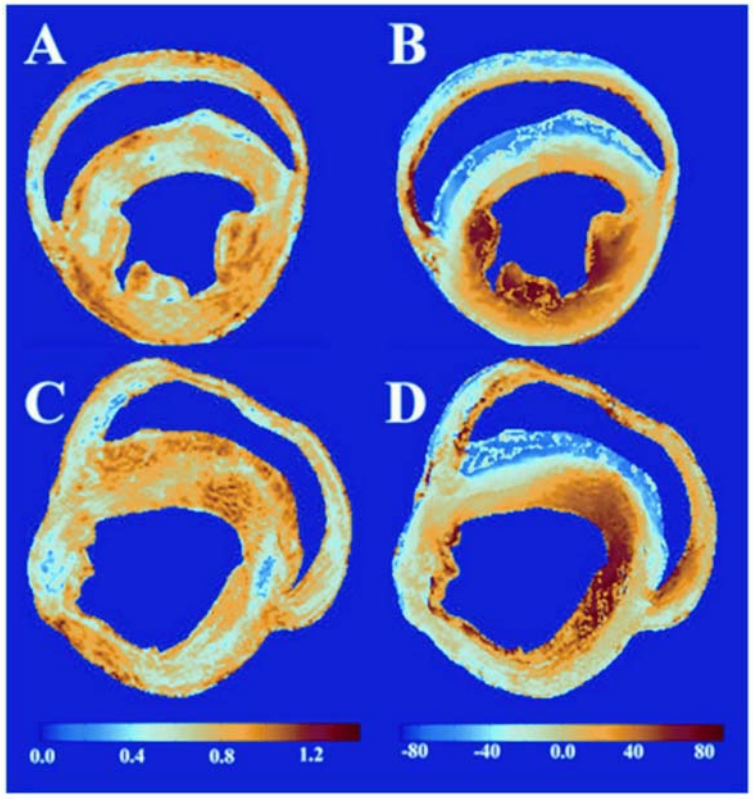


Figure 6. Fiber anisotropy A(x) estimated from DTMRI data, and calculated as $A(\underline{x}) = \sqrt{\frac{[\lambda_1(\underline{x}) - \lambda_2(\underline{x})]^2 + [\lambda_1(\underline{x}) - \lambda_3(\underline{x})]^2 + [\lambda_2(\underline{x}) - \lambda_3(\underline{x})]^2}{\lambda_1(\underline{x})^2 + \lambda_2(\underline{x})^2 + \lambda_3(\underline{x})^2}}$, where $\lambda_{1,2,3}(\underline{x})$ are diffusion tensor eigenvectors at voxel $\underline{x}$, in normal (**A**) and failing (**C**) canine heart. Fiber inclination angle computed using DTMRI in normal (**B**) and failing (**D**) heart. Panels **A** and **B** are the same normal, and panels **C** and **D** the same failing heart.

Representative results from imaging of one normal and one failing canine heart are shown in Figure 6. Figures 6A,C are short-axis basal sections taken at approximately the same level in normal (6A) and failing (6C) canine hearts. These two plots show regional anisotropy according to the indicated color code. Figures 6B,D show the angle of the primary eigenvector relative to the plane of section (inclination angle), according to the indicated color code, for the same sections as in Figures 6A,C. Inspection of these data show: (a) the failing heart

(HF: panels C,D) is dilated relative to the normal heart (N: panels A,B); b) left-ventricular (LV) wall thinning (average LV wall thickness over 3 hearts is 17.5 ± 2.9 mm in N, and 12.9 ± 2.8 mm in HF); (c) no change in RV wall thickness (average RV wall thickness is 6.1 ± 1.6 mm in N, and 6.3 ± 2.1 mm in HF); (d) increased septal wall thickness HF versus N (average septal wall thickness is 14.7 ± 1.2 mm N, and 19.7 ± 2.1 mm HF); (e) increased septal anisotropy in HF versus N (average septal thickness is 0.71 ± 0.15 N, and 0.82 ± 0.15 HF); and (f) changes in the transmural distribution of septal fiber orientation in HF versus N (contrast panels B,D, particularly near the junction of the septum and RV).

3.3. Finite-Element Modeling of Cardiac Ventricular Anatomy

The structure of the cardiac ventricles can modeled be using the finite-element modeling (FEM) methods developed by Nielson et al. (66). The geometry of the heart to be modeled is described initially using a predefined mesh with 6 circumferential elements and 4 axial elements. Elements use a cubic Hermite interpolation in the transmural coordinate (η), and bilinear interpolation in the longitudinal (μ) and circumferential (θ) coordinates. Voxels in the 3D DTMR images identified as being on the epicardial and endocardial surfaces by the semi-automated contouring algorithms described above are used to deform this initial FEM template. Deformation of the initial mesh is performed to minimize an objective function $F(\underline{n})$:

$$F(\underline{n}) = \sum_{d=1}^{D} \gamma_d \left\| v(\underline{\varepsilon}_d) - v_d \right\|^2 + \int_{\Re^2} \{\alpha \nabla^2 \underline{n} + \beta (\nabla^2 \underline{n})^2\} \partial \underline{\varepsilon}\,, \qquad [14]$$

where $\underline{n}$ is a vector of mesh nodal values, v_d are the surface voxel data, $v(\varepsilon_d)$ are the projections of the surface voxel data on the mesh, and α and β are user-defined constants. This objective function consists of two terms. The first describes the distance between each surface image voxel (v_d) and its projection onto the mesh $v(\varepsilon_d)$. The second, known as the weighted Sobelov norm, limits the stretching (first-derivative terms) and bending (second-derivative terms) of the surface. The parameters α and β control the degree of deformation of each element. The weighted Sobelov norm is particularly useful in cases where there is an uneven distribution of surface voxels across the elements. A linear least-squares algorithm is used to minimize this objective function.

After the geometric mesh is fit to DTMRI data, the fiber field is defined for the model. Principal eigenvectors lying within the boundaries of the mesh computed above are transformed into the local geometric coordinates of the model using the following transformation.

$$\underline{V}G = [\underline{F}\ \underline{G}\ \underline{H}]^T\ [R]\underline{V}_S, \qquad [15]$$

where R is a rotation matrix that transforms a vector from scanner coordinates ($\underline{V}_S$) into the FEM model coordinates $\underline{V}_G$ and F, G, and H are orthogonal geometric unit vectors computed from the ventricular geometry as described by Le Grice et al. (67). Once the fiber vectors are represented in geometric coordinates, DTMRI inclination and imbrication angles (α and ϕ) are fit using a bilinear interpolation in the local ε_1 and ε_2 coordinates, and a cubic Hermite interpolation in the ε_3 coordinate. A graphical user interface for fitting FEMs to both the ventricular surfaces and fiber field data has been implemented using the MatLab programming language. Figure 7 shows FEM fits to the epi- and endocardial surfaces of a reconstructed normal canine heart obtained using this software tool. FEM fits to the fiber orientation data are shown on these surfaces as short line segments. We have also developed relational database and data analysis software, named *HeartScan*, to facilitate analysis of cardiac structural and electrical data sets obtained from populations of hearts. HeartScan enables users

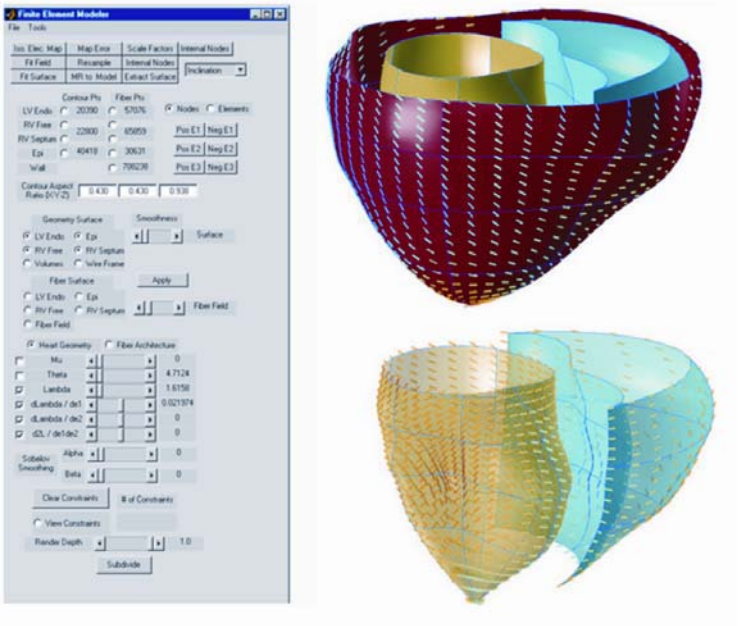


Figure 7. Finite-element model of canine ventricular anatomy showing the epicardial (red), LV endocardial (green), and RV endocardial surfaces. Fiber orientation on each surface is shown by short line segments.

to pose queries (in standard query language, or SQL) on a wide range of cardiac data sets by means of a graphical user interface (Figure 8). These data sets include: (a) DTMRI imaging data; (b) FEMs derived from DTMRI data; (c) electrical mapping data obtained using epicardial electrode arrays; and (d) model simulation data. Query results are either: (a) displayed on a 3D graphical representation of the heart being analyzed; or (b) piped to data-processing scripts, the results of which are then displayed visually. Queries may be posed by direct entry of an SQL command into the Query Window (Figure 8B). This query is executed, and the set of points satisfying this condition are displayed on a wire frame model of the heart being studied (shown in green in Figure 8C). Queries operating on a particular region of the heart may also be entered by graphically selecting that region (Figure 8D). SQL commands specifying the coordinates of the selected voxels are then automatically entered into the Query Window. One example of such a predefined operation is shown in Figure 8E, which shows computation of transmural inclination angle for the region enclosed by the box in Figure 8D.

3.4. Generation of Computational Models from DTMRI Data

The bidomain equations describe the flow of electrical current within the myocardium, between the intracellular and extracellular domains. This approach treats each domain of the myocardial tissue as a continuum, rather than as being composed of discrete cells connected by gap junctions and surrounded by the extracellular milieu. Thus, quantities such as conductivity and transmembrane voltage represent spatial averages. Several excellent reviews have been published detailing the assumptions in, structure of, and solution methods for the bidomain equations (68–70). The following is a brief review of the origins of these equations.

The bidomain equations are derived by applying conservation of current between the intra- and extracellular domains. The equations consist of parabolic Eq. [16] and elliptic Eq. [17], equations that must be satisfied within the myocardium, (a region designated as H):

$$\frac{\partial \nu}{\partial t}(\underline{x},t) = \frac{1}{C_m}\left[-I_{ion}(\underline{x},t) - I_{app}(\underline{x},t) - \frac{1}{\beta}\left(\nabla \cdot M_e(\underline{x})\nabla\phi_e(\underline{x},t)\right)\right], \quad \forall\ \underline{x} \text{ in } H, \quad [16]$$

$$\nabla \cdot M_i(\underline{x})\nabla v(\underline{x},t) = -\nabla \cdot M(\underline{x})\nabla\phi_e(\underline{x},t), \ \forall\ \underline{x} \text{ in } H, \quad [17]$$

and an additional elliptic equation, Eq. [18], that must be satisfied in the bath or tissue surrounding the heart (a region designated as B):

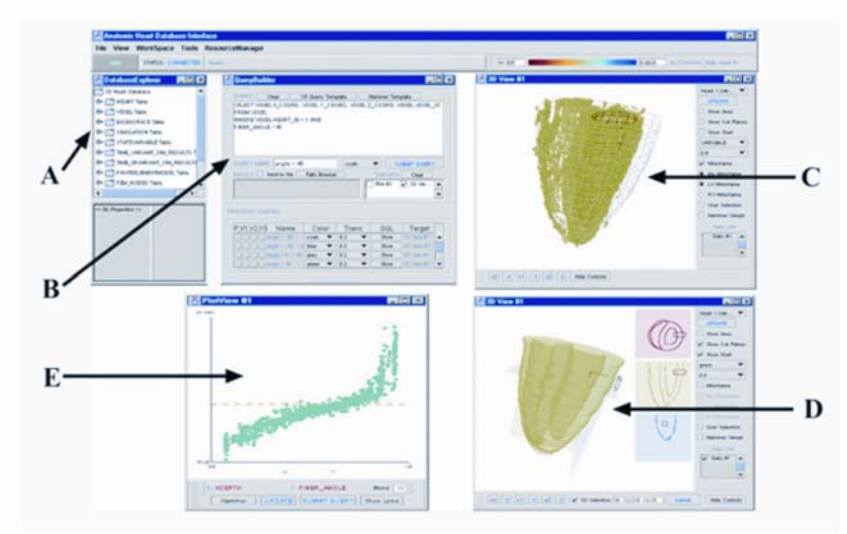


Figure 8. "Screenshot" of the windows by which the user interacts with HeartScan. (**A**) window for viewing data tables; (**B**) SQL query window; (**C**) window for interactive 3D display of heart data; (**D**) pull-down window for user selection of heart regions to query; (**E**) statistics display window.

$$\nabla \cdot M_b(\underline{x}) \nabla \phi_b(\underline{x},t) = 0, \ \forall \ \underline{x} \text{ in } B, \qquad [18]$$

where $\underline{x}$ is spatial position, $\phi_i(\underline{x},t)$ and $\phi_e(\underline{x},t)$ are the transmembrane intra- and extracellular potentials, respectively; $v(\underline{x},t) = \phi_i(\underline{x},t) - \phi_e(\underline{x},t)$ is the transmembrane voltage; C_m is the membrane capacitance per unit area; $I_{\text{ion}}(\underline{x},t)$ is the sum of the ionic currents per unit area through the membrane (positive outward); $I_{\text{app}}(\underline{x},t)$ is an applied cathodal extracellular current per unit area; β is the ratio of membrane area to tissue volume; $M_e(\underline{x})$ and $M_i(\underline{x})$ are the extracellular and intracellular conductivity tensors, with $M(\underline{x}) = M_e(\underline{x}) + M_i(\underline{x})$; $\phi_b(\underline{x},t)$ is the bath potential; and $M_b(\underline{x})$ is the bath conductivity tensor. These parameters may be set, in models of the normal heart, using values described by Pollard et al. (68) and Henriquez et al. (70). Additionally, boundary conditions on the interface between the heart and the surrounding tissue, δH, and the body surface, δB, must be specified. The first boundary condition specifies continuity of potential:

$$\phi_e = \phi_b \ \text{ on } \ \delta H. \qquad [19]$$

The second specifies continuity of current at the interface:

$$\sigma_i \frac{\partial \phi_i}{\partial n} + \sigma_e \frac{\partial \phi_e}{\partial n} = \sigma_b \frac{\partial \phi_b}{\partial n} \quad \text{on} \quad \delta H, \tag{20}$$

where the σ's are the conductivities normal to the interface, and the $\partial/\partial n$ is the normal derivative operator. In order for the problem to be well posed, a third boundary condition on δH is required. While the first two follow necessarily for all electrical phenomena, there are a number of ways to formulate the third boundary condition. Typically, we specify:

$$\sigma_i \frac{\partial \phi_i}{\partial n} = 0 \quad \text{on} \quad \delta H, \tag{21}$$

which has the physical interpretation that at the heart/body interface all intracellular current must flow first through the extracellular space before it flows into the surrounding tissue. A boundary condition at δB for the Laplace equation in ϕ_b is also required. Given that air is a poor conductor, this is simply

$$\sigma_b \frac{\partial \phi_b}{\partial n} = 0 \quad \text{on} \quad \delta B. \tag{22}$$

Finally, an initial condition on the transmembrane voltage must be specified, $v(\underline{x},t=0) = V(\underline{x})$. Then from this, initial conditions on $\phi_e(\underline{x},t=0)$ and $\phi_b(\underline{x},t=0)$ can be found by solving the appropriate elliptic equation. Equations [16]–[22] specify the bidomain problem.

Under some restrictive assumptions, the bidomain equations can be simplified dramatically. If the surrounding tissue is taken to be a good insulator, then $\sigma_b = 0$ in B. Then we have

$$\phi_e = 0 \quad \text{on } \delta H, \tag{23}$$

$$\frac{\partial \phi_e}{\partial \eta} = 0 \quad \text{on} \quad \delta H, \tag{24}$$

$$\frac{\partial \phi_i}{\partial n} = 0 \quad \text{on} \quad \delta H, \tag{25}$$

and the Laplace equation for ϕ_b need not be considered. Additionally, under the assumption of equal anisotropy, namely, that

$$M_i(\underline{x}) = \frac{1}{\kappa} M_e(\underline{x}), \qquad [26]$$

where k is called the anisotropy ratio, Eqs. [16] and [17] uncouple, then requiring only solution of the parabolic equation

$$\frac{\partial \nu}{\partial t}(\underline{x},t) = \frac{1}{C_m}\left[-I_{ion}(\underline{x},t) - I_{app}(\underline{x},t) + \frac{1}{\beta}\left(\frac{\kappa}{\kappa+1}\right)\nabla\cdot(M_i(\underline{x})\nabla\nu(\underline{x},t))\right] \text{ on } H. \qquad [27]$$

Equation [27] is referred to as the monodomain equation.

The conductivity tensors at each point within the heart are specified by fiber orientation and by specific conductivities in each of the local coordinate directions. The conductivity tensor in the local coordinate system, $G_i(x)$, is defined as

$$G_i(\underline{x}) = \begin{bmatrix} \sigma_{1,i} & & \\ & \sigma_{2,i} & \\ & & \sigma_{3,i} \end{bmatrix}, \qquad [28]$$

where $\sigma_{1,i}$ is the longitudinal and $\sigma_{2,i}$ and $\sigma_{3,i}$ are the transverse intracellular conductivities, respectively. This local tensor may be expressed in global coordinates to give the conductivity tensor of Eq. [27] using the transformation

$$M_i(x) = P(x)G(x)P^T(x), \qquad [29]$$

where $P(x)$ is the coordinate transformation matrix from local to global coordinates. $P(x)$ is in turn determined by the underlying fiber organization of the heart, and is obtained using DTMRI as described in §3.2. If only fiber direction information is available, then it is appropriate that conductivities transverse to the fiber long-axis be assumed equal ($\sigma_{2,i} = \sigma_{3,i}$).

Figure 5B shows the results of applying these methods to the analysis of conduction in a normal canine heart. As described previously, Figure 5A shows activation time (color bar, in msec) measured experimentally in response to an RV stimulus pulse applied at the epicardial locations marked in red. Following electrical mapping, this heart was excised, imaged using DTMRI, and an FEM was then fit to the resulting geometry and fiber orientation data sets. Figure 5A shows activation time displayed on this FEM. The stimulus wavefront can be seen to follow the orientation of the epicardial fibers, which is indicated by the dark line segments in Figure 5A. Figure 5B shows results of simulating conduction using a computational model of the very same heart that was mapped electrically in Figure 5A. Results can be seen to agree qualitatively; however, model conduction is more rapid in the region where the RV and LV join. Nonetheless,

these results demonstrate the feasibility of combined experimentation and modeling of electrical conduction in specific imaged and reconstructed hearts.

4. DISCUSSION AND CONCLUSIONS

This chapter has reviewed modeling research in three broad areas: (1) models of single ventricular myocytes; (2) methods for the reconstruction and modeling of ventricular geometry and microanatomy; and (3) integrative modeling of the cardiac ventricles. We have seen that the level of biophysical detail, and hence the accuracy and predictability of current ventricular myocyte models, is considerable. Nonetheless, much remains to be done.

One emerging area of research is modeling of mitochondrial energy production. Approximately 2% of cellular ATP is consumed on each heartbeat. The major processes consuming ATP in the myocyte are muscle contraction, activity of the SR Ca^{2+}–ATPase, and Na–K pumping. Cellular ATP levels also influence ion channel function including the sarcolemmal ATP-modulated K channel (71). Recently, we have formulated an integrated thermokinetic model of cardiac mitochondrial energetics comprising the tricarboxylic acid (TCA) cycle, oxidative phosphorylation and mitochondrial Ca^{2+} handling (72). This model describes the dynamics of the key regulatory effectors of TCA cycle enzymes and the production of NADH and $FADH_2$. These molecules are used by the electron transport chain to establish a proton motive force ($\Delta\mu_H$), which then drives the F_1F_0–ATPase. Mitochondrial matrix Ca^{2+} is also a model state variable. Mitochondrial Ca^{2+} concentration is determined by the Ca^{2+} uniporter and Na^+/Ca^{2+} exchanger activities, and regulates activity of the TCA cycle enzymes isocitrate dehydrogenase (IDH) and α-ketoglutarate dehydrogenase (KGDH). The model is described by twelve ordinary differential equations that represent $\Delta\Psi_m$ (mitochondrial membrane potential) and matrix concentrations of Ca^{2+}, NADH, ADP, and TCA cycle intermediates. The model is able to reproduce experimental data concerning mitochondrial bioenergetics, Ca^{2+} dynamics and respiratory control, relying only on the fundamental properties of the system. The time-dependent behavior of the model, under conditions simulating an increase in workload, closely reproduce the experimentally observed mitochondrial NADH dynamics in heart trabeculae subjected to changes in pacing frequency. The steady-state and time-dependent behavior of the model support the role of mitochondrial matrix Ca^{2+} in matching energy supply with demand in cardiac cells. Further development and testing of this model, its integration into models of the myocyte, and the use of these models to investigate myocyte responses to ischemia, are required.

In real cardiac myocytes, there exist a diversity of mechanisms that act to modulate cellular excitability. This includes α- and β-adrenergic signaling pathways acting through G protein-coupled membrane receptors to modulate the

properties of LCCs, various K^+ channels, and Ca^{2+} transporters such as the SR Ca^{2+}–ATPase. The addition of these modulatory mechanisms to the cell models remains an important goal for the future.

As we have shown, magnetic resonance imaging now offers a relatively rapid way to measure ventricular fiber structure at high spatial resolution. The ability to rapidly acquire fiber orientation data throughout the ventricles in large populations of normal and diseased hearts will enable quantitative statistical comparison of normal and abnormal cardiac structure, and will provide insights into the possible structural basis of arrhythmia in heart disease. Unfortunately, a detailed understanding of the spatial heterogeneities within the heart, such as variation of intercellular coupling, regional expression of ionic currents and Ca^{2+} handling proteins is still unavailable, although significant progress has certainly been made. Understanding and modeling of these spatial heterogeneities remains a challenge for the future.

Finally, the complexity of biological models, including those of the cardiac myocyte, is increasing rapidly. This complexity makes the reliable publication and exchange of models difficult. XML-based markup languages such as CellML (73) and the Systems Biology Markup Language (SBML) (74) are being developed to support error-free exchange of models independently of the hardware and software architectures on which these models will run. An application programming interface for CellML is being developed, and several groups are developing software for automated source code generation from CellML files.

These are indeed exciting times for cardiovascular biology. A national infrastructure supporting the acquisition, distribution and analysis of cardiovascular genomic and proteomic data is now in the formative stage (in particular, the Programs for Genomic Applications and Innovative Proteomics Centers supported by the National Heart, Lung and Blood Institute of the National Institutes of Health). The data and models produced from these efforts will without question enhance our understanding of myocyte and whole-heart function in both health and disease (see the following chapter 3.3 by Glass). Major challenges in data collection, representation, storage, dissemination, and modeling remain. If these challenges are met, we will have the opportunity to create a truly integrated cardiovascular *research community*, the whole of which is far greater than the sum of its parts.

5. ACKNOWLEDGMENTS

This work was supported by grants from the NIH (RO1 HL-61711, RO1 HL-60133, RO1 HL-72488, P50 HL-52307, N01 HV-28180), The Falk Medical Trust, The Whitaker Foundation, and IBM Corporation.

6. REFERENCES

1. Noble D. 1960. Cardiac action and pace maker potentials based on the Hodgkin-Huxley equations. *Nature* **188**:495.
2. Noble D, 1962. A modification of the Hodgin-Huxley equations applicable to Purkinje fiber action and pacemaker potentials. *J Physiol* **160**:317.
3. Moe GK, Mendez C. 1966. Simulation of impulse propagation in cardiac tissue. *Ann NY Acad Sci* **128**(3):766–771.
4. Marban E, Yamagishi T, and Tomaselli GF. 1998. Structure and function of voltage-gated sodium channels. *J Physiol (Lond)* **508**(pt 3):647–657.
5. Kemp TF, Hell KT. 2000. Regulation of cardiac L-type calcium channels by protein kinase A and protein kinase C. *Circ Res* **87**:1095.
6. Chay TR. 1991. The Hodgkin-Huxley Na+ channel model versus the five-state Markovian model. *Biopolymers* **31**(13):1483–1502.
7. Shannon TR, Ginsburg KS, Bers DM. 1998. Reverse mode of the SR Ca pump limits SR Ca uptake in permeabilized and voltage clamped myocytes. *Ann NY Acad Sci* **853**:350–352.
8. DiFrancesco D, Noble D. 1985. A model of cardiac electrical activity incorporating ionic pumps and concentration changes. *Phil Trans Roy Soc B* **307**:353–398.
9. Luo C, Rudy Y. 1991. A model of the ventricular cardiac action potential: depolarization, repolarization and their interaction. *Circ Res* **68**:1501–1526.
10. Luo CH, Rudy Y. 1994. A dynamic model of the cardiac ventricular action potential, I: simulations of ionic currents and concentration changes. *Circ Res* **74**:1071–1096.
11. Noble DS, Noble SJ, Bett CL, Earm YE, Ko WK, So IK. 1991. The role of sodium–calcium exchange during the cadiac action potential. *Ann NY Acad Sci* **639**:334–354.
12. Jafri S, Rice JJ, Winslow RL 1998. Cardiac Ca2+ dynamics: the roles of ryanodine receptor adaptation and sarcoplasmic reticulum load. *Biophys J* **74**:1149–1168.
13. Wagner MB, Golod D, Wilders R, Verheijck EE, Joyner RW, Kumar R, Jongsma HJ, Van Ginneken AC, Goolsby WN. 1997. Modulation of propagation from an ectopic focus by electrical load and by extracellular potassium. *Am J Physiol* **272**(4, pt 2):H1759–H1769.
14. Noble D, Noble SJ. 1984. A model of sino-atrial node electrical activity based on a modification of the DiFrancesco-Noble equations. *Proc R Soc Lond B Biol Sci* **222**(1228):295–304.
15. Demir SS, Clark JW, Murphey CR, Giles WR 1994. A mathematical model of a rabbit sinoatrial node cell. *Am J Physiol* **266**(3, pt 1):C832–C852.
16. Dokos S, Celler B, Lovell N. 1996. Ion currents underlying sinoatrial node pacemaker activity: a new single cell mathematical model. *J Theor Biol* **181**(3):245–272.
17. Courtemanche M, Ramirez RJ, Nattel S. 1998. Ionic mechanisms underlying human atrial action potential properties: insights from a mathematical model. *Am J Physiol Heart Circ Physiol* **275**(1):H301–H321.
18. Nygren A, Fiset C, Firek L, Clark JW, Lindblad DS, Clark RB, Giles WR. 1998. Mathematical model of an adult human atrial cell: the role of K+ currents in repolarization. *Circ Res* **82**(1):63–81.
19. Stern M. 1992. Theory of excitation–contraction coupling in cardiac muscle. *Biophys J* **63**:497–517.
20. Stern M, Song L, Cheng H, Sham J, Yang H, Boheler K. Rios E. 1999. Local control models of cardiac excitation–contraction coupling: a possible role for allosteric interactions between ryanodine receptors. *J Gen Physiol* **113**(3):469–489.
21. Rice JJ, Jafri MS, Winslow RL. 2000. Modeling short-term interval-force relations in cardiac muscle. *Am J Physiol* **278**:H913.
22. Hodgkin AL, Huxley AF. 1952. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiol* **117**:500–544.

23. Winslow R, Greenstein J, Tomaselli G, O'Rourke B. 2001. Computational model of the failing myocyte: Relating altered gene expression to cellular function. *Phil Trans Roy Soc Lond A* **359**:1187–1200.
24. Hille B. 1992. *Ionic channels of excitable membranes*. Sinauer, Sundeland, MA.
25. Horn R, Vandenberg CA. 1984. Statistical properties of single sodium channels. *J Gen Physiol* **84**(4):505–534.
26. Liu S, Rasmusson RL. 1997. Hodgkin-Huxley and partially coupled inactivation models yield different voltage dependence of block. *Am J Physiol* **272**(4, pt 2):H2013–H2022.
27. Irvine L. 1998. *Models of the cardiac Na channel and the action of lidocaine*. Department of Biomedical Engineering. The Johns Hopkins University School of Medicine, Baltimore.
28. Irvine L, Jafri MS, Winslow RL. 1999. Cardiac sodium channel Markov model with temperature dependence and recovery from inactivation. *Biophys J* **76**:1868–1885.
29. Clay JR, DeFelice LJ. 1983. Relationship between membrane excitability and single channel open–close kinetics. *Biophys J* **42**:151–157.
30. Corana A, Marchesi M, Martini C, Ridella S. 1987. Minimizing multimodal functions of continuous variables with the "simulated annealing" algorithm. *ACM Trans Math Software* **13**(3):262–280.
31. Mazhari R, Greenstein JL, Winslow RL, Marban E, Nuss HB. 2001. Molecular interactions between two long-QT syndrome gene products, HERG and KCNE2, rationalized by *in vitro* and *in silico* analysis. *Circ Res* **89**(1):33–38.
32. Mazhari R, Nuss HB, Winslow RL, Marban E. 2002. Ectopic expression of KCNE3 in the heart accelerated cardiac repolarization: a novel approach to gene therapy for long QT syndrome. *J Clin Invest* **109**(8):1083–1090.
33. Greenstein J, Po S, Wu R, Tomaselli G, Winslow RL. 2000. Role of the calcium-independent transient outward current Ito1 in action potential morphology and duration. *Circ Res* **87**:1026.
34. Wagner J, Keizer J. 1994. Effects of rapid buffers on Ca2+ diffusion and Ca2+ oscillations. *Biophys J* **67**(1).
35. Beuckelmann DJ, Nabauer M, Erdmann E. 1992. Intracellular calcium handling in isolated ventricular myocytes from patients with terminal heart failure. *Circulation* **85**:1046–1055.
36. Kaab S, Nuss HB, Chiamvimonvat N, O'Rourke B, Pak PH, Kass DA, Marban E, Tomaselli GF. 1996. Ionic mechanism of action potential prolongation in ventricular myocytes from dogs with pacing-induced heart failure. *Circ Res* **78**(2):262–273.
37. O'Rourke B, Peng LF, Kaab S, Tunin R, Tomaselli GF, Kass DA, Marban E. 1999. Mechanisms of altered excitation–contraction coupling in canine tachycardia-induced heart: experimental studies. *Circ Res* **84**:562–570.
38. Beuckelmann DJ, Nabauer M, Erdmann E. 1993. Alterations of K^+ currents in isolated human ventricular myocytes from patients with terminal heart failure. *Circ Res* **73**:379–385.
39. Näbauer M, Beuckelmann DJ, Überfuhr P, Steinbeck G. 1996. Regional differences in current density and rate-dependent properties of the transient outward current in subepicardial and subendocardial myocytes of human left ventricle. *Circulation* **93**:168–177.
40. Dixon JE, Shi W, Wang H-S, McDonald C, Yu H, Wymore RS, Cohen IS, McKinnon D. 1996. Role of the Kv4.3 K+ Channel in Ventricular Muscle: A Molecular Correlate for the Transient Outward Current. *Circ Res* **79**:659–668.
41. Kaab S, Dixon J, Duc J, Ashen D, Näbauer M, Beuckelmann DJ, Steinbeck D, McKinnon D, Tomaselli GF. 1998. Molecular basis of transient outward potassium current downregulation in human heart failure: a decrease in Kv4.3 mRNA correlates with a reduction in current density. *Circulation* **98**:1383–1393.
42. Takahashi T, Allen PD, Lacro RV, Marks AR, Dennis AR, Schoen FJ, Grossman W, Marsh JD Izumo S. 1992. Expression of dihydropyridine receptor (Ca2+ channel) and calsequestrin genes in the myocardium of patients with end-stage heart failure. *J Clin Invest* **90**:927–935.

43. Studer R, Reinecke H, Bilger J, Eschenhagen T, Bohm M, Hasenfuss G, Just H, Holtz J, Drexler H. 1994. Gene expression of the Na+–Ca2+ exchanger in end-stage human heart failure. *Circ Res* **75**:443–453.
44. Reinecke H, Studer R, Vetter R, Holtz J, Drexler H. 1996. Cardiac Na/Ca exchange activity in patients with end-stage heart failure. *Cardiovasc Res* **31**:48–54.
45. Reinecke H, Studer R, Vetter R, Just H, Holtz J, Drexler H. 1996. Role of the cardiac sarcolemmal Na+–Ca2+ exchanger in end-stage human heart failure. *Ann NY Acad Sci* **779**:543–545.
46. Arai M, Alpert NR, MacLennan DH, Barton P, Periasamy M. 1993. Alterations in sarcoplasmic reticulum gene expression in human heart failure: A possible mechanism for alterations in systolic and diastolic properties of the failing myocardium. *Circ Res* **72**:463–469.
47. Meyer M, Schillinger W, Pieske B, Holubarsch C, Heilmann C, Posival H, Kuwajima G, Mikoshiba K, Just H, Hasenfuss G. 1995. Alterations of sarcoplasmic reticulum proteins in failing human dilated cardiomyopathy. *Circulation* **92**:778–784.
48. Zygmunt AC, Robitelle DC, Eddlestone GT. 1997. Ito1 dictates behavior of ICl(Ca) during early repolarization of canine ventricle. *Am J Physiol* **273**:H1096–H1106.
49. Winslow RL, Scollan DF, Holmes A, Yung CK, Zhang J, Jafri MS. 2000. Electrophysiological modeling of cardiac ventricular function: from cell to organ. *Ann Rev Biomed Eng* **2**:119–155.
50. Wier WG, Egan TM, Lopez-Lopez JR, Balke CW. 1994. Local control of excitation–contraction coupling in rat heart cells. *J Physiol* **474**(3):463–471.
51. Bers DM, Perez-Reyes E. 1999. Ca channels in cardiac myocytes: structure and function in Ca influx and intracellular Ca release. *Cardiovasc Res* **42**(2):339–360.
52. Peterson B, DeMaria C, Adelman J, Yue D. 1999. Calmodulin is the Ca^{2+} sensor for Ca^{2+}-dependent inactivation of L-type calcium channels. *Neuron* **1999**:549–558.
53. Linz KW, Meyer R. 1998. Control of L-type calcium current during the action potential of guinea-pig ventricular myocytes. *J Physiol (Lond)* **513**(pt 2):425–442.
54. Greenstein JL, Winslow RL. 2002. An integrative model of the cardiac ventricular myocyte incorporating local control of Ca(2+) release. *Biophys J* **83**(6):2918–2945.
55. Bers DM, 1993. *Excitation contraction coupling and cardiac contractile force*. Cardiovascular medicine, Vol. 122. Kluwer Academic Press, Boston.
56. Sham, JSK. 1997. Ca2+ release-induced inactivation of Ca2+ current in rat ventricular myocytes: evidence for local Ca2+ signalling. *J Physiol* **500**(2):285–295.
57. Franzini-Armstrong C, Protasi F, Ramesh V. 1999. Shape, size, and distribution of Ca(2+) release units and couplons in skeletal and cardiac muscles. *Biophys J* **77**(3):1528–1539.
58. Cheng H, Lederer WJ, Cannell MB. 1993. Calcium sparks: Elementary events underlying excitation–contraction coupling in heart muscle. *Science* **262**:740–744.
59. Wang SQ, Song LS, Lakatta EG, Cheng H. 2001. Ca2+ signalling between single L-type Ca2+ channels and ryanodine receptors in heart cells. *Nature* **410**(6828):592–596.
60. Basser P, Mattiello J, LeBihan D. 1994. Estimation of the effective self-diffusion tensor from the NMR spin echo. *J Magn Reson B* **103**(3):247–254.
61. Skejskal EaT. 1965. Spin diffusion measurement: spin echoes in the presence of time-dependent field gradients. *J Chem Phys* **69**:1748–1754.
62. Holmes AA, Scollan DF, Winslow RL. 2000. Direct histological validation of diffusion tensor MRI in formaldehyde-fixed myocardium. *Magn Reson Med* **44**:157–161.
63. Hsu EW, Muzikant AL, Matulevicius SA, Penland RC, Henriquez CS. 1998. Magnetic resonance myocardial fiber-orientation mapping with direct histological correlation. *Am J Physiol* **274**:H1627–H1634.
64. Scollan D, Holmes A, Winslow RL, Forder J. 1998. Histologic validation of reconstructed myocardial microstructure from high-resolution MR diffusion tensor imaging. *Am J Physiol* **275**:H2308–H2318.

65. Scollan D, Zhang J, Holmes A, Yung C, Winslow R. 2000. Reconstruction of cardiac ventricular geometry and fiber orientation using GRASS and diffusion-tensor magnetic resonance imaging. *Annals BME* **28**(8):934–944.
66. Nielsen PM, Le Grice IJ, Smaill BH, Hunter PJ. 1991. Mathematical model of geometry and fibrous structure of the heart. *Am J Physiol* **260**(4, pt 2):H1365–H1378.
67. Le Grice IJ, Hunter PJ, Smaill BH. 1997. Laminar structure of the heart: a mathematical model. *Am J Physiol* **272**(5, pt 2):H2466–H2476.
68. Pollard A, Burgess MJ, Spitzer KW. 1993. Computer simulations of three-dimensional propagation in ventricular myocardium. *Circ Res* **72**:744–756.
69. Pollard AE, Hooke N, Henriquez CS. 1993. Cardiac propagation simulation. In *High performance computing in biomedical research*, pp. 289–317. Ed. Pilkington TC, Loftis B, Thompson JF, Woo SLY, Plamer TC, Budinger TF. CRC Press, Boca Raton, 1993.
70. Henriquez CS. 1993. Simulating the electrical behavior of cardiac tissue using the bidomain model. *Crit Rev Biomed Eng* **21**(1):1–77.
71. Nichols CG, Lederer WJ. 1990. The regulation of ATP-sensitive K channel activity in intact and permeabilized rat ventricular cells. *J Physiol* **423**:91–110.
72. Cortassa S, Aon M, Marban E, Winslow R, and O'Rourke B. 2003. An integrated model of cardiac mitochondrial energy metabolism and calcium dynamics. *Biophys J* **84**:2734–2755.
73. Bestetti RB, Moreira-Neto F, Brasil JC, Bombonato R, Sgarbieri RN, Haddad J. 1998. Partial left ventriculectomy: preoperative risk factors for perioperative mortality. *Intl J Cardiol* **67**(2):143–146.
74. Blanco G, Mercer RW. 1998. Isozymes of the Na–K-ATPase: heterogeneity in structure, diversity in function. *Am J Physiol* **275**(5, pt 2):F633–F650.
75. Tomaselli GF and Marban E. 1999. Electrophysiological remodeling in hypertrophy and heart failure. *Circ Res* **42**:270–283.

3.3

CARDIAC OSCILLATIONS AND ARRHYTHMIA ANALYSIS

Leon Glass

Isadore Rosenfeld Chair in Cardiology and Department of Physiology, McGill University, Montreal, Quebec

In current medical practice, the diagnosis and treatment of cardiac arrhythmias in people is carried out without mathematical analysis of the underlying mechanisms of the underlying rhythm. In this article I describe how nonlinear dynamics is being used to formulate mathematical models of cardiac arrhythmias, and to demonstrate the ways the mathematics can be used to predict the changes in rhythms that occur as physiological parameters vary. In spatially heterogeneous cardiac tissue culture, a number of different patterns of spatiotemporal activity can be found, including propagating plane waves, rotating spiral waves, and spiral waves that spontaneously initiate and terminate. These paroxysmal patterns are similar to the paroxysmal rhythms observed during cardiac arrhythmias in people. Mathematical analyses of cardiac arrhythmias can be used to determine automatically if certain arrhythmias, such as atrial fibrillation, are present in individuals. Attempts are also underway to develop new methods to analyze normal and abnormal cardiac activity in patients to better assess the risk of fatal arrhythmias before they occur.

1. INTRODUCTION

Over the course of our lives, our hearts will beat approximately 2×10^9 times. Although we have the impression that the rhythm is quite regular, there is

Address correspondence to: Leon Glass, Department of Physiology, 3655 Promenade Sir William Osler, McGill University, Montreal, Quebec, Canada H3G 1Y6 (glass@cnd.mcgill.ca)

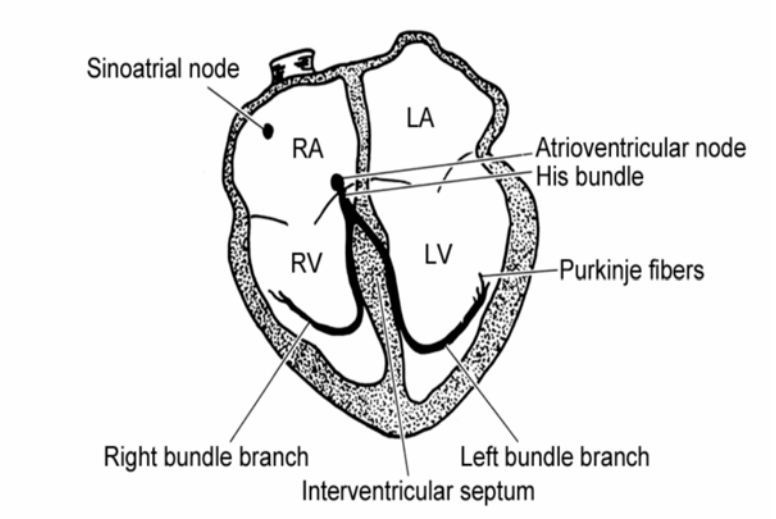


Figure 1. A schematic diagram of the heart. Adapted with permission from Goldberger and Goldberger (1994) (1).

a normal fluctuation with the respiratory cycle, called respiratory sinus arrhythmia. In addition, there is a modulation of the periodicity of the normal heartbeat in response to changing demands of the body and the operation of a variety of feedback and control mechanisms. This normal operation of the heart can be disturbed, leading to abnormal cardiac rhythms that are collectively called cardiac arrhythmias. Cardiac arrhythmias are associated with abnormal initiation of a wave of cardiac excitation, abnormal propagation of a wave of cardiac excitation, or some combination of the two. However, despite the apparent simplicity, cardiac arrhythmias can manifest themselves in a great many different ways, and it is still not always possible to figure out the mechanism of an arrhythmia in any given individual. In this essay, I address some of the advances that have been made in understanding cardiac arrhythmias theoretically, and indicate several directions in which future advances may lead to improved therapies.

In order to appreciate the origin of cardiac arrhythmias, it is first necessary to have a rudimentary knowledge about the spread of the cardiac impulse in the heart (1). The heart is composed of four chambers: the right and left atria, and the right and left ventricles (Figure 1). The atria are electrically connected to each other, but are insulated from the ventricles everywhere except in a small region called the atrioventricular (AV) node. The ventricles are also electrically connected to each other. The rhythm of the heart is set by the sinoatrial node located in the right atrium, which acts as the pacemaker of the heart. From a

mathematical perspective, this pacemaker is an example of a nonlinear oscillator. Thus, if the rhythm is perturbed, for example by delivering a shock to the atria, then in general the timing of subsequent firings of the sinus node may be reset (i.e., they occur at different times than they would have if the shock had not been delivered), but the frequency and amplitude of the oscillation will remain the same. A wave of excitation initiated in the sinus node travels through the atria, then through the atrioventricular node, then through specialized Purkinje fibers to the ventricles. The wave of electrical excitation is associated with a wave of mechanical contraction so that the cardiac cycle is associated with contraction and pumping of the blood through the body. The right and left atria are comparatively small chambers and act as collection points for blood. The right atrium collects blood from the body and the left atrium collects blood from the lungs. The right ventricle pumps blood to the lungs to be oxygenated, whereas the left ventricle pumps blood that has returned to the heart from the lungs to the rest of the body. The right atrium and right ventricle are separated by the tricuspid valve, which prevents backflow of blood during ventricular contraction. Similarly, the left atrium and left ventricle are separated by the mitral valve. In order to pump the blood, the ventricles are comparatively large and muscular.

The electrical events associated with cardiac activity can be easily monitored in an electrocardiogram, which is a measurement of the potential difference between two points on the surface of the body. Since the heart generates waves of electrical activation that propagate through the heart during the cardiac cycle, the deflections on the electrocardiogram reflect cardiac activity. In the normal electrocardiogram, there are several main deflections, labeled the P wave, the QRS complex, and the T wave (Figure 2A) (1). The P wave is associated with electrical activation of the atria, the QRS complex is associated with electrical activation of the ventricles, and the T wave is associated with repolarization of the ventricles. The duration of the PR interval reflects the conduction time from the atria to the ventricles, typically 120 to 200 ms. The duration of the QRS complex reflects the time that it takes for the wave of excitation to activate the ventricles. Because of the specialized Purkinje fibers, the wave of activation spreads rapidly through the ventricles so that the normal duration of the QRS complex is less than 100 ms. The time interval from the beginning of the QRS complex to the end of the T wave, called the QT interval, reflects the duration of the time the ventricles are in the contraction phase. The duration of the QT interval depends somewhat on the basic heart rate. It is shorter when the heart is beating faster. For heartbeats in the normal range the QT interval is typically of the order of 300–450 ms. The rate of the heart is often measured by time intervals between two consecutive R waves. Abnormally fast heart rates, faster than about 90 beats per minute, are called tachycardia, and abnormally slow heart rates, slower than about 50 beats per minute, are called bradycardia.

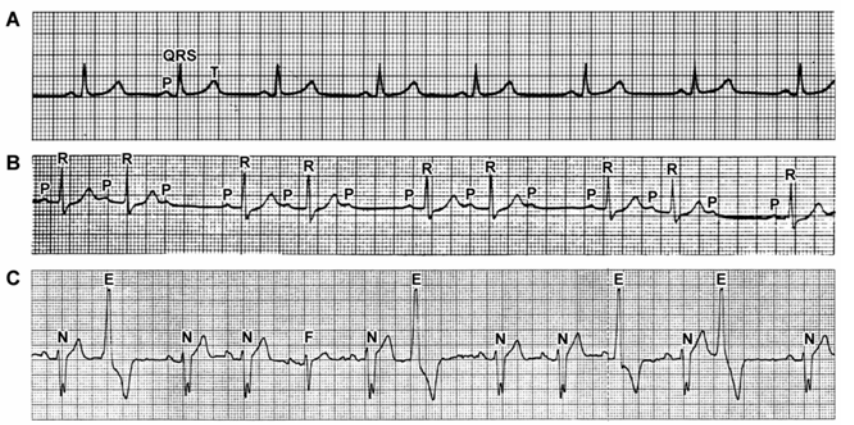


Figure 2. Sample electrocardiograms. In all traces one large box represents 0.2 s. (**A**) The normal electrocardiogram. The P wave, QRS complex, and T wave are labelled. (**B**) 3:2 Wenckebach rhythm, an example of second-degree heart block. There are three P waves for each R wave in a repeating pattern. (**C**) Parasystole. The normal beats, labelled N, occur with a period of about 790 ms, and the abnormal ectopic beats, labelled E, occur with a regular period of 1300 ms. However, when ectopic beats fall too soon after the normal beats, they are blocked. Normal beats that occur after an ectopic beat are also blocked. If a normal and ectopic beat occur at the same time, the complex has a different geometry, labelled F for fusion. In this record, the number of normal beats occurring between ectopic beats is either 4, 2, or 1, satisfying the rules given in the text. Panels **A** and **B** are adapted with permission from Goldberger and Goldberger (1994) (1). Panel **C** is adapted with permission from Courtemanche et al. (1989) (16).

2. TWO ARRHYTHMIAS WITH A SIMPLE MATHEMATICAL ANALYSIS

Even a cursory examination of electrocardiograms from some patients experiencing a cardiac arrhythmia would convince a mathematician or a mathematically oriented physician that there must be an underlying mathematical theory. To illustrate this observation, I consider a class of cardiac arrhythmias associated with conduction defects through the AV node. In Wenckebach rhythms there is a normal sinus rhythm, but not all atrial activations propagate to the ventricles, leading to rhythms in which there are more P waves than QRS complexes. It is common to classify Wenckebach rhythms by a ratio giving the number of P waves to the number of QRS complexes. For example, Figure 2B shows a 3:2 Wenckebach rhythm. In the 1920s, van der Pol and van der Mark developed a mathematical model of the heart as coupled nonlinear oscillators that display striking similarities to the Wenckebach rhythms (2).

Subsequently, a number of studies have demonstrated striking mathematical characteristics of Wenckebach rhythms (3–7). The basis of these formulations is

to assume that the conduction time of the *i*th beat through the AV node, AV_i, is a function F of the recovery time since the passage of the last excitation through the AV node, which is designated VA_{i-1}. Assuming that there is a periodic atrial stimulation, either from the sinus node or by an artificial pacemaker with a period AA, we find (5) that

$$AV_i = F(VA_{i-1}) = F(k \times AA - AV_{i-1}), \quad [1]$$

where k is the smallest integer such that $k \times AA > \theta$, θ being the refractory period of the AV node. Typically, the recovery curve F is a monotonically decreasing curve. In this case, it is possible to demonstrate mathematically that, if the properties of the AV node are fixed, as the frequency of atrial activation is increased, different types of *N*:*M* heart block, where *N* is the number of sinus beats and *M* is the number of ventricular beats in a repeating sequence, can be observed. If there is *N*:*M* heart block at one stimulation frequency and *N*':*M*' heart block at a higher frequency, then *N*+*N*':*M*+*M*' heart block is expected at some intermediate stimulation frequency (4–7). This result provides a mathematical classification complementary to the cardiological classification, and can be confirmed in clinical settings. Experimental and clinical studies have also demonstrated that these basic properties of Wenckebach rhythms can be observed in normal mammalian hearts as the atrial activation rate is increased (5–7). However, careful analysis of data shows that the above generalizations need to be modified. For example, if the atria are stimulated at a rapid rate so that 1:1 conduction is lost, there can often be an evolution of rhythms over a course of several minutes so that fewer atrial activations are conducted to the ventricles over time (8). In other circumstances, as the atria are stimulated at a rapid rate that still is associated with 1:1 conduction, there can be an alternation or other complex fluctuation of conduction time through the AV node (9,10). These effects are associated with changes in the properties of the AV node during rapid activation.

A different type of rhythm that is appealing to mathematicians is called parasystole. In the "pure" case, the normal sinus rhythm beats at a constant frequency, and an abnormal (ectopic) pacemaker in the ventricles beats at a second slower frequency (11). Figure 2C labels the normal (N) beats and the ectopic (E) beats. If the ectopic pacemaker fires at a time outside the refractory period of the ventricles, then there is an abnormal ectopic beat, identifiable on the electrocardiogram by a distinct morphology from the normal beat, and the following normal sinus beat is blocked. If the normal and abnormal beats occur at the same time, this leads to a fusion (F) beat. This simple mechanism has amazing consequences. These can be appreciated by forming a sequence of integers that counts the number of sinus beats between two ectopic beats. In general, for fixed sinus and ectopic frequencies and a fixed refractory period, in this sequence there are at most three integers, where the sum of the two smaller integers is one less than the largest integer. Moreover, given the values of the parameters, it is possible to

predict the three integers. Clinical studies demonstrate the applicability of these results to patients with an artificial parasystolic pacemaker produced by periodically stimulating the ventricles with an intracardiac catheter at a fixed rate (12). The mathematics for this problem is related to the "gaps and steps" problem in number theory (13).

Just as with Wenckebach rhythms, careful examination of data in patients shows ways in which the simple model of pure parasystole is not followed exactly. In some patients, the parasystolic pacemaker is reset, or modulated by the beat originating at the sinus node (14–16). In addition, although the model makes predictions of the exact timing of ectopic beats, in some cases there are unexpected missed ectopic beats. To account for such observations, the presence of stochastic mechanisms that affect the timing of the parasystolic beats has been hypothesized (17).

Both Wenckebach rhythms and parasystole can be used to illustrate the important mathematical concept of universal bifurcations (6). Thus, for both these rhythms, changes in the parameters in the model (e.g., the frequency of the sinus beat, the frequency of the ectopic beat, or the parameters specifying the quantitative properties of the AV node or resetting of the parasystolic focus by the sinus beat) will nevertheless lead to the same sequences of rhythms, but the exact values at which any rhythm appears will be different from person to person. However, the observation that the sequence of dynamic behaviors will be the same is a triumph of mathematics and also provides a fundamental mathematical explanation why the same similar rhythms appear in different people, in whom the anatomical and physiological properties of the heart must necessarily differ. However, both Wenckebach rhythms and parasystole are diagnosed and treated when necessary, by physicians who have no knowledge of the underlying mathematics. Further, it is not clear how knowledge of the mathematics could lead to improvements in the therapy. Thus, at the current time the mathematical analyses are not relevant to the practice of medicine.

3. REENTRANT ARRHYTHMIAS

From a medical perspective, the most important class of arrhythmias are called reentrant arrhythmias. In these arrhythmias, the period of the oscillation is set by the time an excitation takes to travel in a circuitous path, rather than the period of oscillation of a pacemaker (18). In some cases the reentrant circuit can be found in a single chamber of the heart. For example, in typical atrial flutter, there is a wave circulating around the tricuspid valve in the right atrium, and in some patients who have had a heart attack there is reentrant circuit entirely contained entirely in the ventricles. In contrast, in Wolf-Parkinson-White syndrome, there can be excitation following the normal excitation pathways from the atria to the AV node to the ventricles, but then the excitation is conducted retro-

gradely back to the atria via an abnormal accessory pathway between the ventricles and the atria. In all these reentrant arrhythmias, a part of the circuit is believed to be a comparatively thin strand of tissue. Cardiologists often conceive of the reentrant rhythm as traveling in a one-dimensional ring (18). This has an important implication for therapy: "if you cut the ring, you can cure the rhythm." By inserting catheters directly into a patient's heart and delivering radiofrequency radiation to precisely identified loci, the cardiologist destroys heart tissue and can often cure these serious arrhythmias. In these cases, the cardiologist is thinking like a topologist since changing the topology of the heart cures the arrhythmia.

Although the conceptualization of a wave traveling on a one-dimensional ring seems overly simplistic, from perspectives of both mathematics and medicine there are several interesting consequences (19). Experimental systems, simulations, and theoretical analyses have demonstrated that waves circulating on one-dimensional rings may experience an instability such that the circulation is not constant but there can be a complex fluctuating propagation velocity that arises as a consequence of the interaction with the wavefront with its own refractory tail (20–23). In addition, if a single stimulus is delivered to the medium during the course of the reentrant propagation, the propagating wave will either be reset or annihilated (21,24–26). Further, periodic stimulation can lead to the entrainment or annihilation of the propagating wave (25–27). Finally, a sequence of premature stimuli delivered to the heart during normal sinus rhythm can often lead to the initiation of tachycardia.

In some clinical settings, analysis of the resetting, entrainment, and initiation of tachycardias offers clinicians important clues about the mechanism, and consequently can help the cardiologist choose an appropriate therapy (18,28). The ability to induce monomorphic ventricular tachycardia using a sequence of up to three premature stimuli is often taken as an indication of a reentrant mechanism for the tachycardia. In such cases, resetting and entrainment of the induced tachycardia can help the cardiologist localize a site for ablation. For example, using intracardiac catheters cardiologists might attempt to identify an anatomical locus that has the following characteristics during a monomorphic ventricular tachycardia: (i) activation of the site occurs at a fixed time interval before the appearance of a deflection on the surface electrocardiogram; (ii) during periodic pacing at a rate slightly faster than the tachycardia from that site, the time interval from the stimulus to the surface deflection on the electrocardiogram is the same as during the spontaneous rhythm; and (iii) during periodic pacing, the morphology of the tachycardia on a 12-lead electrocardiogram is identical to what is observed during the tachycardia. If these three criteria are satisfied, the cardiologist would identify the site as part of the reentry circuit and would target it for ablation. Recent advances in mapping of arrhythmias and in the ability to identify suitable anatomical sites for ablation based on an understanding of the anatomical substrate of arrhythmias have further advanced the

therapy of tachycardia, even without a detailed mathematical analysis of these arrhythmias (29). Thus, although the precise mathematical analysis of the circulation of reentrant waves on one-dimensional rings is necessarily sophisticated, cardiologists have been able to devise a qualitative understanding of the phenomena that enables them to make clinical decisions without carrying out the mathematics.

Other reentrant arrhythmias are not as well understood and are not as easily treated. Many theoretical and experimental studies have documented spiral waves circulating stably in two dimensions and scroll waves circulating in three dimensions (30,31). Since real hearts are three dimensional, and there is still no good technology to image excitation in the depth (as opposed to the surface) of the cardiac tissue, the actual geometry of excitation waves in cardiac tissue associated with some arrhythmias is not as well understood and is now the subject of intense study. From an operational point of view, I suggest that any arrhythmia that CANNOT be cured by a small localized lesion in the heart will best be described by rotating spiral or scroll waves. Such rhythms include atrial and ventricular fibrillation. In these rhythms, there is evidence that there is strong fractionation (breakup) of excitation waves giving rise to multiple small spiral waves and patterns of shifting blocks (32). Tachycardias can also arise in the ventricles in other patients than those who have experienced a heart attack, or perhaps occasionally in hearts with completely normal anatomy, and in these individuals it is likely that spiral and scroll waves are the underlying geometries of the excitation. A particularly dangerous arrhythmia, polymorphic ventricular tachycardia (in which there is a continually changing morphology of the electrocardiogram complexes), is probably associated with meandering spiral and scroll waves (30).

4. FUTURE PROSPECTS

To date, there have been significant advances in developing a theoretical understanding of the mechanisms underlying complex cardiac arrhythmia and this remains an extremely active area for research. Certain trends are clear. The advances in computational ability, combined with improved understanding of the ionic mechanisms of the action potential, are combining to make feasible simulations of extremely complex mathematical models of cardiac propagation using physiological and anatomical parameters that are increasingly close to those found in real hearts (33,34) (for an excellent review, see this volume, Part III, chapter 3.2, by Winslow). Although it is not yet possible to combine realistic mathematical models of the electrical activity with realistic mechanical models of the heart's pumping, computational advances in modeling the mechanical properties of whole hearts have succeeded in generating flow patterns in the beating heart similar to what is observed (35). Given current trends and the de-

veloping expertise, it is likely that at least over the next decade there will be increasingly realistic and computationally expensive models of electrical and mechanical wave propagation in human heart. Although most studies to date have focused on ventricular tachycardia, ventricular fibrillation, and atrial fibrillation, realistic models should be able to incorporate anatomical and physiological abnormalities that are associated with other arrhythmias, and consequently it is likely that mathematical models of other arrhythmias will likewise be generated.

In parallel with the development of models of arrhythmias, I anticipate work will be carried out on the mathematical analysis and modeling of arrhythmias in people. Although there is a large clinical literature describing various arrhythmias, there is still lots of room for theoreticians to dive into this literature to develop testable models of arrhythmia. One aspect that such models will necessarily have to deal with is the often paroxysmal starting and stopping of arrhythmia. In some cases, if it were possible to predict the onset of the arrhythmia before it occurred, then it might be possible to take corrective steps to avert the arrhythmia. I feel certain that if more theoreticians took time to examine the records of patients experiencing arrhythmias, they would find the data compelling. To understand the clinical data, it will be necessary to examine experimental and mathematical models of the heart in which there is a spontaneous initiation of arrhythmia, in contrast to most current mathematical models in which arrhythmia is induced by delivering a stimulus or series of stimuli to the heart.

Tissue cultures of cardiac cells grown in various geometries often exhibit the spontaneous generation of reentrant arrhythmias (36–39). Figure 3 shows an example of a burst of activity in tissue culture. Mathematical models that incorporate spontaneous activity, excitability, heterogeneity, and a decrease of excitability following rapid activation typically show a range of parameter values in which there is spontaneous initiation and termination of reentrant activity (36) similar to the paroxysmal rhythms observed clinically. Moreover, in heterogeneous excitable media, plane waves may spontaneously break up into spiral waves over intermediate ranges of coupling between the cells (40). Therefore, this approach gives insight into how complex cardiac rhythms may spontaneously arise as a consequence of physiological and anatomical changes that reduce excitability and/or increase heterogeneity. I expect that an ability to manipulate the geometry combined with the potential to manipulate the ionic components of cells using molecular biology techniques, and the use of optical methods to record rhythms over extended spatial areas for extended periods of time, should make tissue cultures an excellent preparation for future studies of complex arrhythmia.

The treatment of cardiac patients has undergone a remarkable development and evolution over the past two decades. The advances are due to many factors, including medications to better control hypertension and blood lipids; a variety

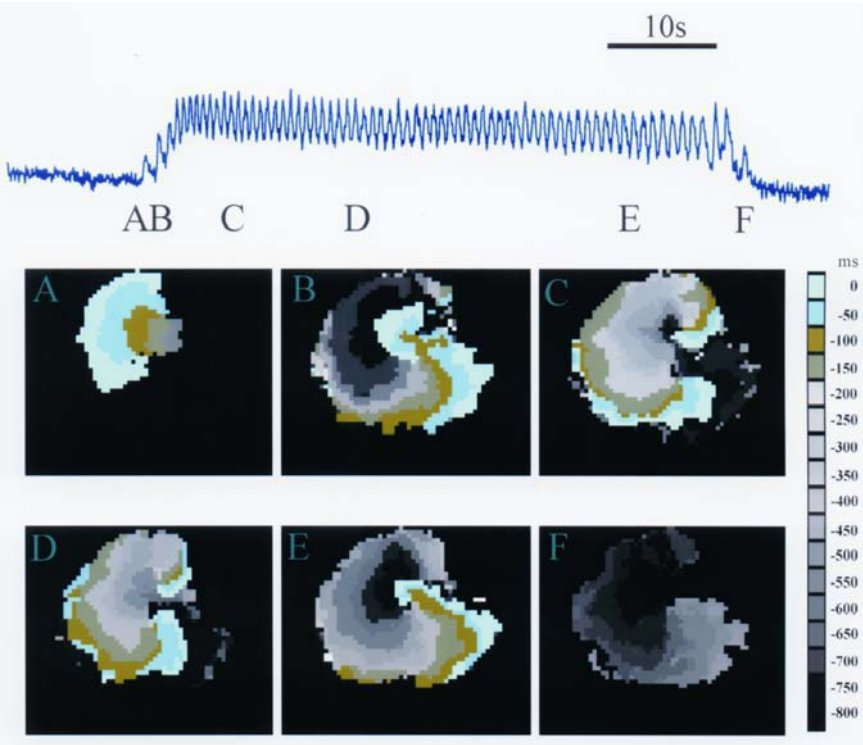


Figure 3. A burst of activity recorded using a calcium-sensitive fluorescent dye in a tissue culture preparation of embryonic chick heart cells. The isochronal maps below the tracing show the evolution of the activity at different stages of the burst. Adapted with permission from Bub, Glass, and Shrier (1998) (36).

of procedures and medications that can reduce the incidence and mitigate the consequences of cardiac infarcts; devices such as pacemakers, antitachycardia pacers, and defibrillators that help to control the cardiac rhythm; and ablation procedures that can help eliminate the anatomical substrate of some types of arrhythmia. Although some of this work involves sophisticated technology, this technology has largely been developed by engineers working in collaboration with cardiologists. To date, the sorts of nonlinear theoretical insights that have been discussed above have not provided a significant impetus or contribution to the advances. An interesting question is whether nonlinear dynamical analyses will be helpful in the development of new medical approaches or the improvement of current approaches. At the moment several new strategies are being explored.

There are many different algorithms that are being developed to help assess the risk of sudden cardiac death. One of these approaches is based on the obser-

vation that alternation of the T wave on the electrocardiogram is correlated with a higher risk of sudden cardiac death (41,42). Since the T wave is associated with action potential duration, and alternation of action potential duration can be associated with a period-doubling bifurcation (43), one possible origin of T-wave alternans is a period-doubling bifurcation induced by rapid cardiac rates. T-wave alternans could also arise from a 2:1 block in some regions of the heart and a 1:1 conduction in other regions. Independent of the cause of the alternans, the attention to this abnormality provides a new medical approach to the assessment of risk of sudden death. Others have noted that reduced heart rate variability is correlated with a higher risk for sudden death (44,45). Since reduced heart rate variability would be expected in patients with impaired ventricular function and elevated sinus rate associated with higher circulating catecholamines, it might not be surprising that reduced heart rate variability is associated with a higher risk of sudden death. It seems clear that there is an enormous amount of information hidden in the fluctuations of the sinus rate and in the dynamics of arrhythmias in patients who have intermittent arrhythmias. For example, arrhythmias with frequent premature ventricular contractions that appear to be the same if looked at superficially are in fact quite different when the dynamics are dissected in detail, and only rarely has it been possible to find a good theoretical understanding of the mechanisms consistent with the observed dynamics over long periods. In order to find quantitative agreement between computed arrhythmias based on mathematical models and observed arrhythmias, it may be necessary to use stochastic models (17,46). Since the opening and closing of single ion channels assume stochastic mechanisms, a physiological basis for stochastic macroscopic behavior may exist. Indeed, recent theoretical models of cardiac cells have employed stochastic models for ion channels dynamics to help interpret experimental data (47) (see also Part III, chapter 3.2, by Winslow).

Algorithms for assessment of arrhythmia may be helpful in other ways. Tateno and I have developed a method to assess the presence of atrial fibrillation based on the timing of ventricular complexes. The basic idea is to compare the histogram of the changes in RR intervals (for a given mean RR rate) with standard histograms collected from patients experiencing atrial fibrillation (48). This type of algorithm might be implemented in a portable device, and so might be helpful in assessing the incidence of atrial fibrillation in a patient at risk or assessing the efficacy of drug therapy in a patient who has atrial fibrillation.

Work is also underway to develop new methods to stimulate the heart to control or terminate arrhythmia (49). Simulations of defibrillation combined with experimental studies might be useful in helping to improve the shapes and placement of electrodes or the optimal wave form for defibrillation. In addition, new methods might be found to regularize or terminate arrhythmias. Current medical devices now use pacing protocols to terminate tachycardias, but it is possible that theoretical analyses and simulation of arrhythmia might yield improved algorithms (25,26). In addition, alternating reentrant cardiac arrhythmias

involving the AV node have been regularized by premature atrial stimulation in both animal models (50) and in people (51). The possibility exists that more complex arrhythmias such as atrial or ventricular fibrillation might also be controlled by stimulation (52,53).

In recent years, there has been an increasingly interdisciplinary research environment in which experts in the mathematical aspects of arrhythmias are being teamed with experts in physiology and clinical problems. It will be interesting to see if the development of our theoretical understanding of the mechanisms of cardiac arrhythmias will lead to a comparable improvement in medical procedures.

5. ACKNOWLEDGMENTS

This work has been supported by the Natural Sciences and Engineering Council of Canada, the Canadian Institute for Health Research, the Canadian Heart and Stroke Foundation, the Mathematics of Information Technology and Complex Systems Centre of Excellence (Canada), and the Research Resource for Complex Physiologic Signals funded by the National Institutes of Health, USA.

6. REFERENCES

1. Goldberger AL, Goldberger E. 1994. *Clinical electrocardiography: a simplified approach*, 5th ed. Mosby, St. Louis.
2. van der Pol B, van der Mark J. 1928. The heartbeat considered as a relaxation oscillation, and an electrical model of the heart. *Philos Mag* **6**:763–765.
3. Roberge FA, Nadeau RA. 1969. The nature of Wenckebach cycles. *Can J Physiol Pharmacol* **47**:695–704.
4. Keener JP. 1981. On cardiac arrhythmia: AV conduction block. *J Math Biol* **12**:215–225.
5. Shrier A, Dubarsky H, Rosengarten M, Guevara MR, Nattel S, Glass L. 1987. Prediction of complex atrioventricular conduction rhythms in humans with use of the atrioventricular nodal recovery curve. *Circulation* **76**:1196–1205.
6. Glass L, Guevara MR, Shrier A. 1987. Universal bifurcations and the classification of cardiac arrhythmias. *Ann NY Acad Sci* **504**:168–178.
7. Guevara MR. 1991. Iteration of the human atrioventricular (AV) nodal recovery curve predicts many rhythms of AV block. In *Theory of heart: biomechanics, biophysics, and nonlinear dynamics of cardiac function*, pp. 313–358. Ed. L Glass, P Hunter, A McCulloch. Springer, New York.
8. Talajic M, Papadatos D, Villemaire C, Glass L, Nattel S. 1991. A unified model of atrioventricular nodal conduction predicts dynamic changes in Wenckebach periodicity. *Circ Res* **68**:1280–1293.
9. Sun J, Amellal F, Glass L, Billette J. 1995. Alternans and period-doubling bifurcations in atrioventricular nodal conduction. *J Theor Biol* **173**:79–91.

10. Christini DJ, Stein KM, Markowitz SM, Mittal S, Slotwiner DJ, Scheiner MA, Iwai S, Lerman BB. 2001. Complex AV-nodal dynamics during ventricular-triggered atrial pacing in humans. *Am J Physiol* **281**:H865–H872.
11. Glass L, Goldberger A, Bélair J. 1986. Dynamics of pure parasystole. *Am J Physiol* **251**:H841–H847.
12. Castellanos A, Fernandez P, Moleiro F, Interian A, Myerburg R. 1991. Symmetry, broken symmetry, and restored symmetry of apparent pure ventricular parasystole. *Am J Cardiol* **68**:256–259.
13. Slater N. 1967. Gaps and steps for the sequence nθ mod 1. *Proc Camb Phil Soc* **63**:1115–1123.
14. Moe GK, Jalife J, Mueller WJ, Moe B. 1977. A mathematical model of parasystole and its application to clinical arrhythmias. *Circulation* **56**:968–979.
15. Ikeda N, Yoshizawa S, Sato T. 1983. Difference equation model of ventricular parasystole as an interaction of pacemakers based on the phase response curve. *J Theor Biol* **103**:439–465.
16. Courtemanche M, Glass L, Bélair J, Scagliotti D, Gordon D. 1989. A circle map in a human heart. *Physica D* **40**:299–310.
17. Schulte-Frohlinde V, Ashkenazy Y, Ivanov PCh, Glass L, Goldberger AL, Stanley HE. 2001. Noise effects on the complex patterns of abnormal heartbeats. *Phys Rev Lett* 87:068104.
18. Josephson ME. 2002. *Clinical cardiac electrophysiology: techniques and interpretations*. Williams and Wilkins, Philadelphia.
19. Rudy Y. 1995. Reentry: insights from theoretical simulations in a fixed pathway. *J Cardiovasc Electrophysiol* **6**:294–312.
20. Frame LH, Simpson MB. 1988. Oscillations of conduction, action potential duration, and refractoriness: a mechanism for spontaneous termination of reentrant tachycardias. *Circulation* **78**:1277–1287.
21. Quan WL, Rudy Y. 1991. Termination of reentrant propagation by a single stimulus: a model study. *Pacing Clin Electrophysiol* **14**:1700–1706.
22. Courtemanche M, Glass L, Keener JP. 1993. Instabilities of a propagating pulse in a ring of excitable media. *Phys Rev Lett* **70**:2182–2185.
23. Vinet A, Roberge FA. 1994. The dynamics of sustained reentry in a ring model of cardiac tissue. *Annu Biomed Eng* **22**:568–591.
24. Glass L, Josephson ME. 1995. Resetting and annihilation of reentrant abnormally rapid heartbeat. *Phys Rev Lett* **75**:2059–2063.
25. Sinha S, Stein KM, Christini DJ. 2002. Critical role of inhomogeneities in pacing termination of cardiac reentry. *Chaos* **12**:893–902.
26. Comtois A, Vinet A. 2002. Resetting and annihilation of reentrant activtiy in a model of a one-dimensional loop of ventricular tissue. *Chaos* **12**:903–923.
27. Glass L, Nagai Y, Hall K, Talajic M, Nattel S. 2002. Predicting the entrainment of reentrant cardiac waves using phase resetting curves. *Phys Rev E* **65**:021908.
28. Stevenson WG, Delacretaz E. 2000. Strategies for catheter ablation of scar-related ventricular tachycardia. *Curr Cardiol Rep* **2**:537–544.
29. Stein KM, Markowitz SM, Mittal S, Slotwiner DJ, Iwai S, Lerman BB. 2002. Anatomic determinants of atrial arrhythmias: new insights from three-dimensional mapping. *Chaos* **12**:740–746.
30. Gray RA, Jalife J, Panfilov A, Baxter WT, Cabo C, Davidenko JM, Pertsov AM. 1995. Nonstationary vortexlike reentrant activity as a mechanism of polymorphic ventricular tachycardia in the isolated rabbit heart. *Circulation* **91**:2454–2469.
31. Winfree AT. 2001. *The geometry of biological time*, 2nd ed. Springer, New York.
32. Fenton FH, Cherry EM, Hastings HM, Evans SJ. 2002. Multiple mechanisms of spiral wave breakup in a model of cardiac electrical activity. *Chaos* **12**:852–892.
33. Qu Z, Kil J, Xie F, Garfinkel A, Weiss JN. 2000. Scroll wave dynamics in a three-dimensional cardiac tissue model: roles of restitution, thickness, and fiber rotation. *Biophys J* **78**:2761–2775.

34. Virag N, Jacquemet V, Henriquez CS, Zozor S, Blanc O, Vesin J-M, Pruvot E, Kappenberger L. 2002. Study of atrial arrhythmias in a computer model based on magnetic resonance images of human atria. *Chaos* **12**:754–763.
35. Kovacs SJ, McQueen DM, Peskin CS. 2001. Modeling cardiac fluid dynamics and diastolic function. *Phil Trans R Soc Lond A* **359**:1299–1314.
36. Bub G, Glass L, Shrier A. 1998. Bursting calcium rotors in cultured cardiac myocyte monolayers. *Proc Natl Acad Sci USA* **123**:10283–10287.
37. Nagai Y, González H, Shrier A, Glass L. 2000. Paroxysmal starting and stopping of circulating waves in excitable media, *Phys Rev Lett* **84**:4248–4251.
38. Bub G, Tateno K, Shrier A, Glass L. 2003. Spontaneous initiation and termination of complex rhythms in cardiac cell culture. *J Cardiovasc Electrophysiol* **14**:S229–S236.
39. Hwang S, Yea K, Lee KJ. 2004. Regular and alternant spiral waves of contractile motion on rat ventricle cell cultures. *Phys Rev Lett* **92**:198103.
40. Bub G, Shrier A, Glass L. 2002. Spiral wave generation in heterogeneous excitable media. *Phys Rev Lett* **88**:058101.
41. Hohnloser H, Klingenheben T, Bloomfield D, Dabbous O, Cohen RJ. 2003. Usefulness of microvolt T-wave alternans for prediction of ventricular tachyarrhythmic events in patients with dilated cardiomyopathy: results from a prospective observational study. *J Am Coll Cardiol* **41**:2220–2224.
42. Verrier RL, Tolat AV, Josephson ME. 2003. T-wave alternans for arrhythmia risk stratification in patients with idiopathic dilated cardiomyopathy. *J Am Coll Cardiol* **41**:2225–2227.
43. Guevara MR, Ward G, Shrier A, Glass L. 1984. Electrical alternans and period-doubling bifurcations. *IEEE Comput Cardiol* **11**:167–170.
44. Goldberger AL, Amaral LA, Hausdorff JM, Ivanov PCh, Peng CK, Stanley HE. 2002. Fractal dynamics in physiology: alterations with disease and aging. *Proc Natl Acad Sci USA* **99**(suppl 1):2466–2472.
45. La Rovere MT, Pinna GD, Maestri R, Mortara A, Capomolla S, Febo O, Ferrari R, Franchini M, Gnemmi M, Opasich C, Riccardi PG, Traversi E, Cobelli F. 2003. Short-term heart rate variability strongly predicts sudden cardiac death in chronic heart failure patients. *Circulation* **107**:565–570.
46. Schulte-Frohlinde V, Ashkenazy Y, Goldberger AL, Ivanov PCh, Costa M, Morley-Davies A, Stanley HE, Glass L. 2002. Complex patterns of abnormal heartbeats. *Phys Rev E* **66**:031901.
47. Krogh-Madsen T, Glass L, Doedel EJ, Guevara MR, 2004. Apparent discontinuities in the phase-setting response of cardiac pacemakers. *J Theor Biol* **230**:499–519.
48. Tateno K, Glass L. 2001. Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and ΔRR intervals. *Med Biol Eng Comput* **39**:664–671.
49. Trayanova N, Eason J. 2002. Shock-induced arrhythmogenesis in the myocardium. *Chaos* **12**:962–972.
50. Hall K, Christini DJ, Tremblay M, Collins JJ, Glass L, Billette J. 1997. Dynamic control of cardiac alternans. *Phys Rev Lett* **78**:4518–4521.
51. Christini DJ, Stein KM, Markowitz SM, Mittal S, Slotwiner DJ, Scheiner MA, Iwai S, Lerman BB. 2001. Nonlinear-dynamical arrhythmia control in humans. *Proc Natl Acad Sci USA* **98**:5827–5832.
52. Garfinkel A, Spano ML, Ditto WL, Weiss JN. 1992. Controlling cardiac chaos. *Science* **257**:1230–1235.
53. Gauthier DJ, Hall GM, Oliver RA, Dixon-Tulloch EG, Wolf PD, Bahar S. 2002. Progress toward controlling *in vivo* fibrillating sheep atria using a nonlinear-dynamics-based-closed-loop feedback method. *Chaos* **12**:952–962.

4.1

HOW DISTRIBUTED FEEDBACKS FROM MULTIPLE SENSORS CAN IMPROVE SYSTEM PERFORMANCE: IMMUNOLOGY AND MULTIPLE-ORGAN REGULATION

Lee A. Segel

Department of Computer Science and Applied Mathematics,
Weizmann Institute of Science, Rehovot, Israel

Complex physiological entities such as the immune system can be regarded as possessing multiple goals. Sensors reflect information on goal performance as well as general physiological state. As is illustrated here with a simple immunological model, in a distributed and generalized version of classical feedback this information can be used to improve system goal attainment. In the immune system it is cytokines that transmit the sensory information. The concept of distributed feedback to improve multiple-goal performance is shown to be of possible relevance to multiple organ dysfunction syndrome and its therapy.

1. INTRODUCTION

The human immune system is composed of about a trillion cells. There are dozens of cell types and hundreds of signalling chemicals. The immune system plays a variety of homeostatic roles, such as helping to direct wound healing, but its main job is controlling the myriad pathogens that roam the body. Pathogens do not passively await immune destruction; they have evolved various evasive

Address correspondence to: Lee A. Segel, Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Ziskind Building #24, 76100 Rehovot, Israel (lee.segel @weizmann.ac.il). [See acknowledgments, p. 435.]

tactics such as the adoption of changing "disguises" and the secretion of molecules that interfere with immune system communication. The trillion-cell immune "army" has no general; signals from the brain modulate the action of the immune system to some extent but do not control its detailed performance.

In this essay I shall survey some of my attempts, with colleagues, to apply theoretical investigations in order to better understand what controls the immune system, with emphasis on ideas that might be of interest to physicians. A unifying concept is the role of information, defined simply as "knowing what's going on," in guiding immune system performance and medical interventions.

Before beginning the main body of this chapter, let me note that a natural analogy of the principal task facing the immune system is the task facing the physician in curing disease, or at least blunting its impact. I have argued that a better analogy is with a public health system (1). Such a system may well have to contend with several diseases at once, selected from a wide spectrum of possibilities. Triage is a key problem: how to allocate the available resources appropriately to meet a changing panoply of challenges. Feedbacks are essential to the proper operation of a public health system. As information flows in concerning the performance of the system, shifts in resource allocation can be made to better contend with changing circumstances.

2. THERAPY AS AN INFORMATION-YIELDING PERTURBATION

Shochat and colleagues (2) constructed a detailed model (about 25 differential equations, almost 40 parameters) for multiple hematopoietic stages, starting from pluripotent stem cells and ending in the generation of monocytes and polymorphonucleic cells (PMN). The role of granulocyte colony-stimulating factor (G-CSF) was taken into account. Values for all but five parameters could be found in the literature. The remaining parameters were estimated by requiring model predictions to match data on blood counts of monocytes and PMN. Of crucial importance were measurements of shifting monocyte and PMN levels by frequent blood counts routinely taken in patients during high-dose chemotherapy followed by autologous peripheral blood stem cell transplantation and injection of G-CSF. (The lymphoid lineage was neglected in the model, since appropriate data were not available.) By choosing parameters to minimize the least-squared difference between the actual blood counts and the theoretical predictions, Shochat et al. produced estimates of parameters that cannot as yet be measured directly in humans, notably stem cell density and division rate. The calibrated model was used to explore alternative strategies to reduce the post-transplant nadir period of deficient immunity. One somewhat surprising suggestion is that G-CSF injections before day 5 post-transplant may have little influence in diminishing the nadir period.

Shochat and colleagues (2) showed how *strong physiological perturbations induced by medical therapy can generate data that can yield both quantitative and qualitative insights, provided the data are compared with predictions of a suitable quantitative model*. Another example is the comparison in (3) of viral counts with model predictions following administration of a protease inhibitor to HIV patients. This and related work established that, although progression to AIDS may take many years, viral turnover is nonetheless rapid, mandating multiple-drug therapy to avoid fast viral evolution to drug resistance.

3. EMPLOYING INFORMATION ON PROGRESS TOWARD MULTIPLE GOALS TO REGULATE THE IMMUNE RESPONSE

3.1. Multiple Goals

Together with R. Bar-Or, I have argued that it is useful to regard the immune system as simultaneously pursuing a variety of overlapping and conflicting goals (4) (see also this volume, Part II, chapter 5, by Krakauer). Examples of such goals are avoiding harm to self, killing dangerous pathogens, and acting quickly against such pathogens, but acting economically. In the face of ever-shifting challenges, the immune system monitors how it is doing with respect to these goals by means of various sensors and uses the information thereby obtained to shift its actions in order to improve performance with respect to the variety of goals. In view of the information gleaned, performance can be improved by selecting suitable effector classes for expansion and others for contraction (Th1 vs. Th2, choice of isotypes). Moreover, within a given effector class, sensor-based information can be used to improve performance (when should a given class be deployed, when expanded, when contracted).

To gain more insight concerning this role of sensors, see (5) for an examination of how partial information from sensors can improve performance in a "team" version of the game Connect Four. A team that can gather and utilize information even slightly more effectively tends to win. If one thinks of a team as a model complex organism, then the results of many contests will lead to the evolution of organisms with gradually improving information capabilities. By monitoring what sensors give superior performance, one is led to the possible characterization of performance in terms of goals, that is, to obtain the ultimate goal of Four in a Row (analogous to survival of the organism) there are a variety of intermediate goals, represented, for example, by board configurations that are worth striving for.

3.2. A Model Immune system: Conflicting Roles of a Noxious Chemical

Here is a sketch of a simple model that illustrates some of the issues. Consider a population of pathogens that in isolation grows exponentially at rate r. Imagine an effector cell E of the immune system that kills pathogens P at a rate kEP (mass action assumption), where k depends on various factors. Suppose in

particular that k is an increasing function of the concentration N of a noxious chemical [$k = k(N)$] that is secreted by E at rate s. (Think of E as macrophages, which require an intracellular but leaky "poison" such as nitric oxide to dispose of ingested pathogens.) The proliferation rate of E is proportional to P (the presence of pathogens induces an immune response); both E and N are supposed to have a finite half-life. We will assume that over evolutionary time immune systems that are described by our model would evolve to diminish the average harm δ that the body sustains by the joint presence of pathogens and noxious chemical. Suppose for simplicity that the rate of doing harm to the body by pathogens P and noxious chemical N is simply $h_P P + h_N N$, where h_P and h_N are constants. The equations of this model are given in the Appendix. There the simplifying assumption is made that $k(N) = aN$, where a is a constant.

Computer simulations of the model demonstrate that there is a rate of secretion s^* that minimizes δ (4). If its secretion rate s is too large, noxious chemical N causes great harm; if s is too small, pathogens P are not controlled. Accordingly, one indeed expects the existence of an optimal secretion rate s^*. But s^* depends on the parameters of the model, particularly pathogen growth rate r. The faster the pathogen grows (the bigger r) the more noxious chemical N should be secreted to kill them, in spite of the harm done by N. Consequently, in the face of the shifting nature of the pathogen population, no "top–down" fixed choice of the secretion rate s can provide effective and economical immune response. Analogous difficulties are faced in selecting immune responses of other types.

One way to deal with the difficulty is to evolve an immune system that can detect telltale signs of a variety of pathogens and hence deal suitably with the different types (6,7). In view of the fast mutability of which many pathogens are capable, I have suggested that this "reflexive" response must be supplemented by an ability to sense how well the response is working and to alter it in view of this sensory information. As will now be shown, our simple model of the action of a noxious but essential chemical N can be modified to illustrate the issues.

3.3. Chemicals that Supply Information Concerning Progress toward Conflicting Goals

In our modified model, we shall assume that evolution has selected immune systems that can be characterized by just two goals: killing harmful pathogens and avoiding harm to self generated by the immune system. These goals are partially contradictory; the pathogen-destroying inflammatory process harms the host. In particular, intracellular pathogens are fought by destroying the cells where the pathogens reside. Suppose that information is available concerning immune performance in the form of a kill chemical K (whose presence is associated with pathogen destruction) and a harm chemical H (whose presence is associated with harm to the host). Possible kill indicators K include N-formyl peptides, palindromic DNA sequences, CD-1 ligating intracellular fragments

(such as mycolic acid) of Gram-negative bacteria, and heat-shock protein (HSP) epitopes that are characteristic of bacteria (1,8). Also see (9) for a discussion of how bacterial CpG DNA motifs (unmethylated cytosine-guanosine dinucleotides), acting as kill indicators, can guide an appropriate choice between Th1 and Th2 helper T cells. An example of a harm chemical is a heparin sulfate fragment resulting from the cleavage of the extracellular matrix by heparanase. Another possibility is a host-specific epitope of heat shock protein. (1).

To monitor progress toward its two goals, the immune system must be able to sense the presence of harmful pathogens (which should upregulate the immune response). In addition, sensors must detect harm due to the immune system (whose presence should downregulate the immune response). One way to accomplish this is to assume that the harm H_P due to pathogens is represented by

$$H_P = H/(1 + kN), \quad k \text{ a constant}, \qquad [1]$$

where the complement to H_P represents the harm H_I done by the immune system, via the noxious chemical N:

$$H_I = H - H_P. \qquad [2]$$

Indeed, we see from [1] that when N is large H_P is small and $H_I = H$, i.e., almost all harm to the host is done by the immune system. By contrast, if N is small, $H \approx H_P$, and almost all the harm is done by the pathogens. "Killing harmful pathogens" can be sensed by monitoring KH_P. If KH_P is large there is much pathogen killing (large K) *and* pathogens are causing much harm (large H_P). Note that new information is obtained by combining simpler information. Note also that such combinations are well within the capability of the complex intracellular interactions that characterize the internal affects of ligating a receptor on a cell surface.

Returning to our simple example, we implement as follows the assumption that the immune response, in this case the noxious chemical secretion rate s, should be elevated by evidence (KH_P) of killing harmful pathogens and should be downregulated by evidence (H_I) that the immune system is causing harm to self:

$$s = s_1 + \frac{s_2 K H_p}{1 + s_3 H_I + s_4 K H_p}. \qquad [3]$$

(In Eq. [3], s_1, s_2, s_3, and s_4 are constants.) To complete our model, we must specify the secretion and decay processes that govern the information chemicals H and K. To that end, let H be produced at a rate that is proportional to the rate $h_P P + h_N N$ at which harm is done to the host, and let K be produced at a rate proportional to the rate $aNPE$ at which pathogens are killed by effectors (compare Eqs. [A4] and [A2]). Suppose further that both H and K have a finite half-life. In the Appendix, these assumptions are translated into Eqs. [A5] and [A6] for the concentrations of K and H.

Computer simulations of Eqs. [A1]–[A6] indeed show, not surprisingly, that monitoring information on progress toward multiple goals can improve immune performance, in the sense of decreasing average harm δ. The system "automatically" adjusts itself for efficient combat when faced by pathogens of varying virulence. Moreover, even a given pathogen is better handled, compared with the control "no information model," by secreting N only when it is needed and turning secretion off when the pathogen threat is no longer serious (4).

I have described a simple model that shows how the deployment of a given immune system arm (effectors working via noxious chemicals) can use sensed information to improve performance. This material is summarized in Table 1. See (4) for a model that illustrates how sensed information can bias an immune system to choose appropriately among a variety of possible effector arms.

Table 1. Possible immune system "goals" and how they might be achieved

SENSING GOALS

Goal	To be sensed
Identify dangerous pathogen	High harm H AND (low noxious N OR high pathogen P)
Identify harm to self	High harm H AND (high noxious N OR low pathogen P)

PERFORMANCE GOALS

Goal	Action
Kill dangerous pathogens	Positive feedback to N secretion when sensors reveal dangerous pathogen and pathogen killing K
Avoid harm to self	Negative feedback to N secretion when sensors reveal dangerous pathogen and pathogen killing K

Note. Actuator effect of sensor information is mediated by cytokines.

4. CYTOKINES

In our model, we postulated "information molecules" *H* and *K* that were elevated when harm was done to the host or when pathogens were being killed. The levels of *H* and *K* were assumed to directly modify cellular behavior. In fact, matters are more complicated. Information molecules typically bind to cellular receptors, which in turn induce the secretion by the cell of signalling molecules that can be lumped under the term "cytokines." To give one example, there is evidence that mycobacterial 65-kD heat shock protein induces T cells to secrete tumor necrosis factor (TNF) and, later, interleukin-6 (IL-6) and IL-8 (10).

4.1. The Observed Actions of Cytokines

Matters might be simple if there were a one-to-one correspondence between the appearance of a cytokine and the triggering of some task or function. But that is not the case. Indeed, consider the following points cI–cIV that characterize the pleiotropic action of cytokines (1,8).

(cI) Upon ligation of receptors, several cytokines are normally produced. Here are three examples: (i) ligation of the LPS (lipopolysaccharide) receptor induces IL-1, 6, 8, 12 and TNFα; (ii) ligation of the Fcγ receptor by IgG induces IL-10; and (iii) ligation of scavenger receptors by apoptotic bodies induces TGFβ, PGE-2 and PAF.

(cII) A cytokine often effects several actions of a given cell (for example, IL-4 influences switches to IgM, IgG, and IgE).

(cIII) Sets of cytokines can induce a variety of different actions of different cells, for example, ligation of the LPS receptor induces B cells to increase antibody production and switch toward the IgG isotype, liver cells to secrete acute phase proteins, hypothalamus cells to elevate body temperature, neutrophils to become mobilized, and vascular cells to be more permeable.

(cIV) A given function is typically affected by several cytokines (the IgG switch is affected by IL-2,4,6 and interferon-γ).

The standard view of cytokines, and it is a correct one, is that *cytokines command*. (The switch from IgM to Ig is "commanded" in B cells, at least in vitro, by the list of cytokines in cII above.) In the language of control theory, in this view cytokines are the front end of the actuator machinery that is activated by sensor-receptors.

4.2. Cytokines as Bearers of Information

I propose an alternative view: that *cytokines transmit information*. I believe that regarding them as bearers of information is another correct way to charac-

terize cytokine actions, and one that is sometimes more useful than characterizing cytokines as inducers of commands.

Insight into the reasons for pleiotropic properties cI–cIV of cytokines is provided when they are regarded as providers of information. For example, it is appropriate that the same cytokine-encoded information induces different responses in different cells, for different cell types should play different roles in different situations. In one situation cell type α should fade away, as it cannot help, while cell type β should multiply and migrate to the site of the threat. In another situation, the roles of α and β should be interchanged. (An analogy is a scenario wherein hearing the chords of a marching band attracts Sunday drivers with children to entertain and repels those on the way to catch a plane.) For example, ligation of the LPS receptor induces the secretion of a panel of cytokines (see cI); these carry the message "Gram-negative bacteria are present." In response, a variety of appropriate actions are induced in different cell types (see cIII). That a given response can be triggered by a variety of cytokines (cIV) can be understood since several different sensed conditions, and hence several different cytokine-coded information packets, could well trigger the same cellular action (e.g., switch from IgM to IgG) as part of an appropriate response.

With respect to their command function, it is easy to find out what cells produce the various cytokines (see, e.g., appendix II in (11). By contrast, increased understanding of the role played by cytokine-borne information requires adding to available knowledge on what cytokines are induced by various receptor ligations.

5. CONTENDING WITH MULTIPLE INDEPENDENT GOALS

In §3.2 and 3.3 it has been implicitly assumed that the two goals—"kill dangerous pathogens" and "avoid harm to self"—are both aspects of a single comprehensive goal of avoiding harm to the host, whether by pathogens or by the immune system. Yet, this assumption may not be ideal, for different types of harm are effected by pathogens and by the immune system itself.

Perhaps it is better to regard the two goals (which are a theoretical construct) as independent. Assuming multiple independent goals certainly seems a sensible way to characterize the interconnected physiological system that deals with respiration, circulation, digestion, metabolism, neurological information processing, etc. Indeed, just the metabolic subsystem is faced with the multiple task of keeping myriad important chemicals at appropriate levels. A related metabolic task is allocating resources between maintenance and growth (12).

We have made progress in constructing a biologically plausible method to improve performance under the assumption that different performance goals are completely independent (N. Rappaport and L. Segel, unpublished). The idea is to adapt the "algorithm" used by bacteria that tend to swim toward certain

chemicals that attract them and to swim away from other chemicals that repel them (*chemotaxis*). The bacteria proceed in more or less straight lines for a certain period. This is called a "run." However, there is always a probability that the bacteria will suddenly "tumble," which amounts to choosing a new run direction at random. If the bacteria sense an increasing concentration of an attractant or a decreasing concentration of a repellent, then the tumble probability decreases. Thus, the bacteria persist longer in favorable behavior. This engenders a chemotactic drift.

The described behavior of chemotactic bacteria has been suggested as the basis for a mathematical algorithm for finding a point that maximizes some function f of three variables x, y, z. (See (13) for a recent exposition.) A "chemotactic" point (x,y,z) runs and tumbles until it approaches a maximum of f.

Suppose now that two different desiderata are simultaneously present, to maximize a function f and also to maximize a second function g. Now let the chemotactic point decrease its tumble probability if f is nondecreasing but g is increasing, or if g is nondecreasing and f is increasing. Then one expects that the point will approach the "Pareto optimum" set, where no change of (x,y,z) can result in an increase of either f or g without decreasing the other. Of course, the principles enunciated work for functions of arbitrary numbers of variables.

The suggested scheme could in principle improve the performance of a biological system in the face of multiple independent goals, goals that alter as conditions change. (See (14) for examples of multiple goal systems, both biological and nonbiological.) Random shifts in tactics could be tried, in attempts to do better with respect to at least one goal and no worse with the others. A tactic that accomplishes this persists longer before the inevitable random shift to something new.

6. RELEVANCE TO BIOMEDICINE

6.1. Cancer Chemotherapy

In principle, any advance in understanding immune system physiology should eventually have medical relevance. In §2 a model of monocyte and PMN hematopoiesis was shown to be of possible relevance to cancer chemotherapy. Reciprocally, observations of the reaction to the chemotherapy supplied data that were essential to calibrate the model.

6.2. Cytokine Therapy

There may be a "long-shot" impact on biomedicine of regarding cytokines as bearers of information (§4.2).

The classical "command view" of cytokines suggests that suitable cytokines can be administered to patients in order to alter harmful physiological conditions. A problem here comes from the pleiotropic nature of cytokine action. Elevation of a cytokine's concentration may have harmful additional effects, aside from the desired response. Moreover, varying backgrounds of other cytokines may alter the "standard" consequence of administering a given cytokine.

There is an alternative approach to "commanding" an immune response that seems to the physician to be appropriate. The physician can try to present information to the patient's physiological systems, perhaps by suitably chosen and timed cytokine doses, perhaps by other means (see below), so that an appropriate response to that information would yield a desired affect. For example, it may be thought appropriate to treat an allergy patient by shifting Th2 dominance to Th1 dominance. Attempts can be made to command this shift by administering cytokines that are known to favor Th1 and/or suppress Th2. Alternatively, one might inject killed bacteria that normally evoke a Th1 response, or inject a panel of cytokines (with proper doses and timings) that are evoked by binding of characteristic epitopes of Th1-inducing pathogens. Countering physiological information that lead to harmful physiological phenomena can be regarded as an alternative version of physicians' classical approach to illness, trying to reverse the harmful symptoms.

6.3. Combating Multiple Organ Dysfunction

Is there any evidence that the "chemotactic scheme" of §5, randomness with drift, has actually been adopted by some organisms to improve performance with respect to multiple goals? Remarkably, there is a hint of such evidence from Buchman (this volume, Part III, chapter 7.2) in multiple organ dysfunction syndrome (MODS). He writes, "analysis suggested that the coupling between heart and lungs was not fixed but rather dynamic.... The inference is that health may be associated with a search through the space of possible interactions to find the one best suited to current physiologic challenges." It might just be that the suggested directed random search for better performance in the face of multiple goals is a fundamental principle of human physiology.

One could even speculate further. Buchman gives as one seeming inadequacy of conventional MODS treatment that mortality is increased when calcium is administered to correct the subnormal values of calcium characteristic of sepsis. In view of such findings, perhaps it is worth thinking of experimental intensive care with a "chemotactic strategy." This would involve continually shifting and partially random choices of therapies to adjust abnormal values of the variety of physiological variables that are traditionally sensed by intensive care monitors. Shifts are less frequent when the present tactic is sensed to be

successful in improving at least one of the multiple independent goals of the multi-organ physiological system.

7. ACKNOWLEDGEMENTS

This work was partially supported by the Santa Fe Institute, both in the short term and also in conversations over the years among the theoretical immunology group.

Sadly, during the course of the production of this textbook, Professor Lee Segel passed away. Both the editors and publisher are immensely grateful for the important chapter Lee was able to contribute. He will be sorely missed throughout the scientific community he influenced so much. [Added in proof.]

8. APPENDIX: EQUATIONS FOR THE MATHEMATICAL MODEL

$$\frac{dN}{dt} = sE - g_N N, \quad \frac{dP}{dt} = rP - aNEP\,, \qquad \text{[A1,2]}$$

$$\frac{dE}{dt} = \mu_p PE - g_E E, \quad \delta = \frac{1}{T}\int_0^T [h_P P(t) + h_N N(t)]dt\,, \qquad \text{[A3,4]}$$

$$\frac{dK}{dt} = c_K(aEPN) - g_K K, \quad \frac{dH}{dt} = C_H(h_P P + h_{_N} N) - g_H H\,. \qquad \text{[A5,6]}$$

9. REFERENCES

1. Segel LA. 2001. How can perception of context improve the immune response. In *Autoimmunity and emerging diseases*, pp. 169–191. Ed. L Steinman. The Center for the Study of Emerging Diseases, Jerusalem.
2. Shochat E, Stemmer S, Segel LA. 2002. Detailed reconstruction of human hematopoietic cell kinetics in steady state and following highly myelotoxic perturbations. *Bull Math Biol* **64**:861–886.
3. Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M. 1995. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* **373**:123–126.
4. Segel LA, Lev Bar-Or R. 1999. On the role of feedback in promoting conflicting goals of the adaptive immune systems. *Immunology* **163**:1324–1349.
5. Segel LA, Safro I. 2003. Collective stochastic versions of playable games as metaphors for complex biosystems: team connect four. *Complexity* **8**:46–55.
6. Borghans JAM, De Boer RJ. 2002. Memorizing innate instructions requires a sufficiently specific adaptive immune system. *J Immunol* **14**:525–532.
7. Mezhitov R, Janeway CA. 1997. Innate immunity: impact on the adaptive immune response. *Curr Opin Immunol* **9**:4.

8. Segel LA. 2001. How does the immune system see to it that it is doing a good job? *Graft* **4**:15–18.
9. Bergmann C, van Hemmen JL, Segel LA. 2002. How instruction and feedback can select the appropriate T helper response. *Bull Math Biol* **64**:425–446.
10. Friedland JS, Shattock R, Remick DG, Griffin GE. 1993. Mycobacterial 65-kD heat shock protein induces release of proinflammatory cytokines from human monocytic cells. *Immunology* **91**:58–62.
11. Janeway Jr CA, Travers P. 1994. *Immunobiology*. Blackwell Scientific, Oxford.
12. Kooijman SA, Segel LA. 2005. How growth affects the fate of cellular metabolites. *Bull Math Biol* **67**:57–77.
13. Miller SD, Marchetto J, Airaghi S, Koumoustakaos P. 2002. Optimization based on bacterial chemotaxis. *IEEE Trans Evol Comput* **6**:16–29.
14. Segel LA, Cohen IR. 2001. *Design principles for the immune system and other distributed autonomous systems*. Oxford UP, Oxford.

4.2

MICROSIMULATION OF INDUCIBLE REORGANIZATION IN IMMUNITY

Thomas B. Kepler

Department of Biostatistics and Bioinformatics and Department of Immunology, Duke University, Durham, North Carolina

The immune system is the key interface between the multicellular host and its unicellular microbial commensals and pathogens. The cells of the immune system live a life balanced between unicellular autonomy—both physiological and genetic—and multicellular cooperation. Transitions between these two modes of operation are induced when the appropriate combination of events occurs and an immune response is triggered. These events may be the detection of tissue damage or of conserved microbial molecular motifs, and the detection of "non-self" epitopes. The transition is mediated by soluble and cell-surface bound signaling molecules, and involves the reorganization of cells from independently moving agents to functional aggregates. I have developed a model to study the properties of such transitions and to inform our thinking about the role of spatial organization in the immune response. It is an agent-based model evolving in continuous time and space, with its agents, representing cells of both microbe and host, interacting via direct contact and via soluble factors. These soluble factors, including host cytokines as well as bacterial chemoattractants and toxins, are represented as continuous fields governed by reaction diffusion partial differential equations. I illustrate the use of the model with an examination of microlocal inflammation.

1. INTRODUCTION

Somewhere around a billion years ago, the differentiated multicellular lifestyle had its debut. This strategy requires that cells abdicate their ability to proliferate unchecked and pass along to progeny the changes accumulating uniquely

Address correspondence to: Thomas B. Kepler, 101A North Building, Research Drive, Durham, NC 27708 (kepler@duke.edu).

in their genomes. This remarkable sacrifice gained these cells the ability to cooperate in the formation of complex spatial structures, which innovation opened new ecological niches and allowed these organisms to develop entirely new ways to make a living. The terminally differentiated somatic cells of plants and animals are provided stable and rich environments within the organism they constitute, and they entrust to their germline brethren the propagation of their common genome (1).

The ecological niches thus made exploitable were opened by virtue of the new functionality made possible by the formation of coherent spatial structures among the cooperating cells. These tissues and organs enabled the utilization of new food sources as well as the distribution of cellular resources to all constituents of the organism. This organizational innovation, in turn, created a new niche for unicellular life. Single-celled organisms are able to take advantage of the rich, climate-controlled environment of the multicellular organism even without cooperating in the soma-germline pact. Some of these associations have led to alternative versions of such a pact, and benefit both the host and the microbe, but not all of them have. There is ample opportunity for the noncooperating microbe to take advantage of the host internal milieu, and thereby become parasitic.

Those plants and animals that are unable to protect themselves from such parasitism suffer loss of replicative fitness relative to those conspecifics that are. The rich diversity and elaboration of various forms of host defense that are found throughout the multicellular phyla testify to the force of the host–pathogen relationship as an agent of selection. Indeed, parasitism is credited, according to one major hypothesis, as the primary reason for the advantage of sexual reproduction (2).

The unicellular lifestyle confers a number of advantages to microbial parasites—they have compact genomes and short generation times (and lack of sentimentality about death). As a result, they are capable of enormous genetic plasticity and rapid adaptation to uncertain and fluctuating environments. These advantages are so powerful that multicellular host defense has developed its own quasi-unicellular organization. The major components of innate immunity in animals are motile cells that circulate throughout the body and monitor their environment for various signs of infection and damage to host tissues. When such signs are encountered, these host cells carry out the effector functions of immunity: killing, disarming, or sequestering the pathogen.

The cells of the immune system also retain the advantages of multicellular tissue organization, however. Immune cells, known in vertebrates as leukocytes, are induced to abandon their strictly independent lives when signals of danger and pathogens are detected, and to reorganize into aggregates, coordinating their activities to a much higher degree than would otherwise be possible. Such behavior is evident in the inducible formation of such complex structures as granulomas in response to, e.g., schistosome ova (3) and myco-

bacteria (4), and germinal centers, where the remarkable process of affinity maturation occurs (5).

This reorganization is mediated by a complex array of signaling molecules. Long-range communication is accomplished by the secretion and binding of cytokines, of which there are over 200 distinct types in humans (6). A subset of the cytokines, the chemokines, is responsible for the regulation of cell migration and trafficking to bring the appropriate cells to the site of infection or the relevant lymphoid tissues (7,8). Short-range interactions are mediated by membrane-bound receptor-ligand systems that are engaged upon direct contact of cells bearing the complementary molecules.

I am developing microsimulation models to study this essential characteristic of immune systems. These models represent the cellular components as agents: computational objects with stochastic internal dynamics that unfold conditioned directly by the state of their "receptors," which states themselves depend on the concentrations of relevant ligands, both soluble and membrane-bound. These cellular agents migrate stochastically, responding to gradients of chemotactic and haptotactic molecules. These molecules, and other soluble factors, are treated as continua, modeled by partial differential equations with appropriate cell types as sources and sinks.

Both the spatial and temporal degrees of freedom are continuous, and the overall dynamics are given by a Markov process.

This chapter will just begin to address the larger questions of spatial organization and inducible reorganization of the immune response during its response to infection. I will consider the earliest stages of the immune response and focus attention on the phenomenon of microinflammation, by which I mean to indicate the dramatic increase in density of phagocytic cells and their aggregation in response to local infection, as well as the local increase in concentration of proinflammatory cytokines that accompanies and mediates this cellular reorganization.

The primary aims of the chapter are twofold. The first is to describe the modeling technique, field-coupled agent modeling (F-CAM), in general—the use of an agent-based approach for cellular components, their activities and motions coordinated through fields of continuous variables, the soluble factors, all evolving in continuous space and time. The second is to call attention to the potentially pivotal role of spatial organization in a phenomenon as uncomplicated (relatively speaking) as the phagocytic response to a sharply circumscribed bacterial infection. The importance of spatial organization is addressed within this volume in a strikingly different way in Part III, chapter 4.3 (by Perelson, Bragg, and Wiegel).

In what follows I will first describe key components of the model in mathematical detail, then touch briefly upon an illustration of its use to first analytically, then via simulation.

2. MODEL

2.1. General Forms

The components of the model that will be described here are the soluble factors, including cytokines, chemokines, their soluble receptors, and bacterial toxins; and cells, including bacteria and phagocytes.

2.2. Cells

The number of cells of any given type is not specified in advance, and will generally change through division, death, emigration and immigration. Each cell has an internal state with continuum and discrete components represented by the pair ($\boldsymbol{s}$,i) where the k-dimensional vector $\boldsymbol{s}$ is the continuum state and i indexes the discrete states. This internal state changes according to a Markov process whose rates are functions of the cell's external environment denoted by the vector $\boldsymbol{E}(t)$. If $f_i(\boldsymbol{s},t)dv$ is the probability of finding the cell within the infinitesimal volume element dv centered at the state ($\boldsymbol{s}$,i) at time t, the evolution of the probability function is given by the equation

$$\frac{\partial}{\partial t} f_i(\mathbf{s},t) = -\sum_{a=1}^{S} \frac{\partial}{\partial s_a}\left[g_{ai}(\mathbf{s},\mathbf{E}(t)) f_i(\mathbf{s},t)\right] + \sum_{j=1}^{S} K_{ij}(\mathbf{s},\mathbf{E}(t)) f_i(\mathbf{s},t)\,, \qquad [1]$$

which can be described by saying that with the discrete state i fixed, the state $\mathbf{s}$ obeys the deterministic ordinary differential equation

$$\frac{ds_a}{dt} = g_{ai}(\mathbf{s},\mathbf{E}(t))\,,$$

so that all of the stochasticity comes from the jumps from one discrete state to another (9).

All cellular types count among their internal variables their post-mitotic age and binary variables indicating vital status and mitotic status. Beyond these, any number of categorical or numerical degrees of freedom can be included.

In practice, the microsimulation generates sample paths of Eq. [1] using the corresponding Ito stochastic differential equation.

2.3. Cell Motility and Chemotaxis

The motions of the cells are described by stochastic processes that I have developed as straightforward generalizations of the Langevin equations.

Let the position of a cell be denoted $\boldsymbol{x}$ and its velocity $\boldsymbol{v}$. Then the motion of this cell is treated using the coupled Ito stochastic differential equations

$$d\mathbf{x} = \mathbf{v}dt, \qquad [2]$$

$$md\mathbf{v} = (F + \phi\mathbf{p} - \beta\mathbf{v})dt + dW_1\,, \qquad [3]$$

and

$$d\mathbf{p} = (\mathbf{I} - \mathbf{p}\mathbf{p}^{\dagger})(\gamma d\mathbf{W}_2 + \chi\mathbf{u}dt) - \gamma^2\mathbf{p}dt\,, \qquad [4]$$

where ***p*** is a unit-vector indicating the direction of the cell's polarization and ***u*** is a vector giving amplitude and direction of the bias due to the chemotactic factors. Eq. [2] simply states the usual relationship between velocity and position. Eq. [3] describes the variation in speed from both deterministic and random sources, given by the external forces ***F*** and the internal motive force $\phi\mathbf{p}$, on the one hand, and the fluctuating forces described by the three-dimensional Wiener process, $d\mathbf{W}_1$. The superscript dagger indicates matrix transpose. Eq. [4] describes biased diffusion of the polarization vector on the surface of the unit sphere. The first term gives the stochastic driving term of this process, with intensity γ; $d\mathbf{W}_2$ is another three-dimensional Wiener process independent of the first. The second term arises "automatically" and can be thought of as preserving the length of **p**. The last term represents bias toward the fixed direction given by the vector **u** whose magnitude gives the strength of the bias. The chemotactic constant χ gives the rate of adjustment toward the preferred direction.

The bias vector is related to the local chemokine concentration through

$$\mathbf{u} = \nabla\frac{c(\mathbf{x})^{\alpha}}{1 + \beta c(\mathbf{x})^{\alpha}} = \frac{\alpha c(\mathbf{x})^{\alpha-1}}{\left[1 + \beta c(\mathbf{x})^{\alpha}\right]^2}\nabla c(\mathbf{x})\,. \qquad [5]$$

2.4. Soluble Factors

I assign arbitrarily to each soluble factor an index represented by a roman letter, and to each individual cell an index represented by a greek letter, for ease of description. Designate by $c_i(\mathbf{x},t)$ the local concentration of soluble factor i at location **x** and time t. This concentration evolves according to reaction-diffusion equations of the form

$$\frac{\partial c_i}{\partial t}(\mathbf{x},t) = \sum_{\mu=1}^{n}\,[\sigma_{i\mu}(t) - \rho_{i\mu}(t)c_i(\mathbf{x},t)]\delta(\mathbf{x} - \mathbf{x}_{\mu}(t)) + \left\{D_i\nabla^2 - \sum_{j=1}^{m}\,R_{ij}c_j(\mathbf{x},t) - \lambda_i\right\}c_i(\mathbf{x},t)\,. \qquad [6]$$

The first term within square brackets on the right-hand side represents the secretion of factor i by individual cell μ. The secretion rate of this factor by cell μ is $\sigma_{i\mu}$, and the position of this cell is $\mathbf{x}_\mu(t)$. The positions of the cellular sources are in turn governed by stochastic differential equations (above). Similarly, the latter term in square brackets represents adsorption of factor i onto cell μ. Both the secretion and adsorption rates are time-dependent, since both depend on the state of cell μ, which itself changes in time. The first term within curly brackets represents diffusion, with diffusion coefficient D_i, where ∇^2 is the Laplacian operator. The sum in the middle of the bracketed term represents removal via complex formation with other soluble factors, where R_{ij} is the effective rate of removal of soluble factor i by complex formation with factor j. The last term represents degradation of factor i or its removal by mechanisms not explicitly treated in the model, such as complex formation with factors treated as existing at constant concentration or adsorption onto cells likewise treated as static.

For the examples in this chapter these partial differential equations were integrated using a semi-implicit forward scheme on a $80 \times 80 \times 80$ cubic grid.

2.5. Specific Forms

I have used this modeling platform to simulate the induction and resolution of microlocal inflammation during bacterial infection. This approach should be regarded as analogous to an in vitro experiment, in the sense that the components are simplified and highly controlled. The components of this specific study are nonmotile bacteria and phagocytes. Each bacterial agent has three discrete states and a continuous age, as well as position. Each bacterial agent secretes a soluble factor chemotactic to phagocytes. Several such factors are familiar, most notably N-formyl-methionyl-leucyl-phenylalanine (fMLP). These bacterial agents are subject to Brownian motion but are not polarized and do not move under their own power. This characteristic is not common to all bacteria, certainly, but many bacteria of medical significance, including *Bacillus anthracis* are nonmotile.

The phagocytic agents have three internal states: quiescent, activated, and refractory. The transition rate from quiescent to activated is zero in the absence of bacteria and proinflammatory cytokine, and is an increasing but saturating function of the local concentration of proinflammatory cytokine. This transition rate is finite and constant when in contact with bacteria; the rates from proinflammatory cytokine and bacterial contact are additive. While in the activated state, the phagocyte secretes proinflammatory cytokine at a constant rate until its transition to a refractory state. The refractory state is characterized by a loss of responsiveness to proinflammatory cytokine and bacterial contact, as well as cessation of PIC secretion. In addition refractory phagocytes shed a soluble receptor for PIC, which binds it and neutralizes its activity.

The soluble factors are the fMLP-like factor secreted by bacteria and chemotactic to phagocytes; the proinflammatory cytokine, modeled on tumor necrosis factor, produced by activated phagocytes, and chemotactic to them, and bound and sequestered by their soluble receptor, and the soluble receptor itself.

The system of interest to us for this chapter consists of two cells types and three soluble factors. The cells are bacteria and phagocytes. The bacterial agents have just the three obligate states: alive (nonmitotic), mitotic, and dead. The age-dependent hazard function for transition from nonmitotic to mitotic is $h(a) = \alpha H(a - T_{\mathrm{CYC}})$, where a is the age, α is a positive constant and T_{CYC} is the minimum cell-cycle time. The bacterial death rate depends on the presence of phagocytes; I neglect natural mortality. In the presence of one or more quiescent phagocytes in contact with the bacterium, the death rate is constant and equals 1/hr. In contact with activated phagocytes, the death rate is 20/hr. Mitotic cells are immediately replaced by a pair of age-zero daughter cells at locations adjacent to the location of the parent cell. Dead cells are immediately removed from the simulation.

For the bacterial agents, Eq. [1] becomes

$$\frac{\partial}{\partial t} f_C(a,t) = \frac{\partial}{\partial a} f_C(a,t) - \alpha H(a - T_{\mathrm{CYC}}) f_C(a,t) - (\mu \sum_{i=1}^{n_p} H(\|x_i - x\| - \theta) f_C(a,t) ,$$

$$\frac{\partial}{\partial t} f_M(a,t) = -\frac{\partial}{\partial a} f_M(a,t) + \alpha H(a - T_{\mathrm{CYC}}) f_C(a,t) ,$$

$$\frac{\partial}{\partial t} f_D(a,t) = \frac{\partial}{\partial a} f_D(a,t) - \alpha + \mu \sum_{i=1}^{n_p} H(\|x_i - x\| - \theta) f_C(a,t) ,$$

where H is the Heaviside function and α is a positive constant. The hazard function for death is zero in the absence of phagocytes, and is constant with value μ when there is a phagocyte (indexed from $i = 1$ to n_p) is within distance θ of the bacterium. The position in space of the bacterium is x and those of the phagocytes x_i.

The phagocytic agents have three internal states, but are never mitotic. These states are quiescent, activated, refractory, and dead. The stochastic transition from quiescent to activated depends on the presence of either bacteria or the proinflammatory cytokine (PIC). In the presence of one or more bacteria, the rate of this transition is 0.5/min. In the presence of the optimal concentration of PIC the rate is 2/min. As a function of the local concentration of PIC, the rate is zero for [PIC] = 0, and increases with [PIC], saturating at large [PIC]. In the activated state, phagocytes kill bacteria more rapidly than do quiescent phagocytes. Activated phagocytes go on to become refractory, in which state they are unresponsive to stimulation. The mean time spent in the activated state is inde-

pendent of local conditions and equal to six hours. In the refractory state, the phagocytes shed a form of the PIC-soluble receptor. The mean time spent in the refractory state is also unaffected by the environment, and equal to two hours. The estimates for these lifetimes are necessarily rough, and indeed the identification of these states as distinct physiological entities is to be taken as a simplification to be overcome as the model is refined.

The simulation volume is a 400-micrometer cube. Both cell types leave this cube, and the simulation, when their stochastic path intersects the boundary of this cube. To account for immigration of phagocytes as well as their emigration, phagocytes are produced temporally as a Poisson process with a uniform spatial distribution at the boundary of the cube.

To account for the increased extravasation of phagocytes induced by PIC, the Poisson intensity of the emigration process is an increasing, saturating function of total PIC concentration within the cube.

The soluble factors are, as mentioned, a proinflammatory cytokine (PIC), a soluble receptor for the PIC (sPICR), and a bacterial chemoattractant (BC). In summary their properties are:

1. **PIC** is (1) secreted by activated phagocytes at a rate of 240 molecules/min. (2) PIC induces activation of quiescent phagocytes. (3) PIC is chemoattractive to phagocytes. (4) PIC strongly enhances extravasation (immigration) of quiescent phagocytes.
2. **sPICR** is produced by refractory phagocytes and binds to PIC in solution in a 1–1 stoichiometry, removing both from further activity. It has no other function.
3. **BC** is produced by the bacteria and is a chemoattractant to phagocytes; it has no other function.

3. RESULTS

3.1. Chemotactic Superposition

For stationary constant sources, the steady-state solution for soluble factors described by Eq. [6] is expressible in closed form:

$$\bar{c}(r) = \frac{\sigma\delta}{4\pi D^2 r} \exp\left\{-r\sqrt{\frac{\delta}{D}}\right\},$$

where r is the distance from the source. Figure 1 shows the net chemotactic signal from Eq. [5] for the case where the responding cell possesses two independent sets of chemotactic receptors. That is, the contours in this figure show the

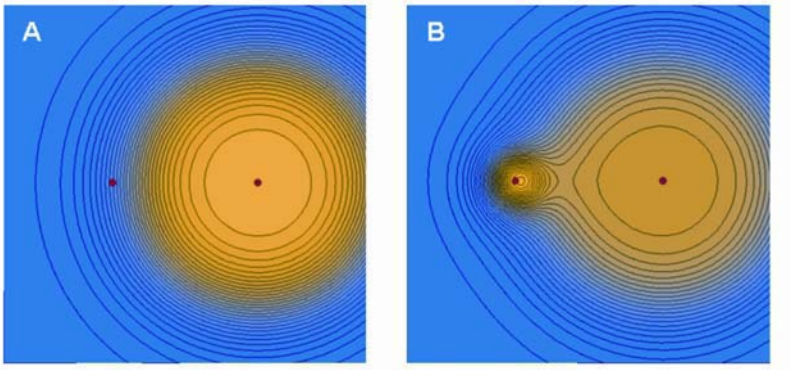


Figure 1. The ligand-bound fraction for the receptors on the surface of a hypothetical probe cell at any location in the simulation volume. The source of chemokine one is at the position of the right-hand red disk. In panel **B** a second chemokine has its source at the position indicated by the left-hand red disk; the strength of the second source is one-tenth that of the first. Curves are contours of equal net receptor occupancy. The induced polarization of the chemotactic probe cell is perpendicular to the contour lines and pointed toward the sources. The computation is for a three-dimensional system; the graph depicts the plane that contains both sources and the probe.

sums of the receptor occupancies for the two types. The gradient of this function is the direction of chemotactically induced cell polarization. Panel A shows the occupancy contours for a single source. Note that the contours flatten markedly near the source. Panel B shows the occupancy contours for the original source, plus a second source (on the right) of another chemotactic factor (on the left). The strength of the left-hand source is just 10% that of the source on the right. For this reason, its effective radius is smaller, but its location, nevertheless, approximates the location of the global maximum, and hence the ultimate destiny of cells obeying Eqs. [2]–[4]. This circumstance makes it possible to use host-derived chemokines (represented by the source on the right) to increase the overall radius of capture of phagocytes and other cells of the immune system, while partially mitigating the effects of redirecting the chemotactic motion away from the pathogen. The experimental results indicate that a second process is involved: the host chemokine receptors are downregulated when occupied at high density.

3.2. Microlocal Inflammation

The innate immune response depicted in these simulations can be decomposed into three phases. The first phase is recognition. Phagocytic cells wander

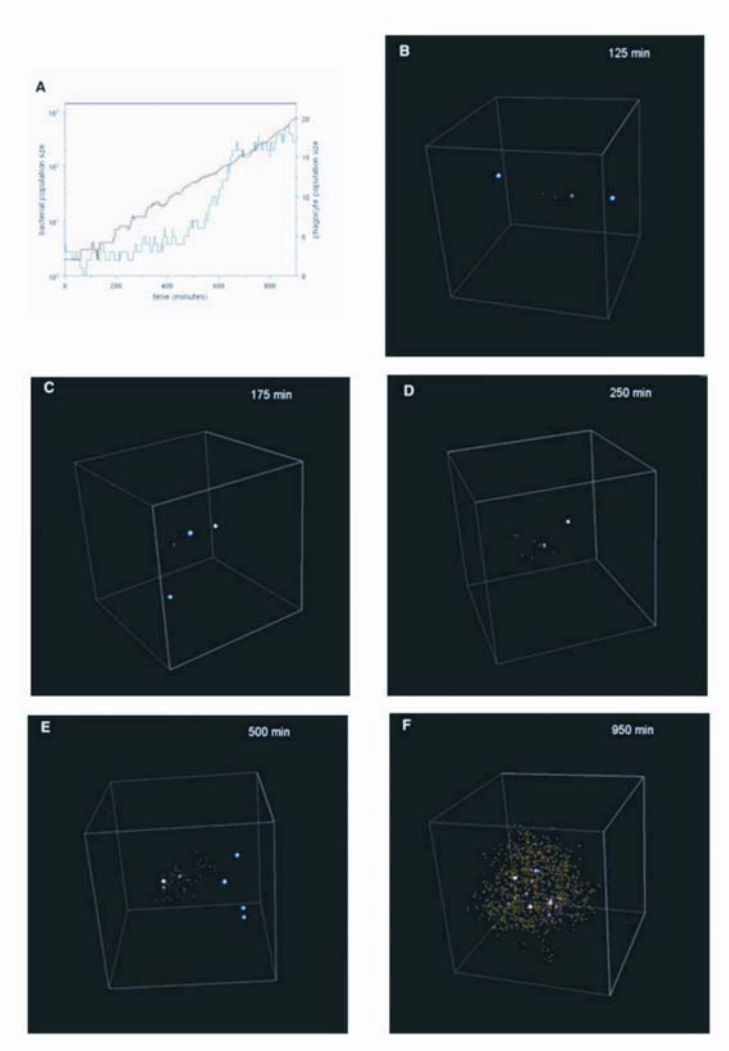


Figure 2. For the simulation run depicted here, the phagocytes produce neither PIC nor sPICR. The mean cell-cycle time for bacteria is 72 minutes. (**A**) The population size as a function of time for a single sample path. The total population for phagocytes is shown in blue; total bacterial population is shown in yellow. (**B–F**) Screen shots of the visualization tool showing the three-dimensional snapshots of the agents' motions. Yellow agents represent bacteria. Phagocyte agents are blue when quiescent, red when activated, and white when refractory. The simulation volume, outlined as a white cube, is rotating from the viewer's perspective.

randomly in the tissue until encountering the bacteria. During this time the bacteria grow unchecked. It is a substantial advantage to the host to detect the bac-

teria as early as possible during this growth. The second phase is characterized by phagocytic activation, an increase in the density of phagocytes to the site of infection, and bacterial killing. The final phase is marked by increased accumulation of anti-inflammatory mediators, decrease of PIC concentration, dissipation of phagocytes, and return to pre-inflammatory homeostasis.

With the parameters and initial conditions in the system described here, the phagocytes typically detect the bacteria within an hour and are capable of halting the infection by killing the bacteria if the mean cycle time is substantially longer than that time, even without the additional recruitment and chemotaxis provided by PIC.

When the bacterial division time becomes comparable to the discovery time, the phagocytes cannot adequately keep pace with the bacterial growth, and the bacterial populations tend to grow virtually unchecked (Figure 2). Note that the phagocyte density does increase in the intermediate stages of the process (Figure 2A). The bacterial chemokines, however, because of their effective lifetime, are of such short range that they cannot increase local influx of phagocytes. Thus, the observed increase is due entirely to slowing of emigration by virtue of phagocyte *capture* by the BC. Phagocytes wander in, but they do not wander back out.

The phagocyte advantage is regained upon inclusion of PIC (Figure 3). Once activated phagocytes produce PIC, the recruitment of new phagocytes is substantially enhanced, and the effective radius of their capture is enlarged. Furthermore, the saturating nature of the chemotactic response to superposed attractants ensures that the host-derived chemokines does not overwhelm the microbe-derived factors and act as a decoy chemotactic destiny, as described above.

On the other hand, the positive feedback loop set up by PIC causes a sustained local inflammation. The resolution of this inflammation requires breaking this loop, a function served by the sPICR. The refractory cells are unresponsive to further stimulation by PIC and shed soluble receptor, which binds and neutralizes the PIC. Figure 4 shows that the evolution of the response with fully sPICR-competent phagocytes is very similar to that of the sPICR-deficient phagocytes in Figure 3 through the first two phases, but differs precisely at the end of the activation phase and throughout the final phase—resolution.

4. DISCUSSION AND CONCLUSION

The use of agent-continuous hybrid modeling for complex cellular systems has really just begun; there is a great deal of work to be done in bringing greater detail to the components, fidelity to their interactions, and scope to the phenomena explored. The extraordinary advances in single-scale manipulation, in vivo cellular and molecular imaging, and of course, computational speed, memory, and software for distributed computing together promise that this effort will

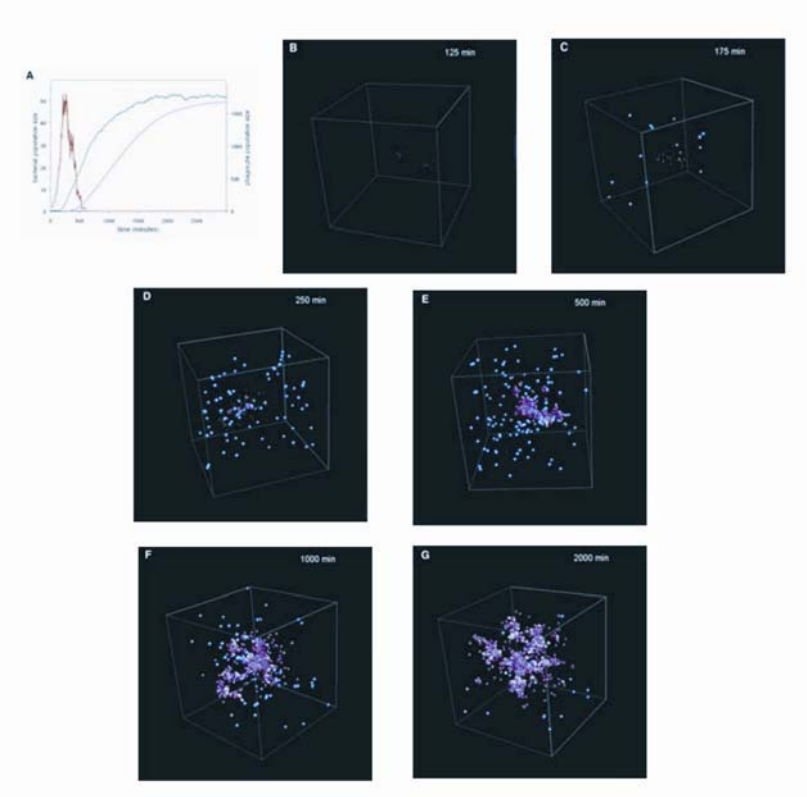


Figure 3. In this simulation run the conditions are as in Figure 1, except that phagocytes produce PIC and the bacterial mean cell-cycle time is 72 minutes. The phagocytes still do not produce sPICR. (**A**) The blue and yellow lines again depict phagocytes and bacteria, respectively; the magenta line represents the total PIC concentration. (**B–G**) As in Figure 1.

result in greatly enhanced confidence in the results of these carefully calibrated models, and that they will provide insight and play the role of valuable adjuncts to empirical investigation for issues as critical and diverse as vaccine efficacy and autoimmune disease.

5. ACKNOWLEDGMENTS

I am grateful, as always, for invaluable discussions shared with the members of the Duke University Laboratory of Computational Immunology—particularly Lindsay Cowell and John Tomfohr—and for access to prepublica-

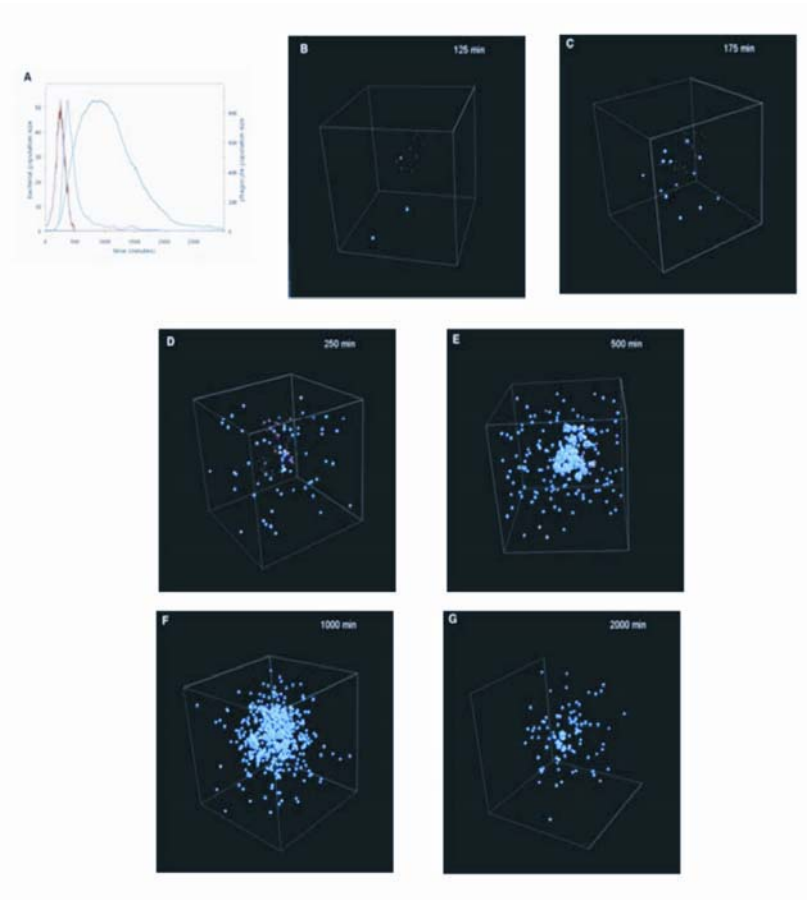


Figure 4. In this simulation run, the conditions are as in Figure 2, but now the phagocytes are sPICR- as well as PIC-competent.

tion data from Mark Miller, then at UC-Irvine; for the collegiality of the members of the Duke Human Vaccine Institute and the financial support of the NIAID through award number AI058227-01, and of the Duke University Center for Translational Research through its NIH award number AI051445-03. I am thankful as well to Tom Deisboeck and Yasha Kresh for their perseverance and patience.

6. REFERENCES

1. Buss LW, 1987. *The evolution of individuality*. Princeton UP, Princeton.
2. Hamilton WD, Axelrod R, Tanese R. 1990. Sexual reproduction as an adaptation to resist parasites. *Proc Natl Acad Sci USA* **87**:3566–3573.
3. Pearce EJ, MacDonald AS. 2002. The immunobiology of schistosomiasis. *Nature Rev Immunol* **2**:499–511.
4. Algood HMS, Chan J, Flynn JL. 2003. Chemokines and tuberculosis. *Cytokine Growth Factor Rev* **14**:467–477.
5. Berek C, Berger A, Apel M. 1991. Maturation of the immune response in germinal centers. *Cell* **67**:1121–1129.
6. Oppenheim JJ, Feldmann M, Durum SK. 2001. *Cytokine reference: a compendium of cytokines and other mediators of host defense*. Academic Press, London.
7. Moser B, Loetscher P. 2001. Lymphocyte traffic control by chemokines. *Nature Immunol* **2**:123–128.
8. Cyster JG. 1999. Chemokines and cell migration in secondary lymphoid organs. *Science* **286**:2098–2102.
9. Kepler TB, Elston TC. 2001. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys J* **81**:3116–3136.

4.3

THE COMPLEXITY OF THE IMMUNE SYSTEM: SCALING LAWS

Alan S. Perelson

Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico

Jason G. Bragg

Department of Biology, University of New Mexico, Albuquerque

Frederik W. Wiegel

Institute of Theoretical Physics, University of Amsterdam, The Netherlands

We provide a brief overview of scaling principles for the immune system. Larger animals tend to live longer than smaller animals and thus their immune systems need to do a better job of protecting them. We then ask how the features of the immune system scale with body mass. A larger animal has more lymphocytes than a smaller one. This implies either more lymphocyte clones or more cells per clone, or both. The immune system also has anatomical features such as lymph nodes. Thus, as animals get larger, do lymph nodes simply get larger or are they more numerous? If appropriate scaling relations can be developed, they could help us understand the relationship between the human immune system and that of other species. A more informed approach to the scaling relationships among immune systems of different organisms would be nothing but helpful.

1. INTRODUCTION

The immune system is a complex system responsible for protection against pathogenic agents. Pathogens can reside in any tissue of the body and the immune system needs to find and respond to them. By necessity the immune system is distributed and the cells and molecules that make up the system move or

Address correspondence to: Alan S. Perelson, Theoretical Division, Los Alamos National Laboratory, MS-K710, Los Alamos, NM 87545 (asp@lanl.gov).

are transported throughout the body. The cells of the adaptive immune system are a type of white blood cell called lymphocytes. These cells are transported by the blood but can exit the blood stream and crawl through the tissues of the body. They return to the blood in the lymph fluid that bathes cells and that is collected in a system of ducts called lymphatics, which ultimately connect with the blood stream.

The best-studied immune systems are those of the mouse and the human. The mouse has about 10^8 lymphocytes, while a human has about 10^{12}. Immune systems are found in all vertebrates, and thus organisms as small as a tadpole (body mass on the order of 10^{-1} g) and as large as an elephant (10^6 g) or a whale (10^8 g) have immune systems. Here we address the question of whether there exist any scaling principles that can guide our understanding of the operation of the immune system in organisms that differ by nine orders of magnitude in mass. Scaling principles are introduced in Part II, chapter 3 (by Savage and West) of this volume.

For much of its history immunology has been a subfield of microbiology and closely linked with medicine. As such, studies of the immune systems of diverse species have not attracted much attention or much funding, and there is a paucity of data about the immune system of most species. We know of only one other group that has theoretically addressed the question of the scaling of the immune system (1). Thus, the goal of this chapter is to raise questions, and provide a brief overview of what is known about scaling laws in the immune system.

1.1. The Protecton Hypothesis

Langman and Cohn (1) suggested that the immune system has a modular structure and is built of basic units called "protectons." Each protecton guarantees an adequate immune response in some volume element of the animal. Thus, in the Langman-Cohn view, a big animal simply has more protectons in its immune system than a small animal. Langman and Cohn also estimated, based on the concentration of antibody needed to protect an animal, that a protecton contains about 10^7 B cells in a volume of about 1 ml. In terms of scaling, the protecton idea suggests that the size of the immune system scales as the mass of the organism.

While modularity is a desirable property of any large complex system, we argue against this strict point of view and the simple scaling proportional to mass ($\sim M$) for a number of reasons:

(1) Transport of lymphocytes depends on the circulatory system. As shown by West, Brown, and Enquist (2), properties of the circulatory system do not scale $\sim M$, but rather as $\sim M^{3/4}$.

(2) The immune system has architectural features, which include a single spleen and a single thymus in mammals of differing size. Thus, at least some components of the immune system are not modular.

(3) A larger animal lives longer than a smaller one. Hence, its immune system has to do a better job of protecting it.

2. SCALING LAWS IN IMMUNOLOGY

It has been observed that the average lifespan (T_0) of a mammal seems to scale with its body mass (M) according to the scaling law $T_0 \sim M^{1/4}$ (cf. (2,3)). A scaling law between a biological variable Y and body mass M is written in the form $Y \sim M^b$, and b is called the scaling exponent. This is shorthand for an approximate, quantitative relation

$$\frac{Y}{Y_0} \cong A\left(\frac{M}{M_0}\right)^b, \qquad [1]$$

where Y_0 is a standard unit with the same dimension as Y, M_0 is a standard unit of mass, and A is a dimensionless constant. Here we shall take the point of view that death is generally not due to failure of the immune system, and thus the mammalian immune system should be designed in such a way that it can protect an organism during a lifetime $T_0 \sim M^{1/4}$.

If the immune system of a larger animal must help keep that animal alive for longer periods than the immune system of a smaller animal, it must be more reliable. A larger animal has more B and T lymphocytes. This implies either more lymphocyte clones or more cells per clone, or both. This suggests the following question: what is the optimal way for the system to balance these two modes of resource allocation—T and B cell diversity versus clone size?

2.1. Scaling of B and T Cell Clone Size

In order to derive the typical size of a lymphocyte clone as a function of M, we follow the model of West et al. (2), in which the circulatory system is represented by a branching tree.

In the West, Brown, and Enquist (or WBE) model the organism is divided into a certain number of small units, each of which is supplied by a single capillary. These units, called *service volumes* or *service units*, are regions that a sin-

gle capillary can supply with oxygen and remove waste products from. The number of service units scales $\sim M^{3/4}$ (2,3), which implies that the volume of a service unit scales as $\sim M^{1/4}$. If we assume that a service unit is spherical, its radius (R) will scale $\sim M^{1/12}$.

Now to connect the WBE model with immunology, we assume that the service volumes for the blood circulation are also the service volumes for immune surveillance, that is, the capillaries allow lymphocytes to exit the circulatory system and explore regions of tissue for foreign molecules and cells, collectively called antigens. This implies that if each clone of B cells or T cells contains at least $\sim M^{3/4}$ cells it can be represented in each of the $\sim M^{3/4}$ service volumes.

One of the essential ideas in the WBE model is that the capillary that supplies a service volume is universal in its properties, such as its diameter. This implies that the amount of blood delivered to the service unit, per unit of time, is independent of M. From the point of view of immune surveillance, most antigens will enter the service unit in the blood. In other words, the number of antigens that enter into a service unit per unit of time is independent of M.

2.2. The Time to Find an Antigen in a Service Unit

Consider a single antigen and one specific lymphocyte, of some clone that is specific for the antigen, both located in the same service unit. This lymphocyte will crawl within the service unit in a more or less random fashion, searching for antigen. How long does it take until it makes first contact with the antigen?

If one describes the random walk of the lymphocyte as spatial diffusion with a diffusion coefficient D, then one can show that T, the average time until first contact between the lymphocyte and antigen, is given by (4)

$$t = \frac{R^3}{3D\ell}, \qquad [2]$$

where ℓ is the sum of the radii of the lymphocyte and the antigen, and R is the radius of the service unit.

If the diffusion coefficient D is independent of M, then as $R \sim M^{1/12}$, $T \sim M^{1/4}$. Thus, if there were only one lymphocyte per clone present in the service unit, the antigen could go undetected by that lymphocyte for a period of time ($\sim M^{1/4}$) that increases with animal size. Since the search time for each clone should scale in the same manner, this result applies to all clones. In order to keep the time until detection a fixed value (smaller than the time during which the antigen could proliferate significantly) the organism has to put $\sim M^{1/4}$ copies of the lymphocyte into this service unit; this would reduce the mean time until first detection by a factor of $\sim M^{1/4}$ to a value independent of M. We conclude that if D is

independent of M the lymphocyte clone size should scale $\sim M^{3/4} M^{1/4} = M$, where the first factor comes from the number of service units and the second one from the number of lymphocytes needed per unit. In (4) we also examine the case in which lymphocyte movement and antigen growth both depend on the basic metabolic rate of an organism and also conclude that for that case clone size $\sim M$.

2.3. Scaling of the Lymphocyte Repertoire

Next, we ask how many different clones of lymphocytes should be present in an animal of mass M. The number of different clones is called the size of the immune repertoire and determines how many different antigens the immune system can recognize.

We assume that antigens mainly enter the body through our intake of food, liquids, and air. The rate at which a mammal consumes food and air is governed by its metabolic rate. One can show that the lifetime total metabolic activity of a mammal scales as $\sim M$ (4), suggesting that a mammal needs to deal with cM antigens during its lifetime, where c is some constant.

In order to assess the probability that an immune system with a repertoire of size N can recognize an antigen, Perelson and Oster (5) introduced the idea of shape-space. In this theory it is assumed each lymphocyte has a receptor that can recognize antigens in a volume v_0 of shape space, which has total volume V. If we let ε be the probability of the immune system, i.e., all N different clones, failing to recognize an antigen, then

$$\varepsilon = \left(1 - \frac{v_0}{V}\right)^N \cong \exp\left(-N\frac{v_0}{V}\right). \qquad [3]$$

The probability of a successful immune response to an infection is then $1 - \varepsilon$, and the probability, P_s, that the organism will successfully repel all cM infections during its lifespan is given by

$$P_s = (1 - \varepsilon)^{cM} \cong \exp(-\varepsilon cM). \qquad [4]$$

This probability should be very near to unity, so we require $\varepsilon cM << 1$ or

$$\varepsilon = \exp\left(-N\frac{v_0}{V}\right) << \frac{1}{cM}. \qquad [5]$$

This in turn implies

$$N >> \frac{V}{v_0} \ln (cM), \quad [6]$$

which shows that N, the repertoire size, should scale ~ln(cM). Thus, this theory suggests that there should be only a weak dependence of repertoire size on a mammal's size.

2.4. Scaling and the Anatomical Features of the Immune System

The lymphocytes of the mammalian immune system not only circulate throughout the body but also accumulate in the spleen and lymph nodes. These tissues act as filters, with the spleen trapping antigens from the blood and lymph nodes trapping antigens that enter the tissues. Because antigens are there, lymphocytes search for and interact with antigens in the "secondary lymphoid tissues." An interesting question then is how should the size or mass of the spleen and lymph nodes scale with body size? Each lymph node "drains" a certain volume of tissue. Thus, as animals get larger, do lymph nodes simply get larger or are they more numerous? An appropriate scaling theory of the immune system should be able to answer these questions. Here, as a first step, we look to see what data are available and if there is any indication of scaling that is more complex than that of scaling simply by mass, i.e., ~M.

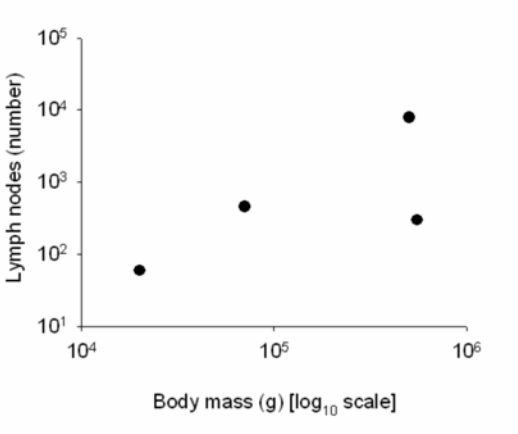


Figure 1. Number of lymph nodes per individual as a function of body mass for dog, human, horse, and cow (in order of mass).

Stahl (6) found that spleen mass scaled with body mass as $\sim M^{1.02}$ across diverse mammal species, or as $\sim M^{0.85}$ when limited to primates only. More recently, Nunn (7) found that spleen mass scaled with body mass as $\sim M^{1.17}$ across primates, based on an analysis that included some of the data used in (6), but which used mean mass values for species as data points, rather than values for individual animals.

Estimates for the number of lymph nodes in dogs, humans, horses, and cows are available (8), and in Figure 1 are plotted against adult body masses typical for these species. While there is some suggestion of an increase in the number of lymph nodes with mass, it is difficult to evaluate the slope reliably with data for so few species, and where body masses and lymph node counts were not available for the same individuals.

Concentrations of lymphocytes in the blood have been compared among primate species, using data compiled by the International Species Information System (9,10). Using a similar database (11), we find that among diverse mammal species, adult blood lymphocyte concentrations scale with adult body mass as $\sim M^{-0.07}$ ($n = 138$, $F = 18.53$, $p < 0.001$, $r^2 = 0.12$), and as $\sim M^{-0.1}$ ($n = 45$, $F = 18.78$, $p < 0.001$, $r^2 = 0.30$) when limited to primates. This implies that if blood volume scales with mass as $\sim M^{1.02}$ (12), total number of lymphocytes will scale as $\sim M^{1.0} M^{-0.1} \sim M^{0.9}$.

Data are also available for lymphocyte output from the thoracic duct (13), which is the major lymphatic vessel through which lymph enters the circulatory system (14). Among nine species of mammals, concentration of lymphocytes in the lymph decreases with body mass as $\sim M^{-0.16}$ (though this relationship is not significant at the $p = 0.05$ level; $n = 9$, $F = 3.49$, $p = 0.10$, $r^2 = 0.33$), while flow from the thoracic duct increases with body mass as $\sim M^{0.89}$ ($n = 9$, $F = 85.10$, $p < 0.001$, $r^2 = 0.92$). It then follows that total lymphocyte output per unit time scales with body mass as $\sim M^{0.73}$ ($n = 9$, $F = 44.40$, $p < 0.001$, $r^2 = 0.86$; Figure 2), which is intriguingly similar to the $M^{3/4}$ scaling laws found in the WBE theory (2,3).

3. CONCLUSIONS

The use of scaling laws in immunology is a field in its infancy. We have tried here to show that scaling laws can provide insights into the properties of immune systems in animals of different sizes. Although we focused our attention on the number and size of lymphocyte clones needed to provide protection to animals of different mass, there is a great need to also understand the anatomical features of immune systems in different mammals. The lymphatic system, with its chains of lymph nodes, is organized to some extent as an inverse branching network, with collecting lymphatics joining together to form larger vessels. Along the way are lymph nodes that filter the fluid as it returns to the circulation. The scaling relations that govern the operation of the circulatory

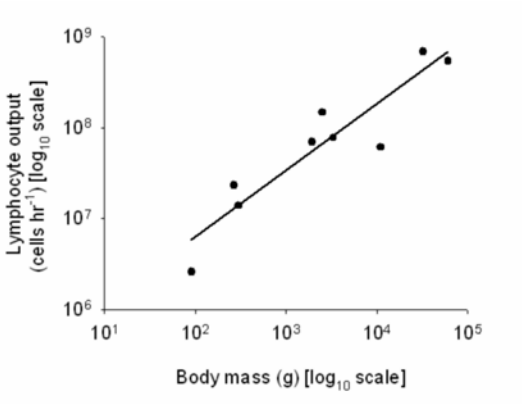


Figure 2. Total lymphocyte output from the thoracic duct plotted as a function of body mass for the hamster, rat, guinea pig, cat, rabbit, monkey, dog, goat, and human (in order of mass). Equation for the fitted line is $y = 211930x^{0.73}$.

system have been well-studied (see (2)). An analogous theory for the lymphatic system still needs to be developed. Hopefully such a theory will address questions such as how does the size and number of lymph nodes scale with animal size? Do big animals have more lymph nodes or bigger lymph nodes than smaller mammals? Also, if scaling relations can be developed, they may open a window into a better understanding of the relationship between the human immune system and that of other species, which are commonly used as "model systems" for studying the effects of drugs and immune system modifiers. A more informed approach to such studies would be nothing but helpful.

4. ACKNOWLEDGMENTS

We thank Geoffrey West for valuable conversations about this work. Portions of this work were done under the auspices of the US Department of Energy under contract W-7405-ENG-36 and supported by NIH grant AI28433. This work was also facilitated by interactions at the Santa Fe Institute and supported at the Santa Fe Institute by a grant from the Joseph P. and Jeanne M. Sullivan Foundation.

5. REFERENCES

1. Langman RE, Cohn M. 1987. The elephant-tadpole paradox necessitates the concept of a unit of B cell function: the protecton. *Mol Immunol* **24**:675–697.
2. West GB, Brown JH, Enquist BJ. 1997. A general model for the origin of allometric scaling laws in biology. *Science* **276**:122–126.
3. Brown JH, West GB. 2000. *Scaling in biology*. Oxford UP, New York.

4. Wiegel FW, Perelson AS. 2004. Some scaling principles for the immune system. *Immunol Cell Biol* **82**:127–131.
5. Perelson AS, Oster G. 1979. Theoretical studies of clonal selection. *J Theoret Biol* **81**:645–670.
6. Stahl WR. 1965. Organ weights in primates and other mammals. *Science* **150**:1039–1042.
7. Nunn CL. 2002. Spleen size, disease risk and sexual selection: a comparative study in primates. *Evol Ecol Res* **4**:91–107.
8. Altman PL,, Dittmer DS. 1974. *Biology data book*, Vol. 3, 2nd ed. Federation of American Societies for Experimental Biology, Bethesda, MD.
9. Nunn CL, Gittleman JL, Antonovics J. 2000. Promiscuity and the primate immune system. *Science* **290**:1168–1170.
10. Nunn CL. 2002. A comparative study of leukocyte counts and disease risk in primates. *Evolution* **56**:177–190.
11. International Species Information System. 2002. *Reference ranges for physiological values in captive wildlife*. Minnesota Zoological Garden, Apple Valley.
12. Stahl WR. 1967. Scaling of respiratory variables in mammals. *J Appl Physiol* **22**:453–460.
13. Altman PL,, Dittmer DS. 1971. *Biological handbooks: respiration and circulation*. Federation of American Societies for Experimental Biology, Bethesda, MD.
14. Goldsby RA, Kindt TJ, Osbourne BA. 2000. *Kuby immunology*, 4th ed. Freeman, New York.

5.1

NEUROBIOLOGY AND COMPLEX BIOSYSTEM MODELING

George N. Reeke Jr.

Laboratory of Biological Modeling,
The Rockefeller University, New York

This chapter gives a brief summary of techniques for modeling neural tissue as a complex biosystem at the cellular, synaptic, and network levels. A sampling of the most often studied neuronal models with some of their salient characteristics is presented, ranging from the abstract rate-coded cell through the integrate-and-fire point neuron to the multicompartment neuron with a full range of ionic conductances. An indication is given of how the choice of a particular model will be determined by the interplay of prior knowledge about the system in question, the hypotheses being tested, and purely practical computational constraints. While interest centers on the more mature art of modeling functional aspects of neuronal systems as anatomically static, but functionally plastic adult structures, in a concluding section we look to near-future developments that may in principle allow network models to reflect the influence of mechanical, metabolic, and extrasynaptic signaling properties of both neurons and glia as the nervous system develops, matures, and perhaps suffers from disease processes. These comments will serve as an introduction to techniques for modeling tumor growth and other abnormal aspects of nervous system function that are covered in later chapters of this book (Part III, §6). Through the use of complex-systems modeling techniques, bringing together information that often in the past has been studied in isolation within particular subdisciplines of neuro- and developmental biology, one can hope to gain new insight into the interplay of genetic programs and the multitude of environmental factors that together control neural systems development and function.

Address correspondence to: George N. Reeke Jr., Laboratory of Biological Modeling, The Rockefeller University, 1230 York Avenue, New York, NY 10021 (reeke@rockefeller.edu)

1. NEURONAL SYSTEMS DYNAMICS

The nervous system is a complex object, and so the modeling of neuronal systems dynamics is necessarily a complex subject. Space here permits only the most superficial survey of some of the more common techniques. References are given to textbooks and original literature where more details can be found. Different approaches are available depending on one's general philosophy about the goals of modeling. Some believe, for example, that models should be used only to test very explicit hypotheses, while others are happy to use models in a looser mode simply to explore the consequences of a particular network architecture or synaptic-efficacy modification rule. In evaluating published models, one should always ask oneself to what extent the assumptions were chosen to give inevitably the desired results (that is, whether the results were "built into" the model from the start) or whether the assumptions were justified for principled reasons other than their success at predicting the desired results. To avoid this potential pitfall, some authors prefer a so-called "bottom–up" approach, in which model features are derived as much as possible from experimental data. Others prefer a "top–down" approach, in which a particular theoretical point of view, for example, one based on a computational metaphor, is used to guide the choice of assumptions. The former class of models is often referred to as "neurally realistic," whereas the latter may be only "neurally inspired." On the other hand, it can be argued that simply putting all the known facts into a model may produce a good emulation of some brain function, but may do little to indicate which particular features of the nervous system make that function possible. In the end, these decisions become a matter for expert debate. For our purposes, we would like to understand better how the brain operates, of course, because this is one of the remaining great questions in modern biology, but more specifically so we can relate the basic principles of brain operation to the disruptions that occur in pathological conditions, particularly tumor growth, a subject treated in Part III, §6 (this volume). This would suggest that a rather detailed "bottom–up" approach would be most appropriate, keeping in mind always that computational limits to our simulations are likely for some time to remain tighter than those imposed by biology on the brain. Therefore, we must be careful to distinguish these two sources of constraint when interpreting brain models.

1.1. Single-Neuron Models

We begin by asking how detailed a model of a single neuron must be in order to make it useful for modeling complex neural systems (1). The basic charge-balance equation that must be satisfied for any volume enclosed by an active membrane is

$$C_m \frac{dV_m}{dt} = \sum_j G_j(t)(E_j - V_m(t)) + I_{\text{ext}}, \quad [1]$$

where C_m is the membrane capacitance, V_m is the membrane voltage, dV_m/dt is the time derivative of V_m, I_{ext} is an externally applied current, $G_j(t)$ is some ionic conductance, E_j is the reversal potential of conductance j, and j runs over the various species of conductances present in the membrane. G_j may vary as a result of dependence on voltage or on the concentration of some chemical species, such as a neurotransmitter, or may represent a constant leak conductance. In principle, any type of neuronal membrane can be modeled with this equation by a suitable selection of relevant G_j functions.

The most basic function of the neuronal membrane that we must be able to simulate is the propagation of an action potential down an axon. Our earliest full understanding of how counterflowing sodium and potassium ions generate the action potential came from the work of Hodgkin and Huxley (2), who proposed a sodium conductance, $G_{\text{Na}} = \bar{G}_{\text{Na}} m^3 h$, a potassium conductance, $G_{\text{K}} = \bar{G}_{\text{K}} n^4$, and a leakage conductance, G_L, where $\bar{G}_{\text{Na}}$, $\bar{G}_{\text{K}}$, and G_L are constant maximum conductances, and h, m, and n are variables specified by equations that capture the voltage dependencies of the Na^+ and K^+ conductances in a semiempirical fashion. To generate accurate models of ion channel function, it may be necessary, for at least two reasons, to allow for varying intracellular and possibly extracellular ion concentrations in the model. First, the reversal potential, E_j, in the treatment above, depends on the ratio of the two concentrations in accord with the Nernst equation:

$$E_j = (RT/z_jF) \ln([X_j]_e/[X_j]_i), \quad [2]$$

where E_j is the reversal potential of the ion X_j ($X = Na^+$, K^+, etc.) associated with conductance j; R is the gas constant; T is the absolute temperature; z is the charge on X_j; F is Faraday's constant; $[X_j]e$ is the extracellular concentration of X_j; and $[X_j]i$ is the intracellular concentration of X_j (3, p. 84). While for many purposes E_j may be taken as constant, because ion concentrations are homeostatically regulated, ions with particularly large variations in concentration, such as calcium, may require a more detailed treatment. Second, the opening of some channels, notably calcium-dependent potassium channels, is controlled by ion concentration, and, obviously, changes in concentration must be modeled if these channels are to be included in the model. Unfortunately, ion concentrations are notoriously difficult to model accurately due to interactions of ions with proteins or other molecules that may buffer their concentrations, poorly known volumes of the spaces to which the ions have diffusional access, and active pumping. In a simple treatment, the ion currents needed to support the electric currents found for channels passing ion X in Eq. [1] can be calculated

and these ion currents used to adjust the intra- and local extracellular concentrations of X at each step of the simulation.

The approach given in Eq. [1] can be applied in an even more detailed manner when it is necessary to deal with the fact that a neuron does not necessarily have the same membrane potential across its entire surface area. In this situation, the cell may be treated as a continuous structure, leading to differential equations for V_m as a function of position as well as time (e.g. (4,5–8)). Alternatively, Eq. [1] can be combined with the passive cable equations developed by Rall (9) to form what is known as a "multicompartment" model (10). The basic idea is to divide the neuron into subdivisions known as compartments. Each compartment has its own V_m and conductance variables, which may be treated as Hodgkin-Huxley channels or by some other formalism. Neighboring compartments are assumed to be connected by cytoplasmic resistances that carry Ohm's law currents proportional to the voltage differences between those compartments. These currents are included as I_{ext} variables in Eq. [1] and the entire system of equations is solved iteratively by computer. The number of compartments depends on the problem to be solved and may range from just two (e.g., dendritic and somatic) to many thousands. In the latter case, compartment geometry may be derived from micrographs of real neurons (10). A typical application is the study of the cerebellar Purkinje cell by De Schutter and Bower (11,12), in which channel parameters were adjusted until a reasonable simulacrum of Purkinje cell function was obtained. However, even with the large number of parameters used in that study, certain aspects of Purkinje cell responses (e.g., the increasing firing rate seen for a time after a stimulating current is turned off) could not be replicated (Figure 1). Thus, the added complexity of multicompartment modeling does not guarantee perfect reproduction of complex neuronal responses, and for many purposes single-compartment models may be considered "good enough."

We now consider in more detail methods that can be applied to single-compartment models or to the individual compartments of multicompartment models. While the Hodgkin-Huxley equations have been extremely influential in showing the way accurately to simulate neuronal membrane function, it must be said that the actual form of the equations for G_{Na} and G_K cannot be justified from first principles. Furthermore, the equations as they stand are incomplete in that, having been derived for the squid giant axon, they do not incorporate the numerous varieties of voltage- and ion-dependent channels found in CNS neurons, and they are computationally burdensome, which renders them impractical for use in large-scale network simulations.

To address the latter limitation, a long line of work has sought simplified versions of the Hodgkin-Huxley treatment that might afford a more rapid, if somewhat less exact, computation of neuronal responses. An early example is the FitzHugh-Nagumo model (13,14), which is analytically tractable and has only two variables. The endpoint of this line of work is the "leaky integrate and

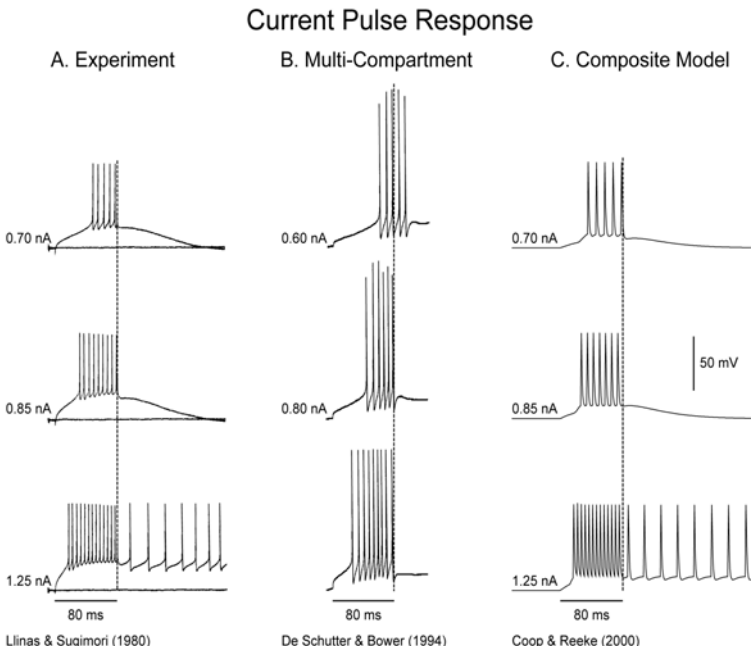


Figure 1. Comparison of multicompartment and composite modeling results with experimental data for a guinea pig cerebellar Purkinje cell stimulated by rectangular current pulses. The dashed vertical lines indicate the offset time of the stimulus pulses. (**A**) Data from a cerebellar slice preparation studied by Llinás and Sugimori (78), reproduced with permission. Note the "plateau" response of the cell with discharge at an increasing rate under 1.25-nA injected current after the stimulus pulse has been turned off (bottom left). (**B**) Data from DeSchutter and Bower (79), reproduced with permission. The spiking response following stimulus offset actually occurs at lower stimulating currents (stimulating current value for bottom panel not given in the original reference). (**C**) Data from a one-compartment version of the composite model of Coop and Reeke (34), reproduced with permission.

fire" (I&F) neuron (15,16), which is easy to simulate but omits most details of cell dynamics. The simplest version of this model eliminates the action potential altogether, replacing it with a simple voltage step on the idea that all action potentials are much alike and already well enough understood. One then has:

$$C\frac{dV}{dt} = -G_L V + I\,, \qquad [3]$$

where the variables are as above, omitting the subscript m, and the voltage scale is chosen such that $V = 0$ is the resting potential. G_L is a leak conductance that allows the voltage to decay to 0 in the absence of input. When the voltage, V,

reaches a threshold, Θ, as a result of integration over the inputs, I, it is reset to a value V_{reset} and integration resumes. These reset events can be passed to other cells in a network model as effective input pulses. This approach fits well with event-based modeling (see below).

An even simpler model, the rate-coded neuron, eschews firing events altogether on the idea that only the average rate of firing over a short time interval or over a small assembly of neurons is relevant for the transmission of information from one cell to another in the nervous system. This idea represents a generalization to the CNS of the early findings of Adrian on sensory neurons (17). The rate-coded model has been very widely used, particularly in the more abstract "neurally inspired" models, such as those described in the classic *Parallel Distributed Processing* books (18,19). The role of rate coding in real nervous systems has become a matter for much current controversy (20–22). Much of the discussion is couched in terms of information theory (23; see also Part II, chapter 1 [by Shalizi], this volume), which provides quantitative measures to help evaluate the various proposals for neural coding schemes (24,25). While the details of these analyses are beyond the scope of the present chapter, it seems safe to answer the question posed at the beginning of this section with the prediction that rate-coded models will inevitably be replaced with firing models that can incorporate effects of proven significance, such as spike synchronization (26,27), after-spike hyperpolarization (28,29), paired-pulse facilitation (30,31), spike-timing dependent plasticity (32,33), and several others, particularly in studies that deal with time-dependent aspects of brain function that go beyond basic pattern recognition. Coop and Reeke (34) describe a composite model in which the action potential and some conductances are computed rapidly by reference to look-up tables, while other conductances are calculated in full. They provide a model of the Purkinje cell using this methodology that can be compared with the full multicompartment model referred to above (Figure 1). Taking another tack, interesting new work shows how the simple I&F model can be enhanced to incorporate slow dynamic effects (35).

1.2. Network Models

To model functional aspects of neuronal circuits, networks of cells communicating via synaptic connections, not just single cells, are required. The lack of detailed knowledge of cell dynamics plus limitations of early computers precluded attempts to model network dynamics in detail, leading to nearly exclusive use of rate-coded cells in the early models. Perhaps because these models were obviously oversimplified from the point of view of the neurophysiologist, the early development of network models occurred mostly in computer science and artificial intelligence (AI) rather than neurophysiology laboratories, and today "neural networks" are in widespread use in communications and data processing applications. This story has been often told (18,19,36–40). However, in recent years network models have gradually evolved towards a greater degree

of biological realism as computers have become more capable and the interests of researchers have turned toward biological understanding, rather than just simulation, of brain functions (41).

A standard for what was to come was established by Traub and Miles with their model of the CA3 network of the hippocampus (42), which featured standardized multicompartment pyramidal and $GABA_A$- and $GABA_B$-type inhibitory cells. Properties of several types of active channels (e.g., Ca^{2+}-dependent K^+ channels) were combined into composite channels to keep the amount of computation to a practical level. Synaptic connections of four basic types took into account axonal delays, a composite activation process, and a first-order relaxation process. This model was very successful in helping to elucidate the roles of the various channel types in complex behaviors such as bursting and network synchronization.

More recently, and perhaps of more relevance to the main subject matter of this book, Tagamets and Horwitz (43) constructed a large-scale model of the visual and forebrain circuitry thought to be involved in delayed match-to-sample tasks in humans. Their goal was to explicate the neural basis for the rCBF signals recorded in PET (positron emission tomography) studies of this and related tasks. Their hypothesis was that the rCBF signal is simply proportional to the sum of the absolute values (whether excitatory or inhibitory) of all the inputs to all the neuronal units in a region of interest. Because this hypothesis did not depend on the complex details of neuronal activation as in the Traub and Miles study, simplified canonical units representing local assemblies of neurons were modeled, representing a total of four areas along the occipitotemporal pathway from lateral geniculate nucleus to prefrontal cortex. Good agreement with experimental rCBF data was obtained, suggesting that synaptic activity indeed accounts for at least a major portion of differential neural metabolic demand, and thus of blood flow, although some influence of glial activity should not be discounted. Arbib et al. (44), working from a similar hypothesis, showed how to calculate simulated functional MRI images from network models, using neural systems for imitative behavior as their exemplar. See also the study of temporal patterns of spontaneous activity in the developing spinal cord in the next chapter of this volume (by Tabak and Rinzel).

1.3. Learning

An essential element of network models is the incorporation of some scheme for adaptive change or "learning." Parallel distributed processing models only really took off with the popularization of the "back-propagation" learning rule (19,45), a form of gradient-descent optimization based on adjusting the strengths of connections between neurons to reduce their contribution to the error measured at the output, as determined by the calculus chain rule for differentiation. However, implementation of this rule requires that nonlocal information be available at each synapse, and this is widely considered to be unavailable

in biological systems, although possible mechanisms for biologically realizable back-propagation have been suggested (46,47). Most biological learning models have instead been based on the proposal of Hebb (48) that "When an axon of cell A ... excite[s] cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells so that A's efficiency as one of the cells firing B is increased" (quoted from (3, p. 1020)). This is often paraphrased as "neurons that fire together wire together," leading to a mathematical formulation such as

$$\Delta c_{ij} = \lceil s_i - \theta_i \rceil \cdot \lceil s_j - \theta_j \rceil, \qquad [4]$$

where Δc_{ij} is the change in the strength, c, of the connection between neurons i and j; $\lceil x \rceil$ represents a negative cutoff function ($\lceil x \rceil = x$ if $x > 0$, otherwise $\lceil x \rceil = 0$); s_i and s_j are the activity levels (here rate-coded) of cells i and j, respectively; and θ_i and θ_j are threshold activity levels for cells i and j, respectively. Although Hebb based his proposal on behavioral data, the rule is now usually related to the phenomenon of long-term potentiation (LTP), which can be observed in numerous neural tissues and cell types (49–51). However, in fact, the Hebb rule as initially stated is unworkable for several reasons.

First, the Hebb rule, naively applied, leads to network instability—it provides for the strengthening, but not the weakening, of connections. (See Part II, chapter 2 [by Socolar], of this volume for methods of analyzing stability, hysteresis, and oscillations in nonlinear dynamical systems.) The stability problem is easily overcome either by normalizing the total strength of all the synapses onto a single cell or network region to a fixed sum, or by replacing the negative cutoff functions in Eq. [4] with any of a variety of specific rules that allow for weakening connections. Such rules are often based on data for long-term depression (LTD), the apparent counterpart to LTP that weakens synapses under certain conditions. The details are complex and depend on cell type and the relative timing of the pre- and postsynaptic signals, but cases have been reported (see references in (52)) in which strong connections that are inactive or only weakly active when the postsynaptic cell is activated are weakened, or, alternatively (53), connections that are active in the absence of postsynaptic activation are weakened. A refinement of these ideas that automatically assures network stability is the Bienenstock-Cooper-Munro (BCM) rule (54), which postulates that Δc_{ij} is zero when s_i is zero, negative for small values of s_i, and positive for values of s_i above a variable crossover threshold (akin to θ_i in Eq. [4]). The threshold is adjusted on a slow time scale in such a way that strengthening is favored when average activity becomes low, while weakening is favored when average activity is high.

More importantly, the Hebb rule does not take into account whether the perception or action produced by the neural circuit where the activity occurs is in fact of use to the organism as a whole. In order for learning to be adaptive, the

animal must not simply strengthen all perceptions and responses that occur for whatever reason, but must apply the Hebb rule or any other learning rule selectively, perhaps regulated by some sort of "value" signal that would play a role similar to that of the error feedback signal in the back-propagation rule, but without its per-connection specificity. This requirement has been clearly set forth by Edelman (55,56) in his theory of neuronal group selection or "neural Darwinism" and given mathematical form and tested in working network models by Reeke et al. (57) and Friston et al. (58). More recently, these ideas have been implemented in synthetic neuronal-network-based control systems for robotic devices performing tasks in a real-world environment (59,60). In its simplest form, value can be implemented by replacing Eq. [4] with a formulation such as

$$\Delta c_{ij} = \left[s_i - \theta_i \right] \cdot \left[s_j - \theta_j \right] \cdot v \,, \qquad [5]$$

where v is the nonspecific value signal, which can be either positive (representing a positive reward for the behavior) or negative (a penalty). Physiologically, the value signal is most likely delivered via widely broadcast modulatory transmitters such as dopamine or acetylcholine, but blood-borne signals (for example, hormones involved in homeostasis) can also be imagined to play a role. (Mapping of reward systems in the brain via MRI imaging is discussed in detail in Part IV, chapter 5 [by Breiter, Gasic, and Makris], this volume). Psychologically, the value signal can be considered not only as responding to intrinsic evaluations of the consequences of behavior, but also as being subject to external manipulation via the administration of rewards and punishments. However, it should be noted that this simple formulation neglects the basic problem known in learning theory as the "credit assignment" problem (61,62), one aspect of which is that by the time the value signal arrives the neural activity levels s_i and s_j may have deviated significantly from the values they had at the time the behavior was produced. For other than trivially short delays in assessing value, some sort of "memory trace" of past activity is required at the cellular level in order for this learning scheme to work. (Part III, chapter 5.3 [by Kolb and Timmann], this volume, contains a discussion of how the classical conditioning of the eyeblink response can be understood in terms of neuronal mechanisms of the kind discussed here operating in the cerebellum.)

1.4. Computational Considerations

In practice, network models are often implemented with *ad hoc* computer codes, especially when new architectural or dynamical principles are being tested. However, where applicable, development is quicker (although execution may be slower) with a general-purpose neuron- or neuronal network-simulation tool that can be programmed with some sort of *problem-description language* to specify the elements needed for a particular simulation. A number of such pro-

grams have been described (63–67). These programs generally provide a network definition phase, in which the user specifies the numbers and types of cells required and the rules for generating connections between them. These connections can be entirely prespecified, perhaps by input matrices, or can be generated at random according to specified statistical rules. The program builds appropriate data structures in memory according to these specifications. Also provided are suitable constructs to define the dynamical properties of the cells in the network. The details of this process vary according to the type of cell being modeled (rate-coded, integrate-and-fire, multicompartment, etc.) and the method of treating synaptic inputs. Full-featured general-purpose simulators will also provide methods for stimulating the network and for recording cell responses and relevant statistics such as mean firing rates.

Greater operating speed, at the cost of significantly longer development time and the need for specialized engineering knowledge, can be obtained by the construction of special purpose simulation hardware, otherwise known as "neuromorphic" devices (see Part IV, chapter 6 [by Northmore, Moses, and Elias], this volume).

The detailed method of updating in network models needs to be considered carefully. The time step must be small enough to provide reasonable accuracy, but as large as possible to minimize computing time. More complex integration schemes, such as Runge-Kutta methods, may allow the time step to be increased relative to straightforward Euler integration but at the cost of greater computational complexity (68). More aggressively, the step size can be adapted to cellular activity, small when rapid changes in membrane potential are occurring, large at other times (69). Taken to an extreme, this idea leads to event-driven modeling (70), in which the equations for V_m are solved analytically in the absence of synaptic input. Detailed simulation is then only necessary when input disturbs the analytical solution. However, in network models, events of interest occur at different times in different cells, making the bookkeeping for such models very difficult.

An additional consideration is the method of updating cell activity levels as seen by other cells. If cell firings are propagated to postsynaptic cells as soon as they are computed, then cells earlier in the update cycle will have an artificial advantage in competitive networks. On the other hand, if cell firings are propagated simultaneously at the end of each update cycle, then artifactual oscillations may be observed in the overall network. These problems can be mitigated by randomly changing the update order in each cycle, but then special care must be taken to allow simulation runs to be replicated, and replication will be particularly difficult or lead to unacceptable serialization in a parallel computing environment.

In summary, a great variety of simulation techniques and software packages for neuronal and neuronal network simulation is available. When contemplating a new modeling project, the prospective modeler should carefully consider the

motivation for the model, the degree of realism that is desired, the amount of detailed physiological and behavioral information that is available for the system under study, whether simulated sensory stimuli can adequately capture the real experimental situation, and the extent of the computing resources and programming skills that are available.

2. FUTURE WORK AND RELEVANCE TO BIOMEDICINE

Most of the work to date in functional neuronal modeling, as summarized above, has totally ignored the geometry of the neural network. This is adequate when the timing of axonal conduction can be neglected, when the supply of metabolic substrates (glucose, oxygen) to neurons can be taken as always adequate, and when development of the adult, functioning system is not at issue. However, questions involving the functional consequences of the embodiment of the nervous system in an organism (some would say the necessity of embodiment to the ontogeny of function) are increasingly of interest and will require the development of a new integrative style of modeling that includes relevant aspects of the geometry of the nervous system and geometric constraints on network connectivity and complexity, as well as constraints on metabolic resources. To take into account more of the relevant mechanical and physicochemical factors in integrated models will require new combinations of tools. In particular, to take locally discontinuous obstacles, forces, and signals into effect, it will be necessary to go beyond differential equations and work with networks of interacting cells. Models of this kind will require, at a minimum, more detailed representations of cell geometry, diffusion equations to handle small-molecule signaling, something like cellular automata rules[1] to model contact signaling, and, especially, finite-element modeling to deal with the operation of mechanical forces across tissue volumes.

2.1. Modeling the Geometry of Nervous System Structures

2.1.1. Computer Representation of Neural Tissue

Reasonably accurate computer representations of brain tissue structure at the cellular level will be essential if growth models are to be integrated with functional models. A key problem is that space must be filled, yet growth and cell division must be accommodated. Local growth has long-range effects because existing nearby structures must be pushed aside, and these in turn push aside more distant structures, and so on until a hard boundary (the cranium), or the edge of the model volume is reached. Space can be filled with hexahedral bricks, but these introduce spurious anisotropies (71). In addition, vasculature

and neurites may be more faithfully represented by branched cylindrical structures. This problem has not yet been solved, but two approaches seem attractive for detailed exploration. First, one can use cylinders for those structures that are cylindrical, impose lumped cohesive forces where the cylinders are in contact to represent areas of cell–cell contact (cylinders of course cannot make areal contacts), then treat the unfilled space as a tortuous fluid matrix where signal molecules can diffuse. Second, one can subdivide the space with Voronoi polyhedra[2] (71), placing the mesh points such that the bounding planes of the polyhedra approximate the surfaces of the cells and capillaries. The former approach has the advantage that it may not be necessary to adjust the geometry of the entire model every time a local movement or mitosis event occurs, since some of the expansion can be accommodated by reduction of the local fluid volume. Also, it provides a natural basis for modeling small- and macromolecular diffusion. On the other hand, mechanical interactions may not be captured faithfully. If finite-element modeling (FEM) is being attempted, the relationship between cellular structures and model elements must be carefully considered. An obvious initial approach is to make the two kinds of boundaries coincide. The Voronoi tessellation approach would appear to be most compatible with the construction of suitable FEM meshes and their periodic adjustment to accommodate growth and the resultant displacement of nearby tissue (but see below).

2.1.2. Geometry from Images

Ideally, one would like to obtain model geometry from imaging studies on real brains, normal or diseased. At present, the only candidates accepted in clinical practice with anywhere near the desired resolution are CT and MRI scanning. As is generally known, CT scanning excels at revealing details of structures delineated by their density, light (ventricles) or heavy (bone), while MRI makes finer distinctions among soft tissues based on their proton composition. Both will have their place, but neither at present has sufficient resolution to support detailed cell-level modeling. Accordingly, in the near term it will be necessary to compromise, using image-derived geometry to delineate tissue boundaries, but generic, randomly generated, cell and vascular geometry on the finest scale.

At a minimum, the tissue types that need to be identified in images include bone, ventricles, blood vessels, and neural white and gray matter. While, as suggested above, detailed individual cell geometry and type identification (neuron, glia, etc.) cannot be obtained from the images now available, one would like at least to obtain some indication of directions of prominent axonal fasciculation, because of its relevance to inter-areal connectivity and possibly to new neurite growth. Automated image segmentation at the level required here is itself an outstanding research problem, although some methods have been published (e.g.

(72)). In the near term, modelers will have to be prepared to segment images manually, possibly with the aid of interactive software, if for no other reason than to validate results obtained with more automated methods.

2.2. Diffusion of Signal Molecules

The equations governing diffusion are well known (73, pp. 700–718). In practice, distinction must be made between certain small molecules, such as nitric oxide (NO), which can diffuse freely through cell membrane barriers (74), and larger molecules, which cannot. To deal with the latter, one must in principle solve the diffusion equations in the highly irregular extracellular volume, where diffusion is significantly hindered for all but the smallest molecules. This can be handled by generalizing the numerical technique used by Gally et al., dividing the tissue into small volume elements and applying mass balance across all the boundaries between the extracellular compartments. In the case of polyhedral meshing, these would be the irregular polygons delineating the compartment boundaries. Fortunately, assuming the individual compartments are small enough for an assumption of spatially constant concentration to apply, only the area, not the detailed shape, of each boundary is required. This information can be computed as part of the meshing process and stored in tables for use during the simulation.

2.3. Contact Signaling

Computationally, contact signaling is a relatively minor problem once the model geometry has been specified. A list of cell–cell contact areas can be maintained and the generation of signal molecules, their movement across each such contact, and their effects on target cells updated at each iteration of the simulation. The problem here is in defining what signals are important for the behavior under study, where and under what conditions those signals are produced, what concentrations are effective in modifying the behavior of target cells, and so on. In general, all that can be said is that detailed rules for signaling are likely to be among the hypothesis-driven model specifications that will need to be varied in order to optimize the faithfulness with which the model mirrors the behavior of real tissue.

2.4. Finite-Element Modeling

The techniques discussed so far address how to model electrical and chemical interactions of cells. This leaves mechanical interactions to be dealt with.

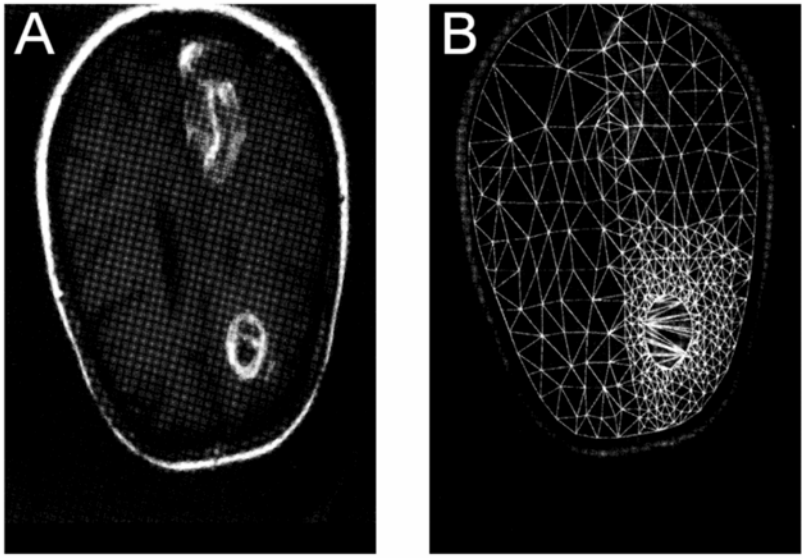


Figure 2. (**A**) Initial MRI scan of a patient with a brain tumor. (**B**) Two-dimensional finite-element mesh constructed from the image in (A). Data from Wasserman et al. (76), reproduced with permission.

While simplified models based on displacement of rigid or incompressible tissue elements might be considered (75), FEM provides the obvious approach to dealing in a more principled way with the inelastic distortions and possibly compression that might be expected to occur during normal or cancerous growth in the nervous system. An early example of a two-dimensional FEM tumor model is shown in Figure 2 (76). FEM can be applied at any desired spatial scale from some upper limit where the mesh elements would grossly violate the assumption of internal homogeneity right down to the subcellular level. Thus, development of software for this application, if properly designed, is an activity that might not have to be iterated as each new generation of models at successively smaller scales comes online.

As with identification of tissue types from images, automated mesh generation is a significant requirement for FEM, if for no other reason than that meshing is entirely too tedious to perform manually in a three-dimensional structure of any useful size. As is the case with tissue type segmentation, meshing routines are available, even in the commercial FEM packages, but software will need to be developed to couple the results of tissue segmentation into the selection of mesh nodes and the assignment of material properties to the mesh ele-

ments. Furthermore, it should be noted that FEM programs work best with a small number of geometrical element types, making them at first glance incompatible with the Voronoi tessellation scheme suggested earlier. This problem can be resolved either by giving up the advantages of the Voronoi approach and reverting to regular mesh elements, or else by first making the Voronoi tessellation and then dividing each of the Voronoi polyhedra into smaller tetrahedra (triangles in a two-dimensional model), an operation that is always guaranteed to be possible. However, tetrahedra are intrinsically rigid unless special tricks are applied and may not make the best elements for FEM of soft tissue.

A further necessity with FEM that needs to be kept in mind is that the calculations lose accuracy as the initial mesh elements become more and more distorted during the course of the simulation. The software must therefore be designed to suspend the calculations at certain intervals and remesh the model volume to provide a new starting point with undistorted mesh elements.

An attractive feature of FEM is that the mesh elements do not need to be of comparable size. Therefore, the level of detail can be varied from place to place within a single model. This suggests that the transition from tissue-level to cell-level modeling might initially be made by surrounding a small volume of detailed cell-level mesh with a larger volume modeled on a cruder mesh scale. The surround could provide a spatially nonuniform external pressure resisting growth, a source for nutrients and oxygen, and a sink for waste products, while the fine-meshed interior would allow one to explore the processes of interest in greater detail. Initially, even this fine-meshed volume might not comprise individual neurons, but eventually, one would want it to do so in order to permit the study of the effects of pressure and distortion on the functioning of a working neural network model. If the FEM neurons were modeled as cylindrical mesh elements, these could be made to correspond with compartments in a multicompartment electrophysiological neural function model (77). One could then study, for example, the effects on function of changes in membrane capacitance, axonal resistivity, ionic equilibria, or channel dynamics caused by the pressure resulting from normal development or from tumor growth. Similarly, neurite bending or shear caused by the growth of a nearby tumor might lead to the choking off of conductance along a neural pathway with functional implications.

3. CONCLUSIONS

The basic physiology of neuronal activation and discharge is reasonably well understood, and a vast literature of modeling studies, using approaches ranging from very abstract to very realistic, is available. While it is true that not all aspects of single neuron function, not to mention network function, can yet be routinely simulated, and certainly not with a single set of parameters applicable to even one type of cell in all experimental situations, nonetheless, there is

not much doubt that further progress can be expected as time passes and computers continue to become more capable. In a less detailed way, modeling of tissue development in general, and that of neural tissue specifically, is a field with a long history of informative studies at the bulk material or cellular automaton level, with continual progress being reported, though often restricted to two dimensions.

It has recently become possible to consider constructing models in which neuronal network development and behavior-generating function are combined. Farther off are models of this kind in which the basic components are individual cells. In such models, it will for the first time be possible to study the combined influences of normal cellular mitotic controls, diffusion of nutrients, angiogenesis, mechanical obstacles to growth, cell–cell signaling by direct contact and by diffusion of signaling molecules, as well as various experimental and therapeutic interventions. Limitations in the acquisition of detailed geometric information from brain images and, at a more practical level, in computer power, will prevent these models in the near future from reaching the size scale of macroscopic brain features. Thus, work will have to continue in parallel with models based on generic detailed tissue properties but more realistic large-scale boundary conditions. The eventual contact of work at the cell and whole-brain or brain-region spatial scales will lead to a new class of models that should greatly increase our understanding of how molecular and mechanical influences interact in normal and cancerous brain development. These models will give us a new understanding of the ecological place of the brain in the whole organism, as well as making possible new approaches to the rational planning of therapies for brain malfunctions.

4. NOTES

1. Cellular automata are models in which abstract units ("cells") arranged on a lattice may exist in one of a finite set of defined states and which undergo changes in state at each time step of the model according to rules that depend only on their own states and the states of a defined set of near neighbors at the previous time step.

2. Polyhedra constructed by erecting planes perpendicular to and bisecting lines between centers of interest. Each plane extends in all directions only to the nearest line of intersection with another such plane.

5. REFERENCES

1. Reeke Jr GN, 1992. Neural net and neuronal nets: how much like the nervous system should a model be? In *Neuropsychology: the neuronal basis of cognitive function*, pp. 15–27. Ed. E Costa, KW Brocklehurst. Thieme Medical, New York.

2. Hodgkin AL, Huxley AF. 1952. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiol Lond* **117**:500–544.
3. Kandel ER, Schwartz JH, Jessel TM. 1991. *Principles of neural science*, 3rd ed. Elsevier, New York.
4. Knight BW, Toyoda JI, Dodge Jr, FA. 1970. A quantitative description of the dynamics of excitation and inhibition in the eye of *Limulus*. *J Gen Physiol* **56**:421–437.
5. Poznanski, RR, Bell J. 2000. Theoretical analysis of the amplification of synaptic potentials by small clusters of persistent sodium channels in dendrites. *Math Biosci* **166**:123–147.
6. Poznanski RR, Bell J. 2000. A dendritic cable model for the amplification of synaptic potentials by an ensemble average of persistent sodium channels. *Math Biosci* **166**:101–121.
7. Poznanski RR. 1988. Membrane voltage changes in passive dendritic trees: a tapering equivalent cylinder model. *IMA J Math Appl Med Biol* **5**:113–145.
8. Reeke GN, Poznanski RR, Lindsay K, Rosenberg J, Sporns O, eds. 2005. *Modeling in the neurosciences*, 2nd ed. CRC Press, Boca Raton, FL.
9. Rall W, Agmon-Snir H. 1998. Cable theory for dendritic neurons. In *Methods in neuronal modeling: from ions to networks*, 2nd ed., pp. 27–92. Ed CI Koch, I Segev. MIT Press, Cambridge.
10. Segev I, Burke R. 1998. Compartmental models of complex neurons. In *Methods in neuronal modeling: from ions to networks*, 2nd ed., pp. 93–136. MIT Press, Cambridge.
11. De Schutter E, Bower JM. 1994. An active membrane model of the cerebellar Purkinje cell, I: simulation of current clamps in slice. *J Neurophysiol* **70**:375–400.
12. De Schutter E, Bower JM. 1994. An active membrane model of the cerebellar Purkinje cell, II: simulation of synaptic responses. *J Neurophysiol* **70**:401–419.
13. FitzHugh R. 1961. Impulses and physiological states in models of nerve membrane. *Biophys J* **1**:445–466.
14. Nagumo JS, Arimato S, Yoshizawa S. 1962. An active pulse transmission line simulating a nerve axon. *Proc IRE* **50**:2061–2070.
15. Ermentrout GB, Chow CC. 2002. Modeling neural oscillations. *Physiol Behav* **77**:629–633.
16. Reich DS, Victor JD, Knight BW, Ozaki T,, Kaplan E. 1997. Response variability and timing precision of neuronal spike trains in vivo. *J Neurophysiol* **77**:2836–2841.
17. Adrian ED. 1926. The impulses produced by sensory nerve endings. *J Physiol (Lond)* **61**:49–72.
18. McClelland JL, Rumelhart DE, PDP Research Group. 1986. *Parallel distributed processing: explorations in the microstructure of cognition*, Vol. 2: *Psychological and biological models*. MIT Press, Cambridge.
19. Rumelhart DE, McClelland JL, PDP Research Group. 1986. *Parallel distributed processing: explorations in the microstructure of cognition*, Vol. 1: *Foundations*. MIT Press, Cambridge.
20. Mainen ZF, Sejnowski TJ. 1995. Reliability of spike timing in neocortical neurons. *Science* **268**:1503–1506.
21. Shadlen MN, Newsome WT. 1998. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J Neurosci* **18**:3870–3896.
22. Maass W, Bishop CM, eds. 1999. *Pulsed neural networks*. MIT Press, Cambridge.
23. Shannon CE. 1948. A mathematical theory of communication. *Bell Syst Tech J* **27**:379–423, 623–656.
24. Rieke F, Warland D, de Ruyter van Steveninck R, Bialek W. 1997. *Spikes: exploring the neural code*. MIT Press, Cambridge.
25. Reeke GN, Coop AD. 2004. Estimating the temporal interval entropy of neuronal discharge. *Neural Comput* **16**:941–970.
26. Grammont F, Riehle A. 2003. Spike synchronization and firing rate in a population of motor cortical neurons in relation to movement direction and reaction time. *Biol Cybern* **88**:360–373.
27. Svirskis G, Hounsgaard J. 2003. Influence of membrane properties on spike synchronization in neurons: theory and experiments. *Netw-Comput Neural Syst* **14**:747–763.

28. Baldissera F, Gustafsson B. 1971. Regulation of repetitive firing in motoneurones by the after-hyperpolarization conductance. *Brain Res* **30**:431–434.
29. Williams SR, Christensen SR, Stuart GJ, Häusser M. 2002. Membrane potential bistability is controlled by the hyperpolarization-activated current I(H) in rat cerebellar Purkinje neurons in vitro. *J Physiol* **539**:469–483.
30. Racine RJ, Milgram NW. 1983. Short-term potentiation phenomena in the rat limbic forebrain. *Brain Res* **260**:201–216.
31. Sokolov MV, Rossokhin AV, Behnisch T, Reymann KG, Voronin LL. 1998. Interaction between paired-pulse facilitation and long-term potentiation of minimal excitatory postsynaptic potentials in rat hippocampal slices: a patch-clamp study. *Neuroscience* **85**:1–13.
32. Bi GQ, Poo MM. 1998. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J Neurosci* **18**:10464–10472.
33. Song S, Miller KD, Abbott LF. 2000. Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neurosci* **3**:919–926.
34. Coop AD, Reeke Jr GN. 2001. The composite neuron: a realistic one-compartment Purkinje cell model suitable for large-scale neuronal network simulations. *J Comput Neurosci* **10**:173–186.
35. Izhikevich EM. 2001. Resonate-and-fire neurons. *Neural Netw* **14**:883–894.
36. Anderson JA, Rosenfeld E, eds. 1988. *Neurocomputing: foundations of research*. MIT Press, Cambridge.
37. Rosenblatt F. 1958. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol Rev* **65**:386–408.
38. Widrow B. 1964. Pattern recognition and adaptive control. *Appl Ind*, pp. 269–277.
39. Minsky ML, Papert SA. 1988. *Perceptrons*, 2nd ed. MIT Press, Cambridge.
40. Churchland PS, Sejnowski TJ. 1992. *The computational brain*. MIT Press, Cambridge.
41. Dayan P, Abbott LF. 2001. *Theoretical neuroscience: computational and mathematical modeling of neural systems*. MIT Press, Cambridge.
42. Traub RD, Miles R. 1991. *Neural networks of the hippocampus*. Cambridge UP, Cambridge.
43. Tagamets MA, Horwitz B. 1998. Integrating electrophysiological and anatomical experimental data to create a large-scale model that simulates a delayed match-to-sample human brain imaging study. *Cerebral Cortex* **8**:310–320.
44. Arbib MA, Billard A, Iacoboni M,, Oztop E. 2000. Synthetic brain imaging: grasping, mirror neurons and imitation. *Neural Netw* **13**:975–997.
45. Werbos P. 1974. Beyond regression: new tools for prediction and analysis in the behavioral sciences. Ph.D. dissertation. Harvard University, Cambridge.
46. Zipser D. 1986. Biologically plausible models of place recognition and goal location. In *Parallel distributed processing*, II: *psychological and biological models*, Vol. 2, pp. 432–470. Ed. JL McClelland, DE Rumelhart, PDP Research Group. MIT Press, Cambridge.
47. Zipser D. 1990. Modeling cortical computation with backpropagation. In *Neuroscience and connectionist theory*, pp. 355–383. Ed. MA Gluck, DE Rumelhart. Lawrence Erlbaum Associates, Hillsdale, NJ.
48. Hebb DO. 1949. *The organization of behavior: a neuropsychological theory*. Wiley, New York.
49. Bliss TV, Lomo T. 1973. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J Physiol* **232**:331–356.
50. Lynch MA. 2004. Long-term potentiation and memory. *Physiol Rev* **84**:87–136.
51. Abraham WC, Williams JM. 2003. Properties and mechanisms of LTP maintenance. *Neuroscientist* **9**:463–474.
52. Christie BR, Kerr DS, Abraham WC. 1994. Flip side of synaptic plasticity: long-term depression mechanisms in the hippocampus. *Hippocampus* **4**:127–135.

53. Artola A, Bröcher S,, Singer W. 1990. Different voltage-dependent thresholds for inducing long-term depression and long-term potentiation in slices of rat visual cortex [comment]. *Nature* **347**:69–72.
54. Bienenstock EL, Cooper LN, Munro PW. 1982. Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *J Neurosci* **2**:32–48.
55. Edelman GM. 1987. *Neural Darwinism: the theory of neuronal group selection*. Basic Books, New York.
56. Edelman GM. 1989. *The remembered present: a biological theory of consciousness*. Basic Books, New York.
57. Reeke Jr GN, Sporns O, Edelman GM. 1990. Synthetic neural modeling: the "Darwin" series of automata. *Proc IEEE* **78**:1498–1530.
58. Friston KJ, Tononi G, Reeke Jr GN, Sporns O,, Edelman GM. 1994. Value-dependent selection in the brain: simulation in a synthetic neural model. *Neuroscience* **59**:229–243.
59. Edelman GM, Reeke Jr GN, Gall WE, Tononi G, Williams D, Sporns O. 1992. Synthetic neural modeling applied to a real-world artifact. *Proc Natl Acad Sci USA* **89**:7267–7271.
60. Krichmar JL, Edelman GM. 2002. Machine psychology: autonomous behavior, perceptual categorization and conditioning in a brain-based device. *Cerebral Cortex* **12**:818–830.
61. Minsky ML. 1961. Steps toward artificial intelligence. *Proc Inst Radio Engineers* **49**:8–30.
62. Sutton RS, Barto AG. 1998. *Reinforcement learning*. MIT Press, Cambridge.
63. Sakai K, Sajda P, Yen SC, Finkel LH. 1997. Coarse-grain parallel computing for very large scale neural simulations in the NEXUS simulation environment. *Comput Biol Med* **27**:257–266.
64. Bower JM, Beeman J. 1993. *The book of Genesis*. Springer Berlin.
65. Goddard NH, Hucka M, Howell F, Cornelis H, Shankar K, Beeman D. 2001. Towards NeuroML: model description methods for collaborative modelling in neuroscience. *Phil Trans Roy Soc London B Biol Sci* **356**:1209–1228.
66. Hines ML, Carnevale NT. 2001. NEURON: a tool for neuroscientists. *Neuroscientist* **7**:123–135.
67. Reeke Jr GN, Edelman GM. 1987. Selective neural networks and their implications for recognition automata. *Int J Supercomput Appl* **1**:44–69.
68. Mascagni MV, Sherman AS. 1998. Numerical methods in neuronal modeling. In *Methods in neuronal modeling: from ions to networks*, 2nd ed., pp. 569–606. Ed. S Koch, I Segev. MIT Press, Cambridge.
69. Borg-Graham LJ. 2000. Additional efficient computation of branched nerve equations: adaptive time step and ideal voltage clamp. *J Comput Neurosci* **8**:209–226.
70. Manor Y, Gonczarowski J, Segev I. 1991. Propagation of action potentials along complex axonal trees: model and implementation. *Biophys J* **60**:1411–1423.
71. Kansal AR, Torquato S, Harsh GRI, Chiocca EA, Deisboeck TS. 2000. Simulated brain tumor growth dynamics using a three-dimensional cellular automaton. *J Theor Biol* **203**:367–382.
72. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM. 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**:341–355.
73. Eisenberg D, Crothers D. 1979. *Physical chemistry with applications to the life sciences*. Benjamin/Cummings, Menlo Park, CA.
74. Gally JA, Montague PR, Reeke Jr GN, Edelman GM. 1990. The NO hypothesis: possible effects of a short-lived, rapidly diffusible signal in the development and function of the nervous system. *Proc Natl Acad Sci USA* **87**:3547–3551.
75. Stamatakos GS, Zacharaki EI, Makropoulou MI, Mouravliansky NA, Marsh A, Nikita KS, Uzunoglu NK. 2001. Modeling tumor growth and irradiation response in vitro—a combination of high-performance computing and web-based technologies including VRML visualization. *IEEE Trans Inf Technol Biomed* **5**:279–289.

76. Wasserman R, Acharya R, Sibata C, Shin KH. 1996. A patient-specific in vivo tumor model. *Math Biosci* **136**:111–140.
77. Koch C, Segev I, eds. 1998. *Methods in neuronal modeling: from ions to networks*, 2nd ed. MIT Press, Cambridge.
78. Llinás R, Sugimori M. 1980. Electrophysiological properties of in vitro Purkinje cell somata in mammalian cerebellar slices. *J Physiol (London)* **305**:171–195.
79. De Schutter E, Bower JM. 1994. An active membrane model of the cerebellar Purkinje cell, II: simulation of synaptic responses. *J Neurophysiol* **71**:401–419.

5.2

MODELING SPONTANEOUS EPISODIC ACTIVITY IN DEVELOPING NEURONAL NETWORKS

Joël Tabak

Laboratory of Neural Control, Section on Developmental Neurobiology, NINDS, National Institutes of Health, Bethesda, Maryland

John Rinzel

Courant Institute of Mathematical Science and Center for Neural Science, New York University, New York

Neuronal networks are extraordinarily complex systems, structurally and dynamically, given the number of elements that compose them, their functional architecture, their plasticity, and their nonlinear mechanisms for signaling over vast ranges of time scales. One approach to understanding how neuronal circuits generate activity is to study developing networks that are relatively simpler, before any experienced-based specialization has occurred. Here, we present a model for the generation of spontaneous, episodic activity by developing spinal cord networks. This model only represents the averaged activity and excitability in the network, assumed purely excitatory. In the model, positive feedback through excitatory connections generates episodes of activity, which are terminated by a slow, activity-dependent depression of network activity (slow negative feedback). This idealized model allowed a qualitative understanding of the network dynamics, which leads to prediction/comprehension of experimental observations. Although the complexity of the system has been restricted to interactions between fast positive and slow negative feedback, the emergent feature of the network rhythm was captured, and it applies to many developing/excitatory networks. An open question is whether this mechanism can help us explain the activity of more complex/mature networks including inhibitory connections.

Address correspondence to: Joël Tabak, Department of Biological Science, BRF/206, Florida State University, Tallahassee, Florida 32306 (joel@neuro.fsu.edu).

1. INTRODUCTION

A key problem in neuroscience is to understand the cellular basis of behavior. Vertebrate neural networks have very diverse architectures, and are composed of large numbers of neurons of different types, connected through various classes of synaptic contacts. In order to discover the principles underlying the activity of neural networks, it may be helpful to understand their history. Therefore, a possible strategy is to study developing networks, as they might be simpler to study but nevertheless operate through similar mechanisms as the more complex, mature networks. One major advantage of this strategy is that many, if not all, developing neural networks generate a similar type of activity that has been called "spontaneous activity." Therefore, the conclusions obtained from the study of one network may be of general application.

In this chapter, we review the characteristics of this spontaneous activity, focusing on the networks of the developing spinal cord. We then present an idealized model of this activity and discuss two important applications of this model. Finally, we discuss the generality of this model and its possible application to more complex networks.

2. SPONTANEOUS ACTIVITY IN DEVELOPING NETWORKS

Early in development, neuronal networks of the central nervous system generate spontaneous activity. It is called spontaneous because it is not provoked by sensory inputs or inputs from other parts of the nervous system, but is generated within each of these circuits in isolation. Spontaneous activity has been well characterized in the developing spinal cord, hippocampus, and retina and has also been described in other circuits (25). Although they have widely different architectures, the features of the spontaneous activity are very similar in all these networks (25). The most characteristic feature of this activity is its *episodic* nature: most if not all neurons of the network become active together for several seconds to a minute; then the network becomes silent for intervals that can last up to several minutes, as illustrated in Figure 1.

Because spontaneous activity is so widespread in the developing nervous system and with striking similarities between different circuits, understanding its mechanisms of generation may provide some general principles of neuronal network function. Furthermore, there is evidence that spontaneous activity can drive the refinement of neuronal circuits (17,46), as this activity usually involves large populations of neurons in a highly correlated fashion (see below) and therefore may lead to strengthening/weakening of synaptic connections through Hebbian mechanisms (28) (see (46) for a discussion of spike-timing-dependent potentiation/depression of synapses related to developing circuits). Finally, the temporal pattern of activity may also regulate the electrical properties of indi-

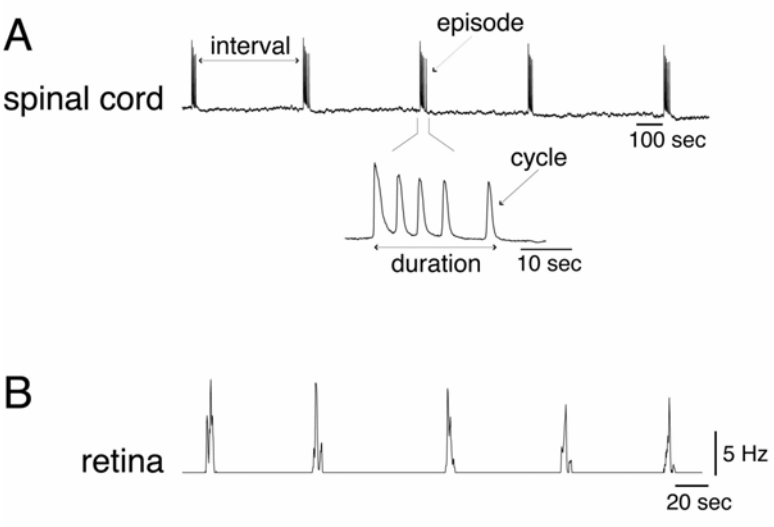


Figure 1. Episodic activity in developing networks. A, Spontaneous activity recorded from the isolated spinal cord of a 7.5-day-old chick embryo. The activity is recorded from a ventral root and represents the synchronous activation of a population of motoneurons. It is characterized by rhythmic episodes lasting up to a minute that are separated by silent intervals lasting up to 20 minutes. The high-frequency (fast) signal corresponding to motoneuron discharge is not visible because of the scale and the low sampling rate (20 Hz). Modified with permission from Tabak et al. (36). (**B**) Spontaneous activity recorded from the isolated retina of a 9-day-old mouse (postnatal). This signal is the rate of neuronal discharge averaged over 29 cells. Note the difference in the time scale with the recording in A. Data reprinted from J. Demas, S.J. Eglen and R.O.L. Wong (unpublished).

vidual neurons (15,22). It is therefore important to understand the mechanisms of generation of spontaneous activity in the developing nervous system. In the following, we focus on the temporal organization of activity through activity-dependent synaptic depression in the developing spinal cord, but suggest that the same features are common to other neural networks.

Spontaneous activity was first observed as spontaneous movements in embryos of diverse animals. Embryonic motility was extensively studied in the chicken embryo (16), as it was easy to observe spontaneous movements through the egg shell, and later to record electrical activity through a small window in the shell. It was shown that these episodic movements were caused by spontaneous electrical activity in the neuronal networks of the spinal cord. More recently, an isolated in vitro preparation of the embryonic chick spinal cord was developed (26), allowing one to record the activity and to manipulate the network (by lesions or pharmacology) at different stages of development.

We can record this activity from a population of motoneurons (the output neurons projecting to the muscles) by suctioning a ventral root (axons from the motoneurons) into an electrode. The activity recorded this way is a combination of two signals. The slow component, illustrated in Figure 1A, represents the depolarization (increase of membrane potential) of the motoneurons, propagating passively along the axons. Superimposed on this slow signal is a fast signal (not visible in the figure) caused by the action potentials generated by the motoneurons. The slow signal shown in Figure 1A is a good indicator of the activity in the whole network, since motoneuron depolarization is caused by synaptic inputs from other neurons in the network. The activity is episodic, with episodes lasting up to a minute separated by intervals of up to 20 minutes. Within an episode, the activity is rhythmic with a cycle frequency of ~0.2–1.0 Hz, markedly decreasing toward the end of the episode. Each cycle can be seen as a network "spike," representing the depolarization of the whole neuronal population. Although the neurons are activated in synchrony, their action potentials are not synchronized.

How is this activity generated? Several key experimental findings provide a working hypothesis.

The first thing to note is that many developing networks can be considered as purely excitatory circuits. This is because the inhibitory neurotransmitters GABA and glycine have excitatory effects early in development (2,6).[1] Indeed, blocking the action of excitatory neurotransmitters acetylcholine and glutamate does not prevent the spinal cord from generating episodic activity (7). It is therefore easy to understand the presence of spontaneous activity in immature circuits. Any event such as a few neurons randomly firing can be amplified by positive feedback through the recurrent excitatory projections, leading to massive activity in the whole network. This explains the activity, but not its episodic pattern. How are the episodes terminated, in the absence of inhibitory connections?

One possibility is that the network "fatigues" during activity, until it is no longer capable to sustain activity. In other words, there is an *activity-dependent* process that depresses the excitability of the network. To demonstrate the presence of such activity-dependent depression process, we have stimulated populations of interneurons through sensory or propriospinal afferents and observed the evoked synaptic response on the motoneurons (11). The synaptic potentials recorded on the motoneurons are decreased after an episode of activity, and progressively increase during the interval between episodes. This suggests that network excitability is depressed by the episodes of activity and that it recovers in the interval between episodes. This depression may be synaptic, that is, activity decreases synaptic efficacy. An activity-dependent synaptic depression process, by reducing the strength of the connections between neurons, would effectively decrease the positive feedback. Another possibility is that this depression acts on the neurons to decrease their excitability, making them less

likely to fire action potentials. Such activity-dependent cellular adaptation process would shift the input/output function of the neurons and make them less responsive to inputs.

Finally, it must be stressed that no particular network structure or cellular properties are thought to underlie the pattern of activity. As mentioned above, spontaneous episodic activity is observed in developing networks with a great variety of structure, and a similar type of activity is observed in dissociated cultures of spinal neurons for which inhibitory transmission has been blocked (35). In the chick embryo, lesion studies have shown that ventral networks of the spinal cord can generate the activity despite transversal or horizontal sectioning (47), and pharmacological studies have shown that episodic activity is still generated when individual neurotransmitters are blocked (7). In addition, there is no evidence that cellular pacemaker properties underlie the rhythmic activity. It is through their (excitatory) network interactions that the rhythm arises, i.e., the rhythmic activity is an emergent property of the network and the dynamics of the recurrent connections.

The main hypothesis is therefore that the spontaneous, episodic activity is generated by a purely excitatory network. This activity depresses network excitability and when excitability is too low the activity stops. In the silent period, network excitability can recover until a new episode starts. In order to test this hypothesis, we have built a very idealized model (schematized in Figure 2A) based on all these observations. This differential-equations model can be analyzed qualitatively using phase-plane and dynamical-systems concepts, so we can understand its dynamical behavior. It allowed us to explain some experimental results and made some predictions, some of which are presented below.

3. MODEL OF SPONTANEOUS ACTIVITY IN THE EMBRYONIC CHICK SPINAL CORD

According to the experimental findings presented above, we model a purely excitatory network whose detailed structure (connectivity pattern, heterogeneity of cell types) is not known but does not seem to be important. We also assume that the membrane properties of the neurons are not important either and all the neurons are active or inactive together. We therefore use a "mean-field" representation of the activity and depression variables, as used by Wilson and Cowan (44). According to this formulation, the activity $a(t)$ of the network is an average of the neuronal firing rate over the population of neurons (see Figure 2A,B). Individual spikes are not modeled (and assumed not to occur synchronously); this firing rate is a temporally coarse-grained representation, that is, averaged over a short period of time.[2] Thus, a can be related to the (pre)synaptic drive, i.e., the amount of synaptic input exciting neurons in the network (29). The model consists of three equations (36):

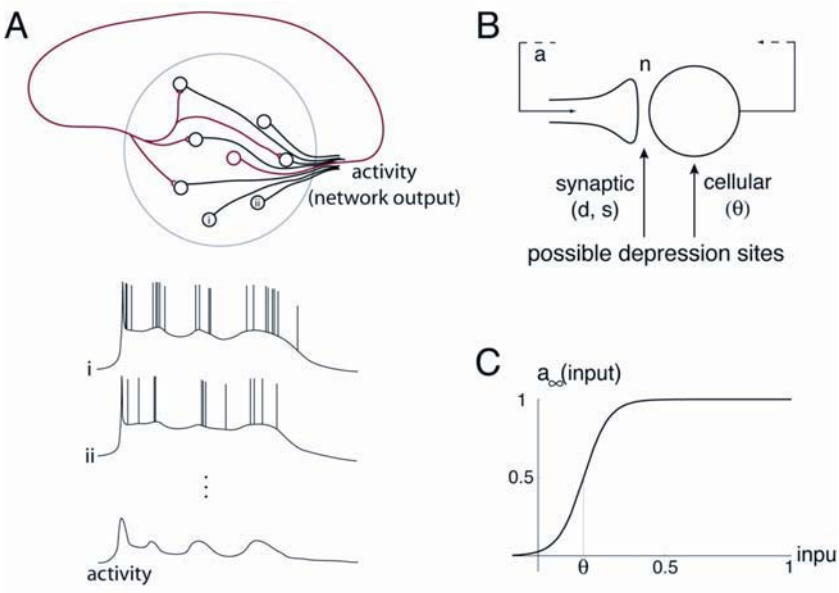


Figure 2. Model of the network generating episodic activity. (**A**) Justification for a mean-field formulation. Neurons are connected to each other through random excitatory connections. Neurons i, ii, ... are active together but their spikes are not synchronized. Spike rate can thus be averaged, giving a smooth curve, the activity (network output) that is injected back into the network. (**B**) Schematic representation of the model network. Network output (a) is fed back to the neuron population through recurrent excitatory synapses. The amount of feedback is proportional to the connectivity (n) and can be reduced by fast (d) and/or slow (s) synaptic depression. Network output can also be modulated by slow variations of the average cellular threshold (θ). (**C**) Sigmoidal input–output function of the network. Note that $a_\infty(0) > 0$, implicitly assuming that a few cells in the network are always discharging at low rate, providing a background "input" to the network. Modified from Tabak et al. (36).

$$\tau_a \dot{a} + a = a_\infty \left(n \cdot s \cdot d \cdot a \right) , \qquad [1]$$

$$\tau_d \dot{d} + d = d_\infty \left(a \right) , \qquad [2]$$

$$\tau_s \dot{s} + s = s_\infty \left(a \right) , \qquad [3]$$

or

$$\tau_\theta \dot{\theta} + \theta = \theta_\infty \left(a \right) . \qquad [3']$$

The first equation describes how activity evolves in a recurrent excitatory network. Basically, a tends to a_∞ with a time constant τ_a. The function a_∞ represents the input–output function of the network. This function depends on the input–output functions of the individual neurons, their distribution across the population, and the dynamics of the synaptic signals. For simplicity we have chosen a sigmoidal function $a_\infty(i) = 1/(1 + e^{(i-\theta)/ka})$, as illustrated in Figure 2C. For low inputs to the network, there is very little output ($a_\infty \approx 0$), until a threshold (θ) is reached; the output then quickly reaches its maximal value ($a_\infty \approx 1$); θ can be seen as an average firing threshold in the neuronal population. Note that the activity a is itself the input to the network—modulated by the effective connectivity factor $n \cdot s \cdot d$—because of the recurrent excitatory connections. The parameter n is the network connectivity, a composite measure of the number of connections per neuron and synaptic strength, which determines the maximal gain of the positive feedback loop created by excitatory connections. As described below, the activity defined by this equation is bistable over a wide range of parameters.

The second equation describes the evolution of the synaptic variable d, which represents a fast depression of the effective connectivity—τ_d ($\approx \tau_a$) is on the order of 100 ms, as in cortical networks (5). When $d = 0$ all synapses are totally depressed while synapses have full strength when $d = 1$; d_∞ is a decreasing function of a, also chosen to be sigmoidal for convenience. The interplay between a and d can create oscillations of the activity. Finally, Eq. [3] describes the variations of the slow ($\tau_s >> \tau_a$) synaptic variable s. This variable also decreases when a is large and increases for low activity, but on a much slower time scale. A possible biophysical mechanism for this slow (time scale minutes) synaptic depression involves the loss of chloride ions by the neurons during an episode, decreasing the excitatory action of gabaergic and glycinergic connections (8). Alternatively, the slow depression could be due to a cellular (not synaptic) process increasing the cellular threshold θ for high levels of activity (Eq. [3']; θ_∞ is an increasing function of a). The slow type of depression, whether synaptic or cellular, is responsible for the episodic nature of the activity (see below).

By reducing the network to a system of three differential equations, we lose the complexity of a population of neurons and only study the mean-field (deterministic) interactions between the fast positive feedback and the slower negative feedback in generating this activity. The advantage is that we are able to qualitatively analyze the dynamics of the system with fast/slow dissection techniques (see, e.g. (30)). We present this dissection in the next section, but in brief the methodology is as follows. The slow variable s is first considered as a parameter, and the dynamic states of the one-variable (a) (or two-variable (a,d), as shown by Tabak et al. (36)) fast subsystem are fully described; the steady state (and oscillatory, in the case of a two-variable fast subsystem) solutions are obtained as a function of s. Then when s is free to follow its autonomous dynamics the full system has solutions that evolve on the slow time scale of s, sampling

Table 1. Values of the parameters used for the models, unless mentioned otherwise in the text or figures

Parameter	s-model	θ-model
n, connectivity	1	1
τ_a, network time constant	1	1
θ, network threshold (half activation)	0.18	variable
k_a, inverse of slope of a_∞ at half activation	–0.05	–0.05
γ_d, time constant of fast synaptic depression d	2	2
θ_d, half activation of d	0.5	0.5
k_d, inverse of slope of d_∞ at half activation	0.2	0.2
γ_s, time constant of slow synapic depression s	500	
θ_s, half activation of s	0.14	
k_s, inverse of slope of s_∞ at half activation	0.02	
τ_θ, time constant of slow cellular adaptation θ		1000
θ_θ, half activation of θ		0.15
k_θ, inverse of slope of θ_∞ at half activation		–0.05

the solution states of the fast subsystem and at critical (bifurcation) points executing rapid transitions between states. This analysis involves numerical bifurcation and branch-tracking methods and forward-in-time integrations, which were carried out using the XPPAUT package (freely available software written by G.B. Ermentrout, http://www.pitt.edu/~phase/). A fourth-order Runge-Kutta scheme with a time step of 0.1 was used to numerically integrate the differential equations. Parameter values used in the simulations are given in Table 1.

4. PROPERTIES AND APPLICATIONS OF THE MODEL

4.1. Bistability of the Excitatory Network with Fixed Synaptic Efficacy

Let us first analyze the properties of the network without depression, that is, we study Eq. [1] and freeze the depression variables ($s = d = 1$). Such a system will reach a steady state ($(\dot{a} = 0)$ for which the activity is defined by $a = a_\infty(n \cdot a)$ (from Eq. [1]). The steady states can be determined graphically as the intersections of the straight line and the curve of $a_\infty(n \cdot a)$ shown in Figure 3A for any value of n. When n is too small ($n = 0.3$ in Figure 3A), there is only one

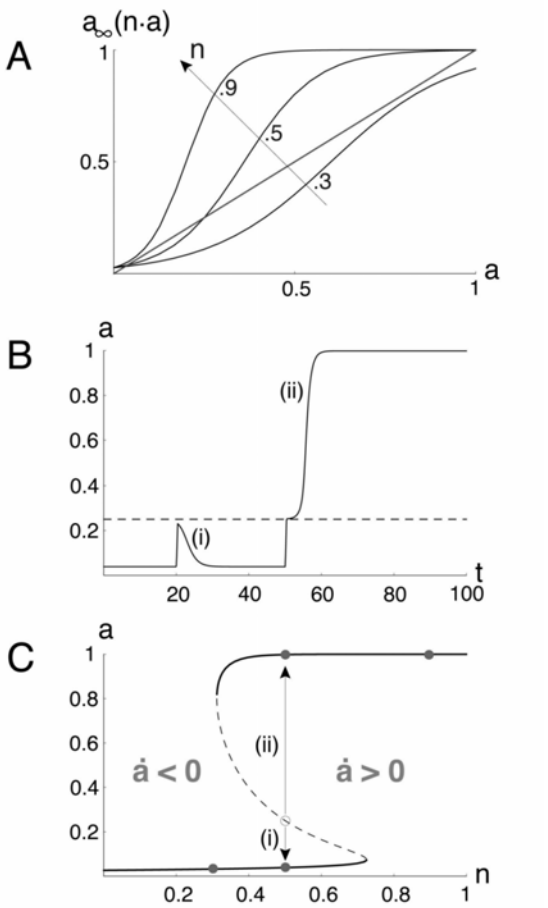


Figure 3. (**A**) Graphical solutions of the equation $a = a_\infty(n \cdot a)$. Depending on the value of n, there can be 1 or 3 solutions. (**B**) Time course of network activity for $n = 0.5$. The network receives external inputs at $t = 20$ and $t = 50$ (arbitrary unit normalized to τ_a). The first input (i) brings the activity just below the middle state (network threshold, dashed line) so activity decreases back to the low state. The second input brings activity just above the network threshold, and then jumps up to the high steady state. (**C**) Diagram showing the possible steady state values of activity for all values of n between 0 and 1. The dashed curve (middle branch) indicates unstable states. The steady states determined in A are represented on the curve by filled (stable) or open (unstable) circles. Modified from Tabak et al. (36).

intersection for a low value of a. Thus, the connectivity is too small to sustain a high level of activity. Even if the network is transiently stimulated, activity will quickly go back to its low level state. On the other hand, for large values of n (n

= 0.9 in Figure 3A), there is one intersection, but at a high activity level. Connectivity in the network is so high that activity is self-sustained. This is expected from an excitatory network: for very low connectivity the network is inactive, while for high connectivity the activity is maintained through positive feedback. It is also known that for intermediate connectivity the low and high activity states can both exist, as described below.

For intermediate values of n (n = 0.5 in Figure 3A), we find three intersections. There are steady states at low, high, and intermediate levels. Note that the middle steady state is *unstable*. This can be easily seen since the slope of $a_\infty(n \cdot a)$ is greater than 1 at this point, therefore if a is slightly increased (respectively, decreased) its derivative $\dot{a} = a_\infty(n \cdot a) - a$ will become positive (resp., negative), which will tend to further increase (resp., decrease) a. The slightest movement away from this steady state will therefore be amplified. This is illustrated in Figure 3B. If the network activity is perturbed from the low state to just below the middle state level (dashed line), activity will decrease back to its low level (i). On the other hand, if the network is kicked to just above the middle state, activity will jump to the high steady level (ii). The middle steady state is thus a *network threshold*, separating the low (inactive) and high (active) states.

We can summarize these results by plotting the activity levels (steady states) calculated when n is varied continuously. We obtain the important diagram shown in Figure 3C. The resulting "S-shaped" curve has 3 *branches*: the lower branch (solid) corresponds to the low activity states, the middle branch (dashed) corresponds to the unstable states, and the upper branch (solid) corresponds to the high activity states. The S-curve defines 2 domains in the ($a - n$) plane. For any value of n, if the activity is such that the point (n,a) is on the right of the curve, then activity will increase until the system reaches the upper branch (high state). Conversely, if (n,a) is to the left of the S-curve, activity will decrease until it reaches the low state.

We can see that for a range of values of n (approximately between 0.31 and 0.73) there are two possible stable states. The network is *bistable*. As we have seen for n = 0.5, a perturbation strong enough to cross the middle branch allows switching between the two stable states. This bistability is the basis for the oscillatory and episodic behavior described below. Imagine the network is in the high-activity state and we slowly decrease the connectivity. The state of the system, defined by a point in the ($a - n$) plane, will be on the upper branch and slowly move to the left, with a minimal decrease of activity. However, when n passes a critical value around 0.31 where the upper and middle states coincide (the "left knee" of the S-curve), the only remaining state is the low-activity state and the network crashes to that state. Now, we slowly increase n, so the state of the system tracks the lower branch, going to the right. Similarly, activity is only going to increase slightly until the "right knee," where the middle and lower states coalesce. Once n is above that point, only the high state remains and the network will jump to its high activity state, terminating the cycle.

4.2. Episodic and Rhythmic Behavior Due to Activity-Dependent Depression of Network Excitability

The network could switch spontaneously between the high and low states according to the above mechanism if we add an activity-dependent mechanism to modulate the connectivity. Therefore, we now let the slow variable s vary according to Eq. [3] (for now, the fast depression variable is still frozen: $d = 1$). Because s_∞ is a decreasing function of a, the new variable s and therefore effective connectivity $n \cdot s$ will tend to decrease when the network is in the high state and increase when the network is in the low state. This may lead to slow oscillations between the high and low states as explained in the previous paragraph and illustrated in Figure 4A.

We can explain the oscillatory behavior of the system geometrically as shown in Figure 4B. This treatment is similar to the phase plane analysis of single neuron excitability pioneered by Fitzhugh (12,30). The variations of the variables a and s (synaptic activity and fraction are not affected by slow depression) define a trajectory in the $(a - s)$ plane, called the *phase plane*. The solid gray S-curve in Figure 4B defines the states of the system for which $\dot{a} = 0$ (cf. Figure 3C) and is called the a-nullcline. The dashed curve defines the states of the system for which $\dot{s} = 0$ and is called the s-nullcline (it is simply the curve $s = s_\infty(a)$). For any value of s, if the activity is below that, curve s will be increasing, while s will be decreasing if a is above the s-nullcline. The steady states of the system (comprised of Eqs. [1] and [3]) are the intersections of the two nullclines. In the case of Figure 4B, there is only one steady state and it is unstable. A necessary condition for the steady state to be unstable is that the intersection occurs on the middle branch of the a-nullcline. If there was an intersection on the upper or lower branch, that intersection would define a stable steady state at high or low activity, which would prevent the episodic behavior. This immediately imposes a constraint on the parameters of the model if episodes are to occur.

Imagine the system is in a state (s,a) on the right of the a-nullcline. The trajectory will quickly go up as if s was constant because $\tau_a << \tau_s$, until it reaches the upper branch of the a-nullcline. Then a will remain constant, while s will decrease, since the system is now above the s-nullcline. The trajectory will thus track the upper branch, going left, until it reaches the left knee of the a-nullcline. Here, a further decrease in s forces the system to leave the a-nullcline, and, being on the left of the nullcline, activity decreases so that the trajectory quickly goes down to the lower branch of the a-nullcline. During this transition, the trajectory crosses the s-nullcline, so s increases. The trajectory will thus track the lower branch going right, until it passes the right knee, causing a new transition upward to the upper state.

So far, the combination of the bistability of the activity (Eq. [1]) and the slow, activity-dependent variations of the effective connectivity (Eq. [3]) creates

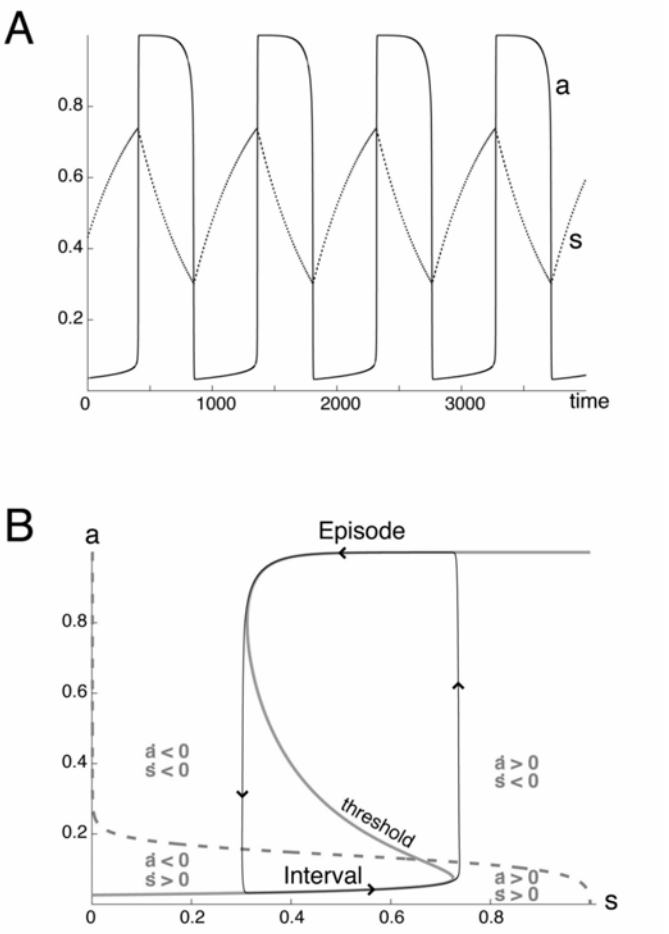


Figure 4. Episodic behavior of the network (Eqs. [1] and [3]). (**A**) Slow oscillatory variations of activity (a, solid curve) and slow depression variable (s, dotted curve) with time. Time is in arbitrary units. (**B**) Phase plane representation of the episodic behavior. The trajectory continuously cycles through the high (Episode) and low (Interval) activity states. The transitions between the two activity levels are very fast because they are governed by the small time constant τ_a, while the evolution at either level is slow since it is governed by a large time constant τ_s. Gray S-shaped curve = a-nullcline; dashed curve = s-nullcline.

slow, spontaneous oscillations between low and high states that mimic the episodic activity observed experimentally. Now, if we allow the fast depression variable, d, to vary (according to Eq. [2]), the system may oscillate quickly during each episode. The system then generates rhythmic episodes as shown in Figure 5. For a complete analysis of the full system using fast/slow dissection technique, see Tabak et al. (36).

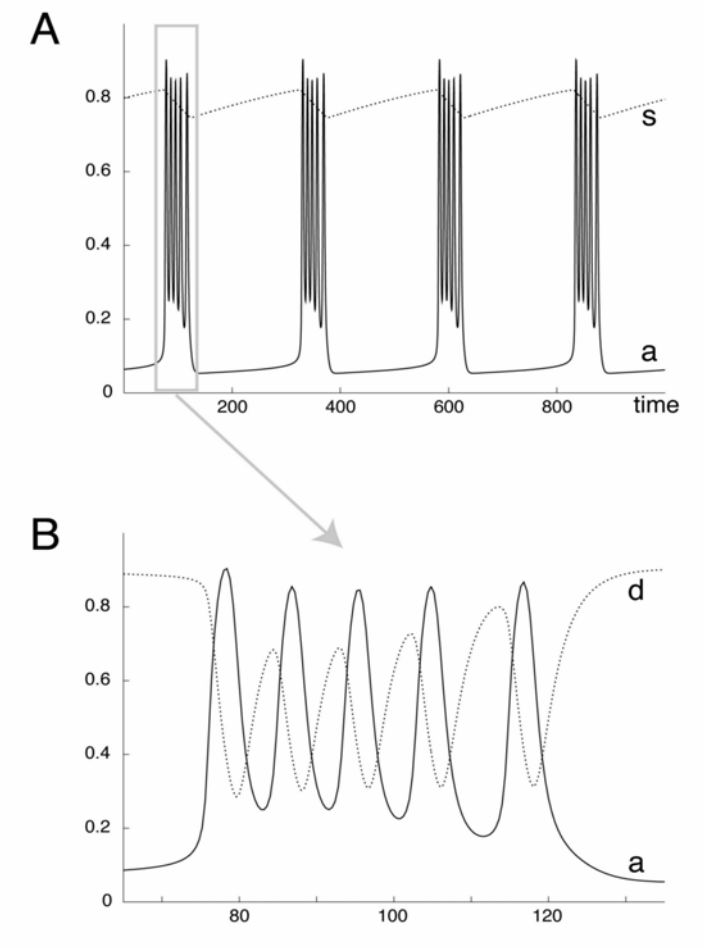


Figure 5. Episodic and rhythmic behavior of the full system (Eqs. [1]–[3]). (**A**) Time variations of *a* and *s*, showing the episodic behavior with fast oscillations. Note that episodes and intervals between episodes are shorter than in Figure 4A. This is mostly because episodes terminate at a higher value of *s* when fast depression is present. (**B**) Detail of an episode on a faster time scale, showing the fast oscillations of *a* and *d*.

Finally, note that if we had used Eq. [3'] instead of Eq. [3], that is, if we had considered a cellular adaptation mechanism instead of a synaptic depression mechanism to drive the episodic behavior, we would obtain a similar dynamical behavior of the system. At this point, we do not see a qualitative distinction between these two mechanisms (cellular or synaptic) of episode generation. In the next sections, we denote these models the *s*-model (synaptic depression) and the

θ-model (cellular adaptation), and we will show particular parameter variations from which differences in their behaviors will emerge.

4.3. Relationship Between Episode Duration and Inter-Episode Interval

Can we trigger the network before synaptic strength (or firing threshold, if we consider the θ-model) has fully recovered? We can answer this question directly by looking at Figure 4B. At any time during the inter-episode interval (while the system is tracking the lower branch of the S-curve), if we transiently "stimulate" the network so that activity increases above the threshold, the phase point will move to the high state, initiating an episode. However, the episode will start from a lower value of s than a spontaneous episode, therefore the system will track a shorter segment of the upper branch before reaching the left knee and falling back to the low state (whatever the value of s for which we triggered an episode, the critical value of s at which the episode terminates is always the same). Therefore, the triggered episode is shorter than a spontaneous episode. More precisely, the longer we wait to artificially trigger an episode, the longer the episode is, as illustrated in Figure 6A,B.

This model prediction is testable experimentally. Indeed, we have shown that it is possible to trigger episodes by stimulating sensory nerves afferent to the spinal cord, and that the duration of the stimulated episodes increases with the interval between the triggered episode and the end of the previous (spontaneous) episode, as shown in Figure 6C,D (37). In other words, the longer we let the network excitability recover, the longer the triggered episode is. This suggests that in the experimental preparation, as in the model, there is a critical value of network excitability for which all triggered episodes terminate.

In Figure 6D we have also plotted the durations of spontaneous episodes (gray dots) against the recovery interval that just preceded the episodes. Although their range is different, the relationship is the same as for stimulated episodes, suggesting again that all episodes (spontaneous or triggered) terminate at a fixed level of network excitability. To confirm this finding we have also looked at the relationship between episode duration and the *following* interval and found no correlation (37). This lack of correlation suggests that there is no "memory" of the system's state once an episode is terminated, supporting our finding that all episodes terminate at a fixed level of network excitability.

Figure 6D shows that spontaneous episodes occur after various intervals. Unlike our simple model, episodes therefore can start at various levels of network excitability. Episode initiation is a stochastic event. Although a higher level of excitability means a higher probability of triggering an episode, the network needs a triggering event in order for an episode to start, and this is where randomness is introduced. On the other hand, as our data suggest, all episodes terminate at the same value of network excitability; episode termination is

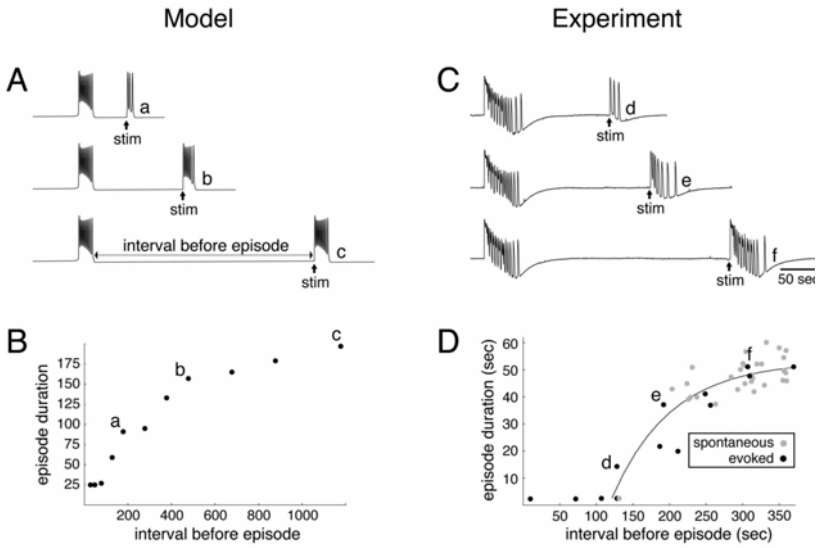


Figure 6. Relationship between episode duration and interval preceding the episode. (**A**) Time course of activity generated by the *s*-model for different intervals between a spontaneous episode and a triggered (stim) episode. (**B**) Plot of episode duration against preceding interval for the model; a, b, c correspond to the traces shown in A. (**C**) Time course of activity generated by a spinal cord obtained from a 10-day-old chick embryo. Stimulations (stim) were applied at different time intervals after a spontaneous episode. Traces were high-pass filtered at 0.01 Hz. (**D**) Plot of episode duration against preceding interval for evoked (black circles) and spontaneous episodes (gray circles); d, e, f correspond to the traces shown in C. Modified from Tabak et al. (36).

a deterministic event. This seems to be a property of many developing or hyperexcitable systems that generate episodic activity, since the same relationships (positive correlation between episode duration and preceding—but not following—interval) was found in the developing retina (14), developing cortical networks (27), hippocampal slices (33), and disinhibited spinal networks (32,35). It is therefore tempting to suggest that similar mechanisms operate in the generation of episodic activity by all these different networks.

4.4. Recovery of the Activity after Blockade of Excitatory Connections

One of the most surprising features of the spontaneous activity is its robustness to pharmacological perturbation. When synaptic transmission mediated by the excitatory neurotransmitters glutamate and acetylcholine is blocked using pharmacological agents, the spontaneous activity stops for a long period of time

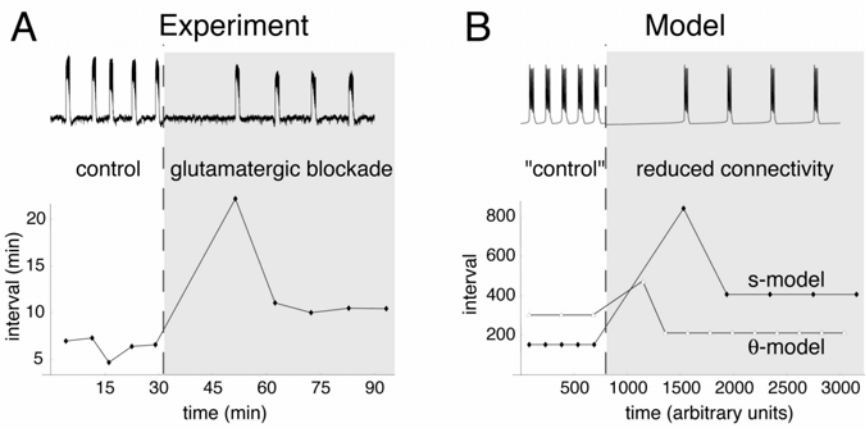


Figure 7. Recovery of activity after partial block of excitatory connections. (**A**) Results from a 10-day-old chick, showing activity (upper trace) and inter-episode intervals in control and after blockade of some glutamatergic connections (100 μM APV). (**B**) Model results before ("control") and after n was decreased from 1.2 to 0.9 (–25%). The upper trace was obtained with the s-model. On the interval plot, results from both the s-model (filled diamonds) and θ-model (open triangles) are shown. Modified from Tabak et al. (36).

(compared to the inter-episode intervals) but then recovers, as shown in Figure 7A (1,7). The activity is then thought to depend exclusively on gabaergic and glycinergic synapses, which have, early in development, an excitatory role. It must be noted that the recovery is not gradual, as the new inter-episode intervals are not gradually decreasing to a new level but remain at a fixed level immediately after the activity has recovered (Figure 7A; note that after recovery the intervals are larger than in control conditions). We could imagine an additional process through which the lack of activity is detected and compensated by slowly increasing network excitability until activity reaches its control level. However, this slow homeostatic process may cause a progressive—not abrupt—decrease of inter-episode intervals, unlike the experimental observation.

Surprisingly, the s-model can explain this recovery without introducing any additional variable into the model. If we decrease the connectivity parameter (n), in order to mimic the blockade of a fraction of the connections, the s-model reacts exactly like the chick spinal cord: activity first stops, then recovers with a fixed inter-episode interval (Figure 7B; (37)). As for the experimental result, the new inter-episode intervals were larger than before the blockade. In addition, episode duration was only slightly affected, also in agreement with the experiments. On the other hand, the θ-model behaved in a different way: activity recovered quickly after the reduction in connectivity and the intervals were then slightly *lower* than in "control" (Figure 7B, "θ-model"). In addition, episode

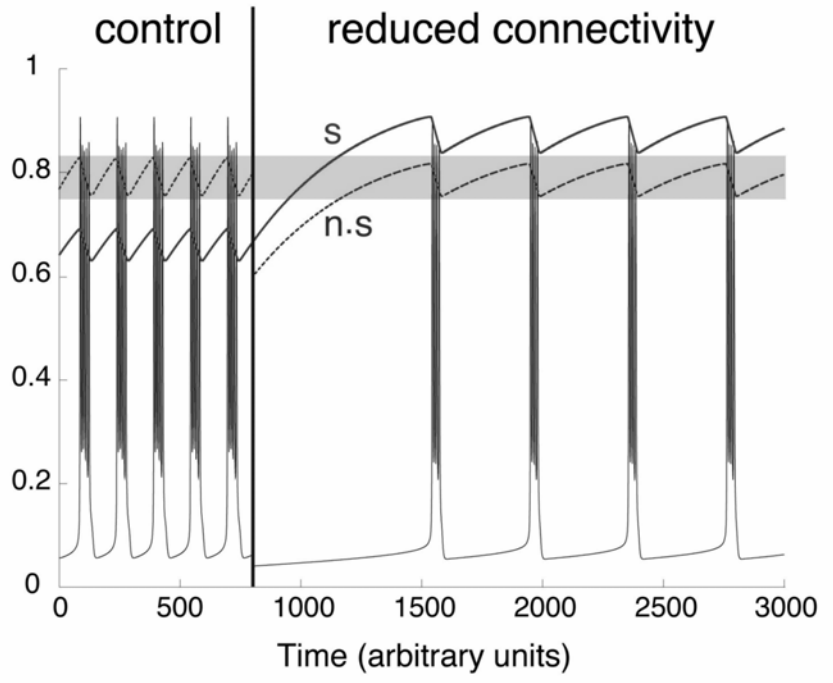


Figure 8. Time course of activity, synaptic strength (s), and effective connectivity ($n \cdot s$) produced by the s-model before ("control") and after ("reduced connectivity") n was decreased by 25% (from 1.2 to 0.9). Activity stops after n was decreased, because the effective connectivity becomes too small to support network activity. This allows s to increase beyond its control level, until $n \cdot s$ reaches its control level. Activity then reoccurs, although with larger inter-episode intervals. Modified from Tabak et al. (36).

duration was markedly reduced for the recovered activity (not shown). This is an important qualitative difference between the s- and θ-models. This difference in how the models react to a decrease in connectivity suggests that the episodic nature of the spontaneous activity in the chick cord is due mostly to synaptic depression, not cellular adaptation.

In order to understand the recovery of activity, we plotted again the activity generated by the s-model in Figure 8, together with the slow depression variable (s) and the effective connectivity ($n \cdot s$). As described previously, s and therefore $n \cdot s$ decrease during the episodes and increase during the inter-episode intervals. When the connectivity (n) is suddenly decreased by 25% (from 1.2 to 0.9), this causes a corresponding decrease of $n \cdot s$, the effective connectivity in the network, i.e., the effective gain of the positive feedback due to excitatory connections. Activity is therefore blocked until this gain can reach back to its "control"

value. This can happen because in the absence of activity s keeps increasing towards its asymptotic value, 1. As soon as $n \cdot s$ reaches its control value, episodes of activity reoccur. The intervals between episodes are somewhat longer than in control, because the exponential increase of s during the intervals is slower since s is now closer to its asymptotic value. Therefore, the activity can recover after a moderate decrease of connectivity because the system *compensates* by decreasing the level of depression (increasing s).

This predicts that the unblocked synapses see their availability or efficacy increase relative to their control level. To verify this prediction, we have stimulated a pathway (between adjacent ventral roots) that does not contain glutamatergic synapses. Indeed, we have shown that the strength of the response was increased after blockade of glutamatergic connections and subsequent recovery of the activity (37). The developing spinal circuits are therefore able to approximately maintain their level of activity following the blockade of some of their connections. This is very important since the temporal pattern of activity may be important in the development of network and cellular properties (15,34). During development, some cells and connections may be lost, while other synapses may see their efficacy increased. Through the very mechanism that regulates its patterned activity (activity-dependent depression), the developing spinal cord is able to compensate for these changes and therefore maintain its activity level within a certain operating range.

5. DISCUSSION AND FUTURE WORK

We have presented an idealized model of spontaneous activity in developing neural networks. Despite its mean-field approach, the model captures the emergent nature of the phenomenon. Although we implicitly assumed that a small proportion of cells were active, none have pacemaker capabilities, so the ensemble interactions are crucial for generating the episodic rhythm, not just for synchronizing cellular oscillators.

The model presented in this chapter was developed to understand the spontaneous activity in the developing spinal cord. However it is general enough to apply to spontaneous activity in other developing circuits. Indeed, other modeling and experimental studies have suggested that similar mechanisms can explain the spontaneous episodic activity in developing retinal (4,14,24) and cortical networks (27). These mechanisms involve fast positive feedback through excitatory connections together with a slow activity-dependent depression of network excitability. They may therefore be common to many developing networks. This gives us a framework to study developing and excitatory networks. Can this framework be helpful to the study of mature neural networks?

One critical fact is that with maturation, networks acquire functional inhibitory connections, which stops the episodic activity. Mature networks become segregated into subserving different functions, with a great variety of patterns of activity. Can the concept of recurrent excitatory network still be used to understand mature networks? In the following we review some evidence from experimental and modeling studies suggesting that the "immature" mechanisms of burst generation could be conserved in more mature networks.

First, episodic bursts of activity can be generated in "mature" networks that are disinhibited by pharmacological block of their inhibitory connections. This was shown in in vitro spinal (3,10) and hippocampal (45) networks. Therefore, the immature mechanism is still potentially present in mature networks and can be unmasked. Furthermore, episodic activity can be generated through the same mechanism in some networks with functional inhibitory connections if inhibition is not too strong or if the networks are rendered more excitable (20,21,33,38–41). Thus, bursting activity could be evoked in mature networks through neuromodulators that would decrease synaptic inhibition and/or raise cellular excitability.

Although demonstrating that a mechanism of activity can be uncovered does not mean that this mechanism is in fact used during the normal function of a network, there are several examples suggesting such possibility. A model of the spinal circuit for swimming in the lamprey is based on two excitatory subnetworks that generate bursts using a cellular adaptation mechanism. These two units are connected by mutual inhibitory connections, ensuring that the pattern of rhythmic bursts is in alternation between left and right sides (18). Therefore, it is possible that the rhythmic locomotor activity is simply a faster version of the spontaneous activity, with inhibition in the mature network simply allowing the coordination between left and right sides, as well as between flexors and extensors in higher vertebrates. Inhibition would also ensure that the locomotor network is not always "on," but only activated when necessary. Coupled rhythm-generating circuits control many functions like locomotion, respiration, and chewing (9), and it is therefore important to understand how these circuits generate oscillatory activity.

Another example suggesting that the immature mechanism may play a role in mature networks comes from studies in cortex of anesthetized cat. Timofeev and colleagues (39) showed that isolating a small slab of cortex led to episodic bursts of activity (about 5 bursts per minute), with a mechanism similar to the one presented herein. However, when they recorded from a larger network, they observed the 1H-z oscillation that is observed during sleep. This led to the suggestion that the cortical sleep (<1 Hz) oscillations and the episodic bursts in small slabs could be generated through the same type of mechanism.

Finally, an application of this type of activity regarding neural computation was presented by Loebel and Tsodyks (23). This processing has for its basis the short "population spikes" generated by networks of mostly excitatory neurons

with synapses subject to a fast depression (a small fraction of inhibitory cells didn't disrupt this activity). These population spikes are generated through a similar mechanism as the episodes described here, but on a much faster time scale. During a population spike, most neurons fire only once, but they become transiently synchronized (41). Such synchrony among neuronal populations could play an important role in cortical information processing, and Loebel and Tsodyks have suggested some applications of the transient synchronization allowed by population spikes, by considering whether transient input signals elicited a response (i.e., a population spike) or not from the network. Success in triggering a population spike depends on the level of depression of the synapses in the network at the time of the stimulus. In other words, the response of the network to an input depends on the (short-term) history of network activity (23).

On the other hand, it is possible that the episodic activity, which is characteristic of developing networks, is an abnormal mode of activity in mature networks. For example, the episodic activity in cortical slabs mentioned above is due to the much decreased number of synaptic connections once the slab is isolated (39). Also, some forms of epileptic activity resemble the spontaneous episodic activity of developing systems. Indeed, epileptic events can be caused by an impairment of inhibitory synapses, or by an anomalous level of excitation. An example of "epileptic activity in vitro" was shown by Staley et al. (33), who recorded spontaneous episodic bursts of activity in hippocampal slices that were disinhibited or rendered hyperexcitable. They showed that these bursts were regulated by synaptic depression, as in the *s*-model. Alternatively, a model by Traub and Dingledine (40) of "epileptic" bursts in hyperexcitable hippocampal networks proposed that bursts are terminated by a slow hyperpolarizing current, that is, a cellular type of depression.

It is critical to know whether these bursts are terminated by a synaptic or cellular process, if one wants to choose an appropriate pharmacological treatment. As a thought experiment, suppose that we want to suppress "epileptic bursts" in a hyperexcitable hippocampal network. Should we target cellular excitability or synaptic connections? If, for example, the bursts are terminated by synaptic depression as in the Staley et al. (33) experiments, our results with the *s*-model suggest that we should use a pharmacological agent that blocks excitatory synapses, as this will increase the interval between each burst. Decreasing cellular excitability would also increase the interburst interval, but it would increase burst duration as well. Similarly, the effectiveness of a drug potentiating inhibitory synapses in order to stop the bursts would depend on the type of depression mechanism that terminates the bursts and on the effects of the inhibitory connections—phasic or tonic.

Much recent work has been aimed at understanding the role of inhibition in neuronal networks. Although our approach has emphasized excitatory networks in vertebrates, it should be pointed out that circuits of inhibitory neurons may also produce oscillations (31,42,43). A great deal of future work should be con-

cerned with adding inhibitory cells in our network models. Because the effects of inhibition can depend on many factors (local versus diffuse connectivity, phasic versus tonic, etc.), this will certainly require the use of cell-based ensemble models (as by Timofeev et al. (39) and Tsodyks et al. (41)), instead of the mean-field type of model presented here.

It will be particularly important to understand how network organization is changed as gabaergic synapses switch from excitatory to inhibitory during development, a transition that may be driven by network activity itself (13). This transition occurs in parallel to developmental changes in excitatory connectivity and cellular properties, changes which may also, in part, be due to spontaneous activity. But how does activity modify these network properties? Our understanding of the spontaneous activity in developing systems will facilitate the study of the role of this activity for network maturation. It will become necessary to identify the long-term mechanisms of activity-dependent plasticity operating in developing networks. These "learning rules" will then be added to our models of spontaneous activity, allowing us to study how activity in a network leads to changes in that network, changes which in turn will affect activity (see (19) for an example of how synaptic plasticity leads to changes in the patterns of activity). This effect of activity on itself, by way of modifying network properties, is one striking feature of neural network complexity.

6. ACKNOWLEDGMENTS

We thank Jay Demas for providing his unpublished data on mouse retinal activity. We also thank Cristina Marchetti for her comments on the manuscript.

7. NOTES

1. When a neuron discharges an action potential, it releases neurotransmitters from its synaptic terminals onto postsynaptic neurons. Some neurotransmitters will increase the membrane potential of the postsynaptic neurons, which makes the postsynaptic neurons more likely to fire action potentials. This is what we mean by "excitatory connection." Other transmitters will decrease postsynaptic membrane potential or increase membrane conductance (shunt) such that the postsynaptic neuron is less likely to generate action potentials; this type of connection is called "inhibitory."

2. We use a "top–down," "rate-coded" approach, as defined in the previous chapter 5.1. by Reeke.

8. REFERENCES

1. Barry MJ, O'Donovan MJ. 1987. The effects of excitatory amino acids and their antagonists on the generation of motor activity in the isolated chick spinal cord. *Brain Res* **433**:271–276.
2. Ben-Ari Y. 2002. Excitatory actions of gaba during development: the nature of the nurture. *Nature Rev Neurosci* **3**:728–739.
3. Bracci E, Ballerini L, Nistri A. 1996. Spontaneous rhythmic bursts induced by pharmacological block of inhibition in lumbar motoneurons of the neonatal rat spinal cord. *J Neurophysiol* **75**:640–647.
4. Butts DA, Feller MB, Shatz CJ, Rokhsar DS. 1999. Retinal waves are governed by collective network properties. *The J Neurosci* **19**:3580–3593.
5. Chance FS, Nelson SB, Abbott LF. 1998. Synaptic depression and the temporal response characteristics of V1 cells. *J Neurosci* **18**:4785–4799.
6. Cherubini E, Gaiarsa J-L, Ben-Ari Y. 1991. GABA: an excitatory transmitter in early postnatal life. *Trends Neurosci* **14**:515–519.
7. Chub N, O'Donovan MJ. 1998. Blockade and recovery of spontaneous rhythmic activity after application of neurotransmitter antagonists to spinal networks of the chick embryo. *J Neurosci* **18**:294–306.
8. Chub N, O'Donovan MJ. 2001. Post-episode depression of GABAergic transmission in spinal neurons of the chick embryo. *J Neurophysiol* **85**:2166–2176.
9. Cohen A, Rossignol S, Grillner S, eds. 1988. *Neural control of rhythmic movements in vertebrates*. Wiley, New York.
10. Darbon P, Scicluna L, Tscherter A, Streit J. 2002. Mechanisms controlling bursting activity induced by disinhibition in spinal cord networks. *Eur J Neurosci* **15**:671–683.
11. Fedirchuk B, Wenner P, Whelan PJ, Ho S, Tabak J, O'Donovan MJ. 1999. Spontaneous network activity transiently depresses synaptic transmission in the embryonic chick spinal neurons. *J* **Neurosci**:19, 2102–2112.
12. Fitzhugh R. 1961. Impulses and physiological states in models of nerve membrane. *Biophys J* **1**:445–466.
13. Ganguly K, Schinder AF, Wong ST, Poo MM. 2001. GABA itself promotes the developmental switch of neuronal GABAergic responses from excitation to inhibition. *Cell* **105**:521–532.
14. Grzywacz N, Sernagor E. 2000. Spontaneous activity in developing turtle retinal ganglion cells: statistical analysis. *Visual Neurosci* **17**:229–241.
15. Gu X, Spitzer NC. 1995. Distinct aspects of neuronal differentiation encoded by frequency of spontaneous Ca^{2+} transients. *Nature* **29**:784–7876.
16. Hamburger V, Balaban M. 1963. Observations and experiments on spontaneous rhythmical behavior in the chick embryo. *Dev Biol* **7**:533–545.
17. Katz LC, Shatz CJ. 1996. Synaptic activity and the construction of cortical circuits. *Science* **274**:1133–1138.
18. Hellgren-Kotaleski J, Lansner A, Grillner S. 1999. Neural mechanisms potentially contributing to the intersegmental phase lag in lamprey, II: hemisegmental oscillations produced by mutually coupled excitatory neurons. *Biol Cybern* **81**:299–315.
19. Izhikevich EM, Gally JA, Edelman GM. 2004. Spike-timing dynamics of neuronal groups. *Cerebral Cortex* **14**:933–944.
20. Latham PE, Richmond BJ, Nelson PG, Nirenberg S. 2000. Intrinsic dynamics in neuronal networks, I: theory. *J Neurophysiol* **83**:808–827.

21. Latham PE, Richmond BJ, Nirenberg S, Nelson PG. 2000. Intrinsic dynamics in neuronal networks, II: experiment. *J Neurophysiol* **83**:828–835.
22. Li M, Jia M, Fields RD, Nelson PG. 1996. Modulation of calcium currents by electrical activity. *J Neurophysiol* **76**:2595–2607.
23. Loebel A, Tsodyks M. 2002. Computation by ensemble synchronization in recurrent networks with synaptic depression. *J Comput Neurosci* **13**:111–124.
24. Meister M, Wong RO, Baylor DA, Shatz CJ. 1991. Synchronous bursts of action potentials in ganglion cells of the developing mammalian retina. *Science* **252**:939–943.
25. O'Donovan MJ, 1999. The origin of spontaneous activity in developing networks of the vertebrate nervous system. *Curr Opin Neurobiol* **9**:94–104.
26. O'Donovan MJ, Landmesser LT. 1987. The development of hindlimb motor activity studied in the isolated spinal cord of the chick embryo. *J Neurosci* **7**:3256–3264.
27. Opitz T, De Lima AD, Voigt T. 2002. Spontaneous development of synchronous oscillatory activity during maturation of cortical networks in vitro. *J Neurophysiol* **88**:2196–2206.
28. Paulsen O, Sejnowski TJ. 2000. Natural patterns of activity and long-term synaptic plasticity. *Curr Opin Neurobiol* **10**:172–179.
29. Pinto DJ, Brumberg JC, Simons DJ, Ermentrout GB. 1996. A quantitative population model of whisker barrels: re–examining the Wilson-Cowan equations. *J Comput Neurosci* **3**:247–264.
30. Rinzel J, Ermentrout GB. 1998. Analysis of neural excitability and oscillations. In *Methods in neuronal modeling*, 2nd ed., pp. 251–292. Ed. C Koch I Segev. MIT Press, Cambridge.
31. Rinzel J, Terman D, Wang X, Ermentrout B. 1998. Propagating activity patterns in large-scale inhibitory neural networks. *Science* **279**:1351–1355.
32. Rozzo A, Ballerini L, Abbate G, Nistri A. 2002. Experimental and modeling studies of novel bursts induced by blocking Na^+ pump and synaptic inhibition in the rat spinal cord. *J Neurophysiol* **88**:676–691.
33. Staley KJ, Longacher M, Bains JS, Yee A. 1998. Presynaptic modulation of CA3 network activity. *Nature Neurosci* **1**:201–209.
34. Stellwagen D, Shatz CJ. 2002. An instructive role for retinal waves in the development of retinogeniculate connectivity. *Neuron* **24**:673–685.
35. Streit J, Tscherter A, Heuschkel MO, Renaud P. 2001. The generation of rhythmic activity in dissociated cultures of rat spinal cord. *Eur J Neurosci* **14**:191–202.
36. Tabak J, Senn W, O'Donovan MJ, Rinzel J. 2000. Modeling of spontaneous activity in the developing spinal cord using activity-dependent depression in an excitatory network. *J Neurosci* **20**:3041–3056.
37. Tabak J, Rinzel J, O'Donovan MJ. 2001. The role of activity-dependent network depression in the expression and self-regulation of spontaneous activity in the developing spinal cord. *J Neurosci* **21**:8966–8978.
38. Tscherter A, Heuschkel MO, Renaud P, Streit J. 2001. Spatiotemporal characterization of rhythmic activity in rat spinal cord slice cultures. *Eur J Neurosci* **14**:179–190.
39. Timofeev I, Grenier F, Bazhenov M, Sejnowski TJ, Steriade M. 2000. Origin of slow cortical oscillations in deafferented cortical slabs. *Cerebral Cortex* **10**:1185–1199.
40. Traub RD, Dingledine R. 1990. Model of synchronized epileptiform bursts induced by high potassium in CA3 region of rat hippocampal slice: role of spontaneous epsps in initiation. *J Neurophysiol* **64**:1009–1018.
41. Tsodyks M, Uziel A, Markram H. 2000. Synchrony generation in recurrent networks with frequency-dependent synapses. *J Neurosci* **20**:RC50.

42. Wang XJ, Buzsaki G. 1996. Gamma oscillations by synaptic inhibition in a hippocampal interneuronal network model. *J Neurosci* **16**:6402–6413.
43. Whittington MA, Traub RD, Jefferys JG. 1995. Synchronized oscillations in interneuron networks driven by metabotropic glutamate receptor activation. *Nature* **373**:612–615.
44. Wilson HR, Cowan JD. 1972. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys J* **12**:1–24.
45. Wong RK, Traub RD. 1983. Synchronized burst discharge in disinhibited hippocampal slice, I: initiation in the CA2–CA3 region. *J Neurophysiol* **49**:442–458.
46. Zhang LI, Poo MM. 2001. Electrical activity and development of neural circuits. *Nature Neurosci* **4**(suppl.):1207–1214.
47. Ho S, O'Donovan MJ. 1993. Regionalization and intersegmental coordination of rhythm-generating networks in the spinal cord of the chick embryo. *J Neurosci* **13**:1354–1371.

5.3

CLINICAL NEURO-CYBERNETICS: MOTOR LEARNING IN NEURONAL SYSTEMS

Florian P. Kolb

Institute of Physiology, University of Munich, Munich, Germany

Dagmar Timmann

Department of Neurology, University of Duisburg-Essen, Essen, Germany

Although the understanding of complex cerebellar function is still a matter of discussion, new imaging techniques have provided evidence that the human cerebellum is critically involved in motor learning. For associative plastic motor processes this evidence has been obtained by comparison of classically conditioned eyeblink results from cerebellar patients with those from corresponding control subjects, the former showing typically reduced incidence levels of conditioned responses. Of particular interest was that ***non-associative***-motor-related processes such as habituation are also affected characteristically in cerebellar patients. Aside from these motor-related functions, we also review evidence that the cerebellum may be involved as well in ***non-motor*** visuomotor associative learning. The common denominator for impaired function may be an inadequate error-detection or error-correction capability, a putative function of the olivo-cerebellar system. In the final section of this chapter, we review computational models based on feedback-error learning.

1. INTRODUCTION

1.1. Types of Learning

Changes in the activity of the central nervous system are characteristic and desirable in developing systems, as has been pointed out in previous chapter 5.2 by Tabak and Rinzel (Part III, this volume), which presents a model for spon-

Address correspondence to: Florian P. Kolb, Institute of Physiology, University of Munich, Pettenkoferstrasse 12, 80336 Munich, Germany (F.Kolb@LMU.de).

taneous activity in the developing spinal cord, also applicable to other developing systems. Other types of changes following possibly similar rules are seen in tumor growth and other negative aspects of nervous systems; a novel model for that aspect is presented in previous chapter 5.1 by Reeke.

Learning based on plastic changes is a multilevel, multifaceted phenomenon, and its analysis depends on the complexity of the learning situation. In a visuospatial delayed-response task, a monkey must retain the previously given information about the position of a raisin in one of two food wells, both then covered with identical pieces of cardboard. A screen is then lowered for a "delay" of approximately 10 s and then raised again. To obtain the reward, the monkey must choose the position at which the food had been placed prior to the delay. The marine polychete worm *Nereis pelagica* contracts in response to a sudden vibration or a shadow. Both behavioral reactions are typical examples of learning. The analysis of very simple learning in a very simple nervous system is attractive from the point of dissecting and understanding it. The analysis of learning in complex systems may be more "interesting" and relevant to learning in humans. The examples given differ markedly in their complexity, and their analyses thus require completely different experimental approaches.

The general meaning of learning is associated with acquisition and retention of either facts or behaviors that were not present in the organism beforehand. The type of learning is closely related to the corresponding memory. The memory can be classified (66) into two main categories, one of which, the declarative explicit memory, and the other, the non-declarative implicit memory. The explicit memory refers to recollection of facts and events, and requires the integrity of the medial temporal lobe. The implicit memory is related to the ability to perform procedures (procedural memory) and seems to be independent of the medial temporal lobe. The corresponding location within the brain is a matter of an ongoing debate, although data have been accumulated providing evidence that the cerebellum is involved. This chapter focuses on the implicit memory and the corresponding procedures for acquisition of motor behavior.

Although the category "motor learning" is defined loosely only (8), Ito (31) provided helpful interpretations: "motor learning" implies both "adaption" and "learning"—processes that are not always clearly distinct from each other, whereas "learning" implies more than "adaptation." Both procedures must be able to adjust parameters to maintain optimal control performance under changing circumstances with the adjustment as a progressive process and sequential exploration of optimal conditions. During repetitive trials an adaptive system will always make the same sequential exploration, whereas in a learning system exploration will be improved from trial to trial such that the system will achieve the optimal point faster and with more accuracy than in the previous trial (31). Adaptation utilizes preceding experience within an ongoing experiment, whereas learning is based on experiences acquired in preceding experiments. Consequently, learning is based not only upon adaptive mechanisms but also on

additional features of maintaining and utilizing results from earlier experiments (31). Adams (1) specializes the learning aspect to improvement in the quality of motor performance with respect to accuracy, speed, and minimal energy, whereas skill is more than the ability to perform and often peaks in individual different proficiencies. Mechanisms different from these are "habituation" or "sensitization," which represent the change in response amplitudes to repetitively given identical stimuli. The amplitudes may increase (sensitization) or decrease (habituation) due to an increasing (sensitization) or decreasing (habituation) mismatch between stimulation and an updated representation of the outside world.

Adaptation, habituation, and sensitization may be summarized as nonassociative procedures. Correspondingly, other types of learning involve the formation of association among stimuli or among stimuli and actions (e.g. (20)). Results of associative processes can be studied experimentally by classical and instrumental conditioning. In instrumental conditioning the subject learns relations among actions and their outcome. This is different from classical conditioning, in which the subject learns relations among stimuli. Two stimuli—the conditioning stimulus (CS) and the unconditioned stimulus (US)—are paired with each other, so that the CS comes to evoke a conditioned response (CR), which is similar to the unconditioned response (UR) elicited by the US. An effective protocol is delay conditioning, in which the CS is preceding US onset and coterminates with the US. Classical conditioning of the eyeblink reflex is one of the most studied experimental approaches of simple associative learning in mammals (e.g. (11)). In a typical protocol, tone is the CS and corneal air puff the unconditioned stimulus. The unconditioned stimulus elicits closure of the eyelid in humans as well as extension across the cornea of the nictitating membrane (the internal eyelid) in rabbits. With repetitive presentation of paired CS–US (tone–air puff), subjects learn to blink in response to the tone prior to onset of the air puff or to the tone alone. Animal and human lesion studies could show that elemental delay conditioning of eyeblink is dependent on the cerebellum but independent of the cerebral cortex and hippocampus. In more complex forms of conditioning other brain regions are engaged. For example, in trace conditioning, the CS starts and ends before the US starts, and the subjects must hold offline information about the CS before US onset. In addition to the cerebellum, trace conditioning is dependent on other brain regions including the hippocampus.

"Despite continuing work on the structure and function of the cerebellum, there is still no consensus as to what it does and how it does it" (67). The cerebellar cortex appears as a structure similar to a rectangular lattice of high precision. Together with its extraordinary double innervation via a low-frequency system (climbing fiber system) of high transmission probability and a high-frequency system characterized by an enormous amount of convergence (mossy fiber-granule cell system), scientists have been inspired to numerous and even

speculative ideas about its function. From the beginning of the 1960s theoretical approaches, hypotheses, and models of the cerebellum have been created. Some models were derived from the strategic position of the cerebellum within closed or open loops of structures assumed to be involved in active movements (e.g. (29,3,59)). Others were based on the unique anatomical microarchitecture of the cerebellum (e.g. (12,48,2)) or on the requirement to transform motor planning into motor execution coordinates (56). More recent models assume the cerebellum to be an adaptive controller, or, more exactly, that the cerebellum could be modeled as an adaptive controller, as pointed out by Barlow (6). In his recently published book, he reviews different types of models as well as the experimental sources these models were derived from (6).

In the subsequent paragraphs and based on experimental approaches, we will first review the evidence in human lesion and neuroimaging studies about involvement of the human cerebellum in classical conditioning of the eyeblink response. In addition, based on the recent discussion of the cerebellum's involvement in non-motor functions, we will review evidence that the cerebellum may be involved in visuomotor associative learning as well as in habituation processes (69). Different hypotheses and models of cerebellar function in learning will be discussed in the theoretical part of this chapter.

1.2. Anatomy of the Cerebellum

The cerebellum is located behind and below the cerebral hemispheres, overlying the brain stem (for a review, see (14,77)). The cerebellum consists of two large hemispheres and has, in contrast to most other parts of the brain, a midline structure, the vermis. In the anterior-posterior direction, the cerebellum is subdivided into the anterior, posterior, and flocculonodular lobes. The three lobes are subdivided into several lobules. Larsell has introduced a numbering system, based on comparative studies on phylogenetic similarities (44), which consists of Roman numerals in the vermis and the prefix H in the hemispheres.

Each half of the cerebellum contains four distinct nuclei. The fastigial nucleus is located most medially, followed by the globose and emboliform nucleus and, most laterally, the dentate nucleus. The emboliform and globose nuclei in humans most likely resemble the anterior and posterior interposed nuclei in animals.

On the basis of efferent projections from the cerebellar cortex to the cerebellar nuclei, Jansen and Brodal (32), and later Chambers and Sprague (16), suggested a subdivision into three longitudinal (sagittal) zones: a medial zone (vermis) projecting to the fastigial nucleus, an intermediate (paravermal part of the cerebellar hemisphere) zone projecting to the interposed nuclei and a lateral (lateral part of the cerebellar hemisphere) zone projecting to the dentate nucleus.

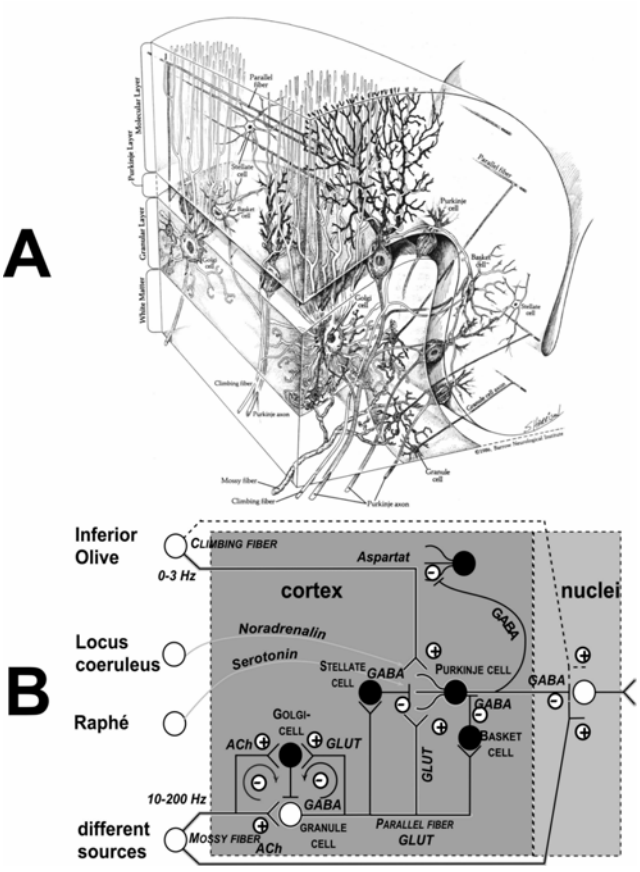


Figure 1. Anatomy. (**A**) Three-dimensional drawing of a folium of the cerebellar cortex. Reprinted with permission from S. Harrison (*BNI Quarterly*, Journal of St. Jospeph's Hospital, Barrow Neurological Institue, Phoenix, AZ, **2**(2), 1986). (**B**) Schematics of the microarchitecture of the cerebellar cortex and nuclei with different elements, including the synaptic connection with "+" for excitatory and "–" for inhibitory synaptic transmission.

The cerebellar cortex is a uniform structure. It is divided into three distinct layers: the molecular layer, the Purkinje cell layer, and the granular layer (Figure 1). The cerebellar cortex contains five types of neurons: (1) Purkinje, (2) granule, (3) Golgi, (4) stellate, and (5) basket cell. The molecular layer is the outermost layer and contains primarily the ascending part of the axons of the granule cells, which bifurcate rectangularly, sending fibers along the direction of the folium, and are thus termed parallel fibers. Moreover, there are also dendrites of the Purkinje and Golgi cells, climbing fibers, and two types of interneurons

(stellate and basket cells). The large pear-shaped cell bodies of Purkinje cells are aligned side by side in a single layer, known as the Purkinje cell layer. The extensive dendritic tree of a Purkinje cell extends into the molecular layer in a single plane. The granular layer is the innermost layer and primarily contains densely packed, small granule cells as well as a few larger interneurons (Golgi cells).

The cerebellar cortex receives afferent input from most parts of the peripheral and central nervous system. The most important afferent fibers consist of mossy fibers and climbing fibers. Climbing fibers originate solely from the inferior olive in the brain stem. Mossy fibers originate from different brain stem nuclei and neurons in the spinal cord. The mossy fiber afferents terminate in the granular cell layer. Mossy fiber contacts with dendrites of granule and Golgi cells are known as cerebellar glomeruli. The mossy fibers alter the activity of the Purkinje cell via the parallel fibers. Each Purkinje cell receives inputs from numerous granule cells, and each granule cell collects inputs from several mossy fibers. Climbing fibers directly contact the dendrites of the Purkinje cell. Each Purkinje cell receives input from a single climbing fiber, and one climbing fiber contacts 1 to 10 Purkinje cells.

The Purkinje cell is the only output neuron of the cerebellar cortex. Purkinje cells have a GABAergic, inhibitory action on the cerebellar nuclear neurons. Both mossy and climbing fiber afferents have an excitatory action on the Purkinje cell. The excitatory input is modulated by inhibitory interneurons (stellate and basket cells in the molecular layer, and Golgi cells in the granular layer).

2. EXPERIMENTAL APPROACHES AND BEHAVIORAL DATA

2.1. Tests for Associative Processes

2.1.1. Eyeblink Conditioning

There are numerous studies that provide clear evidence for critical involvement of cerebellar structures in classical conditioning of the eyeblink reflex. The parts of the cerebellar cortex and cerebellar nuclei involved in eyeblink conditioning have been assessed carefully in animal models. Animal lesion studies, most in the rabbit, indicate that the ipsilateral interposed nucleus and Larsell lobule H VI are of particular importance for acquiring conditioned responses (for a review see (9,71,81)).

In addition to electrophysiological studies, functional brain imaging has been the most widely used technique to study the involvement of structures in behavior of humans. This is discussed in more detail in Part IV, chapter 5, by Breiter, Gasic, and Makris (this volume). From several human lesion and functional brain imaging studies there is evidence that the human cerebellum is also

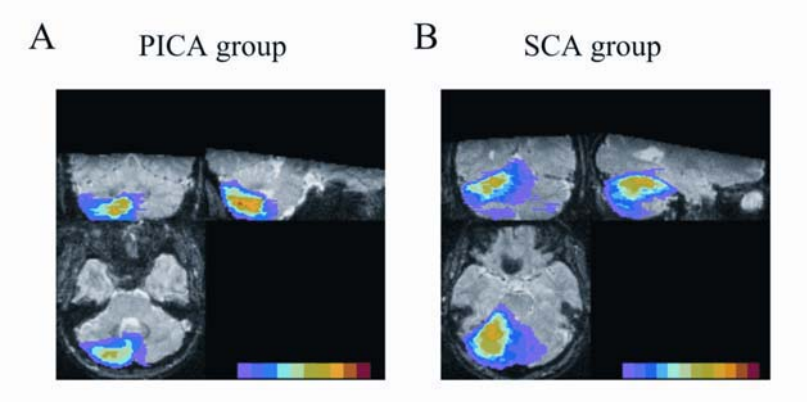


Figure 2. Lesions of PICA (**A**) and SCA (**B**) patients superimposed on axial, sagittal, and coronal stereotaxically normalized MR images of the cerebellum of a healthy 26-year-old female subject. All unilateral lesions are superimposed on the left cerebellum, with right-sided lesions flipped to the left. The number of overlapping lesions is illustrated in color. PICA group: from violet (n = 1) to red (n = 11). Data of two patients are not included for technical reasons. Note the center of overlap (orange, n = 10) in the posterior inferior cerebellum within lobules VIIB and VIIIA. Note that some lesions affected the lower and inferior part of the dentate nucleus (dark blue, n = 4; light blue, n = 5), with the interposed nucleus being preserved. SCA group: from violet (n = 1) to red (n = 12). Data from two patients are not included for technical reasons. Note the center of overlap (light green, n = 9; darker greens, n = 8 and 7) in the superior cerebellum within hemispheric lobules VI and Crus I.

involved in classical eyeblink conditioning. Eyeblink conditioning has been shown to be impaired in patients with cerebellar lesions (10,17,75,80). In addition, studies using positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) revealed learning-related changes of activity in the cerebellum during eyeblink conditioning in healthy human subjects (e.g. (63,58)).

A recent human lesion study conducted by our group investigated classical delay eyeblink conditioning in 27 patients with primarily unilateral lesions, particularly infarcts of the superior cerebellar artery (SCA) and the posterior inferior cerebellar artery (PICA) (22). The extent of the cortical lesion (i.e., which lobules were affected) and possible involvement of the cerebellar nuclei was determined by 3D-magnetic resonance (MR) imaging. Figure 2B shows the lesions of all SCA patients and Figure 2A of all PICA patients superimposed on MR images of the cerebellum of a healthy subject.

The cerebellar areas known to be most critical in eyeblink conditioning based on animal data (i.e., Larsell lobule H VI and interposed nuclei) are commonly supplied by the superior cerebellar artery (4). Therefore, we hypothesized that conditioning of the eyeblink reflex was impaired in patients with lesions

including the territory of the SCA, but preserved in patients with lesions restricted to the territory of the PICA.

Comparing acquisition of classically conditioned eyeblink responses (CR-incidence) between patients and sex- and age-matched controls revealed five main findings. First, the ability to acquire classically conditioned eyeblink responses was reduced in cerebellar patients (Figure 3A,B). Second, in patients with unilateral cerebellar lesions conditioning deficits were present ipsilaterally (Figure 3B). These two results strengthen findings described in the previous human literature in a larger patient sample with more accurate MR-based description of the cerebellar lesion. The three other results provide evidence that some additional findings described in the previous animal literature are transferable to humans. First, deficits of eyeblink conditioning were most prominent in patients with lesions of the superior cerebellum including hemispheric lobule VI and/or Crus I. Mean total CR-incidences were most clearly reduced on the affected side compared to the unaffected side in cerebellar patients with lesions including the territory of the SCA (affected side = 13.0%, SD 6.31; unaffected side = 33.9%, SD 17.7; see filled and open bars in Figure 3D). In cerebellar patients with lesions restricted to the territory of the PICA, the differences in CR-incidences comparing the affected and unaffected side were less pronounced (mean total percentage CR-incidence affected side = 19.0%, SD 14.3; unaffected side = 27.6%, SD 25.7; Figure 3C). Finally, eyeblink conditioning deficits were not significantly different in patients with pure cortical lesions compared to patients with additional nuclear impairment nor in patients with unilateral and bilateral lesions. In brief, data indicated that a unilateral cortical lesion within the superior cerebellum was sufficient to significantly reduce eyeblink conditioning in humans. Similar to findings in animal models, these areas overlap with regions involved in unconditioned eyeblink control in humans. A recent fMRI study conducted by our group showed that areas within ipsilateral lobules Crus I and VI are most active during evocation of the unconditioned eyeblink in healthy human subjects (18).

In sum, involvement of the human cerebellum in eyeblink conditioning is a robust finding in various human lesion and functional brain imaging studies despite differences in the cerebellar patients' pathology and differences in the experimental design. Comparable to the results from eyeblink conditioning studies, the classically conditioned lower limb withdrawal reflex requires an intact cerebellum as well, as has been shown for humans (73) and for animals (40).

2.1.2. Visuomotor Associative Learning

Involvement of the human cerebellum in associative learning does not seem restricted to conditioning of avoidance reactions. Cerebellar patients have been shown to be impaired in visual motor associative tasks, which required the link-

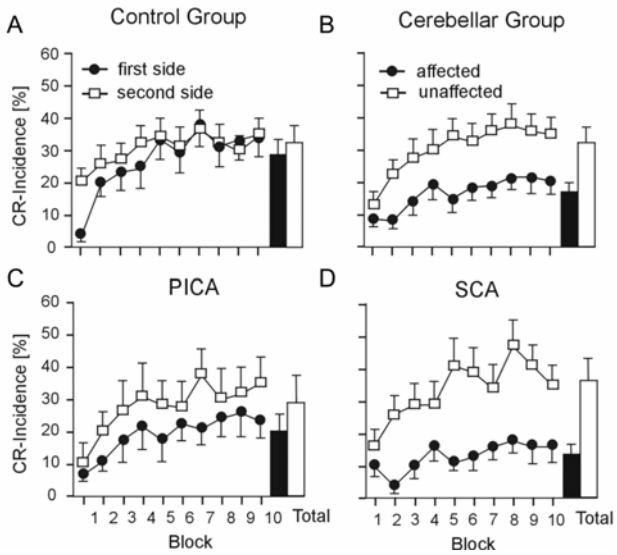


Figure 3. Eyeblink conditioning. Mean percentage CR-incidences ± SE for each of 10 blocks and across all blocks (total) in a group of sex- and age-matched controls (**A**) in relation to a group of cerebellar patients (**B**), the latter separated in patients with lesions restricted to the PICA territory (**C**) and patients with lesions including the SCA territory (**D**). Filled squares and columns represent the eye tested first in controls and the ipsilesional eye in patients. Open squares and columns represent the eye tested second in controls and the contralesional eye in cerebellar patients. (Block = 10 tone–airpuff trials; Total = mean total percentage CR-incidence). Adapted with permission from Fig. 4 in (22).

age of a visual stimulus and motor response (15,76). Involvement of the human cerebellum in procedural learning of a visuomotor sequence has been found by others (51,23).

In a study by our group (19), a visuomotor associative learning task was examined in patients with pure cerebellar disease. Patients were matched to controls with respect to age, education, IQ, and visual memory. A simple reaction time and visual scanning task were performed to assess motor background variables. The patients' ability to learn the association of a color and a numeral was significantly impaired (Figure 4) regardless of the amount of motor performance deficits. A reasonable assumption was that the cerebellum is involved in this kind of associative learning.

In a subsequent study (74), we were able to show that visuomotor associative learning deficits were not due to increased attentional demands in cerebellar patients. It may be that the execution of the motor component of a task (to push a button) took up more attentional resources as compared to controls, which in turn reduced the resources for the cognitive components of the task.

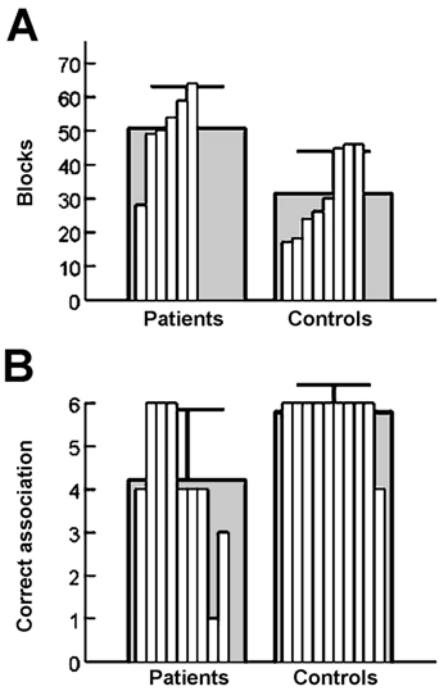


Figure 4. Comparison of the cerebellar patients and control groups in the visuomotor associative learning task. Means and standard deviations of (**A**) the number of blocks required to reach the learning criterion and (**B**) the number of correct associations between numerals and colors produced after the learning task was concluded. The effects of both variables are significant ($p < 0.05$). Each small bar represents the value of one subject. Adapted with permission from Fig. 3 in Drepper et al. (19).

Subjects had to learn the association between pairs of color squares. To change the motor demands of the task, subjects had to press a target button once or three times (one vs. three key presses). If motor execution of the task affected the ability to learn the association of color pairs, deficits in associative learning should be more pronounced in the more difficult motor condition.

Association of the color pairs enabled subjects to predict the correct side of a motor response. Cerebellar patients were less able than controls to learn the association between pairs of colors and, therefore, less able to reduce reaction times based on predictive knowledge of the side of the response.

Cerebellar patients, however, were able to reduce decision times as much as controls in control conditions that did not require associative learning. Therefore, deficits in cerebellar patients' ability to reduce decision times in the associative learning part were not due to a general inability to further reduce decision times based on motor performance deficits.

Moreover, motor performance deficits were not related to deficits in associative learning. First, there was no significant interaction between motor demands (one vs. three key presses) and the effects of learning. Second, the effects of impaired visuomotor learning in the cerebellar group were most prominent when reaction times were normalized for motor performance deficits. Finally, the findings of our previous study were confirmed: background motor variables—i.e., simple reaction times, visual scanning times, and clinical ataxia scores—did not relate to impaired associative learning.

In monkey cerebellar lesions studies, however, Passingham and his group found that visuomotor associative learning was not impaired when controlling for motor performance deficits (53,54). Likewise, a recent fMRI study in healthy human subjects showed that cerebellar activation during learning a visuomotor sequence was due to motor performance but not learning itself (66). Our studies required the association of two visual stimuli and the linkage of the correct color pair and motor response. In Nixon and Passingham's (53,54) and Seidler's (66) studies, discrete associations of one visual stimulus and a motor response were learned. The role of the cerebellum may be different in stimulus–stimulus–response and stimulus–response associations.

In sum, the findings of the two studies by our group suggest that the cerebellum is not only involved in conditional learning of avoidance reactions but also in visuomotor associative learning tasks. Further experiments are needed to differentiate between the role of the cerebellum in sequencing incoming visual stimuli and its role in helping to build an association between them and/or the motor response.

2.2. Tests for Non-Associative Processes

2.2.1. Habituation during Startle

Both animal and human lesion data suggest that the cerebellum is involved in habituation of unspecific aversive reaction, i.e., the acoustic startle response. The acoustic startle response is a protective behavioral reaction consisting of muscle contractions of the eyelid, the neck, and the extremities that is elicited by sudden loud acoustic stimuli. Lesion studies in the rat have shown an involvement of the cerebellar vermis in long-term habituation of the acoustic startle response, but not in short-term habituation (45–47). Similar findings were observed in a human lesion study of our group (49). Patients with midline cerebel-

lar lesions due to surgery were studied. Subjects received 40 acoustic startle stimuli daily for five successive days. Data were analyzed for response decrement within the training session of one day (short-term habituation) and for a decrease in the startle response across the five training days (long-term habituation). Long-term habituation of the blink component of the acoustic startle response recorded at the orbicularis oculi muscles was significantly impaired in patients with cerebellar lesions compared with control subjects, whereas short-term habituation was preserved in both groups. Findings of involvement of the human cerebellum in habituation of the startle response are further supported by PET studies in healthy human subjects (72,57). These findings of impaired startle habituation provide evidence that cerebellum is involved in non-associative learning.

2.3. Tests for Mixed, Associative and Non-Associative Processes

2.3.1. Associative and Non-Associative Processes in Postural Reflexes

The eyeblink or the lower limb withdrawal reflex represents protective avoidance reactions. An important driving force for an organism to establish conditioned responses (CRs) preceding the unconditioned stimulus (US) is thus to increase protection against the harmful US. This is in fact accomplished in classically conditioned eyeblink experiments with an air puff as US but is in contrast to the methodological-conceptual issues suggested by Gormezano and Kehoe (25), who claimed that subjects should not profit from the CR. Postural reflexes represent compensatory responses of an organism to unexpected external perturbations of the body equilibrium. In contrast to nociceptive reflexes, postural reflexes must be classified as protective and non-nociceptive and, thus, must be based on a driving force different from avoidance. To analyze these reflexes Nashner introduced a method using a dynamic platform that allows reproducible translatory or rotational perturbations (e.g. (52)). Unexpected movements of a tilting platform (toes-up) evokes a passive deviation of the body with characteristic changes in the activation of corresponding leg and/or trunk muscles. Following repetitive perturbations, the reflex patterns may alter minimally only, whereas the amplitude of the responses may change. Different plastic processes have been discussed to be involved. Functional habituation of postural responses induced by "toes-up" rotations has been studied in detail by Hansen et al. (26), who reported for the gastrocnemius muscle (GA) an initial reduction due to habituation of a startle-like response and a subsequent more gradual decay (26). "Habituation" or "adaptation" of automatic postural responses is a functional mechanism that allows subjects to minimize energy expenditure (36). Such decays or attenuations of responses are based presumably on non-associative processes. Classical conditioning of postural reflexes, however, re-

quires an underlying associative process, and there is some evidence that the cerebellum is involved comparably, as in classical conditioning of eyeblink (22) and lower limb withdrawal reflexes (42,43).

We studied classical conditioning of postural reflexes in a group of young and healthy subjects (42) and in a group of 8 cerebellar patients suffering from diffuse degeneration disorders who were compared with a group of sex- and age-matched controls (43). Subjects were standing on a tilting platform allowing rotational movements, representing the US, with the rotation axis through the ankle. In paired trials a preceding auditory signal was given as CS. Subjects were tested in 70 US-alone trials, in 80 trials with paired stimuli, followed by a brief section of 20 US-alone trials. EMG signals were recorded from the main muscle groups of the leg (tibial anterior muscle (TA), GA, rectus femoris muscle, and biceps femoris muscle). In a study on a group of young and healthy subjects (42), it turned out that 22% of the subjects established CR in all main groups of the leg muscles tested and thus responded to the unexpected perturbations similarly as in experiments on classical conditioning of the eyeblink or withdrawal reflex (73). The preponderant proportion of the subjects, however, established CR in the GA only but developed a substantial decay of the UR amplitude in the TA, which was much larger compared with that observed during US alone trials (Figure 5B). Subjects with CR in ***all*** muscles (strategy-I subjects) were assumed to apply a different strategy than those with CR in the GA only (strategy-II subjects). In cerebellar patients, however, no such strategy was observed. Stacks of typical TA responses recorded during US-alone and paired trials obtained from a healthy strategy-I subject, from a strategy-II subject and from a cerebellar patient are shown in Figure 5. Patients hardly established any CR, and if they did show responses within the CS-US window, the responses were not time-locked. Consequently, the resulting CR-incidence did not follow a characteristic exponential learning curve.

The platform was equipped with strain gauges to record vertical forces exerted by the subjects. From these forces the center of vertical pressure (CVP) was calculated (43). The CVP represents the final outcome of all muscles involved during a given trial with excursions primarily in the sagittal plane. Figure 6 shows the deviations of the CVP in the sagittal and frontal planes for the controls and the group of cerebellar patients. Consequently, if a subject established a muscle-related CR there was also a deviation in the CVP prior to the US, which has to be attributed to a CVP-related CR. This deviation was to the rear and was small compared with the deviation to the front due to the perturbation representing the UR. In contrast to the generally accepted rules for classical conditioning (25), the deviation of this CVP-based CR is in the opposite direction to the UR and can be interpreted as a preceding, preparatory, and compensatory shift of the CVP. As mentioned above, in cerebellar patients no muscle activity-based-CR has been observed. Although we recorded from all main muscle groups of the leg, there is the general problem of missing muscles involved

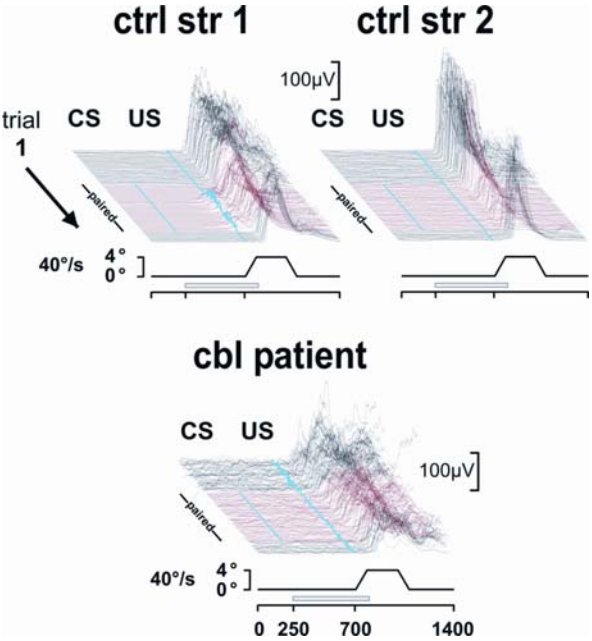


Figure 5. Postural responses from the tibial anterior muscle elicited by a tilting platform (US) obtained from two control subjects exhibiting strategy I and II, and a cerebellar patient without presenting one of these strategies. Stack plots with the first (1) trial on the top and the last trial at the bottom of each stack plot. Stack plots are constructed from 70 US-alone trials, 80 CS-US trials, and 20 US-alone trials. Analysis time is 1400 ms. The conditioning stimulus (CS) and unconditioned stimulus (US) are marked by cursors. The movement function with its characteristic parameters, representing the US, the onset and duration of CS, marked by a shaded bar, are below the stack plots. Note that during paired trials the subject using strategy I established CR whereas the subject using strategy II did not but exhibited remarkable decay of the UR. The cerebellar patient did not profit from the occurrence of the CS.

in the motor task studied. On the other hand, in cerebellar patients we did not find a CVP-based CR, which thus coincided with a lack of the corresponding muscle activity. During conditioning controls reduced their deviations in both directions, whereas that of the patients remained unchanged.

In the initial ***US-alone*** trials the decay of UR amplitudes is most likely due to a habituation process. The mean UR amplitude at the end of the US-alone trials was 79.1 ± 10.9% for the controls and 84.4 ± 5.9% for the patients with respect to the mean amplitude (set to 100%) of the first 10 trials at the beginning of the session. Although the mean values just missed significance, there was a clear tendency to a smaller decay in cerebellar patients. During paired trials the UR attenuations were larger in both groups, but the final difference was signifi-

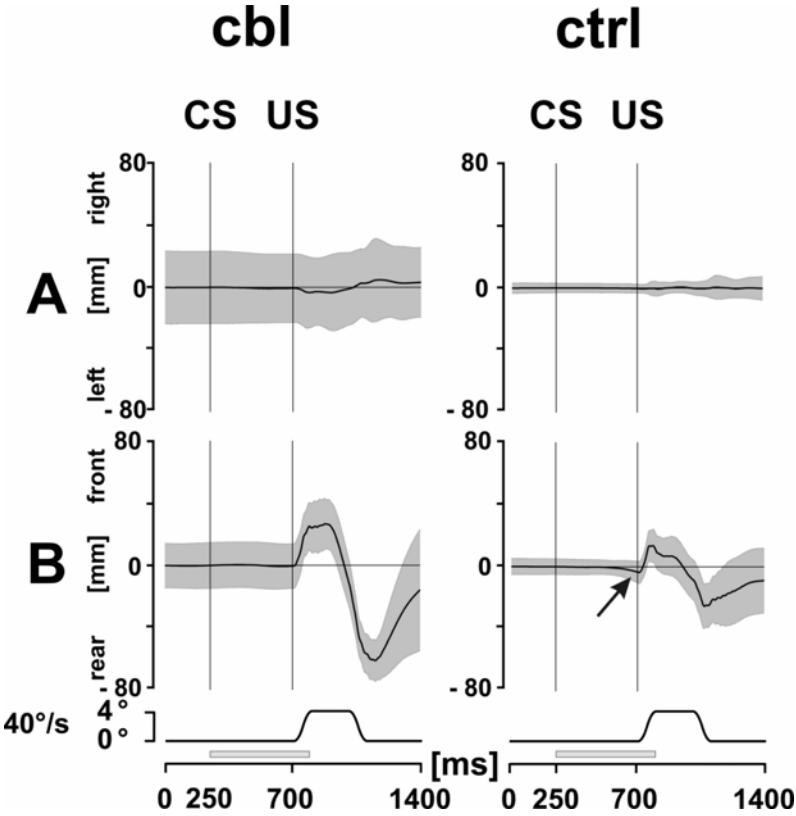


Figure 6. Spatiotemporal components of the trajectories of the center of vertical pressure of the cerebellar (left column) and the control group (right column) are shown. (**A**) *x*-components (right-left), (**B**) y-components (front-rear), including the standard deviations (shaded areas) of the trajectories. The arrow within the CS–US window indicates the conditioned response in the control group. Note that the standard deviation is much higher in the cerebellar patients for both directions.

cantly larger in the controls (final level in controls = 14.5 ± 3.4%, patients = 68.8 ± 16.9%; (43)). Although less pronounced, a similar result has been found in a previous study on the classically conditioned lower leg withdrawal reflex (41). During paired trials the final mean UR amplitude was 48.9 ± 19.3% for the cerebellar patients, 70.0 ± 5.9% for the group of cerebellar patients showing extracerebellar symptoms, and 42.8 ± 17.6% for the controls (41).

As mentioned above, the plastic process responsible for decay during US-alone trials is of a non-associative type and is most likely due to habituation. During paired trials an additional associative plastic process must be assumed,

interacting with or superimposing onto the habituation. Since cerebellar patients did not benefit from the occurrence of the CS, and thus hardly established CR either in the muscles or in the CVP, and since UR decay was smaller in US-alone trials and particularly in paired trials, the cerebellum is assumed to be critically involved in both associative and non-associative processes.

3. THEORETICAL APPROACHES

3.1. Hypotheses and Considerations on Cerebellar Functions

Our ideas on cerebellar function are based on our early experimental findings in well-trained behaving monkeys with characteristic Purkinje cell activity during active (pursuit) and passive hand movements (7). The complex spike activity representing climbing fiber activity, recorded from a Purkinje cell in the intermediate part of the cerebellum, responds most sensitively to the beginning of passive hand movements, whereas during active moments the same unit stops any increased complex spike activity preceding the movement at onset. This canceling of the feedback signal from the periphery seen during passive movements led to an assumption that the inferior olive, the origin of climbing fibers, could be one site at which information about "intended" (efference copy) and "ongoing" (reafference signal) movements are compared and nullified if both signals are equal according to the "Reafferenz" hypothesis (28). If both signals are not equal, the olivary cell would inform the cerebellum about an error in the performance of the movement (38). This coincided well with a comparator hypothesis of the inferior olive suggested by Miller and Oscarrson (50). The concept of the climbing fiber system as an "event marker" (61) seemed to us to be too restrictive since this system is able to transmit precise information about parameters up to the third derivative of passive hand movements (39). An error-detecting system may receive, while an error-correcting system (55) ***must*** receive all this detailed movement-related information. This will not exclude other preolivary (peripheral, or more central) structures involved in error detecting. Although there are numerous hypotheses on how error or control error signals may look, or how they are compiled (31), the common denominator is that the inferior olive, and thus the climbing fiber system, is involved in handling errors occurring during voluntary or reflexive movements.

The concept of interaction between the descending motor commands and feedback information ascending from the periphery at the level of the intermediate part of the cerebellum coincides well with one part of Allen's and Tsukahara's cerebral–cerebellum cooperative hypothesis (3). The intermediate cerebellum is assumed to update ongoing movement on the basis of sensory feedback information, whereas the lateral cerebellum, receiving information from association areas and (almost) none from the periphery, and linked within

the loop via specific thalamic nuclei to premotor areas, is thought to participate in planning the movement or in establishing a motor program that has to be seen as a set of muscle commands, with its sequence planned prior to the beginning of the movement without interaction with the periphery.

In the context of our current experimental approaches to motor learning, there is a commonly accepted concept known as the Brindley-Marr-Albus-Ito model. This actually comprises four distinct hypotheses, and we will focus on these only. Brindley (13) was the first to suggest that the cerebellum as a site of motor learning and a principal agent in the learning of motor skills. Marr (48) proposed a detailed theory in which the cerebellum learns motor skill (action) and learns to maintain posture and balance (reflexes). Accordingly, the Purkinje cell learns to execute elemental movements due to the activity patterns provided by parallel fiber synapses contacting the Purkinje cell dendrites at a given time. The instruction arises in the cerebral cortex, and is conveyed to inferior olive neurons, which in turn activate the corresponding Purkinje cell via the climbing fibers, and thus determine the contexts given by the parallel fiber activity pattern. After the Purkinje cell has learned to recognize the corresponding context, the context alone is sufficient to activate the Purkinje cell, causing the next elemental movement. The crucial step in learning is assumed to take place at the only modifiable parallel fiber/Purkinje cell synapse, due to a facilitation of the simultaneous firing of climbing fibers and presynaptic parallel fiber activity, similar to a Hebbian synapse (27). In the case of eyeblink conditioning, Thompson and colleagues (70) proposed that CS-information (tone) reaches the cerebellar cortex and nuclei via the mossy fibers, and US-information (air puff) via the climbing fibers. The cerebellar cortex and the cerebellar nuclei are assumed to be possible sites of association and plasticity in this type of learning with the interposed nuclei of particular importance.

Albus' theory (2) is similar to, but extends, that of Marr (48). As Marr, he assumes the mossy fiber–granule cell–parallel fiber system to be a pattern recognition data processing system; however, as more like the classical Perceptron (60), performing a fan-out operation facilitating pattern discrimination and learning speed. Moreover, the modifiability of synapses is not restricted but includes synapses of stellate and basket cells and, in contrast to Marr, with the pattern storage accomplished by weakening synaptic weights. The learning process is assumed to be associative, as in classical conditioning. The post-climbing pause (or inactivation pause) in the Purkinje cell is regarded as an unconditioned response to the unconditioned stimulus represented by a climbing fiber activation. The conditioning stimulus is the current parallel fiber activity at the time of climbing fiber activation. After learning, the conditioning stimulus alone should be able to evoke an inactivation pause, similar to that evoked previously by climbing fibers. Such a pause should thus be accomplished by a weakened parallel fiber contact rather than by a strengthened one (2).

Ito provided the first experimental support for the Marr-Albus model by showing that conjunctive local stimulation of parallel fibers and climbing fibers results in a long-lasting depression of parallel fiber–Purkinje cell transmission (30). The results were obtained from a field potential analysis providing a corresponding significant depression of the n_2 potential lasting for more than 1 hr. These experimental data provided the basis for plasticity in the cerebellum.

3.2. Models and Simulations of Selected Cerebellar Functions

With a simple non-adaptive functional model, including adjustable vectors of weights and nonlinear elements such as thresholds, we were able to simulate experimentally recorded activity patterns in cerebellar elements of the cat observed during passive movements (37).

The plan of the model is based on three different levels of computation (Figure 7A). At the receptor level (Figure 7A1), ideal linear receptors are assumed to encode the stimulus $\phi(t)$ function (passive movement) as well as their first- ($d\phi/dt$) and second-derivative functions ($d^2\phi/dt^2$), either directly (Figure 7A, left column) or reciprocally (Figure 7A, medial column). Specific rectifiers split positive (pos = positive half-wave rectifier) from negative (neg = negative half-wave rectifier) velocities or accelerations. At the second, ***multiplicative level*** (Figure 7A2), two types of patterns are obtained, one of which, termed the ***ON*** response, generated by the product of ***low***-threshold signals transmitting $\phi(t)$ and ***low***-threshold signals transmitting ($d\phi/dt$) and the other, the ***OFF*** response, generated by the product of ***high***-threshold signals transmitting $\phi(t)$ and ***high***-threshold inhibiting signals transmitting ($d\phi/dt$). In view of the positive or negative velocities, $ON\oplus$ and $ON\ominus$ or $OFF\oplus$ and $OFF\ominus$ patterns can be obtained for conceivable synaptic connections (Figure 7A1, third column). Omitting weighing factors, an $ON\oplus(t)$ pattern is obtained by

$$ON\oplus(t) = \text{pos}(d\phi/dt) * \phi(t), \qquad [1]$$

and an $OFF\oplus$ pattern by

$$OFF\oplus(t) = \text{neg}(d\phi/dt) * \phi(t). \qquad [2]$$

Assuming a linear, ramp-shaped function for $\phi(t)$, its constant positive velocity pos($d\phi/dt$) acts as a gate such that only in the presence of the positive velocity the ramp shape function is transmitted, resulting in an $ON\oplus$ pattern. Correspondingly, in $OFF\oplus$ patterns the linear, ramp-shaped function $\phi(t)$ is conveyed completely except during the period when a negative velocity neg($d\phi/dt$) is present. At the third, ***additive level*** (Figure 7A3), the sums of the different *ON* and *OFF* patterns are obtained. Sums of $ON\oplus$ and $OFF\ominus$ patterns are characteristic for

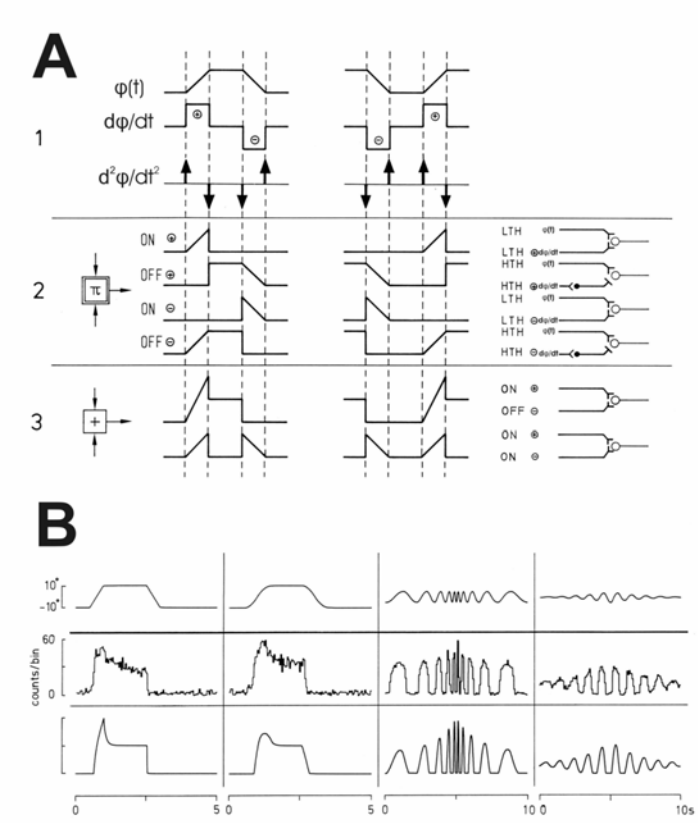


Figure 7. (**A**) Basic plan of a simple non-adaptive functional feedback model (37). The stimulus function is a linear, ramp-shaped passive movement of the paw around the wrist joint. Different types of ideal linear receptors encode at a first (1) level of computation the position $\phi(t)$, the velocity $d\phi/dt$, and the acceleration $d^2\phi/dt^2$ of the movement, either directly (left column) or reciprocally (middle column). At the second (2) multiplicative level of computation the product of low-threshold (LTH) signals transmitting $\phi(t)$ and LTH-signals transmitting $d\phi/dt$ form *ON*⊕ or *ON*⊖ patterns, depending on the positive or negative velocity. Analogously, high-threshold (HTH) signals form corresponding *OFF* patterns. At the third (3) additive level, different *ON*⊕ and *OFF*⊖ patterns are summed. Conceivable synaptic connections are shown in the right column. (**B**) Physiological responses (histograms of extracellularly recorded spikes of a cerebellar granule cell in counts per bin, middle row) and results from the computer simulation of the model (bottom row) to different stimulus functions (top row). Stimulus function in the first column: linear ramp-shaped function, starting from a low holding position of –10°, proceeding with a constant velocity of 40°/s to a high holding position of 10°, with 0° as the horizontal plane. Second column: ramp of identical amplitude but of quadratic increase and decrease of the position. Sinusoidal movement function of constant amplitude with logarithmically increasing and decreasing frequency (third column) and that of damped oscillation at a constant frequency (fourth column). Modified with permission from (36).

patterns obtained during extracellular recordings from cerebellar elements. For the computer simulation of the model, acceleration signals were included as well as amplitude limiters and dead time elements. The responses of such a model are

shown in Figure 7B, with different input functions in the top row, histograms of extracellularly recorded responses of a single cerebellar granule cell in the middle row, and simulated responses in the lower row. For an expansion to an *n*-dimensional system (representing *n* input receptors and afferent fibers), specific matrix operations are required (37).

Most modern models, however, are adaptive and in their simplest version so-called linear combiners. They utilize error signals originating from the current output signal and the desired response to adjust weight vectors by iteratively minimizing the error using a least-mean-square criterion (e.g. (78)). Such an adaptive control model of the cerebellum, described as a hierarchical neural network model, is that of Kawato et al. (33). The original model was designed for control and learning of voluntary movements, and consists of three parts: (1) the main descending motor pathway from the motor cortex and the somatosensory feedback pathway to the motor cortex; (2) the spino-cerebellum-magnocellular nucleus ruber system as an internal model of the ***forward*** dynamics of the musculoskeletal system; and (3) the cerebro-cerebellum-parvocellular red nucleus system as an internal model of ***inverse*** dynamics of the musculoskeletal system with synaptic plasticity included in the cerebellar cortex. In this hierarchical network the association cortex sends the desired movement pattern to the motor cortex. The actual movement pattern is measured by proprioceptors and returned to the motor cortex. Feedback control can thus be achieved using the error in the movement trajectory. The cerebro-cerebellum and the parvocellular part of the nucleus ruber system receive information from wide areas of the cortex, but not from the periphery, and thus monitor the desired trajectory and the motor command. For ***inverse*** dynamics, the input and output of the (nonlinear) system are inverted, resulting in an input being the desired trajectory and an output being the motor command. Once the inverse dynamics model is established by iterative motor learning, it can compute an appropriate motor command directly from the desired trajectory.

The more recent model of feedback error learning in conjunction with acquisition by the cerebellum of the inverse of a controlled limb (or, generally, controlled object) is based on (1) the cerebrocerebellum, (2) the intermediate part of the cerebellum including the vermis, and (3) the flocculus (24,34). Their feedback-error-learning approach (Figs. 1C and 3 in (34)) is discussed in more detail here. The controlled object, e.g., a limb, produces a realized trajectory Θ after receiving the motor command $\tau(t)$:

$$\tau(t) = \tau_c(t) + \tau_n(t), \qquad [3]$$

which results from the sum of the feedback motor command $\tau_c(t)$ and the feedforward command $\tau_n(t)$, the latter generated by an inverse model. The inverse model receives the desired trajectory Θ_d (e.g., from the association cortex) and

monitors the feedback torque $\tau_c(t)$ for the error signal. τ_n is calculated from the desired trajectory Θ_d and the synaptic weights ω by

$$\tau_n = \Phi\,(d^2\Theta_d/dt^2,\ d\Theta_d/dt,\ \Theta_d, \omega). \qquad [4]$$

The inverse model is trained during motor control using a feedback motor command as the error signal. The feedback controller transforms the trajectory error into the motor-command error. The shape of the *m*-dimensional vector function Φ depends on the type of neural network that actually constitutes the feedforward controller. The vectors τ, τ_c and τ_n are *m*-dimensional, the vectors Θ_d and Θ are *n*-dimensional, and ω is an *l*-dimensional vector. The synaptic modification rule providing the plasticity of the feedback-error-learning scheme is represented in a general form as

$$d\omega/dt = (\partial \tau_n/\partial \omega)^{\tau}{}_{\tau_c}. \qquad [5]$$

Aside from the theoretical considerations, the authors (34) relate their components to neural circuits, as proposed in a review by Allen and Tsukahara (3). For the limb as the controlled object the association cortex provides the desired trajectory Θ_d, which is sent via pontine nuclei to the lateral cerebellum and simultaneously to the motor cortex. Via transcortical loops the motor cortex is informed about the realized trajectory Θ. From the difference $\Theta_d - \Theta$, motor cortex neurons act as feedback controllers and calculate the negative feedback command u_{fb} (corresponding to $\tau_c(t)$ in the model), which is sent to motor cortex output neurons. A copy of this signal is sent as an error signal to inferior olive neurons, evoking complex spikes in cerebellar Purkinje cells. The output of the corticonuclear complex is the feedforward motor command u_{ff} (corresponding to $\tau_n(t)$ in the model), which is sent via the dentate nucleus and the thalamus to the motor cortex. Summation of u_{ff} and u_{fb} is performed in the motor cortex, forming the final motor command u descending the corticospinal tract.

The corresponding assumption with the climbing fiber response as the error signal coincides well with suggestions derived from our experimental observations (7,38). An inverse model, inverting the inputs and outputs of the controlled object, results in an ideal feedforward controller also able to perform transformation of coordinates. The necessity for such a transformation was suggested by Pellionisz and Llinas (56), who provided corresponding algorithms. The model assumes that sensorimotor learning and transformation of coordinates occur initially in the cerebral cortex. The resulting movement is clumsy and produces a motor command error that is sent by the climbing fiber activity to the cerebellum, where procedural learning then occurs. Such a model allows adaptive modification of the vestibulo-ocular reflex, the adaptive control for posture and locomotion and learning control for voluntary movements (34).

As mentioned above, a typical paradigm for analyzing specific aspects of motor learning is the method of classical conditioning of the eyeblink or withdrawal reflexes. For determining the corresponding area within the cerebellum, a transient inactivation of the cerebellar nuclei was performed in well-trained animals by injection of the $GABA_A$ agonist muscimol (e.g., in the rabbit (11) and in the cat (40)). The feedback error cerebellar learning model of Kawato (35) simulates such a situation as follows: the signal from the cerebral cortex is conveyed to a summing point projecting to spinal motor centers and to the cerebellum including an inverse model, the output of which represents the feedforward motor command in motor coordinates. In a "well-trained" situation, the resulting trajectory is equal to the desired trajectory such that the output of the feedback controller from the spinal motor centers representing the error signal is typically zero. After injection with the "cerebellar nuclei blocked," the output of the cerebellum is zero and the motor command generated by the cerebral cortex is the only signal at the summing point. During recovery of the cerebellum, the contribution of the motor cortex progressively attenuates. Although this simulation by Kawato (35) does not include cerebellar pathology, such as ataxia, it shows automatic substitution of the cerebral cortex when cerebellar functions are disabled.

An update of these models is summarized by Wolpert et al. (79). In their cerebellar-feedback-error-learning model (an inverse internal model), the trajectory error trains the internal modal such that the actual trajectory (motor output) finally will be fairly close to the desired trajectory. Further, the cerebellum also generates a forward representation of the motor apparatus, known as a forward model. This allows for simplification: the inputs of such a forward model are the current state of, e.g., the arm (processed reafference signal) and the efference copy of the motor command producing an estimate of the new state of the arm. Such a predictor system has to take into account corresponding "transport delays," which may be long with respect to movement duration. For the motor learning aspect, a computational model is provided that includes multiple ***paired*** forward and inverse modules. The inverse model generates the motor command representing the desired trajectory, whereas a single forward model predicts the consequence of a performed action and can thus be used to estimate responsibility. This, however, can be achieved earlier when a movement has been initiated within corresponding sensory contextual cues and the result of the action is known (79).

A general model of the cerebellum presented by Arbib et al. (5) can simulate results from prism adaptation in control subjects and patients during dart-throwing (68). The model includes a movement pattern generator, the concept of modifiable synapses (29), and, again, an assumption that the inferior olive conveys the error signal via climbing fibers. The simulation results coincide well with experimental results in humans (5). The same group presented a model in

adaptive control of saccadic gain, once again with the inferior olive as an error detector (64,65).

Summarizing, modern adaptive models and their simulation are based on different approaches with respect to the different assumed functions of the cerebellum. Successful simulations are presented using different kinds of olivary error handling as a common denominator and providing results that are in good agreement with behavioral data.

4. RELEVANCE FOR PATIENTS AND THERAPY

Adaptation and the general aspects of motor learning in complex nervous systems, as in humans, are important aspects of currently applied physiotherapy. However, since an understanding of the physiology, and particularly of the pathophysiological basics, of, say, ataxia, is still very limited, so that adequate pathology-related therapy is unfortunately restricted. Results using different experimental approaches and concepts derived from theoretical considerations support the importance of processing error signals in complex systems. There is good evidence that the olivo-cerebellar system is critically involved in both, in online corrections, and in longer-lasting learning processes. Understanding these complex processes will hopefully help to overcome the current limitations in physiotherapy.

5. ACKNOWLEDGMENTS

We would like to thank Dr. J. Davis for helpful and critical comments. This work was supported by grants from Wilhelm Sander-Stiftung (94.090.1-3) and from the Deutsche Forschungsgemeinschaft (DFG Ko 903/1-1; DFG Ti 239/2-1,2,3).

6. REFERENCES

1. Adams JA. 1987. Historical review and appraisal of research on the learning retention, and transfer of human motor skills. *Psychol Bull* 101(1):41–74.
2. Albus JS. 1971. A theory of cerebellar function. *Math Biosci* 10:25–61.
3. Allen GI, Tsukahara N. 1974. Cerebrocerebellar communication systems. *Physiol Rev* **54**:957–1006.
4. Amarenco P, Hauw JJ, Caplan LR. 1993. Cerebellar infarctions. In *Handbook of cerebellar diseases*, pp. 251–290. Ed. R Lechtenberg. Marcel Dekker, New York.
5. Arbib MA, Schweighofer N, Thach WT. 1995. Modeling of the cerebellum: from adaptation to coordination. In *Motor control and sensory motor integration: issues and directions*, pp. 11–36. Ed. DJ Glencross, JP Pick. Elsevier, Amsterdam.
6. Barlow JS. 2002. *The cerebellum and adaptive control*. Cambridge UP, Cambridge.

7. Bauswein E, Kolb FP, Leimbeck B, Rubia FJ. 1983. Simple and complex spike activity of cerebellar Purkinje cells during active and passive movements in the awake monkey. *J Physiol* **339**:379–394.
8. Bloedel JR, Bracha V, Kelly TM, Wu JZ. 1991. Substrates for motor learning: does the cerebellum do it all? In *Activity-driven CNS changes in learning and development*, pp. 305–318. Ed. JR Wolpaw, JT Schmidt, TM Vaughan. New York Academy of Sciences, New York.
9. Bloedel JR, Bracha V. 1995. On the cerebellum, cutaneomuscular reflexes, movement control and the elusive engrams of memory. *Behav Brain Res* **68**:1–44.
10. Bracha V, Zhao L, Irwin KB, Bloedel JR. 2000. The human cerebellum and associative learning: dissociation between the acquisition, retention and extinction of conditioned eyeblinks. *Brain Res* **860**:87–94.
11. Bracha V, Webster ML, Winters NK, Irwin KB, Bloedel JR. 1994. Effects of muscimol inactivation of the cerebellar interposed-dentate nuclear complex on the performance of the nictitating membrane response in the rabbit. *Exp Brain Res* **100**:453–468.
12. Braitenberg V, Atwood RP. 1958. Morphological observations on the cerebellar cortex. *J Comp Neurol* **109**:1–33.
13. Brindley GS. 1964. The use made by the cerebellum of the information that it receives from sense organs. *Ibro Bull* 3:80.
14. Brodal P. 1998. *The central nervous system: structure and function*, 2nd ed. Oxford UP, New York.
15. Canavan AG, Sprengelmeyer R, Diener HC, Homberg V. 1994. Conditional associative learning is impaired in cerebellar disease in humans. *Behav Neurosci* **108**:1–11.
16. Chambers WW, Sprague JM. 1955. Functional localization in the cerebellum, I: organization in longitudinal cortico-nuclear zones and their contribution to the control of posture, both extrapyramidal and pyramidal. *J Comp Neurol* **103**:105–129.
17. Daum I, Schugens MM, Ackermann H, Lutzenberger W, Dichgans J, Birbaumer N. 1993. Classical conditioning after cerebellar lesions in humans. *Behav Neurosci* **107**:748–756.
18. Dimitrova A, Weber J, Maschke M, Elles H-G, Kolb FP, Forsting M, Diener HC, Timmann D. 2002. Cerebellar and brainstem areas controlling human eyeblink responses as revealed by fMRI. *Hum Brain Mapping* **17**:100–115.
19. Drepper J, Timmann D, Kolb FP, Diener HC. 1999. Non-motor associative learning in patients with isolated degenerative cerebellar disease. *Brain* **122**:87–97.
20. Dudai Y. 2002. *Memory from A to Z: Keywords, concepts and beyond*. Oxford UP, Oxford.
21. Eccles JC, Ito M, Szentagothai J. 1967. *The cerebellum as a neuronal machine*. Springer, Berlin.
22. Gerwig M, Dimitrova A, Kolb FP, Maschke M, Brol B, Kunnel A, Böring D, Thilmann AF, Forsting M, Diener HC, Timmann D. 2003. Comparison of eyeblink conditioning in patients with superior and posterior inferior cerebellar lesions. *Brain* **126**:71–94.
23. Gomez-Beldarrain M, Garcia-Monco JC, Rubio B, Pascual-Leone A. 1998. Effect of focal cerebellar lesions on procedural learning in the serial reaction time task. *Exp Brain Res* **120**:25–30.
24. Gomi H, Kawato M. 1992. Adaptive feedback control models of the vestibulocerebellum and spinocerebellum. *Biol Cybern* **68**:105–114.
25. Gormezano I, Kehoe EJ. 1975. Classical conditioning: some methodological-conceptual issues. In *Handbook of learning an cognitive processes*, Vol. 2: *Condition and behavior theory*, pp. 143–179. Ed. WK Estes. Lawrence Erlbaum Associates Hillsdale, NJ.
26. Hansen PD, Woollacott MH, Debu B. 1988. Postural responses to changing task conditions. *Exp Brain Res* **73**:627–636.
27. Hebb DO. 1949. *The organization of behaviour*. John Wiley, New York.
28. Holst von E, Mittelstaedt H. 1950. Das Reafferenzprinzip. *Naturwissenschaften* **37**:464–476.

29. Houk JC, Singh SP, Fisher C, Barto AG. 1990. An adaptive sensorimotor network inspired by the anatomy and physiology of the cerebellum. In *Neural networks for control*, pp. 301–348. Ed. WT Miller, RS Sutton, PJ Werbos. MIT Press, Cambridge.
30. Ito M, Kano M. 1982. Long-lasting depression of parallel fiber-Purkinje cell transmission induced by conjunctive stimulation of parallel fibers and climbing fibers in the cerebellar cortex. *Neurosci Lett* **33**:253–258.
31. Ito M. 1984. *The cerebellum and neural control*. Raven Press, New York.
32. Jansen J, Brodal A. 1940. Experimental studies on the intrinsic fibers of the cerebellum, II: the corticonuclear projection. *J Comp Neurol* **73**:267–321.
33. Kawato M, Furukawa K, Suzuki R. 1987. A hierarchical neural-network model for control and learning of voluntary movement. *Biol Cybern* **57**:169–185.
34. Kawato M, Gomi H. 1992. A computational model of four regions of the cerebellum based on feedback-error learning. *Biol Cybern* **68**:95–103.
35. Kawato M. 1996. Learning internal models of the motor apparatus. In *The acquisition of motor behavior in vertebrates*, pp. 408–430. Ed. JR Bloedel, TJ Ebener, SP Wise. MIT Press, Cambridge.
36. Keshner EA, Allum JH, Pfaltz CR. 1987. Postural coactivation and adaptation in the sway stabilizing responses of normals and patients with bilateral vestibular deficit. *Exp Brain Res* **69**:77–92.
37. Kolb FP. 1983. Results from a simulation model describing a biological, sensory feedback information system. *Informatik-Fachberichte* **71**:588–594.
38. Kolb FP, Rubia FJ. 1980. Information about peripheral events conveyed to the cerebellum via the climbing fiber system in the decerebrate cat. *Exp Brain Res* **38**:363–373.
39. Kolb FP, Rubia FJ, Bauswein E. 1987. Cerebellar unit responses of the mossy fibre system to passive movements in the decerebrate cat. *Exp Brain Res* **68**:234–248.
40. Kolb FP, Irwin KB, Bloedel JR, Bracha V. 1997. Conditioned and unconditioned forelimb reflex systems in the cat: involvement of the intermediate cerebellum. *Exp Brain Res* **114**:255–270.
41. Kolb FP, Timmann D, Baier PC, Diener HC. 2000. Classically conditioned withdrawal reflex in cerebellar patients, 2: impaired unconditioned responses. *Exp Brain Res* **130**:471–485.
42. Kolb FP, Lachauer S, Maschke M, Timmann D. 2002. Classical conditioning of postural reflexes. *Pflugers Arch* **445**:224–237.
43. Kolb FP, Lachauer S, Maschke M, Timmann D. 2004. Classical conditioning of postural reflex in cerebellar patients. *Exp Brain Res* **159**:163–179.
44. Larsell O. 1947. The development of the cerebellum in man in relation to its comparative anatomy. *J Comp Neurol* **87**:85–129.
45. Leaton, RN, Supple Jr WF. 1986. Cerebellar vermis: essential for long-term habituation of the acoustic startle response. *Science* **232**:513–515.
46. Leaton RN, Supple Jr WF. 1991. Medial cerebellum and long-term habituation of acoustic startle in rats. *Behav Neurosci* **105**:804–816.
47. Lopiano L, de'Sperati C, & Montarolo, P.G. 1990. Long-term habituation of the acoustic startle response: role of the cerebellar vermis. *Neuroscience* **35**:79–84.
48. Marr D. 1969. A theory of cerebellar cortex. *J Physiol* **202**:437–470.
49. Maschke M, Drepper J, Kindsvater K, Kolb FP, Diener HC, Timmann D. 2000. Involvement of the human medial cerebellum in long-term habituation of the acoustic startle response. *Exp Brain Res* **133**:359–367.
50. Miller S, Oscarsson O. 1970. Termination and functional organization of spino-olivocerebellar paths. In *The cerebellum in health and disease*, pp. 172–200. Ed. WS Fields, WD Willis. Warren H. Green, St. Louis.
51. Molinari M, Leggio MG, Solida A, Ciorra R, Misciagna S, Silveri MC, Petrosini L. 1997. Cerebellum and procedural learning: evidence from focal cerebellar lesions. *Brain* **120**:1753–1762.

52. Nashner LM. 1976. Adapting reflexes controlling the human posture. *Exp Brain Res* **26**:59–72.
53. Nixon PD, Passingham RE. 1999. The cerebellum and cognition: cerebellar lesions do not impair spatial working memory or visual associative learning in monkeys. *Eur J Neurosci* **11**:4070–4080.
54. Nixon PD, Passingham RE. 2000. The cerebellum and cognition: cerebellar lesions impair sequence learning but not conditional visuomotor learning in monkeys. *Neuropsychologia* **38**:1054–1072.
55. Oscarsson O. 1980. Functional organization of olivary projection to the cerebellar anterior lobe. In *The inferior olivary nucleus: anatomy and physiology*, pp. 279–289. Ed. O Creutzfeldt, C de Montigny, Y Lamarra. Raven Press, New York.
56. Pellionisz A, Llinas R. 1979. Cerebellar coordination: covarant analysis and contravariant synthesis via metric tensor: a tensorial approach to the geometry of brain function. *Neurosci Abs* 5:342.
57. Pissiota A, Frans O, Fredrikson M, Langstrom B, Flaten MA. 2002. The human startle reflex and pons activation: a regional cerebral blood flow study. *Eur J Neurosci* **15**:395–398.
58. Ramnani N, Toni I, Josephs O, Ashburner J, Passingham RE. 2000. Learning- and expectation-related changes in the human brain during motor learning. *J Neurophysiol* **84**:3026–35.
59. Robinson DA. 1976. Adaptive gain control of vestibuloocular reflex by the cerebellum. *J Neurophysiol* **35**:954–969.
60. Rosenblatt F. 1958. The perceptron: a probabilistic model for information storage and organization in the brain. *Psych Rev* **65**:386–408.
61. Rushmer DS, Roberts WJ Augter GK. 1976. Climbing fiber responses of cerebellar Purkinje cells to passive movement of the cat forepaw. *Brain Res* **106**:1–20.
62. Schmahmann JD. 1997. *The cerebellum and cognition*. Academic Press, San Diego.
63. Schreurs BG, McIntosh AR, Bahro M, Herscovitch P, Sunderland T, Molchan SE. 1997. Lateralization and behavioral correlation of changes in regional cerebral blood flow with classical conditioning of the human eyeblink response. *J Neurophysiol* **77**:2153–2163.
64. Schweighofer N, Arbib MA, Dominey PF. 1996a. A model of the cerebellum in adaptive control of saccadic gain, I: the model and its biological substrate. *Biol Cybern* **75**:19–28.
65. Schweighofer N, Arbib MA, Dominey PF. 1996b. A model of the cerebellum in adaptive control of saccadic gain, II: simulation results. *Biol Cybern* **75**:29–36.
66. Seidler RD, Purushotham A, Kim SG, Ugurbil K, Willingham D, Ashe J. 2002. Cerebellum activation associated with performance change but not motor learning. *Science* **296**:2043–2046.
67. Squire LR, Zola-Morgan S. 1991. The medial temporal lobe memory system. *Science* **253**:1380–1386.
68. Thach WT, Goodkin HP, Keating JG. 1992. The cerebellum and the adaptive coordination of movement. *Annu Rev Neurosci* **15**:403–442.
69. Thach WT. 1998. What is the role of the cerebellum in motor learning and cognition? *Trends Cogn Sci* **2**:331–337.
70. Thompson RF, Kim JJ. 1996. Memory systems in the brain and localization of a memory. *Proc Natl Acad Sci USA* **93**:13438–13444.
71. Thompson RF, Bao S, Chen L, Cipriano BD, Grethe JS, Kim JJ, Thompson JK, Tracy JA, Weninger MS, Krupa DJ. 1997. Associative learning. *Int Rev Neurobiol* **41**:151–189.
72. Timmann D, Musso C, Kolb FP, Rijntjes M, Jüptner M, Müller SP, Diener HC, Weiller C. 1998. Involvement of the human cerebellum during habituation of the acoustic startle response: a PET-study. *J Neurol Neurosurg Psychiatry* **65**:771–773.
73. Timmann D, Baier PC, Diener HC, Kolb FP. 2000. Classically conditioned withdrawal reflex in cerebellar patients, 1: impaired conditioned responses. *Exp Brain Res* **130**:453–470.
74. Timmann D, Drepper J, Maschke M, Kolb FP, Böring D, Thilmann AF, Diener HC. 2002. Motor deficits cannot explain impaired cognitive associative learning in cerebellar patients. *Neuropsychologia* **40**:788–800.

75. Topka H, Valls-Sole J, Massaquoi SG, Hallett M. 1993. Deficit in classical conditioning in patients with cerebellar degeneration. *Brain* **116**:961–969.
76. Tucker J, Harding AE, Jahanshahi M, Nixon PD, Rushworth M, Quinn NP, Thompson PD, Passingham RE. 1996. Associative learning in patients with cerebellar ataxia. *Behav Neurosci* **110**:1229–1234.
77. Voogd J, Glickstein M. 1998. The anatomy of the cerebellum. *Trends Neurosci* **21**:370–375.
78. Widrow B, Lehr MA. 1990. Thirty years of adaptive neural networks: perceptron, madaline, and backpropagation. *Proc IEEE* **78**:1415–1442.
79. Wolpert DM, Miall RC, Kawato M. 1998. Internal models in the cerebellum. *Trends Cogn Sci* **2**:338–347.
80. Woodruff-Pak DS, Papka M, Ivry RB. 1996. Cerebellar involvement in eyeblink classical conditioning in humans. *Neuropsychology* **10**:443–458.
81. Yeo CH, Hesslow G. 1998. Cerebellum and conditioned reflexes. *Trends Cogn Sci* **2**:322–330.

6.1

MODELING CANCER AS A COMPLEX ADAPTIVE SYSTEM: GENETIC INSTABILITY AND EVOLUTION

Kenneth J. Pienta

University of Michigan Comprehensive Cancer Center, Ann Arbor

Cancer is the second leading cause of death in the United States, claiming over 500,000 lives annually. While we now understand that cancer is a disease of genetic mutation, it is still difficult to describe how cancer arises from normal cells. Describing carcinogenesis in terms of a complex adaptive system reacting to the forces of Darwinian evolution gives a framework to understand tumorigenesis. This understanding is leading to new paradigms of cancer therapy, including multidisciplinary approaches to attack the cancer as a heterogeneous group of diseases as well as the development of aptamer molecular evolution techniques to design therapeutics to evolve as the cancer mutates.

1. INTRODUCTION

Generally, we consider evolution the fundamental strategy of life at the level of the organism. It is how we became who we are via interplay of genetic variation and phenotypic selection (1). The premise of evolution is that genes, and hence gene variants, are selected because they encode functions that in some way improve the chance of organism survival (2,3). This premise can be passed onto the level of the cancer cell. A tumor can be considered to be an organism or species that is able to speed up the evolutionary process by millions of years to select properties that help it survive and thrive within the macrocosm of the human body (4–6).

Address correspondence to: Kenneth J. Pienta, University of Michigan Comprehensive Cancer Center, 1500 East Medical Center Drive, 7303 CCGC, Ann Arbor, MI 48109-0946 (kpienta@umich.edu).

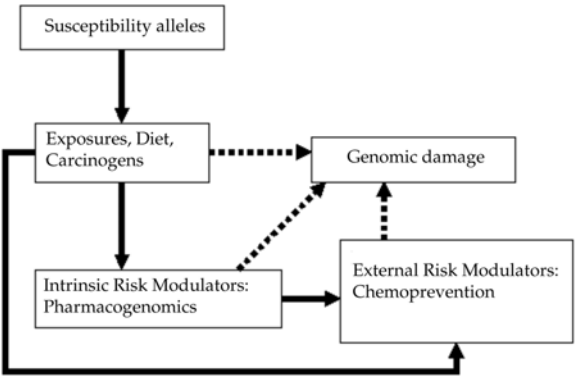


Figure 1. Cancer is a result of gene–environment interactions that lead to genetic mutations in pieces of DNA that in turn lead to survival advantage. Every person inherits a different set of genes from their parents. Some genes carry with them an inherent risk or susceptibility to cancer. Within this genetic background, we are exposed to multiple different carcinogens in the form of diet, infections, chemicals, radiation, etc. These exposures are processed by the body to varying extents. The carcinogen can directly cause DNA damage, or its risk may be modulated by intrinsic modulators. For example, each person processes the chemicals in tobacco smoke differently based on the genetic doses of modifying enzymes. In addition, the relative risk of exposures can be altered by extrinsic modulators, such as the antioxidants found in chemoprevention agents. Finally, the damaging factor must mutate a relevant part of the DNA. Many mutations occur in sequences of DNA that do not provide a survival advantage but occur rather in survival-neutral or deleterious genome sequences.

2. CANCER RISK IN THE CONTEXT OF AN EVOLUTIONARY PARADIGM

How then does a cancer cell evolve from a normal cell (see Figure 1). At the most basic level, it is the potential result of an accumulation of DNA damage that may count toward a survival advantage (2). Mutations to the genome must occur in places where they (a) do not lead to the death of the cell, (b) do not occur in a sequence of DNA that does not change behavior, and (c) occur in a place that conveys a growth or survival advantage. Meaningful DNA damage is the result of gene–environment interactions on multiple levels. First, cells may inherit "susceptibility" for damage from parental alleles. This can be at a very recognizable and measurable level—for example, a damaged DNA repair enzyme in Li-Fraumeni syndrome (4). Within this genetic background, the cells are assaulted by a variety of genome damaging exposures. These include radiation, viruses, microbes, carcinogens, chemicals, hormones, and other agents too numerous to list. But these risk factors to the genome are modulated in two important ways prior to their ability to damage the DNA.

First, the factors must pass through a phalanx of both organ- and non-organ-specific intrinsic risk modulators. Intrinsic risk modulators are inherited traits that do not contribute directly to DNA damage but modulate the environment to

which the cells are exposed. Examples include how well metabolizing enzymes function to modulate drug and hormone activity (pharmacogenomics) as well as how well a hormone such as testosterone binds to the androgen receptor based on the number of CAG repeats in the promoter region (7). In addition, before the damaging agent can cause mutation, it must evade extrinsic risk modulators. Extrinsic risk modulators are best characterized by chemoprevention agents such as antioxidants. Dietary factors such as selenium and vitamin E have been demonstrated to remove damaging oxygen radicals from the intracellular environment by catalyzing their breakdown to water (8,9). If the damaging agent escapes all of these potential protective mechanisms, it still must damage the DNA in a susceptible place that will allow a survival advantage (2,4). Most mutations to the DNA are either deleterious or neutral—very few are adaptive (1). In bacteria, for example, it is estimated that only one in 10,000 mutations provide an adaptive advantage (1,10). It is probable that this ratio would be greatly higher for the much more complex human genome.

These gene–environment interactions that contribute to cancer can be understood in the context of any number of evolutionary paradigms (Table 1). In breast cancer, a woman may inherit a mutation in the allele that contains BRCA-1, a gene important in maintaining normal breast cell function. This starts the

Table 1. Comparison of cancer cells and members of an animal herd: an evolution/natural selection paradigm

	Examples of contributors to mutations in cancer cells	Examples of contributors to successful selection and evolution in individual members of a herd
Susceptibility allele	Loss of BRCA1: increases chance of developing breast cancer	Loss of gene to make horns
Exposures	Diet, carcinogens, radiation, viruses, microbes, inflammation, chemicals, hormones	Predators, weather, diet, viruses, microbes, water supply
Intrinsic modulators	Drug metabolizing pathways	Length of legs, strength of muscles
Extrinsic modulators	Antioxidants, cancer screening, i.e., PAP smears	Size of the herd, place in the herd when attacked, ability of the herd to migrate in response to changes in environment

cell down the cancer pathway. Similarly, an antelope could inherit a rare allele and be born an albino, immediately putting it at a disadvantage to the other, camouflaged, members of the herd. Cancer cells are subject to a wide variety of genotoxic perturbations that could potentially cause mutation and selective pressure. These are mirrored by the same types of insults that a herd of animals must survive—for example, changes in weather, ability to withstand infections. These risks are modulated by inherent factors: in cells, for example, drug metabolizing enzymes; in animals, muscle fiber length (running speed). The risks are also modulated by extrinsic agents. Are there chemoprevention agents present for cells? For animals such agents include the presence of other protective species, the ability to migrate, and the number of adult males present to ward off attack.

3. CANCER EVOLUTION IN THE CONTEXT OF RECENT HUMAN EVOLUTION

Each cancer and the cancer cells that comprise it have a distinct phenotype; however, cancers do share a group of common characteristics (4,11,12). A tumor is the result of a collection of cancer cells that are actively acquiring mutations which allow the emergence of a successful clone of cells. This is a highly inefficient process, and tumors are filled with clones of cells that will not survive long term and are undergoing apoptosis (programmed cell death) as a result of harmful mutations and hypoxia, as well as other insults including immune surveillance. Some cells manage to acquire enough mutations and acquire the characteristics of a successful cancer cell. This can be compared, at least on one level, to the evolution of human civilization. A key difference between these two types of evolution is that we believe human beings evolved societies as a result of conscious decisions that increased our chances for species survival. To understand cancer clonal expansion, we have to explain cancer cell growth and survival in terms of an unconscious process. This is much more likely to be modeled by early evolution, when we pulled ourselves out of the sea and became multicellular organisms. However, the exercise of comparing a successful cancer cell successfully colonizing a new metastatic site to human civilization and colonization is worthwhile (see Table 2).

3.1. Unlimited Replicative Potential

Cancer cells are immortal. This does not mean that each cell itself lives forever (just like humans), but that the cell population doubles without limit and creates uncontrolled clonal expansion. In non-cancerous cells, a cell can double approximately 50 times before it undergoes senescence and dies (13). This has been termed the Hayflick number and is the result of an internal cell doubling

Table 2. Comparison of the process by which a cancer cell acquires the traits necessary for metastasis and how humans successfully colonize

Trait to allow growth and dissemination	Cancer cell—clonal expansion (unconscious)	Human organism—civilization expansion (conscious)
Unlimited replicative potential	Asexual reproduction, activation of telomerase	Sexual reproduction, desire for survival
Adaptation	Genetic instability, natural selection	Evolution, natural selectio
Protection from death	Loss of apoptotic pathway activation	Safety in numbers, city walls, castles
No growth inhibition	Anchorage independent growth	Ability to move about as individuals or groups without constraint
Nutrient supply	Stimulate new blood vessel growth	Building of water reservoirs and aqueducts to bring water to the population
Population expansion	Activation of proteases to break down surrounding tissue	Expansion/invasion into neutral territory
Evasion of enemies	Evasion of the immune surveillance system, e.g., as cells circulate prior to establishing themselves in a new organ	Avoiding contact with hostile forces that want to prevent colonization, e.g., warships trying to prevent colonial expansion
Successful colonization	Adaptation to the use of growth factors in the new environment and applying all of the traits above in a new environment	Building a new site, learning to eat new foods, and applying all of the traits outlined above in a new environment

clock, termed telomeres, built onto the end of each chromosome (14). Telomeres are specific strands of DNA that shorten with each cell division. At a critical shortened length, the cells undergo apoptosis, or programmed cell death. Cancer cells reactivate an enzyme, telomerase, that maintains the length of telomeres with each cell division by adding base pairs back onto the telomeres, thereby maintaining length integrity.

3.2. Adaptation, Mutation, and Natural Selection

A fundamental characteristic of cancer is the generation of tumor cell heterogeneity, i.e., cells with multiple mutated phenotypes, through a mechanism of

genetic instability (15–20). There are multiple ways that genetic instability can be generated (chromosomal instability and microsatellite instability) and observed. For example, tumor cells exhibit karyotypes that are grossly changed in quantity and quality from the complement of normal cell chromosomes.

Radman and colleagues have suggested that two different models can explain mutations in evolution (1). In one model, there is a low mutation rate in a very large population. In the second model, there is a high mutation rate in a limited population with coincident intense recombination, permitting a rare adaptive mutation to become separated from frequent deleterious mutations (1). The latter type of evolution can be seen in bacterial populations under stress. It is likely that the evolution of cancer is a combination of these two models. The initial mutations within a cell destined to become cancer happen as a result of a low mutation rate within a large population of cells. These mutations occur as a result of the interplay between susceptibility alleles and the environment, as outlined above. Within the expanding clone, a mutation eventually occurs that induces a "mutator phenotype" with coincident high mutation rates and the generation of tumor cell populations with a heterogeneous set of properties over a relatively short period of time. While this mutator phenotype may occur as a result of chance, it may also be facilitated by the exposure of the cells to stresses, such as hypoxia, as the size of the tumor increases. Indeed, it has been demonstrated that hypoxia induces genetic instability in cancer cell populations (21,22). The emergence of the mutator phenotype rapidly selects cells with the most robust survival advantages. This robust phenotype can be observed clinically. A cancer can be in remission for many years and then present with metastatic disease that quickly kills the patient over a matter of weeks or months (23).

3.3. Protection from Death

There are multiple redundant pathways in place to maintain the fidelity of cellular systems to prevent mutation and damage. More often than not, deleterious mutations lead to initiation of programmed cell death. Teleologically, this is built into systems to protect the rest of the cell population. There are multiple apoptotic pathways within cells in response to different types of cellular damage (24,25). Cancer cells have acquired mutations that allow damage to occur and accumulate without activating apoptotic pathways. It is almost unbelievable, the amount of genetic ruin, mutation, and rearrangement that a cancer cell can accumulate and still be viable, functional, and robust (26).

3.4. No Inhibition of Growth

For an organism or organ such as the liver to function in a coordinated fashion, it must control the individual cells that compose it, just as a society must. But for a human population to grow and expand, it must outfit groups to leave the population base and find new areas to populate. In cancer cells, this growth

inhibition is controlled by anchorage-dependent growth and maintenance. If a society sends out an individual to explore who is ill-equipped, that explorer will likely perish. If a normal cell becomes disconnected from its neighbors or the basement membrane that it resides on, apoptosis is triggered and the cell dies. Cancer cells have acquired mutations that allow them to grow independent of attachment to a basement membrane or to other cells (27–29). This anchorage independence releases the cell from communicating with its neighbors and breaks down the fundamental fidelity of the organism system. Several cell attachment proteins have been identified that have been demonstrated to be altered in cancer cells. These mutations also allow the cancer cell the freedom to leave the primary tumor environment and start down the path of metastasis (30).

3.5. Ability to Ensure a Nutrient Supply

A group of cancer cells undergoing clonal expansion can only become approximately a cubic millimeter in size (20 population doublings, one million cells) without a blood supply to oxygenate the cells (31). A critical step in successful cancer development is the release of factors such as vascular endothelial growth factor (VEGF) from the cancer cells to attract new blood vessel growth (neovascularity of angiogenesis) (32,33). This is a good example of how cancer cells, even in the presence of tumor cell heterogeneity, must unconsciously cooperate with each other. No single cell produces enough VEGF to stimulate the growth of a new blood supply by itself. Enough individual cells or clones must then have the ability to each secrete VEGF into the surrounding environment to allow a gradient of growth factor to be established to attract new blood vessels.

3.6. Population Expansion and Growth beyond Natural Boundaries

Cancer rarely kills its host because of its growth in one single organ. The majority of such cancers can be successfully treated by surgery and/or radiation. Even untreated, a solitary cancer can grow in a primary organ for years before becoming clinically evident. Cancer kills because it spreads (metastasizes) to other organs. This certainly requires the mutations that allow uncontrolled growth, anchorage independent growth, apoptosis evasion, and new blood vessel growth. But it also requires the acquisition of several other adaptation properties. Even though the cancer cell does not require its neighbors to grow, to be lethal it has to acquire properties that allow it to leave the primary tumor environment. For the cancer cell population to grow, it must break down its surrounding tissue environment. This periphery of the tumor is the most oxygenated and has the richest nutrient gradients. For the cancer cells to keep expanding into this environment, there must be a selective pressure for cells that can invade into that environment. It has been demonstrated that cancer cells secrete high amounts of proteases that break down the confining extracellular matrix of surrounding tissue (34). This allows continued growth of the clonal

populations without starvation. It also allows cells to find their way into the circulation and lymphatic system and spread to other parts of the organism. What is not clear is whether these types of mutations are the result of selective pressure or the simple result of an intrinsically unstable genetic system (see §3.2, above).

3.7. Evasion of Enemies during Growth and Expansion

At every level in its life, the cancer cell, and its daughter clones, must evade the immune system. The immune system is a remarkably adaptable system that seeks out and destroys foreign and harmful agents within the organism. Cancer cells have developed several ways to evade this immune system surveillance (35). In fact, in some ways, it appears that cancer cells flourish in lymph nodes, the waystations for the white blood cells that the body uses to fight infection and foreign bodies. The first question of every cancer evaluation is whether the cancer is in nearby lymph nodes? How it survives in this hostile environment is unclear. Many cancer cells have lost the proteins (antigens) on their cell surface that allow the body to recognize them as foreign. Other cancer cells secrete such cytokines as transforming growth factor beta (TGFβ), which inhibits the function of immune system cells (36).

3.8. Successful Colonization (Successful Metastasis)

All of the acquired mutations, whether acquired through selective pressure via adaptation to continued hostile environmental hurdles or by chance accumulation, result in a cancer cell clonal population that successfully metastasizes and grows in multiple new organ sites (4,30,37). This clearly resembles colonial expansion, and if the cancer was a set of thinking individuals, this is what one would expect. A final required trait is the ability to survive and flourish in new environments. This requires adapting to use the growth factors that the new environment is rich in. For example, prostate cancer cells grow well in the bone marrow, partly because transferrin is a potent growth factor for them and is present in high amounts in the bone (38).

4. MODELING CANCER AS A COMPLEX ADAPTIVE SYSTEM AT THE LEVEL OF THE CELL

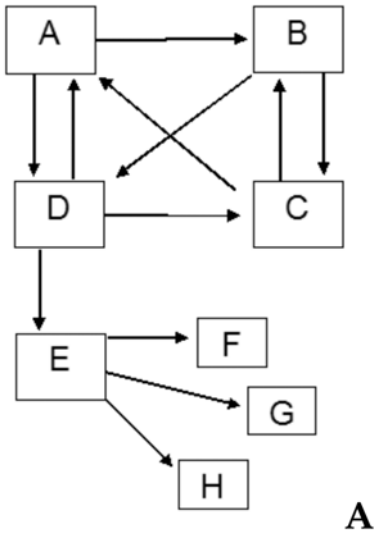
Cancer cells acquire the multiple traits necessary to survive within the greater macroenvironment of the host. We can also model the tumor, i.e., the collection of cancer cells, as acting in concert to function as a complex adaptive system—one that exhibits emergent properties (see Table 3). In this model, the individual cancer cells act as individual agents of the complex adaptive system (6,39–41). Each cell can act independently, but may also interact to create the tumor with the resultant properties.

Table 3. Cancer modeled as a complex adaptive system (CAS); these elements allow emergence of the CAS

Elements of a complex adaptive system (CAS)	Corresponding elements of a CAS in cancer
Agents: set of active components that interact selectively	Cells
Building blocks: provide a mechanism for generating a wide range of rules, tags, and internal models from a small number of parts	The genes that cancer cells draw on to acquire the properties that are necessary for survival; this often requires the activation of genes that are normally turned off in normal tissue
Aggregation: components group together according to similar abilities	Cells with similar adaptive mutations survive while others undergo apoptosis and die
Nonlinearity: a property resulting from conditional (nonadditive) interactions between agents	One cell cannot produce enough VEGF to stimulate new blood vessel growth to supply the tumor with nutrients but many cells together can
Flow: a property mediated by the movement of agents within the CAS; this can be represented by a series of IF/THEN rules	IF a cell produces proteases, THEN the tissue microenvironment will be broken down and a cell will be able to escape its local environment
Diversity: a property resulting when agents compete and adapt to fill available "niches" within the system	Genetic instability gives cells adaptive advantages that allow for clonal expansion and survival of the fittest
Tagging: a mechanism that facilitates interactions between and among components	The tissue matrix of the cancer cells allows dynamic remodeling of the system
Internal Model: a mechanism for providing agents with anticipatory actions	Cancer cells turn on genes that allow them to use multiple growth factors from a variety of different organ microenvironments—key to successful metastasis

4.1. Cells Are the Agents of the Cancer Complex Adaptive System

Complex systems are organized as a finite number of states, which can be defined by Boolean networks. A Boolean network is an array of elements, which has a particular rule associated with it, linked by a finite number of inputs. As the number of elements and links increases, the number of initial states of the system also increases. By cycling through the network (i.e., applying the elements rules as influenced by their links), however, one finds that the number of states the system occupies is limited to certain specific state-cycles (attractors)



(see Figure 2) (6,15,42). By taking the square root of the number of elements in a network, one can approximate the number of attractors. Therefore, Boolean networks obey a power law. Kauffman used these networks to show how the size of an organism's genome is related to the number of cell types it generates (42). For example, a sponge has approximately 10,000 genes and about 12 cell types. Humans have about 30,000 to 40,000 genes and over 250 cell types. A Boolean network with 100,000 elements, with each element linked to two others, has the potential of $10^{30,000}$ states. In fact, only 370 states are realized. Each of these states is an attractor; likewise, each cell type in a human body is a state-cycle attractor of the genome. A state-cycle attractor is defined by certain boundary conditions. In the cell, it has been proposed that these boundary conditions are defined by the ribonucleic acid (RNA)–protein complex termed the nuclear matrix (15,43). The nuclear matrix, therefore, may define the boundary conditions of a cell. Perturbation of a steady-state attractor through mutation may upset the genetic stability and cause the cell to enter the carcinogenic cascade (a new state, *E*). This state *E* is fundamentally unstable and results in a new set of attractors, i.e., cell types. These cell types are manifested as tumor cell heterogeneity. Cancer, then, is the result of multiple perturbations (i.e., mutations) to a cell that result in a redefinition, or perhaps even loss, of its boundary conditions. For a further discussion of modeling tumors as complex biosystems, see chapter 6.3, by Mansury and Deisboeck (Part III, this volume).

Initial States of Elements				Successor States of Elements			
A	B	C	D	A	B	C	D
0	0	0	0	0	0	0	0
0	0	0	1	1	0	1	0
				1	1	0	0
				0	0	1	1
0	0	1	1	1	0	0	0
				0	0	0	0
0	1	0	0	0	0	1	0
				1	0	0	0
				0	0	0	0
1	0	0	0	0	0	0	0
0	0	1	1	1	0	1	0
				1	1	0	0
				0	0	1	1
0	1	0	1	1	0	1	0
				1	1	0	0
				0	0	1	1
1	0	0	1	1	0	1	0
				1	1	0	0
				0	0	1	1
1	0	1	0	1	1	0	0
				0	0	1	1
				1	0	1	0
1	1	0	0	0	0	1	1
				1	1	0	0
				0	0	1	1
0	1	1	0	1	0	1	0
				0	0	1	1
				1	1	0	0
1	1	1	0	1	1	1	1
0	1	1	1	1	0	1	0
				0	0	1	1
				1	1	0	0
1	1	0	1	1	0	1	1
				1	1	1	0
				1	1	1	1
1	0	1	1	1	1	1	0
				1	1	1	1
1	1	1	1	1	1	1	1

E
2

E	F	G	H
2	3	4	5

B

Figure 2. The alteration of a four-element AND/OR Boolean network by cancer. The Boolean network diagrammed consists of four binary elements. Each element receives signals from two other elements, and sends signals to two elements. In this network, elements A and C obey the Boolean OR function and elements B and D obey the Boolean AND function. An element can either be "on" (1) or "off" (0), depending on the rule it obeys and the binary states of the elements from which it receives signals. For example, element A is active when either element C *or* D is active. In contrast, element B is active only when both elements A and C are active. In a four-element Boolean network there are 16 possible initial states. Using these values, and applying Boolean logic to each element, the successor states of each element can be determined. These successor states describe the behavior of the system. The 16 initial states, along with the successor states for each, are shown on the truth table to the left of the diagram. Although the number of initial states is numerous, the system eventually settles into only *three* state-cycles. These three cycles are shaded in the truth table. Perturbation of a steady-state attractor through mutation may upset the genetic stability and cause the cell to enter the carcinogenic cascade (a new state, *E*). This state *E* is fundamentally unstable and results in a new set of attractors, i.e., cell types. These cell types are manifested as tumor cell heterogeneity.

4.2. Genes Are the Building Blocks that Cell Structure and Function Is Based on

The six feet of DNA molecule that is present within each cell is segmented into genes that encode the proteins that interact with each other to form the structure and function of the cell (12). In normal cells, this structure and function is tightly controlled. In cancer, however, mutation leads to abnormal cellular functions and structural abnormalities. For a further discussion of the importance of linking cell structure to function, see chapter 2.1, by Huang, Sultan, and Ingber (Part III, this volume).

4.3. Cells with Similar Adaptive Mutations Aggregate into Clonal Populations

There is no question that the transformation of a normal cell to a cancer can be viewed as an evolutionary process and that the tumor can be viewed as a separate species (1–6,11,17,44). With the realization that a single tumor is an assembly of heterogeneous cells, it seems more appropriate to view each clonal population within the tumor as a different species (23,41,45). The members of each clone have a unique karyotype, morphology, and evolutionary fitness within the context of the global ecosystem: the human body. In this system, a tumor is a local ecosystem in which various species, clones, are in competition. As each tumor grows, it is a collection of clones that live and die. Each cubic centimeter of tumor (one gram) contains a billion individuals within it. If one assumes no death, this is equivalent to 35 generations from one aberrant cell. After ten more generations, these billion individuals have increased to a trillion. The population of a single tumor, therefore, surpasses the population history of mankind on the planet. The clone, or clones, that survive this growth are the most fit, and can spread (i.e., metastasize) to other local ecosystems (i.e., other organs). Their "success" eventually leads to a global ecological disaster: host death. Carcinogenesis is simply the act of speciation and the populating of the human global ecosystem.

4.4. Cancer Cells Acting in Concert Produce Properties with Growth Advantages

A primary tumor is a collection of cells that maintain contact and communication with cellular heterogeneity. Therefore, although only a single tumor may exist, it may be subdivided, on a cellular level, into separate populations (the clones). For these cells or clonal populations to survive, they must exert properties to help each other survive. A good example of this is the stimulation of new blood vessel growth (neoangiogenesis), which results in the sprouting of new blood vessels to the tumor, with subsequent nutrient flow to a growing tumor mass that would otherwise starve.

4.5. Cancer Cells Can Be Defined by a Set of IF/THEN Rules of Varying Complexity

A fundamental property of complex adaptive systems is flow—the ability to model much of its actions as a set of IF/THEN rules. Previously, we and others have demonstrated how the cell-signaling cascades of cells can be modeled as a series of biocircuits within the cell that can be perturbed by mutation (6,8,46). A classic circuit consists of wires leading to switches and/or circuits that guide electricity. For example, the electrical wiring of a house is connected to the power company, which feeds into the main junction box of the house, which then sends electricity to several main circuits. Each of these circuits in turn split to turn on appliances and plugs. These circuits are often termed "nodes." A prime example of a biocircuit in a cancer cell is hypoxia-inducible factor 1α (HIF-1α) (see Figure 3). As a tumor grows, it cannot extend beyond a millimeter in size before it outstrips its blood supply and the cells are starved for oxygen. Hypoxia serves as a tag to turn on HIF-1α. HIF-1α then turns on more than 40 different pathways to promote cell survival (47). Cell survival is accomplished by turning on pathways that promote cell energy metabolism in anoxic environments, blood vessel growth to decrease hypoxia, and a series of genes that promote metastasis—allowing the cell to migrate to areas of normal oxygenation. In this manner, HIF-1α functions as a "hypernode" in the biocircuit. If HIF-1α is mutated, the pathways to promote survival are never turned on and the cells undergo programmed cell death, or apoptosis. If the HIF-1α is counteracted by cytostatic factors, the tumor may exist in a quiescent steady state. If other growth factors are present fueling the cancer cells, proliferation is allowed to occur and the tumor mass grows. If HIF-1α is turned on, growth factors are present and the immune system escape mechanisms are in place, the cancer cells can metastasize. IF/THEN rules can also be applied at the level of the cells themselves. For example, IF a cell produces proteases, THEN it will break down the surrounding tissue matrix environment. These rules can be applied to each of the fundamental alterations that are necessary to form a lethal cancer cell (Table 2).

4.6. Genetic Instability Gives Rise to the Diversity of Cancer Cells: Tumor Cell Heterogeneity

The mutations that lead to formation of a tumor facilitate further changes in the cancer cells' genetic makeup. The genetic instability inherent in a tumor allows populations of cells to adapt rapidly to new conditions (see chapter 6.2, by Solé, Gonzales Garcia, and Costa (Part III, this volume)). This helps explain how cancers avoid the immune system, become resistant to certain drugs, and how they are able to metastasize. The strategy undertaken by a tumor appears to

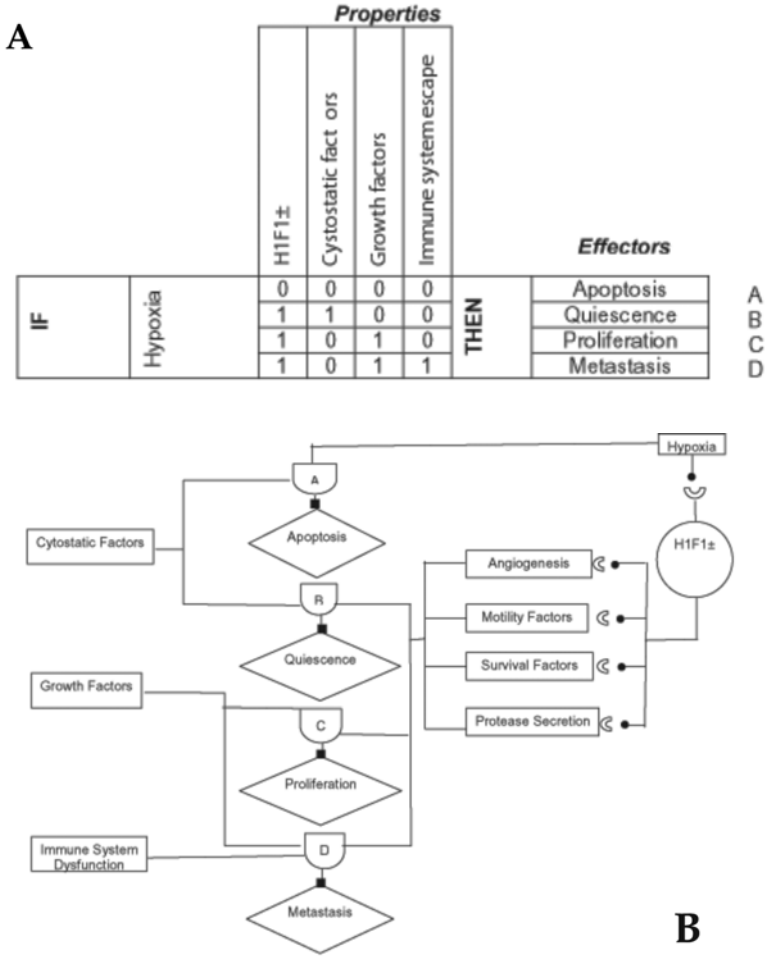


Figure 3. (**A**) A hypoxia-inducible factor 1α rule table, and (**B**) tokenized signal transduction. The rules outlined in the table have been translated into a tokenized signaling pathway using AND gates. These tables and pathways can be used to represent biocircuits. HIF-1α represents a critical node in the cell's response to hypoxia.

repeat features of evolutionary history. During the Cambrian period there was a great explosion of body types (48). Fossils from this epoch exhibit a far greater variation in gross morphology than exists today. Likewise, a tumor, due to its genetic pliability, can try innumerable cellular phenotypes, "searching" for the one that can thrive in the current environment (host organ) or spread to different environments (metastatic target organ), while discarding unfit cells. The fact that

tumors exhibit high death rates supports this contention (23). Most of the cells in a tumor die because they are incapable of forming strategies that allow them to survive in their current environment.

4.7. Complex Adaptive Systems Change How Strongly They Interact with Others in a Way that Maximizes the Average Fitness of the System

Tagging can, and does, occur at multiple levels within any system. At the level of the biocircuit within the cell, a tag can represent a phosphorylated or ubiquinated protein which signals that it should be recycled. At the level of the immune system, tagging can represent an antigen on a cell surface that allows white blood cells to recognize it as "self." Metastasis of a tumor can be taken as proof that the cells comprising that tumor have altered their interactions and connections not only with adjacent tumor cells, but also with the cells that form the lining of blood vessels. Metastasis requires active interactions between the cancer cells themselves and their environment. For cancer cells to enter the bloodstream, their connection with other cancer cells must be weakened. In the bloodstream, cancer cells bind to each other as well as platelets to survive the ambient turbulence. To escape the bloodstream, the cancer cells must then successfully bind to the endothelial cells of the target organ (30,37). All of these actions occur by altering the expression of cell–cell adhesion molecules in a dynamic fashion.

4.8. Tumor Cell Heterogeneity Provides Growth Advantages that May Appear to Be "Anticipatory"

The word "anticipatory" carries the connotation that a complex adaptive system is conscious of its actions. On the contrary, the strength of modeling through a complex adaptive system is that it needs no conscious thought process to form complicated, rule-based systems. The culmination of genetic instability and tumor cell heterogeneity is the acquisition of mutations requisite for a robust and lethal cancer. Cancer can do this because it can recapitulate evolution at a rate almost beyond comprehension.

5. CONCLUSION: APPLYING COMPLEXITY THEORY TOWARD A CURE FOR CANCER

The ultimate question is whether understanding cancer in terms of evolution and complexity theory can help us cure the disease. "Cancer" is a complicated

set of diseases arising in a variety of organs; however, these diseases share the similar properties outlined here. Currently, approximately half of all cancers are cured by surgical removal, radiation, or chemotherapy. The other half are lethal because they have metastasized (and are thus not removable) and because they are resistant to known therapies (a result of tumor cell heterogeneity).

What implications does the complex adaptive nature of cancer have for future research and treatment? It may be possible to turn a molecular process of therapeutic evolution against the evolutionary power of the cancer cell by designing a therapeutic approach that mimics and counters tumor evolution at a molecular level so that drug diversity can negate tumor cell heterogeneity and take away the advantage the cancer cell has to overcome present treatments. At a very simple level, the cancer could select its own drugs. This could be accomplished by using a randomized library of RNA sequences, termed aptamers, and permit the lethal cancer cells to bind to the aptamers with the highest affinity and specificity (49–55). These specific aptamers are amplified and then conjugated to radionuclides and cytotoxic drugs.

This is a novel approach to the treatment of resistant cancers. This technique essentially floods the cell with billions of random RNA sequences and allows the cancer cell to select out specific molecules to bind that it is expressing. Aptamers are modified oligonucleotides that are isolated by the systematic evolution of ligands by an exponential enrichment (SELEX) process. They are globular molecules that can recognize and bind with high affinity to a variety of cellular constituents. They are intermediate in size between small peptides and single-chain antibody fragments. One of their main advantages for cancer targeting and therapy is their small size compared to antibodies, which can result in improved cancer tissue permeation and delivery of lethal agents (54,55). Molecular evolution using random libraries of polymers might be used to select high-affinity binding components specific for prostate tumor cells. This pitting of molecular evolution against tumor evolution will permit a wide diversity of tightly binding synthetic ligands to match the biological diversity of the tumor cells. One type of these polymers that can be used includes highly diverse RNA molecules synthesized with random sequences and that are relatively inert to RNAse hydrolysis. A 15-mer of random nucleotides produces over a billion different RNA aptamers. These mixtures of aptamers can be differentially selected for their ability to bind tightly to cancer tissue while not binding to normal tissue. The specific tumor binding aptamers can then be amplified by reverse transcriptase and PCR to enrich the population of tight binding aptamers for the tumor cell. This process can be cycled over and over (see Figure 4). Lupold and colleagues have applied this concept to target prostate cancer cells (49). They were able to select two specific aptamers to an important prostate cancer marker, prostate-specific membrane antigen, from an initial 40-mer library of approximately 6×10^{14} random-sequence RNA molecules for their ability to bind to a recombinant protein representing the extracellular 706 amino

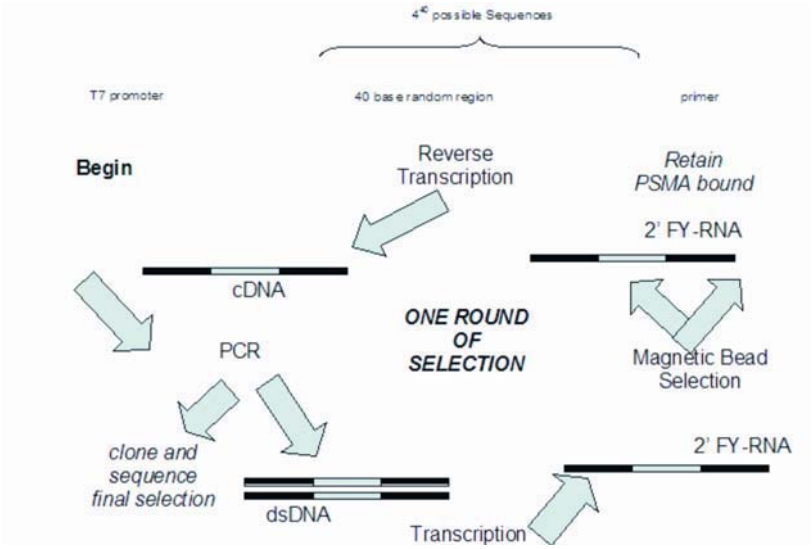


Figure 4. Construct of a random aptamer library with 40 random bases providing 6×10^{14} different aptamers. Different aptamers binding to tumor antigens are then selected and amplified through successive rounds of selection to create a highly purified population.

acids of PSMA, termed xPSM. Six rounds of in vitro selection were performed, enriching for xPSM binding as monitored by aptamer inhibition of xPSM N-acetyl-alpha-linked acid dipeptidase (NAALADase) enzymatic activity. After six rounds of selection, 95% of the remaining total aptamer pool consisted of only two sequences. These two aptamers, termed xPSM-A9 and xPSM-A10, were cloned and found to be unique, sharing no consensus sequences. The affinity of each aptamer for PSMA was quantitated by its ability to inhibit the enzymatic activity of PSMA. Aptamer xPSM-A9 inhibits PSMA noncompetitively with an average Ki of 11/9 nM. Distinct modes of inhibition suggest that each aptamer identifies a unique extracellular epitope of xPSM. One aptamer was truncated from 23.4 to 18.5 kDa and specifically binds LNcaP human prostate cancer cells expressing PSMA, but not to PSMA-devoid PC-3 human prostate cancer cells. These are the first reported RNA aptamers selected to bind a tumor-associated membrane antigen and the first application of RNA aptamers to a prostate-specific cell markers. These aptamers may be used clinically by modification to carry imaging agents and therapeutic agents that are directed to prostate cancer cells.

Within a single tumor, cells are heterogeneous. Just as important, tumor types are heterogeneous between patients. This approach of selected aptamers is

applicable to both types of heterogeneity. While is it expected that some aptamers may be common to all types of lethal cancers, it cannot be taken as a given. Every tumor may be different. However, these strategies give us the opportunity to explore customized therapy for individual patients. Ultimately, one would like to create a specific aptamer library for a particular patient. This could be particularly useful in the surgical patient. Cancer tissue would be used to generate a patient-specific library. This patient-specific library would then be used systemically to scavenge and destroy micrometastases. If and when the tumor progresses, samples from the metastatic lesions could be used to generate new libraries. In summary, therapeutic evolution should be able to outpace biologic evolution.

6. REFERENCES

1. Radman M, Matic I, Taddei F. 1999. Evolution of evolvability. *Ann NY Acad Sci* **870**:146–155.
2. Greaves M. 2002. Cancer causation: the Darwinian downside of past success? *Lancet Oncol* **3**:244–251.
3. Nowell PC. 1976. The clonal evolution of tumor cell populations. *Science* **194**:23–28.
4. Hanahan D, Weinberg RA. 2000. The hallmarks of cancer. *Cell* **100**:57–70.
5. Nesse RM, Williams GC. 1998. Evolution and the origins of disease. *Sci Am* **279**:58–65.
6. Schwab ED, Pienta KJ. 1996. Cancer as a complex adaptive system. *Med Hypotheses* **47**:235–241.
7. Coffey DS. 2001. Similarities of prostate and breast cancer: evolution, diet, and estrogens. *Urology* **57**:31–38.
8. Pathak SK, Sharma, RA, Mellon JK. 2003. Chemoprevention of prostate cancer by diet-derived antioxidant agents and hormonal manipulation [review]. *Int J Oncol* **22**:5–13.
9. Farinati F, Cardin R, Della Libera G, Herszenyi L, Marafin C, Molari A, Plebani M, Rugge M, Naccarato R. 1994. The role of anti-oxidants in the chemoprevention of gastric cancer. *Eur J Cancer Prev* **3**(suppl 2):93–97.
10. Taddei F, Radman M, Maynard-Smith J, Toupance B, Gouyon PH, Godelle B. 1997. Role of mutators in adaptive evolution. *Nature* **387**:700–702.
11. Marusic M. 1991. Evolutionary and biological foundations of malignant tumors. *Med Hypotheses* **34**:282–287.
12. Pienta KJ, Partin AW, Coffey DS. 1989. Cancer as a disease of DNA organization and dynamic cell structure. *Cancer Res* **49**:2525–2532.
13. Neumann AA, Reddel RR. 2002. Telomere maintenance and cancer—look, no telomerase. *Natl Rev Cancer* **2**:879–884.
14. Rubin H. 20:6. 2002. The disparity between human cell senescence *in vitro* and lifelong replication *in vivo*. *Nature Biotechnol* **20**:75–81.
15. Pienta KJ, Ward WS. 1994. An unstable nuclear matrix may contribute to genetic instability. *Med Hypotheses* **42**:45–52.
16. Nowak MA, Komarova NL, Sengupta A, Jallepalli PV, Shih Ie M, Vogelstein B, Lengauer C. 2002. The role of chromosomal instability in tumor initiation. *Proc Natl Acad Sci USA* **99**:16226–16231.
17. Anderson GR, Stoler DL, Brenner BM. 2001. Cancer: the evolved consequence of a destabilized genome. *Bioessays* **23**:1037–1046.
18. Hoglund M, Gisselsson D, Sall T, Mitelman F. 2002. Coping with complexity: multivariate analysis of tumor karyotypes. *Cancer Genet Cytogenet* **135**:103–109.

19. Kerbel RS, Cornil I, Korczak, B. 1989. New insights into the evolutionary growth of tumors revealed by southern gel analysis of tumors genetically tagged with plasmid or proviral DNA insertions. *J Cell Sci* **94**:381–387.
20. MacPhee DG. 1991. The significance of deletions in spontaneous and induced mutations associated with movement of transposable DNA elements: possible implications for evolution and cancer. *Mutation Res* **250**:35–47.
21. Yuan J, Narayanan L, Rockwell S, Glazer PM. 2000. Diminished DNA repair and elevated mutagenesis in mammalian cells exposed to hypoxia and low pH. *Cancer Res* **60**:4372–4376.
22. Reynolds TY, Rockwell S, Glazer PM. 1996. Genetic instability induced by the tumor microenvironment. *Cancer Res* **56**:5754–5757.
23. Coffey DS, Isaacs JT. 1981. Prostate tumor biology and cell kinetics-theory. *Urology* **17**(suppl):40–53.
24. Hussein MR, Haemel AK, Wood GS. 2003. Apoptosis and melanoma: molecular mechanisms. *J Pathol* **199**:275–288.
25. Bowen AR, Hanks AN, Allen SM, Alexander A, Diedrich MJ, Grossman D. 2003. Apoptosis regulators and responses in human melanocytic and keratinocytic cells. *J Invest Dermatol* **120**:18–55.
26. Hoglund M, Gisselsson D, Hansen GB, Sall T, Mitelman F. 2002. Multivariate analysis of chromosomal imbalances in breast cancer delineates cytogenetic pathways and reveals complex relationships among imbalances. *Cancer Res* **62**:2675–2680.
27. Abraham S, Zhang W, Greenberg N, Zhang M. 2003. Maspin functions as tumor suppressor by increasing cell adhesion to extracellular matrix in prostate tumor cells. *J Urol* **169**:1157–1161.
28. Su ZZ, Gopalkrishnan RV, Narayan G, Dent P, Fisher PB. 2002. Progression elevated gene-3, PEG-3, induces genomic instability in rodent and human tumor cells. *J Cell Physiol* **192**:34–44.
29. Kondoh N, Shuda M, Arai M, Oikawa T, Yamamoto M. 1988. Activation of anchorage-independent growth of HT1080 human fibroblasts. *Mutat Res* **199**:273–291.
30. Cooper CR, Chay CH, Gendernalik JD, Lee HL, Bhatia J, Taichman RS, McCauley LK, Keller ET, Pienta KJ. 2003. Stromal factors involved in prostate carcinoma metastasis to bone. *Cancer* **97**:739–747.
31. Folkman J. 2002. Role of angiogenesis in tumor growth and metastasis. *Semin Oncol* **29**:15–18.
32. van Nieuw Amerongen GP, Koolwijk P, Versteilen A, van Hinsbergh VW. 2003. Involvement of RhoA/Rho kinase signaling in VEGF-induced endothelial cell migration and angiogenesis in vitro. *Arterioscler Thromb Vasc Biol* **23**:211–217.
33. Chang L, Kaipainen A, Folkman J. 2002. Lymphangiogenesis new mechanisms. *Ann NY Acad Sci* **979**:111–119.
34. Chung AS, Yoon SO, Park SJ, Yun CH. 2003. Roles of matrix metalloproteinases in tumor metastasis and angiogenesis. *J Biochem Mol Biol* **36**(1):128–137.
35. Nambu Y, Beer DG. 2003. Altered surface markers in lung cancer: lack of cell-surface Fas/APO-1 expression in pulmonary adenocarcinoma may allow escape from immune surveillance. *Methods Mol Med* **74**:259–266.
36. Ivanovic VV, Todorovic-Rakovic N, Demajo M, Neskovic-Konstantinovic Z, Subota V, Ivanisevic-Milovanovic O, Nikolic-Vukosavljevic D. 2002. Elevated plasma levels of transforming growth factor-beta(1) (TGF-beta(1)) in patients with advanced breast cancer: association with disease progression. *Eur J Cancer* **39**:454–461.
37. Keller ET, Zhang J, Cooper CR, Smith PC, McCauley LK, Pienta KJ, Taichman RS. 2001. Prostate carcinoma skeletal metastases: cross-talk between tumor and bone. *Cancer Metastasis Rev* 20:333–349.
38. Rossi MC, Zetter BR. 1992. Selective stimulation of prostatic carcinoma cell proliferation by transferrin. *Proc Natl Acad Sci USA* **89**:6197–6201.

39. Schwab ED, Pienta KJ. 1997. Modeling signal transduction in normal and cancer cells using complex adaptive systems. *Med Hypotheses* **48**:111–123.
40. Holland J. 1995. *Hidden order: how adaptation builds complexity*. Addison-Wesley, New York.
41. Chinnaiyan AM, Coffey DS, Forrest S, Goldberg E, Holland J, Kepler T, Maley C, Mitchell M, Montie JE, Morowitz M, Nelson WG, Omenn G, Perelson AS, Pienta KJ, Rubin MA, Scardino P, Shapiro JA, Wheeler T. 2002. Merging bottom-up and top-down approaches to study prostate cancer biology. *Complexity* **7**:22–30.
42. Kauffman SA. 1991. Antichaos and adaptation. *Sci Am* **265**:78–84.
43. Pienta KJ, Murphy BC, Getzenberg RH, Coffey DS. 1993. The tissue matrix and the regulation of gene expression in cancer cells. *Adv Mol Cell Biol* **7**:131–156.
44. Temin HM. 1988. Evolution of cancer genes as a mutation-driven process. *Cancer Res* **48**:1697–1701.
45. Lewin RS. 1993. *Complexity: life at the edge of chaos*. Collier Books, New York.
46. Bao JZ, Davis CC, Schmukler RE. 1993. Impedance spectroscopy of human erythrocytes: system calibration and nonlinear modeling. *IEEE Trans Biomed Eng* **40**:364–378.
47. Semenza GL. 2003. Targeting HIF-1 for cancer therapy. *Nature Rev Cancer*. **3**:721–732.
48. Kerr RA. 2002. Evolution: a trigger for the Cambrian explosion? *Science* **298**:1547.
49. Lupold SE, Hicke BJ, Lin Y, Coffey DS. 2002. Identification and characterization of nuclease-stabilized RNA molecules that bind human prostate cancer cells via the prostate-specific membrane antigen. *Cancer Res* **62**:4029–4033.
50. Coffey DS. 2002. Understanding the cancer biology universe: enigmas, context and future prospects. *Cancer Biol Ther* **1**:564–567.
51. Faria M, Ulrich H. 2002. The use of synthetic oligonucleotides as protein inhibitors and anticode drugs in cancer therapy: accomplishments and limitations. *Curr Cancer Drug Targets* **2**:355–368.
52. Cerchia L, Hamm J, Libri D, Tavitian B, de Franciscis V. 2002. Nucleic acid aptamers in cancer medicine. *FEBS Lett* **528**:12–16.
53. Lato SM, Ozerova ND, He K, Sergueeva Z, Shaw BR, Burke DH. 2002. Boron-containing aptamers to ATP. *Nucleic Acids Res* **30**:1401–1407.
54. Tuerk C, Gold L. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**(4968):505–510.
55. Hicke BJ, Marion C, Chang YF, Gould T, Lynott CK, Parma D, Schmidt PG, Warren S. 2001. Tenascin-C aptamers are generated using tumor cells and purified protein. *J Biol Chem* 276:48644–48654.

6.2

SPATIAL DYNAMICS IN CANCER

Ricard V. Solé

ICREA-Complex Systems Laboratory, Universitat Pompeu Fabra, Barcelona, Spain, and The Santa Fe Institute, Santa Fe, New Mexico

Isabel González García and José Costa

Department of Pathology, Yale University School of Medicine, and Yale Comprehensive Cancer Center, New Haven, Connecticut

The relevance of spatial constraints to cancer growth and development are explored by means of simple models. It is shown that the explicit introduction of space into the picture allows the observation of new phenomena, such as the coexistence in time of different cells populations that would exclude each other without the presence of local interactions. The implications for our understanding of cancer and its possible treatments are discussed.

1. INTRODUCTION

Cancer can be defined as a semiautonomous growth of tissue that spreads to eventually compromise the vital functions of the host. Cancer is estimated to be responsible for seven million deaths worldwide yearly and is the second leading cause of mortality in the United States, where just over half a million new cases are registered each year. The epidemiological projections indicate that lifelong cumulative risk for an individual is one of every two males and one of every three females (1). Sustained improvements in prevention, early diagnosis, and therapy have resulted in a constant decrease of the absolute overall cancer

Address correspondence to: Ricard V. Solé, ICREA Complex Systems Laboratory, Universitat Pompeu Fabra, Dr Aiguader 80, 08003 Barcelona, Spain (ricard.sole@upf.edu).

mortality by 2% (2) per year, and progress in therapy has achieved cures for a significant number of patients.

In the last two decades, the success of the reductionist program of research has yielded unprecedented advances in our understanding of the basic mechanisms of cancer formation at the molecular and cell biological levels (3). This new knowledge enables us to begin to treat the more complex aspects of tumor formation at the next level of organizational complexity in mammals: the tissue level. Understanding the process of cancer formation and progression at the tissue level is relevant because many of the medical and biological characteristics of cancer depend on the organ or site where the tumor arises, in other words, the tissue the cancer comes from.

The order necessary for normal tissue function results from the stability, within homeostatic boundaries, of a complex network of interacting molecules that control intra- and intercellular regulatory circuits. The structural and functional integrity of the tissues is ensured by a compartmental organization (4). Each compartment of differentiated cells is maintained by a set of stem cells that, through asymmetrical division, ensure self-renewal and generate a proliferative population that expands exponentially. For a normal tissue, we can describe the interactions of cells using concepts and terms not too distant from those used in community ecology. The integrity of the tissue is guaranteed by the cooperation of its individual units, formed by the clonal cell population derived from a stem cell in charge of maintaining each unit. The entire tissue system functions under a continuous and stable turnover only disturbed by interaction with the environment and the process of aging. In the absence of external disturbances, each cell is programmed to play out a developmental program. In each tissue compartment the changes in form and function of the cells follow a transformational mode. Each individual of the ensemble follows a predetermined trajectory of change, and the proportions of each constitutive element of the ensemble remain constant. Mutations occur infrequently in stem cells responsible for maintaining the integrity of tissue compartments, and they are unlikely to be transmitted to the progeny, because normal cells are capable of repairing genetic damage resulting from errors during DNA replication. In addition, many mutations will result in deleterious effects and cause premature cell death by apoptosis. Thus under normal conditions, given the lack of genetic variation, selection is a weak force molding the populations that constitute a mature adult tissue. Competition among somatic cells is avoided, but clonal patches will increase in size when a neighboring clone fails and disappears.

In contrast, both tumor formation and tumor progression (the increase in biological malignancy with time) can be envisioned as microevolutionary processes during which change occurs in a variational mode (see also preceding chapter 6.1 by Pienta). Cancer is a genetic disease affecting somatic cells, and the emergence of a tumor requires accumulation of several mutations in a single cell. Mutations in oncogenes and tumor suppressor genes, including lack of

regulation of the epigenetic control of expression, disable key nodes in the networks controlling cell growth and differentiation. The rate of mutation can be increased by failure to maintain genetic integrity. This failure results in genetic instability caused either by deficiencies in the DNA repair machinery or by chromosomal instability. Defects in DNA repair can cause inherited susceptibility for certain cancers in humans, and chromosomal instability, manifested by loss of heterozygosity, is commonly observed in human cancerous tissue (5). The search for the genes responsible for chromosomal instability is being very actively pursued, and their identification will elucidate both how their products function and the role they play in tumor formation and progression (6,7).

Clonal evolution has been and continues to be regarded as the principal mechanism underlying tumor formation and progression (8). Implicit in the microevolutionary view of these processes is the notion of competition between the variants generated by mutation. Selection, playing on a checkerboard of mutations, can be a factor determining the particular genotype realized in a tumor cell population. Thus, the ecology of cancer cell populations that constitute the tumoral tissue is likely to depart significantly from that observed in normal tissues.

2. POPULATION DYNAMICS

Under a population-based view, two leading processes exploit mutational events: competition among different clones and expansion of those with larger growth rates. By applying the standard theory of ecological competition (9,10), coexistence is allowed by a limited range of conditions, defined by the strength of the interspecific interactions. Given two distinct populations, both will be present at low competition rates. Otherwise, one of the populations will win and exclude the second one. Thus, unless the parameters that define competition are properly tuned to avoid the parameter range of competitive exclusion, the population will tend to homogeneity.

Recent progress in the area of spatial ecology supports the idea that heterogeneity may be much more common than predicted by competition theory. Competitors that may exclude each other under defined experimental conditions may coexist under a spatially explicit framework (11,12). This is largely a consequence of the constraints imposed by space under limited dispersal. If two species that are good competitors involve individuals with short-range movement, the fact that their direct impact on other individuals is limited to their nearest neighbors slows down or even cancels the expected effects of exclusion. The final result of this spatially extended competition scenario is a patchy distribution of the two species that can be generalized to many different coexisting species. This occurs under a wide range of conditions, and it is not the result of a specific choice in the model (13,14). The only requirement is that spatial effects

must be able to suppress the theoretically predicted global effects of competitive replacement. Within the context of tumor growth, the space constraints are particularly important since cell movement within the tumor tissue is limited and cell–cell interactions, including paracrine effects, are short range.

We review here some key results and consider the theoretical arguments and computer simulations that support the view that the expected population structure to be found in a tumor is a heterogeneous spatial distribution of genotypes and phenotypes. At a time when molecular therapies are being implemented, the understanding of tumor heterogeneity in general theoretical terms can contribute to the design and understanding of multi-drug molecular therapies and to the prediction of treatment response.

3. COMPETITION IN TUMOR CELL POPULATIONS

A rich literature explores the growth of tumors assuming a homogeneous structure of the cancer cell population, without detailed reference to competition among the different cell types that compose a tumor (e.g., neoplastic cells and stromal cells). Under such approximation, the growth in the total number of cells N can be represented by a one-dimensional differential equation:

$$\frac{dN}{dt} = \phi_{\mu}(N)\,, \qquad [1]$$

where the right-hand side would contain a density-dependent behavior (i.e., size-dependent growth) and such parameter(s) as replication rate μ. A first approximation is given by a linear function $\phi_{\mu}(N) = \mu N$ and leads to an exponential increase in the number of tumor cells. Under conditions of limited resources, this model has to be modified by a term expressing the effect of a population control mechanism. The standard example is provided by a logistic equation:

$$\frac{dN}{dt} = \phi_{\mu}(N) = \mu N\left(1 - \frac{N}{K}\right), \qquad [2]$$

which can be solved, giving the time-dependent solution

$$N(t) = \frac{K}{\left(\dfrac{K - N_0}{N_0}\right)} e^{-\mu t}\,, \qquad [3]$$

indicating that after a long time the population stabilizes at some equilibrium level K, i.e., the carrying capacity. The carrying capacity is here fixed and would constitute a predefined limit to tumor growth. During tumor formation, such a limit is likely to be imposed by physical or nutritional constraints, but, particularly during the early stages of tumor formation, it may be overcome by the adaptive capacity conferred by mutation. Constant mutation, favored by the mutator phenotype, introduces competition among diverse cell populations and requires modification of the previous model. The fact that a tumor is composed of neoplastic cells and of stromal cells leads to the notion of competition between these two groups (15).

The consideration of a two-species competition model brings to the fore how spatial constraints modify the expectations derived from nonspatial models. Within the ecological context, two species will compete for given resources. In the case of a malignant tumor the competition will take place among different populations of cells. Gatenby (15) used the standard Lotka-Volterra system to model tumor progression:

$$\frac{dN_1}{dt} = r_1 N_1 \left(1 - \frac{N_1 - \alpha N_2}{K_1}\right), \qquad [4]$$

$$\frac{dN_2}{dt} = r_2 N_2 \left(1 - \frac{N_2 - \beta N_1}{K_2}\right). \qquad [5]$$

Here r_1 and r_2 are the intrinsic growth rates of the tumor population and normal host stromal population, respectively. The corresponding carrying capacities are denoted K_1 and K_2, and are defined as the maximum allowed population size that could occupy the tissue space and be supported by the environment. Here α and β are the interspecific competition coefficients. They measure the effects on stromal cells caused by the presence of the tumor cells and the effect of the stromal cells on the neoplastic cells, respectively. In this type of scenario, alternative models proposed by Gatenby incorporate different types of functional responses and consider availability of resources as a separate variable (16).

Before further exploring the relevance of space considerations within the context of tumor growth, we wish to consider one particular instance of the previous model that is defined by symmetric species competition, where the coefficients are the same, i.e., $\alpha = \beta$. If we consider these to be tumor cell populations with similar biological characteristics, the coexistence point is given by

$$N_1^* = N_2^* = \frac{(1+\beta)K}{1-\beta}. \qquad [6]$$

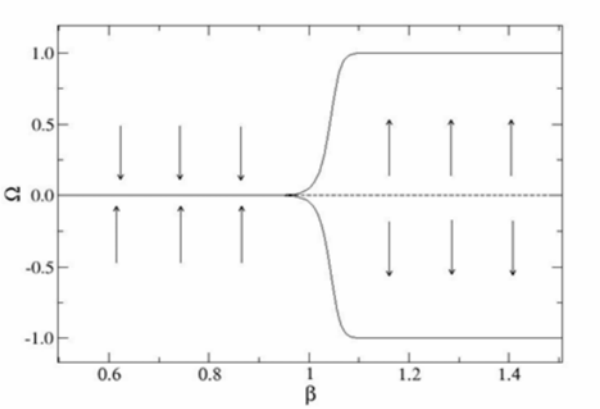


Figure 1. Symmetry breaking in a competition model. Here the parameter Ω measures the difference between the two populations (with a carrying capacity normalized to one). This parameter is plotted against the competition parameter, measuring the inhibitory effect of one population on the other.

Coexistence takes place and is stable provided that $\beta < 1$. Otherwise, one of two exclusion points,

$$(N_1^*, N_2^*) = (K, 0), \quad [7]$$

$$(N_1^*, N_2^*) = (0, K), \quad [8]$$

will be reached. Which population wins depends only on the initial conditions: the first population that increases in size over the other will take over. This is illustrated in Figure 1, which depicts the difference, $\Omega = |N_1 - N_2|$, between both populations at equilibrium starting from two different initial conditions in which each population has a slightly large population than the other.

Using different competition rates β, we observe a sharp transition at $\beta_c = 1$. For low subcritical competition levels, populations coexist and have the same size. At high competition rates, exclusion takes place. Once an initial difference between the populations is created, exclusion takes place. For this symmetric system, the two possible choices define two attractors of the dynamics. The situation is qualitatively shown in Figure 2. Here the original state is represented by the ball in the middle of the two valleys, representing the two possible attractors. The ball can roll down in two different directions, breaking the original symmetry and choosing one of two possible outcomes.

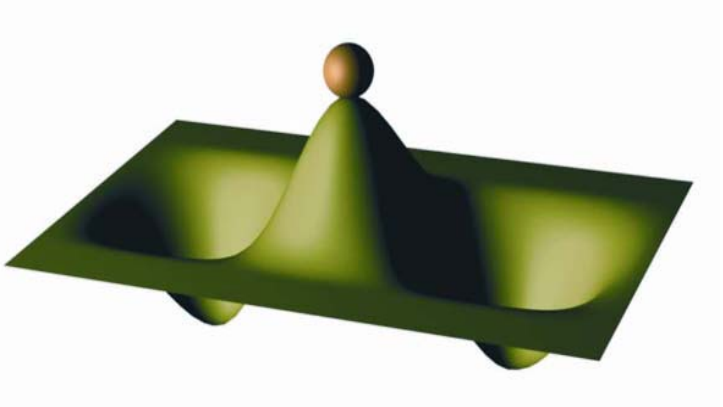


Figure 2. Symmetry breaking. This is a mechanical analogure of the pattern presented in Figure 1. The initial state of the system (the initial population sizes of the two competing cell populations) is represented by the ball at the top. Any small initial difference will be amplified, and eventually the ball will end up in one of the two minima. The initial symmetry is broken and one of the possible exclusion points has been chosen.

4. COMPETITION WITH SPATIAL DYNAMICS

By introducing diffusion in the competing species (clones), simple models can be generalized to spatially extended ones. The standard procedure is to add a diffusion term to the Lotka-Volterra equation.

The new model now takes the form

$$\frac{dN_1}{dt} = r_1 N_1 \left(1 - \frac{N_1 - \alpha N_2}{K_1}\right) + D_1 \nabla^2 N_1 , \qquad [9]$$

$$\frac{dN_2}{dt} = r_2 N_2 \left(1 - \frac{N_2 - \beta N_1}{K_2}\right) + D_2 \nabla^2 N_2 , \qquad [10]$$

where $D_i \nabla^2 N_i$ are the diffusion terms for each clone. Each population spreads at a rate D_i (the diffusion rate), and we have, in three dimensions,

$$\nabla^2 N = \frac{\partial^2 N}{\partial x^2} + \frac{\partial^2 N}{\partial y^2} + \frac{\partial^2 N}{\partial z^2} . \qquad [11]$$

Gatenby and Gawlinsky (30) used a similar model to study tumor invasion. The tumor host interface was treated as two competing cell populations in a low pH

microenvironment. The latter is known to exist in tumors, and the Gatenby-Gawlinsky model analyzed the effects of tumor-induced acidity in the tissue microenvironment. The model not only contributed to an understanding of tumor progression but also generated testable clinical predictions.

The spatially extended competition model can be simulated on a given spatial domain using appropriate boundary conditions. A standard procedure begins with an initial condition where the two populations start displaying the same local population plus a small noise term. In other words, we start with an almost homogeneous system very close to the coexistence point. The outcome is illustrated in Figure 3. Here two populations start growing and eventually exclude each other. The exclusion strongly depends on the initial local conditions; small initial advantages are rapidly enhanced and at a local scale competitive exclusion is completed. Together with competition there is diffusion: locally advantaged populations might disperse toward adjacent areas vacated by dying cells and create patches of homogeneous populations. Since both clones follow identical rules, the resulting situation is a set of patches that at the global scale achieve effective coexistence.

One possible limitation of the previous scenario is the initial condition where both populations are scattered over the surface. This circumstance could fit an experimental setup based on seeding a cell culture with a suspension of cells composed of two well-mixed populations but does not seem to apply to in-vivo adult tissues where small clones occupy a defined space compartment. When we simulate tumor growth starting from a spatially nonhomogeneous initial condition, the resulting "tumor" is highly heterogeneous and the two populations coexist on a global basis despite local exclusion.

These simple models explain how the local character of interactions imposed by tissue architecture constrain competitive interactions. Thus, we may be confronted by a situation common to other communities that largely evolve through competition such as plants. As stated by Tilman (17),

> Plants compete only with individuals living sufficiently nearby but each could cast shade on or have roots that overlap with the other. Because of poor dispersal ability, low local abundance, or chance events, however, many plant species may be absent from such a neighborhood and have their abundance be recruitment limited. Like a team that fails to appear at a sporting event, a species that is locally absent has forfeited any chance of competitive victory at the site. This can allow inferior competition to win by default. If there is recruitment limitation, the winners of local competition are not necessarily the best competitors that exist in the region, but the best competitors that happened to colonize a particular site. This can lead to essentially unlimited diversity.

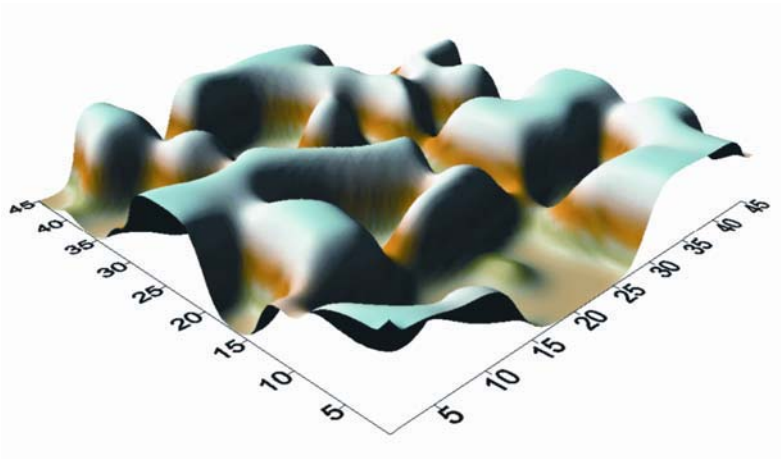


Figure 3. Spatial distribution of one of the two competing species in a two-species competition model with diffusion. Here the vertical axis indicates the relative amount of individuals of this species against the two-dimensional space. Starting from a roughly uniform state, where both species are equally distributed (except for a small noise), local exclusion takes place in some areas. Those regions where the first is abundant match those where the second is depleted, and vice versa.

These concepts have been applied to the study of cancer biology only recently (18), and in the next section we explore some of the implications of applying metapopulation dynamics models to the study of human cancer.

5. METAPOPULATION DYNAMICS AND CANCER HETEROGENEITY

The term *metapopulation* was coined by Richard Levins to describe a population consisting of many local populations (19,20). Metapopulation models are well adapted to describe systems composed of a set of patches of habitats. Visualizing tumor tissue as an ensemble of populations, differing by the mutations present in each group of cells makes the application of metapopulation dynamic models to the study of cancerous tissues intuitively appealing.

The notion that invasive cancers are composed of diverse populations stems from the sometimes striking variation in cell phenotype found in human tumors. It is not unusual to observe nodules of unique appearance growing within larger nodules of tumor groups of distinct morphology apparently coexisting next to each other. Complementing these morphological observations, several groups of

investigators conducting molecular analysis of tumor tissue have described a significant degree of genetic heterogeneity in tumors of various types (21–23).

It is significant that the heterogeneity at the molecular level involves genes that play a important role in the pathogenesis of the tumor state and in the maintenance of the neoplastic phenotype. Because of this, one may argue that alleles at these loci should be under selective pressure, so that one would expect to find the allele present in every cell of the tumor. Many studies that have been interpreted as demonstrating clonality of cancer genes in tumors do not deal with the caveat that they examined large mixed populations in aggregate and that, given the sensitivity of the detection technologies used in most cases, the results are likely to show a dominant allele, but do not critically exclude the presence of a quilted pattern of cell populations harboring allelic diversity in cancer genes. Most recent studies demonstrating genetic heterogeneity are predicated on using microdissection as a means of procuring relatively small and well-defined cell populations.

In attempting to generate models that will illuminate the phenomenology of tumor cell heterogeneity, it is crucial to simplify by focusing, sometimes arbitrarily, on a subset of biological characteristics of the tumor cell. Some key properties of cancerous tissues result in the increased mutational rate of tumor cells. These include their increased proliferative activity (which can overcome elevated rates of cell death) and prolongation of the lifespan of a cell by a variety of mechanisms that include avoidance of senescence. Other properties of tumors, such as the capacity to invade or stimulation of tumor angiogenesis, are undoubtedly of crucial importance to tumor pathophysiology and will need to be incorporated in more refined versions of metapopulation models of tumor growth. Is there an ecologic analogue for the two strategies found in tumor cells, e.g., high replicative rate and avoidance of death? The answer is affirmative, given the fact that two negatively correlated strategies for space occupation and persistence are observed in a community. The first, high colonization, is simply a strategy of effectively occupying available space. In tumors this stems from the high replication rate of tumor cells compared to normal cells. The second involves local extinction. It is known that some plants will be more tolerant to existing local conditions than others because they have developed mechanisms to tolerate such adverse conditions as shading from their neighbors. Inability to trigger apoptosis or independence of the need for survival signals from the environment are documented strategies evolved by tumor cells during tumor formation. In many cases there may be a tradeoff between both strategies: good colonizers are less able to persist, whereas poor colonizers are best adapted to local conditions, less prone to extinction, and thus better competitors.

A simple metapopulation model can be constructed to illustrate the importance of each ingredient. Let p be the fraction of patches occupied by a given species; the Levins model will describe the evolution in time of the metapopulation by

$$dP/dt = CP(1 - P) - mP, \quad [12]$$

where c is the colonization rate and m the extinction rate. Both are local rates and describe the ability of the given species to occupy neighboring patches and to get locally extinct. The model has two equilibrium points: $p^* = 0$ and

$$P^* = 1 - m/c. \quad [13]$$

It can be seen from the last expression that the species will persist provided that $c > m$. In other words, there is a minimum requirement for colonization rates in relation to extinction rates. The choice of strategy for a population will in part depend on the environmental constraints and the contingencies that may influence these constraints.

In cancer both strategies can be adopted, leading to coexistence of diverse cell populations (18). The Levins model enables us to explore what type of tradeoffs between competing populations allow them to coexist. Let us consider two competitors, which for simplicity we will assume to be ordered in a hierarchical way. The superior competitor will be more likely to colonize available adjacent patches, and the second competitor will be less likely to die.

We can map these two strategies in the cancer context by choosing two genes in epithelial cells: one coding for the receptor of tumor growth factor beta (TGFBR-2), and the second encoding a pro-apoptotic protein named BAX. These two genes will allow us to couple phenotypic traits associated with replication and senescence with the underlying genetic traits. Loss of function of TGFBR-2 makes cells grow faster, while loss of BAX function increases cell longevity. In terms of a model approach, this means that more mutations in TGFBR-2 will map into increased proliferation rates, whereas mutations in BAX will reduce cell mortality. Homozygous mutants will presumably have a stronger phenotype, although the biological consequences of haploinsufficiency in tumor suppressor genes are only emerging (18,24,25) and have not yet been explored in much detail.

These constraints can be implemented as follows according to Levins dynamics:

$$dP_1/dt = c_1P_1(1 - P_1) - m_1P_1. \quad [14]$$

From this equation we conclude that the first population senses as available habitat all spaces not occupied by $S - 1$, but perhaps occupied by $S - 2$. The second population is described by:

$$dP_2/dt = c_2P_1(1 - P_1 - P_2) - m_2P_2 - c_1P_1P_2. \quad [15]$$

Now the space available to the second species is reduced to $1 - P_1 - P_2$, i.e., those patches occupied by $S - 1$ cannot be invaded. This simple model predicts that coexistence will occur provided that some critical tradeoffs are satisfied. Specifically, it can be shown that coexistence will take place if

$$c_1 > m_1, \tag{16}$$

$$c_2 > [c_1(c_1 - m_2 - m_1)]/m_1. \tag{17}$$

Essentially what these inequalities define is a set of requirements that the two competitors have to verify in order to not exclude each other. The inferior colonizer, for example, will be able to persist if it is able to exploit local resources more efficiently. Since energy resources have to be distributed in different survival strategies, it is not difficult to understand that such tradeoffs will be expected to be common.

Of course, the previous models can seem too simplistic to say anything relevant. But we have learned from years of modeling and data analysis that simple models often capture the underlying causes responsible for the observed patterns. A step beyond the previous metapopulation models is to consider space explicitly (13,20). This was done within the context of cancer heterogeneity for the spatial dynamics of the two genes mentioned above (TGFBR-2 and BAX). A three-dimensional space was considered in terms of a cubic lattice (18). Each lattice point was occupied by a single cell, whose internal state was defined in terms of two copied of the two genes. The reproduction and death rates of each cell were determined by the state of their two key genes and their mutations. Mutations in TGFBR-2 triggered phenotypic responses in terms of increased replication, whereas mutations in BAX led to decreased levels of apoptosis. Mutations within this framework are independent random events not coupled with environmental factors.

Available experimental data provided well-defined frequencies of each mutation for each tumor analyzed. The simulation model, starting from a small group of non-mutated cells, could produce a whole spectrum of possible outcomes through its growth dynamics. In order to determine the most likely combination of growth and death rates compatible with the available information, a search algorithm was used in order to find the optimal parameter set compatible with experimental data. In this way, we can go beyond the limitations imposed by a high-dimensional parameter space and simply leave the algorithm to search for candidate solutions.

The model revealed a very good parameter combination able to reproduce the observed frequencies of mutants and their spatial distribution (18). In Figure 4 we show an example of the model outcome for the optimal parameter combination. As expected from the competition models described in the previous sections, spatial heterogeneity was achieved because of a combination of tradeoffs

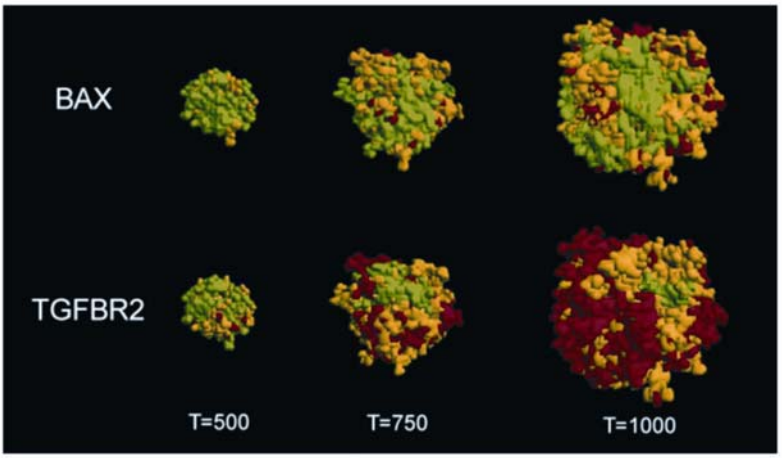


Figure 4. An example of the 3D simulation model using optimal parameters. Here the spatial distribution of the two genes under consideration is shown. Green, yellow, and red correspond to wild type, one, and two mutations, respectively. BAX displays lower levels of mutation, whereas TGFBR-2 presents larger red zones, indicating that homozygous mutations are widespread.

and the constraints imposed by local interactions. Although the same model achieves homogeneity under an absence of spatial constraints, it fails to do so when a realistic spatial structure is introduced. The model thus provides an accurate description of the basic rules operating through tumor progression and also predicts a number of features (such as the relative mutation levels associated with each mutation) that can be eventually tested from experiments.

6. DISCUSSION

A study of the geography of genetic diversity present in human tumors with a mutator phenotype revealed an unsuspected degree of diversity. Remarkably, tumor cells could be found with a wild-type genotype in the midst of the tumoral tissue. The ability to reproduce the results using a metapopulation dynamic model suggests that the same fundamental principles that explain the lack of dominance in complex ecosystems are at play in neoplasia and underlie the maintenance of clonal diversity in human tumors. Study of the metapopulation models further suggest that, in order to generate the degree of diversity observed, the mutational rate can remain stable throughout the natural history of

the tumor and that no differences in neutrality of interaction or separated time scales are necessary for different groups.

Although after prolonged evolution tumor systems are likely to tend toward homogeneity, diversity is the preferred landscape state in most intermediate or advanced tumors. Homogeneity may be achieved theoretically, but only after death of the host. Detailed studies of preneoplastic tissues reveal that diversity is the rule. When tissues at risk for developing a tumor are studied, diversity within cancer gene loci is commonly found.

Quantitative studies of the frequencies of alleles for different cancer genes indicate that the normal physiology of tissues includes random fluctuation of mutated alleles present at very low frequencies. Theoretical work by the group of Nowak (26) indicates that the best protection against mutations that favor overgrowth of a clone is a tissue homeostasis that is locally regulated by small compartments of cells in tissues. The small size of the compartments, which protects against the emergence of populations harboring a mutated oncogene or tumor suppressor gene, favors the emergence of genetic instability (e.g., chromosomal instability). If confirmed, these results suggest that genetic instability is a characteristic of small tumor clones (oncodemes) from the earliest stages of the neoplastic process. In fact, many expanded oncodemes may coexist in tissue, at very low frequencies, and a clinically significant malignant neoplasm may never emerge. Modeling of the preneoplastic states found in tissue is in keeping with observations of mutated allele frequencies found in somatic cells (27).

The idea that tumors are composed off a heterogeneous ensemble of tumor cell populations suggests that the biological behavior of tumors depends on the composition of the tumor tissue rather than on a common characteristic of "the tumor cell." In very diverse tumors it is most likely that at least one cell type will harbor a mutation that renders the cell resistant to therapy. Chemotherapy may in many instances introduce a disturbance in the system, and disturbance, an external agent of mortality, can modulate the effect of spatial heterogeneity on biological diversity (28). The relationship of disturbance to diversity could explain why a drug, if applied too often or too rarely, can cause an increase in tumor diversity. Another agent of disturbance that is likely to have significant effects on the natural evolution of a tumor is ischemia. Loss of perfusion in large areas is likely to modify the degree of diversity in their immediate vicinity, and such events are not unusual throughout the natural history of tumors. It is of interest to consider whether additional ecological principles can illuminate tumor pathophysiology. It seems not too risky to speculate that productivity, the flow of energy through a system, is determined in the case of tumors by the flow of blood through the tumor tissue. In bacterial systems, productivity is known to affect diversity in the same way that disturbance does (29).

As more and more therapies are designed to hit specific molecular targets, it will be important to know the degree of heterogeneity in tumors that are to be treated. Ideally, one would want to monitor the tumor-cell cell composition as

the therapy has its effect. Knowing the degree of diversity may help design the best multi-modality therapies for a given tumor, taking into account the effects of disturbance on diversity. Perhaps one day we will be able to determine what therapeutic strategies are going to optimally alter tumor growth.

7. ACKNOWLEDGMENTS

IGC acknowledges the support of a Ramon y Cajal research contract. This work was also supported by grant NIH 113004 and by the Santa Fe Institute. Dr. Isabel Gonzalez is now at the Department of Molecular Pathology, IDIBAPS, Villarroel 170, 08036 Barcelona, Spain.

8. REFERENCES

1. Jemal A, Thomas A, Murray T, Thun M. 2002. Cancer statistics 2002. *CA Cancer J Clin* 52:23–45.
2. National Cancer Institute. 2004. *Cancer mortality maps and graphs* (http://www3.cancer.gov/atlasplus/charts.html).
3. Hahn WC, Weinberg RA. 2002. Modelling the molecular circuitry of cancer. *Natl Rev Cancer* **2**:331–341.
4. Mintz B. 1971. Clonal basis of mammalian differentiation. *Symp Soc Exp Biol* **25**:345–370.
5. Nowak MA, Komarova NL, Sengupta A, Jallepalli PV, Shi IeM, Vogelstein B, Lengauer C. 2002. The role of chromosomal instability in tumor initiation. *Proc Natl Acad Sci USA* **99**:16226–16231.
6. Cahill DP, Lengauer C, Yu J, Riggins GJ, Willson JK, Markowitz SD, Kinszler KW, Vogelstein B. 1998. Mutations of miotic checkpoint genes in human cancers. *Nature* **392**:300–303.
7. Bardelli A, Cahill DP, Lederer G, Speicher MR, Kinzler KW, Vogelstein B, Lengauer C. 2001. Carcinogen-specific induction of genetic instability. *Proc Natl Acad Sci USA* **98**:5570–5775.
8. Nowell PT. 1976. The clonal evolution of tumor cell populations. *Science* **194**:23–28.
9. May RM. 1975. *Stability and complexity in model ecosystems*. Princeton UP, Princeton.
10. Case TJ. 2000. *An illustrated guide to theoretical ecology*. Oxford UP, Oxford.
11. Solé RV, Bascompte J. 1992. Stability and complexity in spatially extended two-species competition. *J Theor Biol* **159**:469–480.
12. Tilman D. 1994. Competition and biodiversity in spatially structured habitats. *Ecology* **75**:2–12.
13. Hanski, I. 1999. *Metapopulation dynamics*. Oxford UP, Oxford.
14. Durrett R, Levin SA. 1994. Stochastic spatial models. *Phil Trans Roy Soc London B* **343**:329–350.
15. Gatenby RA. 1996. Application of competition theory to tumor growth: implications for tumour biology and treatment. *Euro J Cancer* **32**:722–726.
16. Gatenby RA. 1995. Models of tumor–host interactions as competing populations. *J Theor Biol* **176**:447–455.
17. Tilman D. 1999. Diversity by default. *Science* **283**:495–496.
18. Gonzalez-Garcia I, Solé RV, Costa J. 2002. Metapopulation dynamics and spatial heterogeneity in cancer. *PNAS* **99**:13085–13089.
19. Levins, R. 1970. Extinction. *Lect Notes Math* **2**:75–107.
20. Bascompte J, Solé RV. 1998. *Modeling spatiotemporal dynamics in ecology*. Springer, Berlin.

21. Baisse B, Bouzourene H, Saraga E-P, Bosman F-T, Benhattar J. 2001. Intratumor genetic heterogeneity in advanced human colorectal adenocarcinoma. *Intl J Cancer* **93**:346–352.
22. Barnetson R, Jass J, Tse R, Eckstein R, Robinson B, Schnitzler M. 2000. Mutations associated with microsatellite unstable colorectal carcinomas exhibit widespread intratumoural heterogeneity. *Genes Chromosomes Cancer* **29**:130–136.
23. Samowitz W-S, Slattery M-L. 1999. Regional reproducibility of microsatellite instability in sporadic colorectal cancer. *Genes Chromosomes Cancer* **26**:106–114.
24. Celeste, A., Difilippantonio J, Difilippantonio MJ, Fernandez-Capetillo O, Pilch DR, Sedelnikova OA, Eckhaus M, Ried T, Bonner WM, Nussenzweig A. 2003. H2AX Haploinsufficiency Modifies Genomic Stability and Tumor Susceptibility. *Cell* **114**:371–383.
25. Bassing CH, Bassing CH, Suh H, Ferguson DO, Chua KF, Manis J, Eckersdorff M, Gleason M, Bronson R, Lee C, Alt FW. 2003. Histone H2AX: a dosage-dependent suppressor of oncogenic translocations and tumors. *Cell* **114**:359–370.
26. Michor F, Iwasa Y, Komarova NL, Nowak MA. 2003. Local regulation of homeostasis favors chromosomal instability. *Current Biology* **13**:581–584.
27. Tuck D, Costa J. Unpublished observations.
28. Buckling A, Kassen R, Bell G, Rainey P. 2000. Disturbance and diversity in experimental microcosms. *Nature* **408**:961–964.
29. Kassen R, Buckling A, Bell G, Rainey PB. 2000. Diversity peaks at intermediate productivity in a laboratory microcosm. *Nature* **406**:508–511.
30. Gatenby RA, Gawlinski ET. 2003. The glycolytic phenotype in carcinogenesis and tumor invasion: insights through mathematical models. *Cancer Res* **15**:3847–3854.

6.3

MODELING TUMORS AS COMPLEX BIOSYSTEMS: AN AGENT-BASED APPROACH

Yuri Mansury and Thomas S. Deisboeck

Complex Biosystems Modeling Laboratory, Harvard-MIT (HST)
Athinoula A. Martinos Center for Biomedical Imaging,
Massachusetts General Hospital, Charlestown

We argue that tumors behave as *complex dynamic self-organizing* and *adaptive biosystems*. In this chapter, we present a numerical *agent-based model* of malignant brain tumor cells in which both time and space are discrete yet environmental variables are treated as a continuum. Simulations of this *multiscale* algorithm allow us to investigate the molecular, microscopic, and multicellular patterns that *emerge* from various interactions among cells and between the cells and their environments.

1. INTRODUCTION

Studies of multicellular organisms recently experienced a paradigm shift into a framework that views these biological life forms as *complex systems*. In studies of malignant tumors, such a paradigm shift is accompanied by growing evidence that these tumors behave as dynamic self-organizing and adaptive biosystems (see (1–3) and chapter 6.1 by Pienta (Part III, this volume)). The present chapter reviews the applications of insights from complex system research in studies of malignant brain tumor cells, such as glioblastoma multiforme (GBM). Understanding the emerging behavior of malignant cancer cells with the use of

Address correspondence to: Thomas S. Deisboeck, Complex Biosystems Modeling Laboratory, Harvard-MIT Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, East CNY-2301, 13th Street, Building 149, Charlestown, MA 02129 (deisboec@helix.mgh.harvard.edu).

numerical simulations is the current state of the art in modeling of the (brain) tumor as a multicellular complex system. New advances in computer microprocessors as well as programming tools have significantly improved the speed with which these simulations can be performed. The "agent-based" approach (see also this volume, Part II, chapter 1 by Shalizi), in which the smallest unit of observation is the *individual* cancer cell, offers many advantages not possessed by, for example, the continuum model that has been proposed for cancer in the preceding chapter 6.2 by Solé, Gonzales Garcia, and Costa. The principal motivations for using an agent-based model to examine the spatio-dynamic behavior of a malignant brain tumor can be listed in the following non-exhaustive list.

First, to date, conventional clinical imaging techniques can only detect the presence of malignant behavior after the tumor has reached a critical size larger than a few millimeters in diameters. Hence, long before the disease process can be diagnosed on image, the tumor likely has already started to invade the adjacent brain parenchyma, thus seriously undermining the options of cytoreductive therapy. A computational model can therefore be useful in helping to better understand these critical *early stages* of tumor growth. During such initial stages, only a relatively small number of tumor cells have emerged in the system; hence, a continuum model based on the dynamical behavior of tumor "lumps" (each lump representing a large population of tumor cells) will fail to capture the early growth process that is highly path dependent on the discrete history of each individual cell.

Second, an agent-based model is suitable to examine these aggregate (i.e., macroscopic) patterns that result from the microscopic (i.e., local) interactions among many individual components. This *micro–macro* perspective is indispensable in a model of cancer cell heterogeneity that is driven by molecular dynamics. On one hand, the bottom–up approach is useful due to the ability of cancer cells to proliferate rapidly during tumorigenesis, which leads to intense competition for dominance among distinct tumor clones (4). On the other hand, the collective behavior of the network of individual cancer cells may result in emerging large-scale multicellular patterns, which calls for a system-level outlook. Indeed, the assumed rapid nonlinear growth of brain tumors during the initial stage and the subsequent invasion into regions of least resistance, most permission, and highest attraction would indicate an "emergent" behavior that is the hallmark of a complex dynamic self-organizing system (3).

Third, such agent-based models can easily handle both space and time *simultaneously*. In a realistic model of malignant tumor systems, space must be taken into account explicitly because there are only limited numbers of locations exhibiting an abundance of nutrients and low tissue consistency. Thus, there should be a fierce spatial competition among tumor cells to reside in such favorable locations. On the other hand, time serves as a constraining variable since future prognosis of the host patients are critically dictated by the past history of events. Small changes during the initial stage of tumorigenesis may induce tu-

mor cells to experience a genetic switch (5), which in turn can transform a benign growth into an aggressively expanding, malignant tumor.

From the methodological standpoint, the advantage of using a numerical platform is that it circumvents the need to solve analytically the underlying mathematical equations. That is, instead of relying on theorems and closed-form solutions, the statistical properties of the system are estimated by spinning the model forward a sufficient number of times. Such desirable advantages of an agent-based model, however, do not come without a "price" for the user.

First, as for any other theoretical approach that is based on a numerical platform, calibration of the model parameters using experimental data is often difficult due to (i) the lack of the latter in some cases, or (ii) the fact that they have been collected over a wide range of different experimental setups, thus rendering combinations of the results nontrivial. Without proper prior experiments, however, robustness of the model prediction must be verified by exhaustive exploration of the relevant parameter space, which is often of high-dimensional order and therefore computationally can be very resource intensive. If, in addition, the model also contains stochastic elements that can potentially have a significant effect on the outcome, then a Monte Carlo simulation must be performed across various random seeds to ensure robustness.

Second, even if data are available, translating an in-vivo or in-vitro experimental setup into an operational in-silico model can be a formidable exercise. The challenge is to confine the number of cellular characteristics and environmental variables into a manageable few, preferably those that are most pertinent to the question posed by the researcher. This of course is a critical step and has to be carefully balanced against *over*simplification. Nonetheless, stripping a complex biological organism to its bare essentials is necessary to render any model tractable, which in turn allows one to establish cause–effect relationships.

On balance, however, it is clear that, as long as the researcher is aware of the model limitations, the potential benefits of an agent-based framework far outweigh its deficiencies. Typically, a realistic tumor model exhibits the following features:

- An agent-based model treats both *space and time explicitly* and in a *discrete* manner. Discretized time allows the assessment of tumor progression at various time steps.

- Explicit inclusion of *environmental variables* that have proved critical in guiding tumor proliferation and invasion—such as nutrient sources, mechanical confinements, toxic metabolites, and diffusive biochemical attractants.

- *Variable* grid lattice: allowing more than one cell to share the same location, capturing the spatial and resource competition among the tumor cells themselves. Such cellular clustering arguably guides the overall spatiotemporal behavior of the tumor system (6).

More recently, in an attempt to emulate the in-vivo setup in an even more realistic way, sophisticated agent-based models of brain tumors have introduced further modifications. Among the novel features of these recent models are:

- *Nonlinear feedback* effects through two types of interactions, namely: (i) local interactions among cancer cells themselves, and (ii) interactions between cancer cells and their surrounding environment (6–9) represented by, for example, an adaptive grid lattice.

- *Heterogeneous* cell population as represented by the emergence of distinct subpopulations, each with different cell clones (10).

- *Phenotypic dichotomy* between cell proliferation and invasion as supported by recent experimental findings (11,12).

To put this in perspective, in the following section, we briefly review previous works on tumor modeling.

2. PREVIOUS WORKS

Based on the choice of methodology, most of the existing models of brain tumors derive their results from either solving analytically a set of mathematical equations or from performing Monte Carlo simulations using a numerical platform. The latter follows the longstanding tradition of *cellular automaton* (CA) models, albeit with considerably richer specification of tumor behavior. For an excellent review article of cellular automata approaches to the modeling of biological systems, see (13). In the modeling of complex systems, CA models have proved to be a versatile tool. For example, (14) and (15) treat the automata as abstract dynamical systems. In addition, (16) presents and (17) reviews CA applications that are biologically motivated. (18) presents an early work in tumor modeling that employs a three-dimensional cellular automaton model to investigate tumor growth. In their model, automaton rules were designed to capture the nutritional requirements of tumor growth. Being early in the field, their minimalist model did not yet consider the impact of other important environmental variables such as mechanical confinements and toxic metabolites. More recent efforts include (19), which utilizes a CA model to study growth progression. (20) also constructed a CA model to successfully generate a growth profile of tumor cells that follows the well-known Gompertz law. They examine the dynamics of tumor growth in the presence of immune system surveillance and mechanical stress generated from within the tumor, but they do not explicitly consider the influence of growth stimulants and inhibitors. (21) employs the CA approach to investigate the effects of location-specific autocrine and paracrine factors on tumor growth and morphology. Neither (20) and (21) explicitly consider the mechanics of how tumor cells evaluate the attractiveness of location.

Rather, a static probability distribution function is exogenously assumed to determine whether events such as migration, proliferation, or cell death can occur. Such external rules, imposed in a top–down manner, however, rule out the possibility for virtual cells to be true autonomous "decision-making" agents. All these CA applications typically assume discreteness in time and space. The discrete treatment of time and space is often desirable not only because it models biological systems more realistically but, more importantly, because it allows for examination of tumor progression over time and across space. The latter is invariably the basis of any clinician's prognosis in practice. However, there are many variables of interest that are less discrete in nature, such as nutrient sources, toxic metabolites, and mechanical confinements. For these types of variables, their dynamic evolution in the extracellular matrix can be better described using the continuous Navier-Stokes or reaction-diffusion equations. In cases where the spatiotemporal evolution of the tumor is closely linked to environmental conditions, an approach combining both the discrete (for time and space) and continuum (for environmental conditions) elements of a tumor system would offer a very sensible and promising alternative. For this reason, more contemporary modeling efforts extend the CA framework into *agent-based* models that still simulate time and space discretely, yet treat many of the biological components of interest as continuous variables, thus avoiding the need to transform these variables into unrealistic integer states as in a traditional CA model. Recent contributions that have made use of discrete-continuum intersections include (6–9,22,23).

Another useful method for classifying existing models is determining whether the focus is on the proliferation or migratory behavior of tumor cells, or on both growth and invasion. Many previous studies have focused on either the proliferative growth of the tumor (24,25) or on the invasive behavior (26,27). More recent modeling efforts have attempted to place equal emphasis on both cell proliferation and cell motility. The work that has been done in this area of research with a dual focus on both growth and invasion includes (6–9,28,29).

Proliferation-focused studies can be further subdivided into those that employ deterministic-continuum and those that make use of stochastic-discrete models. In the former, the tumor system is viewed as an aggregation of (multicellular) tumor lumps, and the variable of interest is invariably tumor volume, which is a continuous variable. As an example of a continuum platform, (24) develops a deterministic mathematical framework to generate the growth pattern of multicellular tumor spheroids that follows the Gompertz law. On the other hand, a stochastic-discrete approach typically employs a cellular-level agent-based model to enable the explicit examination of chance elements in the behavior of individual cells. For example, using a three-dimensional cellular automaton model, (25) shows that macroscopic tumor behavior can emerge from local interactions at the microscopic level. (22) presents an attempt to bridge these two approaches by introducing random elements into a continuum model. Note

that, although their model allows cell migration through continuous diffusion, at the core theirs is proliferation-focused, with a cursory treatment on tumor cell invasion. Researches that focus on *invasive* expansion can also be broken down into deterministic-continuum (e.g. (26,27,30)) and stochastic-discrete approaches (e.g. (23)). In a deterministic-continuum model, multicellular patterns typically emerge from the dynamic evolution of population density functions satisfying second-order nonlinear differential equations of reaction-diffusion and wave propagation. For example, (30) successfully simulates the branching pattern on a tumor surface using a continuum model whereby migrating cells follow the gradients of diffusive substrates.

As an attempt to generalize, (28) considers both proliferation and migration within a three-dimensional diffusion framework. All these continuum models emphasize the interaction of cells with the environment, but usually cannot identify the individual cell *itself*. In addition, incorporating stochastic cell behavior within a reaction-diffusion framework is a daunting task. Perhaps more importantly, such models are not suitable for modeling the early stages of tumorigenesis when only a small number of tumor cells are present. At that stage, the progression of tumor growth depends on the discrete history of each individual cell and its local interactions with the environmental variables as well as with neighboring tumor cells. (23) therefore develops a *hybrid* discrete-continuum model to account for the importance of tumor cells to be treated as discrete units. This study shows that the formation of (experimentally observed) "branches" on the surface of a multicellular tumor spheroid may require both heterotype and homotype *chemoattraction*, i.e., toward distinct signals that are released by nutrient sources as well as produced by the tumor cells themselves (i.e., paracrine). (29) employs a discrete model to replicate the spatiotemporal pattern of malignant cell invasion into the surrounding extracellular matrix (ECM). However, in their model cancer cells migrate to minimize their collective energy expenditures, implying that each cell endeavors to minimize the surface energy of the entire tumor domain. Their model therefore does not qualify as a true agent-based framework since the latter by definition must be based on individual-level "decisions" rather than community-level considerations on the part of tumor cells. Nonetheless, their work underscores the critical role of *minimal energy expenditure* for tumor expansion, which also influences our work (6–9).

For example, in the agent-based model presented in (6), due to a cascading information structure, tumor cells gradually "learn" information content about a particular location in two stages. In the first stage, signal content is global (based on the assumption that cells can upregulate their receptor sensitivity) but incomplete, while in the second stage detailed local information is complete. Guiding the migratory behavior of tumor cells is the principle of "least resistance, most permission,[1] and highest attraction," which classifies the attractiveness of the microenvironmental conditions. The key finding of this study is the emergence

of a phase transition leading to two distinct spatiotemporal patterns depending on the dominant search mechanism (related to the cells' energy expenditure required by the particular environmental conditions). In brief, if global search is dominant, the result is an invasive tumor system operating with a few large clusters that expands rapidly but also dies off quickly. By contrast, if local search is dominant, the result is many small cell clusters with a longer lifetime but much slower spatiotemporal velocity. Building on this work, (7) focuses on search *precision* and implements local search only. The main finding there is that a less than perfect search can yield the fastest spatiotemporal expansion, thus indicating that multicellular tumor systems might be able to exploit the in vivo, presumably (chemically and mechanically) rather "noisy" microenvironment. In the following section we describe the mathematical model in more detail.

3. MATHEMATICAL MODEL

In our model, a tumor cell can proliferate, migrate, become quiescence, or undergo cell death depending on (a) its own specific location, (b) other tumor cells sharing the same location and competing for resources, (c) the onsite levels of microenvironmental variables, and (d) the state of neighboring regions. To discuss the nature of the relationships among tumor cells and between tumor cells and their environment in our model, we first introduce a number of theoretical notions.

3.1. Definitions

3.1.1. Population Dynamics

Let $\eta_{j,t}$ represent the (discrete) number of tumor cells residing in location j at time t. In our model, both space and time are discrete, i.e., the x-y coordinates of j and time t are nonnegative integers. The population of tumor cells in any location changes due to (i) proliferation of new offsprings, (ii) net migration (i.e., in-migration minus out-migration), and (iii) cell death.

3.1.2. Clustering

A cluster is the *spatial agglomeration* of virtual tumor cells in contiguous sites. These virtual clusters represent cell aggregates observed in actual experiments involving malignant brain tumors (3,31). More formally, we define a cluster of tumor cells, C, as a federation of contiguous regions, each of which contains at least one viable tumor cell $\eta_{j,t} > 1$. To this, we add another qualification that for a group of regions to be a legitimate cluster C it must be the case that their collective population size exceeds a certain minimum size: the union

of contiguous, nonempty locations $\bigcup_{j\in C}\eta_{j,t}$ represents a cluster if and only if $\sum_{j\in C}\eta_{j,t} > \overline{\eta}$. For example, in (7,8), we fixed $\overline{\eta}$ = 5, which means an agglomeration of more than five adjacent cells (i.e., their locations share a common border) qualifies as a cluster. A tumor cell is defined to be on the surface of a cluster if (i) its location j belongs to a cluster, i.e., $j \in C$, and at the same time (ii) there must exist an empty location I in j's neighborhood, formally: $\eta_{i,t} = 0$ and $i \in j$'s neighborhood.

3.1.3. Measure of Distance

Since our model explicitly takes the geography of a brain tumor into consideration, a distance measure is necessary. We chose the L-infinity metric of distance for the simple reason that it can be conveniently implemented in a computer algorithm. Specifically, given two points A and B in the two-dimensional grid lattice with coordinates (x_A,y_A) and (x_B,y_B), respectively, then the distance between these two points is computed as d_{AB} = Max[abs($x_B - x_A$),abs($y_B - y_A$)].

3.1.4. Local Neighborhood

The set of locations, which are adjacent to a tumor cell's current site, constitute the local neighborhood of that cell. In our previous works, the local neighborhood consists only of those locations sharing a common border with the tumor cell's current location j, i.e., those locations i that are within one unit of distance away: $d_{ij} \le 1$. Typically, we adopt the notion of a *Moore* neighborhood, which includes locations in the north, south, east, and west of the tumor cell's current location, as well as in the NE (northeast), NW (northwest), SE (southeast), and SW (southwest) directions.

3.2. Cell Behavior and Environment

The local environmental variables in our model are captured by nutrient supplies, mechanical confinements, and deposits of detrimental toxic metabolites. The generic term "nutrients" can be interpreted as representing glucose, as it is the principal source of energy for the brain. Indeed, (32) demonstrated that the extent of glioma malignancy is highly correlated with the expression of the GLUT3 glucose transporter. However, nutrients here can be also interpreted, for example, as epidermal growth factor (EGF), which has been shown to stimulate and guide the invasion of glioma tumor cells in vitro (33). The active migration

of tumor cells toward these EGF signals is facilitated by the cells' specific epidermal growth factor receptor (EGFR), known to be overexpressed in primary human glioblastomas (34). In the case of glucose, tumor cells convert their nutrient uptake to lactic acid. Thus, relative to normal tissue, tumor tissues are often found to experience a several-fold increase in both glucose uptake and lactic acid production (35,36). The result of the escalating production of lactate and hydrogen ions is a thin layer of an acidic environment that surrounds tumor cell colonies. For example, (37) reports decreasing levels of pH, while (38) reports decreasing levels of pO_2 on the surface of tumor spheroids. Both the lower level of pO_2 and the reduced pH should render the microenvironment less viable and thus less "attractive." Finally, mechanical confinements in our model represent the stress or pressures exerted by the surrounding tissues, which hampers the ability of tumor cells to grow and invade the parenchyma. Experimentally, a number of studies have shown that these mechanical properties of the tissue environment indeed influence both the proliferating and migratory behavior of tumor cells (39, 40).

3.2.1. Proliferation

Either proliferation or migration of tumor cells is allowed to occur if a number of criteria are fulfilled. The algorithm proceeds as follows. First, for every viable tumor cell, we determine whether its location j belongs to a "legitimate" tumor cluster (see §3.1.2 above for details). If the tumor cell does not reside in a cluster, then it is eligible to either proliferate or migrate, though not both. If, on the contrary, the cell is a member of a cluster, then we next check whether the tumor cell is located on the surface of a cluster. If the "cluster surface" condition is satisfied, a tumor cell can proliferate if its onsite levels of nutrients and toxic metabolites are within certain demarcating thresholds (see Figure 1). Let ϕ_j and τ_j denote site j's level of nutrients and toxic metabolites, respectively. Then proliferation may occur if the level of nutrients is higher than the upper nutrient threshold, $\phi_j > \phi_U$, while at the same time, the level of toxic metabolites is below the lower toxicity threshold, $\tau_j < \tau_L$.

Even if the environmental conditions are favorable (i.e., high nutrients and low toxicity), there is still a chance that proliferation will fail to occur. We model this element of stochasticity by assuming that the probability to proliferate is proportional to the onsite levels of nutrients:

$$\Pr_{\text{proliferate},j} = \phi_j / (\phi_j + k_{\text{prolif}}), \qquad [1]$$

where k_{prolif} represents a parameter that controls the likelihood of cell proliferation. Higher k_{prolif} implies a lower *probability* to proliferate because it is inversely proportional to the capability of tumor cells to proliferate. Equation [1] states

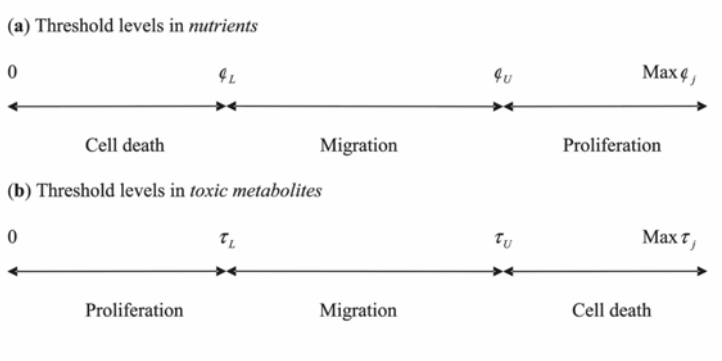


Figure 1. Dual threshold levels in (a) nutrients and (b) toxic metabolites.

that even when nutrients are sufficient and toxic metabolites are low, due to the probabilistic nature of proliferation there are some eligible agents that generate no daughter cells. Note that at very high levels of nutrients, the probability of proliferation approaches unity: $Pr_{proliferate,j} \rightarrow 1$ as $\phi_j \rightarrow \infty$. Conversely, even when the environmental conditions are not sufficient (i.e., due to $\phi_j \leq \phi_U$ or $\tau_j \geq \tau_L$), we may admit a small probability for the disadvantaged cells to proliferate, hence allowing for an element of cellular "adaptation" under a less-than-optimal environment.

3.2.2. Migration

Tumor cells are allowed to migrate and thus invade adjacent regions if: (a) they are eligible to proliferate but do not due to chance, or (b) they are not eligible to proliferate because of lesser environmental conditions, yet they reside on the tumor surface and their location exhibits one of these three conditions: (i) $\tau_j \leq \tau_L$ but $\phi_L < \phi_j < \phi_U$, or (ii) $\phi_j \geq \phi_U$ but $\tau_L < \tau_j < \tau_U$, or (iii) $\phi_L < \phi_j < \phi_U$ and $\tau_L < \tau_j < \tau_U$. Notice that in our model no tumor cell can perform both proliferation and migration at the same time; it must be one or the other. This trait *dichotomy* has been shown experimentally. Specifically, (11) shows that a tumor cell at a given time and location experiences only one of the key activities (either proliferation or migration) at the maximum level, but not both. More recently, the same authors supported their conclusion reporting distinct gene-expression profiles for both phenotypes (12).

3.2.3. Search Process

At each time period t = 0, 1, 2, ..., a tumor cell that is eligible to migrate then "screens" the surrounding regions in its neighborhood to determine whether there is a more attractive location. The biological equivalent for this mechanism (and the cell behavior it induces) is cell-surface "receptor–ligand" interaction. Each eligible tumor cell then ranks the *attractiveness* of a neighboring location based on the following real-valued function:

$$L_j(G,\phi_j,\tau_j,p_j) = G_j + q_\phi \cdot \phi_j - q_\tau \cdot \tau_j - q_p \cdot p_j + \sum_{i\in\{j\text{'s neighborhood}\}} G_i \,. \qquad [2]$$

Equation [2] states that the value of location j depends on (i) the function G_j (see Eq. [3] below), which represents this location's attractiveness, which is due to its cell population, as well as on (ii) the onsite environmental factors. Specifically for the latter, the parameters q_ϕ, q_τ, and q_p capture the contributions of nutrient supplies ϕ_j, toxic metabolites τ_j, and mechanical pressures p_j, respectively. The last term in Eq. [2] captures the "neighborhood effect" due to cells that exert influence, and are influenced by other cells located in adjacent locations. As we have detailed in §3.1.4, our previous works have utilized the concept of the "Moore neighborhood," which includes only those adjacent locations at most one unit of distance away. The explicit form of G_j here is specified as a non-monotonic function of the population density, η_j, discounted by the distance of location j from the evaluating tumor cell:

$$G_j = (\eta_j - c\,\eta_j^2)\exp(-\rho \cdot d_j^2/2). \qquad [3]$$

According to Eq. [3], a tumor cell is attracted to locations that already accommodate a number of other tumor cells, implementing the biological concept of a "paracrine" attraction. However, there is also a negative "crowding out" effect represented by the parameter c, such as if the location j's population of tumor cells expands beyond a maximum (i.e., beyond the point where $\partial G_j/\partial \eta_j = 0$), then the attractiveness of that location starts to decline due to, for example, limited carrying capacity and spatial competition. Because of such a maximum threshold cell density, G_j is nonmonotonic: first, it increases, then at the maximum $\eta_j = \eta_{\max}\big|_{\partial G_j/\partial \eta_j = 0}$ it starts to decrease as the population of tumor cells grows further. Importantly, Eq. [3] also contains a geographical dimension since the value of G_j is discounted at an increasing rate proportional to its squared distance d_j^2 from the evaluating tumor cell's current location. The parameter ρ thus captures the metabolic energy required for a single cell to move across regions.

We postulate that such *energy expenditures* are determined by both an intrinsic factor and an external effect:

$$\rho = r_\rho p_j, \quad [4]$$

where the intrinsic factor r_ρ corresponds to the inverse capability of tumor cells for spatial movement, while the external effect is attributable to mechanical confinements p_j. Note that higher values of r_ρ imply greater costs of spatial movement in terms of cells' energy expenditures; it is thus an inverse capability term.

3.2.4. Cell Death

As shown in Figure 1, cancer cells that experience dwindling supplies of nutrients or escalating levels of toxic metabolites can either turn quiescent, a reversible state, or undergo cell apoptosis/death. In our model, cell death is a stochastic event such that if either of two conditions occur—(i) nutrient reserves fall below the lower threshold $\phi_j < \phi_L$, or (ii) levels of toxic metabolites rise above the upper threshold $\tau_j > \tau_U$, then for tumor cells residing in such locations, the likelihood of death becomes positive and is proportional to the toxicity levels:

$$\Pr_{\text{death},j} = \tau_j/(k_\tau + \tau_j), \quad [5]$$

where k_τ is the parameter representing the inverse sensitivity of cell death to toxicity level. Note that higher values of k_τ correspond to lower probability of death; hence the "inverse" term. Other than quiescence, cell death is a nonreversible event. In fact, cell death is particularly imminent for the nonproliferating and non-migrating quiescent cells trapped inside a cluster due to their inability to escape to more favorable locations. Eventually, these dead cells start to form an emerging central necrotic region, a hallmark of highly malignant brain tumors (25).

3.2.5. Nutrient Sources

The supplies of nutrients in our model can be either *replenished* (e.g., through neighboring blood vessels) or *non-replenished* (representing scattering traces of nutrient in the intercellular space). In every period t, the change in the levels of nutrients in every location is due to the current (new) production of nutrients, $g_\phi\phi$, diffusion from the surrounding lattice sites, $\nabla\cdot(D_\phi\nabla\phi)$, and nutrient depletion, $r_\phi\eta$:

$$\frac{\partial\phi}{\partial t} = g_\phi\phi + \nabla\cdot(D_\phi\nabla\phi) - r_\phi\eta\,. \quad [6]$$

where the parameter g_ϕ represents the fixed rate of nutrient production, D_ϕ is the diffusion coefficient of nutrients, and r_ϕ controls how fast a tumor cell metabo-

lizes the on-site nutrient sources. For replenished sources of nutrients we set $g_\phi > 0$, while for non-replenished sources $g_\phi = 0$. Nonetheless, this setup allows for a feedback between the two distinct types of nutrient sources: the initially non-replenished intercellular nutrients may get recharged through diffusion from the replenished source.

3.2.6. Mechanical Confinements

As a first approximation, in (6), we assumed a static distribution of mechanical confinements. It is a "static" distribution in the sense that the levels of tissue resistance remain constant over time regardless of cell behavior. Subsequently, in (7), we adopted the more realistic notion of an *adaptive grid lattice* such that locations that have been traversed by migrating tumor cells experience a reduction in mechanical confinements:

$$\frac{\partial p}{\partial t} = -r_p \eta \,. \qquad [7]$$

Equation [7] specifies that mechanical confinements "decay" at the rate of r_p per viable cell regardless of its phenotype, so that for a given constant cell population $\overline{\eta}_j$, mechanical confinements go to zero after $p_{t,j} /(r_p \cdot \overline{\eta}_j)$ time steps. Over the course of a simulation, however, it is rarely the case that the cell population remains constant, and thus the time when mechanical confinements disappears at a particular lattice site is a stochastic variable. Biologically, the fall in mechanical confinements represents the degeneration of extracellular matrix due to cell invasion and secretion of proteases, i.e., matrix-degrading enzymes (41). According to Eq. [7], the presence of any viable tumor cell (regardless of phenotype) reduces tissue consistency. However, we argue that only invasive cells are capable of taking advantage of the deformed grid lattice by following the path of declining resistance. Assuming an underlying tendency of invasive cells to limit their energy expenditure as captured by Eq. [4], this process would in turn encourage even more tumor cells to invade the host tissue further following the paths that traversed by their peers.

3.2.7. Toxic Metabolites

In our model, the levels of toxic metabolites (which here can represent a combination of specific inhibitory soluble factors released by tumor cells, lysosomal content from dying tumor cells or tissue hypoxia) evolve according to the following function:

$$\frac{\partial \tau}{\partial t} = \nabla \cdot (D_\tau \nabla \tau) + r_\tau \eta \,. \qquad [8]$$

In other words, the change in the level of toxicity at time t, $\tau_{t,j}$, is due to (i) diffusing toxicity from the surrounding regions, $\nabla\cdot(D_\tau\nabla\tau)$, where D_τ is the diffusion coefficient of toxic metabolites, and augmented by (ii) the rate of toxic accumulation, r_τ, multiplied by the population of tumor cells in that location, η_j. The last term thus contains the assumption that a greater population of tumor cells leads to faster accumulation of toxic metabolites. In fact, (42) has shown that in EMT6-cells detrimental accumulation of toxicity in the form of lactic acid results in inhibition of onsite proliferation. Yet at the same time, the authors of (43) found that higher toxic levels can also stimulate active migration of tumor cells, and that, although hypoxia induces growth arrest in ovarian carcinoma cells, they still exert proteolytic (Type IV collagenase) activity, which is required to maintain their invasive properties. Similarly, (44) described recently a hypoxia-induced migration of human U-138MG glioblastoma cells using an in-vivo model. Together, these studies indicate that migration is stimulated when onsite accumulation of toxicity acts as repulsion, ultimately forcing tumor cells out of their current location.

4. SPECIFICATIONS OF THE MODEL

In recent work, we have extended the "core" agent-based modeling platform above to examine specific scientific questions.

4.1. Search Mechanism

In (6), we propose the concept of *global vs. local* search. This theoretical notion represents the existence of two different cell-surface receptors directing the chemotactic movement of virtual tumor cells with two distinctively different lower signal detection thresholds (or with two distinct intracellular amplification strengths). The first type of receptor, employed during global search, can be thought of exhibiting a lower signal detection threshold and as such is more sensitive to diffusive signals emitted from distant locations. On the other hand, the second type of receptor (involved during local search) exhibits an elevated level of the lower signal detection threshold and thus is arguably employed to capture stronger signals coming from the local neighborhood of a cell's current location. The potential tradeoff between global and local search is captured by the parameter r_ρ. Smaller values of r_ρ imply lower energy costs of spatial movement, ρ (see Eq. [4]), and thus increase the scope of global search by conferring higher mobility to tumor cells. In contrast, larger values of r_ρ promote a shift toward a more local search in the neighborhood of the cells' original location due to the higher costs of spatial movement in terms of energy expenditures.

4.2. Search Precision

In (7), we introduced *noise* into cellular signal reception, such that the extent of the noise in the signal is captured by the search precision parameter, denoted as Ψ, which is positive and between zero and one, $\Psi \in [0,1]$. Here, the search precision Ψ represents the likelihood with which tumor cells evaluate the attractiveness of a location without error. As Eq. [2] specifies, the error-free value of a location is jointly influenced by the onsite levels of nutrient, mechanical confinement, and toxicity. Formally, let T_j be the attractiveness of location j as evaluated by a tumor cell using its signal receptors, and let L_j be the error-free evaluation of location j. The extent of search precision is introduced to the model such that, due to the noise in the signals, the attractiveness of location j is evaluated without error only Ψ proportion of the time:

$$T_j = \Psi \cdot L_j + (1 - \Psi) \cdot \varepsilon_j, \qquad [9]$$

where $\varepsilon_j \sim N(\mu,\sigma^2)$ is an error term that is normally distributed with mean μ and variance σ^2. As a concrete example, a 70% search precision implies that Prob $[T_j = L_j] = 0.7$, i.e., the attractiveness of location j is evaluated without error in seven out of ten trials on average. At one extreme, $\Psi = 1$ represents the case where tumor cells consistently evaluate the permissibility of a location without error. At the other extreme, $\Psi = 0$ represents the case when tumor cells always perform a *random-walk* motion, thus completely ignoring the guidance of the gradients of environmental variables.

4.3. Structure–Function Relationship

In (8), we investigated the emerging structural patterns of a multicellular tumor as represented by the *fractal dimensions* of the tumor surface. We then examine the link between the tumor's fractal dimensions and the cancer system's dynamic performance (i.e., functionality) as captured by its average velocity of spatial expansion. The tumor's fractal dimension, which characterizes the irregularities of the tumor surface, is determined using the box-counting method (45). This method quantitatively measures the extent of surface roughening at the tumor–stromal border due to both proliferation and migration of malignant tumor cells. The choice of fractal dimensions as a measure of structural pattern is motivated by the idea that the morphology of a tumor surface depends on the scale of observation. Let $SA(l)$ be the entire surface area of the tumor that is computed by counting the number of boxes, $N(l)$, each of size s, which are needed to cover the entire area. Then it is the case that $SA(l) = N(l) \cdot l^2$. If the dimension of the tumor surface is indeed fractal, then we will find that:

$$N(l) \sim l^{-d_f}, \qquad [10]$$

where $d_f = \lim_{s \to 0}[\ln N(l) / \ln (1/l)]$ stands for the fractal dimension of the tumor surface.

4.4. Molecular Level Dynamics

In (9), we augment our 2D agent-based model with the molecular level dynamics of alternating *gene expression* profiles. Specifically, in that study we analyze the impact of environmental factors on gene expression changes, which in turn have been found to accompany the phenotypic cellular "switch" from proliferation to migration. For reasons of tractability, we focus on the behavior of two genes, namely Tenascin C and PCNA, which have been chosen on the basis of their reportedly active roles during proliferation and migration of glioma cells. Tenascin C is an extracellular matrix glycoprotein overexpressed in malignant in-vivo gliomas. (46) has shown that tumor cell motility is stimulated when human SF-767 glioma cells are placed on Tenascin C. On the other hand, PCNA stands for "proliferating cell nuclear antigen," whose gene expression markedly rises during neuroepithelial cell proliferation (47). Experimental findings have shown that an increase in the gene expression of Tenascin (hereafter *gTenascin*) is associated with an increase in both levels of nutrients and tissue hypoxia, or (for our purposes, in more general terms) toxicity. Accordingly, in our model, *gTenascin* is computed as a simple continuous positive function of both the normalized levels of nutrients $\hat{\phi}_j$ and toxic metabolites $\hat{\tau}_j$: *gTenascin* = $b_T \hat{\phi}_j \hat{\tau}_j$. On the other hand, the literature suggests that an increase in the gene expression of PCNA (hereafter *gPCNA*) is associated with an increase in nutrients and a *decrease* in toxicity. Accordingly, in our model we compute *gPCNA* as a positive function of nutrients yet negatively affected by toxicity, *gPCNA* = $b_T \hat{\phi}_j / \hat{\tau}_j$. The molecular modules of both *gTenascin* and *gPCNA* enable our agent-based model to generate a virtual *time-series profile* of both gene expressions as they relate to the proliferative, migratory and quiescent tumor cell phenotype. A particular aim of a time-series analysis (see also Part II, chapter 1, by Shalizi, this volume) is to ascertain whether a dynamic series exhibits intertemporal long-range autocorrelations. The presence of such autocorrelations indicates the potential use of past historical values to forecast future outcomes. For that purpose, we applied *detrended fluctuation analysis* (DFA), which Peng et al. (48) developed as a robust method to detect long-range correlations in various DNA sequences. If the statistical properties of a time series exhibits a *random walk* (no autocorrelations across time), then DFA would yield an autocorrelation measure $\alpha = 0.5$. In contrast, DFA would detect a long-range autocorrelation by a value of α that significantly deviates from the random walk value, i.e., $\alpha \neq 0.5$.

5. BASIC MODEL SETUP

In our previous works, our space of observation is a torroidal square grid lattice representing a virtual, two-dimensional slice of brain parenchyma. On that grid lattice we introduce fields of environmental variables: nutrients, toxic metabolites, and mechanical confinements. As we have discussed briefly above in §§3.2.1, 3.2.2, and 3.2.4, we impose dual thresholds in the levels of nutrients (ϕ_L,ϕ_U) and toxicities (τ_L,τ_U), which in turn trigger the onset of migration, proliferation, and cell death. To summarize, a tumor cell in location j proliferates if $\phi_j > \phi_U$ and $\tau_j < \tau_L$, i.e., sufficiently high levels of nutrients (above upper threshold) and low levels of toxicity (below lower threshold) lead to a positive probability of proliferation, the extent of which in turn is proportional to the levels of nutrients. If these optimal conditions for growth are not satisfied, then a tumor cell can migrate whenever it is located on the tumor surface, and one of these three conditions hold: (i) $\tau_j \le \tau_L$ but $\phi_L < \phi_j < \phi_U$, or (ii) $\phi_j \ge \phi_U$ but $\tau_L < \tau_j < \tau_U$, or (iii) $\phi_L < \phi_j < \phi_U$ and $\tau_L < \tau_j < \tau_U$. Whether the cell actually migrates depends on whether there is a more attractive location in its local neighborhood as determined by the combined impact of nutrients, mechanical resistance, and toxicity. Finally, if either the nutrient levels are below the lower limit, $\phi_j < \phi_L$, or the toxic levels are above the upper threshold, $\tau_j > \tau_U$, then cell death is imminent with a probability proportional to the levels of toxic metabolites. In the following, we detail how the environmental variables are initialized in our numerical model.

5.1. Initial Levels of Environmental Variables

5.1.1. Nutrients

The initial distribution of nutrients is modeled as follows. We assume that there are three distinct types of nutrient sources that can be distinguished based on their geographical distribution. Specifically, there are two *replenished* primary sources of nutrients in the initial setup: (i) at the center of the grid, and (ii) at the center of the northeast (NE) quadrant. Everywhere else, nutrients are initially *non*-replenished. In order to add to the site's growth-permissive environmental conditions, the first primary source is placed next to the "crater" of mechanical confinements and bell-shaped distributed with the peak located at the tip of the pressure crater. The second source can be found to the NE of the grid center and is also bell-shaped distributed with the peak level located at the center of the NE quadrant. This second nutrient source exhibits significantly higher levels than the first one such that the peak of this second source is five times (5×) the peak of the first one. This arrangement ensures a chemoattractive gradient where the two sources can be thought of as representing two distinct-

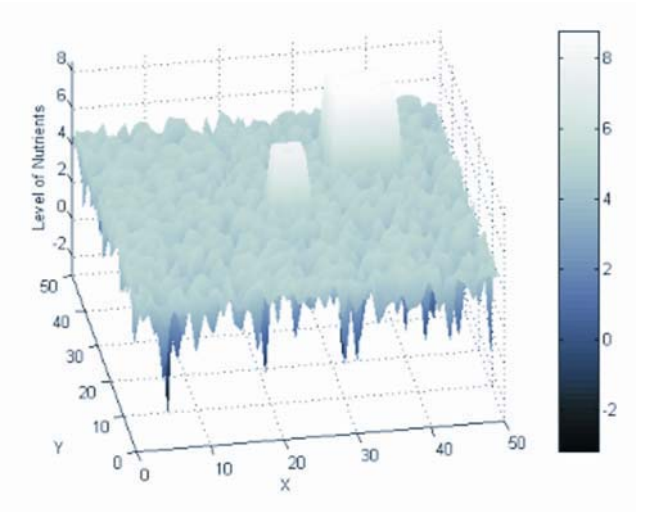


Figure 2. Initial distribution of nutrients in a 50 × 50 square grid lattice. The XY axes represent the two-dimensional spatial coordinates. The twin mounds correspond to the two replenished nutrient sources: the first peaks at the tip of the mechanical pressure crater (see Figure 3), the second at the center of the NE quadrant. The vertical bar on the right shows the color scale used to measure the levels of nutrient sources.

sized blood vessels within the brain parenchyma. In addition to these two primary sources, there are non-replenished ("interstitial") nutrient substrates that are distributed randomly in a uniform manner at a much lower level than the first source, such that the minimum nutrient level of the first source is fifty times (50×) larger than the maximum level of the non-replenished site. Figure 2 illustrates the initial distribution of nutrients. As the simulation progresses, nutrients diffuse from replenished sources to non-replenished ones due to tumor-induced alterations in mechanical confinements (see Eq. [7]).

5.1.2. Mechanical Confinements

In several of our previous works (7–9), we assume that the center of the square lattice (where the initial seed of tumor cells is placed) corresponds to a small "crater" of mechanical confinement. In this stress or pressure crater, mechanical confinements are relatively low, reflecting a growth-permissive anatomical condition at the initial site of tumorigenesis, with the lowest pressure located at the center of the lattice. In (6) we assume that the peak

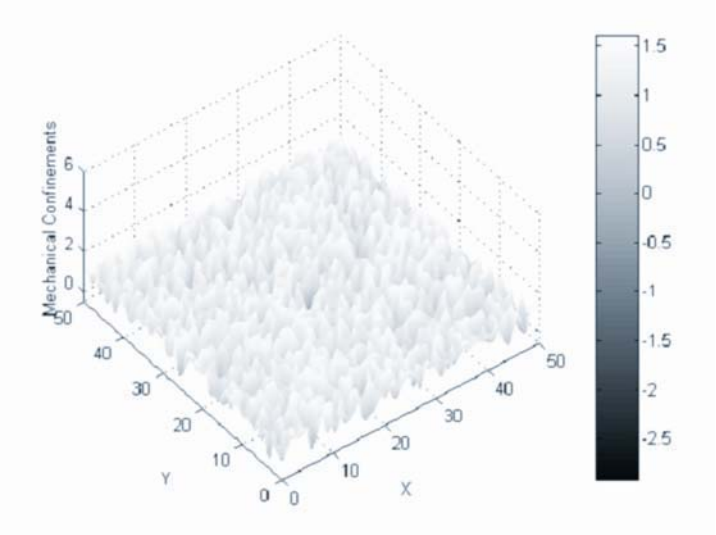


Figure 3. Initial levels of mechanical confinements in a 50 × 50 square grid lattice. The XY axes correspond to the spatial coordinates of the grid locations. The center of the pressure "crater" is placed at coordinates (x = 25, y = 25). The vertical color bar on the right shows the *grayscale* used to measure the intensity of mechanical pressures.

pressure corresponds to the site of the second nutrient source at the NE quadrant (Figure 2). Thus, in effect, the second nutrient source is "defended" by higher mechanical confinements, representing a normal blood vessel within a noncancerous parenchymal environment. Our subsequent works assume that, outside of the low-pressure crater, mechanical confinements are distributed randomly in a uniform manner. Figure 3 illustrates the initial levels of mechanical confinements as assumed in (7) and (8).

5.1.3. Toxicity

The initial levels of toxicity can be assumed to be zero everywhere on the lattice (6–8), or we can introduce an initial uniform distribution of hypoxia that is already in the system from the beginning (9). *Hypoxia* (due to limitation of oxygen diffusion in tissue) is known to be an important component of tumor biology. For example, it has been shown that the growth of multicellular spheroids is adversely affected by diffusion-limited pO_2 (49). Our "initial hypoxia" setup was inspired by the fact that locations farther away from blood vessels

experience hypoxic conditions due to a lack of oxygen related to its diffusion limit of 100–200 μm in tissue (50).

5.2. Structural Measures

The structural, macro-level characteristics of the tumor system can be captured by: (i) the *size* of the tumor clusters (within the invasive system), (ii) the *fractal dimension* of the tumor surface, and (iii) the average overall tumor *diameter*. The average size of tumor cell clusters is computed as the total number of viable (alive) tumor cells divided by the total number of clusters in the entire tumor system.

5.3. Performance ("Functional") Measures

We have examined the following quantifiable measures characterizing the performance of the tumor system: (i) the time it takes for the first tumor cell to invade a nutrient-rich region, and (ii) the lifetime of the tumor system. The first performance measure tracks the time that elapsed between $t = 0$ and the moment where the first tumor cell reaches the peak of the second nutrient peak (see Figure 4). A related measure is the *average velocity*, $\langle v \rangle$, of the tumor system, which we compute as $\delta_{\text{init,quad3}}$ (i.e., the distance between the initial location of the tumor cells' seed and the peak of the second nutrient source at the center of the NE quadrant), divided by the time it takes for the first tumor cell to invade the peak of the second nutrient, $\langle v \rangle = \delta_{\text{init,quad3}}/t_{\text{quad3}}$. Because the numerator is constant in any given simulation, $\delta_{\text{init,quad3}} = \delta$, the average velocity is directly yet inversely proportional to the time for the first cell to reach quadrant three, $\langle v \rangle \approx 1/t_{\text{quad3}}$. The *lifetime* of the tumor system spans between $t = 0$ and the time where the last tumor cell finally dies off due to complete depletion of nutrient sources and detrimental accumulation of toxic metabolites. The simulation is thus terminated when there are no longer viable cells in the system (i.e., in biological terms, when the entire tumor system has turned necrotic).

6. RESULTS

Typically, with a total of 16 parameters, exhaustive exploration of the entire parameter space is nearly impossible. It is thus important from the outset to confine the parameters of interest into a manageable few. For that purpose, we have kept a number of parameters fixed in all of our previous works. Those constant parameters and their specific values are listed in Table 1.

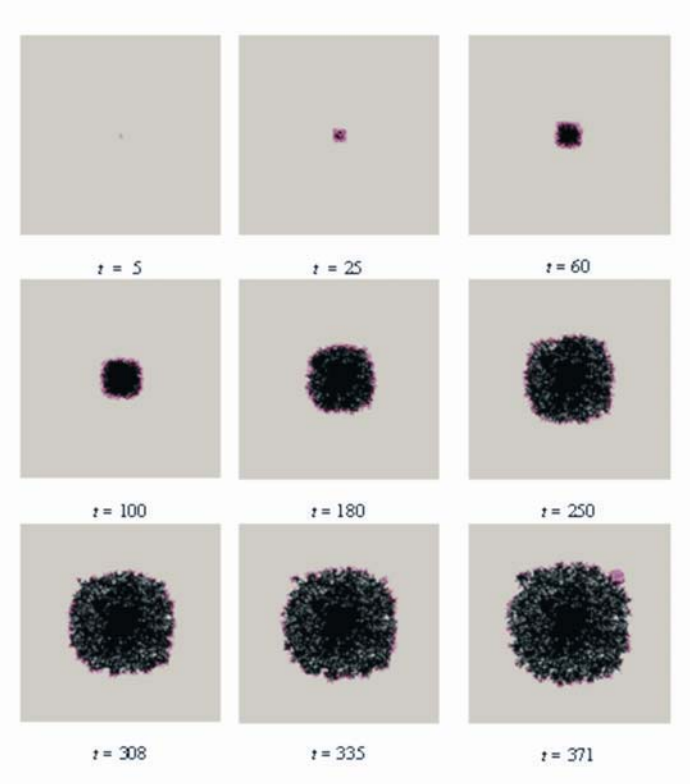


Figure 4. Spatiotemporal progression of a virtual malignant brain tumor in a 200 × 200 square grid lattice. At t = 25, an expansive rim (*orange* as depicted), consisting of both invasive and proliferative cells, starts to emerge on the tumor surface. At t = 60, the proliferative rim has become prominent, while at the tumor core a black necrotic region has been established. At t = 335, a "bulging" macrostructure at the NE tip starts to emerge. At t = 371, this "bulge" becomes a prominent structural feature just prior to the tumor's successful invasion of the second nutrient source.

In the following, we discuss the parameters that we have examined in our previous work to establish robust numerical results.

6.1. Search Mechanism

In (6), we employ a 2D model in which tumor cells are capable of searching both globally and locally. We performed simulations for various values of (i) the (inverse) intrinsic capability to migrate, r_ρ, and (ii) the rate of nutrient depletion,

Table 1. Fixed parameters in the model along with their corresponding denotation and specific values

Parameter	Equation	Value
The positive impact of nutrients, q_ϕ	2	10
The negative impact of toxicity, q_τ	2	–10
The negative impact of resistance, q_p	2	–10
The negative effect of overcrowding, c	3	0.1
Inverse sensitivity of cell death to toxicity level, k_τ	5	10
The rate of nutrient production, g_ϕ	6	0.1
The rate of nutrient depletion, r_ϕ	6	0.01
Diffusion coefficient of nutrients, D_ϕ	6	0.001
The rate of mechanical resistance reduction, r_p	7	0.001
The rate of toxicity accumulation, r_τ	8	0.02
Diffusion coefficient of toxic metabolites, D_τ	8	0.001

r_ϕ. As discussed in §4.1, the parameter r_ρ captures the extent of local search relative to global search. In particular, as $r_\rho \to \infty$, global search is completely eliminated, leaving only local search as the sole probing mechanism. At the other extreme, as $r_\rho \to 0$, global search becomes the dominant *cell receptor mechanism* since tumor cells can invade any location with no spatial constraint. Therefore, with lower values of r_ρ we expect to see acceleration of the tumor's average velocity since it takes less time for the first tumor cell to reach the second nutrient source. Interestingly, we found that the performance of the tumor system exhibits a *phase transition* at a critical value r_ρ*. At $r_\rho > r_\rho$*, where local search is the dominant probing mechanism, raising r_ρ results in slower average velocity yet longer lifetime of the tumor system. In contrast, at lower values $r_\rho < r_\rho$*, where global search dominates, incrementing r_ρ by a small amount (and thus introducing a modicum of local search) actually increases the average velocity (i.e., shortens the time it takes for the first tumor cell to invade the peak of the second nutrient source) yet decreases the lifetime of the tumor system. The term "phase transition" here thus corresponds to the nonlinear behavior of the tumor system: starting at low values of r_ρ, increasing that value will accelerate the tumor's spatial expansion, yet there is a threshold level r_ρ* beyond which the velocity starts to decline as r_ρ increases further toward maximum local search. The choice of r_ρ also reflects the potential tradeoff between the velocity of spatial expansion (i.e., a clinically relevant measure for the aggression of the tumor system) and the

lifetime of the very same tumor, making it more difficult to rank the *fitness* of the tumor system itself. Another parameter that we chose to examine in (6) is the rate of *nutrient depletion* r_ϕ (see Eq. [6]), because the metabolic uptake of tumor cells is an important parameter of interest from the experimental standpoint. We found that *self-organizing* behavior emerges as we simultaneously vary both r_ρ and the rate of nutrient depletion, r_ϕ. Specifically, we show that at slower metabolism rates, $r_\phi < r_\phi^*$, raising r_ρ (i.e., encouraging more local search) results in self-organization of tumor cells into increasingly smaller clusters. The formation of smaller cluster sizes, however, disappears as r_ϕ is increased even further. That is, when nutrients are rapidly depleting due to the high metabolism rates $r_\phi > r_\phi^*$, tumor cells reorganize into clusters that are insensitive to variations in r_ρ. Perhaps most interestingly, it turns out that this self-organization behavior at a high cellular metabolism rate leads to improved performance of the entire tumor system by way of both accelerating its average velocity and also longer lifetime. Although a faster average velocity is to be expected since tumor cells are "forced" to migrate as nutrient levels experience a rapid depletion, from a biological perspective it is intriguing that such aggressive virtual tumor cells actually survive longer under these rather adverse microenvironmental conditions.

6.2. Search Precision

In (7), we vary the search precision parameter, $\Psi \in [0,1]$, to study the impact of cellular signal sensitivity on the performance of the tumor system. As defined in Eq. [9], higher Ψ represents the proportion of time in which tumor cells correctly assess the attractiveness of a location. We found that, unexpectedly, the maximum average velocity of the tumor systems always occurs at *less* than 100% search precision. At the outset, one would expect that a strictly nonrandom search procedure (with $\Psi = 1$) should optimize the average velocity of the tumor system. In fact, we found that, although initially it is true that decreasing randomness results in increasing average velocity, there is a threshold level beyond which the velocity starts to decline if randomness is reduced further. Such a *phase transition* corresponding to a 70% search precision (i.e., 30% chance of committing an error in signal reception or processing) actually elevates average velocity to its maximum, and hence yields optimal performance of the tumor. We also experimented with varying both the extent of search precision Ψ and k_{prolif}, the parameter controlling the probability to proliferate (see Eq. [1]). Recall that higher values of k_{prolif} implies lower proliferation rates as it renders tumor cells less likely to produce offsprings. As expected, spatial velocity increases even more at higher values of k_{prolif} due to the dichotomy assumption between proliferation and invasion as supported experimentally (11). What is not expected is that increasing k_{prolif} is also accompanied by a shift toward higher search precision in order to reach maximum velocity; this is an *emergent* behav-

ior that was not hard-coded into the algorithm implementation. Nonetheless, it was never the case that maximum velocity occurs under a 100% search precision. As $k_{prolif} \to \infty$, proliferation becomes virtually impossible, and the entire tumor simply dies out before it is able to reach the second nutrient source.

6.3. Structure–Function Relationship

From previous in-vitro studies, it has been suggested that malignant gliomas form an invasive branching structure on their surface (3), most likely with higher fractal dimensions than noncancerous tissues, as has been shown also for other cancers (51,52). This complex pattern appears to function as a facilitator of rapid tissue infiltration into the surrounding brain parenchyma. The precise mechanism underlying this "structure–function" relationship, however, remains unclear. In vivo, the difficulty is imposed by the limited spatial resolutions of current imaging techniques that prohibit the monitoring of the structure–function relationship in brain tumor patients over several time points. The contribution of (8) is, using an agent-based model, to show that a numerical analysis can satisfactorily reveal the hypothetical link between the observed structural patterns of malignant brain tumors and their functional properties within a multicellular framework. We investigated the relationship between the structure of the tumor surface, measured by its fracticality d_f (see Eq. [10]), and the neoplasm's dynamic functional performance, measured by the average expansive velocity. As expected, the tumor accelerates its spatial expansion when the "rewards" for tumor cells following their peers along traversed pathways increase. "Rewards" refer to lesser energy expenditure of the succeeding cells and can be thought of as being directly related to, for instance, the secretion of tissue-degrading proteases, and hence represent a "molecular" dimension. Yet, surprisingly, such an increase in average velocity is also accompanied by a concomitant increase in the tumor's surface fracticality, indicating an *emerging structure–function relationship* that is not inherently assumed at the cellular level. Interestingly, using our model we found no correlations between tumor diameter and its surface fracticality; the former increases in an almost monotonic manner, while the latter shows a complex nonlinear behavior marked by intermittent peaks and troughs.

6.4. Molecular Level Dynamics

Finally, combining this *micro–macro* platform explicitly with the *molecular* modeling level (9), we perform time series of the gene expression of both Tenascin and PCNA. At the *macroscopic* level, Figure 4 shows the progression of the tumor system at various time points t. At $t = 180$, the directed invasion at the NE tip of the tumor rim starts to become prominent. Subsequently, $t = 308$

corresponds to the "breakpoint" in the gene expression profile of Tenascin C and PCNA. At $t = 335$, a "bulging" structure at the NE tip starts to emerge, pointing toward the peak of the second nutrient source. Finally, the bulge becomes a prominent structural feature of the NE quadrant at $t = 371$, just before the first tumor cell successfully invades the peak of the second nutrient source. At the *molecular* level, we found that this emergence of a structural asymmetry in the rim of the growing tumor is accompanied by a positive correlation between tumor diameter and the gene expression of Tenascin C, and at the same time a negative one between the former and PCNA expression. To determine the dominant phenotype responsible for this micro–macro link, we next examine the gene expression profiles separately for proliferating, migrating, and quiescent cells. We found that Tenascin C expression is always higher among the migratory phenotype than among their proliferating peers, while the converse is true for the expression of PCNA, i.e., it is always upregulated among the proliferative cells. Intriguingly, detrended fluctuation analysis (DFA) analyses indicate that the time series of gene expression of the combined tumor cells (i.e., including all phenotypes), the long-range autocorrelation indicates non-random-walk *predictability* as represented by $\alpha = 1.32$ for *gTenascin* and $\alpha = 1.06$ for *gPCNA*. However, when DFA is applied separately to migrating and to proliferating cells, the resulting values of α reveal the time-series properties of random walk behavior (i.e., with $\alpha = 0.5$).

7. DISCUSSION, CONCLUSIONS, AND FUTURE WORK

The underlying hypothesis of our work is that malignant tumors behave as *complex dynamic self-organizing* and *adaptive biosystems*. In this chapter, we have presented a numerical agent-based model of malignant brain tumor cells in which both time and space are discrete yet environmental variables are treated realistically as continuous. Simulations of this model allow us to infer the statistical properties of the model and to establish the cause–effect relationships that emerge from various interactions among and between the cells and their environments. The key findings of our works can be briefly summarized as follows.

In (6), we showed the nonlinear dynamical behavior of a virtual tumor system in the form of *phase transitions* and *self-organization*. The phase transitions indicate that if global search is dominant, then lowering r_ρ (i.e., higher mobility) can actually result in slower overall velocity of the tumor system, while self-organization in the form of smaller clusters can contribute to a longer lifetime of the tumor system. Subsequently, in (7), we demonstrated that tumor systems can achieve maximum velocity at *less* than 100% search precision. This finding challenges the conventional wisdom that an error-free search procedure would maximize the velocity of a tumor system dependent on receptor-based mobility. In a follow-up paper examining the structure–function relationship, (8) used

numerical analysis to examine the link between the tumor surface structural pattern, measured by its fractal dimensions, and the spatiotemporal expansion velocity. In particular, we found a positive correlation between these two measures, i.e., higher fracticality of the tumor's surface corresponds to an accelerating spatial expansion. Finally, using a truly multiscale model (9), we found that biopsy specimens containing all available tumor phenotypes (proliferating, migrating, and quiescent cells) are of more predictive value than separate gene-expression profiling for each distinct phenotype. Our multiscale model also confirmed that it is the invasive phenotype and not the proliferative one that drives the tumor system's spatial expansion, as indicated by the strong correlations between the tumor diameter and the gene expression profiles of migrating tumor cells.

Most importantly, several of these simulation results have already been corroborated by experimental and clinical findings:

- For example, in (6), we showed that at higher rates of nutrient depletion, tumor cells can not only exhibit maximum spatial expansion velocity under specific circumstances, but actually survive longer, hence exhibiting a lower apoptosis rate. Interestingly, this result is supported by recent findings from (12), which reported for migratory brain tumor cells in vitro not only an increase in expression of genes implicated in cell motility but also, concomitantly, a decrease in the expression of apoptosis-related genes.

- In (7), we reported that introducing a modicum of randomness in signal processing can actually improve the performance of the tumor. At the same time, however, we also found that an increasing invasive potential appears to require a higher search precision of the cells in order to reach maximum velocity. This emergent property indicates a more significant role for cell-surface receptor mechanisms in the invading cell population of more aggressive neoplasms. This concept has been supported by a recent in situ hybridization (FISH) study investigating EGF-R expression in the infiltrative zone of a human glioblastoma specimen (53).

- In yet another *clinical* study supporting our findings (54), it has been revealed that there is no apparent relationship between tumor size and survival of the host patient. Indeed, these authors found significant statistical correlations between the tumor's volumetric surface and survival time, thus substantiating their claim that a better measure of the tumor's invasive capability is its surface conditions rather than its entire size. Their results thus corroborated our work in (8), which found no significant correlations between tumor

diameter and surface roughness. At the same time, our simulations indicated that an increase in the tumor's velocity of spatial expansion, $<v>$, indeed corresponds to an increase in the surface fracticality d_f.

In the future, there are several extensions that can be pursued as follow-up projects. As genomics data become increasingly available, our micro–macro approach should provide a very helpful starting point for investigating the crucial relationship between the molecular level—e.g., gene expression changes—and the performance of the tumor system on a macroscopic scale. Our current version of the model includes two "key" genes only since the precise role of other potentially critical genes involved in the gene–protein regulatory network remain largely unknown. If such information becomes available in the future, our agent-based model can easily be extended to include more genes and proteins involved in subcellular signaling cascades, as we have recently shown for the case of EGF-R (55). In addition, since we currently assume a monoclonal population of tumor cells, if a more realistic model is desired one will have to consider a heterogeneous multiclonal population of tumor cells. In that context of pursuing a more biologically accurate model, currently underway is extension of our 2D framework into a 3D version, which will be better suited to simulate the progression of an in-vivo tumor, and will also have potential clinical applications related to intraoperative navigation techniques (see this volume, Part IV, chapter 8, by Heilbrun).

In summary, *multiscale agent-based modeling* is a powerful tool for investigating tumors as complex dynamic biosystems. This innovative approach has a high potential to lead to paradigm-shifting insights into tumor biology, which in turn is a first step toward improving diagnostic tools and therapeutic strategies and thus, ultimately, patient outcome.

8. ACKNOWLEDGMENTS

This work has been supported by grants CA 085139 and CA 113004 from the National Institutes of Health and by the Harvard–MIT (HST) Athinoula A. Martinos Center for Biomedical Imaging and the Department of Radiology at Massachusetts General Hospital. Y.M. is the recipient of an NCI-Training Grant Fellowship from the National Institutes of Health (CA 09502).

9. NOTE

1. "Permission" refers to *haptotaxis*, i.e., enhanced cell movement along a solid substrate, which however, is not explicitly modeled here.

10. REFERENCES

1. Chignola R, Schenetti A, Chiesa E, Foroni R, Sartoris S, Brendolan A, Tridente G, Andrighetto G, Liberati D. 1990. Oscillating growth patterns of multicellular tumour spheroids. *Cell Prolif* **32**:39–48.
2. Coffey DS. 1998. Self–organization, complexity and chaos: the new biology for medicine. *Nature Med* **4**:882–885.
3. Deisboeck TS, Berens ME, Kansal AR, Torquato S, Stemmer-Rachamimov AO, Chiocca EA. 2001. Pattern of self-organization in tumor systems: complex growth dynamics in a novel brain tumor spheroid model. *Cell Prolif* **34**:115–134.
4. Gonzalez-Garcia I, Sole RI, Costa J. 2002. Metapopulation dynamics and spatial heterogeneity in cancer. *PNAS* **99**:13085–13089.
5. Waliszewski P, Molski M, Konarski J. 1998. On the holistic approach in cellular and cancer biology: nonlinearity, complexity, and quasi determination of the dynamic network. *J Surg Oncol* **68**:70–78.
6. Mansury Y, Kimura M, Lobo J, Deisboeck TS. 2002. Emerging patterns in tumor systems: simulating the dynamics of multicellular clusters with an agent-based spatial agglomeration model. *J Theor Biol* **219**:343–370.
7. Mansury Y, Deisboeck TS. 2003. The impact of "search precision" in an agent-based tumor model. *J Theor Biol* **224**:325–337.
8. Mansury Y, Deisboeck TS. 2004. Simulating "structure–function" patterns of malignant brain tumors. *Physica A* **331**:219–232.
9. Mansury Y, Deisboeck TS. 2004. Simulating the time series of a selected gene expression profile in an agent-based tumor model. *Physica D* **196**:193–204.
10. Kansal AR, Torquato S, Chiocca EA, Deisboeck TS. 2000. Emergence of a subpopulation in a computational model of tumor growth. *J Theor Biol* **207**:431–441.
11. Giese A, Loo MA, Tran N, Haskett D, Coons S, Berens ME. 1996a. Dichotomy of astrocytoma migration and proliferation. *Intl J Cancer* **67**:275–282.
12. Mariani L, Beaudry C, McDonough WS, Hoelzinger DB, Demuth T, Ross KR, Berens T, Coons SW, Watts G, Trent JM, Wei JS, Giese A, Berens ME. 2001. Glioma cell motility is associated with reduced transcription of proapoptotic and proliferation genes: a cDNA microarray analysis. *J Neuro-Oncol* **53**:161–176.
13. Ermentrout GB, Edelstein-Keshet L. 1993. Cellular automata approaches to biological modeling. *J Theor Biol* **160**:97–133.
14. Kauffman S. 1984. Emergent properties in random complex automata. *Physica D* **10**:145–156.
15. Wolfram S. 1984. Cellular automaton as models of complexity. *Nature* **311**:419–424.
16. Kauffman S. 1990. *The origins of order*. Oxford UP, New York.
17. Forest S. 1990. Emergent computation: self-organizing, collective, and cooperative phenomena in natural and artificial computing networks. *Physica D* **42**:1–11.
18. Duechting W, Vogelsaenger T. 1983. Aspects of modelling and simulating tumor growth and treatment. *J Cancer Res Clin Oncol* **105**:1–12.
19. Chowdhury D, Sahimi M, Stauffer D. 1991. A discrete model for immune surveillance, tumour immunity and cancer. *J Theor Biol* **152**:263–270.
20. Qi A-S, Zheng X, Du C-Y, An B-S. 1993. A cellular automaton model of cancerous growth. *J Theor Biol* **161**:1–12.
21. Smolle J, Stettner H. 1993. Computer simulation of tumor cell invasion by a stochastic growth model. *J Theor Biol* **160**:63–72.
22. Delsanto PP, Romano A, Scalerandi M, Pescarmona GP. 2000. Analysis of a "phase transition" from tumor growth to latency. *Phys Rev E* **62**:2547–2554.

23. Sander LM, Deisboeck TS. 2002. Growth patterns of microscopic brain tumors. *Phys Rev E* **66**:051901.
24. Marusic M, Bajzer Z, Freyer JP, Vuk-Pavlovic S. 1994. Analysis of growth of multicellular tumour spheroids by mathematical models. *Cell Prolif* **27**:73–94.
25. Kansal AR, Torquato S, Chiocca EA, Deisboeck TS. 2000. Simulated brain tumor growth dynamics using a three-dimensional cellular automaton. *J Theor Biol* **203**:367–382.
26. Tracqui P. 1995. From passive diffusion to active cellular migration in mathematical models of tumor invasion. *Acta Biotheor* **43**:443–464.
27. Perumpanani AJ, Sherratt JA, Norbury J, Byrne HM. 1996. Biological inferences from a mathematical model for malignant invasion. *Invasion Metastasis* **16**:209–221.
28. Swanson KR, Alvord Jr EC, Murray JD. 2002. Virtual brain tumours (gliomas) enhance the reality of medical imaging and highlight inadequacies of current therapy. *Brit J Cancer* **86**:14–18.
29. Turner S, Sherratt JA. 2002. Intercellular adhesion and cancer invasion: a discrete simulation using the extended Potts model. *J Theor Biol* **216**:85–100.
30. Habib S, Molina-Paris C, Deisboeck TS. 2003. Complex dynamics of tumors: modeling an emerging brain tumor system using a set of coupled reaction-diffusion equations. *Physica A* **327**:501–524.
31. Tamaki M, McDonald W, Amberger VR, Moore E, Maestro RFD. 1997. Implantation of C6 astrocytoma spheroid into collagen type I gels: invasive, proliferative, and enzymatic characterizations. *J Neurosurg* **87**:602–609.
32. Boado RJ, Black KL, Pardridge WM. 1994. Gene Expression of GLUT3 and GLUT1 glucose transporters in human brain tumors. *Brain Res Mol Brain Res* **27**:51–57.
33. Chicoine MR, Madsen CL, Silbergeld DL. 1995. Modification of human glioma locomotion in vitro by cytokines EGF, bFGF, PDGFbb, NGF, and TNFα. *Neurosurgery* **36**:1165–1171.
34. Sang HU, Espiritu OD, Kelley PY, Klauber MR, Hatton JD. 1995. The role of the epidermal growth factor receptor in human gliomas, I: the control of cell growth. *J Neurosurg* **82**:841–846.
35. Jabour BA, Choi Y, Hoh CK, Rege SD, Soong JC, Lufkin RB, Hanafee WN, Maddahi J, Chaiken L, Bailet J. 1993. Extracranial head and neck: PET imaging with 2-[F-18] fluoro-2-deoxy-D-glucose and MR imaging correlation. *Radiology* **186**:27–35.
36. Yonekura Y, Benua RS, Brill AB, Som P, Yeh SD, Kemeny NE, Fowler JS, MacGregor RR, Stamm R, Christman DR, Wolf AP. 1982. Increased accumulation of 2-deoxy-2-[18F] fluoro-D-glucose in liver metastases from colon carcinoma. *J Nucl Med* **23**:1133–1137.
37. Acker H, Holtermann G, Boelling B, Carlsson J. 1992. Influence of glucose on metabolism and growth of rat glioma cells (C6) in multicellular spheroid culture. *Intl J Cancer* **52**:279–285.
38. Sutherland RM, Sordat B, Bamat J, Gabbert H, Bourrat B, Mueller-Klieser W. 1986. Oxygenation and differentiation in multicellular spheroids of human colon carcinoma. *Cancer Res* **46**:5320–5329.
39. Helmlinger G, Netti PA, Lichtenbeld HC, Melder RJ, Jain RK. 1997. Solid stress inhibits the growth of multicellular tumor spheroids. *Nature Biotechnol* **15**:778–783.
40. Oishi Y, Uezono Y, Yanagihara N, Izumi F, Nakamura T, Suzuki K. 1998. Transmural compression-induced proliferation and DNA synthesis through activation of a tyrosine kinase pathway in rat astrocytoma RCR-1 cells. *Brain Res* **781**:159–166.
41. Nakano A, Tani E, Miyazaki K, Yamamoto Y, Furuyama J. 1995. Matrix metalloproteinases and tissue inhibitors of metalloproteinases in human gliomas. *J Neurosurg* **83**:298–307.
42. Marx E, Mueller-Klieser W, Vaupel P. 1988. Lactate-induced inhibition of tumor cell proliferation. *Intl J Radiation Oncol Biol Phys* **14**:947–955.
43. Krtolica A, Ludlow JW. 1996. Hypoxia arrests ovarian carcinoma cell cycle progression, but invasion is unaffected. *Cancer Res* **56**:1168–1173.
44. Plasswilm L, Tannapfel A, Cordes N, Demir R, Hoeper K, Bauer J, Hoeper J. 2000. Hypoxia-induced tumour cell migration in an in vivo chicken model. *Pathobiology* **68**:99–105.

45. Barabasi A-L, Stanley HE. 1995. *Fractal concepts in surface growth*. Cambridge UP, Cambridge.
46. Giese A, Loo MA, Norman SA, Treasurywala S, Berens ME. 1996b. Contrasting migratory response of astrocytoma cells to tenascin mediated by different integrins. *J Cell Sci* **109**:2161–2168.
47. Cavalla P, Schiffer D. 1997. Cell cycle and proliferation markers in neuroepithelial tumors. *Anticancer Res* **17**:4135–4143.
48. Peng C-K, Buldyrev SV, Havlin S, Simons M, Stanley HE, Goldberger AL. 1994. Mosaic organization of DNA nucleotides. *Phys Rev E* **49**:1685–1689.
49. Mueller-Klieser W, Sutherland RM. 1982. Oxygen tensions in multicell spheroids of two cell lines. *Brit J Cancer* **45**:256–264.
50. Carmeliet P, Jain RK. 2000. Angiogenesis in cancer and other diseases. *Nature* **407**:249–257.
51. Cross SS, Bury JP, Silcocks PB, Stephenson TJ, Cotton DW. 1994. Fractal geometric analysis of colorectal polyps. *J Pathol* **172**:317–323.
52. Cross SS. 1997. Fractals in pathology. *J Pathol* **182**:1–8.
53. Okada Y, Hurwitz EE, Esposito JM, Brower MA, Nutt CL, Louis DN. 2003. Selection pressures of TP53 mutation and microenvironmental location influence EGFR gene amplification in human glioblastomas. *Cancer Res* **63**:413–416.
54. Valery CA, Marro B, Boyer O, Duyme M, Mokhtari K, Marsault C, Klatzmann D, Philippon J. 2001. Extent of tumor–brain interface: a new tool to predict evolution of malignant gliomas. *J Neurosurg* **94**:433–436.
55. Athale C, Mansury Y, Deisboeck TS. 2005. Simulating the impact of a molecular "decision-process" on cellular phenotype and multicellular patterns in brain tumors. *J Theor Biol* **233**:469–481.

7.1

THE COMPLEXITY OF DYNAMIC HOST NETWORKS

Steve W. Cole

David E. Geffen School of Medicine, UCLA AIDS Institute, the Norman Cousins Center, and UCLA Molecular Biology Institute, University of California, Los Angeles

Mathematical epidemiology has generally modeled host populations as homogeneous networks with static linkage structures. However, real hosts react biologically and behaviorally to disease in ways that dynamically alter network connectivity. This chapter summarizes results from agent-based modeling of disease-reactive social networks and their impact on the propagation of infections. Results show that simple distributed rules about "sickness behavior" can interact with nonhomogeneous social networks to destabilize propagation kinetics, augment biological immune responses, and favor the evolution of biological structures and cultural norms that function as a network-level social immune response.

1. INTRODUCTION

Disease is generally analyzed as a biological process, but sickness is also an experience. This chapter analyzes the impact of that experience on the course of epidemics with an eye toward its evolutionary significance. A large research literature has sought to understand the "lifestyle strategies" of parasites (1), but the hosts they colonize are often conceptualized as vacuous mobile resource patches that incubate pathogens and disseminate them randomly throughout

Address correspondence to: Steve W. Cole, Division of Hematology-Oncology, Department of Medicine, 11-934 Factor Building, David Geffen School of Medicine at UCLA, Los Angeles CA 90095-1678 (coles@ucla.edu).

society before prematurely expiring. We all have our problems, but this seems a bit severe. In reality, most organisms change their behavior during periods of illness (2,3). These responses are often analyzed in terms of their benefits for the afflicted, but they also have the potential to create a type of "social immune response" that protects healthy individuals by altering patterns of interpersonal contact. At the aggregate level, these changes amount to transient distortions in the structure of social networks. Disease-reactive networks can be induced by a centralized authority (e.g., quarantine), but they can also develop as an emergent property of simple behavioral rules operating at the individual level (e.g., avoid sick people). In fact, vertebrate biology appears to have evolved molecular signaling pathways to generate disease-reactive behavior without any explicit reasoning by a host. Diagnosis and treatment also change the functional connectivity of a disease transmission network, as do variations in host resistance or pathogen virulence. In homogeneous social networks, individual disease-reactive behavior aggregates into fairly simple nonlinear feedback at the population level. Vertebrate social structures are actually quite heterogeneous, and dynamic linkage can produce highly complex behavior in such heavily structured systems. The present studies seek to map those dynamics and understand their significance for population survival. The findings that emerge suggest that the neural substrates of disease-reactive behavior may have evolved in tandem with biological immune responses to "strategically" alter host population structures under the ecological press of socially transmitted disease.

2. MODEL

The results presented here come from a series of epidemics simulated in ActiveHost—an agent-based modeling system for analyzing interactions between biological and behavioral determinants of health. The general architecture is summarized in Figure 1, and described in greater detail in the Appendix.

The basic modeled unit is an individual host—an "agent" that interacts with other agents in some way that can transmit disease. Each agent maintains a list of potential interaction partners and realizes those interactions according to a specified and potentially variable probability during each time unit. The aggregated set of "source–target" linkages constitutes the disease transmission network. Additional linkage networks may also be provided to transmit disease-blocking interventions, behavioral responses to illness, or changes in social contact patterns. Each host maintains a submodel specifying pathogenetic processes on the cellular and molecular level, and this model can delete the host from the live population after a specified period of infection. The pathogenesis model can include periods of latent infectiousness (host is not recognizably sick but can transmit disease), overt illness (visibly sick and transmits disease), disease remission/recovery (infectiousness ceases after a period), immunity (loss of infectibility), and reactivation (infectiousness resumes after a period of recov-

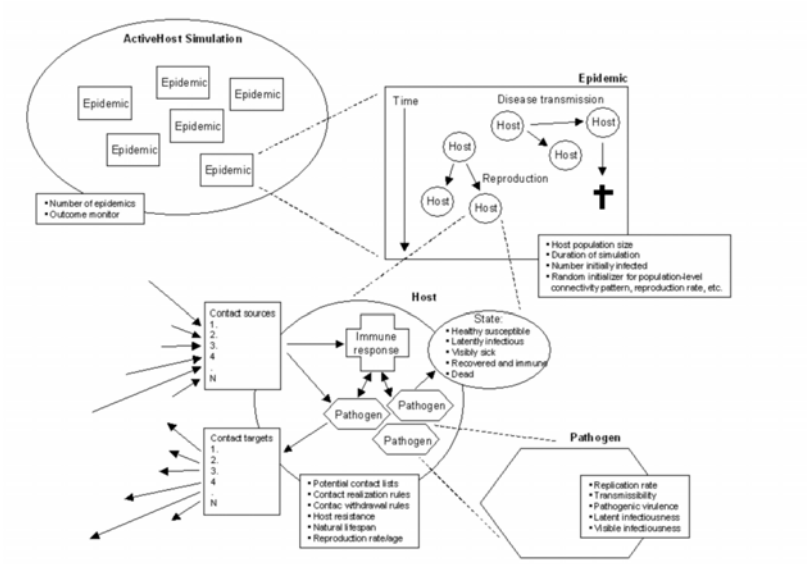


Figure 1. ***General architecture of the ActiveHost modeling system***. Each simulated epidemic is based on a collection of hosts interacting with one another in ways that transmit disease. Agents maintain a list of potential contacts that are realized according to probabilistic rules that constitute the structure of the disease transmission network. Network dynamics occur as agents modulate contact realization probabilities in response to simple, ecologically realistic rules, such as "interact with friends more often than strangers," "avoid people who are sick," or "maintain a constant number of social contacts." Transmitted pathogens provoke immune responses, replicate, and die with the host after a defined period of infection if they cannot spread to a new host. Each simulated object maintains a set of relevant parameters (bulleted list items), and the overarching simulation object compiles mortality trajectories, population-wide values of evolving parameters, and other outcomes over an arbitrary number of stochastic repetitions.

ery). Hosts also have a natural lifespan, and the system includes a framework for modeling evolutionary dynamics via age-dependent reproduction and noisy transmission of physiologic and behavioral parameters from parent to child. The aggregate collection of hosts is aged for a fixed number of time units, and the primary outcomes monitored are the host population size, the pattern of realized social linkages, and the frequency of various host states (sick, latently infected, overtly infected, dead, immune, etc.).

3. RESULTS

The present analyses focus on two ways in which real disease transmission networks depart from the assumptions underlying conventional epidemiologic

network models; (i) nonhomogeneous connectivity, and (ii) dynamic network structure. After defining the consequences of each departure in isolation, our primary interest focuses on their synergistic effects on host survival and evolutionary dynamics.

3.1. Nonhomogeneous Host Networks

Most dynamic models in epidemiology implicitly assume that disease spreads within a homogeneous network of randomly linked hosts (4). This approach fares reasonably well in some cases (e.g., mosquito-borne malaria), but it fails to accurately forecast the spread of illnesses that depend on close physical contact, such as HIV or hepatitis, or those involving a major behavioral risk component, such as malignant melanoma or lung cancer. The AIDS epidemic in particular motivated the consideration of more structured social networks in which each agent is connected to only a few others (sparse connectivity) and individuals are clustered in their pattern of social contact (blocked networks). Much attention has focused on "small-world" networks in which a few individuals are linked to many others, and "giant components" can develop to virtually ensure a transmission path between any two individuals (5,6). Other systems analyzed include (a) "tribe" or "subculture blocks," in which small groups of individuals are highly interconnected and only sparsely linked to other groups, (b) "continuous bands," in which individuals are linked within smooth adjacency neighborhoods, and (c) "defector blocks," in which most links are reciprocal but a limited number of infidelities connect one member of a pair to another pair. Figure 2 presents examples of these structures along with epidemic simulations showing their impact on disease propagation. In Figure 2A-2F each point represents a contact with the potential to transmit disease from a source agent (horizontal axis) to a target agent (vertical axis). All targets are connected to at least one source, and each realizes a randomly selected one of its social links per unit time. All other epidemiologic parameters remain constant (hosts 1 and 2 are initially infected, probability of infection given exposure = 100%, and individuals are infectious for 2 time units). The only thing that differs is the number and pattern of the potential contacts available for realization.

In structured networks such as those depicted in panels 2B–2F, disease dissemination rates differ substantially from those seen under the random homogeneous mixing (panel 2A). Disease propagates through alternative contact structures at very different rates despite the fact that the total number of links realized per unit time is equivalent (1 per host in all cases). Thin gray lines represent 100 realized mortality trajectories for each system, and heavy black lines show the average. Trajectories achieving a flat slope before 100% mortality indicate epidemiologic "burnout" (pathogen extinction), whereas those reaching 100% indicate host extinction. When each host reallocates one potential contact

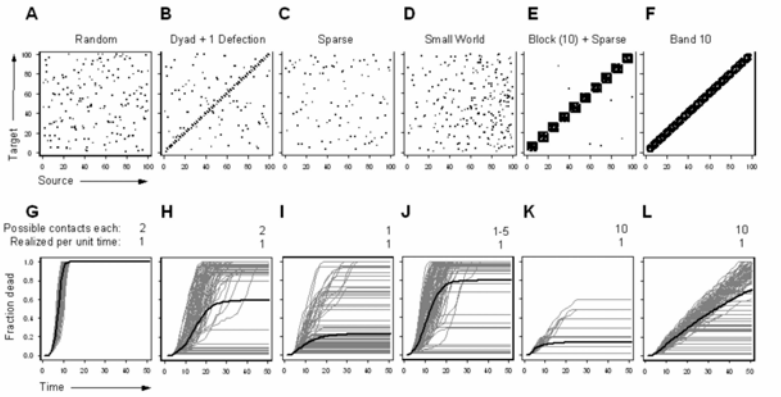


Figure 2. ***Social structure and disease propagation***. Connectivity structures are plotted for a homogeneous randomly linked population (**A**) and several structured alternatives, including a reciprocal binary system with defectors (**B**), a sparse random network (**C**), a small-world network of one-to-many mappings (**D**), a block structure with random interconnections among blocks (**E**), and a continuous adjacency band (**F**). Points represent contacts with the potential to transmit disease from a source (horizontal axis) to a target (vertical axis), and all targets are connected to at least one source. Disease propagates through alternative contact structures at very different rates despite the fact that the total number of links realized per unit time is equivalent (1 in G–L). In G–L thin lines represent realized mortality trajectories for each system, and heavy lines show the average. Trajectories achieving a flat slope before 100% mortality have burned out (pathogen extinction), whereas those reaching 100% indicate host extinction.

to a stable dyad (2H) rather than a random partner (2G), population survival rates increase substantially. However, such effects are not equivalent to reducing the total number of potential contacts (2I) because the network retains the capacity for occasionally generating far leaps in disease distribution. Even a small number of highly connected individuals can undermine a population's protection from disease, as in 2J's small-world network where possible contact numbers for each individual follow a power-law distribution between 1 and 5. In contrast, organization of social contacts into highly clustered blocks can profoundly retard disease propagation even when total possible links are fivefold greater than those of a randomly connected network (10 vs. 2 in 2K vs. 2G). Smooth adjacency networks with the same number of links show an intermediate phenotype (2L), with disease decelerated relative to a random homogeneous system (2G) but still marching inexorably through the population due to the absence of clearly delineated subgroups that can bottle up and extinguish the infection. Structured networks provide lots of opportunities to halt an epidemic, but how quickly or even whether this happens depends critically on small random variations in realization of a few links between members of different subpopulations (e.g., in Figure 2J,K). Because population-wide disease penetrance depends on a

small number of stochastic events, it is difficult to accurately predict the course of an epidemic from its early trajectory. Repeated simulations show wildly different kinetics in panels 2H–2L. Under the homogeneous mixing assumption, epidemiologic kinetics seem highly predictable (2G), but more realistic network structures reveal much greater variability. Often, it is not even possible to predict whether the host population or the pathogen will perish first (2H, 2J, and 2K).

3.2. Dynamic Network Structures

In addition to structural heterogeneity, social contact networks also show temporal heterogeneity in realized linkages. Two basic processes drive these dynamics. First, the development of biological immunity or death removes individuals from the system of transmitters following a certain period of infectiousness. Second, the social interactions that transmit disease are intrinsically dynamic in their own right. We each know hundreds of people, but on any given day we interact with only a small number of them. The links we do realize are clustered in both time and social space because we typically interact with a small and stable social core (e.g., family members and immediate coworkers). The vast majority of potential links are realized only rarely. This "small-world" temporal structure implies a functional decrease in network connectivity per unit time, but it is not equivalent to removing low-frequency links because the network retains a capacity for occasional far leaps in disease distribution. Temporally sparse social contact is modeled by generating a fixed set of possible contacts for each individual and realizing a constant number of contacts per unit time according to a specified probability model (Figure 3).

Temporal link dynamics can profoundly impact the propagation of disease even through highly vulnerable social structures such as the small-world network (all links realized in Figure 3A vs. a random 50% in 3B; note differential frequency of host population extinction). Compared to a uniform probability of realizing any possible contact (3B), increased probability of realizing more proximal links results in considerably enhanced survival despite the fact that the total number of realized links remains constant (3C). These examples come from epidemics initiated by 3 infected individuals in the midst of a 200-host population with a small-world contact distribution ranging from 1 to 5 possible contacts per individual (496 total links), and an infectious duration of 1 time unit. In Figure 3B, each link is realized with a probability of 50% per unit time for all individuals. In Figure 3C the probability of realizing each link is an average of 50% that varies between 0 and 1 depending on the squared social distance between source and target. In addition to slowing the mean disease penetrance trajectory, sociospatial link heterogeneity also greatly increases variability in outcomes.

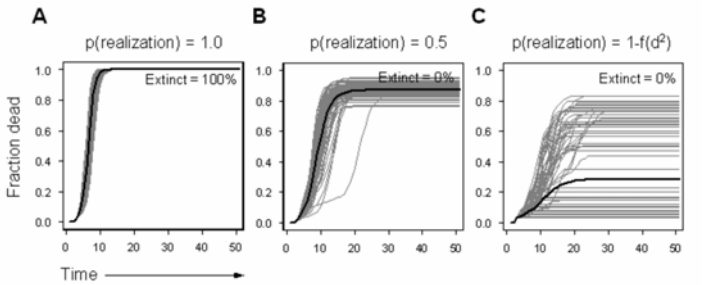


Figure 3. ***Temporal sparseness***. Dynamic social contact is modeled by generating a fixed set of possible contacts for each individual and realizing a constant number per unit time according to a specified probability model. Reducing the total number of realized contacts from 100% (**A**) to 50% (**B**) ensures survival of at least some members of the population, even in highly vulnerable populations such as this small-world network. However, the sociospatial structure of linkage realization probabilities also has an important effect as shown in (**C**), where the total number of contacts realized per unit time is eequivalent to (**B**), but the probability of realizing each link is inversely proportional to the distance between individuals in social space.

3.2.1. Disease-Reactive Dynamics

The effects of intrinsic temporal variability are amplified by illness-induced changes to social interaction. One factor that would seem to play a major role is a healthy person's conscious avoidance of the sick, either at the behest of health authorities or through their own spontaneous social quarantines. However, the potential value of this mechanism is undermined by the fact that many pathogens are transmissible for days or even years before any signs of illness emerge to provoke social withdrawal (e.g., upper respiratory viral infections, HIV, or the "infectious" habit of smoking). Most visible symptoms are generated by the immune response, rather than the pathogen, and thus require at least a day or two to develop. Quarantines also demand extreme vigilance on the part of a large number of hosts if they are to effectively protect a population, or even a specific individual. Given the high degree of clustering in social networks, A can infect B quite certainly by transmitting disease to their mutual friends C, D, and E, no matter how studiously B avoids A. Thus, B's health depends on the simultaneous diligence of C, D, and E, and all require some overt sign of disease to trigger withdrawal from A. Figure 4 shows the results of epidemiologic simulations in which uninfected hosts probabilistically reduce contact with infected individuals once signs of illness appear. In this case, hosts are latently infectious for 2 time units before becoming overtly infectious for one more time unit before dying.

Relative to a constant-contact network (4A), social systems that dynamically withdraw contact from overtly sick individuals (4B) experience considerable advantages in population survival even when the vast majority of indi-

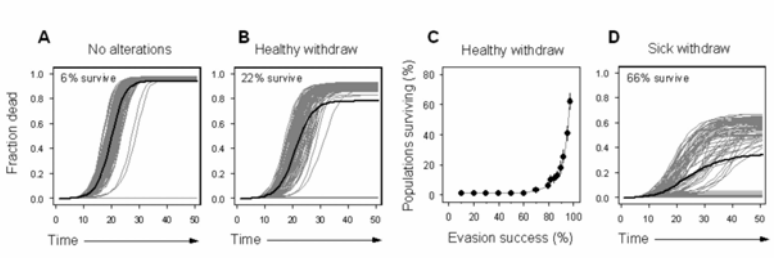


Figure 4. ***Social quarantine***. To model avoidance of sick individuals by the healthy, uninfected individuals reduce the probability of realizing a contact with sick individuals in (**B**). However, the probability reduction must be substantial to protect a population from extinction (**C**). Social withdrawal by infected individuals (**D**) is even more effective in containing an epidemic because they often detect illness before signs are apparent to others (e.g., after 1 time unit of infection, vs. 2 in **B**).

viduals are killed. However, uninfected individuals must detect and avoid those who are infectious with an extremely high rate of success to halt an epidemic. Figure 4C shows that under the conditions of panels A and B, uninfected individuals must detect and evade at least 80% of infected individuals to avoid consistent host population extinction.

Surprisingly, the network dynamics that most decisively contain disease do not stem from the self-protective behavior of the healthy, but from the involuntary behavior of the sick. It has recently been discovered that proinflammatory cytokines—the signaling molecules that initiate an immune response—also prompt the brain to unleash an integrated package of "sickness behaviors" that immobilize us with fatigue, malaise, and myalgia, and substantially crimp our social and reproductive motivation (2,3). Sickness behavior has generally been analyzed in terms of its advantage for individual recovery, but its most significant contribution may lie in protection of the group. Even small reductions in contact can substantially impede the spread of infection through a sparse network. Figure 5B shows an example in which infected individuals withdraw 10% of their social contacts at random, resulting in substantial increases in population survival relative to a comparable scenario in which the sick do not withdraw (5A).

Self-generated quarantines are more efficient than socially imposed ones because we generally feel sick sooner than we look it, and this allows a faster blockade of transmission links. Figure 4D emphasizes this point by comparing the effects of a fixed probabilistic reduction in contact generated by the sick individual one time unit following infection with the same probabilistic reduction generated by healthy individuals once they recognize illness at two time units post infection (4B). A sick host's self-generated quarantines are also more efficient than social quarantines because the motivation for altered behavior

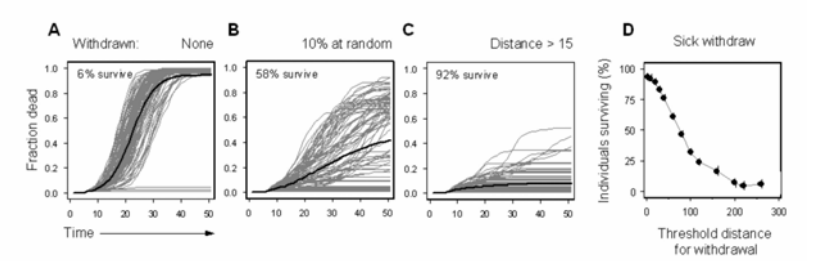


Figure 5. ***Sickness behavior***. When sick individuals withdraw from even a small number of social contacts (10% in **B**), mortality rates decline substantially in highly vulnerable population structures such as this small-world network (compare to 0% withdrawal in **A**). If the same number of voluntary withdrawals is selectively deployed toward distant contacts (**C**), population protection is even greater. Considerable population survival benefits accrue even if individuals maintain contact with large number of individuals in their vicinity (supporting survival of the afflicted) and defer contact only with quite distant interaction partners (**D**).

emerges from a biologically impacted person who tracks (is) the source of infection, rather than depending upon the simultaneous conscientiousness of many potential targets who face an unrealized threat. Interestingly, sick people are more likely to defer contact with strangers or low-frequency partners than they are to avoid their core social contacts (especially family). From an evolutionary perspective, this would seem to set our closest genetic relatives at a competitive disadvantage. However, it also prevents large jumps of disease through social space, and thus efficiently protects the population as a whole. Figure 5C illustrates the comparative advantage of redeploying social contacts away from socially distant partners during times of illness. The same number of links is withdrawn in Figures 5B and 5C. The only difference involves making the probability of withdrawal depend upon social distance in 5C. Distance-dependent contact reduction confers major population survival benefits, even when reductions are limited to very distant interaction partners (shown parametrically in Figure 5D). Such results imply there may be considerable selective pressure for biological mechanisms that reduce social contact during sickness. All simulations in Figure 5 were conducted in a small-world network of 1000 agents, each realizing a single randomly selected contact per unit time from a set of possible contacts that vary from 1 to 10 according to a power law. The pathogen is infectious for 1 time unit before a stricken individual feels ill and another 3 time units subsequently, and each of the 300 hosts has a resistance of .5 (1 = 100% probability of infection given exposure). Disease easily penetrates the nonreactive small world of Figure 5A to kill an average 94% of the population within 50 time units. Even a small (10%) reduction in social contact rates by sickened individuals can increase survival frequencies by about tenfold (5B), and these pro-

tective effects are amplified when sick individuals selectively withdraw contacts from their most socially distant partners (5C).

All disease-reactive network dynamics fundamentally stem from individual biological immune responses (see Part III, section 4, this volume). Inflammatory biology generates the illness signs that prompt the healthy to withdraw from the sick, the sickness behaviors that prompt the sick to withdraw from the healthy, and the leukocyte responses that modulate host resistance. Even when the immune response fails to save the individual from disease (e.g., Ebola virus), it may still effectively protect a population by triggering changes in social contact. An ironic corollary is that the diseases most disastrous for an individual may be the least dangerous to society as a whole because their spectacular visibility induces the most pronounced changes in network structure. A more sobering corollary suggests that pathogens that acquire the capacity to undermine behavioral responses to illness might enjoy a powerful selective advantage. Sickness behavior and the reception of cytokine signals by the brain appear to constitute one example in which evolution has encoded a critical emergent property of the entire social network in the molecular biology of the individual.

3.2.2. Host Resistance Dynamics

In addition to changing social behavior, individual physiologic processes also influence host resistance to disease. One example of this involves the effects of physical or psychological stress, which can impair leukocyte function and render individuals more vulnerable to infectious diseases (7,8). Reduced resistance is tantamount to increasing the number of exposures that can transmit full-blown disease, and thus functionally increases the connectivity of a disease-transmission network in the vicinity of a stressed individual. Dynamic host resistance can be modeled by varying the probability of infection given an exposure, as shown in Figure 6. In addition to intra-individual dynamics, between-host heterogeneity in resistance can also have significant implications. Figure 6B shows that assigning all individuals a constant resistance of .5 (50% probability of infection given exposure) produces rapidly bifurcating disease trajectories, with either hosts or pathogens quickly going extinct. In Figure 6C individual resistance varies randomly between 0 and 100%, with the same mean level (50%) as in 6B. This damps the explosive kinetics of 6B, producing fewer population extinctions despite the fact that the mean trajectory remains constant. Under these conditions, the addition of illness-reactive link dynamics can be especially decisive. If uninfected individuals evade one visibly sick contact per unit time, disease penetrance is substantially retarded (Figure 6D). However, reducing contact with sick individuals can also have unintended negative effects if host resistance depends in part on the number of social contacts realized. Social relationships are major sources of individual sustenance, and host resistance

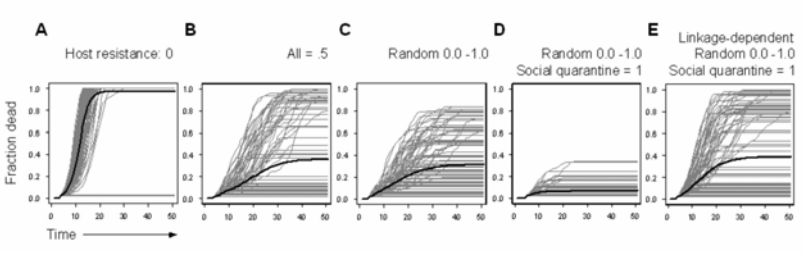


Figure 6. ***Dynamic host resistance***. Variation in individual disease resistance is modeled by varying the probability of infection given exposure. Compared to a population with no resistance (**A**), a constant resistance of .5 (**B**) (50% probability of infection given exposure) substantially reduces disease propagation. Populations with interindividual variation in resistance enjoy even greater protection (**C**), despite an equivalent population-wide average. In **C**, resistance is randomly realized on the uniform interval 0–1, for a mean .5. Under these conditions, illness-reactive link dynamics can be highly decisive as in (**D**), where uninfected individuals can also evade one visibly sick contact per unit time. However, when host resistance depends in part on the number of social contacts realized (**E**), protective effects of reducing exposure can be offset by increased individual vulnerability to infection via remaining contacts (who may be infectious but not visibly sick).

diminishes in people with little social contact (9,10). Any attempt to decrease disease exposure by reducing social interaction may thus be offset by increased vulnerability to infection. In Figure 6E, healthy individuals withdraw from those who are visibly sick just as in 6D, but they suffer a fractional reduction in disease resistance as a result. This dynamic creates the opportunity for explosive acceleration of an epidemic, particularly in clustered populations where much of one's social network may fall ill simultaneously. Total contact levels can be maintained by redeploying links to new partners, but this increases the real connectivity of the network and is counterproductive when the redeploying agents are asymptomatically infected. The reemergence of explosive epidemic dynamics in Figure 6E underscores how seemingly subtle aspects of social behavior can have highly amplified impacts on epidemic behavior in the context of dynamic host networks.

3.2.3. Transmission of Resistance in Multilevel Networks

In addition to altering social contact with those already ill, host networks also respond to disease by developing preventive interventions (e.g., safe sex, antibiotics, and vaccines). However, the networks distributing such interventions are often structured differently from those distributing disease. For example, the socioeconomic network that controls access to antiretroviral medications is quite incongruent with the network that currently transmits HIV. Such misalignments can be analyzed by superimposing a second "intervention network" upon a

population of hosts already connected by a dynamic exposure network. More realistic variants might include a host-specific proclivity to utilize the intervention, which may depend upon a third "media" network distributing perceived vulnerability. Multilevel networks provide a platform for analyzing a variety of sociocultural dynamics that may impact physical health, including socially selective stressors, differential access to medical care, culturally motivated avoidance of diagnosis or treatment (e.g., failure to be tested for HIV due to stigma), social transmission of health risk behaviors (e.g., smoking), and the globalization of personal behavior (e.g., homogenization of health beliefs, social values, lifestyle, and behavior, as described by Garrett (12)). Multilevel networks also provide a context for analyzing interactions between host behavior and the biology of developing disease, such as gene *x* diet interactions in atherosclerosis or the evolution of pathogens and immune responses within behaviorally structured niches. Evolving variants of ActiveHost, for example, mimic observed data in developing more powerful immune systems for sexually promiscuous individuals (11). In the context of disease-reactive social behavior, evolutionary analyses also show a strong selective pressure for the development of social norms that isolate individuals during times of illness. These norms need to be transmissible from parent to child for population-level selection, but they need not be genetically encoded. In fact, dissemination of such norms via superimposed intervention networks enjoys considerable advantage over genetic transmission due to the enhanced speed of norm dispersal.

3.3. Synergistic Complexity

Disease-reactive social behavior creates a temporal sparseness to social networks that combines with structural sparseness to create transient social firewalls at the interface between infected and uninfected subgroups. This has the net effect of discretizing continuous disease dynamics. A pathogen that kills all members of a subpopulation before they can convey infection to the superpopulation does not suffer a quantitative reduction in penetrance; it becomes extinct. Sparse dynamics can cut the other way, of course, with a few random links carrying the potential to connect an isolated outbreak to a system-wide giant component (the "patient zero" problem (12), as illustrated in Figures 2I–2K). These quantal dynamics constitute the primary reason that linear algebraic models perform poorly in predicting the course of emerging epidemics. Linear statistical models forecast the future range of an epidemic based on its past variation, but reactive host networks show increasingly jumpy dynamics as the size of an epidemic grows. Figure 7 illustrates the complex kinetics that emerge when highly structured networks are combined with disease-reactive linkage. Figure 7A shows results from a highly infectious epidemic spreading through a population of 300 agents organized into interconnected blocks of 3 (families), with one

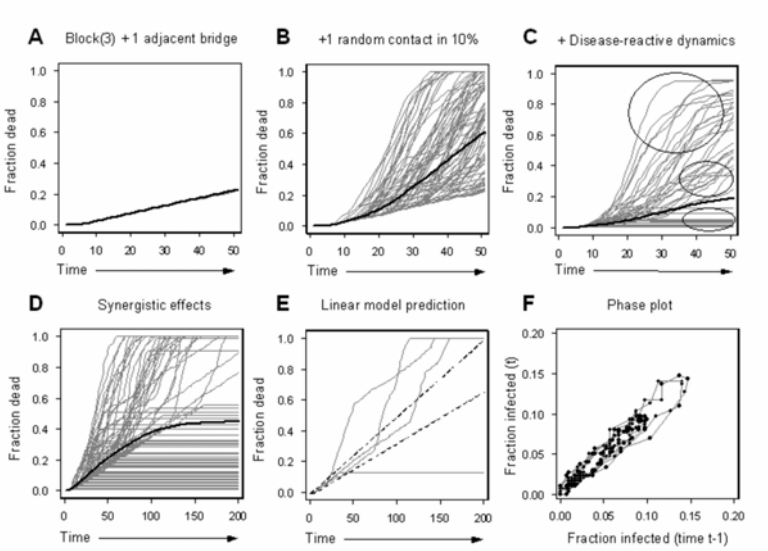


Figure 7. ***Complex epidemiologic trajectories in dynamic structured networks***. The combined effects of highly structured populations and illness-reactive link dynamics is modeled in 100 families of 3 agents, with one member of each family also capable to transmitting disease to an adjacent family (**A**). Family systems are highly sensitive to the introduction of a small number of random linkages as in (**B**), where addition of a single additional random contact for every 30th individual results in substantial acceleration in mean disease penetrance (heavy line) and the collapse of predictability. When distance-dependent disease-reactive network dynamics are superimposed (**C**), mean penetrance rates return to basal levels but the dynamic regime remains highly unstable and produces 3 attractor trajectories including explosive growth, a slow steady population burn, and rapid pathogen extinction. Note that the mean trajectory does not coincide with any of the regimes actually observed. (**D**) "Knife-edge" dynamics emerge in the same system when sick individuals withdraw at random from 50% of their potential contacts (instead of selectively avoiding those most distant as in **C**). Host/pathogen equilibrium is virtually impossible to attain under these circumstances and one population or the other rapidly becomes extinct. Which one occurs is difficult to predict on the basis of the epidemic's early behavior (**E**). Linear statistical analyses fail to accurately forecast epidemic trajectories due to highly unsmooth derivatives (dashed lines represent a 95% prediction interval based on ARIMA 1,1,0 time-series analysis of the first 30 observations). However, these "catastrophic" dynamics are neither chaotic nor random, as shown by a strong autoregressive component in the phase plot of the number of infected hosts at time t vs. $t - 1$ (**F**). Note the absence of classically chaotic (smooth-curved) or random stochastic (scattered) dynamics.

agent in each block maintaining a potential connection to one member of an adjacent block. Figure 7B shows the effects of adding a single additional random contact possibility for every 30th individual (10% of the population), which substantially accelerates disease penetrance and causes a profound collapse in predictability (Figure 7B). When distance-dependent disease-reactive link dynamics are superimposed in Figure 7C, mean penetrance returns to basal levels but the dynamic regime remains highly unstable. Three attractors emerge including (a) explosive depletion of the majority of hosts, (b) a slow steady burn through the population, and (c) rapid extinction of the pathogen with predomi-

nant host survival. Note that the mean trajectory that is typically modeled does not correspond to any of the regimes actually observed. Figure 7D shows "knife-edge" dynamics emerging in the same system when sick individuals randomly withdraw from 50% of their potential contacts (instead of selectively avoiding those most socially distant, as in 7C). Host/pathogen equilibrium is virtually impossible to attain under these circumstances, and one population or the other rapidly becomes extinct. However, given the dynamic instability of the system, it is difficult to forecast which population will go extinct first on the basis of the epidemic's early behavior. As a result, Figure 7E shows how conventional linear statistical analyses fail to accurately forecast observed disease trajectories due to their highly unsmooth derivatives (dashed lines represent a 95% prediction interval based on ARIMA 1,1,0 time-series analysis of the first 30 mortality prevalence observations). Figure 7F gives a phase space for the epidemic (number of infected hosts at time t vs. $t - 1$), which shows kinetics that are neither classically chaotic (smooth-curved) or randomly stochastic (scattered), but migrate noisily around the autoregressive major diagonal. In all of the models in Figure 7, contacts are realized at an average rate of 1 per unit time with a probability inversely proportional to the square of the social distance (summing to 100% per unit time), and infected agents can transmit disease for 1 time unit before the appearance of illness and 3 units thereafter. Potential network reactions to disease (7C–7F) include sick individuals withdrawing contact with partners more than 2 units of social space distant (i.e., outside their own block of 3) and healthy individuals avoiding overtly sick individuals with a success rate of 50%. Hosts begin with a resistance of .2 (probability of infection, given exposure), which drops to .05 for those with no social contact in the previous time epoch. From a public-health perspective, Figure 7E is the key result, showing that initially smooth disease trajectories provide a poor basis for predicting the subsequent spread of disease in the face of realistic network structures and link dynamics. Sparsity-driven discretization drives this unpredictability by generating frequent opportunities for the bifurcation of disease trajectories, as observed in Figures 2H, 2J, 3C, 4D, 5B, 6B, and 6E. Even epidemics that have thoroughly "burnt into" a population can suddenly sinter out or explode because they are maintained in a perpetual state of knife-edge criticality by reactive network dynamics (e.g., Figure 7D).

3.4. Evolutionary Consequences

In the context of such highly leveraged systems, weak interventions can have powerful effects (13). The strength of an intervention is often analyzed in terms of its individual impact, but the key to protecting a network lies in an intervention's breadth and consistency. Perfect protection of an individual has little epidemiologic impact if disease can reach the same destination through another

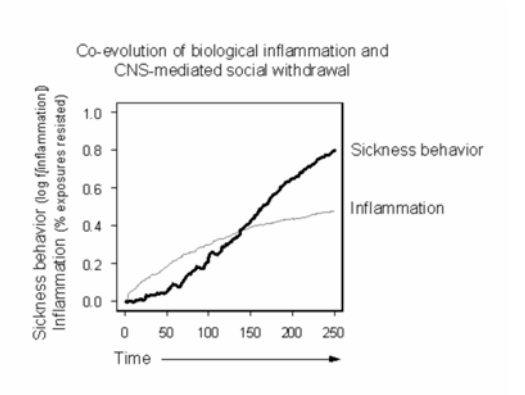


Figure 8. ***Evolution of disease-reactive social behavior***. In this evolutionary model, hosts reproduce and pass a noisy copy of two individual characteristics to their offspring: (i) an inflammatory response governing the probability of infection following exposure, and (ii) a sickness behavior parameter that drives probabilistic social withdrawal by the sick after infection. Inflammation is linearly related to resistance (50% of exposures are resisted when inflammatory response = .5), and it is costly in the sense that individual lifespan is shortened by its square (.5 inflammation induces a 25% reduction in average lifespan). The light line shows strong selective pressure for increased inflammatory responses that begins to decelerate at ~40% resistance as its cost mounts. The sickness behavior parameter (dark line) shows slow initial growth that begins to accelerate as inflammatory responses become more pronounced (the log of the mean sickness gain parameter is plotted for comparison with the linear inflammatory parameter). Once inflammatory responses begin to reach their cost-induced limits, considerable selective advantage begins to accrue to those who reduce social behavior during inflammatory reactions.

path (recall the stringent efficiency requirements for a successful quarantine in Figure 4A). On the other hand, even weak individual protection can generate strong herd immunity if it is widely distributed (4,14). The preeminent value of consistency suggests that there should be strong selective pressure to develop heritable genetic structures that reduce social contact during infection, even if the individual impact is small. The evolving ActiveHost model of Figure 8 shows just this phenomenon in a model monitoring the emergence of disease-reactive behavioral dynamics via neural reception of cytokine signals. In this model, a population of 100 individuals, each randomly linked to 5 others, is attacked by a fatal infection that is transmissible for 2 time units prior to the appearance of illness signs and 18 time units subsequently. The host population begins with a inflammatory response of 0 (all exposures result in infection), and parents pass on to their progeny their own resistance level modified by a small noise factor. Inflammation is linearly related to resistance (50% of exposures are resisted when inflammatory response = .5) and inflammation is costly in the

sense that individual lifespan is shortened by the square of its intensity (.5 inflammation results in a 25% reduction in average lifespan) (15). Sickness behavior is modeled as a multiplicative link between the magnitude of the inflammatory response and the fractional reduction in social contacts realized during inflammation (sickness behavior sensitivity parameter = 1.0 initially for all individuals, with the parental sensitivity value passed on to progeny with 5% noise, as described above for the inflammatory response). Each individual produces 2 progeny at random times between 13 and 40 years of age and, in the absence of infection, dies of natural causes at a normally distributed age with a mean of 40 and standard deviation of 10. The population mean value of the inflammatory response parameter and the sickness behavior sensitivity parameter are averaged over 20 simulations of 250 time units (~10 generations). The gray line in Figure 8 reveals a strong selective pressure for increased inflammatory response that begins to decelerate at ~40% resistance as the costs of septic shock, autoimmunity, and infertility begin to offset protection from infectious disease. In contrast, the sickness behavior sensitivity parameter (black line) shows slow initial growth that subsequently accelerates as inflammatory responses become pronounced (the log of the mean sickness gain parameter is plotted for comparison with the linear inflammatory parameter). During the early evolution of the immune system (0–100 time units), there is little selective pressure to link social behavior with the more directly effective inflammatory responses. The sickness behavior sensitivity parameter multiples the effects of biological inflammation, so it fails to evolve much while the basic inflammatory response is weak. However, once inflammatory responses begin to reach their cost-induced limits (15), benefits begin to accrue to those who reduce social contact in response to inflammation. Interestingly, in models that suppress the emergence of sickness behavior, the biological immune response evolves more rapidly and reaches a higher asymptotic equilibrium. This suggests that behavioral immune response contributes significantly to total host protection. Such results also imply that we may have been spared higher rates of inflammatory disease by the emergence of CNS-mediated sickness behavior. This is certainly consistent with the observation that vertebrate physiology dedicates substantial molecular resources to communication between the disease-sensing immune system and the behavior-controlling nervous system (2,16,17). Similar selective pressures may also have shaped mammalian brains to prefer small clustered social systems rather than large herds. Neither structural sparsity or disease-reactive linkage slows an epidemic much in isolation, but their combination can be decisive. In the model of Figure 7, for example, the joint effects of a clustered social structure and disease-reactive connectivity are equivalent to a 60-fold increase in the strength of biological immunity. The sick are unlikely to take much consolation from the fact that their immune systems synthesize most of their suffering to protect others, but they may find some comfort in considering the more malevolent blood they have been spared.

4. DISCUSSION AND CONCLUSIONS

The combined effect of structured social networks and disease-reactive contact dynamics is to force smoothly growing epidemics into unstable kinetic regimes where they can be easily extinguished. Sociospatial sparseness interacts with temporal sparseness to generate decisive qualitative boundaries from the mass action of individually quantitative behaviors. The emergent social immune response need not involve any central coordination, but can easily arise from the widespread distribution of comparatively simple rules regarding the behavior of sick individuals. The computational simplicity of these rules and their aggregated survival consequences provide considerable pressure for the evolution of molecular crosstalk between the nervous and immune systems. Indeed, the present analyses suggest that sickness behavior has coevolved in concert with the biological immune response to provide a synergistic set of defenses against infectious pathogens. The results of these analyses also suggest that our biological immune system might have become considerably more menacing had it not developed a sociobehavioral ally.

The present studies examine "sickness behavior" in the context of agent-based models, but illness-reactive behavior can also be analyzed in more traditional algebraic simulations by varying the contact rates that mix susceptible and infected individuals (4). Such variations can damp epidemics that would otherwise oscillate, shift the basal prevalence of disease, or kick stable epidemics into truly chaotic behavior, all depending upon the exact specification of the feedback function (nonlinear? time-lagged?) and whether or not it affects other parameters aside from the mixing rate (e.g., whether falling contact rates also reduce host resistance). However, agent-based analyses have several advantages over algebraic models in forecasting epidemic trajectories, particularly in terms of realistic confidence bounds. First, complex real-world social structures are more easily encoded in the explicit interaction matrices shown in Figure 2 than they are in analytically tractable continuous functions that modulate population-wide mixing rates. This is especially helpful in assessing the impact of small behavior changes generated in reaction to locally available information. Agent-based models also provide an opportunity to analyze network-mediated distribution of recursive operators that reshape individual behavior, host–pathogen dynamics, or linkage matrices depending on the realized course of an epidemic (e.g., dispersing and reconstituting groups). A third advantage is the natural discreteness of agent-based models in regions of spatiotemporal sparseness. As noted above, this is key to understanding the epidemic-extinguishing behavior of dynamic host networks, and natural discretization reduces the likelihood that minor, seemingly ignorable boundary conditions will propagate into large prediction errors. Figure 7D shows a prototypic example—an epidemic simultaneously subject to all of the influences considered above, including a complex clustered social structure with a small number of intergroup contacts, behavioral

reactions to disease by both sick and uninfected individuals, and heterogeneous basal host resistance that varies as a function of social contact. The observed disease trajectories show knife-edge dynamics, with epidemics burning slowly through a population for a variable period of time before either collapsing or exploding. These "time-bomb" kinetics are critical to recognize, but they would not be apparent from the "expected-value" disease trajectories generated by traditional algebraic models (e.g., the falsely consoling prediction limits of Figure 7E). The qualitative significance of variability is also apparent in the contrast between Figure 7A and 7C. Mean trajectories are comparable, so conventional epidemiologic models would suggest little difference. However, host populations consistently survive in panel A whereas large segments of society are often annihilated in C. The significance of that difference transcends public health to reach the level of evolutionary extinction.

Throughout their evolutionary histories, vertebrates and their parasites have each shaped the other's development (1,14,17,18). The present studies suggest that a similar reciprocal dynamic may have occurred in the evolution of the immune and nervous systems. As biochemical crosstalk between these two systems becomes increasingly appreciated, one teleologic perspective has emerged to suggest the immune system inhibits social behavior to maximize its own claim on physiologic resources (3,8). The present analyses support an alternative view in which biologically induced sickness behavior generates an emergent social immune response that operates in synergy with leukocytes to defend its genome at a species-wide level. From this perspective, the jaggedly unpredictable disease trajectories seen in many of these studies testify to the close fight between socially defended hosts and their socially predatory pathogens. Both the immune system and the nervous system have evolved under the weight of this pressure, and the present results suggest that they are more likely to collaborate than compete in response.

5. ACKNOWLEDGMENTS

This work was supported by the National Institutes of Allergy and Infectious Disease (AI49135, AI52737), the James L. Pendelton Charitable Trust, and a visiting scholarship from the Santa Fe Institute for Complex Systems.

6. APPENDIX

This appendix provides more detail on the implementation of ActiveHost as summarized in Figure 1. This agent-based modeling system is composed of four basic objects: (a) a **SimulationSystem** that creates multiple instances of a given

epidemic and summarizes the results, (b) an **Epidemic** object representing the realized course of a single epidemic, (c) a **Host** object that interacts with other Hosts as part of a disease transmission network, and (d) a **Pathogen** object that can be transmitted among Hosts and interacts with the Host to modify its health status. The basic properties and operations of each object are summarized below:

Object: SimulationSystem

Properties

- Number of simulations (user input integer)

Operations

- *Run Simulation*
  - For each simulation cycle, create and initialize an **Epidemic** object (described below)
  - Collect results on host health status as **Epidemic** iterates over time (described below)

- *Plot Results* for each simulation (from the **Epidemic**'s *Summarize Course* operation below) and compute summary statistics across the ensemble of simulations
  - Mean fraction of **Host** population uninfected, infected, visibly sick, and dead at each timepoint
  - Mean fraction of **Host** population alive at end of simulation
  - Mean time to **Host** population extinction
  - Mean time to **Pathogen** population extinction
  - Peak prevalence of infection over time
  - Median time of peak infection prevalence

Object: Epidemic

Properties

- Duration of simulation (user input integer time units)
- Initial number of **Hosts** infected (user input integer)
- Realized **Host** population trajectory (**Hosts** alive at each time unit—product of the *Run Epidemic* operation below)
- Realized **Host** health status trajectory (number and percentage uninfected, latently infected, visibly infected, and dead at each time point—product of the *Run Epidemic* operation below)
- Final **Host** population status (population survives or perishes within duration of simulation—product of the *Run Epidemic* operation below)

Operations

- *Initialize Epidemic*
  - Create the designated number of **Host** objects
  - Allocate potential linkages to each **Host** according to a user-specified rule
    - Random (each **Host** is targeted by N randomly selected source agents, where N is a user-specified integer)
    - Reciprocal (all contact links are bidirectional)
    - Small-world (**Hosts** are randomly assigned varying numbers of links according to a power law as described (5,6), such that small numbers of **Hosts** are highly connected and most hosts are sparsely connected. User specifies the distribution in terms of the total number of links in the contact network.)
    - Block (user-specified groups of N **Hosts** all share reciprocal linkage)
    - Block +1 (as above, with one member of each block also reciprocally linked to the next adjacent block)
    - Band (each **Host** is linked to a user-specified N socially adjacent Hosts)
    - Superimposed combinations of the options above
  - Allocate infections to a user-specified number of **Hosts**
  - Allocate all other user-specified **Host** characteristics (each detailed in **Host** below)
    - Natural lifespan
    - Reproduction age
    - Contact realization rules
    - **Host** resistance to infection
    - **Host** resistance dynamics rule
- *Run Epidemic*
  - For the number of time units specified by duration of simulation, execute the *Age 1 Time Unit* operation for each **Host** (described below)
  - Update the realized host population and health status trajectories after each time
- *Summarize Course*
  - Provide the final trajectory of realized **Host** population size, realized **Host** health status frequencies, and final **Host** population status

Object: Host

Properties

- Social position (integer ranging from 1 to initialized **Host** population size)
- Current age (increases by 1 with each call of the *Age 1 Time Unit* operation below)
- Health status (currently uninfected, latently infectious, subjectively sick and infectious, visibly sick and infectious, recovered, or dead)
- Host resistance to infection (probability of remaining uninfected following contact with an infected source **Host**; parameter values range from 0% to 100% and may be modified by the dynamic host resistance rule below)
- Dynamic host resistance rule
  - **Host** resistance remains constant at its initialized value, or
  - Drops by a user-specified factor when no contacts are realized at time $t - 1$
- Potential contact list
  - Sources of contact (list of **Hosts** that can transmit disease to this **Host**)
  - Targets of contact (list of **Hosts** that can be infected by this **Host**)
- Contact realization rules (combinations of the following):
  - Constant per unit time (realize user-specified N contacts per unit time)
  - Social quarantine (user-specified probability that an uninfected **Host** will avoid contact with a visibly infected source **Host**)
    - Withdraw contact from all "visibly infectious" source **Hosts**, or
    - Withdraw contact according to a constant user-supplied probability (e.g., avoid contact with a visibly infectious **Host** with 80% success)
  - Sickness behavior (user-specified option for infected **Hosts** to withdraw from social contact once they become subjectively sick)
    - Withdraw all contacts with other **Hosts**
    - Withdraw contact according to a constant user-supplied probability (e.g., avoid contact with probability .80)
    - Withdraw contact with a probability proportional to this **Host's** social distance from the target (contact avoidance probability is a linear or quadratic function of |source social position – target social position|)

    - Maintain constant contact
        - If social quarantine rules prevent contact with an infected source **Host**, this user-specified option forces the target **Host** to seek another source **Host** for interaction.
- Reproductive ages (user-specified number of parturitions at integer **Host** ages randomly selected from a uniform distribution with user-specified minimum and maximum. Call the *Reproduce* operation below when current age = one of the reproductive ages)
- Inflammatory response
    - Factor modifying initialized host resistance parameter such that realized resistance = CDF^{1}_{Normal}[Inverse CDF^{2}_{Normal}(initialized resistance) + Inverse CDF^{2}_{Normal}(inflammatory factor)], with inflammatory factor being a **Host**-specific constant that is either initialized at a user-specified constant or transmitted with noise from the parent as described below in *Reproduce*
    - Costly in that natural lifespan decreases the square root of the inflammatory response parameter, such that realized natural lifespan = initialized natural lifespan $*$ (1/inflammatory response factor squared)
- Sickness behavior sensitivity parameter
    - Factor modifying the initialized sickness behavior sensitivity parameter such that the realized probability of avoiding social contact when sick = initialized contact probability rule result $*$ sickness behavior sensitivity parameter $*$ inflammatory factor, with sickness behavior sensitivity being a **Host**-specific constant that is transmitted with noise from the parental Host as described below in *Reproduce*
- Natural lifespan (health status becomes dead when age >= this value; initialized for each **Host** as a random integer from a normal distribution with user-specified mean and standard deviation, and potentially modified by the inflammatory response parameter above)

Operations

- *Age 1 Time Unit*
    - If health status of this **Host** is not dead, then
        - Increase current age by 1
        - Update **Host** resistance based on number of contacts realized at time $t - 1$
        - Survey this **Host's** resident **Pathogen** objects' *Progress Disease* operator to update **Host's** current health status

        - Realize contacts at time *t* according to the current contact realization rule (including effects of Social quarantine and Sickness behavior)
        - Scan realized contacts for infectious status, and for each, become infected with its **Pathogen** object with probability = (1-host resistance)
        - Reproduce if current age $\in$ {reproduction ages}

- *Reproduce*
    - Clone the current **Host** and initialize is properties:
        - Current age = 0
        - Health status = uninfected
        - Initialize lifespan, reproduction ages, potential contact list, contact realization rules, inflammatory response, host resistance, and dynamic host resistance rules as for parent
        - If the user elects noisy reproduction, initialize:
            - Progeny inflammatory response = CDF^{1}_{Normal}[Inverse CDF^{2}_{Normal}(parental inflammatory response value) + a random deviate from a normal distribution with mean = 0 and standard deviation = user-specified noise factor]
            - Progeny sickness behavior sensitivity = CDF_{Normal} [Inverse CDF_{Normal} (parental sickness behavior sensitivity value) + a random deviate from a normal distribution with mean = 0 and standard deviation = user-specified noise factor]
    - Add progeny to the **Host** population

Object: Pathogen

Properties

- Infectious potential (user-specified probability of achieving infection given exposure, 0–100%)
- Pathogenic status (replicating or latent, based on duration of time since initial infection)
- Growth rate (user-specified multiplicative growth factor per unit time)
- Duration of latent infectiousness (user-specified integer number of time units in which **Pathogen** is infectious without causing visible sickness)
- Duration of visible sickness (user-specified integer number of time units in which **Pathogen** is infectious and causes visible sickness)

- Fatal vs. recoverable infection (user-specified option: does infected **Host** die after infectious period?)
- Duration of latency (user-specified option: Can an infected **Host** cease being infectious after initial infectious period but resume being infectious at some later time point? If so, "reactivation from latency" is realized as a random integer from a uniform probability distribution with a user-supplied minimum and maximum value)

Operations

- *Replicate*
  - Expand the **Pathogen** population in this **Host** by a factor = user-assigned growth rate / current host resistance
- *Progress Disease*
  - With each execution of the **Host's** *Age 1 Time Unit* operation, update the Pathogenic status by comparing duration of time since infection with initialized values of (a) duration of latent infectiousness, (b) duration of visible sickness, (c) duration of latency, and (d) fatal vs. recoverable infection.

7. NOTES

1."CDF_{Normal}" denotes the Normal (mean = 0, SD = 1) cumulative distribution function mapping a real value from $-\infty$ to $+\infty$ into the probability range 0–1.

2."Inverse CDF_{Normal}" denotes inversion of the CDF_{Normal} function to map a probability between 0 and 1 into a Standard Normal Deviate (mean = 0, SD = 1) value ranging from $-\infty$ to $+\infty$

8. REFERENCES

1. Dobson AP. 1988. The population biology of parasite-induced changes in host behavior. *Q Rev Biol* **63**:139–165.
2. Dantzer R, Bluthé RM, Castanon N, Chauvet N, Capuron G, Goodall G, Kelley KW, Konsman JP, Layé S, Parnet P, Pousset P. 2001. Cytokine effects on behavior. In *Psychoneuroimmunology*, 3rd ed., pp. 703–727. Ed. R Ader, DL Felten, N Cohen. Academic Press, New York.
3. Hart BL. 1988. Biological basis of the behavior of sick animals. *Neurosci Biobehav Rev* **12**(2):123–137.
4. Anderson RM. 1982. *The population dynamics of infectious diseases*. Chapman and Hall, New York.
5. Newman MEJ, Watts DJ, Strogatz SH. 2001. Random graph models of social networks. In *Technical report of the Santa Fe Institute*. Santa Fe Institute, Santa Fe, NM.
6. Watts DJ, Strogatz SH. 1998. Collective dynamics of "small-world" networks. *Nature* **393**:440–442.
7. Ader R, Cohen N, Felten D. 1995. Psychoneuroimmunology: interactions between the nervous system and the immune system. *Lancet* **345**:99–103.

8. Sapolsky RM. 1994. *Why zebras don't get ulcers: a guide to stress, stress-related diseases, and coping*. Freeman, New York.
9. Cassel J. 1976. The contribution of the social environment to host resistance. *Am J Epidemiol* **104**(2):107–123.
10. House JS, Landis KR, Umberson D. 1988. Social relationships and health. *Science* **241**:540–545.
11. Nunn CL, Gittleman JL, Antonovics J. 2000. Promiscuity and the primate immune system. *Science* **298**:1168–1169.
12. Garret L. 1994. *The coming plague: newly emerging diseases in a world out of balance*. Farrar, Straus and Giroux, New York.
13. Gladwell M. 2000. *The tipping point: how little things can make a big difference*. Little, Brown, Boston.
14. Burnet M. White DO. 1972. *Natural history of infectious disease*. Cambridge UP, London.
15. Weiss U. 2002. Nature insight: inflammation. *Nature* **420**:846–891.
16. Blalock JE. 1994. The syntax of immune-neuroendocrine communication. *Immunol Today* **15**:504.
17. Pulendran B, Palucka K. Banchereau J. 2001. Sensing pathogens and tuning immune responses. *Science* **293**:253–256.
18. May RM, Anderson RM. 1983. Epidemiology and genetics in the coevolution of parasites and hosts. *Proc Roy Soc London B* **219**:281–313.

7.2

PHYSIOLOGIC FAILURE: MULTIPLE ORGAN DYSFUNCTION SYNDROME

Timothy G. Buchman

Edison Professor of Surgery, Washington University School of Medicine, St. Louis, Missouri

Functional integrity of complex organisms (including man) requires physiological adaptation to ordinary and extraordinary stress. When stress exceeds adaptive capacity, one or more physiologic systems "fails"; without intervention, the organism dies. Clinical medicine offers system-specific supports that have proven necessary but often insufficient to promote recovery of function despite anatomic integrity and relief from the inciting stress. Either the underlying relationships of the physiological adaptive systems have been substantially altered or the depth of the basin of attraction described as multiple organ dysfunction is sufficiently deep to make escape improbable using current organ support strategies. Experimental alternative organ system support strategies that emulate healthy biological variability have accelerated recovery of dysfunctional organ systems.

1. INTRODUCTION

Humans, like other life forms, can be viewed as thermodynamically open systems that continuously consume energy to maintain stability in the internal milieu in the face of ongoing environmental stress. In contrast to simple unicellular life forms such as bacteria, higher life forms must maintain stability not only in individual cells but also for the organism as a whole. To this end, a collection of physiologic systems evolved to process foodstuffs; to acquire oxygen

Address correspondence to: Timothy G. Buchman, Washington University School of Medicine, 1 Barnes-Jewish Plaza, 6104 Queeny Tower, St. Louis, MO 63110 (buchman@msnotes.wustl.edu).

and dispose of gaseous waste; to eliminate excess fluid and soluble toxins; and to perform other tasks. These systems—labeled respiratory, circulatory, digestive, neurological, and so on—share several features.

- Physiological systems are spatially distributed. Food has to get from mouth to anus. Urine is made in the kidneys but has to exit the urethra. Blood may be pumped by the heart but has to reach the great toe.
- Physiologic systems are generally space-filling and structurally fractal. Each cell has demand for nutrients; each cell excretes waste. Information has to travel from the brain throughout the body. While not every physiologic system is self-similar at all levels of granularity, there is typically a nested architecture that facilitates function: microvillus to villus to intestinal mucosa; alveolus to alveolar unit to bronchial segment; capillary to arteriole to artery.
- Physiologic systems are functionally integrated. After eating, blood is redistributed to the gut and splanchnic circulation. When alveoli become atelectatic, blood is shunted away from these hypoxic regions. Ingestion and delivery of excess fluid to the circulation is quickly followed by augmented production of urine.
- Physiologic systems have characteristically variable time signatures that lose their variability in aging and disease (1). Instantaneous cardiac and respiratory rates vary from one event (heart beat or breath) to the next. Many hormones exhibit not only diurnal variation, but also a superimposed pattern or irregular pulses. The product of physiologic systems—such as gait, which combines neural, musculoskeletal, cardiac, and respiratory systems into a semivoluntary activity (one usually does not think about putting one foot in front of the next)—display characteristically variable time signatures.

The first three features of physiologic systems have medical consequences. Both aging and illness can compromise one or more physiologic systems. Management of such compromise was, until about fifty years ago, directed exclusively toward minimizing the performance demand placed on the system. For example, in advancing pulmonary insufficiency, patients were progressively confined to home, to chair, and finally to bed. Each organ system had a critical level of compromise, and once the compromise exceeded the critical level, the patient simply died.

Two major advances in the last half-century have changed the clinical trajectory. The first major advance was the development of mechanical supports for failing mechanical systems. Ventilators (respirators), ventricular assist devices, and renal replacement therapies (dialysis machines) have come into widespread clinical use. These supports have evolved to the point where individual

system failures can be managed by outpatient use of these devices: many patients thus gain years of useful life. The second advance was the development of organ transplantation. Beginning with blood components (transfusion is a type of transplantation) and progressing to kidney, heart, lung, pancreas, and intestine, component replacement is increasingly frequent and relatively safe. Both advances—mechanical support and tissue transplantation—compromise immune system function, but this compromise is usually a good trade for survival.

It is hardly surprising that organ dysfunction and failures accumulate. Like an aging automobile with worn bearings, cracked hoses, and leaky engine valves, many humans eventually acquire an illness to which they cannot successfully respond even with medical care—a bleed into the brain, a metastatic cancer, or a high-speed motor vehicle crash. They die with multiple organs failing to perform their appropriate function. The subject of this chapter, however, is a syndrome of widespread, progressive, and disproportionate multiple organ dysfunction (MODS) that rapidly accumulates following a minor or modest insult (2,3). Despite timely and appropriate reversal of the inciting insult—whether a pneumonia, intraabdominal abscess, pancreatitis, or simply the stress of an anesthetic and elective surgery—many patients develop the syndrome. Mortality is proportional to the number and depth of system dysfunction (4), and the mortality of MODS after, for example, repair of ruptured abdominal aortic aneurysm, is little changed despite three decades of medical progress (2,5). Unfortunately, MODS remains the leading cause of death in most intensive care units.

2. PREVIOUS WORK

2.1. MODS: The Phenotype

Autopsy findings in patients who succumb to MODS are surprisingly bland. Tissue architecture is preserved, cells do not appear abnormal, and there is no widespread thrombosis. At least anatomically, the body appears to be largely intact. (The exception is lymphatic tissues, which are often exhausted through accelerated programmed cell death [apoptosis] (6).) Nor does organ function appear to be irretrievably lost: among MODS survivors, many—especially younger survivors—experience return of multiple organ performance to levels approaching that which they enjoyed prior to the syndrome (7). These two observations—anatomic integrity and the potential for near-complete recovery—led to replacement of the old descriptor ("multiple organ failure") with the current and more apt label of "multiple organ dysfunction syndrome." Using the jargon of information technology, the focus shifted from the hardware to the software (8).

Bearing in mind that MODS is only three decades old (multiple organ supports had to be developed and used in enough patients before it could be ob-

served that the patients were dying despite the treatments), its phenotype has changed somewhat as physicians have tried to preempt its occurrence and progression. In its original—and perhaps purest—form, organ dysfunctions would accumulate in a more or less predictable sequence (2,3). The lungs would fail first, and the patient would require intubation and mechanical ventilation. A few days later, evidence of gut and liver failure would appear—patients would fail to absorb nutrients and fail to manufacture critical proteins such as clotting factors. Artificial nutrition and transfusion medicine were therefore administered. A few days after that, kidney failure would become apparent and the patient would require dialysis. Not only was the dysfunction sequential, but this particular sequence precisely mirrored the sequence in which organs matured in fetal life—kidneys first, then the liver/gut, and last, the lungs. For this reason, multiple organ failure began to be recast as organ systems "falling off line," each functionally separating from the whole. For this reason, physicians initiate organ-specific support earlier and earlier—at the first sign of dysfunction. This strategy of early intervention has muddied the failure sequence, unfortunately with little effect on outcome: four-organ failure is still quite lethal.

2.2. Physiologic Stability

How do organisms maintain function in the face of external stress? There appear to be two general ways (9). One relies on purpose-specific mechanisms that have arisen and been refined in the course of evolution. At the resolution of the individual cell, the "stress response"—originally called the "heat shock response" because it was observed in polytene chromosomes of *Drosophila* cells exposed to high temperatures—activates specific transcription factors, modulates RNA splicing, and applies selection filters to translation. The phenotype of this stress response is marked alteration in protein synthesis while the cell becomes (temporarily) refractory to additional external stimuli. At the resolution of the intact organism, circulating blood sugar levels are maintained by the secretion of insulin and of glucagon, which sequester and mobilize (respectively) carbohydrates. Such engineered mechanisms have been identified at all levels of granularity, and their product was termed "homeostasis" by Walter B. Cannon early in the twentieth century. A central dogma of medical care as articulated by Cannon instructs the physician to render "external aid" when homeostatic mechanisms are overwhelmed by disease. This has been translated by the medical community into the "fix-the-number" imperative: if the bicarbonate level is low, give bicarbonate; if the urine output is low, administer a diuretic; if the bleeding patient has a sinking blood pressure, make the blood pressure normal. Unfortunately, such interventions are commonly ineffective and even harmful (10,11). For example, sepsis—a common predecessor of MODS—is often ac-

companied by hypocalcemia. In controlled experimental conditions, administering calcium to normalize the laboratory value increases mortality (12).

The other fount of stability is network structure. At the same time that Cannon described homeostasis, Lawrence J. Henderson—a colleague of Cannon's at Harvard—studied the physiology of blood with an emphasis on its capacity to buffer changes in pH brought about by addition of fixed acids or carbon dioxide (13). Henderson recognized and provided a mathematical foundation for this buffer system, which even today is known eponymously as the Henderson-Hasselbach equation. He observed that the mere presence of a system of interacting components conferred stability. Chauvet has provided a framework for the study of formal biological systems and has reached two important conclusions. First, such stability is not unique to blood or other buffers but rather is an expected consequence of interactions among biological elements. Second, nesting such systems—this may be thought of as organelles into cells, cells into tissues, tissues into organs, and so on—confers further stability onto formal biological systems (14).

Biological stability must not be confused with invariance—survival of organisms from bacteria to man requires adaptive capacity. Purpose-specific homeostatic mechanisms typically include engineering features such as negative feedback that provide such adaptive responses. Self-aggregating networks are another matter: once a fitness maximum (or energy minimum) is reached, simulations of such networks often assume a trivial and biologically useless trajectory in state space, such as occupying a single point or endlessly traversing a few points. How might the characteristic variability and long-range correlations of biological signals (for example, heart rate variability; see this volume, Part III, chapters 3.3 (by Glass) and 7.3 (by Lipsitz) for additional information) arise?

3. MODEL

3.1. Coupling and Uncoupling

Several years ago, we suggested that the networks of organ systems that collectively constitute macrophysiology are not only coupled, but also that the couplings are intrinsically unstable (15). Subsequently, Schaefer and colleagues presented data in support of this conjecture: analysis of the interaction between cardiac and respiratory cycles in healthy athletes at rest suggested that the coupling between heart and lungs was not fixed but rather dynamic (16). These athletes' organs would couple (for example, 5 heartbeats for 2 respirations), then uncouple, then recouple at the same or a different ratio (for example, 6 heartbeats for every 2 respiratory cycles). The inference is that health may be associ-

ated with a search through the space of possible interactions to find the one best suited to current physiologic challenges.

Experimental manipulation of the connections suggests that physiologic optima do indeed exist. Hayano and colleagues experimentally interrogated the relationship among cardiac cycles, respiratory cycles, and the vagally mediated respiratory sinus arrhythmia that reflects central respiratory drive and the lung inflation reflex in dogs (17). These investigators electrically paced the diaphragm, applied electrical stimuli to the vagus nerve to simulate normal, absent, or inverse respiratory sinus arrhythmia, and measured the matching of lung ventilation with perfusion, which is critical to healthy physiology. The data showed that normal respiratory sinus arrhythmia (i.e., physiologic coupling of respiratory and cardiac cycles) minimized wasted ventilation (dead space) and perfusion (shunt fraction), whereas the inverse arrhythmia was physiologically much less efficient. These investigators suggest that respiratory sinus arrhythmia is an intrinsic resting function of the cardiopulmonary system that provides a continuous fitness maximum for the coupled heart–lung system (18).

3.2. MODS: Uncoupled Oscillators?

MODS is not a disease but rather a syndrome, a common pathway that is all too often final. Yet some patients do recover. Two features of recovery are invariant. First, the time to recover is significantly longer than the time to become ill. Second, measured physiologic parameters do not retrace their paths, implying hysteresis in the clinical trajectory. These features led to speculation that MODS did not follow a specific event, but rather reflected a more general phenomenon of network failure at multiple levels of granularity. What kind of networks might fail at the level of organ physiology? We observed that most organs had characteristic varying time signals, and further speculated that network failure might represent failure of the uncoupling/recoupling process of these biological oscillators (19). Several lines of evidence support such a conjecture.

4. RESULTS

First, it is possible to directly estimate coupling among select physiologic systems from common continuous clinical measurements such as heart rate and blood pressure. Goldstein's studies of critically ill children as diverse as those with sepsis (20) and with severe head injury (21) suggest loss of heart rate/blood pressure coupling as patients deteriorate, and recovery of transfer function as the patients themselves recover. Second, Pincus' conjecture—that loss of variability implies greater system isolation (uncoupling) between systems that contain stochastic components—allows for additional inferences based on heart rate infor-

mation alone (22). Godin and colleagues demonstrated precisely such loss of variability in humans experimentally exposed to bacterial endotoxin, a common predecessor of MODS (23). Seiver and Szaflarski (24) reported a startling loss of physiologic variability with the appearance of monotonous sinusoidal variation in cardiac output among critically ill humans. Winchell and Hoyt have shown that loss of heart rate variability in critically ill patients is a predictor of death (25).

5. IMPLICATIONS FOR TREATMENT

If MODS is the clinical expression of network recoupling failure, then therapy might logically be directed toward facilitating that recoupling. Paradoxically, severe illness prompts physicians to suppress biologic variation in many organ systems. For example, ventilators are set to fire at fixed intervals, catecholamines are infused at fixed rates, fixed composition nutrition is administered without interruption, venovenous hemofiltration is conducted at a fixed rate around the clock, and so on. Such rigidity invites perceptions of therapeutic success: "the patient in bed 21 is now stable as a rock." Perhaps the more important question is whether such therapeutic rigidity promotes or suppresses clinical recovery.

Although no trials have been performed on patients with MODS, reports have begun to appear in which normal physiologic variability has been synthetically applied to the function of mechanical ventilators. Gas exchange and respiratory mechanics are improved by biologically variable ventilation not only in models of lung injury but also in healthy lungs (26,27). More to the point, a group of patients at risk for MODS—those undergoing surgical repair of abdominal aortic aneurysm—also enjoyed better lung function when the ventilation algorithm included simulated biological variation (28). Part of the improvement may reflect enhancement of the respiratory sinus arrhythmia (30).

During cardiac surgery, the perfusion of the body is supported by the "heart–lung" machine. This perfusion, called cardiopulmonary bypass, can be continuous, pulsatile at fixed sinusoidal frequency, or aperiodically pulsatile. Mutch and colleagues have demonstrated improved brain blood flow characteristics with the aperiodic (biologically variable) algorithm versus the conventional clinical techniques of constant or periodic flow (29). Suboptimal brain blood flow is clinically associated with cognitive impairment, a manifestation of central nervous system "organ failure."

While none of these data directly address MODS, they raise the disquieting possibility that conventional therapeutic rigidity that applies fixed or strictly periodic inputs to the network of dysfunctional biological systems may actually hinder recovery. The need for trials comparing monotonous versus biologically variable algorithms applied to existing therapies is evident.

6. SUMMARY AND PERSPECTIVE

The list of diseases that are *associated* with breakdown of network interactions and appearance of highly periodic dynamics continues to grow: epilepsy, fetal distress, sudden cardiac death, Parkinson's disease, and obstructive sleep apnea are among recent additions. Herein, we have suggested that breakdown of network interactions may actually *cause* disease, and when this breakdown is widespread the clinical manifestation is multiple organ dysfunction syndrome. If this hypothesis is correct, then network dysfunction might be expected at multiple levels of granularity, from organ systems to intracellular signal molecules. Restoration of network integrity may be a reasonable therapeutic goal, and a more permissive approach to clinical support (including algorithms that simulate biological variability) might facilitate restoration of network complexity, which now appears essential to health.

7. ACKNOWLEDGMENTS

This work was supported, in part, by award GM48095 from the National Institutes of Health and by an award from the James S. McDonnell Foundation.

8. REFERENCES

1. Goldberger AL, Amaral LA, Hausdorff JM, Ivanov PCh, Peng CK, Stanley HE. 2002. Fractal dynamics in physiology: alterations with disease and aging. *Proc Natl Acad Sci USA* **19**(suppl 1):2466–2472.
2. Tilney NL, Bailey GL, Morgan AP. 1973. Sequential system failure after rupture of abdominal aortic aneurysms: an unsolved problem in postoperative care. *Ann Surg* **178**(2):117–122.
3. Baue AE. 1975. Multiple, progressive, or sequential systems failure: a syndrome of the 1970s. *Arch Surg* **110**(7):779–781.
4. Cook R, Cook D, Tilley J, Lee K, Marshall J. 2001. Canadian Critical Care Trials Group: multiple organ dysfunction: baseline and serial component scores. *Crit Care Med* **29**(11):2046–2050.
5. Bown MJ, Nicholson ML, Bell PR, Sayers RD. 2003. The systemic inflammatory response syndrome, organ failure, and mortality after abdominal aortic aneurysm repair. *J Vasc Surg* **37**(3):600–606.
6. Hotchkiss RS, Swanson PE, Freeman BD, Tinsley KW, Cobb JP, Matuschak GM, Buchman TG, Karl IE. 1999. Apoptotic cell death in patients with sepsis, shock, and multiple organ dysfunction. *Crit Care Med* **27**(7):1230–1251.
7. Garcia Lizana F, Manzano Alonso JL, Gonzalez Santana B, Fuentes Esteban J, Saavedra Santana P. 2000. Survival and quality of life of patients with multiple organ failure one year after leaving an intensive care unit. *Med Clin (Barc)* **114**(suppl 3):99–103.

8. Tjardes T, Neugebauer E. 2002. Sepsis research in the next millennium: concentrate on the software rather than the hardware. *Shock* **17**(1):1–8.
9. Buchman TG. 2002. The community of the self. *Nature* **420**(6912):246–251.
10. Mehta RL, Pascual MT, Soroko S, Chertow GM. 2002. PICARD Study Group: diuretics, mortality, and nonrecovery of renal function in acute renal failure. *JAMA* **288**(20):2547–2553.
11. Bickell WH, Wall MJ Jr, Pepe PE, Martin RR, Ginger VF, Allen MK, Mattox KL. 1994. Immediate versus delayed fluid resuscitation for hypotensive patients with penetrating torso injuries. *New Engl J Med* **331**(17):1105–1109.
12. Zaloga GP, Sager A, Black KW, Prielipp R. 1992. Low dose calcium administration increases mortality during septic peritonitis in rats. *Circ Shock* **37**(3):226–229.
13. Chambers NK, Buchman TG. 2001. Shock at the millennium, II: Walter B. Cannon and Lawrence J. Henderson. *Shock* **16**(4):278–284.
14. Chauvet GA. 1993. Hierarchical functional organization of formal biological systems: a dynamical approach, I: the increase of complexity by self-association increases the domain of stability of a biological system. *Phil Trans R Soc Lond B Biol Sci* **339**(1290):425–444.
15. Godin PJ, Buchman TG. 1996. Uncoupling of biological oscillators: a complementary hypothesis concerning the pathogenesis of multiple organ dysfunction syndrome. *Crit Care Med* **24**(7):1107–1116.
16. Schafer C, Rosenblum MG, Kurths J, Abel HH. 1998. Heartbeat synchronized with ventilation. *Nature* **392**(6673):239–240.
17. Hayano J, Yasuma F, Okada A, Mukai S, Fujinami T. 1996. Respiratory sinus arrhythmia: a phenomenon improving pulmonary gas exchange and circulatory efficiency. *Circulation* **94**(4):842–847.
18. Hayano J, Yasuma F. 2003. Hypothesis: respiratory sinus arrhythmia is an intrinsic resting function of cardiopulmonary system. *Cardiovasc Res* **58**(1):1–9.
19. Buchman TG. 1996. Physiologic stability and physiologic state. *J Trauma* **41**(4):599–605.
20. Ellenby MS, McNames J, Lai S, McDonald BA, Krieger D, Sclabassi RJ, Goldstein B. 2001. Uncoupling and recoupling of autonomic regulation of the heart beat in pediatric septic shock. *Shock* **16**(4):274–277.
21. Goldstein B, Toweill D, Lai S, Sonnenthal K, Kimberly B. 1998. Uncoupling of the autonomic and cardiovascular systems in acute brain injury. *Am J Physiol* **275**(4 Pt 2):R1287–R1292.
22. Pincus SM. 1994. Greater signal regularity may indicate increased system isolation. *Math Biosci* **122**(2):161–181.
23. Godin PJ, Fleisher LA, Eidsath A, Vandivier RW, Preas HL, Banks SM, Buchman TG, Suffredini AF. 1996. Experimental human endotoxemia increases cardiac regularity: results from a prospective, randomized, crossover trial. *Crit Care Med* **24**(7):1117–1124.
24. Seiver AJ, Szaflarski NL. 2003. Report of a case series of ultra low-frequency oscillations in cardiac output in critically ill adults with sepsis, systemic inflammatory responses syndrome, and multiple organ dysfunction syndrome. *Shock* **20**(2):101–109.
25. Winchell RJ, Hoyt DB. 1996. Spectral analysis of heart rate variability in the ICU: a measure of autonomic function. *J Surg Res* **63**(1):11–16.
26. Boker A, Graham MR, Walley KR, McManus BM, Girling LG, Walker E, Lefevre GR, Mutch WA. 2002. Improved arterial oxygenation with biologically variable or fractal ventilation using low tidal volumes in a porcine model of acute respiratory distress syndrome. *Am J Respir Crit Care Med* **165**(4):456–462.
27. Mutch WA, Eschun GM, Kowalski SE, Graham MR, Girling LG, Lefevre GR. 2000. Biologically variable ventilation prevents deterioration of gas exchange during prolonged anaesthesia. *Br J Anaesth* **84**(2):197–203.
28. Boker A, Haberman CJ, Girling L, Guzman RP, Louridas G, Tanner JR, Cheang M, Maycher BW, Bell DD, Doak GJ. 2004. Variable ventilation improves perioperative lung function in patients undergoing abdominal aortic aneurysmectomy. *Anesthesiology* **100**(3):608–616.

29. Mutch WA, Warrian RK, Eschun GM, Girling LG, Doiron L, Cheang MS, Lefevre GR. 2000. Biologically variable pulsation improves jugular venous oxygen saturation during rewarming. *Ann Thorac Surg* **69**(2):491–497.
30. Mutch WA, Graham MR, Girling LG, Brewster JF. 2005. Fractal ventilation enhances respiratory sinus arrhythmia. *Resp Res* **6**(1):41–50.

7.3

AGING AS A PROCESS OF COMPLEXITY LOSS

Lewis A. Lipsitz

The Hebrew Rehabilitation Center for the Aged,
Beth Israel Deaconess Medical Center, and
Harvard Medical School, Boston

Healthy physiologic function requires the integration of complex networks of control systems and feedback loops that operate on multiple scales in space and time. When measured continuously, the output of physiologic systems is highly complex, resulting in dynamic behavior that can be described using techniques derived from fractal analysis. These fractal-like physiologic processes enable an organism to adapt to the exigencies of everyday life. During normal human aging the degeneration of various tissues and organs, and the interruption of communication pathways between them, results in a loss of complexity of physiologic systems and, consequently, a reduced capacity to adapt to stress. Therefore, relatively minor perturbations such as new medications, a viral illness, or emotional trauma may result in serious disability and death. Fortunately, a number of novel interventions may be able to restore healthy dynamics in elderly individuals and enhance their ability to adapt to a variety of external stimuli.

1. INTRODUCTION

Classical research in the field of aging has been largely reductionistic, defining the process of aging as a linear decline in many organs and physiologic systems until functional disability results (Figure 1) (1). However, aging is a nonlinear, multidimensional process that is associated not only with changes in individual systems but, probably more importantly, alterations in the

Address correspondence to: Lewis A. Lipsitz, MD, Hebrew Rehabilitation Center for the Aged, 1200 Centre Street, Boston, MA 02131 (Lipsitz@mail.hrca.harvard.edu).

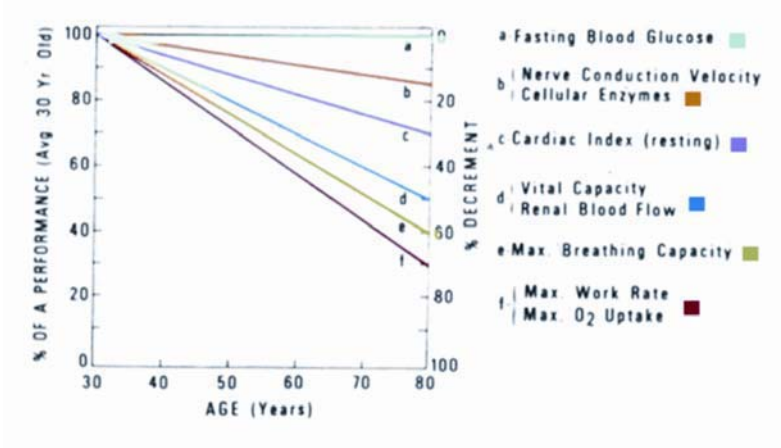


Figure 1. Age-related decrements in physiologic performance from Shock et al., the Baltimore Longitudinal Study on Aging (1). Classical research in aging examined the mean values of various physiologic functions in different age groups or over time, thus failing to recognize the complex dynamics of these processes.

connections and interactions between system components. The complex mechanisms by which these components interact to enable an organism to perform a variety of functions necessary for survival is the subject of physiology.

Healthy physiologic processes require the integration of complex networks of control systems and feedback loops that operate on multiple scales in space and time (2). For example, physiologic systems exist at molecular, subcellular, cellular, organ, and systemic levels of organization. Continuous interplay between the electrical, chemical, and mechanical components of these systems ensures that information is constantly exchanged, even as the organism rests. These dynamic processes give rise to a highly adaptive, resilient organism that is prepared to respond to internal and external perturbations.

Recognition of the dynamic nature of regulatory processes challenges the concept of homeostasis, which is taught by physiologists as a function of all healthy cells, tissues, and organs to maintain static or steady-state conditions in their internal environment (3). However, with the introduction of techniques that can acquire continuous data from physiologic processes such as heart rate, blood pressure, nerve activity, or hormonal secretion, it has become apparent that these systems are in constant flux, even under so-called steady-state conditions. Dr. Eugene Yates introduced the term *homeodynamics* to convey the fact that the high level of bodily control required to survive depends on a dynamic interplay of multiple regulatory mechanisms rather than constancy in the internal environment (4).

Although it is often difficult to separate the effects of aging from those of disease and such lifestyle changes as reduced physical activity, healthy aging in the absence of such confounding factors appears to have a profound impact on physiologic processes. Because of the progressive age-related degeneration of various tissues and organs, and the interruption of communication pathways between them, complex physiologic networks break down, become disconnected, and lose some of their capacity to adapt to stress.

There is considerable redundancy in many of the biologic and physiologic systems in higher organisms; for example, humans have far more muscle mass, neuronal circuitry, renal nephrons, and hormonal stores than they need to survive. This creates a *physiologic reserve* that allows most individuals to compensate effectively for age-related changes. Because the network structure of physiologic systems also enables alternate pathways to be used to achieve the same functions, physiologic changes that result from aging alone usually do not have much impact on everyday life. However, these changes may become manifest at times of increased demand, when the body is subjected to high levels of physiologic stress. For this reason, elderly individuals are particularly vulnerable to adverse events such as falls, confusion, and incontinence when exposed to environmental, pharmacologic, or emotional stresses.

2. MEASURES OF COMPLEXITY LOSS

With the development of monitoring devices that can measure the output of regulatory processes on a moment-to-moment basis, it has become apparent that the dynamics of many systems lose complexity with advancing age (5). The continuous heart rate time series of a healthy young subject and healthy elderly subject shown in Figure 2 provides a good example. The average heart rate over the 8-minute period of recording is 64.7 beats per minute in the young subject and 64.5 in the older subject—nearly identical. Furthermore, the standard deviation of the heart rate is also nearly identical: 3.9 in the young subject and 3.8 in the elderly subject. However, it is apparent that the dynamics of the heart rate time series are strikingly different. Until recently, scientists have lacked the tools to describe these dynamics. However, new advances in the fields of nonlinear dynamics, chaos theory, and complex systems (see Part II, chapters 1 (by Shalizi) and 2 (by Socolar) this volume) have provided new ways to quantify the aging process and understand its mechanisms. One particularly useful concept that can be used to quantify the complexity of various anatomic structures or physiologic processes is the concept of fractals.

The definition of a fractal, first described by Mandelbrot (6), is a geometric object with "self-similarly" over multiple measurement scales. For example, many anatomic structures demonstrate self-similarity in their structures and have fractal properties. The branches upon branches upon branches of bronchi in the

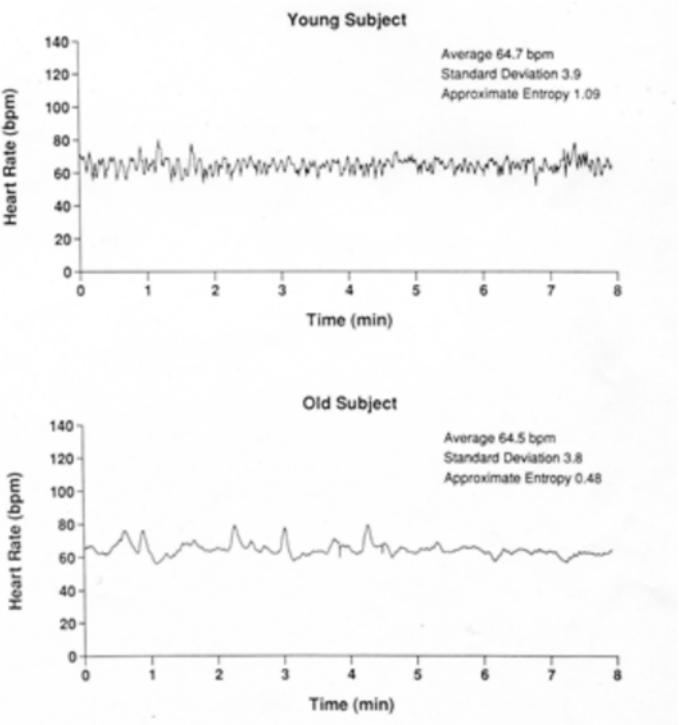


Figure 2. Continuous heart rate time series over 8 minutes for a healthy young subject (top graph) and a healthy elderly subject (bottom graph). Note the similar average heart rate and standard deviation of heart rate, but different dynamics as quantified by Approximate Entropy. Reprinted with permission from (5).

respiratory tree (7) or smaller and smaller vessels in the circulatory system (8) look similar whether they are measured in microns, millimeters, centimeters, or meters. In fact, the smaller the measuring device, the larger the length of a fractal object. This property is known as "power-law scaling" because a smaller measuring device leads to an exponential (i.e., "the power") increase in the length of a fractal object. The output of dynamic physiologic processes such as heart rate, measured over time rather than space, also have fractal properties (9). Their oscillations appear self-similar when observed over seconds, minutes, hours, or days. Furthermore, they demonstrate power-law scaling in the sense that with a smaller frequency of oscillation of these signals their amplitude increases exponentially.

The heart-rate time series demonstrates complex irregularity that has been described as a fractal process because it looks similar whether it is plotted over days, hours, or minutes. In fact, the heart rate time series has "1/*f*" or power-law scaling, in that the amplitude of oscillations (A) is proportional to the inverse of oscillation frequency, according to the formula $A \cong 1/f^{\beta}$. The exponent β can be derived from the slope of the log–log transformation of the Fourier power spectrum. A fractal process (most complex) has a slope of 1 (i.e., amplitude is inversely proportional to frequency over a wide range of frequencies, indicating the presence of long-range correlations (self-similarity) in the data). A loss of complexity occurs as the slope approaches 0 (white noise; i.e., there is no relation between amplitude and frequency) or 2 (Brownian noise; i.e., the relation between amplitude and frequency occurs only over a short range, then rapids falls off, indicating a loss of long-range correlations).

Unfortunately, computation of the power spectrum using Fourier analysis requires stationary data, which most physiologic signals are not. Another particularly useful technique that minimizes the effect of nonstationarities in the data is "detrended fluctuation analysis" (DFA), which has been well validated in a number of dynamic systems (10). The DFA algorithm is a two-point correlation method that computes the slope of the line relating the amplitude of fluctuations to the scale of measurement, after detrending the data. The root-mean-square fluctuation of the integrated and detrended data are measured in observation windows of different size and then plotted against the size of the window on a log–log scale. The slope of the regression line that relates log-fluctuation to log-window size quantifies the complexity of the data (1 = fractal, 0.5 = random, 1.5 = random-walk).

Other indicators of complexity loss in physiologic systems include an increase in periodicity (e.g., the tremor of Parkinson's disease), increased randomness (e.g., atrial fibrillation of the heart), and loss of long-range correlations (e.g., stride-length changes during gait).

It is important to recognize that complexity and variability are not the same. For example, a high-amplitude sine wave signal is quite variable, but not at all complex. Alternatively, an irregular low-amplitude signal such as the heart rate of a healthy young subject shown in Figure 2, can be quite complex but much less variable. Similarly, irregularity and complexity are not the same. Traditional entropy-based algorithms such as Approximate Entropy (11), used to quantify the regularity of a time series, indicate greater irregularity (which has been interpreted as greater complexity) for certain uncorrelated random signals associated with pathologic processes. However, as highlighted by the recent work of Costa et al. (12), these algorithms fail to account for the multiple time scales over which healthy physiologic processes operate. By calculating the entropy of heart-rate dynamics over multiple time scales (multiscale entropy analysis), these investigators were able to distinguish healthy from pathological (e.g., atrial fibrillation and congestive heart failure) dynamics, and showed

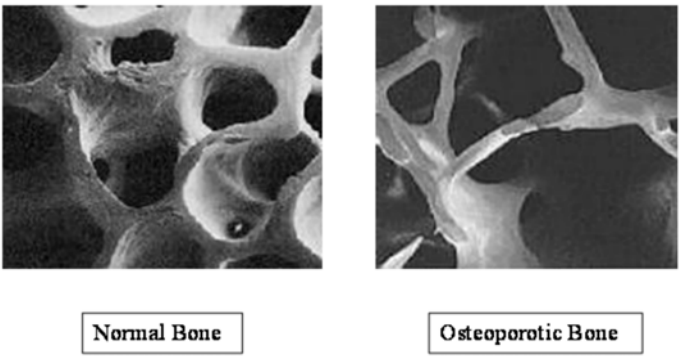


Figure 3. Loss of complexity in bone. Note the loss of connectivity and fractal-like architecture in bone with age-related osteoporosis.

consistently higher multiscale entropy values for dynamic processes with long-range (1/*f*-type) correlations, compared to those with uncorrelated random noise. Therefore, this measure appears useful for quantifying the complexity of a time series and distinguishing it from irregularity due to uncorrelated noise.

3. EXAMPLES OF COMPLEXITY LOSS WITH AGING

There are numerous examples of the progressive loss of complexity in the fractal architecture of anatomic structures and the dynamics of physiologic processes with advancing age. As shown in Table 1, structures such as neurons, bone trabeculae, kidney glomeruli, and elastic fibers or dermal papillae beneath the skin all lose structural complexity with aging (5). The loss of connectivity in bone trabeculae, characteristic of osteoporosis, is shown in Figure 3. Not only anatomic structures, but physiologic processes also lose complexity with aging. These include heart rate (13), blood pressure (13), and respiratory dynamics (14), center of pressure trajectories when measured on a balance platform (Figure 4), and gait dynamics (15) (Table 2). Moreover, complexity loss has been shown to be associated with a variety of diseases and adverse outcomes (Tables 3 and 4). On the basis of these observations, we have hypothesized that the age-related loss of complexity in physiologic systems results in an impaired ability to adapt to stress and the ultimate development of disease and disability (2).

In response to a given stress or perturbation, physiologic systems mount a specific adaptive response that restores the organism to a new dynamic equilibrium. Therefore, the dynamics observed during resting and stimulated conditions are often quite different. For example, when glucose is ingested, insulin is

Table 1. Decreased complexity of anatomic structures with aging

Structure	Measure	Age effect
Neurons	Dendritic arbor	Loss of branches and neural connections
Bone	Trabecular meshwork	Trabecular loss and disconnection
Kidney	Glomerular capillary tuft	Degeneration and loss of capillaries
Subepidermis	Elastic fibers and dermal papillae	Loss and collapse of subepidermal structure

Table 2. Decreased complexity of physiologic systems with aging

System	Measure	Age effect
Heart rate dynamics	1/*f* slope, DFA, & approx. entropy of interbeat intervals	Decreased fractal scaling, more regular dynamics
Blood pressure dynamics	1/*f* slope, DFA, & approx. entropy of BP fluctuations	Decreased fractal scaling, more regular dynamics
Respiratory dynamics	DFA of interbreath intervals	Loss of long-range correlations in elderly males
Postural control	1/*f* slope of center-of-pressure trajectories	Decreased fractal scaling
Gait dynamics	1/*f* slope, DFA	Decreased fractal scaling

secreted in pulsatile fashion in order to promote glucose metabolism. When an individual stands up, blood pools in the lower extremities, and blood pressure suddenly falls. In response, the sympathetic nervous system secretes norepinephrine to restore the blood pressure to its resting state. Thus, the complex dynamics observed during resting conditions differ from the more focused singular response that occurs during stress. The dynamics of this adaptive response have been referred to as "reactive tuning" (2). In the field of nonequilibrium statistical mechanics, the relationship between the correlation properties of the fluctuations of a system and its relaxation to equilibrium is described by the fluctuation-dissipation theorem (16). If the complex interactions of physiologic systems during rest enable an organism to mount a focused adaptive response during a perturbation, the loss of complexity in resting dynamics may indicate an impaired ability to adapt to stress and a predisposition to functional decline.

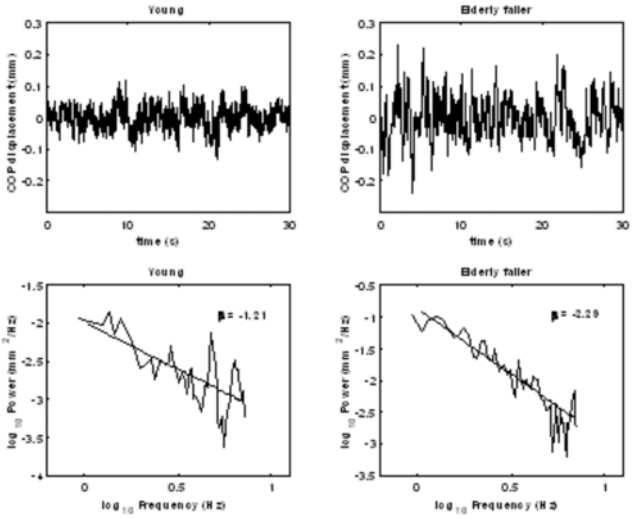


Figure 4. Center of pressure displacements (top graphs) and their corresponding log-transformed power spectra (bottom graphs) for a 30-year-old healthy woman (left) and a 69-year-old woman with previous falls (right), obtained while the subjects stood on a force-plate for 30 seconds. The slope of the regression line through the power spectra represents the scaling exponent β, which is greater in the elderly faller than in the young subject, indicating a loss of complexity. Reprinted with permission from (2).

4. MECHANISMS OF PHYSIOLOGIC COMPLEXITY

A variety of mechanisms probably underlie the complexity of physiologic systems, including neuronal networks in the nervous system, biochemical pathways in metabolic control systems, signaling pathways within and between cells, genetic switches, and transcription control elements. Two experiments highlight the importance of the autonomic nervous system in generating the complexity of heart rate dynamics. As shown in Figure 5, when autonomic nervous system influences on the heart are eliminated through the administration of the muscarinic receptor blocker atropine and the beta-receptor blocker propranolol, the complex dynamics observed under control conditions are lost.

In another study, baby pigs were shown to develop increasing heart rate complexity as they matured from 8 to 33 days after birth (17) (Figure 6). During this period of time, the heart becomes innervated by sympathetic nerves from the right stellate ganglion. When the right stellate ganglion is denervated at birth, heart rate complexity does not develop. Thus, during healthy development, complexity appears to emerge in physiologic systems such as heart rate, and with senescence system complexity is lost.

Table 3. Decreased complexity in disease

Disease (ref)	Measure
Alzheimer's disease (25)	EEG correlation dimension is reduced
Emphysema (26)	Reduced 1/*f* slope of size distribution of terminal airspace clusters
Vent. fibrillation (27)	Reduced HR fractal scaling (1/*f* slope*) and long-range correlations (DFA**)
Atrial fibrillation (28)	Increased RR regularity ($ApEn^{+}$) and decreased long-range correlations (DFA)
CHF (10) and CAD (29)	Reduced HR fractal scaling (1/*f*) and long-range correlations (DFA)
Aging and Huntington's disease (15)	Loss of stride interval long-range correlations (DFA)
Breast cancer (30)	Fractal dimension of mammographic mass

*1/*f* = the slope of the log-transformed power spectrum (see text). **DFA = detrended fluctuation analysis, a technique that quantifies the fractal-like correlation properties of time-series data (see text). $^{+}$ApEn = Approximate entropy, a measure of regularity in time series data (11). ApEn quantifies the (logarithmic) likelihood that a series of data points that are a certain distance apart for a given number of observations remain within the same distance on the next incremental comparisons. Reprinted with permission from (2).

Table 4. Long-term consequences of complexity loss

Study authors/ref.	Subjects	Complex system	Complex measure	Adverse outcome
Ho et al. Circ 1997 (31)	52 CHF pts.+ 52 matched controls	RR interval (2° ECG)	DFA	Mortality = 1.9 yr
Huikuri et al. Circ 1998 (32)	347 random elders >65 yr	Heart rate (24° ECG)	1/*f* slope	Mortality = 10 yr
Makikallio et al. AJC 1999 (27)	Case-control, post-MI +/– VF	RR interval (24° ECG)	DFA 1/*f* slope	Ventricular fibrillation
Huikuri et al. Circ 2000 (29)	446 MI pts w/ decreased LV fxn	RR interval (24° ECG)	DFA	Mortality = 1.9 yr
Makikallio et al. AJC 2001 (33)	499 CHF pts. w/ EF< 35%	RR interval (24° ECG)	DFA	Mortality = 1.8 yr
Colantonio et al. J Geron 1993 (34)	87 elderly stroke survivors	Psycho-social fxn.	Social network index	Fx'l loss, NH adm.
Fratiglioni et al. Lancet 2000 (35)	Popn. of 1203 elders >75 yrs.	Psycho-social fxn.	Social network	Dementia

Reprinted with permission from (2).

5. LOSS OF COMPLEXITY AS A PATHWAY TO FRAILTY IN OLD AGE

These considerations have led to the theory that loss of complexity is the physiologic basis of frailty in old age (2). This is illustrated in Figure 7. During youth, a multitude of physiologic systems interact to produce a highly complex output signal (such as heart rate), which is associated with a high level of physical function. With progressive aging, many of the physiologic inputs and their

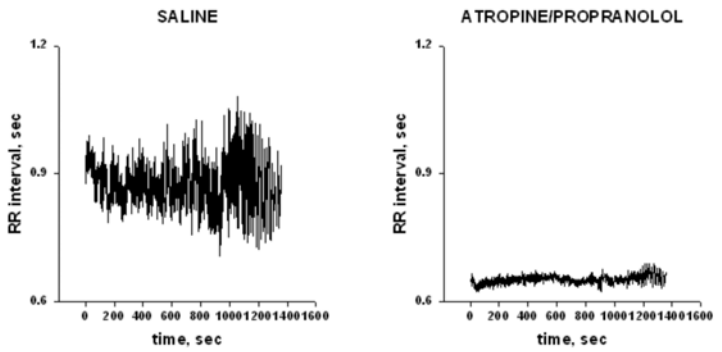


Figure 5. The effect of autonomic blockade with atropine and propranolol on RR interval dynamics in a healthy human subject.

connectivity are lost, resulting in a less complex output signal. This loss of complexity is associated with a decline in functional ability. Finally, late in life, physiologic inputs are diminished to the extent that the output becomes periodic or random, and the individual crosses a frailty threshold. At this point the ability to adapt to stress is lost and relatively minor perturbations such as new medications, a viral illness, or emotional trauma may result in serious disability and death.

6. INTERVENTIONS TO RESTORE COMPLEXITY IN PHYSIOLOGIC SYSTEMS

Fortunately, a number of novel interventions may be able to restore healthy dynamics in elderly individuals and enhance their ability to adapt to a variety of external stimuli. Some single interventions that have multiple system effects have already shown to improve functional ability in older individuals. These include exercise (18), which can improve physical and mental function, and medications such as beta-blockers, which reverse many of the physiologic consequences of congestive heart failure. Other multifactorial interventions that address the multiple systems that are impaired in individuals with syndromes such as falls (19) or delirium (20) have also proven useful in clinical trials. Recently, Dr. Collins and colleagues demonstrated that low levels of mechanical or electrical noise can be used to enhance somatosensation and thus improve postural control in healthy elderly subjects and patients with diabetic neuropathy or stroke (21). New dynamic drug delivery systems provide more physiologic drug

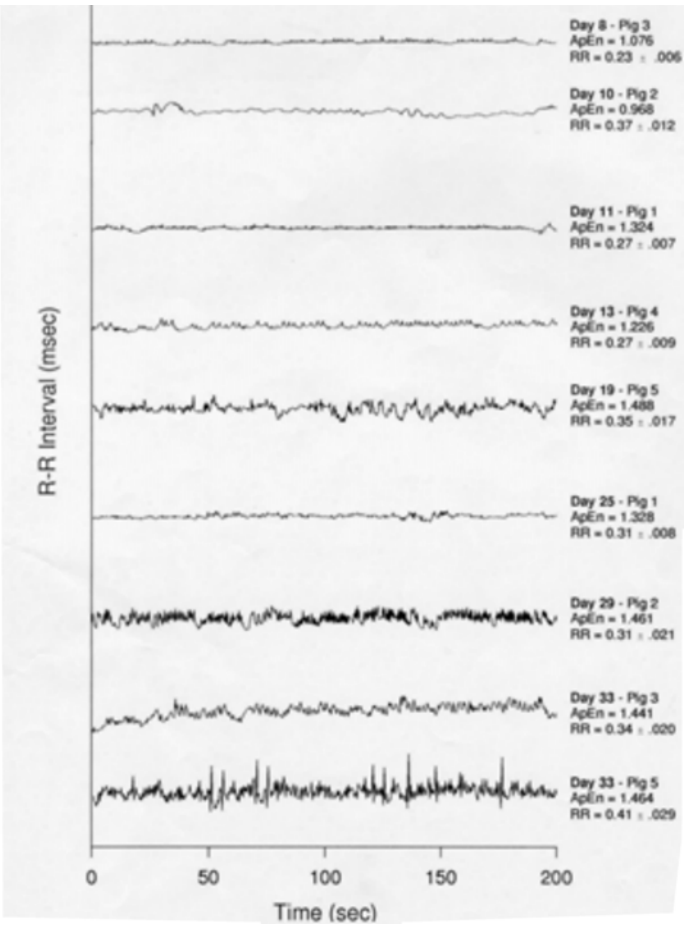


Figure 6. Interbeat interval (RR interval) time series over 200 seconds in 5 piglets as they mature from 8 to 33 days of life. Note the increasing amplitude (standard deviation) and irregularity (approximate entropy, ApEn) of the interbeat intervals with advancing postnatal age. Reprinted with permission from (17).

administration and thereby enhance physiologic effects. For example, insulin given in oscillatory fashion has a greater effect on glucose metabolism than when given continuously (22). Similarly, when parathyroid hormone is give intermittently rather than continuously, it increases bone mass in osteoporotic patients (23). Finally, external dynamic control techniques such as pacing procedures to terminate cardiac arrhythmias have proven useful in cardiac patients (24).

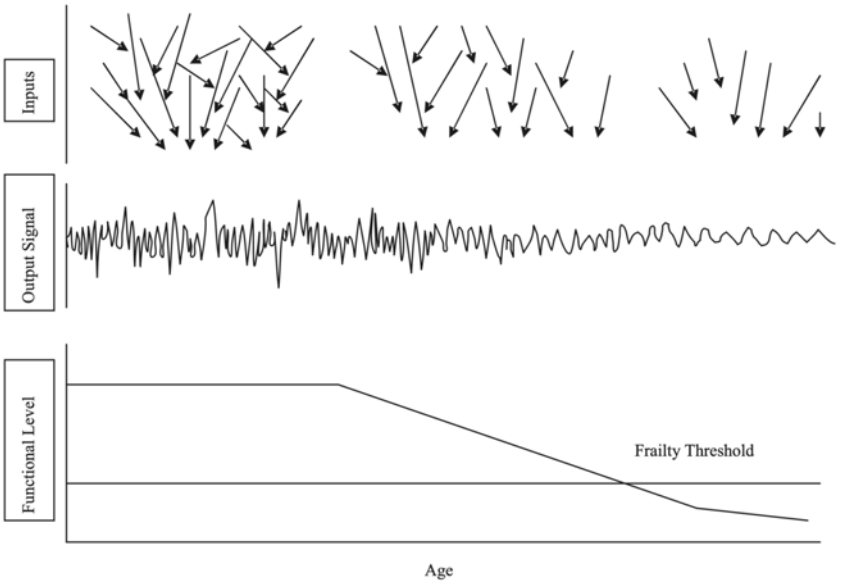


Figure 7. The physiologic basis of frailty. Multiple interacting physiologic inputs (top) produce highly irregular, complex dynamics (middle), which impart a high level of functionality (bottom) on an organism. As the inputs and their connections degrade, the output signal becomes more regular and less complex, resulting in functional decline. Ultimately, with continued loss of physiologic complexity, function may fall to the critical level below which an organism can no longer adapt to stress (the frailty threshold). Reprinted with permission from (2).

7. CONCLUSION

In conclusion, aging results in a number of changes in anatomic structures and physiologic control processes that result in a reduction in system complexity and a loss of ability to adapt to common stresses in the external and internal environment. This loss of complexity may be the physiologic basis of frailty. New interventions aimed at restoring complex dynamics may be able to enhance physiologic adaptation and prevent the onset of disease and disability.

8. ACKNOWLEDGMENTS

This work was supported by Grants AG04390 and AG08812 from the National Institute on Aging, Bethesda, Maryland. Dr. Lipsitz holds the Irving and

Edyth S. Usen and Family Chair in Geriatric Medicine at the Hebrew Rehabilitation Center for Aged.

9. REFERENCES

1. Shock NW, Greulich RC, Andres R, Arenberg D, Costa PT, Lakatta EG, Tobin JD, eds. 1984. *Normal human aging: the Baltimore longitudinal study of aging*. US Department of Health and Human Services, Baltimore.
2. Lipsitz LA. 2002. The dynamics of stability: the physiologic basis of functional health and frailty. *J Gerontol Biol Sci* **57A**:B115–B125.
3. Guyton AC. 1991. *Textbook of medical physiology*. Saunders, Philadelphia.
4. Yates FE. 1993. Self-organizing systems. In *The logic of life: the challenge of integrative physiology*, pp. 189–218. Ed. CAR Boyd, R Noble. Oxford UP, New York.
5. Lipsitz LA, Goldberger AL. 1992. Loss of "complexity" and aging: potential applications of fractals and chaos theory to senescence. *JAMA* **267**:1806–1809.
6. Mandelbrot BB. 1983. *the fractal geometry of nature*. Freeman, New York.
7. McNamee JE. 1991. Fractal perspectives in pulmonary physiology. *J Appl Physiol* **71**:1–8.
8. Goldberger AL, Rigney DR, West BJ. 1990. Chaos and fractals in human physiology. *Sci Am* **262**:43–49.
9. Goldberger AL. 1996. Non-linear dynamics for clinicians: chaos therory, fractals, and complexity at the bedside. *Lancet* **347**:1312–1314.
10. Peng C-K, Havlin S, Stanley HE, Goldberger AL. q1995. uantification of scaling exponents and crossover phenomena in nonstationary heart-beat time series. *Chaos* **5**:82–87.
11. Pincus SM, Goldberger AL. 1994. Physiological time-series analysis: what does regularity quantify? *Am J Physiol* **266**:H1643–H1656.
12. Costa M, Goldberger AL, Peng C-K. 2002. Multiscale entropy analysis of complex physiologic time series. *Phys Rev Lett* **89**:068102.
13. Kaplan DT Furman, MI Pincus, SM, Ryan SM, Lipsitz LA. 1991. Aging and complexity of cardiovascular dynamics. *Biophys J* **59**:945–949.
14. Peng C-K, Mietus JE, Liu Y, Lee C, Hausdorff JM, Stanley HE, Goldberger AL, Lipsitz LA. 2002. Quantifying fractal dynamics of human respiration: age and gender effects. *Ann Biomed Eng* **30**:683–692.
15. Hausdorff JM, Mitchell SL, Firtion R, Peng CK, Cudkowicz ME, Wei JY, Goldberger AL. 1997. Altered fractal dynamics of gait: reduced stride-interval correlations with aging and Huntington's disease. *J Appl Physiol* **82**:262–269.
16. Lauk M, Chow CC, Pavlik AE, Collins JJ. 1998. Human balance out equilibrium: nonequilibrium statistical mechanics in posture control. *Phys Rev Lett* **80**:413–416.
17. Lipsitz LA, Morin RJ, Gagnon M, Kiely DK, Medina A. 1997. Vasomotor instability preceding tilt-induced syncope: does respiration play a role? *J Appl Physiol* **83**:383–390.
18. Schuit AJ, van Amelsvoort LG, Verheij TC, Rijneke RD, Maan AC, Swenne CA, Schouten EG. 1999. Exercise training and heart rate variability in older people. *Med Sci Sport Exer* **31**:816–821.
19. Tinetti ME, Baker DI, McAvay G, Claus EB, Garrett P, Gottschalk M, Koch ML, Trainor K, Horwitz RI. 1994. A multifactorial intervention to reduce the risk of falling among elderly people living in the community. *New Engl J Med* **331**:821–827.
20. Inouye SK, Bogardus Jr ST, Charpentier PA, Leo-Summers L, Acampora D, Holford TR, Cooney Jr LM. 1999. A multicomponent intervention to prevent delirium in hospitalized older patients. *New Engl J Med* **340**:669–676.

21. Liu W, Lipsitz LA, Montero-Odasso M, Bean J, Kerrigan DC, Collins JJ. 2001. Noise-enhanced vibrotactile sensitivity in older adults, patients with **stroke**:and patients with diabetic neuropathy. *Arch Phys Med Rehab* **83**, 171–176.
22. Sturis J, Scheen AJ, Leproult R, Plonsky KS, Van Cauter E. 1995. 24-hour glucose profiles during continuous or oscillatory insulin infusion: demonstration of the functional significance of ultradian insulin oscillations. *J Clin Invest* **95**:1464–1471.
23. Neer RM, Arnaud CD, Zanchetta JR, Prince R, Gaich GA, Reginster JY, Hodsman AB, Eriksen EF, Ish-Shalom S, Genant HK, Wang O, Mitlak BH. 2001. Effect of parathyroid hormone (1–34) on fractures and bone mineral density in postmenopausal women in osteoporosis. *New Engl J Med* **344**:1434–1441.
24. Christini DJ, Stein KM, Markowitz SM, Mittal S, Slotwiner DJ. 2001. Nonlinear-dynamical arrhythmia control in humans. *Proc Natl Acad Sci USA* **98**:5827–5832.
25. Jelles B, van Birgelen JH, Slaets JP, Hekster RE, Jonkman EJ, Stam CJ. 1999. Decrease of non-linear structure in the EEG of Alzheimer patients compared to healthy controls. *Clin Neurophysiol* **110**:1159–1167.
26. Mishima M, Hirai T, Itoh H, Nakano Y, Sakai H, Muro S, Nishimura K, Oku Y, Chin K, Ohi M, Nakamura T, Bates JH, Alencar AM, Suki B. 1999. Complexity of terminal airspace geometry assessed by lung computed tomography in normal subjects and patients with chronic obstructive pulmonary disease. *Proc Natl Acad Sci USA* **96**:8829–8834.
27. Makikallio TH, Koistinen J, Jordaens L, Tulppo MP, Wood N, Golosarsky B, Peng CK, Goldberger AL, Huikuri HV. 1999. Heart rate dynamics before spontaneous onset of ventricular fibrillation in patients with healed myocardial infarcts. *Am J Cardiol* **83**:880–884.
28. Vikman S, Makikallio TH, Yli-Mayry S, Pikkujamsa S, Koivisto AM, Reinikainen P, Airaksinen KE, Huikuri HV. 1999. Altered complexity and correlation properties of R–R interval dynamic before the spontaneous onset of paroxysmal atrial fibrillation. *Circulation* **100**:2079–2084.
29. Huikuri HV, Makikallio TH, Peng CK, Goldberger AL, Hintze U, Moller M. 2000. Fractal correlation properties of R–R interval dynamics and mortality in patients with depressed left ventricular function after an acute myocardial infarction. *Circulation* **101**:47–53.
30. Velanovich V. 1996. Fractal analysis of mammographic lesions: a feasibility study quantifying the difference between benign and malignant masses. *Am J Med Sci* **311**:211–214.
31. Ho KKL, Moody GB, Peng CK, Mietus JE, Larson MG, Levy D, Goldberger AL. 1997. Predicting survival in heart failure case and control subjects by use of fully automated methods for deriving nonlinear and conventional indices of heart rate dynamics. *Circulation* **96**:842–848.
32. Huikuri HV, Makikallio TH, Airaksinen KE, Seppanen T, Puukka P, Raiha IJ, Sourander LB. 1998. Power-law relationship of heart rate variability as a predictor of mortality in the elderly. *Circulation* **97**:2031–2036.
33. Makikallio TH, Huikuri HV, Hintze U, Videbaek J, Mitrani RD, Castellanos A, Myerburg RJ, Moller M. 2001. Fractal analysis and time- and frequency-domain measures of heart rate variability as predictors of mortality in patients with heart failure. *Am J Cardiol* **87**:178–182.
34. Colantonio A, Kasl SV, Ostfeld AM, Berkman LF. 1993. Psychosocial predictors of stroke outcomes in an elderly population. *J Gerontol* **49**:S261–S268.
35. Fratiglioni L, Wang H-X, Ericsson K, Maytan M, Winblad B. 2000. Influence of social network on occurrence of dementia: a community-based longitudinal study. *Lancet* **355**:1315–1319.

Part IV

ENABLING TECHNOLOGIES

1

BIOMEDICAL MICROFLUIDICS AND ELECTROKINETICS

Steve Wereley

Associate Professor, School of Mechanical Engineering, Purdue University, West Lafayette, Indiana

Carl Meinhart

Associate Professor, Mechanical and Environmental Engineering Department, University of California, Santa Barbara

Flow phenomena are of great importance in the study of biological systems: both natural organisms and biomedical devices. Although scientists and engineers have an excellent understanding of transport processes at large length scales, the study of transport processes at cellular length scales and smaller is just beginning. Considering the importance of biological activities at and below the cellular level (see also this volume, Part III, chapter 2.1, by Huang, Sultan, and Ingber), it is critical to understand the microfluidic environment in which these processes occur and how we can manipulate them. Recent strides in micrometer- and nanometer-scale diagnostic techniques have allowed exploration of flow phenomena at length scales comparable to single cells, and even smaller. One of the most useful means of manipulating fluids and suspended species such as cells, DNA, and viruses is with electric fields. Electrokinetic phenomena are important at micron length scales, and can be used to manipulate fluid and particle motion in microfluidic devices. This chapter will briefly review the various methods of electrokinetic fluid and particle manipulation, then review the recently developed microfluidic diagnostic processes available for assessing flow behavior at micron length scales, and finally discuss in detail advances in electrothermal and dielectrophoretic fluid and particle manipulation.

Address correspondence to: Steve Wereley, School of Mechanical Engineering, 585 Purdue Mall, Purdue University, West Lafayette, IN 47907 (wereley@purdue.edu).

1. INTRODUCTION

Flow phenomena are of great importance in the study of biological systems, both natural organisms as well as biomedical devices. Most major life processes occur in an aqueous environment. Scientists and engineers already have an excellent understanding of fluid mechanics at length scales of millimeters and larger. Considering the importance of cellular activities, such as protein production and DNA reproduction, it becomes critical to understand the micrometer- and nanometer-scaled fluidic (i.e., microfluidic) environment in which these processes occur and how we can manipulate that environment. Without such understanding, we have only an incomplete picture of how the fluid transport processes occur in biological and biomedical systems and cannot assess how behavior changes at a cellular level will affect the behavior of the organism as a whole. A full understanding of the complex systems science comprising biological systems is not possible without understanding the transport processes at the smallest of length scales. Recent strides in micrometer- and nanometer-scale diagnostic techniques have allowed exploration of flow phenomena at length scales comparable to single cells, and even smaller. New fabrication tools have enabled therapeutic and analytical biomedical devices to be constructed that interact with biological components on their intrinsic length scale. One of the most useful means of manipulating fluids and suspended species such as cells, DNA, viruses, etc., is with electric fields. Electrokinetic phenomena are important at micron length scales, and can be used to manipulate fluid and particle motion in microfluidic devices. Electrokinetics can be broadly classified into DC and AC electrokinetics, as shown in Table 1. DC electrokinetic phenomena include electrophoresis and electroosmosis. Electrophoresis has been widely used in capillary gel electrophoresis for fractionation of DNA, and capillary zone electrophoresis for separation of chemical species (21). Nanogen Inc. (San Diego, CA) uses DC electrophoresis from individually addressable electrodes to control the motion of DNA molecules—first concentrating and separating target particles from the sample, then combining with target oligonucleotides at a specific location in an array of spot electrodes (4).

Electroosmotic flow is generated when microchannels with glass walls filled with aqueous solutions naturally produce electric double layers (15). In the presence of an external electric field, the electrical charge in the double layers exhibits a Coulomb force, causing the ions to migrate parallel to the channel wall. The movement of the ions induces fluid motion in the channel, creating electroosmotic flow. Electroosmosis is widely used for sample injection and transport in microchannels in commercial systems manufactured by such companies as Aclara and Caliper (1,2).

AC electrokinetics, in contrast, has received limited attention in the microfluidics community. AC electrokinetics refers to induced particle and/or fluid motion resulting from externally applied AC electric fields. One primary advantage of AC electrokinetics is that the zero mean alternating fields significantly

Table 1. Classification of AC and DC electrokinetic phenomena

Type of force	AC electrokinetics	DC electrokinetics
Body force on fluid	Electrothermal	
Surface force on fluid	AC electroosmosis	Electroosmosis
Force on suspended particles	Dielectrophoresis	Electrophoresis

reduce electrolysis. Further, the characteristic voltages necessary to accomplish useful work are typically of the order of tens of volts—much smaller than what is required for DC electrokinetics. AC electrokinetics can be classified into three broad areas: dielectrophoresis (DEP), electrothermal forces, and AC electroosmosis (17).

2. DC ELECTROKINETICS

2.1. Electroosmosis

Electroosmosis is a good place to begin discussing electrokinetic effects because the geometries involved can be idealized to considering a liquid in contact with a planar wall. When a polar liquid, such as water, and a solid surface are brought into contact, the surface acquires an electric charge. The surface charge attracts oppositely charged ionic species in the liquid that are strongly drawn toward the surface, forming a very thin tightly bound layer of ions, called the Stern layer, in which the ions in the liquid are paired one for one with the charges on the surface. Thermal energy prevents the ions from completely neutralizing the surface charge. The surface charge not neutralized by the Stern layer then influences the charge distribution deeper in the fluid, creating a thicker layer of excess charges of the same sign as those in the Stern layer, called the diffuse or Gouy-Chapman layer. Together these two layers are called the electric double layer, or EDL. Because of the proximity of charges, the Stern layer is fixed in place while the diffuse layer can be moved. In particular, the diffuse layer has a net charge and can be moved with an electric field. Consequently, the boundary between the Stern layer and the diffuse layer is called the shear surface because of the relative motion across it. The potential at the wall is called the wall potential ϕ_w, and the potential at the shear plane is called the zeta potential ζ. This situation is shown in Figure 1 and is typical of the charge distributions observed in many microfluidic devices. Both glass- (10) and polymer-based (18) microfluidic devices tend to have negatively charged or deprotonated

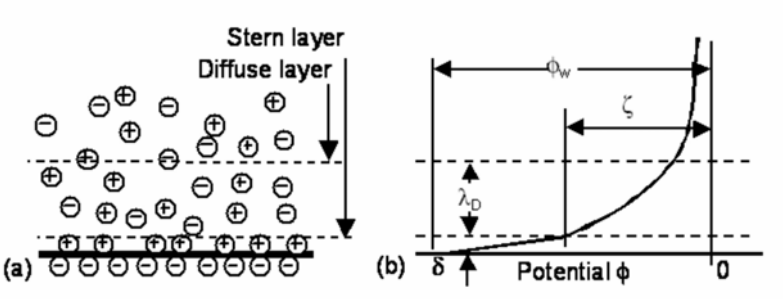


Figure 1. Sketch of the electric double layer showing (**a**) the Stern layer and the diffuse layer and (**b**) the resulting potential.

surface chemistries, which means that the EDL is positively charged. The governing equation for the electric potential ϕ is found to be the Poisson-Boltzmann equation:

$$\frac{d^2\phi}{dy^2} = \frac{2Fzc_\infty}{\varepsilon}\sinh\frac{zF\phi}{KT}, \qquad [1]$$

where c_∞ is the concentration of ions far from the surface, z is the charge number (valence) of each ion, $\varepsilon = \varepsilon_r\varepsilon_0$ is the dielectric constant of the liquid, ϕ is the electric potential, T is the absolute temperature, K is Boltzmann's constant, and F is Faraday's constant. This equation is clearly nonlinear and difficult to solve. However, the relative thickness of the EDL is usually small enough in micron-sized systems that the hyperbolic sine term can be replaced by the first term in its Taylor series—just its argument. This approximation is called the *Debye-Hückel limit* of thin EDLs and it greatly simplifies Eq. [1] to

$$\frac{d^2\phi}{dy^2} = \frac{\phi}{\lambda_D^2} \quad \text{where} \quad \lambda_D^2 = \frac{\varepsilon KT}{2z^2F^2c_\infty}, \qquad [2]$$

where λ_D is called the *Debye length* of the electrolyte. The solution to this ordinary differential equation is quite straightforward and found to be

$$\phi = \phi_w \exp\left(-\frac{y}{\lambda_D}\right). \qquad [3]$$

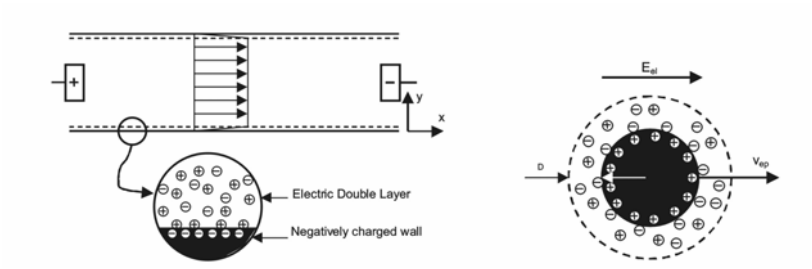


Figure 2. Schematic representation of electroosmosis (left) and electrophoresis (right)

Hence, the Debye length represents the 1/*e* decay distance for the potential as well as the electric field at low potentials.

This potential can be added into the governing equation of fluid mechanics, namely the Navier-Stokes equation, to calculate the flow produced by the electroosmotic effect. Consider the geometry shown in Figure 2 where electroosmotic flow is established in a long chamber of constant cross-section. Combining the appropriate form of the Navier-Stokes equation with the potential distribution in Eq. [3], we get

$$u_{\text{eof}} = \frac{\varepsilon E_{\text{el}} \zeta}{\eta}, \qquad [4]$$

where the component of the flow due to electroosmosis is denoted u_{eof}, the dynamic viscosity of the liquid is η, and ζ is the zeta potential, or the potential at the location of the shear plane just outside the Stern layer. This equation is known as the *Helmholtz-Smoluchowski equation* and is accurate when the Debye layer is thin relative to the channel dimension. Because of the typical low Reynolds number behavior of electrokinetic flows, the velocity field can be directly added to that obtained by imposing a pressure gradient on the flow to find the combined result of the two forces. Obtaining solutions for the flow when the Debye length is large generally requires resorting to numerical solutions because the Debye-Hückel approximation is not valid when the Debye layer is an appreciable fraction of the channel size.

For the purpose of comparing the effectiveness of several different electroosmotic channel/solution combinations, the *electroosmotic mobility*, μ_{eo} is defined as

$$\mu_{\text{eo}} = \frac{u_{\text{eof}}}{E_{\text{el}}}. \qquad [5]$$

Electroosmotic mobility is a useful empirical quantity that aids in predicting flow velocities expected for different imposed electrical fields. In the absence of appreciable Joule heating, the proportionality is very good.

2.2. Electrophoresis

This phenomenon is closely related to the electroosmosis phenomenon discussed above and relies on interaction of the EDL with an electric field to manipulate particles. The analysis of particles moving in fluids necessarily includes some drag model to account for the effect of the fluid drag on the particle. Because the electrophoretically manipulated particles tend to be small and slow moving, inertia is not important to the particle's motion and a very simple Stokes drag model is used to approximate the fluid drag on the particle. Further, the particle is assumed to be nonconducting, which is reasonable because even materials that would normally be conducting tend to become polarized by the applied field and behave as nonconductors.

There are two cases of importance in electrophoresis, when the Debye length is small compared to the radius of the particle and when it is large. The electrophoretic motion of molecules oftentimes meets the limit of Debye length large compared to the effective size of the molecule simply because molecules can be very small. In addition, with the emergence of gold and titania nanoparticles, and fullerenes, this limit becomes a very important one for nanotechnology. The expression for the electrophoretic velocity u_{ep} becomes

$$u_{ep} = \frac{q_s E_{el}}{6\pi\eta r_0} = \frac{2}{3}\frac{\varepsilon\zeta E_{el}}{\eta}, \qquad [6]$$

where the first form of the equation is well suited to molecules in which the total charge $q = q_s$ of the molecule may be known (valence number) rather than some distributed surface charge. The second form of the equation is more appropriate for very small particles for which the zeta potential ζ might be known. This form of the equation is called the Hückel equation.

The limit of small Debye length compared to particle radius is an appropriate limit to consider for particles in excess of 100 nm. Examples of these types of particles include polystyrene latex spheres used to "tag" biomolecules as well as single-cell organisms, which tend to have diameters measured in microns. When the Debye length is small compared to particle radius, the EDL dynamics are approximately reduced to the flat-plate scenario discussed in the case of electroosmosis. Hence, the equation of motion becomes

$$u_{ep} = \frac{\varepsilon\zeta E_{el}}{\eta}, \qquad [7]$$

which is simply the Helmholtz-Smoluchowski equation from the electroosmosis phenomenon. One interesting thing to note about Eqs. [6] and [7] is that, even though they are developed for opposite limiting cases, they differ only by the constant factor 2/3. When the Debye length is neither large nor small relative to the particle radius, the dynamics of the particle motion are significantly more difficult to calculate. However, even in these cases Eq. [7] is still a reasonable estimate of particle velocity.

As with the case of electroosmosis, the effectiveness of electrophoresis is quantified using an *electrophoretic mobility* parameter defined as

$$\mu_{ep} = \frac{u_{ep}}{E_{el}}, \qquad [8]$$

where μ_{ep} can be thought of as motion produced per unit field.

2.3. Applications

Because of the net negative charge associated with DNA molecules, electrophoresis can be used to manipulate DNA molecules. A classic example is capillary gel electrophoresis, where an electrical field is used to pull tagged DNA molecules through a gel matrix. The gel effectively filters the DNA molecules according to size, since the shorter DNA segments can travel through the gel much quicker than the longer segments.

Nanogen's biochip (8) is an example of using electrophoresis-enhanced hybridization of DNA in a microfluidic chip (see Figure 3). The electrodes have a positive potential, thereby inducing the DNA molecules toward specific hybridization sites. The microfluidic chip shown in Figure 3 (8), contains 100 microlocation test sites, which are approximately 80 μm in size.

3. AC ELECTROKINETICS

AC electrokinetics has received limited attention in the microfluidics community compared to its DC counterpart. AC electrokinetics refers to induced particle and/or fluid motion resulting from externally applied AC electric fields. A primary advantage of AC electrokinetics is that the alternating fields significantly reduce electrolysis at the electrodes. In addition, the characteristic voltages are typically on the order of tens of volts, which are typically much smaller than those used in DC electrokinetics. AC electrokinetics can be classified into three broad areas: dielectrophoresis (DEP), electrothermal forces, and AC electroosmosis (17).

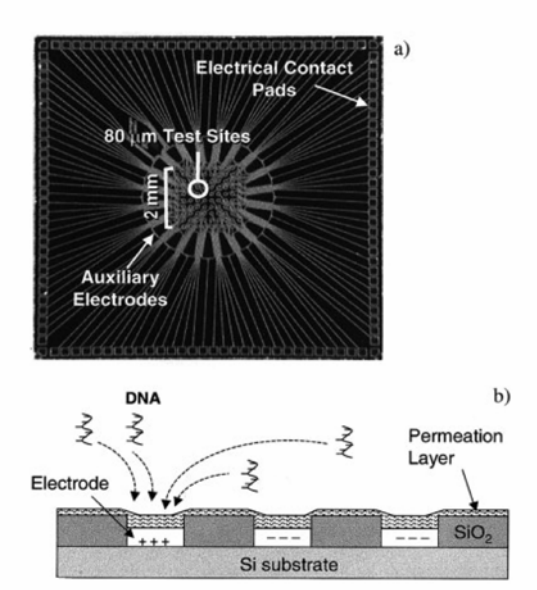


Figure 3. Active microelectronic DNA chip device and DNA transport. (**a**) Basic structure of an active microelectronic array that contains 100 microlocation test sites. (**b**) Basic scheme for electrophoretic transport of charged molecules (DNA, RNA) on the active microelectronic array test sites. Reprinted with permission from Gurtner et al. (8).

3.1. Dielectrophoresis (DEP)

Dielectrophoresis, or DEP, is a force on particles in a nonuniform electric field arising from differences in dielectric properties between the particles and the suspending fluid. The time-averaged force on a homogeneous sphere of radius r_p can be approximated as

$$\vec{F}_{\text{DEP}} = 2\pi\varepsilon_m r_p^3 \,\text{Re}(K)\nabla E_{\text{rms}}^2 \,. \qquad [9]$$

Here Re(K) is called the dielectrophoretic mobility and is the real part of K, the Clausius-Mosotti factor,

$$K = \frac{\varepsilon_p - \varepsilon_m}{\varepsilon_p + 2\varepsilon_m} \,. \qquad [10]$$

The Clausius-Mosotti factor depends on the complex permittivity of particle and medium. Complex permittivity is

$$\varepsilon^* = \varepsilon - j\sigma / \omega \,, \qquad [11]$$

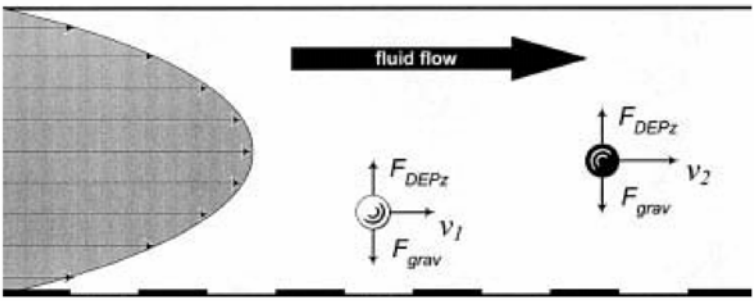


Figure 4. Schematic of field flow fractionation. DEP electrodes on the bottom microchannel surface create a nonuniform electric field. The cells are levitated from negative DEP force. The cells levitated in the center of the channel are advected faster than cells near the channel walls. This provides a mechanism for separating cells based upon their electrical properties. Reprinted with permission from Gascoyne and Vykoukal (5).

where $j = \sqrt{-1}$, ε is the electrical permittivity, σ is the electrical conductivity, and ω is the angular field frequency. In this way, the DEP force depends not only on the dielectric properties of the particle and medium, but also on the frequency of the applied field. For a sphere, the real part of K is bounded as $-0.5 < \mathrm{Re}(K) < 1.0$. Positive DEP occurs for $\mathrm{Re}(K) > 0$, where the force is toward the high electric field and the particles collect at electrode edges. The converse of this is negative DEP, which occurs when $\mathrm{Re}(K) < 0$, where the force is in the direction of decreasing field strength, and particles are repelled from electrode edges. Since the dielectrophoretic force scales with the cube of particle size, it is effective for manipulating particles of the order of one micron or larger. DEP has been used to separate blood cells and to capture DNA molecules (13,22).

DEP has limited effectiveness for manipulating proteins that are on the order of 10–100 nm (3). However, for these small particles DEP force may be both augmented and dominated by the particle's electrical double layer, particularly for low-conductivity solutions (5).

DEP has been used to manipulate macromolecules and cells in microchannels. For example, Miles et al. (13) used DEP to capture DNA molecules in a microchannel flow. Gascoyne and Vykoukal (5) present a review of DEP with emphasis on manipulation of bioparticles. An example of a cancer cell separation device is shown in Figure 4. Here, interdigitated DEP electrodes are fabricated on the surface of a microchannel. Cells are transported through the channel using pressure-driven flow. Negative DEP forces levitate the cells in the microchannel at varying heights, depending on the electrical properties of the cell. Since the velocity profile in the microchannel is parabolic, cells that are levitated in the center of the channel advect downstream faster than cells near

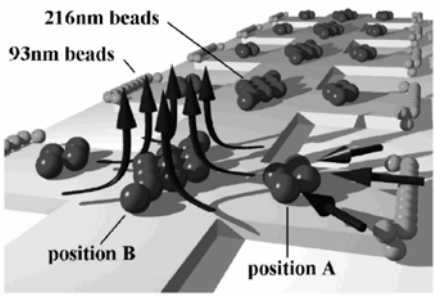


Figure 5. Size-selective movement of submicron beads based on the balance of electrothermal force and DEP. Reprinted with permission from Green and Morgan (7).

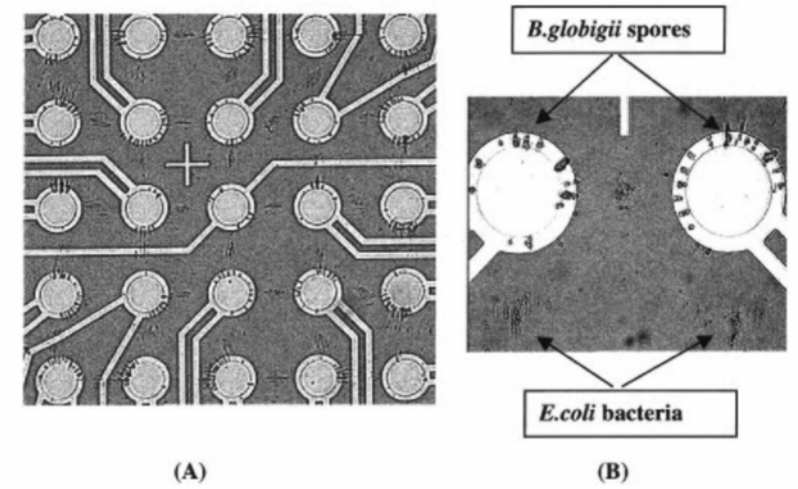


Figure 6. (**A**) Images representing the microscale separation of *B. globigii* spores and heat-killed *E. coli* bacteria on the 5 × 5 array. The electrodes in the array were addressed with an AC voltage at 50 kHz and 5 V (p-p). The spores and bacteria were suspended in a 280-mM mannitol solution having a conductivity of 20 μS/cm. (**B**) Expanded view showing that the spores were collected on the electrodes and the bacteria were repelled from the electrodes. Reprinted with permission from Huang et al. (9).

the microchannel surface. Therefore, cancerous and noncancerous cells can be separated and identified based on their electrical properties. This separation technique is known as flow field fractionation (FFF). A schematic of this separation technique is shown in Figure 4 (taken from (5)).

The combination of DEP with electrothermal flow and AC electroosmosis is discussed in detail by Green et al. (7), who demonstrated how, in the absence

of pressure-driven flow, different-sized particles can be separated based on the balance of DEP force and fluid drag force from an electrothermally generated flow. Figure 5 depicts how particles can be separated by varying sizes under the influence of DEP and electrothermal flow. By varying the frequency (up to 500 kHz) and the voltage (up to 10 V peak-to-peak), the stable position of the larger beads can be moved from the electrode edges to position A, or to position B.

DEP has also been demonstrated by Huang (9), (Nanogen Inc., San Diego) to concentrate a dilute sample of *E. coli* cells by 20-fold and to separate *E. coli* cells from *B. globigii* cells. A picture of the microfabricated electrode structure and captured bacteria is shown in Figure 6.

3.2. Electrothermally Driven Flow

Electrothermal body forces are created by nonuniform Joule heating of the medium. The Joule heating is a source term in the temperature equation, and creates spatial variations in conductivity and permittivity, which in turn create Coulomb and dielectric body forces in the presence of an externally applied electric field. The resulting fluid motion can be determined by solving the Navier-Stokes equation with the electrothermal body force. Electrothermally driven flow can be simulated by solving for the quasistatic electric field in a specific geometry. The nonuniform electric field gives rise to nonuniform temperature fields through Joule heating. Ignoring unsteady effects and convection, and balancing thermal diffusion with Joule heating yields

$$k\nabla^2 T + \sigma E^2 = 0, \qquad [12]$$

where T is temperature and E^2 is the magnitude squared of the electric field, given by $\vec{E} = -\nabla V$, where k and σ are the thermal and electrical conductivity.

Gradients in temperature produce gradients in permittivity and conductivity in the fluid. For water $(1/\sigma)(\partial\sigma/\partial T) = +2\%$ and $(1/\varepsilon)(\partial\varepsilon/\partial T) = -0.4\%$ per degree Kelvin. These variations in electric properties produce gradients in charge density and perturb the electric field. Assuming the perturbed electric field is much smaller than the applied electric field, and that advection of electric charge is small compared to conduction, the time-averaged electrothermal force per unit volume for a non-dispersive fluid can be written as (17)

$$\vec{F} = -0.5\left[\left(\frac{\nabla\sigma}{\sigma} - \frac{\nabla\varepsilon}{\varepsilon}\right)\vec{E}_{\text{rms}}\frac{\varepsilon\vec{E}_{\text{rms}}}{1+(\omega\tau)^2} + 0.5\left|\vec{E}_{\text{rms}}\right|^2\nabla\varepsilon\right], \qquad [13]$$

where $\tau = \varepsilon/\sigma$ is the charge relaxation time of the fluid medium and the incremental temperature-dependent changes are

$$\nabla\varepsilon = \left(\frac{\partial\varepsilon}{\partial T}\right)\nabla T, \quad \nabla\sigma = \left(\frac{\partial\sigma}{\partial T}\right)\nabla T\,. \qquad [14]$$

The first term on the right-hand side of Eq. [13] is the Coulombic force, and is dominant at low frequencies. The second term is the dielectric force, and is dominant at high frequencies. The crossover frequency scales inversely with the charge relaxation time of the fluid, and typically occurs at around several MHz.

The electrothermal force shown in Eq. [13] is a body force on the fluid. The motion of the fluid can be determined by solving the Stokes equation for zero Reynolds number fluid flow, such that

$$0 = -\nabla P + \mu\nabla^2\vec{u} + \vec{F}_{\mathrm{ET}}\,, \qquad [15]$$

where $\vec{u}$ is the fluid velocity, P is the pressure in the fluid, and μ is the dynamic viscosity of the fluid.

3.3. AC Electroosmosis

AC electroosmosis arises when the tangential component of the electric field interacts with a double layer along a surface. It becomes less important with increasing electric field frequency. For example, in an aqueous saline solution with an electrical conductivity of $\sigma = 2 \times 10^{-3}$ S/m, it is predicted that AC electroosmosis will not be important above 100 kHz (16).

3.4. Numerical Simulations of Electrothermal Flow

AC electrokinetics can be used to manipulate fluid motion and to enhance the sensitivity of certain biosensors (20). The finite-element package CFD-ACE$^+$ (CFD Research, Huntsville, AL) was used to simulate electrothermally induced flow and subsequent enhanced binding in the cavity. First, the quasistatic potential field for two long electrodes along the cavity wall is calculated (Figure 7a). The Joule heating of the fluid from this electric field produces local changes in temperature. Figure 7b shows the temperature field resulting from Joule heating. From this temperature field, the electrothermal force, $\vec{F}_{\mathrm{ET}}$, can be estimated from Eqs. [13] and [14]. The fluid motion can be calculated using the Stokes equation, Eq. [15]. Figure 7c shows the resulting velocity field. The velocity of the ETF is on the order of 500 μm/s and characterized by a pair of counter-rotating vortices. This fluid motion will effectively stir the analyte, moving it toward the immobilized antibodies.

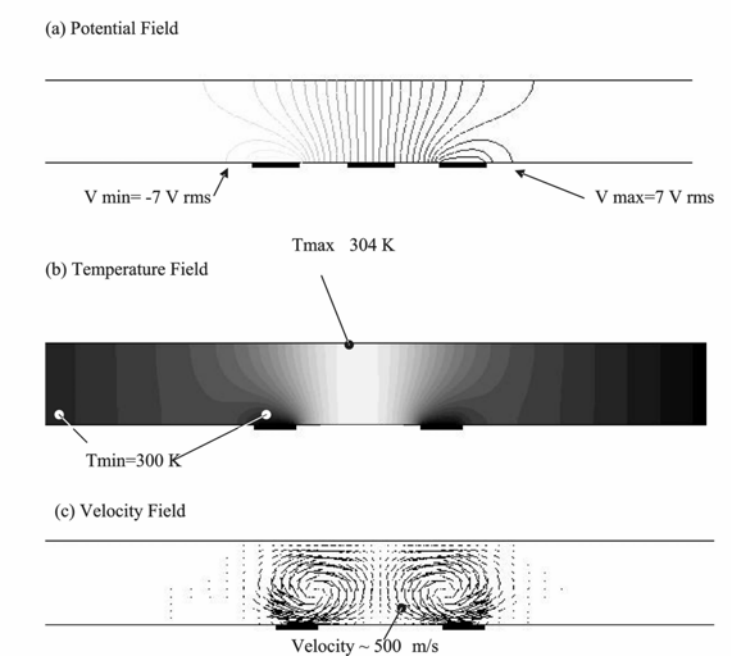


Figure 7. Simulation of ETF in a 2000 × 40 μm cavity: (**a**) Quasistatic electric potential field, calculated from two electrodes with potentials of +/– 7 V_{rms} (1O-V peak-to-peak). (**b**) Temperature field resulting from a balance of Joule heating and thermal diffusion. The fluid has an increase in temperature between the electrodes; electrodes conduct heat to the environment. (**c**) Velocity vectors from 2D simulation of electrothermally generated fluid motion.

The convective scalar equation can be used to calculate the effect of electrothermally induced fluid motion on the analyte concentration in the cavity and the analyte binding on a cavity wall:

$$\frac{\partial C}{\partial t} + \vec{u} \cdot \nabla C = D\nabla^2 C \,, \qquad [16]$$

where C is the concentration of antigen in the outer flow, $\vec{u}$ is the fluid velocity, D is the diffusivity of the antigen, and t is the time. Following the model given by Myszka et al. (14), the rate of association is $k_a C(R_T - B)$, where k_a is the association constant, C is the concentration of antigen at the surface, and $R_T - B$ is the available antibody concentration. The rate of dissociation is $k_d B$, where k_d is the dissociation constant and B is the concentration of bound antigen. The time

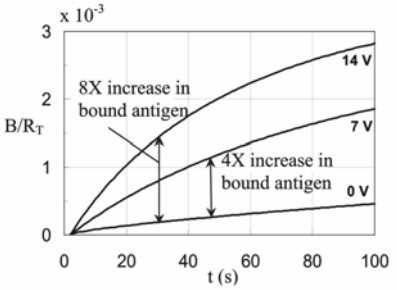


Figure 8. Numerical simulation of dimensionless binding curves for non-enhanced (0 V) and enhanced (7 V, 14 V) transport. The differences in the two curves show an increase in binding rate, which yields a factor of 4 higher binding for 7 V and a factor of 8 higher binding after 30 seconds for 14-V applied root-mean square potential. The binding improvement for the 14-V case decreases to around sixfold after 100 seconds: the binding is no longer completely transport-limited.

rate of change of antigen bound to the immobilized antibodies is equal to the rate of association minus the rate of dissociation:

$$\frac{\partial B}{\partial} = k_a C(R_T - B) - k_d B \,. \qquad [17]$$

The rate of antigen binding to immobilized antigen, $\partial B/\partial t$, must be balanced by the diffusive flux of antigen at the binding surface, $y = 0$, such that

$$\frac{\partial B}{\partial t} = D \left. \frac{\partial C}{\partial y} \right|_{y=0} . \qquad [18]$$

Equations [16]–[18] are solved with an initial antigen concentration C_0 = 1 nM and an immobilized antibody concentration R_T = 1.7 nM cm (i.e., one molecule per 100 nm^2). The binding rates for three conditions, 0-, 7-, and 14-V root-mean-square voltage, are shown in Figure 8. The 0-V case corresponds to the passive case, which is the result of pure diffusion. This is the standard mode of most immobilized assays, such as ELISA. The 7- and 14-V curves correspond to the result of electrothermally driven flow enhancing transport of antigen to the immobilized antibodies. The curves in Figure 8 show that a factor of up to 8 (800%) improvement in sensitivity (or response) is obtained by using AC electrokinetics.

4. EXPERIMENTAL MEASUREMENTS OF ELECTROKINETICS

Two examples of using electrokinetics to manipulate small particles will be presented here. Since both examples employ the μPIV technique to quantify the response of small particles to the electrically induced forces, a brief introduction to the technique will also be given. The first example, electrokinetic flow, illustrates the electrothermal effect and the second illustrates the dielectrophoretic effect.

4.1. Microparticle Image Velocimetry (μPIV)

μPIV is a technique that has been developed recently to measure the velocity of small scale flows in a spatially resolved manner (19). Figure 9 shows the typical layout of a μPIV system. The flow is illuminated by either a broad-wavelength continuous light source, such as a mercury vapor lamp, or a pulsed laser, such as a frequency-doubled Nd:YAG. Normally, μPIV is used to measure the velocity of small-scale flows by measuring the motion of small tracer particles either naturally present in the flow or artificially added to the flow. In the following examples, however, the motion of the fluid is not the primary subject of study, but rather the motion of the suspended particles in response to an electrically applied force. Regardless of whether the fluid motion or particle motion

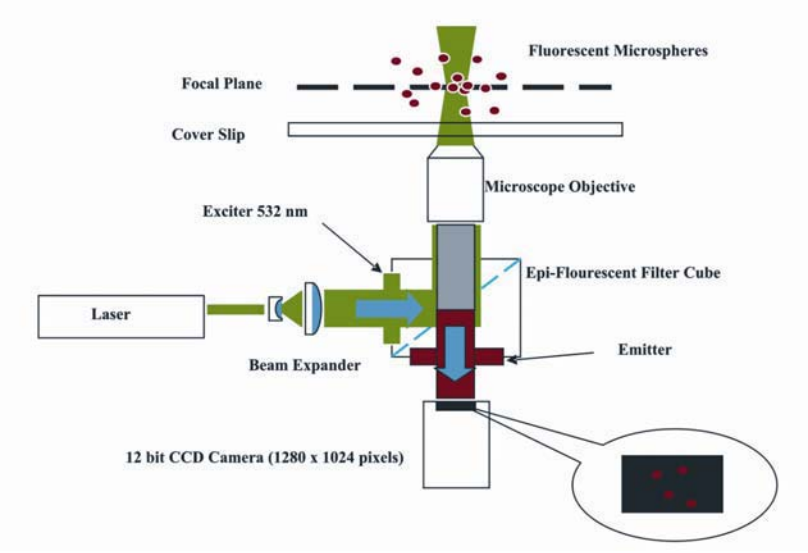


Figure 9. Diagram of typical μPIV system.

is being studied, the technique is the same. The small particles are observed with a microscope. The particles are typically coated with a fluorescing dye to enable epifluorescent imaging. Images are captured with a precise time delay from one image to the next. Consecutive pairs of images are divided into many small interrogation regions. The corresponding interrogation regions from each of the two original images are cross-correlated to determine the most likely relative displacement of particles in the interrogation regions in the form of a cross-correlation peak. Repeating this procedure thousands of times produces the spatially resolved measurements of fluid or particle motion seen in the following sections.

4.2. Electrothermal Effect

Micro-PIV experiments using polystyrene spheres in an optically accessible flow cell with wedge-shaped electrodes have been conducted. The trajectories of 1-μm diameter polystyrene particles suspended in sugar solution were measured in a device consisting of two brass electrodes sandwiched between two glass wafers. An AC potential of 10 V_{rms} at 10 kHz was applied to the electrodes. The particle-velocity field is measured quantitatively using μPIV following Meinhart et al. (11), and is shown in Figure 10a.

The experimental results compare well to numerical solutions of electrothermally driven flow: fluid motion is simulated by solving the Stokes equation, subject to an electrothermal force (Eq. [13]). The velocity of suspended 1-μm particles relative to the fluid medium can be estimated by balancing the two dominant particle forces: Stokes drag force and DEP force. The numerically simulated particle velocity field is shown in Figure 10b. For these parameters, according to model results, the DEP was negligible in comparison with motion generated through electrothermal flow. The results are described in detail by Meinhart et al. (12). The agreement between simulations and experiments may indicate that electrothermal forces are important in the microfluidic devices tested. However, in these numerical simulations, the effect of AC electroosmosis is not modeled.

4.3. Dielectrophoretic Effect

The second set of experiments were designed to isolate the effects of dielectrophoresis from electrothermal motion. A channel measuring 350 μm wide by 12 μm deep is shown in Figure 11. The bright regions in the image are platinum electrodes, while the dark regions are areas without platinum or electrode gaps. The entire imaged region is covered by a thin layer of silicon dioxide, which insulates the electrodes from the fluid medium to suppress the Joule heating that

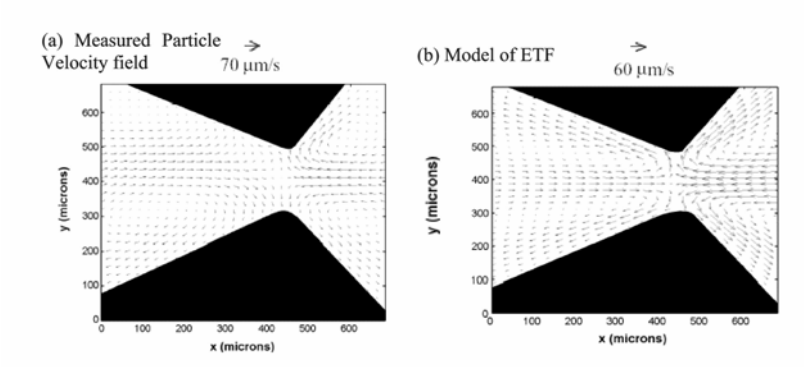


Figure 10. AC electrokinetically driven particle motion: (**a**) experimentally measured particle velocity field using micro-PIV, and (**b**) numerically simulated particle velocity arising from electrothermal fluid motion and viscous drag. This suggests that for this regime ETF is dominant compared to DEP.

typically causes electrothermal effects. The electrodes are arranged in interdigitated pairs so that in Figure 11 the first and third electrodes are always at the same potential as each other. The second and fourth electrodes are also at the same potential as each other but can be at potential different from the first and third electrodes. An alternating electric potential is applied to the interdigitated electrodes to create an electromagnetic field with steep spatial gradients. Particle motion through the resulting electric field gradients causes polarization of the microspheres, resulting in a DEP body force that repels particle motion into increasing field gradients.

Six sets of experiments were performed, each using a different voltage. All experiments used the same flow rate of 0.42 μl/hr (equivalent to a Reynolds number of 3.3×10^{-4}) and the same AC frequency of 580 kHz. The voltages were chosen between 0 and 4 V. Charge-neutral fluorescent polystyrene particles measuring 0.69 μm in diameter (Duke Scientific) were suspended in the flow. Sets of images 800 images each were acquired using a Photometrics CoolSNAP HQ interline monochrome CCD camera from Roper Scientific. This camera is capable of 65% quantum efficiency around the 610-nm wavelength, which is the emission wavelength of the red fluorescing microspheres. Images were captured at a speed of 20 frames per second. The microscope used in these experiments was an epifluorescent Nikon E600 with a Nikon "CFI W FLUOR 60X" water-immersion objective lens having a numerical aperture of 1.00. Epifluorescent imaging was used because the silicon base of the device is highly reflective, resulting in strong background noise. The broad-spectrum mercury vapor light source is bandpass filtered to admit wavelengths of approximately

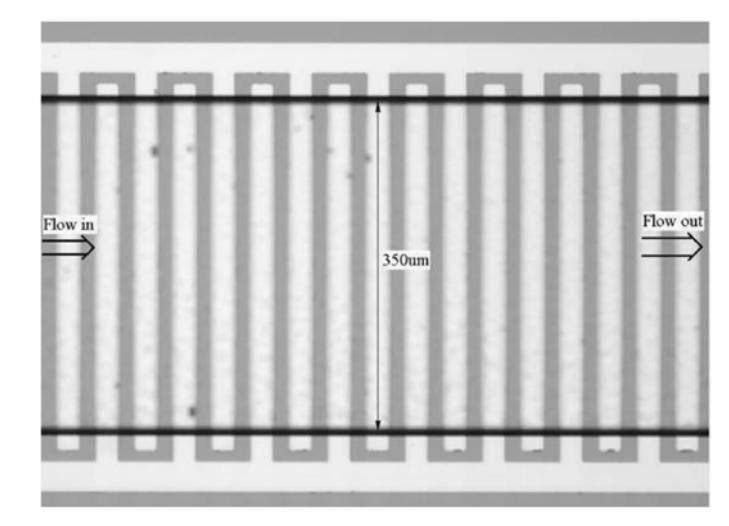


Figure 11. Dielectrophoretic device. (Courtesy of Bashir and Li, Purdue University).

540 nm. This wavelength excites the red dye on the latex microspheres, which emits light at approximately 610 nm. The epifluorescent filter cube then bandpass filters the light directed to the camera to admit wavelengths of approximately 610 nm. Thus, the 540-nm light reflected by background features that do not fluoresce is removed.

A particle velocity increase can be seen from 0.5 to 2.0 volts, and a velocity decrease is seen from 2.0 to 4.0 V. In the 2.5-V case, low-velocity regions between electrodes are first noticeable. In the 3.5- and 4.0-V cases, large numbers of trapped particles result in near-zero velocities. Figure 12 shows the experimental results with a plot of the particle velocity as a function of position within the device for each of the six voltage cases measured. The three lowest voltages share a trend of decreasing particle velocity in the downstream direction. The simplest explanation for this behavior is that, with each electrode a particle encounters, it lags the fluid velocity a little more. The cumulative effect of encountering a series of electrodes is a gradual slowing of the particles.

Another interesting result apparent from Figure 12 is that the average particle velocity initially increases as the voltage increases from 0.5 to 2.0 V. This phenomenon is explained by particles being pushed away from the channel bottom (where the electrodes are located) into the faster areas of fluid flow near the center of the channel. For the higher-voltage measurements, the effect of particles being hindered by field gradients is compounded by particles being pushed

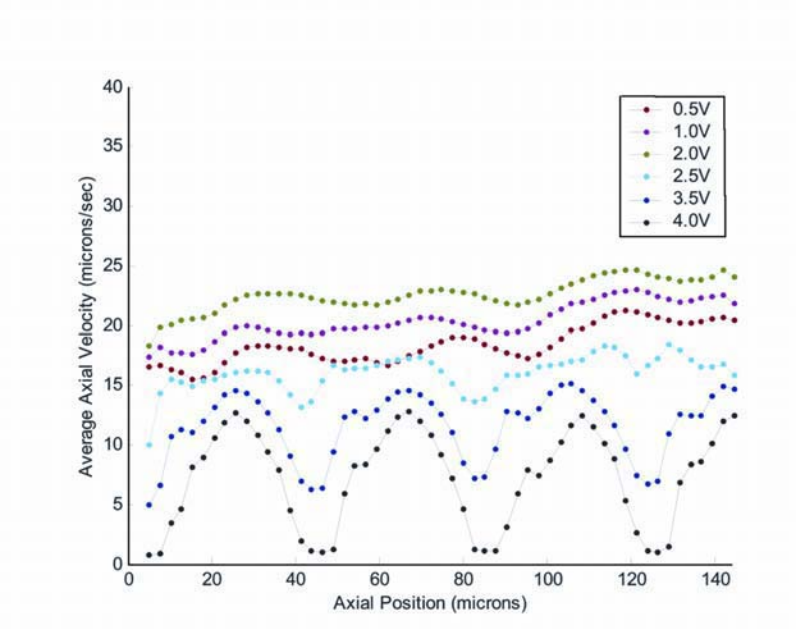


Figure 12. Average axial velocity results from PIV results for all voltage cases. Note that flow is from right to left in this figure, opposite that in Figure 11.

beyond the high-speed portion of the flow profile and toward the low-speed upper wall of the device. It can be qualitatively confirmed that particles are pushed to the upper wall of the channel by observing the particle shapes from the higher-voltage cases. Many different particle shapes are evident in Figure 13a, in which the voltage is 0.5 V. These many shapes represent how the particle images change with distance from the focal plane, which is focused on the electrodes. Particles in focus appear as small round dots while particles out of focus (near the upper device wall) appear less bright and have rings around them from the diffraction pattern. Figure 13b shows particle images for a 4-V voltage. Two phenomena are obvious here: first, the particles tend to cluster at electrode edges, and, second, almost all particle images are out of focus, indicating that nearly all particles are found near the top wall of the device, remote from the electrodes.

5. CONCLUSIONS

The results from §§4.2 and 4.3 show that the motion of suspended particles—or any discrete elements in the fluid such as cells, proteins, and DNA—in

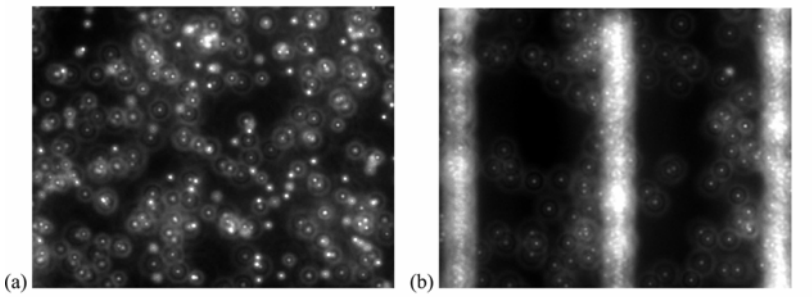


Figure 13. Raw particle images for the case of 0.5 (**a**) and 4.0 V (**b**).

response to applied electric fields can be quite complicated. It is becoming increasingly popular to manipulate such suspended particles with electrical fields. To fully understanding complexity science applied to biomedicine, one must understand how particles behave at micrometer and nanometer length scales. Experiments and modeling are often needed to determine not only the dominant phenomenon (DEP vs. ETF) but also how the particles are distributed within the flow. Particle distribution is of great importance in biological systems. Sometimes they need to be trapped near a wall for analysis and sometimes they need to be kept away from walls to reduce fouling of a device.

6. REFERENCES

1. Bousse L, Cohen C, Nikiforov T, Chow A, Kopf-Sill AR, Dubrow R, Parce JW. 2000. Electrokinetically controlled microfluidic analysis systems [review]. *Annu Rev Biophys Biomol Struct* **29**:155–181.
2. Chien R-L. 2002. Simultaneous hydrodynamic and electrokinetic flow control, *Micro Total Anal Syst* **1**:386–388.
3. Deval J, Tabeling P, Ho C-M. 2002. A dielectrophoretic chaotic mixer. *Proc IEEE MEMS Workshop*, pp. 36–39.
4. Forster AH, Krihak M, Swanson PD, Young TC, Ackley DE. 2001. A laminated, flex structure for electronic transport and hybridization of DNA. *Biosensors Bioelectron* **16**(3):187–194.
5. Gascoyne PRC, Vykoukal J. 2002. Particle separation by dielectrophoresis [review]. *Electrophoresis* **23**(13):1973–1983.
6. Green NG, Ramos A, Morgan H. 2000. AC electrokinetics: a survey of sub-micrometre particle dynamics. *J Phys D Appl Phys* **33**(6):632–641.
7. Green NG, Morgan H. 1998. Separation of submicrometre particles using a combination of dielectrophoretic and electrohydrodynamic forces. *J Phys D Appl Phys* **31**:L25–L30.
8. Gurtner C, Tu E, Jamshidi N, Haigis RW, Onofrey TJ, Edman CF, Sosnowski R, Wallace B, Heller MJ. 2002. Microelectronic array devices and techniques for electric field enhanced DNA hybridization in low-conductance buffers. *Electrophoresis* **23**:1543–1550.

9. Huang Y, Ewalt KL, Tirado M, Haigis TR, Forster A, Ackley D, Heller MJ, O'Connell JP, Krihak M. 2001. Electric manipulation of bioparticles and macromolecules on microfabricated electrodes. *Analyt Chem* **73**(7):1549–1559.
10. Hunter RJ. 1981. Zeta potential in colloid science. Academic Press, London.
11. Meinhart CD, Wereley ST, Santiago JG. 1999, PIV measurements of a microchannel flow. *Exp Fluids* **27**(5):414–419.
12. Meinhart CD, Wang D, Turner K. 2002. Measurement of ac electrokinetic flows. In *Proceedings of the 10th international symposium on flow visualization*, Kyoto, Japan, August 26–29.
13. Miles R, Belgrader P, Bettencourt K, Hamilton J, Nasarabadi S. 1999. Dielectrophoretic manipulation of particles for use in microfluidic devices, MEMS, I: microelectromechanical systems (MEMS). In *Proceedings of the ASME international mechanical engineering congress and exposition*, Nashville, TN, November 14–19, 1999.
14. Myszka DG, He X, Dembo M, Morton TA, Goldstein B. 1998. Extending the range of rate constants available from BIACORE: interpreting mass transport-influenced binding data. *Biophys J* **75**(2):583–594.
15. Probstein RF. 1994. *Physiochemical hydrodynamics, an introduction*, 2nd ed. Wiley Interscience, New York.
16. Ramos A, Castellanos A, Gonzales A, Morgan H, Green N. 2002. Manipulation of bioparticles in microelectrode structures by means of nonuniform ac electric fields. In *Proceedings of ASME International mechanical engineering congress and exposition*, New Orleans, LA, November 17–22.
17. Ramos A, Morgan H, Green NG, Castellanos A. 1998. Electrokinetics: a review of forces in microelectrode structures. *J Phys D Appl Phys* **31**:2338–2553.
18. Roberts MA, Rossier JS, Bercier P, Girault HH. 1997. UV laser machined polymer substrates for the development of microdiagnostic systems. *Anal Chem* **69**:2035–2042.
19. Santiago JG, Wereley ST, Meinhart CD, Beebee D. Adrian RJ. 1998. A particle image velocimetry system for microfluidics. *Exp Fluids* **25**(4):316–319.
20. Sigurdson M, Meinhart C, Wang D, Lui X, Feng J, Krishnamoorthy S, Makhijiani VB. 2002. Transport enhancement in tunable laser cavity sensor In *ASME IMECE'02 MEMS Symposium*, New Orleans, LA.
21. Thormann W, Lurie I, McCord B, Mareti U, Cenni B, Malik N. 2001. Advances of capillary electrophoresis in clinical and forensic analysis (1999–2000). *Electrophoresis* **22**:4216–4243.
22. Wang X-B, Vykoukal J, Becker F, Gascoyne P. 1998. Separation of polystyrene microbeads using dielectrophoretic/gravitational field-flow-fractionation. *Biophys J* **74**:2689–2701.

2

GENE SELECTION STRATEGIES IN MICROARRAY EXPRESSION DATA: APPLICATIONS TO CASE-CONTROL STUDIES

Gustavo A. Stolovitzky

IBM Computational Biology Center, Yorktown Heights, New York

Over the last decade we have witnessed the rise of the gene expression array assay as a new experimental paradigm to study the cellular state at the whole genome scale. This technology has allowed considerable progress in the identification of markers associated with human disease mechanisms, and in the molecular characterization of diseases such as cancer, by careful characterization of genes involved directly or indirectly in the disease. A typical gene expression experiment provides scientists with an enormous amount of data. Analysis of these data, and interpretation of the ensuing results, have attracted the attention of many researchers, who have developed new ways of interrogating the expression data. In this chapter we will review some of these recent efforts, emphasizing the need to make use of batteries of methods rather than one method in particular, as well as the need to properly validate results with independent data sets. The application of DNA array technology for use in disease diagnostics will be exemplified in the case of chronic lymphocytic leukemia.

1. INTRODUCTION

DNA microarrays constitute one of the most powerful high-throughput technologies in molecular biology today. It has emerged in recent years as a powerful tool that provides a glimpse into the complexities of cellular behavior

Address correspondence to: Gustavo A. Stolovitzky, IBM Computational Biology Center, 1101 Kitchawan Road, Yorktown Heights, NY 10598 (gustavo@us.ibm.com).

through the window of transcriptional activity (see also this volume, Part III, chapter 2.1, by Huang, Sultan, and Ingber). As is the case with other nascent high-throughput technologies (such as protein arrays, single nucleotide polymorphism profiling) the completion of a DNA microarray experiment requires a concerted effort between the data producers and the data analysts. In this sense, both the wet lab protocols as well as the methods used in the analysis of the ensuing data are two faces of the same coin that enable this emerging technology reach its full potential.

Gene expression array experiments present us with vast amounts of data containing substantial biological information. In order to obtain the most from these data, a considerable variety of approaches for statistical and algorithmic analyses have been developed. Each one of these analyses, if sufficiently different, can provide important biological clues. In this sense, gene expression data have become a paradise for statisticians. Evidence of this is the exponential growth in the number of publications on gene expression analysis during the last few years (1). Many useful resources for gene expression data, links, and literature have flourished on the internet (see, e.g. (2–6) and the links therein) to support the growing needs of the field.

There are many types of questions that can be explored with microarray experiments. Some of the common themes in DNA array data analysis—including gene selection, clustering of similarly expressing genes, class prediction, and pathway inference—have recently been reviewed in (1,7). In this chapter we concentrate on supervised statistically based methods to identify genes that show differences between two classes of tissues, a problem known as gene or feature selection. Among these methods, we shall discuss univariate and multivariate techniques for gene selection. In the former, genes are selected on the basis of their individual merits to separate between two or more classes of tissues (typically cases and controls). In the latter we deal with the differential behavior of groups of genes in distinct tissue types. Both types of analysis yield important information and the possibility of rich interpretations. For example, a gene that is identified by univariate methods as strongly transcribed in cancer patients compared to control subjects could be a good candidate for an oncogene, or the result of chromosomal instabilities that resulted in more copies of that gene. A multivariate methodology, instead, can reveal subtler changes of ensembles of genes that work in coordination, such as may be the case in a pathway that is deregulated in the transformed cells. Multivariate analyses are better aligned with a complex-systems perspective of the data, in that we explore interactions between genes rather than genes in isolation. After all, it is the coordinated activity of interacting genes that gives the cellular environment its complex behaviors.

Once we have selected genes that express differentially in cases and controls, there remains the need to establish the validity of the results. Two methods are typically used to provide some degree of validation for the selected genes. In

one method, which can be called *validation by classification*, the selected genes are validated if they perform adequately in the prediction of the class of tissue that a set of unknown samples belongs to. Typically, these unknown samples are part of the original data set, but are left out in the gene selection phase for later validation purposes. In a second approach, which could be called *validation by statistical significance*, genes are chosen if they behave sufficiently different from what would be expected if there were no distinction between cases and controls (the null hypothesis). We proposed an alternative validation method in (1), a sort of hybrid between the previous two methods, in which the selected genes are validated if they show consistency in their behavior in a different data set (different laboratories and may be different technology, but the same types of tissues).

The typical outcome of a gene selection process is a list of genes that show differential expression in cases and controls. If the data are analyzed using more than one method, it is likely that the resulting gene lists will be different, albeit overlapping. We will discuss how to deal with lists of genes coming from different algorithms, and the advantages or disadvantages of creating the union list or the intersection list of these genes. It is clear, however, that the explosion in the number of methods to analyze expression data should be complemented with a convergent effort in which different algorithms are used and their results combined.

The lists of genes generated by the algorithms discussed in this chapter will have to be organized, possibly with the help of literature search techniques (8) or by systematically relating the selected genes with existing biological information (9,10), in order to bring the results of the microarray technology to meaningful applications. Among these applications we can mention the identification of potential drug targets (11), the discovery of disease specific genes (12), toxicogenomics (13), disease prognosis (14), and the molecular taxonomy of diseases (15,16). It has been suggested, indeed, that microarrays will be routine practice in clinical diagnostics within the next decade or so. Making this happen will surely necessitate a larger number of samples in clinical trials and proof of the robustness of the technology (16). Towards the end of this chapter we will present an example in clinical diagnostics which shows that the technology is indeed reaching a state of maturity, both in terms of the algorithms used for gene selection and in terms of the DNA array technology.

2. PREVIOUS WORK: GENE SELECTION METHODS IN MICROARRAY DATA

To organize the presentation, we will separate the discussion of gene selection algorithms into univariate and multivariate methods. In either case, the genes selected as informative need to be validated in one way or another. Two

methods have been used in the literature to provide some degree of validation for the selected genes. In one method, which could be called *validation by classification*, the selected genes are validated if they perform adequately in the prediction of the type of tissue (case or control) for samples excluded from the training set but whose class is known. In a second approach, which could be called *validation by statistical significance*, genes are chosen if they behave sufficiently different from what would be expected if there were no class distinction between the case and control samples (the null hypothesis). We will discuss other possibilities for validation later on. Before describing the recent literature on gene selection, however, a few nomenclature conventions are necessary.

2.1. Nomenclature

Throughout this section it will be assumed that we are dealing with an assay in which M samples were hybridized to their respective M arrays, each containing N gene probes. Of the M samples, M_1 cases are of class 1 (C_1) and M_2 cases are of class 2 (C_2), where by class 1 and 2 we mean cancer and control, or cancer of type 1 and cancer of type 2, etc. The values of the expression measured for the ith gene in the kth sample of class c will be denoted by $X_{ik}^{(c)}$. In many algorithms data are preprocessed by different normalizations and transformations. In these cases we shall still denote by $X_{ik}^{(c)}$ the resulting gene expression values after the preprocessing steps. (For a review of normalization considerations see (17).) The sample mean and standard deviation of gene i in class c will be respectively denoted $\mu_{i(c)}$ and $\sigma_{i(c)}$.

2.2. Selecting Genes One at a Time: Univariate Methods

2.2.1. t-Score-Based Statistics

One of the most common univariate analyses uses the t-statistic (or t-score), which for gene i can be written as:

$$t_i = \frac{\mu_{i(1)} - \mu_{i(2)}}{\sqrt{\sigma_{i(1)}^2 / M_1 + \sigma_{i(2)}^2 / M_2}} . \qquad [1]$$

This statistic measures the difference between the sample means in cases and controls in units of the standard deviation of this difference. If the two samples are normally distributed, or if M_1 and M_2 are large, the theoretical distribution of the t-score is known. In the former case t_i would be distributed according to the t-distribution, and in the latter case the distribution of t_i

asymptotically approaches normality. In both cases, the *t*-test can be used to test the null hypothesis that the two samples have the same mean, i.e., the gene is not differentially expressed. In typical gene expression experiments, however, we should not assume that the data are normally distributed, especially if the sample sizes M_1 and M_2 are small. In these cases we can still use the *t*-score, but we need to resample its distribution (18–20). Analyses based on *t*-scores can also be validated by classification. To do this, the gene selection has to be complemented with classification schemes such as *k*-nearest neighbors, decision trees, support vector machines, and naive Bayes (21). In these cases, the classification methods take as input the genes whose *t*-scores rank highest, but the informative nature of the genes is assessed according to whether we can classify unseen samples correctly.

Other *t*-type statistics have been proposed. One of the most widely used is the signal-to-noise ratio (SNR) score, used first in an early seminal paper in gene expression array research (22). Its definition,

$$\mathrm{SNR}_i = \frac{\mu_{i(1)} - \mu_{i(2)}}{\sigma_{i(1)} + \sigma_{i(2)}}, \qquad [2]$$

is appealing because of its simplicity and its intuitive interpretation: it measures the degree of overlap of the *i*th gene distribution in class 1 and class 2.

2.2.2. More Methods of Univariate Gene Selection

Aside from *t*-score-based methods, there have been many other univariate methods of gene selection reported in the recent literature. In (23), for example, information-theoretic ideas were used to design a gene selection method in which a gene is selected if there exists a gene expression value out of the *M* values which partitions the patients in such a way that the entropy of the proportions of cases and controls determined at each side of the partition is minimized. A maximum likelihood ratio approach was taken in (24) to rank genes in the order of most discriminating to least discriminating between two classes. Many other methods of gene selection have been proposed. These include the "ideal discriminator method" (which can be mapped to *t*-type statistics) (19), the "Wilcoxon rank sum test" (19), χ^2 statistics (23), a correlation-based feature selection (23), the Bayesian variable selection approach (25), and the Use-Fold approach, where genes are selected whose fold changes are greater than the corresponding assay noise (26), among others.

2.3. Selecting Genes Many at a Time: Multivariate Gene Selection

2.3.1. The Use of Singular Value Decomposition in Gene Expression Data Analysis

Aside from the univariate methods discussed in the previous section, there have been a number of gene selection methods that explicitly use the high-dimensional nature of the gene expression space. Among these, one of the earliest methods tried in the gene expression arena was singular value decomposition (SVD), probably because of its extensive use in other applications to cluster, visualize, and classify high-dimensional data. The early application of SVD to gene expression research (27–29) used the yeast cell cycle data (3). The application of SVD for gene selection in case-control studies has been considered (30,31) but has not yet been explored in depth. The importance of normalization, and the use of alternative normalizations to highlight differential behavior between groups of genes in different classes in cancer-control gene expression data, was considered in (32). As more work addresses the applicability of SVD and related dimensional-reduction analyses to the exploration of gene expression data (33,34), we shall see more systematic methods of gene selection using this multivariate technique (35). At the heart of the SVD methods for gene expression analysis is the notion of simultaneously clustering groups of genes and patients. This aim can also be achieved using alternative methods, as proposed in (36–39).

2.3.2. Other Methods of Multivariate Gene Selection

The recent literature also contains other interesting work on multivariate gene selection. These methods choose genes on the basis of their collective synergy to separate classes, and it can happen that many genes selected on the basis of multivariate analysis would have been deemed nonsignificant in terms of their individual differential expression. In (40) gene clusters were constructed iteratively according to the ability of the average cluster expression to discriminate between cancer and control in such a way that average cluster expression is uniformly low for one class and high for the other. The genes selected in clusters are subsequently validated by classification. Evolutionary algorithms have also been proposed (41) for gene selection where, as in (40), the selected genes were chosen as a collective and not on the basis of their individual ability to discriminate between classes.

The two previous approaches are examples of heuristics designed to overcome the difficulty of choosing all the potential 2^N groups of genes (with N of the order of 10,000) and evaluate each group's predictive ability. However, one

could explore all the pairs and triads and perhaps tetrads of genes, and decide their predictive power. This is the approach taken in (42), where the predictive ability of all the sets with less than 4 genes is tested, and those sets that perform the best (and above an error threshold) in a validation-by-classification task are selected as genes of interest for further biological study.

Another multivariate algorithm that has been used in different applications in the recent literature ((43–45); see also (39)) is Genes@Work, a gene expression pattern discovery approach. Genes@Work searches for patterns that differentiate one particular phenotype from another phenotype chosen as a reference or control. Each pattern consists of a group of genes observed to act consistently over a subset of the samples in the phenotype set (formed by either the cases or the controls). All subsets of genes and all subsets of experiments that satisfy some given pattern parameters are searched. These patterns can be found in a computationally efficient and exhaustive manner by algorithms that avoid searching the complete combinatorial space of possible patterns (46). Each pattern can be assigned a p-value, and the selected genes are the union of all the genes that participate in at least one statistically significant pattern. Thus, Genes@Work is an approach validated by statistical significance.

All of the methods presented in this section are interesting in that they interrogate the data from different perspectives. In this sense, a method can rescue as positives those genes that may have been left off as false negatives by other methods. We will explore the value of combining different gene selection methods in the following section.

3. COMBINING SELECTION METHODS PRODUCES A RICHER SET OF DIFFERENTIALLY EXPRESSED GENES

In this section we describe the application of a combination of gene selection methods to identify interesting genes in lymphoma data. In particular, we seek genes that differentiate between two types of lymphomas: diffuse large B-cell lymphoma (DLBCL, the most common lymphoid neoplasm) and follicular lymphomas (FL). FL is frequently characterized by a transformation to DLBCL, and therefore a comparative study of the gene expression profiles of these two lymphomas has been considered in the recent literature (47,48). In this section we compare the gene expression profile of these two cancers to exemplify the use of a combination of gene selection techniques.

Gene expression data for FL and DLBCL have been analyzed by Whitehead Institute (WI) researchers in (49), where the 50 largest and positive scoring genes (genes more expressed in DLBCL than in FL) and the 50 largest and negative scoring genes (genes less expressed in DLBCL than in FL) were selected using the signal-to-noise ratio method (SNR) described in §2.2.1. Each of these 100 genes appear to have a statistical significance better than 1% when its

SNR was assessed against the distribution of the SNRs of a similarly ranked gene in 500 class-label permutation experiments (see (49) and supplementary materials). We will use these 100 genes in combination with genes arising from alternative gene selection methods.

We analyzed the same lymphoma data set considered in (49) (hereafter called the WI data), with *t*-statistics as defined in §2.2.1, to obtain the 100 best-scoring genes according to *t*-score. The false discovery rate (a measure of significance that avoids the flood of false positives arising from multiple comparisons that we incur in microarray experiments (50)) corresponding to these 100 genes was estimated to be 5×10^{-6}. This estimate is based on an assumption that the *t*-scores are normally distributed under the null hypothesis that all genes are similarly distributed in DLBCL and FL. This was checked by random permutation of the DLBCL and FL labels in the data. Indeed, the pooled probability density of the *t*-scores of all the genes after randomization of the labels has an average of 0.03, a standard deviation of 1.03, and a kurtosis of 3.2, indicating reasonable resemblance to a Gaussian distribution. The 100 genes found using this *t*-statistics have an overlap of 42 genes with the 100 best genes found in (49) based on the SNR.

Both the SNR ratio and the *t*-score methods choose genes on the basis of a univariate criterion. There may be genes whose statistical significance according to the SNR or *t*-score method is small but whose significance would be larger if a multivariate approach were used. To explore this possibility we applied our multivariate gene expression pattern discovery algorithm, Genes@Work, to generate groups of markers that express differentially in DLBCL and FL. We applied Genes@Work to the WI data.[1] The union of genes that participated in at least one pattern with the parameters described in (46) resulted in 100 genes.

Figure 1 summarizes the information of the genes discovered specifically by each method. There are a total of 210 genes, of which only 17 were reported by the three methods. Genes@Work chose 52 genes that neither the SNR ratio nor the *t*-score method chose. The SNR method chose 34 genes that neither Genes@Work nor the *t*-score methods found. Similarly, the *t*-score method found 51 genes that neither of the other methods found. The fact that the 100 most significant genes reported by each of the methods considered found genes that the other methods did not is the result of the specific questions with which each method interrogates the data. In Genes@Work, a gene must correlate with other genes through a pattern to be reported. On the other hand, when selected by SNR, each gene is considered in isolation and the overlap of its distribution in DLBCL and FL must be small. Finally, the sample averages in DLBCL and FL must differ beyond the standard error to qualify as a gene selected by the t-statistics method.

The question may arise as to whether the combination of methods we advocated above is really necessary. Indeed, if by slightly relaxing the threshold of significance of any method we could engulf most of the genes discovered by a

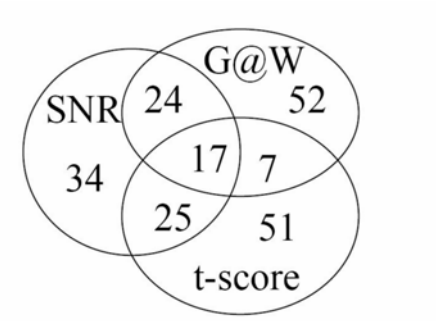


Figure 1. Venn diagrams of the set of genes identified in the analysis of diffuse large B-cell lymphoma and follicular lymphoma by each of three methods: the signal-to-noise ratio (SNR), Genes@Work (G@W), and the *t*-score. The numbers indicate the number of genes in each of the sets.

different method, then the argument that the two methods interrogate the data differently would be questionable. We mentioned earlier that using the *t*-test with a false discovery rate (FDR) of 5×10^{-6} we discovered 100 genes. Out of these 100 genes, 24 were also discovered by Genes@Work (Figure 1). How much should we relax the FDR in the *t*-score to engulf 90% of the genes discovered by Genes@Work? The answer to this question is very telling. To get 90% of the genes discovered by Genes@Work using the *t*-score method, the total number of genes discovered by the *t*-statistics should grow to be 1,839, corresponding to an FDR of 0.5. This FDR value is too permissive: at this false discovery rate we expect that half of the 1,839 are false positives! Similar results can be reached by exploring the other comparisons. We conclude that each of these methods interrogate the data in its own specific way. Obviously, only when the methods used are sufficiently different from each other does the combination of algorithms contribute novelty above the application of just one of the methods.

3.1. The Intersection or the Union?

It may be argued that the "best" genes differentiating between the two classes under study are the 17 genes within the intersection of all the methods. Even though this argument makes intuitive sense, it is not necessarily true. Indeed, in each of the methods used above genes were selected under a relatively strict *p*-value, which controls the specificity but not the sensitivity of the method. In other words, when we are very stringent in preventing false genes

from being selected, it is very likely that we will fail in terms of not allowing relevant genes to join our list of differentially expressing genes. A small *p*-value allows us to have a low false positive rate, but will probably make us incur a high false negative rate. By interrogating the data with several methods that address distinct aspects of the data, each method can rescue as true positives the genes left out by the other methods as false negatives. At the same time, we can keep a low rate of false positives by using low *p*-values for each of the participating methods. The ultimate proof of the soundness of the genes selected is, in any case, a biological validation of the selected genes. We address that issue next.

3.2. Validation of Gene Selection by Consistency with Independent Data

It was discussed earlier on that the typical validation strategies used in gene selection schemes are validation-by-classification and the validation by statistical significance (§2). An interesting alternative is what we have called the *validation-by-consistency* method (1), in which the selected genes are validated if they show consistency in their behavior in a different data set (different laboratories and maybe different technology, but the same types of tissues).

In (49), for example, 30 of the 100 markers were verified as informative by using the gene voting scheme introduced in (22), which is a validation-by-classification approach. All the 100 markers also passed a test of statistical significance, and were thus validated by statistical significance. In our case we could state that the statistical significance of the patterns found by Genes@Work was stringent given that the *p*-value of the least significant pattern was 10^{-10}. However, statistical significance need not mean biological relevance. Furthermore, when we combine several methods to discover differentially expressing genes, the gene composite resulting from the use of the different methods lacks an error estimate, even if there was a clear validation approach used in each of the chosen methods. Thus, it is desirable to have a means by which to validate the composite, a task for which validation by consistency can be extremely useful and telling.

We shall exemplify the validation-by-consistency approach in the set of genes found by merging the three methods described in the previous section. The heat map, or Eisen plot (51), of the 210 genes discovered by applying a combination of the *t*-score, the SNR, and the Genes@Work methods on the DLBCL/FL data generated at the Whitehead Institute can be viewed to the left of the yellow line in Figure 2. Two groups of genes can be easily visualized: the ones that overexpress in DLBCL compared to FL (mostly red in DLBCL and mostly blue in FL), and the ones that underexpress in DLBCL compared with FL (mostly blue in DLBCL and mostly red in FL). A stringent validation for these genes would be to check that the same neat separation is achieved in a

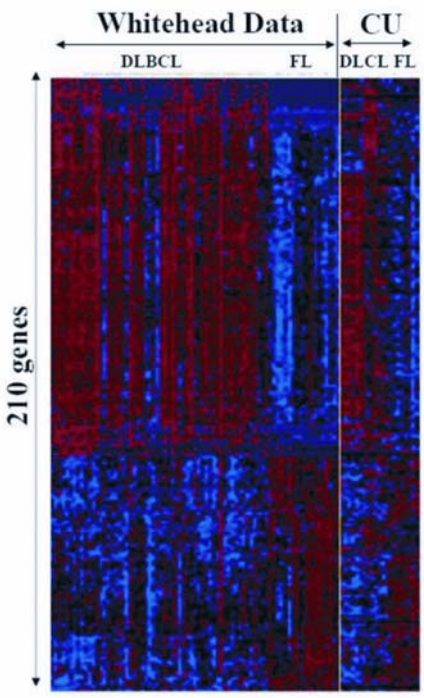


Figure 2. Heat map of the genes differentially expressed in diffuse large B-cell lymphoma data (DLBCL) and follicular lymphoma data (FL). Each row corresponds to a different gene, whereas each column is a different patient. For each gene, blue indicates downregulation and red upregulation. For gene selection we used a combination of the SNR, *t*-score, and Genes@Work methods. The data used for gene selection were produced at the Whitehead Institute (WI) (49) and are drawn to the left of the yellow line. The genes upregulated an downregulated in DLBCL vs. FL in the WI data show a consistent behavior in an independent data set produced at Columbia University (CU) (52), as seen in the heat map to the right of the yellow line.

gene expression data set produced from patients with the same diseases, but using different chips, different technicians, different patients, and different labs. These differences are likely to break the systematic errors that can arise if experiments are performed in the same lab, and the validation is therefore more telling about the underlying biology.

On the right side of the yellow line in Figure 2 we can observe how the genes selected using the WI data separate an independent data set, obtained in R. Dalla-Favera's lab at Columbia University (the CU data). This independent data set contains 14 DLBCL samples and 7 FL samples, and was previously used in (52). We can see that the CU data reproduce the gene expression profile found in the WI data. It is important to notice that the WI data were collected using the

Affymetrix GeneChip® HUGeneFL array, which predates the Hu95Av2 Affymetrix GeneChip® array used for the collection of the CU data. The last one is not just an anecdotic comment. The fact that different GeneChip® arrays were used in the comparison (different in gene probes, location of probes in the chip, number of probes, etc.) and yet a highly similar behavior was obtained in the CU data using the genes selected from the WI data is indicative that the composite of genes selected using different methods is biologically relevant in that it contains reproducible information about the DLBCL and FL phenotypes. *The selected genes are not a statistical artifact.*

The qualitative validation of the genes selected in the WI data and tested in the CU data revealed by Figure 2 needs to be complemented by a quantitative assessment of the likelihood that this similarity occurs by chance. We can estimate the statistical significance of the comparison by counting the number of consistent genes across the two data sets. We say that a gene is consistent if the sign of the difference of the average expression in DLBCL and FL is the same in both data sets. A p-value for the number of consistent genes can be estimated as the probability that the same or a larger number of genes found to be consistent in the two data sets will be found consistent if the genes are chosen at random. This p-value can be calculated from a binomial distribution with a probability parameter equal to 0.5 (i.e., the null hypothesis assumes that in the CU data set each gene had equal probability of overexpressing or underexpressing in DLBCL vs. FL). The p-value of the consistency for these selected features across these two data sets can be computed independently for the genes upregulated in DLBCL vs. FL, those downregulated, and the full set of genes. The results are shown in Table 1, where we can see that the resulting p-values are extremely small. Notice that this validation is neither a validation by statistical significance within the same data set nor a validation by classification. It is something in between those two validation schemes, in which a test data set is assessed for consistency with a training data set. From Table 1 we conclude that the gene expression profile identified by the three methods is highly reproducible in an independent data set. Indeed, out of the 210 selected genes, 184 (88%) showed a consistent behavior in the CU data set. This high percentage of coincidence is extremely unlikely to be found purely by chance, and confirms the informative nature of the genes selected by our gene selection methods.

4. GENE EXPRESSION ARRAYS CAN BE USED FOR DIAGNOSTICS: A CASE STUDY

The practice of both combining gene selection methods on the one hand, and of validating the selected genes across laboratories on the other, will likely be used frequently in the future. This is bound to be the case because of the existence of more available data sets on the same types of tissues in the public

Table 1. Consistency of the expression levels for the genes selected from the Whitehead Institute (WI) data in the Columbia University data (CU) corresponding to Figure 2

Profile	No. of genes in WI data	No. of genes consistent in in CU data	No. of genes inconsistent in CU data	*p*-value
DLBCL up/ FL down	129	115 (89%)	14 (11%)	1×10^{-20}
FL up/ DLBCL down	81	69 (85%)	12 (15%)	6×10^{-11}
Overall profile	210	184 (88%)	26 (12%)	5×10^{-30}

Three groups of genes are considered: those that are upregulated in diffuse large B-cell lymphoma (DLBCL) over follicular lymphoma (FL), those with the opposite behavior, and the union of the previous two sets. 88% of the genes selected from the WI data show a consistent behavior in the CU data. The likelihood that this occurs by chance is extremely low, as shown in the *p*-value column.

domain, and because of the meaningful nature of the exercise. Testing in an independent data set the genes selected in our data is basically a way to do diagnostics. In other words, if we trust our methods of data mining, and the gene expression array technology, then the genes found to express differentially in a data set can be used as markers for diagnosis. If the chosen genes are truly informative about the disease under study, one should in principle be able to decide whether an array with a sample from a previously unseen subject corresponds to an individual affected by the disease or to a healthy individual. In this section we will present proof that this can be successfully done.

In (45), we used univariate and multivariate methods to discover genes that differentiate between chronic lymphocytic leukemia (CLL) patients and normal B cells. CLL is the most common leukemia in the United States and is a significant cause of morbidity and mortality in the older adult population. The underlying cause of CLL, however, remains unknown. In this regard, gene expression profiling has been used successfully by a variety of investigators to discover genetic differences between tumor cells and normal counterpart cells (45,52,53). This information is proving to be very useful in understanding tumor cell biology. In (45), Affymetrix Hu95A GeneChips were used to profile the gene expression from 38 CLL patients and from 10 healthy age-matched individuals to identify key genetic differences between CLL and normal B cells. The univariate methods selected 37 gene probes, whereas the multivariate method yielded 54 gene probes. Only 10 of the 81 total probes were identified in com-

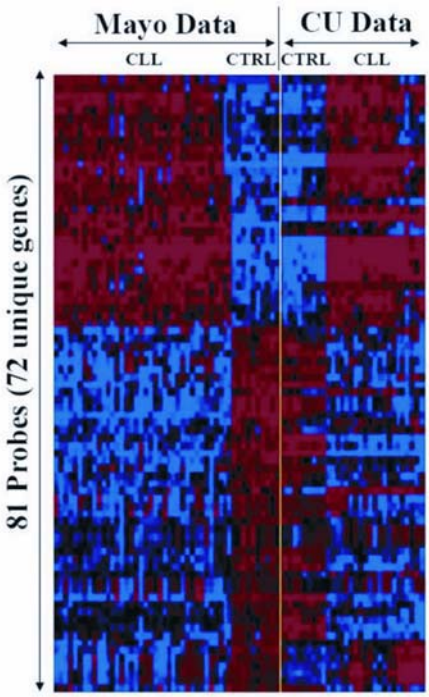


Figure 3. Identification of genes differentially expressed in CLL and normal B cells. Each row corresponds to a different gene, whereas each column is a different patient. The data to the left of the yellow line represent the supervised cluster analysis of 38 control CLL samples and 10 CLL samples using a combination of gene selection methods in data collected at the Mayo Clinic (Mayo data) (45,46). Rows correspond to genes and color changes within a row indicate expression levels relative to the average of the sample population (blue and red are low- and high-expression levels, respectively). The data to the right of the yellow line represent 10 controls and 21 CLL patients studied in a Columbia University (CU data) (45,52,53). The genes identified in the Mayo data that discriminate between normal B and CLL cells also discriminate normal B from CLL cells in the CU data.

mon by the two methods. This indicates that each of these methods interrogates the data in specific and different ways, as was discussed at length in a different example from the previous section.

The results of the analyses referred to above are shown in the colored matrix to the left of the yellow line in Figure 3, where we show an Eisen plot with the 81 probes (72 unique genes) and their differential expression in the CLL/control data collected in the Mayo Clinic (45) (the Mayo data). In Figure 3 the genes and the patients are ordered according to a hierarchical clustering organization. The differential behavior of these genes in the CLL cases and normal

controls is conspicuous. However, the fact that these markers clearly separate CLL patients from normal subjects in the Mayo data is expected given that these genes were specifically selected to separate cases from controls *in this data set*. A much more stringent verification of whether these markers express differentially in CLL versus control would be to establish an equivalent differential expression in a completely independent data set. The study of Klein et al. (52) (performed in R. Dalla-Favera's laboratory at Columbia University) provides such an independent data set, the Columbia University (CU) data. The colored matrix to the right of the yellow line in Figure 3 clearly shows that the 81 probes found to differentially express in the analyses of the Mayo data have the same qualitative behavior in the CU data. The fact that the majority of the genes that underexpress in CLL versus control (as well as the genes that overexpress in CLL versus control) in the Mayo data also do so in the CU data is a definitive indication of the informative nature of these genes in the context of CLL. This is not only a biological validation that the genes that arose from the gene selection algorithms are truly differentially expressed in CLL with respect to control, but also serves to validate the reproducibility of the DNA array technology. As stressed in the previous section, this kind of validation, which we called validation by consistency, is likely to become much more widespread as more gene expression data produced in different laboratories pertaining to the same case/control studies become available.

The previous discussion naturally leads to the question of whether the gene expression values of these 81 probes can be used to create a diagnostic method to determine whether or not a subject is affected by CLL. The idea in this case is to create a decision rule based on the gene expression values of these 81 probes. The flourishing field of machine learning (54,55) provides a number of techniques to determine decision rules. Many of these learning techniques have been applied to gene expression research. Among them we can mention nearest-neighbor classifiers (e.g. (14)), neural networks (e.g. (56)), and support vector machines (e.g. (57)). The latter has proved to be a very powerful method for separating two classes. In addition, it has an intuitive geometrical interpretation.

We shall use a support vector machine classifier to show that the disease state of the subjects in the Columbia data can be perfectly predicted using the Mayo data. Let us briefly explain how support vector machines operate. Figure 4a exemplifies a two-dimensional space, and in it we have two classes of points: the cases (solid circles) and controls (open circles). In its simplest conception, a support vector machine will attempt to compute a line that perfectly separates the cases from the controls, and whose distance to the closest point (or points) in each class is maximal (the optimal hyperplane). If this hyperplane exists the problem is said to be linearly separable, but this need not be the case in general. Some additional constraints are necessary if the problem is not linearly separable (58), but we shall not discuss them in this chapter. Once the optimal hyperplane has been found in the training set (in our case, the Mayo data), then we

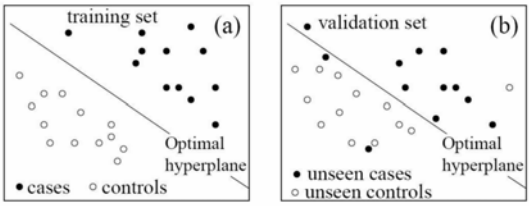


Figure 4. Schematic of the way support vector machines operate. (**a**) A set of points belonging to two classes (cases or controls) are used to create a decision boundary (the optimal separating hyperplane, diagonal line) that optimally separates between the two classes. (**b**) A validation set of previously unseen examples is classified on the basis of the decision boundary calculated using the training set. The points that fall below the optimal hyperplane are deemed to be controls, and those that fall above are deemed to be cases.

have a decision rule, which simply states that new points that fall in the region where the cases (respectively, controls) fell in the training set will be deemed to belong to the class of cases (respectively, controls). This is illustrated in Figure 4b, where an independent validation data set is plotted. If we use the optimal hyperplane as the decision boundary, we can count that two cases fall on the control side, whereas one control falls on the case side. In the example of Figure 4b, we have a false positive (FP) count of one (one control deemed to be a case) and a false negative (FN) count of two (two cases deemed to be controls). Similarly, the number of true positives (TP) is 10 (i.e., ten cases deemed to be cases), and the number of number of true negatives (TN) is 12 (controls deemed to be controls).

We created a decision rule using the Mayo data (similar to Figure 4a, but in 81 dimensions), and then applied it to the CU data as a validation set (as schematized in Figure 4b). When trained on the Mayo data, the support vector machine will attempt to find an 80-dimensional hyperplane that divides the 81-dimensional space into two sides, leaving all the case points on one side of the plane and all the control points on the other side. The genes selected for the Mayo data allowed us to find an optimal hyperplane that perfectly separates the Mayo data, i.e., the Mayo data are linearly separable. When we apply the decision boundary learned from the Mayo data to the Columbia data, we find that *all the Columbia subjects are perfectly classified*, that is, all the CLL patients segregate to the same side of the plane as the CLL patients in the Mayo Clinic. In like manner, all the control subjects fall on the other side of the plane. (This classification task was performed within the environment of the Genes@Work software.) This perfect classification indicates that the group of genes selected by our gene selection algorithms contains enough information to determine the status of health of a previously unseen patient.

The fact that the disease state from the Columbia subjects could be predicted without error should not be taken lightly. This is a case in which the training data was taken in an institution, and the validation data was taken from another setting. This makes the set of patients have different ethnic and geographical backgrounds, the gene expression chips correspond to different batches, the protocols followed for the gene expression assay differ, and the technicians doing the work have different styles. The fact that we could do perfect diagnosis on the Columbia subjects in spite of these differences constitutes proof of the principle that this set of genes can be used as markers in a gene expression-based diagnostics procedure.

5. DISCUSSION AND CONCLUSIONS

Transcription data comprise only the tip of the iceberg of the highly complex cellular systems that translate chemistry into life. Even so, gene expression data have a very intricate structure that requires sophisticated tools for its study. Experimental noise in gene expression data makes it necessary to measure a large number of samples to increase the statistical power of the analyses to "fish" for the signal in a sea of noise. Furthermore, the signal itself contains variability inherent to a biological system. Thus, the potential information that could be extracted from gene expression experiments is hidden in the data in more than one way. Different algorithms that probe different facets of the data are needed to discover the hidden patterns and can have their place in the gene expression analysis tool kit, especially if each of these methods is tailored to look for a particular statistical order or structure in the data. The complex gene–gene interactions that give rise to interesting cellular behavior are probably best captured with multivariate techniques, in which the unit of analysis is groups of genes rather than genes in isolation.

Validation of the genes selected by our algorithms can be accomplished using different strategies such as validation by statistical significance and validation by classification. But other methods exist that have not been sufficiently explored. In §3.2 we have shown that the ability of the genes determined in one data set to differentiate between class and control in an independent data set (generated in different laboratories and using different technologies) constitutes a rather stringent validation.

We advocated a multipronged approach to gene expression data analysis, consisting of combining different gene selection methods. There is a risk associated with this choice, as a tradeoff exists between *specificity* and *sensitivity*. By combining different statistical filters, we certainly accomplish higher sensitivity, yet in doing so we typically sacrifice specificity due to the acceptance of some outcomes that otherwise would have been excluded. This is a classic dilemma in statistics that can only be resolved in terms of the application context. For in-

stance, if we are seeking a set of diagnostic markers we should probably look for a small yet robust set of genes that strongly differentiates between case and control. This is the basis for DNA array-based diagnostics, of which we gave a successful example in the previous section. In this context, if we keep adding markers, the performance of the classification will eventually deteriorate. However, if we are interested in discovering biological processes that behave differentially between case and control populations, we should probably be rather encompassing in our differentiating features.

The lists of genes generated by gene selection methods will have to be organized with the help of literature search techniques, ontologies, or other sources of biological information. These lists will likely trigger hypothesis-driven research, which will give rise to a more mechanistic understanding of basic biological processes and of the ways in which these cellular processes are affected in disease.

6. ACKNOWLEDGMENTS

I am deeply indebted to my collaborators Yuhai Tu, Andrea Califano, Jorge Lepre, and Ulf Klein for many discussions on how to best analyze gene expression data. Thanks also go to Riccardo Dalla-Favera for providing the CLL, DLBCL and FL data from Columbia University, and Diane Jelinek and Neil Kay for providing the Mayo Clinic CLL data, and for their support on the biological side of the work presented in this chapter.

7. NOTE

1. For nonprofit or academic institutions, a free version of Genes@Work can be downloaded for noncommercial research purposes only from http://www.research.ibm.com/FunGen.

8. REFERENCES

1. Stolovitzky G. 2003. Gene selection in microarray data: the elephant, the blind men and our algorithms. *Curr Opin Struct Biol* **13**:370–376.
2. ArrayExpress database on World Wide Web: http://www.ebi.ac.uk/arrayexpress/
3. Stanford Microarray database on World Wide Web: http://genome. www4. stanford.edu/MicroArray/SMD/
4. GenomeWeb Gene Expression and Microarrays on World Wide Web: http://www.hgmp.mrc.ac.uk/GenomeWeb/nuc-genexp.html
5. YF Leung's Microarray Links on World Wide Web: http://ihome.cuhk.edu.hk/%7Eb400559/array.html

6. Bibliography on Microarray Data Analysis on World Wide Web: http://www.nslij-genetics.org/microarray/
7. Slonim DK. 2002. From patterns to pathways: gene expression data analysis comes of age. *Nature Genet* **32**:502–508.
8. Chaussabel D, Sher A. 2002. Mining microarray expression data by literature profiling. *Genome Biol* **3**:RESEARCH0055.
9. Khatri P, Draghici S, Ostermeier GC, Krawetz SA. 2002. Profiling gene expression using onto-express. *Genomics* **79**:266–270.
10. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, Groop LC. 2003. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genet* **34**:267–273.
11. Clarke PA, te Poele R, Wooster R, Workman P. 2001. Gene expression microarray analysis in cancer biology, pharmacology, and drug development: progress and potential. *Biochem Pharmacol* **62**:1311–1336.
12. Carr KM, Bittner M, Trent JM. 2003. Gene-expression profiling in human cutaneous melanoma. *Oncogene* **22**:3076–3080.
13. Salter AH, Nilsson KC. 2003. Informatics and multivariate analysis of toxicogenomics data. *Curr Opin Drug Discov Devel* **6**:117–122.
14. Pomeroy SL, Tamayo P, Gaasenbeek M, Sturla LM, Angelo M, McLaughlin ME, Kim JY, Goumnerova LC, Black PM, Lau C, Allen JC, Zagzag D, Olson JM, Curran T, Wetmore C, Biegel JA, Poggio T, Mukherjee S, Rifkin R, Califano A, Stolovitzky G, Louis DN, Mesirov JP, Lander ES, Golub TR. 2002. Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* **415**:436–442.
15. Service RF. 2003. Genetics and medicine: recruiting genes, proteins for a revolution in diagnostics. *Science* **300**:236–239.
16. Ardekani AM, Petricoin III EF, Hackette JL. 2003. Molecular diagnostics: an FDA perspective. *Expert Rev Mol Diagn* **3**:129–140.
17. Quackenbush J. 2002. Microarray data normalization and transformation. *Nature Genet* **32**:496–501.
18. Pan W. 2002. A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments. *Bioinformatics* **18**:546–554.
19. Troyanskaya OG, Garber ME, Brown PO, Botstein D, Altman RB. 2002. Nonparametric methods for identifying differentially expressed genes in microarray data. *Bioinformatics* **18**:1454–1461.
20. Pan W, Lin J, Le CT. 2002. How many replicates of arrays are required to detect gene expression changes in microarray experiments? a mixture model approach. *Genome Biol* **3**:research0022.
21. Li J, Liu H, Downing JR, Yeoh AE, Wong L. 2003. Simple rules underlying gene expression profiles of more than six subtypes of acute lymphoblastic leukemia (ALL) patients. *Bioinformatics* **19**:71–78.
22. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. 1999. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**:531–537.
23. Liu H, Li J, Wong L. 2002. A comparative study on feature selection and classification methods using gene expression profiles and proteomics patterns. *Genome Informatics* **13**:51–60.
24. Li W, Yang Y. 2002. Zipf's law in importance of genes for cancer classification using microarray data. *J Theor Biol* **219**:539–551.
25. Lee KE, Sha N, Dougherty ER, Vannucci M, Mallick BK. 2003. Gene selection: a Bayesian variable selection approach. *Bioinformatics* **19**:90–97.

26. Tu Y, Stolovitzky G, Klein U. 2002. Quantitative noise analysis for gene expression microarray experiments. *Proc Natl Acad Sci USA* **99**:14031–1436.
27. Holter NS, Mitra M, Maritan A, Cieplak M, Banavar JR, Fedoroff NV. 2000. Fundamental patterns underlying gene expression profiles: simplicity from complexity. *Proc Natl Acad Sci USA* **97**:8409–8414.
28. Alter O, Brown PO, Botstein D. 2000. Singular value decomposition for genome-wide expression data processing and modeling. *Proc Natl Acad Sci USA* **97**:10101–10106.
29. Alter O, Brown PO, Botstein D. 2003. Generalized singular value decomposition for comparative analysis of genome-scale expression data sets of two different organisms. *Proc Natl Acad Sci USA* **100**:3351–3356.
30. Nielsen TO, West RB, Linn SC, Alter O, Knowling MA, O'Connell JX, Zhu S, Fero M, Sherlock G, Pollack JR, Brown PO, Botstein D, van de Rijn M. 2002. Molecular characterisation of soft tissue tumours: a gene expression study. *Lancet* **359**:1301–1307.
31. Misra J, Schmitt W, Hwang D, Hsiao LL, Gullans S, Stephanopoulos G. 2002. Interactive exploration of microarray gene expression patterns in a reduced dimensional space. *Genome Res* **12**:1112–1120.
32. Kluger Y, Basri R, Chang JT, Gerstein M. 2003. Spectral biclustering of microarray data: coclustering genes and conditions. *Genome Res* **13**:703–716.
33. Liebermeister W. 2002. Linear modes of gene expression determined by independent component analysis. *Bioinformatics* **18**:51–60.
34. Antoniadis A, Lambert-Lacroix S, Leblanc F. 2003. Effective dimension reduction methods for tumor classification using gene expression data. *Bioinformatics* **19**:563–570.
35. Bicciato S, Luchini A, Di Bello C. 2003. PCA disjoint models for multiclass cancer analysis using gene expression data. *Bioinformatics* **19**:571–578.
36. Lazzeroni L, Owen A. 2002. Plaid models for gene expression data. *Statistica Sinica* **12**:61–86.
37. Cheng Y, Church GM. 2000. Biclustering of expression data. *Proc Int Conf Intell Syst Mol Biol* **8**:93–103.
38. Getz G, Levine E, Domany E. 2000. Coupled two-way clustering analysis of gene microarray data. *Proc Natl Acad Sci USA* **97**:12079–12084.
39. Califano A, Stolovitzky G, Tu Y. 2000. Analysis of gene expression microarrays for phenotype classification. *Proc Int Conf Intell Syst Mol Biol* **8**:75–85.
40. Dettling M, Buhlmann P. 2002. Supervised clustering of genes. *Genome Biol* **3**:RESEARCH0069.
41. Deutsch JM. 2003. Evolutionary algorithms for finding optimal gene sets in microarray prediction. *Bioinformatics* **19**:45–52.
42. Kim S, Dougherty ER, Barrera J, Chen Y, Bittner ML, Trent JM. 2002. Strong feature sets from small samples. *J Comput Biol* **9**:127–146.
43. Klein U, Tu Y, Stolovitzky GA, Keller JL, Haddad Jr J, Miljkovic V, Cattoretti G, Califano A, Dalla-Favera R. 2003. Transcriptional analysis of the B cell germinal center reaction. *Proc Natl Acad Sci USA* **100**:2639–2644.
44. Kuppers R, Klein U, Schwering I, Distler V, Brauninger A, Cattoretti G, Tu Y, Stolovitzky GA, Califano A, Hansmann ML, Dalla-Favera R. 2003. Identification of Hodgkin and Reed-Sternberg cell-specific genes by gene expression profiling. *J Clin Invest* **111**:529–537.
45. Jelinek DF, Tschumper RC, Stolovitzky GA, Iturria SJ, Tu Y, Lepre J, Shah N, Kay NE. 2003. Identification of a global gene expression signature of B-chronic lymphocytic leukemia. *Mol Cancer Res* **1**:346–361.
46. Lepre J, Rice JJ, Tu Y, Stolovitzky G. 2004. Genes@Work: an efficient algorithm for pattern discovery and multivariate feature selection in gene expression data. *Bioinformatics* **7**:1033–1044.
47. Martinez-Climent JA, Alizadeh AA, Segraves R, Blesa D, Rubio-Moscardo F, Albertson DG, Garcia-Conde J, Dyer MJ, Levy R, Pinkel D, Lossos IS. 2003. Transformation of follicular

lymphoma to diffuse large cell lymphoma is associated with a heterogeneous set of DNA copy number and gene expression alterations. *Blood* **101**:3109–3117.

48. Lossos IS, Alizadeh AA, Diehn M, Warnke R, Thorstenson Y, Oefner PJ, Brown PO, Botstein D, Levy R. 2002. Transformation of follicular lymphoma to diffuse large-cell lymphoma: alternative patterns with increased or decreased expression of c-myc and its regulated genes. *Proc Natl Acad Sci USA* **99**:8886–8891.
49. Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RC, Gaasenbeek M, Angelo M, Reich M, Pinkus GS, Ray TS, Koval MA, Last KW, Norton A, Lister TA, Mesirov J, Neuberg DS, Lander ES, Aster JC, Golub TR. 2002. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nature Med* **8**:68–74.
50. Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* **100**:9440–9405.
51. Eisen MB, Spellman PT, Brown PO, Botstein D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* **95**:14863–14868.
52. Klein U, Tu Y, Stolovitzky GA, Mattioli M, Cattoretti G, Husson H, Freedman A, Inghirami G, Cro L, Baldini L, Neri A, Califano A, Dalla-Favera R. 2001. Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells. *J Exp Med* **194**:1625–1638.
53. Rosenwald A, Alizadeh AA, Widhopf G, Simon R, Davis RE, Yu X, Yang L, Pickeral OK, Rassenti LZ, Powell J, Botstein D, Byrd JC, Grever MR, Cheson BD, Chiorazzi N, Wilson WH, Kipps TJ, Brown PO, Staudt LM. 2001. Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia. *J Exp Med* **194**:1639–1647.
54. Baldi P, Brunak S. 2001. *Bioinformatics, the machine learning approach*. MIT Press, Cambridge.
55. Hastie T, Tibshirani R, Friedman JH. 2001. *The elements of statistical learning*. Springer, New York.
56. Mateos A, Dopazo J, Jansen R, Tu Y, Gerstein M, Stolovitzky G. 2002. Systematic learning of gene functional classes from DNA array expression data by using multilayer perceptrons. *Genome Res* **12**:1703–1715.
57. Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, Haussler D. 2000. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* **16**:906–914.
58. Vapnik V. 1998. *Statistical learning theory*. Wiley-Interscience, New York.

3

APPLICATION OF BIOMOLECULAR COMPUTING TO MEDICAL SCIENCE: A BIOMOLECULAR DATABASE SYSTEM FOR STORAGE, PROCESSING, AND RETRIEVAL OF GENETIC INFORMATION AND MATERIAL

John H. Reif, Michael Hauser, Michael Pirrung, and Thomas LaBean

Departments of Computer Science, Ophthalmology, and Chemistry, Duke University, Durham, North Carolina

A key problem in medical science and genomics is that of the efficient storage, processing, and retrieval of genetic information and material. This chapter presents an architecture for a Biomolecular Database system that would provide a unique capability in genomics. It completely bypasses the usual transformation from biological material (genomic DNA and transcribed RNA) to digital media, as done in conventional bioinformatics. Instead, biotechnology techniques provide the needed capability of a Biomolecular Database system without ever transferring the biological information into digital media. The inputs to the system are DNA obtained from tissues: either genomic DNA, or reverse-transcript cDNA. The input DNA is then tagged with artificially synthesized DNA strands. These "information tags" encode essential information (e.g., identification of the DNA donor, as well as the date of the sample, gender, and date of birth) about the individual or cell type that the DNA was obtained from. The resulting Biomolecular Database is capable of containing a vast store of genomic DNA obtained from many individuals (multiple army divisions, etc.). For example, the DNA of a million individuals requires about 6 pedabits (6×10^{15} bits), but due to the compactness of DNA a volume the size of a conventional test tube with a few milliliters of solution could contain that entire Biomolecular Database. Known procedures for amplification and reproduction of the

Address correspondence to: John H. Reif, Department of Computer Science, Room D223, LSRC Building, Duke University, Durham, NC 27708 (reif@cs.duke.edu).

resulting Biomolecular Database are discussed. The Biomolecular Database system has the capability of retrieval of subsets of stored genetic material, which are specified by associative queries on the tags and/or the attached genomic DNA strands, as well as logical selection queries on the tags of the database. We describe how these queries can be executed by applying recombinant DNA operations on the Biomolecular Database, which have the effect of selection of subsets of the database as specified by the queries. In particular, we describe how to execute these queries on this Biomolecular Database by the use of biomolecular computing (also known as DNA computing) techniques, including execution of parallel associative search queries on DNA databases, and the execution of logical operations using recombinant DNA operations. We also utilize recent biotechnology developments (recombinant DNA technology, DNA hybridization arrays, DNA tagging methods, etc.), which are quickly being enhanced in scale (e.g., output via DNA hybridization array technology). The chapter also discusses applications of such a Biomolecular Database system to various medical sciences and genomic processing capabilities, including: (a) rapid identification of subpopulations possessing a specific known genotype, (b) large-scale gene expression profiling using DNA databases, and (c) streamlining identification of susceptibility genes (high-throughput screening of candidate genes to optimize genetic association analysis for complex diseases). Such a Biomolecular Database system may provide a revolutionary change in the way that these genomic problems are solved.

1. INTRODUCTION

1.1. Motivation: The Need for a Compact Database System for Storing, Processing, and Retrieving Genetic Information/Material

Recent advances in biotechnology (recombinant DNA techniques such as rapid DNA sequencing, cDNA hybridization arrays [see also this volume, Part IV, chapter 1, by Meinhart and Wereley], cell sorters, etc.) have resulted in many benefits in health fields. However, these advances in biotechnology have also brought risks and considerable further challenges. The *risks* include the use of biotechnology for weaponry, for example, diseases (or environmental stresses) engineered to attack and disable military personnel. The *challenges* include the difficulties associated with the acquisition, storage, processing, and retrieval of individual genetic information. In particular, it is apparent that the sequencing of the human genome is not sufficient for many medical therapies, and one may instead require information about the specific DNA of the diseased individual, as well as information concerning the expression of genes in various tissue and cell types. In the scenario of biological warfare, such individual-specific information can be essential for therapies or risk mitigation (e.g., identification of individuals likely to be susceptible to a particular biological attack). To do this, there must be a capability to store this biological information, and also a capability to execute queries that identify individuals who contain certain selected subsequences in their DNA (or transcribed RNA). Hence, what is needed is essentially a database system capable of storing and retrieving biological material and information.

This biological information is quite data-intensive; the DNA of a single human contains about 6 gigabits of information, and the number of genes that potentially may be expressed may total approximately 30,000 (up to 15,000 genes may be expressed in each particular cell type, and there are thousands of cell types). The DNA of a single individual contains about 3×10^9 bases, which (with 4 bases) is 6×10^9 bits. The DNA of a million individuals (e.g., a large military force) therefore requires 6 pedabits (a pedabit is 10^{15} bits). The expression information for a few dozen cell types in each of a million individuals may also require multiple pedabits. Although the acquisition of such a vast DNA databank may be feasible via standard biotechnology, the rapid transfer of the DNA of such a large number of individuals into digital media seems infeasible, due to the tedious and time-consuming nature of DNA sequencing. Even if this large amount of information could be transferred into digital media, it certainly would not be compact: current storage technologies require considerable volume (at least a few dozen cubic meters) to store a pedabit. Furthermore, even simple database operations on such a large amount of data require vast computational processing power (if executed in a few minutes).

1.2. Overview of the Biomolecular Database System

This chapter presents the architecture of a Biomolecular Database system for the efficient storage, processing, and retrieval of genetic information and material. It completely bypasses the usual transformation from biological material (genomic DNA and transcribed RNA) to digital media, as done in conventional bioinformatics. Instead, biotechnology techniques provide the needed capability of a Biomolecular Database system, without ever transferring the biological information into a digital media. It may provide a potentially unique and revolutionary capability in genomics.

1.2.1. DNA: An Ultra-Compact Storage Media

The storage media of this database system is comprised by the strands of DNA, which are (in comparison to RNA) relatively stable and non-reactive: they can be stored for a number of years without significant degradation. In particular, the genetic information can be stored in the form of DNA strands containing fragments of genomic DNA as well as appended strands of synthesized DNA ("information tags") encoding information relevant to the genomic DNA. This Biomolecular Database is capable of containing a vast store of genomic DNA obtained from many individuals (e.g., multiple divisions of an army). We can provide the store with a redundancy (i.e., a number of copies of each DNA in the database) that ranges from a few hundred or thousand downwards to perhaps 10,

as the stringency of the methods increase. As mentioned above, the DNA of a million individuals contains 6 pedabits, but due to the compactness of DNA a volume the size of a conventional test tube can contain the entire Biomolecular Database. A pedabit of information can be stored (with tenfold redundancy) in less than a few milligrams of dehydrated DNA, or when hydrated may be stored within a test tube containing a few milliliters of solution.

1.2.2. Construction of the Biomolecular Database System

The inputs to the system are DNA obtained from tissues: either genomic DNA, or reverse-transcript cDNA obtained from mRNA expressed from the DNA of a particular cell type. The Biomolecular Database is constructed as follows:

a. The input DNA strands are first fragmented, e.g., they may be partially digested into moderate-length sequences by the use of restriction enzymes. We describe a variety of methods for fragmentation protocols, and compare them by their distribution of strand lengths, and the predictability of the end sequences of the fragmented DNA.

b. The DNA are then tagged with artificially synthesized DNA strands. These "information tags" encode essential information (e.g., identification of DNA donor, as well as the date of the sample, gender, date of birth) about the individual or cell type that the DNA was obtained from. These "information tags" are represented by a sequence of distinct DNA words, each encoding variables over a small domain. We describe and test tagging protocols based on primer extension and utilizing the predictability of the end sequences of the fragmented DNA.

1.2.3. Processing Queries in the Biomolecular Database System

The chapter then discusses how to execute queries on the resulting Biomolecular Database. The system makes use of biomolecular computing (also known as DNA computing) methods to execute these queries, including the execution of parallel associative search queries on DNA databases, and the execution of logical operations using recombinant DNA operations. We also describe the use of conventional biotechnology (recombinant DNA technology, DNA hybridization arrays, DNA tagging methods, etc.), for example, output is via DNA hybridization array technology.

These queries include retrieval of subsets of the stored genetic material, which are specified by associative queries on the tags and/or the attached genomic DNA strands, as well as logical selection queries on the tags of the database. These queries are executed by applying recombinant DNA operations on this Biomolecular Database, which have the effect of selection of subsets of the database as specified by the queries. We describe two distinct methods for processing logical queries: a surface-based primer-extension method, as well as a solution-based PCR method. Query processing is executed with vast molecular-level parallelism by a sequence of biochemical reactions requiring a length of time that remains nearly invariant with respect to the size of the database up to extremely large numbers (e.g., up to 10^{15}). This is because the key limitation is the time for DNA hybridization, which is done in parallel on all the DNA. Output of the queries would be accomplished via DNA hybridization array technology.

1.2.4. Computer Simulations and Software

We describe computer simulations and software that can be used for the analysis and optimization of the experimental protocols. In particular, we describe the use of computer simulations for the design of hybridization targets for readout of information tags and SAGE tags by microarray analysis. We also discuss the scalability of these methods to do logical query processing within Biomolecular Databases of various sizes.

1.2.5. Applications

The chapter also discusses applications of a Biomolecular Database system to provide various genomic processing capabilities, including: (a) rapid identification of subpopulations possessing a specific known genotype, (b) large-scale gene expression profiling using Biomolecular Databases, and (c) streamlining identification of susceptibility genes: high-throughput screening of candidate genes to optimize genetic association analysis for complex diseases (see, e.g., this volume, Part III, section 6, on cancer). Such a Biomolecular Database system provides a revolutionary change in the way that these genomic problems can be solved, with the following advantages: (i) avoidance of sequencing for conversion from genomic DNA to digital media, (ii) extreme compactness and portability of storage media, (iii) use of vast molecular parallelism to execute operations, and (iv) scalability of the technology, requiring a volume that scales linearly with the size of the database, and a query time that is nearly invariant of that size. These unique advantages may potentially provide a number of opportunities for a variety of applications beyond medicine, since they also impact

defense and intelligence in the biological domain. Applications discussed include reasonable scenarios in (a) medical applications (e.g., in oncology, rapid screening, among a selected set of individuals, for expressed genes characteristic of specific cancers; see also preceding chapter 2 by Stolovitzky), (b) biological warfare (e.g., for biological threat analysis, rapid screening of a large selected set of personnel for possible susceptibility to natural or artificial diseases or environmental stresses, via their expressed genes), and (c) intelligence (e.g., identification of an individual, out of a large selected subpopulation, from small portions of highly fragmented DNA).

1.3. Organization of the Chapter

In this section we have provided a brief medical science motivation for a Biomolecular Database system, and a brief overview of the system. In §2 we briefly discuss relevant conventional biotechnologies and briefly overview the biomolecular computing (also known as DNA computing) field. In §3 we describe in detail our Biomolecular Database system. In that section we make use of various relevant biomolecular computing methods, including the use of word designs for synthetic DNA tags, execution of parallel associative search queries on DNA databases, and the execution of logical operations using recombinant DNA operations. In §4 we discuss a number of genomic processing applications of Biomolecular Database systems. In §5 we conclude with a review of potential advantages of Biomolecular Database systems.

2. REVIEW OF BIOTECHNOLOGIES FOR GENOMICS AND THE BIOMOLECULAR COMPUTING FIELD

2.1. Conventional Biotechnologies for Genomics

There have been considerable commercial biotechnological developments in the last few decades, and many further increases in scale can reasonably be expected over the next five years. For example, the DNA hybridization array technology developed by Affymetrix Inc. (the capability is currently up to 400,000 output spots, and within 5 years a projected 1,000,000 outputs) can be adapted for output of queries to conventional optical/electronic media. Other biotechnology firms (e.g., Genzyme Molecular Oncology Inc.) have developed competing biotechnologies.

2.1.1. Genomics

In the research field known as *genomics*, there are a number of main areas of focus, each with somewhat different goals. These include:

1. **DNA sequencing**. *Sequencing* is the determination of the specific base pair sequence making up the DNA. This tells us all the possible genes that a given organism may express—its genetic makeup. In conventional bioinformat-

ics, it is generally assumed that the genes discussed have been previously sequenced and placed in a computer database.

2. **Gene expression analysis**. *Expression analysis* attempts to determine which genes are being expressed in a given tissue or cell type at a specific moment in time. The objective, to identify all the genes that are being expressed, is challenging because of the great complexity of the mixture of mRNA being analyzed—each cell may express as many as tens of thousands of genes. SAGE Tagging and cDNA hybridization arrays, as discussed below, are techniques for determining comprehensive gene expression data for a given cell type or tissue. The technique of differential expression analysis compares the level of gene expression between two different samples. Variations in the level of expression of individual genes or groups of genes can provide valuable clues to the underlying mechanism of the disease process. There are a number of methods currently used to obtain comprehensive gene expression data.

a. *cDNA hybridization arrays*. A cDNA hybridization array is composed of distinct DNA strands arrayed at spatially distinct locations. A cDNA hybridization array operates by hybridizing the array with fluorescent-tagged probes made from mRNA, which anneal to its DNA strands. This generates a fluorescent image-defining expression, which provides a very rapid optical readout of expressed genes. However, cDNA hybridization arrays are generally manufactured for use with a given set of expressed genes, for example, those of a given cell type. The design and manufacture of cDNA hybridization arrays for a given expression library of a size over 10,000 can be quite costly and lengthy. Affymetrix has recently developed an oligonucleotide array, known as a UniversalChip, that is not specialized to any gene library; it consists of 2000 unique probe sequences that exhibit low cross-hybridization and broad sampling of sequence space It can be used with fluorescent-tagged probes made from DNA rather than mRNA. This technology can be used for output in a Biomolecular Database system.

b. *Serial analysis of gene expression* (SAGE) is a technique for profiling the genes present in a population of mRNA. By the use of various restriction enzymes, SAGE generates, for each mRNA, a 10-base tag that usually uniquely identifies a given gene. In the usual SAGE protocol, the resulting SAGE tags are blunt-end ligated together and the results are sequenced. The sequencing is faster than sequencing the entire set of expressed genes because the tags are much shorter than the actual mRNA they represent. Once sequencing is complete, the tag sequences can be looked up in a public database to find the corresponding gene. Using the sequence data and the current UniGene clusters, a computer processing stage determines the genes that have been expressed. SAGE can be used on any set of expressed genes and it is not specialized to a particular set. This technology can be adapted for use for additional information tags appended to the DNA in a Biomolecular Database. Genzyme Molecular Oncology Inc. is the developer of this SAGE technology.

c. *Differential expression analysis* is a technique for finding the difference in gene expression, for example, between two distinct gene types. Lynx Therapeutics Inc. has developed a randomized tagging technique for differential expression analysis. The randomized tagging techniques of Lynx Therapeutics may be adapted to determine the difference between two Biomolecular Database subsets.

2.2. Relevant Biomolecular Computing Techniques: Biomolecular Computing

In the field known as *Biomolecular Computing* (and also known as *DNA Computing*), computations are executed on data encoded in DNA strands, and computational operations are executed by use of recombinant DNA operations. Surveys of the entire field of DNA-based computation are given in (50,52).

The first experimental demonstration of Biomolecular Computing was offered by Adleman (1), who solved a small instance of a combinatorial search problem known as the Hamiltonian path problem. Considerable effort in the field of Biomolecular Computing methods has been made to solve Boolean satisfiability problems (SAT) problems, that is, the problem of finding Boolean variable assignments that satisfy a Boolean formula. Frutos and colleagues (24), Faulhammer et al. (23), and Liu et al. (37) applied surface-chemistry methods and Pirrung et al. (47) improved their fidelity. Adleman's group (11) recently solved an SAT problem with 20 Boolean variables using gel-separation methods. While the 20 Boolean variables size problem is impressive, Reif (52,53) has pointed out that the use of Biomolecular Computing to solve very large SAT problems is limited to at most approximately 80 variables, so is not greatly scalable in terms of number of variables.

The use of Biomolecular Computing to store and access large databases, in contrast, appears to be a much more scalable application. Baum (7) first discussed the use of DNA for information storage and associative search; Lipton (36) and Bancroft and coworkers (6) also discussed this application. Reif and LaBean (54) developed and Reif et al. (55) experimentally tested the synthesis of very large DNA-encoded databases with the capability of storing vast amounts of information in very compact volumes. Reif et al. (55) tested the use of DNA hybridization to do fast associative searches within these DNA databases. Reif (50) also developed theoretical DNA methods for executing more sophisticated database operations on DNA data, such as database join operations and various massively parallel operations on the DNA data. Gehani and Reif (27) investigated methods for executing DNA-based computation using microfluidics technologies. In addition, Gehani et al. (28) describe a number of methods for DNA-based cryptography and countermeasures for DNA-based steganography systems as well as discuss various modified DNA steganography systems that appear to have improved security. Kashiwamura and colleagues (34) describe the use of nested PCR to do hierarchical memory operations.

Suyama et al. (60) and Sakakibara and Suyama (59) have developed biomolecular computing methods for gene expression analysis. Garzon and colleagues (25) recently analyzed the efficiency and reliability of associative search in DNA databases, and Chen and colleagues (13) discussed DNA databases with natural DNA based on the prior work of Reif et al. (55) and the present work.

3. A BIOMOLECULAR DATABASE SYSTEM

3.1. Overview

The inputs to the system are natural DNA obtained from tissues: either genomic DNA or reverse-transcript cDNA obtained from mRNA expressed from the DNA of a particular cell type. A short piece of synthetic DNA is added to each natural DNA strand. This piece of synthetic DNA, called an information tag, is used to code information about the original piece of DNA. This information can include the age or gender of the person from whom the DNA came, or the clinical symptoms of individuals suffering from a disease. In a typical application, the Biomolecular Database consists of a mixture of DNA strands from many different people (or other organisms). This Biomolecular Database system is capable of storage, processing, and retrieval of genetic information and material. Individual molecules of DNA in the Biomolecular Database can be selected and removed from the mixture on the basis of the information encoded in their information tag. This chapter describes several innovative biological applications for Biomolecular Databases; in particular, we discuss the application of our Biomolecular Database system to a number of genomic information processing applications.

3.2. Biological Inputs

The inputs to the system are DNA obtained from tissues. This input DNA is typically either (i) genomic DNA, or (ii) reverse-transcript cDNA obtained from mRNA expressed from the DNA of a particular cell type. (To ensure stability and non-reactivity, we suggest that the database be composed of DNA rather than RNA.)

3.3. Preprocessing the DNA

Biochemical operations can be used to partially digest the DNA by restriction enzymes (ensuring the resulting DNA strands are of modest size), and then label the resulting genomic DNA fragments with synthetic DNA information tags.

3.3.1. Fragmentation of the input DNA

The creation of a Biomolecular Database must involve some degree of fragmentation of genomic DNA. While this may at first seem a very simple step, it is in fact critical to later processing that this fragmentation step be done in a highly controllable way. We describe several methods to produce DNA strands of the desired length. The methods are required to produce a predictable distribution of lengths, and to ensure that at least one of the resulting ends has a defined sequence.

a. *Mechanical shearing*. This is a method that produces a certain size distribution; however, it is not so useful in our context since the resulting ends have undefined sequences.

b. *Reagent-less methods to create breaks*. Pirrung, Zhao, and Harris (48) developed a nucleoside analogue whose backbone can be cleaved by long-wavelength UV light, and specific photocleavable T analogues could be used (analogous to the dUTP method). However, again it is not so useful in our context since the resulting ends have undefined sequences.

c. *Controlled digestion of high-MW DNA by DNAse I*. This is another method that can be used to produce DNA of a specific size range. It relies on careful monitoring of reaction progress and does not produce specific sequences at the ends of the fragments to enable ligation or PCR processes.

d. *Digestion of DNA with restriction endonucleases*. This offers the advantage that known sticky ends are generated.

e. *"Rare Cutting" endonucleases*. These can be used to produce DNA fragments of larger size. The recognition sequence of such enzymes is as large as 8 bp, meaning that, on average, DNA is cut to $1/(0.25^8)$ or 65 kb. In many situations fragments larger than 65 kb may be desired; for example, complicated loci with many introns might comprise as much as 100 kb, and that is just for one gene.

f. *PCR methods for fragmentation*. One attractive alternative is to use PCR. Random-primed PCR has been used to amplify the whole genome of a single sperm (4,35,67). The challenge in using this strategy is to create long amplicons. In principle, amplicon size in random-primed PCR is a function only of the average distance between two inward-facing hybridized primers, which is then a function only of the primer concentration and temperature. Modest flexibility exists in the hybridization temperature in PCR, so a fruitful strategy to make long amplicons is to lower the primer concentration. In order to efficiently amplify with a low primer concentration, the primers should have a high melting temperature (T_m). Increasing length and G/C content increase primer T_m. Random-primed PCR can therefore be examined with novel conditions and primer designs to maximize the amplicon length. Lengthening the random primers by oligonucleotide synthesis is straightforward. Making the primers G/C-rich is challenging, as G/C-rich templates are known to be more difficult to amplify

owing to increased secondary structure. Substitution of *nonstandard bases* such as deazaG for G alleviates this difficulty.

g. *UV-sensitive nucleoside analogues in cell growth for fragmentation*. Another approach is to grow immortalized lymphoblast cell lines in the presence of UV-sensitive nucleoside analogues Pirrung et al. (48). These analogues can be incorporated into the DNA of the cells, which could subsequently be cleaved by exposure to UV light. The concentration of the analogues determines the frequency of their incorporation, and the size of the resulting fragments.

3.4. Creation of the Tagged Biomolecular Database

These DNA tags are composed of a concatenation of short subsequences, which encode scalar data values. For example, the information tags may contain the individual's unique ID and the cell type (in the case of reverse-transcript cDNA obtained from the RNA expressed by a particular cell type) of the genomic DNA and may also encode other useful information (e.g., sex and birthdate of the individual). The tagging can be done using known methods, for example, a primer-extension reaction, using the fact that one of the ends of the genomic cDNA can be predicted by the use of the appropriate initial fragmentation process, and further designing the tags with ends complementary to these sticky ends resulting from the fragmentation process. The resultant database elements have tags on each 5'- and 3'-end. This can be done so that each Biomolecular Database strand bears a universal amplification (primer) sequence at the extreme 5'- and 3'-ends.

3.5. DNA Word Design for the Information Tags

A key problem is the design of a lexicon of short DNA sequences (DNA words) for the information tags in our Biomolecular Database. (Our DNA "information tag" sequences are in general a subset of such a lexicon). Careful word design is crucial for optimizing error control in the queries executed within the Biomolecular Database. Good word design can be used to minimize unwanted secondary structure and to minimize mismatching by maximizing binding specificity. There are conflicting requirements on word design: as strand length decreases (which is desirable), the difference between distinct information words decreases (not desirable). Prior work in DNA word design includes a four-base mismatch word design used for surface-based DNA computing (29), and Frutos et al. (24) showed that surface morphology may be an important factor for discrimination of mismatched DNA sequences. A three-base design was used by Cukras and coworkers (17). Evolutionary search methods for word designs are described by Deaton et al. (18). Other DNA word designs are de-

scribed by Baum (8), Deaton and colleagues (19), Mir (39), and Garzon et al. (25). Laboratory experiments of word designs are described by Kaplan et al. (33), and ligation experiments are described by Jonoska and Karl (32). Wood (66) considers the use of error-correcting codes for word design and to decrease mismatch errors. One can utilize and improve on these methods for DNA word design, including evolutionary search methods and error-correcting codes. Hartemink and colleagues (30) describe an automated constraint-based procedure for nucleotide sequence selection in word design. In designing the DNA tags used in the database, one needs to determine how many residues should be used for each data block of the tag sequence (the tag sequence on the database strands binds to the probe sequence on the query strand), and then decide how many words are required at each block position (determined by the number of values available to the variable).

The range of possible sentences entailed by a word-block construction scheme is shown in Figure 1. For each block position in the sequence one word is chosen from the word set and synthesized on the growing DNA strand. Separate reaction vessels are used for each word in the block, so that all word choices are utilized but only one is present on any particular strand. For example, the arrows indicate the trace that results in the sentence: word1A–word2D–word3A–word4B. A particular bead is drawn through a particular path in the set of possible word choices, but all possible paths are populated with beads, so all possible DNA sentences are synthesized. Each bead contains multiple copies of a single DNA sequence that can be synthesized by the well known mix-and-split synthesis scheme. Figure 2 shows the scaling of library diversity with increasing sentence length (block count) and increasing number of available words within each block. Diversity is calculated by raising the word count to the exponential power given by block count (i.e., diversity = $[\text{word count}]^{\text{block count}}$). As a simple example, to achieve a total diversity greater than one million with sentences containing 6 blocks, for example, one would require a set of 10 word choices per block. Also, to achieve a total diversity of 12^{14} with sentences containing 14 blocks, for example, also requires a set of 12 word choices per block. In designing a DNA-encoded database one must consider several important factors including the following. (i) The overall length of the oligonucleotide sequences used for matching is critical because sequence length directly affects the fidelity and melting temperature of DNA annealing. (ii) The Hamming distance (the number of changes required to morph one sequence into another) is another critical consideration. One would like to maximize the Hamming distance between all possible pairs of encodings in the database in order to minimize near-neighbor false-positive matching. One strategy for maintaining sequence distance is to assign block structures to the sequences with sets of allowed words (subsequences) defined for each block. (iii) Another important consideration is the choice of the words themselves and the grouping of words into sets within the blocks. Sentence length, desired library diversity, and word-pair distance

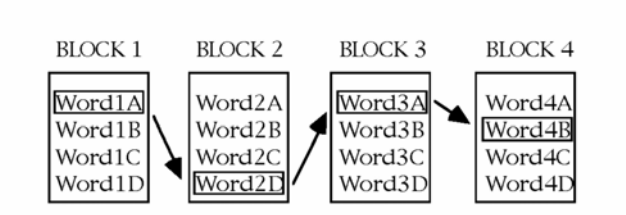


Figure 1. Range of possible sentences entailed by a word-block construction scheme.

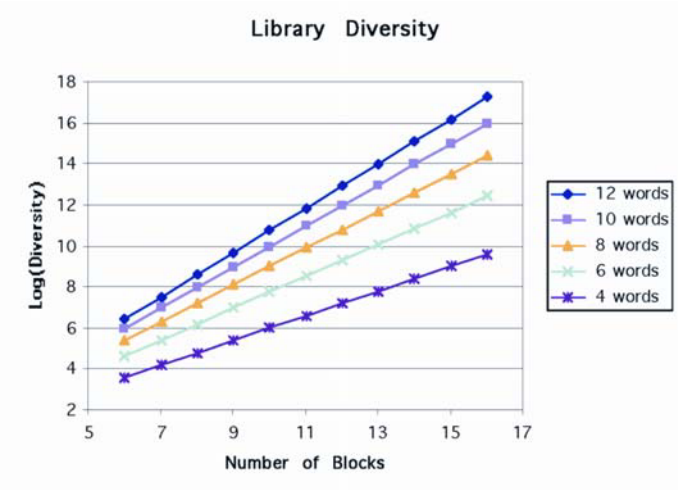


Figure 2. Scaling of library diversity with increasing sentence length (block count) and increasing number of available words within each block.

constraints all affect the choices of words in the lexicon. Word design can be maximized by careful design of words, the lexicon, and database elements, as well as experimental tuning of annealing conditions (e.g., temperature-ramp rate, pH, and buffer and salt concentrations). A useful tool in this task is the computer simulations of DNA hybridization, known as BIND, developed by Hartemink et al. (30).

3.6. Additional Tagging Methods for the DNA Strands

Various sophisticated tagging techniques have been developed by the biotechnology industry for expression analysis and differential expression analysis..

These include the SAGE tagging of Genzyme Molecular Oncology Inc. and the randomized tagging techniques of Lynx Therapeutics Inc.

a. *Serial analysis of gene expression (SAGE)* is a technique developed by *Genzyme Molecular Oncology Inc*., for profiling the genes present in a population of mRNA. By the use of various restriction enzymes, SAGE generates, for each mRNA, a 10-base tag that usually uniquely identifies a given gene. In the standard SAGE protocol, the resulting SAGE tags are blunt-end ligated and the results are sequenced. Such sequencing is faster than sequencing the entire expressed genes because the tags are much shorter than the actual mRNA they represent. Once sequencing is complete, one may look up the tag sequences in a public database to find the corresponding gene. Using the sequence data and the current UniGene clusters, a computer processing stage determines the genes that have been expressed. SAGE can be used on any set of expressed genes; it is not specialized to any particular set. This technology can be adapted for use as additional information tags appended to the DNA in our database.

b. *Differential expression analysis* is a technique developed by Lynx Therapeutics Inc. for finding the difference in gene expression, for example, between two distinct cell types. The randomized tagging techniques of Lynx Therapeutics Inc. can be adapted to determine the difference between two DNA database subsets.

c. *Hybrid methods*. One can modify these methods and extend them to apply to the tagged DNA strands of our database. This requires considerable changes in the protocols, due to unwanted hybridization that may occur as a result of combination of synthetic tags with genomic DNA in our database strands. However, these modified methods can provide further powerful capabilities, for example, the capability for fingerprinting (creating short DNA tags that are nearly unique IDs for longer DNA strands of the database), identification of expressed genes of selected DNA strands, and also the capability for differential expression analysis of distinct selected subsets of the Biomolecular Database.

3.7. Amplification and Reproduction of Biomolecular Databases

Once a Biomolecular Database is created, it is important to be able to accurately replicate it, as it may be consumed during the course of interrogation. Prudence suggests maintaining each database in an archive, and querying only daughter databases prepared from the archival forms. Since each database member is designed to bear a universal amplification (primer) sequence at the extreme 5'- and 3'-ends, database replication can be performed using PCR. Because the length of the DNA strands in the database might be quite substantial, including both biological DNA information and many flanking tag sequences, the ability to produce full-length amplicons with long templates is

crucial to maintaining the fidelity of the databases. "Long accurate" PCR techniques (62,63), using novel thermostable proofreading polymerase enzymes such as *Pfu*, are currently capable of amplifying loci of up to ~40 kb. While powerful, the database design should not be limited to this length by the method of database replication, and it may be easier to enable PCR to produce amplicons of somewhat longer length. One simply needs to enhance by a moderate multiple the (amplicon) length that can be reliably amplified.

Optimized choice of amplicons may be achieved by exploiting two principles: experimental design (9,10,21) and combinatorial chemistry (44,45). Continuous variables that affect PCR reactions include the temperature of the initiation (hot start), annealing, extension, and dissociation steps, and the concentration of buffer components, additives, nucleotides, primers, and template. These variables compose a multidimensional space. A pervasive challenge in science and technology is identifying specific values for each parameter affecting multivariable processes that result in globally optimum performance and avoid local maxima. Commercial software enables the design of experiments that much more reliably and quickly lead to the global optimum. Noncontinuous variables that affect PCR reactions include the identity of the template, primers, and polymerase. An optimum combination of these molecules can be found only by systematic screening for each. For tractable numbers of combinations, all can be examined explicitly. When the diversity space expands beyond that domain, "indexing" techniques are available that permit optimum performers to be identified even when in a mixture with lower performers (46). A selection of variable-length primers can be examined, including those incorporating modified bases (deazapurines, 2'-OMe RNA) that suppress primer consumption by dimerization. A selection of commercial polymerase enzyme systems can be examined, including MasterAmp™ Taq, ThermalAce,™ Advantage-Tth,™ AdvanTaq™, and KlenTaq™/Pfu. A selection of templates should be examined, including whole viral genomes, bacterial artificial chromosomes (BACs), yeast artificial chromosomes (YACs), and the smallest yeast chromosome (225 kb). Analysis of the products of these reactions is challenging due to shearing of large DNA molecules by conventional sieving matrices. Pulsed-field gel electrophoresis (12,41) can therefore be used with amplicons of this size.

3.8. Associative Search in Biomolecular Databases

3.8.1. DNA-Based Associative Search

Eric Baum (7) first proposed the idea of using DNA annealing to do parallel associative search in large databases encoded as DNA strands. The idea is very appealing since it represents a natural way to execute a computational task in massively parallel fashion. Moreover, the required volume scales only linearly

with the database size. However, there were further technical issues to be resolved. For example, the query may not be an exact match with any data in the database, but DNA annealing affinity methods work best for exact matches. Reif and LaBean (54) described improved biotechnology methods to do associative search in DNA databases. These methods adapted some information processing techniques (error-correction and VQ coding) to optimize input and output (I/O) to and from conventional media, and to refine the associative search from partial to exact matches.

Reif and colleagues (55) developed and then experimentally tested a method for executing associative searches in DNA databases of encoded images, and this method was tested using an artificially synthesized DNA database. Prior to that project, the idea of using DNA annealing to do parallel associative search in synthetic DNA databases had never been experimentally implemented. They detailed a study involving the design, construction, and testing of large databases for storage and retrieval of information within the nucleotide base sequences of artificial DNA molecules. The databases consisted of a large collection of single-stranded DNA molecules that was immobilized on polymer beads. Each database strand carried a particular DNA sequence, consisting of a number of sequence words drawn from a predetermined lexicon. They made a number of experimental databases of artificially synthesized DNA sequences designed for encoding digital data, scaled in increasing sizes. Each DNA strand of the database is single stranded, and encodes a number that provides the index to the database element. They used an extensive computer search for the design of the DNA word libraries, to ensure a significant Hamming distance between distinct words and allow for annealing discrimination. They constructed their largest synthetic databases in two phases. In the first phase they constructed an initial DNA database by combinatorial, mix-and-split methods on plastic microbeads. This constituted by far the largest artificially constructed synthetic databases of this sort. The next phase was the development of a construction method for much larger synthetic databases by combining pairs of the synthetic database strands so as to square the size of the database to approximately 10^{15} distinct data elements (each represented redundantly by over 10 identical strands of DNA). Even with this greater than tenfold redundancy, the DNA database using this construction method is extremely compact, and requires only 10 milligrams of DNA.

3.8.2. Associative Search via PCR

PCR methods can be used for associative search queries in Biomolecular Databases (in particular, on the words of the tagged portions of the Biomolecular Database strands), using known and modified PCR techniques previously developed by Reif and coworkers (55). They describe experiments for executing

associative search queries within the above described synthetic DNA databases. Associative search queries were executed by hybridization of a database DNA strand with a complementary query strand. Discrimination in annealing experiments was enhanced by the library design, which guaranteed a minimum Hamming distance between distinct sequences. In their initial annealing experiments for processing associative search queries, they employed fluorescently labeled query strands and then performed separation of fluorescent versus nonfluorescent beads by Fluorescence-Activated Cell Sorting (FACS). They also experimentally tested variants of conventional PCR techniques for executing associative search queries, and, in addition, developed a PCR technique for associative search in the pairwise constructed DNA database.

3.8.3. Analysis of Associative Search

Similar error analysis and experimental testing methods can be employed in our proposed generalizations of this prior work (55) to tagged genomic DNA. It would be informative to measure rates of various search errors including: false positives from near-neighbor mismatches, partial matches, and nonspecific binding, as well as false negatives from limit-of-detection problems. It is desirable to directly measure the limits of detection, and to measure the ability to retrieve rare sequences within databases of high strand diversity.

3.9. Logical Query Processing in Biomolecular Databases

Biochemical operations can be used to execute query operations on this Biomolecular Database, so as to retrieve subsets of the Biomolecular Database. Each of the information strands of the database encodes a sequence of data values v_1, v_2, ..., v_k, where the ith value v_i ranges over a small finite domain D_i (e.g., D_i typically would range over 10 or less possible values, each encoded by a distinct fixed-length DNA sequence). Retrieval can be specified by logical queries on the tags of the database as well as associative queries on the attached genomic DNA strands. The associative searches can be executed by recombinant DNA operations, for example, variants of PCR combined with surface-chemistry methods and/or solution-based methods. The logical queries include the following: (i) SELECTION—selects DNA strands of a given ID or cell type; and (ii) Logical SELECTION—executes logical queries that select those genomic DNA strands whose information strands satisfy a specified logical query formula, whose logical conjunctives include AND as well as OR. These logical conjunctives are applied to selective predicates of the form "Tag(i) = v", where Tag(i) is the ith portion of the information tag of a DNA strand of the database, and v is a fixed value over the domain D_i. (The Boolean NOT of a selective

predicate of the form "Tag(i) = v" is not applied directly (since PCR and similar methods do not allow this) by instead applying the OR of selective predicates of the form "Tag(i) = u" for all possible u in D_i that are not equal to v). These selection operations can be executed by the use of recombinant DNA operations, applying and improving on logical processing methods developed in the field of DNA computing. Furthermore, one can provide the additional operation of selective amplification of the DNA populations. If these amplification operations are also executed, the logical selection and amplification operations result in a test tube whose selected DNA is vastly amplified. After the amplification process is completed, the output strands should vastly predominate all other strands of the Biomolecular Database. Other database operations that can be implemented by biochemical operations include database unions and limited joins (50).

3.9.1. Scalability of Our Query Processing

These operations can be executed in a scalable way. The required volume never grows significantly; the volume is a fixed linear function of the number of elements of the database. (The constant multiple here is the degree of redundancy with which DNA strands are used to store database elements; we expect that one can allow between a few hundred and possibly as few as 10 DNA strands to encode a given database element.) The number of required DNA hybridization steps grows only linearly with the size of the query formula. So the time for executing a query grows just linearly with the length of the query formula, which in practice is of very modest size (as compared to the size of the database, which can be enormous)—say 20 or so variables. Hence the key time limitation is the time for DNA hybridization. But DNA hybridization time is nearly invariant of the size of the database even if the hybridization is execution on an enormous amount of DNA (up to extremely large database sizes, say 10^{15}). However, there are considerable technical challenges in the design of the protocols—for example, biological data strands may be originally dsDNA while search protocols would function best with ssDNA (hence the protocols need to either form ssDNA or be modified appropriately). A key additional technical challenge in scaling the technology is the scale and number of resulting molecular biology reactions, requiring many tedious laboratory steps, particularly in the case of extremely large database sizes. This can be addressed by subsequent automation. We discuss two distinct methods for logical query processing: the first uses primer-extension techniques on solid support, previously developed (47) to solve SAT problems, and the other uses solution-based PCR amplification techniques. The second has greater potential for scalability due to the fact that it is solution based (so the chemistry operates in 3D, rather than being con-

strained to a surface). In both cases, one can apply DNA hybridization array technology for output of query results.

3.9.2. Executing Queries into the Biomolecular Database via Primer-Extension Techniques on Solid Support

A number of DNA computing researchers have developed microarray methods for DNA-based computing, exploiting the high fidelity of the primer-extension reaction to detect complementarity between primer libraries of all solutions to SAT problems and logical queries as templates (see, e.g., the work of Faulhammer et al. (23) and Liu et al. (37), and also that of Pirrung et al. (47), which improved fidelity). Primer extension is a two-step process, involving first annealing of a template molecule to the primer, the efficiency of which is directly related to sequence complementarity throughout the primer/template complex (see Figure 3). Second, a polymerase enzyme binds to the primer/template complex and adds a nucleotide or nucleotides complementary to the base

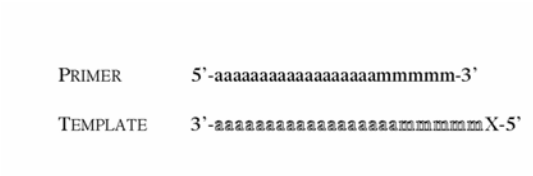


Figure 3. A primer and template used in the primer-extension method for logical queries.

t h e b a s e X, the first unpaired base at the 5'-end of the template, only when there is a perfect match in the last portion of base pairs of the primer/template complex. It is important to emphasize that while primer extension was in this case performed on a DNA microarray, the elementary step of a polymerase chain reaction (PCR) is also a primer-extension process and thus is subject to the same stringent sequence requirements. The variables (primers) in the SAT computation of (47) were composed of two portions, which can be considered the message (the last few bases "m" at the 3'-end of the oligonucleotide) and the address (a sequence of bases "a" at the 5'-end). With the base-4 encoding of DNA, a message sequence is capable of encoding 10 Boolean variables. For the experiment of (47) all addresses were the same, as only one SAT problem was being addressed. However, this need not be the case. Using similar designs, it is possible to design up to approximately 20 blocks of distinct address sequences, which con-

catenated form the tag. Each of these blocks of distinct address sequences should exhibit no cross-hybridization under stringent conditions (a Hamming distance of at least 5), thereby enabling independent encoding and primer extension and therefore interrogation of up to 20 distinct attributes, each with up to 10 scalar values.

Use of the primer-extension method for logical queries into the Biomolecular Database is most efficient if performed in solution rather than on a microarray. This creates a challenge in terms of product detection and identification. The following method enables both to be accomplished. An example is presented for the fate of one molecule, though it is appreciated that all molecules in the library are subjected to the same process in parallel. The database member is a DNA molecule that has been created by the methods described earlier, with a biological DNA sequence flanked by one or more created tag sequences, which are to be the templates in a primer-extension reaction. The "bottom" strand is interrogated in this example. Primers with the following structure (shown in expanded form below the database element in Figure 4) are created to interrogate each tag. Complements to the address sequence in the tag/template are the same in each primer. Also common to each primer is a "barbed tail" in the form of a 5'-psoralen group. The irradiation of psoralens with long-wavelength UV is widely used to cross-crosslink duplex DNA (31,43,64). The message sequence

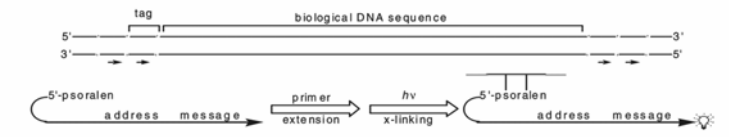


Figure 4. Obtaining a bottom strand in single-strand form, hybridized to a cDNA microarray, via photochemical cross-linkage of psoralen.

must be unique to each variable value, meaning that up to 10 primers are prepared per variable. The primer is also designed to address a unique X base in the tag/template to be interrogated. The primer-extension reaction is performed using a dideoxynucleotide terminator complementary to X and bearing a fluorescent dye with a unique and readily imaged emission spectrum. The dye color is specific to the variable, with the same dye/terminator being used for all interrogations of that variable. Multiple tags can be interrogated simultaneously because their dyes are different. The challenge at this stage is to read out the tags (based on the color(s) of the incorporated fluorescent dye(s)) in the context of the biological DNA. While the primer is still bound to the template, the psoralen

is photochemically cross-linked to the bottom strand of the library member, preserving the color of the dye. The bottom strand can then be obtained in single-strand form, which is hybridized to a cDNA microarray (see Figure 4). The color(s) of the array element complementary to the biological DNA identify the outcome of the queries of the tag sequences connected with it.

This concept could be applied in a similar fashion to a sequential (nested) PCR process by omitting the terminators and psoralen and providing one primer for the top strand and one for the bottom strand in each PCR. The eventual production of a full-length amplicon is dependent on the complementarity of *each* of the primers (logical AND) with its cognate tag sequence. This approach lends well to the use of DNA hybridization array technology for output of query results, providing distinct special locations for distinct outputs.

Another approach for executing Boolean queries on a Biomolecular Database is to use the gel separation-based method for SAT from Braich et al. (11), who succeeded in solving a 20-variable Boolean satisfiability problem. Although the queries would be executed on the tag portions of the DNA strands of the database, it is not clear how the efficiency of these separation methods would be affected by the genomic portion of the DNA strands in the databases.

3.9.3. Executing Queries in the Biomolecular Database via PCR Amplification Techniques

Another approach for Logical Query Processing is to use a variant of PCR amplification. The goal of this query processing is to selectively amplify only those DNA sequences (the *output strands*) whose information tags satisfy a given logical query. After the amplification process is completed, the output strands would vastly predominate all other strands of the Biomolecular Database.

3.9.4. Initialization Before Logical Query Processing

First, operations are executed that generate, from each DNA strand in the database, a new strand containing a concatenation of multiple copies of the Watson-Crick complement of the original strand. This can be done by a known sequence of routine recombinant DNA operations known as rolling circle replication (38). This begins by a circularization of each strand on the database, and then a primer-extension reaction on the circularized strand that repeatedly replicates the complement of the DNA strand to form a repeated sequence, followed by denature and separation of the result. The length of the resulting DNA strands is predictable (via the time duration and various parameters, including temperature) only to a degree, but it is predictable enough to allow us to con-

struct strands that we *expect* to have at least a given repeat length (as required by the below protocol). Recall that we have assumed that each database DNA strand is also redundantly represented by a number (ranging from up to a few thousand down to as few as 10) of identical DNA strands. This redundancy aids us here, since this initialization procedure results in a Biomolecular Database, where most of the redundant strands are lengthened by at least the given multiplying factor.

Figure 5 illustrates a scheme for processing genomic DNA into this database format which might include the following steps: (i) Cleave dsDNA into manageable pieces. (ii) Append prefixes to both ends of both strands. Heat denature dsDNA. Anneal to circularizing oligo. (iii) Ligate ssDNA circles. (iv) DNA polymerase reaction with circular templates to produce linear ssDNA containing multiple concatenated database entries. Note that the process of converting the DNA into database format may have unintended effects on the representation of entries in the database due to uneven amplification. Artificial bias may take the form of variations in the number of copies present on the average strand (distribution of strand lengths) or differences in the number of strands present for a given database entry. These protocols need to be optimized to take into account these possible affects.

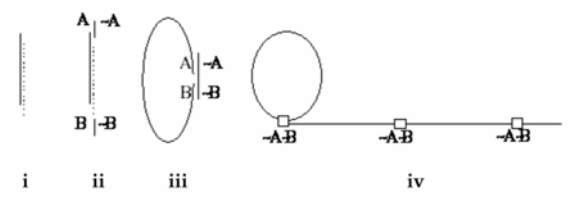


Figure 5. Scheme for processing genomic DNA into database format.

Multiple copies of the database entry are required on a single ssDNA strand, so that when Boolean variables recorded in the prefixes (A and B in Figure 5) are queried by primer binding and PCR, information recorded farther out toward the ends of the strand is not lost by failure to be copied (PCR only amplifies sequence physically between primer binding sites). The goal is to keep at least a few copies of the prefix information internally within the database strands so that information is not lost to subsequent rounds of query.

3.9.5. Logical Query Processing Using Repeated PCR Operations

We assume that the logical query is presented as the logical AND of a list of K logical clauses (each clause needs to be satisfied), where each clause consists

of the logical OR of a list of literals (the literals can be Boolean variables or their negation), one of which needs to be satisfied. Each clause *C* in the formula is processed in turn, selectively amplifying only those DNA strands whose Boolean variables satisfy at least one literal of that given clause. To do so, one adds PCR primers encoding the literals of that clause *C* and their Watson-Crick complements. Then a series of primer-extension reactions are executed that replicate only those DNA strands (or their Watson-Crick complements) that have subsequences that encode one of the literals of clause *C*. This process, applied as a series of PCR cycles, thus amplifies only those DNA strands whose Boolean variables satisfy at least one literal of that given clause, so that they vastly predominate all other strands of the Biomolecular Database. (**A technical note**: on each cycle, the amplified strands undergo loss of the material prefixing the primer's location, but the initial step of concatenating to each DNA strand in the database multiple copies of the strand ensure that is not a problem.) After the process is completed for each of the clauses in turn, the output strands that satisfy all the clauses would vastly predominate all other strands of the Biomolecular Database. This method for processing a logical query in the database is exquisitely sensitive: to get a result, one requires that the initial database have no more than 10 identical strands of DNA that satisfy the query. Again, DNA hybridization array technology can be used for output of query results, providing distinct special locations for distinct outputs.

3.9.6. Scalability

As discussed above, our query processing is executed with vast molecular-level parallelism by a sequence of biochemical reactions requiring a time that remains nearly invariant of the size of the database up to extremely large database sizes (e.g, up to 10^{15}). This is because the key limitation is the time for DNA hybridization, which is done in parallel for all the DNA.

3.10. Management of Errors

The logical and associative searches used to select specific molecules and sets of molecules from Biomolecular Databases are not 100% specific or effective. There may be several different kinds of errors: false negatives (appropriate DNA strands are present but not selected, either because of a lack of sensitivity or depletion of the relevant sequences from the database), false positives (inappropriate DNA strands are selected along with desired strands), errors based on degradation of the Biomolecular Database contents, and errors resulting from poorly designed queries, based on incomplete understanding of complex biological parameters. These kinds of errors can affect the results of the applica-

tions described here. For example, false negative errors can prevent finding existing individuals with the desired genetic variant. This is less serious than a false positive result, which could lead to sending a nonresistant individual into a contaminated area, under the false belief that he is genetically protected from a biological agent. A similar type of error could arise from database degradation (as, for example, from repeated error-prone duplication of the database). This type of error can be easily eliminated by follow-up confirmatory screening of that single individual's DNA. In general, it would be best to use the Biomolecular Database for very powerful and rapid selections based on genetic information, but then to confirm all results on individual DNA samples. This would require maintenance of individual stocks of DNA for each individual. This is a relatively large task, but well within current technology. An LIMS (laboratory information maintenance software) system and robotic liquid handling capacity are musts for this type of storage. There could also be errors of magnitude. These errors result from preferential amplification of one DNA strand over another. This kind of error is particularly troublesome, for it would skew allele frequencies. It may be necessary for us to monitor the frequency and extent of such errors and develop Boolean search strategies that minimize them. The final type of error, based on an incomplete understanding of the human genome, can only be rectified by continued research in other fields. This type of error could result from incomplete knowledge of the way in which genetic variants are distributed among different racial and ethnic groups. For example, the well-described *ccr5* variant that prevents HIV infection has been detected to date only among white males. If one were to select for nonexistent African American females expressing this variant, one might well obtain a small number of false positives. This type of error could also arise from mistaken assumptions. A given genetic variant might protect Hispanic females from infection by a given biologic agent, but oriental males carrying the same variant might be fully susceptible to infection because of another independent genetic variant.

3.11. Computer Simulations

Reif et al. (52,53) have made computer simulations of their methods for DNA-based associative search. They constructed computer software (viewable on the web) that provide a simulation of the entire experimental process, including conversion of this attribute database into a DNA database using DNA chips, the PCR method for associative search in the DNA database (using a software simulation of the kinetics of DNA hybridization), and, finally, conversion of the result of this query (using extensions of techniques described in (55)) into conventional media by use of a DNA expression array. Our computer simulation software to the above-described query processing provides a basis for future

software to simulate and optimize experimental protocols for query processing in Biomolecular Database systems.

4. APPLYING OUR BIOMOLECULAR DATABASE SYSTEM TO EXECUTE GENOMIC PROCESSING

There is tremendous potential to apply Biomolecular Databases to the solution of a number of biological problems. The huge amount of data provided by the sequencing of the human genome has outstripped many conventional methods for DNA analysis.

4.1. Genomic Processing Applications

We now discuss applications of such a Biomolecular Database system to provide key genomic processing capabilities. Three basic kinds of applications are discussed, which demonstrate different ways in which the massive parallelism of Biomolecular Databases can be used: (1) *Rapid identification of individuals either susceptible to or resistant to chemical or biological agents*. We describe the selection of a group of DNA molecules based on a common property, and then use the information tags to identify the individuals selected. (2) *Large-scale gene expression profiling using Biomolecular Databases*. Expressed genes from multiple tissues are represented in a Biomolecular Database, from which they can be selected individually or in groups for subsequent expression analysis. (3) *High-throughput screening of candidate genes to optimize genetic association analysis for complex diseases such as heart disease or Parkinson's disease*. Pools of individuals are selected through use of the information tags appended to each DNA molecule in the database. The pools so selected are then subjected to genetic analysis. We describe in detail these three applications that concern genomic information processing, and constitute important genomic processing applications of Biomolecular Database systems for medical science.

4.1.1. Rapid Identification of All Individuals Possessing a Specific Known Genotype

A single known genotype can confer properties making the individual either susceptible to or resistant to a particular chemical or biological agent found in the environment. It is certainly possible with existing biotechnology (e.g., hybridization experiments) to screen individuals for a given genotype. This is done one individual at a time, and is thus a relatively slow process. In addition, the cost of traditional genotyping of an individual ranges from $300 to over $1,000.

(At least one genotype databank has executed genotyping of approximately 1,000 individuals at considerable expense and time; however, for experimental purposes, that databank provides examples of previously executed individual genotyping at a cost of $0.50–1.00 per sample.) Clearly, an effort to screen a large number of individuals (say a million) would be slow and very expensive. In contrast, the methodology described herein is a *selection* for individuals with a certain genotype rather than a screen. It is correspondingly faster and less expensive. There is currently no available methodology for selection of specific genotypes. Many drugs that are quite effective in treating disease are very toxic to a small portion of the population. Currently, such drugs are removed from the market to avoid these rare but fatal adverse reactions. Such an approach is very costly from the standpoint of untreated disease. The removal of drugs from the marketplace because of rare fatal reactions is very costly in terms of untreated individuals, as well as the money spent on bringing those drugs to the market in the first place. Improved methods for identification of individuals at risk for adverse reactions would eliminate this cost. The capability of screening large numbers of individuals for a given genotype could also avoid a tremendous potential loss of life in the event of battlefield release of biological weapons or chemical agents.

As another example of a clinical application, one could construct a Biomolecular Database made from the blood samples of people suffering from Alzheimer's disease and their families, with the goal of finding genes that may increase people's risk of contracting the condition. The information tags could be used to select specific groups of molecules from this database. These molecules, which come from people with similar clinical symptoms, can then be used to test a large number of possible Alzheimer's disease genes. Genes that yield promising results could then be tested on the large number of individual samples from which the Biomolecular Database was made. The advantage of this approach is that it allows very efficient use of the limited DNA samples, and it is a good way to look at lots of different combinations of clinical features.

4.1.2. Large-Scale Gene Expression Profiling Using Biomolecular Databases

One may wish to determine the entire set of genes expressed by a particular cell type for a population of individuals who suffer debilitating effects due to a (perhaps unknown) chemical or biological agent. This may allow us to determine if there is a single or small number of genotypes that characterize susceptibility to that agent, over that population. Gene expression profiling is a labor-intensive and slow process. The conventional methods used are as follows: (a) *cDNA Hybridization Arrays*, which are 2D arrays of DNA spotted onto a solid support in an addressable way such that the spatial location of a spot identifies to the sequence of the DNA bound there. The input cDNA is labeled with a fluo-

rescent dye with a unique and readily imaged emission spectrum. After annealing on this array, the fluorescent cDNA provide a visual readout of the expression. (b) *SAGE libraries*, prepared by extraction from cDNA of very short tag sequences that characterize the expressed gene, followed by concatenation of a number of these tag sequences together for sequencing. Then computer software (using prior information on the relation between these tags and the original expressed cDNA) is used to determine which genes are being expressed. Gene expression profiling can require the development of new cDNA hybridization arrays, or the construction and sequencing of SAGE libraries. The methods for parallel analysis of large numbers of samples described here would streamline this process. In addition, the readout of SAGE data by microarray hybridization would result in significant savings of time and money as compared to the standard method of sequencing SAGE libraries. It would enhance our understanding of both complex disease processes and acute responses to biologic agents.

As another example of a clinical application, one could construct a Biomolecular Database made from a large group of healthy people, with the goal of finding people who are naturally resistant to certain germs, or who respond in certain ways to prescription drugs. One could study the selection of DNA strands from this mixture that have a specific sequence change in a specific gene that is known to change a person's resistance to germs or their response to drugs. Once these strands are isolated, the information tags would be examined to identify the people who have that change in their genes. This would be an extremely useful way to identify people who could have a bad reaction to a drug commonly used to treat disease. It could also be very useful in discovering people who are resistant to naturally occurring diseases or those caused by agents released during germ warfare.

4.1.3. Streamlining Identification of Susceptibility Genes

In terms of high-throughput screening of candidate genes to optimize genetic association analysis for complex diseases, consider the problem of genomic characterization of those individuals who first were infected by a biological agent, and then died. The death may often have been due to complications involving additional "complex" diseases, such as heart disease. Hence, mortality resulting from a chemical or biological agent attack may often have been due to complications involving a preexisting disease such as heart disease. Mortality can thus often only be predicted by considering both the individual's susceptibility to that agent, as well as that to various preexisting "complex" diseases. For many "complex" diseases, susceptibility often depends on a number of single-nucleotide polymorphisms (SNPs) in the human genome. Research into the genetic causes of complex disease is currently very expensive, and progress is slow, and complex diseases are quite common, affecting large propor-

tions of the population. Delays in understanding the genetic basis of these diseases slow the development of improved treatments at a significant financial and human cost. In the last 2 to 3 years, there have been several large-scale efforts to identify single-nucleotide polymorphisms in the human genome. The SNP consortium, a non-profit foundation composed of the Wellcome Trust and eleven pharmaceutical and technological companies, has agreed to deposit all SNPs they discover to public databases such as dbSNP, the SNP database maintained by the National Center for Biotechnology Information (NCBI). The number of entries in this database has increased from several thousand to over two million within the last 3 years. This sudden increase in the number of polymorphic markers has completely overwhelmed current methods for SNP genotyping and high-throughput screening. It has also become apparent that the incidence of single-nucleotide polymorphisms varies widely from one region of the genome to another, and large numbers of SNPs must be screened to analyze each candidate gene. Even with unlimited funds and the capacity for genotyping, serious challenges to the family-based association screening would remain, because the individual screening of a large number of SNPs would quickly exhaust the amount of DNA that can be easily obtained from a single individual. This problem is compounded by the sample cost of preparing pools of DNA from multiple individuals by simple mixing: once samples are mixed they cannot be separated again, and leftover, pooled DNA is wasted. Indexing of a Biomolecular Database can be of significant assistance in this regard. Large numbers of different groups of individuals can be selected from the Biomolecular Database by logical queries on the information tags. These pools can be used for allelic frequency determinations, and any remaining DNA can be added back to the remaining database.

As another example of a clinical application, one could use Biomolecular Databases to help discover what genes are turned on in a specific tissue of the body. Genes that are needed in the brain may not be expressed in the muscles, and genes needed in the muscles may not be needed in the liver. For this reason, measuring what genes are turned on in a specific tissue can help us understand what the possible functions of those genes might be. Biomolecular Databases would provide increased efficiency for these approaches.

4.2. Further Applications

The applications described above could be of critical value to the United States in the event of a terrorist release of a biological or chemical agent, as in the following brief scenario. A biological agent is released by a terrorist group in an American city or another populated area. The city is evacuated, but it becomes necessary to traverse a potentially contaminated area, or to revisit a known contaminated area. Clearly, any personnel sent into this area,

even with protective gear, are at risk for infection. A Biomolecular Database query would be initiated to identify personnel who possess a known genetic variation that prevents or mitigates infection. The personnel actually sent into the contaminated area could then be selected from the list of genetically resistant individuals.

As an alternative anti-terrorist application, suppose a large population (e.g., that of a city) has been exposed to a given biological or chemical agent. It then becomes apparent that a subgroup of individuals require significantly more aggressive medical therapy to survive, but for logistical reasons such aggressive therapy cannot be provided to ALL exposed individuals. Stored DNA from resistant and susceptible individuals can be used to determine the status of specific groups of genetic markers as described in application *C* (markers are chosen based on biological and medical inferences). In this way, a series of markers diagnostic for increased susceptibility can be identified. This type of analysis is called class discovery, and has been applied to the treatment of breast cancer and leukemia, among other disorders. However, the use of Biomolecular Databases can greatly streamline this work. Once diagnostic markers have been identified, the techniques worked out in application 4.1.3 can identify individuals in need of more aggressive care.

5. DISCUSSION AND CONCLUSIONS

We have described Biomolecular Databases constructed from DNA for rapid genetic analysis of large populations of individuals and complex diseases involving multiple genetic loci. They may improve on conventional methods in size of database and speed of search with the Biomolecular Databases system.

5.1. Comparison with Biomolecular Computing Methods for SAT Problems

As described above, these selection operations can be executed by the use of recombinant DNA operations, using logical processing methods developed in the field of DNA computing. The methods used in DNA computing to solve combinatorial search problems such as the Boolean satisfiability (SAT) problem have the disadvantage that they require a volume that scales exponentially with the size of the problem (i.e., the number of Boolean variables). This is because the search space of possible Boolean variable assignments scales exponentially. In contrast, logical queries are executed only on the information tags of the existing database, so the volume only scales linearly with the number of strands of the Biomolecular Database.

5.2. The Key Advantages of Biomolecular Databases

The key advantages of Biomolecular Databases appear to be:

a. *Bypassing of conventional impasses*: In particular, the avoidance of sequencing for conversion from DNA (genomic DNA and transcribed RNA) to digital media.

b. *Ultra-compact storage media*: The extreme compactness and portability of the storage media—a pedabit of information can be stored (with tenfold redundancy) in less than a few milligrams of dehydrated DNA, or when hydrated may be stored in a few milliliters of solution. A Biomolecular Database is capable of containing the DNA of a million individuals (6 pedabits of information) in a volume the size of a conventional test tube.

c. *Massive molecular parallelism*: Although one query may require a number of minutes, it is operated on vast numbers of data items (DNA strands), implying a processing power of vast molecular parallelism with at least a few hundred teraflops. The operations can operate in parallel on an entire population of DNA.

d. *Scalability*: The technology requires volume that scales linearly with the size of the database, and a query time that remains nearly constant up to extremely large database sizes.

e. *Limitations*: The Biomolecular Database technology is limited to applications of a biological nature (where the data are DNA or easily convertible to DNA), and the operations are limited to logical queries in the Biomolecular Database, associative searches, and some essential database operations. It is not intended that the technology compete in any direct way with conventional high-performance computers. Instead, the objective is to bypass conventional bioinformatics methodology by processing biological material (genomic DNA and transcribed RNA) in "wet" media, rather than digital media.

5.3. Scalability of Biomolecular Databases Systems

The key parameters of Biomolecular Database are: (a) N = the number of distinct elements of the Biomolecular Database, (b) v = the number of variables (each ranging over 10 possible values) used in queries, and (c) k = the number of individuals in application studies.

For our practical genomic applications of Biomolecular Databases to be fully realized in practice: (i) the database size N should grow to extremely large values (with a long-term goal of approximately 10^{15}), but (ii) for these applications the number of variables v needs only to grow to moderately small constant values (with a long-term goal of approximately $v = 14$), since for the genomic applications considered only a limited number of values need to be recorded in the information tag per database element. The relative difficulty of obtaining

human genomic material limits the number of individuals k in possible studies to approximately 1000, which is the size of the largest genomic database we are aware of for which one can legally obtain samples of genomic DNA. However, this figure of $k = 1000$ is by no means a limit on the capability of Biomolecular Database technology. In particular, these genomic databases are quickly growing in size, and it may be projected to grow by a number of multiples in a few years. Furthermore, military sources of human genomic DNA may be obtainable, providing alternate routes to obtain the samples of genomic DNA required in large-scale studies.

6. ACKNOWLEDGEMENTS

This work has been supported by grants from NSF to J.H.R. (EIA-00-86015, EIA-0218376, EIA-0218359) and DARPA/AFSOR to J.H.R. (F30602-01-2-0561).

7. REFERENCES

1. Adleman L. 1994. Molecular computation of solution to combinatorial problems. *Science* **266**:1021–27.
2. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, Staudt LM. 2000. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**(6769):503–511.
3. Alon U, Barkai N, Notterman DA, Gish K, Ybarra S, Mack D, Levine AJ. 1999. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc Natl Acad Sci USA* **96**(12):6745–6750.
4. Arnheim, N, Li HH, Cui XF. 1990. PCR analysis of DNA sequences in single cells: single sperm gene mapping and genetic disease diagnosis. *Genomics* **8**:415–419.
5. Bach E, Condon A, Glaser E, Tanguay C. 1996. Improved models and algorithms for DNA computation. In *Proc. 11th annual IEEE conference on computational complexity*, *J Comp Syst Sci*, pp. 290–299.
6. Bancroft C, Bowler T, Bloom B, Clelland CT. 2001. Long-term storage of information in DNA. *Science* **293**(5536):1763–1765.
7. Baum EB. 1995. How to build an associative memory vastly larger than the brain. *Science* **268**:583–585.
8. Baum EB. 1996. DNA sequences useful for computation. In *DNA sequences useful for computation*, *Proc. 2nd DIMACS workshop on DNA-based computing*, Princeton. AMS DIMACS Series, **44**:235–241. Ed. LF Landweber, E Baum. See (http://www.neci.nj.nec.com/homepages/eric/seq.ps.)

9. Box, GEP. 1978. *Statistics for experimenters: an introduction to design, data analysis, and model building*. Wiley, New York.
10. Box, GEP. 1987. *Empirical model-building and response surfaces*. Wiley, New York.
11. Braich RS, Chelyapov N, Johnson C, Rothemund PWK, Adleman L. 2002. Solution of a 20-variable 3-SAT problem on a DNA computer. *Science* **296**(5567):499–502.
12. Cantor CR, Smith CL, Mathew MK. 1988. Pulsed-field gel electrophoresis of very large DNA molecules. *Annu Rev Biophys Biophys Chem* **17**:287–304.
13. Chen CJ, Deaton R, Wang Y.. 2003. A DNA-based memory with in vitro learning and associative recall, *Proc. 9th annual meeting on DNA-based computers*, pp. 127–136.
14. Clayton SJ, Scott FM, Walker J, Callaghan K, Haque K, Liloglou T, Xinarianos G, Shawcross S, Ceuppens P, Field JK, Fox JC. 2000. K-ras point mutation detection in lung cancer: comparison of two approaches to somatic mutation detection using ARMS allele-specific amplification. *Clin Chem* **46**:1929–1938.
15. Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MA. 1993. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **261**(5123):921–923.
16. Corder EH, Saunders AM, Risch NJ, Strittmatter WJ, Schmechel DE, Gaskell Jr PC, Rimmler JB, Locke PA, Conneally PM, Schmader KE, Small GW, Roses AD, Haines JL, Pericak-Vance MA. 1994. Protective effect of apolipoprotein e type 2 allele for late onset Alzheimer disease. *Nature Genet* **7**:180–184.
17. Cukras AR, Faulhammer D, Lipton, RJ, Landweber LF. 2000. Molecular computation: RNA solutions to chess problems, *Proc Natl Acad Sci USA* **97**:1385–1389.
18. Deaton R. Murphy RE, Rose JA, Garzon M, Franceschetti DR, Stevens Jr SE. 1997. A DNA-based implementation of an evolutionary search for good encodings for DNA computation. In *Proc. IEEE Conference on Evolutionary Computation*, ICEC-97, pp. 267–271.
19. Deaton R, Garzon M, Rose JA, Franceschetti DR, Murphy RC, Stevens Jr SE. 1998. Reliability and efficiency of a DNA-based computation. *Phys Rev Lett* **80**:417–420.
20. Deaton R, Murphy RC, Garzon M, Franceschetti DR, Stevens Jr SE. 1999. Good encodings for DNA-based solutions to combinatorial problems. In *Proc. DNA-based computers*, II: DIMACS Workshop 10–12 June. Ed LF Landweber and EB Baum. DIMACS Series in Discrete Mathematics and Theoretical Computer Science **44**:247–258.
21. Deming SN. 1987. *Experimental design: a chemometric approach*. Elsevier, New York.
22. DeRisi J, van den Hazel B, Marc P, Balzi E, Brown P, Jacq C, Goffeau A. 2000. Genome microarray analysis of transcriptional activation in multidrug resistance yeast mutants. *FEBS Lett* **470**(2):156–160.
23. Faulhammer D, Cukras AR, Lipton RJ, Landweber. 2000. Molecular computation: RNA solutions to chess problems. *Proc Natl Acad Sci USA* 97:1385–1389.
24. Frutos AG, Thiel AJ, Condon AE, Smith LM, Corn RM. 1997. DNA computing at surfaces: 4 base mismatch word design. In *Proc. 3rd DIMACS meeting on DNA-based computers*, University of Pennsylvania, Philadelphia, June.
25. Garzon M, Deaton R, Neathery P, Murphy RC, Franceschetti DR, Stevens Jr SE. 1997. On the Encoding Problem for DNA Computing. In *Proc. 3rd DIMACS meeting on DNA-based computers*, University of Pennsylvania, Philadelphia, June.
26. Garzon M, Neel A, Bobba K. 2004. Efficiency and reliability of semantic retrieval in DNA-based memories. In *DNA computing, 9th international workshop on DNA-based computers*. Ed. J Chen, JH Reif. *Lect Notes Comput Sci* **2943**:157–169.
27. Gehani A, and Reif JH. 1999. Microflow bio-molecular computation. In *Proc. 4th DIMACS workshop on DNA-based computers*, University of Pennsylvania, June 1998. Series in Discrete Mathematics and Theoretical Computer Science. Ed. H Rubin. American Mathematical Society, Providence, RI. Also appeared in special issue of *Biosystems: J Biol Inform Processing Sci* 52:(1–3):197–216.

28. Gehani A, LaBean TH, Reif JH. 2000. DNA-based cryptography. In *5th DIMACS workshop on DNA-based computers*, MIT, June 1999. Series in Discrete Mathematics and Theoretical Computer Science. Ed. E Winfree. American Mathematical Society, Providence, RI.
29. Gray JM, Frutos TG, Berman AM, Condon AE, Lagally MG, Smith LM, Corn RM. 1996. *Reducing errors in DNA computing by appropriate word design*. Draft paper, University of Wisconsin, Department of Chemistry, October 9.
30. Hartemink A, Gifford D, Khodor J. 1998. Automated constraint-based nucleotide sequence selection for DNA computation, In *Proc. 4th DIMACS workshop on DNA-based computers*, University of Pennsylvania, June 1998.
31. Helene C, Thuong NT. 1991. Design of bifunctional oligonucleotide intercalator conjugates as inhibitors of gene expression. *Nucleic Acids Symp Ser* **24**:133–137.
32. Jonoska N, Karl SA. 1997. Ligation experiments in computing with DNA. In *Proc. IEEE Conference on Evolutionary Computation*, ICEC-97, pp. 261–265.
33. Kaplan P, Cecchi G, Libchaber A. 1996. DNA-based molecular computation: template–template interactions in PCR. In *Proc. 2nd DIMACS workshop on DNA-based computing*. Ed. LF Landweber and EB Baum. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, **44**:94–102.
34. Kashiwamura S, Yamamoto M, Kameda A, Shiba T, Ohuchi A. 2003. Hierarchical DNA memory based on nested PCR. In *Proc. 8th DIMACS workshop on DNA-based computing*, Sapporo, Japan, June 10–13. Ed. M Hagiya, A Ohuchi. *Lect Notes Comput Sci* **2568**:112–123.
35. Li HH, Cui XF, Arnheim N. 1990. Analysis of DNA sequences in individual gametes: application to human genetic mapping. *Prog Clin Biol Res* **340C**:207–211.
36. Lipton RJ. 1996. DNA computations can have global memory. In *Proc. 2nd DIMACS workshop on DNA-based computing*. Ed. LF Landweber and EB Baum. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, **44**:259–266.
37. Liu Q, Liman W. Frutos AG, Condon AE, Corn RM, Smith LM. 2000. DNA Computing on surfaces. *Nature* **403**:175–179.
38. Lizardi P, Huang X, Zhu Z, Bray-Ward P, Thomas DC, Ward DC. 1998. Mutant detection and single molecule counting using isothermal rolling circle replication. *Nature Genet* **19**:225–232.
39. Mir KU. 1996. A restricted genetic alphabet for DNA computing. In *Proc. 2nd DIMACS workshop on DNA-based computing*. Ed. LF Landweber and EB Baum. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, **44**:243–246.
40. Niculescu AB, Segal DS, Kuczenski R, Barrett T, Hauger RL, Kelsoe JR. 2000. Identifying a series of candidate genes for mania and psychosis: a convergent functional genomics approach. *Physiol Genomics* **4**(1):83–91.
41. Olson MV. 1989. Separation of large DNA molecules by pulsed-field gel electrophoresis: a review of the basic phenomenology. *J Chromatogr* **470**:377–383.
42. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D. 2000. Molecular portraits of human breast tumours. *Nature* **406**(6797):747–752.
43. Pieles U, Englisch U. 1989. Psoralen covalently linked to oligodeoxyribonucleotides: synthesis, sequence specific recognition of DNA and photo-cross-linking to pyrimidine residues of DNA. *Nucleic Acids Res* **17**:285–99.
44. Pirrung MC. 1995. Combinatorial libraries: chemistry meets Darwin. *Chemtracts Org Chem* **8**:5.
45. Pirrung MC. 1997. Spatially-addressable combinatorial libraries. *Chem Rev* **97**:473.
46. Pirrung MC, Chau JH-L, Chen J. 1996. Indexed combinatorial libraries: non-oligomeric chemical diversity for the discovery of novel enzyme inhibitors. In *Combinatorial chemistry: a high-tech search for new drug candidates*, pp. 191–206. Ed. SR Wilson, R Murphy. John Wiley & Sons, New York.

47. Pirrung MC, Connors RV, Montague-Smith MP, Odenbaugh AL, Walcott NG, Tollett JJ. 2000. The arrayed primer-extension method for DNA microchip analysis: molecular computation of satisfaction problems. *J Am Chem Soc* **122**:1873.
48. Pirrung MC, Zhao X, Harris SV. 2001. A universal, photocleavable, DNA base: nitropiperonyl 2'-deoxyriboside (dP*). *J Org Chem* **66**:2067.
49. Quillent C, Oberlin E, Braun J, Rousset D, Gonzalez-Canali G, Metais P, Montagnier L, Virelizier JL, Arenzana-Seisdedos F, Beretta A. 1998. HIV-1-resistance phenotype conferred by combination of two separate inherited mutations of CCR5 gene. *Lancet* **351**(9095):14–18.
50. Reif, J.H. 1998. Paradigms for biomolecular computation. Paper presented at 1st international conference on unconventional models of computation, Auckland, New Zealand, January. In *Unconventional models of computation*, pp. 72–93. Ed. CS Calude, J Casti, MJ Dinneen. Springer, New York.
51. Reif JH. 1999. Parallel Molecular Computation: Models and Simulations. In *Proc. 7th annual ACM symposium on parallel algorithms and architectures* (SPAA'95), Santa Barbara, CA, July 1995, pp. 213–223. Published in *Algorithmica*, special issue on *Comput Biol* **25**(2):142–176.
52. Reif JH. 2002. The emergence of the discipline of biomolecular computation in the US. Invited paper presented in a special issue on *Biomolecular Computing, New Generation Computing*, ed. M Hagiya, M Yamamura, T Head, **20**(3):217–236.
53. Reif, JH. 2002. Perspectives: successes and challenges. *Science* **296**:478–479.
54. Reif JH. LaBean TH. 2001. Computationally inspired biotechnologies: improved dna synthesis and associative search using error-correcting codes and vector-quantization, In *Proc. 6th DIMACS workshop on DNA-based computers*, Leiden, The Netherlands, June 13–17, 2000. *Lect Notes Comput Sci* **2054**:145–172.
55. Reif JH, LaBean TH, Pirrung M, Rana VS, Guo B, Kingsford C, Wickham GS. 2002. Experimental construction of very large-scale DNA databases with associative search capability. In *Proc. 7th DIMACS workshop on DNA-based computers*, Tampa, FL, June 10–13, 2001. *Lect Notes Comput Sci* **2340**:231–247.
56. Risch N, Merikangas K. 1996. The future of genetic studies of complex human disorders. *Science* **273**(5281):1516–1517.
57. Robinson BH, Seeman NC. 1987. The design of a biochip: a self-assembling molecular-scale memory device. *Prot Eng* 1:295–300.
58. Roweis S, Winfree E, Burgoyne R, Chelyapov NV, Goodman MF, Rothemund PWK, Adleman LM. 1998. A sticker-based model for DNA computation, *J Comput Biol* 5:615–629.
59. Sakakibara Y, Suyama A. 2000. Intelligent DNA chips: logical operation of gene expression profiles on DNA computers. *Genome Informatics* **11**:33–42.
60. Suyama A, Nishida N, Kurata K, Omagari K. 2000. Gene expression analysis by DNA computing. *Curr Comput Mol Biol* **30**:12–13.
61. Szatmari I, Aradi J. 2001. Telomeric repeat amplification, without shortening or lengthening of the telomerase products: a method to analyze the processivity of telomerase enzyme. *Nucleic Acids Res* **29**:E3.
62. Taylor GR, Logan WP. 1995. The polymerase chain reaction: new variations on an old theme. *Curr Opin Biotechnol* **6**:24–29.
63. Taylor GR, Robinson P. 1998. The polymerase chain reaction: from functional genomics to high-school practical classes. *Curr Opin Biotechnol* **9**:35–42.
64. Wellinger RE, Lucchini R, Dammann R, Sogo JM. 1999. In vivo mapping of nucleosomes using psoralen-DNA crosslinking and primer extension. *Methods Mol Biol* **119**:161–173.
65. Winfree E. 1998. Whiplash PCR for O(1) computing. In *Proc. 4th DIMACS workshop on DNA-based computers*, University of Pennsylvania, June 1998.
66. Wood DH. 1998. Applying error-correcting codes to DNA computing. In *Proc. 4th DIMACS workshop on DNA-based computers*, University of Pennsylvania, June 1998, pp. 109–110.

67. Zhang L, Cui X, Schmitt K, Hubert R, Navidi W, Arnheim N. 1992. Whole genome amplification from a single cell: implications for genetic analysis. *Proc Natl Acad Sci USA* **89**:5847–5851.
68. Zhao R, Gish K, Murphy M, Yin Y, Notterman D, Hoffman WH, Tom E, Mack DH, Levine AJ. 2000. Analysis of p53-regulated gene expression patterns using oligonucleotide arrays. *Genes Dev* **14**(8):981–993.

4

TISSUE ENGINEERING: MULTISCALED REPRESENTATION OF TISSUE ARCHITECTURE AND FUNCTION

Mohammad R. Kaazempur-Mofrad
Department of Bioengineering, University of California at Berkeley

Eli J. Weinberg and Jeffrey T. Borenstein
Draper Laboratory, Cambridge, Massachusetts

Joseph P. Vacanti
Harvard Medical School, Cambridge, Massachusetts

Replacement of a lost or failed organ is a long-standing problem in medicine. Research in the field of tissue engineering is progressing rapidly towards the replacement of numerous organs. Each organ is a complex system, and analysis of an organ requires understanding of phenomena over a range of length scale. This chapter provides an overview of the multiscale analysis currently used and under development in the field of tissue engineering.

1. INTRODUCTION

The loss or failure of a functioning organ can sometimes only be treated by replacement of the entire organ. As long as humans have practiced medicine, we have used all available technologies to improve our methods of organ replacement. Ancient cultures whose most advanced materials were wood and leather used these to fabricate leg prostheses. Developments in fine metalworking led to

Address correspondence to: Mohammad Reza Kaazempur-Mofrad, Department of Bioengineering, University of California, Berkeley, 483 Evans Hall #1762, Berkeley, CA 94720 (mofrad@berkeley.edu).

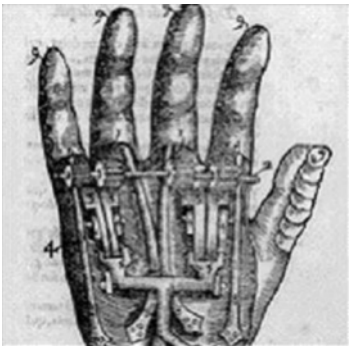


Figure 1. Iron hand design by Ambroise Paré, c. 1550 (83).

Figure 2. Mechanical heart valve manufactured by Edwards Lifesciences (84).

metal prostheses, beginning around 200 BCE in Rome with an iron hand fabricated for General Marcus Sergius (19). In the 1500s, the developing field of clockmaking introduced springs and gears to prosthesis makers, and in the 1550s surgeon/barber Ambroise Paré incorporated these innovations into jointed leg, arm, and hand prostheses (61). Figure 1 is an illustration of the prosthetic hand that Paré built.

The advent of anesthesia in the 1840s and antiseptic techniques in the 1860s allowed surgeons to consider repair and replacement of internal organs, and surgeons began to attempt organ transplants into humans from other humans, cadavers, or animals. In the last century, new materials led to the creation of many synthetic organ replacements: artificial knees, hips, and heart valves (Figure 2) are now commonplace. Advances in mechatronics have given us much-publi-

cized whole-heart prostheses, from the Jarvik-7 in 1982 to the AbioCor in 2001 (47), and the less-publicized but much more common ventricular assist devices and hepatic assist devices (53).

While many methods of organ replacement have been developed and are used successfully every day, most of these techniques are imperfect, as they cannot permanently restore full organ function. Four major methods for organ replacement have been conceived: taking tissue from a donor (transplantation), moving tissue from one location on a patient to another (autografting), fabricating a synthetic organ (prosthesis), or growing living replacement tissue (synthesis). Transplantation of major organs has been spectacularly successful, but requires a donor who is a close genetic match to the patient. More than 65,000 people are on transplant waiting lists in the Unites States alone, and many will die before a suitable organ becomes available (76). Autografting is useful in replacing damaged skin (11) and nerves (58) but has limited application elsewhere. External (Figure 3) and dental prostheses have been widely successful, and implanted prostheses generally can only provide temporary function.

An implanted prosthesis, such as a hip replacement, may function for fifteen years or more before damaging the implant site or becoming sufficiently damaged to lead to failure, as is often seen in hip prostheses (80). Additionally, implantation of synthetic materials may cause clotting, calcification, or infection (74), and synthetic materials are unable to grow with a growing patient. An implant constructed from living tissue will not have these problems and, once implanted, can grow to operate seamlessly with the rest of the body.

As technological advances have always led to development of new organ-replacement techniques, modern technologies in cell biology and genetic manipulation allow now for the development of living tissue implants. This field, known as "tissue engineering," has already shown promising successes in a variety of physiological systems. The most widely researched and most successful engineered tissue thus far has probably been the effort to grow replacement skin for burn victims. A number of groups have produced "skin-equivalents" in the lab (3), and other researchers have had success regenerating skin at the site of the injury in human patients (79). The tissue engineering of heart valves has also been widely researched (30,70,73), with successful aortic valve replacement in sheep reported (70) (Figure 4). Promising progress has been reported in regeneration of nerves (26), and in fabrication of tissue-engineered bladders (16), stomach (54), trachea (42), cartilage (43), and other tissues. These tissues are all relatively simple, and researchers are now developing more complex tissue-engineered organs like the liver and kidney (47).

To construct a tissue-engineered organ, the researcher must understand the components of the organ and how they interact. All organs, regardless of complexity or location in the body, consist of four components: cells, scaffolds, signals, and nutrients. The cells are the familiar living building block of a tissue. They provide any active behavior or functionality of a tissue, such as the con-

Figure 3. Example of a modern prosthetic leg from Otto Bock Healthcare (85).

traction of a muscle, the filtration of a kidney, or the metabolic behavior of a liver. Cells exist connected by a scaffold of solid and fluid. The solid is mostly extracellular matrix (ECM) and the fluid is mainly blood (of course, blood contains cells that have their own functions), while other solids and fluids exist in smaller quantities. In some tissues the volume and mass of the scaffold greatly exceeds that of the cells; in structural tissues such as blood vessels, heart valves, and bone, the scaffold provides the main functionality and the cells mostly serve to maintain the scaffold. Signals are molecules that communicate with the cells, and that the cells use to communicate with each other. Some signals are soluble and are transported between cells by fluid, and other signals are insoluble and may be attached to the surface of a cell or the scaffold. Nutrients are molecules needed to sustain the cell or are produced by the cell. The circulatory system exists to transport nutrients, especially oxygen, to and from all of the cells in the body.

Figure 4. Tissue-engineered valve conduit explanted after 5 weeks in vitro (70).

For a tissue-engineered organ to perform, the cells in the organ must be interacting correctly with the other components. The tissue engineer must first understand these interactions, and then design the organ so that the interactions can correctly and optimally take place. The interactions can take place at a wide range of length scales: nutrients and signals affect cells at a submicron level, cells interact with each other and with the surrounding scaffold at the micron level, and cells can react to mechanical stimuli that occur at the millimeter-to-centimeter range (compare with Part III, chapter 2.1, by Huang, Sultan, and Ingber, this volume). Designing and fabricating an organ requires a knowledge of the diffusive and convective transport that puts the cells in their desired locations and circulates the signals and nutrients; transport can take place from the micron to centimeter length scales. The final interaction of concern is that between the implant and the rest of the body, which may have effects at the centimeter or even meter scale.

Creation of a functional tissue-engineered organ requires an understanding of numerous behavior at length scales from submicron up to a meter. This chapter reviews some of the experimental and theoretical techniques being used over the range of scales.

2. TISSUE-ENGINEERING INVESTIGATIONS AT VARIOUS LENGTH SCALES

This chapter surveys the field of tissue engineering, giving examples of relevant physical behaviors. The theoretical and experimental tools used to in-

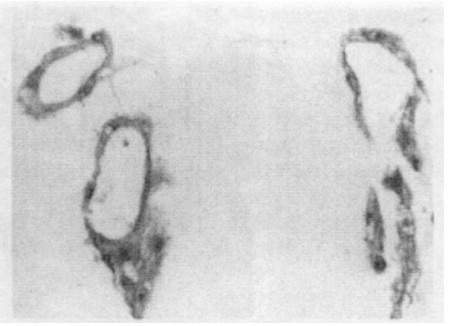


Figure 5. Hepatocytes form bile duct-like tubules when appropriate growth factors are applied (9).

vestigate these behaviors are discussed, starting with the smallest scales and proceeding to the larger scales.

2.1. Molecular (Submicron) Scale

2.1.1. Cell Signaling

Cells have receptors on their surfaces that interact with various extracellular molecular signals (see also Part III, chapters 2.2 (by Subramanian and Narang) and 2.3 (by Goentoro and Shvartsman), this volume). These signals can be flowing in fluid outside the cell (soluble signals) or can be attached to extracellular matrix or a substrate (insoluble signals). Some of these molecules have been isolated and identified and are now widely available for researchers to use in controlling their cell cultures.

A class of signals that is particularly useful to the tissue engineer is the family of molecules known as growth factors. This family of factors includes platelet-derived growth factor (PDGF), epidermal growth factor (EGF), vascular endothelial growth factor (VEGF), keratinocyte growth factor (KGF), and many others. As the name implies, growth factors affect the growth behavior of cells. The presence of a growth factor may initiate cell proliferation and cell organization into structures, and any growth factor will affect different cells in different ways. The effect of growth factors on a cell type is examined in vitro by culturing the cell in a dish or on a scaffold and monitoring the cell behavior under varying amounts of growth factor. Block et al. grew liver cells (hepatocytes) on a collagen gel matrix, and showed that adding hepatocyte growth factor (HGF) causes the cells to organize in tubules akin to the bile ducts found in the liver (9)

(see Figure 5). To investigate the effect of VEGF-B, Silvestre et al. implanted collagen matrices into live rats and injected the growth factor into some of the matrices (71). The injections promoted the formation of new blood vessels, known as angiogenesis, in the matrices.

One of the molecular interactions that is most important to implants is the immune response (see also this volume, Part III, section 4). The immune response is what rejects foreign transplants into the body, and the trigger for this response is a set of cell-surface molecules that the body recognizes as being foreign. While in theory engineered tissues should have the native set of surface molecules and not be subject to an immune response, experiments have shown that even autograft tissues can produce unexpected immune reactions (28).

Another large class of molecules is cell adhesion molecules (CAMs). These are responsible for cells binding to each other, to extracellular matrix, and to any other surface. There has been a large amount of research into CAMs, so that we now know which CAMs are present on the surface of each cell and what extracellular components each CAM binds (62), information vital to producing tissues that are strongly bound together.

Many other families of molecules involved in cell signaling are known to exist, and there are likely signaling molecules that have not yet been discovered. Recent research shows that cell signaling acts in cascades, where signals introduced at different times and in different combinations have widely varying effects on a cell's development. While our understanding of cell signaling is rapidly improving, it will be many years before all cell signaling pathways are known.

2.1.2. Submicron Solid Mechanics

A theoretical model for the solid deformation of a biological tissue is based on the microstructure of that tissue. Until recently, researchers could only examine the bulk behavior of tissues. In the past decade such methods as optical tweezers and scanning-force microscopy have been developed that allow investigation of tissue behavior at the molecular level (12). New understanding of molecular behavior allows for improved models of bulk behavior.

Investigations of molecular behavior are now common. Evans and Ritchie used atomic force microscopy (AFM) to investigate the strength of molecular adhesion bonds (Figure 6), and also developed computational models to predict this behavior (18) (Figure 7).

Molecular experiments are useful in increasing our understanding of molecular interactions, but perhaps more important, molecular models can now be scaled up to predict bulk mechanical behavior of tissues. Kwan and Woo created a constitutive model for aligned collagenous tissue by taking a simple stress–strain model for a single collagen fiber and calculating the effect of many of

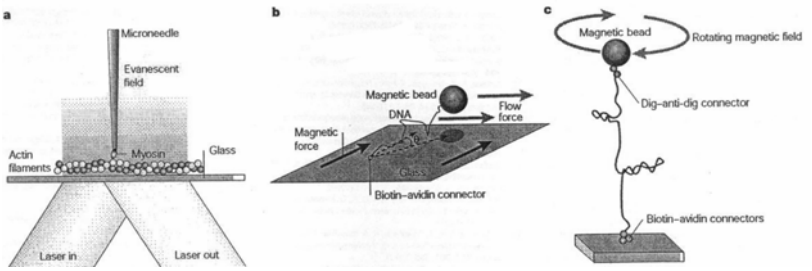


Figure 6. Diagram of atomic force microscopy used to measure molecular bonds (12).

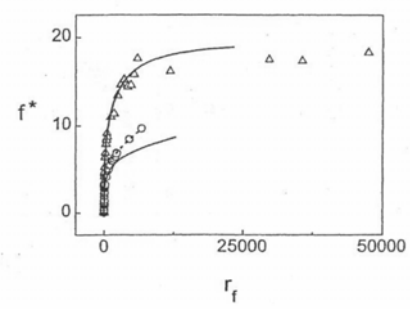


Figure 7. Experimental (circles and triangles) and theoretical (curves) force versus distance data for molecular adhesion measurements (18).

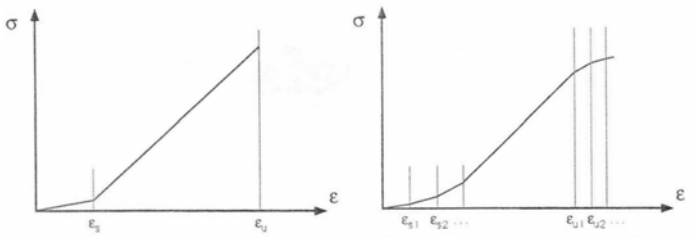


Figure 8. Stress–strain curves for a single collagen fiber (left) and for a network of collagen fibers (right) (45).

these fibers linked together to obtain a constitutive model of the bulk material (45) (Figure 8).

Many models use statistical models, Gaussian or Langevin, to represent the mechanics of a single molecule, and then consider the behavior of a unit cell made of a small group of molecules to determine the bulk behavior (Figure 9). Bischoff et al. used a Langevin model of a single fiber and an eight-chain unit

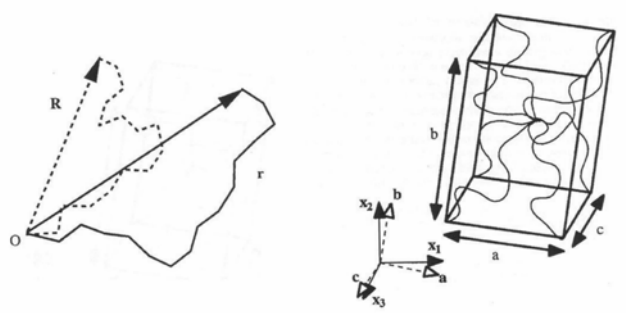


Figure 9. Diagram of single fiber (left) and a unit cell consisting of multiple fibers (right) (7).

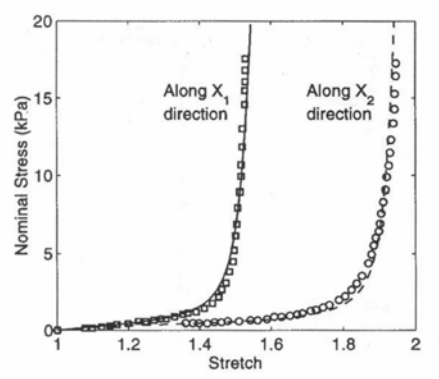


Figure 10. Experimental (box and circle) and theoretical (solid and dotted curve) results for stretching of rabbit skin, showing excellent agreement between model and experiment (7).

cell model to create a continuum model that accurately predicts behavior of skin (7) (Figure 10).

2.2. Micron Scale

2.2.1. Cell–Cell Interactions: Co-Culture

In nature, cells very rarely live exclusively with cells of their own type. Cells of different types interact with each other, and the right cell types must be combined to form a functioning tissue. Even as the mechanisms by which cells

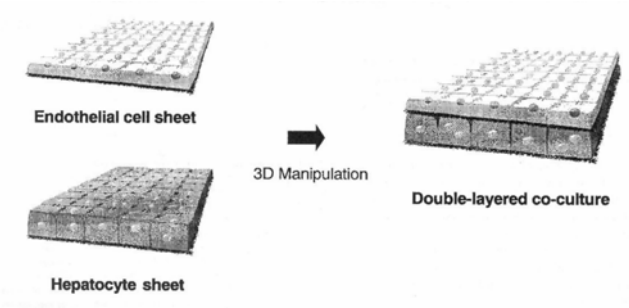


Figure 11. Harimoto's scheme for multilayer co-culture (27).

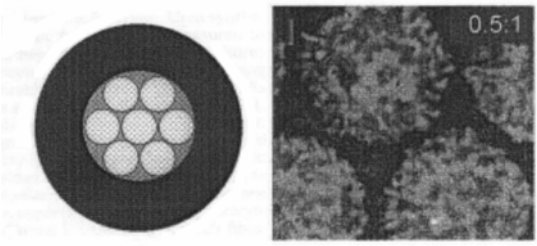


Figure 12. Bhatia's scheme for co-culture, with geometrically distinct areas on the same layer (5).

types interact with each other remain largely unknown, researchers can still examine the effects of these interactions.

The cell that provides most of the liver's metabolic function is the hepatocyte, but hepatocytes grown on their own lose function after a few weeks. Harimoto et al. developed techniques for growing hepatocytes layered with endothelial cells, and this improved hepatocyte function compared to hepatocytes alone (27) (Figure 11).

Bhatia et al. performed similar experiments in microdevices and showed that epithelial cells also enhance hepatocyte function and survival (6) (Figure 12), and also showed that the effect of hepatocyte–endothelial cell co-culture varies depending on how many of each cell type are included and how much contact area there is between the two cell types (5).

Other experiments have had similar results with different cell types, both inside and outside of the body (8), telling us that the fabrication of functional living tissue will require control of numerous cell types.

2.2.2. Cell–Scaffold Interactions

Much of the body is made up of extracellular matrix (ECM), which is a network of fibrous proteins produced by fibroblasts and smooth muscle cells. The ECM is not inert, as it has a significant effect on the cells growing on it. In tissue engineering, researchers need to understand how the ECM affects the cells, and also how growing the cells on a different material, such as a biodegradable polymer or a culture dish, will be different.

To examine the effects of ECM on cell behavior, different components of the ECM can be isolated and cell growth experiments can be run while adding the different components. Mooney et al. did this with hepatocytes and found that these cells do not normally proliferate in a dish, but will if such ECM components as laminin, fibronectin, and collagen are added (60). Similar work was done by Gerlach et al. showing that hepatocyte function improved when grown on collagen (22). Extracellular matrix can have effects other than aiding cell proliferation: cells can also migrate over the ECM. Fibroblasts grown on three-dimensional collagen matrices are observed to sprout dendrites and pull themselves along the fibers of the ECM (24) (Figure 13).

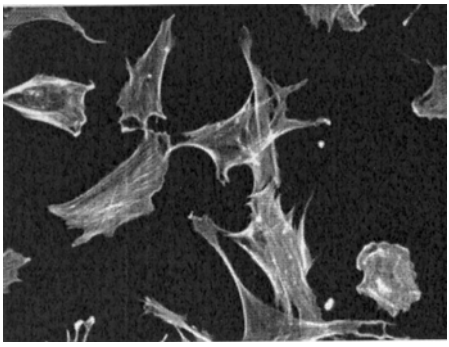


Figure 13. Fiboblasts extend dendrites along fibers of extracellular matrix (23).

In addition to understanding a cell's response to contact with ECM, it is important for tissue engineering purposes to examine cellular responses to other materials, since there are many synthetic materials that can help in constructing an organ. This sort of experiment is performed by growing the cells in contact with the material in question, and the results can be interpreted visually or chemically. The behavior of the cells varies widely depending on what material and preparation methods are used (29).

2.3. Millimeter-Centimeter Scale

A living tissue is subject to continual stresses applied by the tissues around it and the fluids flowing throughout. These stresses have profound effects on the cells in the tissue. To control the behavior of these cells, a tissue engineer must often apply these forces in the lab, which requires both a knowledge of what stresses and strains a cell will respond to and how to apply them.

2.3.1. Effects of Solid Deformation on Cells

Cells living on an extracellular matrix are stressed and strained as the matrix is deformed. Often these deformations influence the behavior of the cell that a tissue engineer is trying to control, a process known as mechanotransduction. This is most evident in the fabrication of structural tissues such as cartilage, heart valves, and blood vessels.

Structural tissues are composed almost entirely of extracellular matrix components: heart valves and blood vessels are usually over 80% type I collagen in combination with a few other ECM components and a small percentage of cells. The cells, smooth muscle cells and fibroblasts, produce the ECM. For a tissue to have the correct mechanical properties, the ECM must be produced in adequate quantities and aligned in a certain direction. It is the forces applied by the ECM on the cells that control the cellular production of ECM.

Applying deformation to tissues and observing the results can elucidate the mechanotransduction process. McKnight and Frangos grew human vascular smooth muscle cells on a collagen matrix and applied cyclic uniaxial stretch at various strain rates (57). The smooth muscle cells produce aligned collagen only when subjected to strains and strain rates similar to those they would be subjected to physiologically. Chapman et al. performed similar work and found that physiological cyclic stretching inhibits cell growth (13). These works together suggest that applying physiologically appropriate stretching to a smooth muscle cell will keep the cell in a steady state where it can be used to produce ECM without the cell proliferating.

Articular cartilage, which bears the loads in all synovial joints, is also produced by cells, in this case chondrocytes. Physiologically, cartilage is rarely subjected to tension but is often compressed. Efforts in tissue engineering of cartilage have focused logically on the response of chondrocytes to compression. Sah et al. have shown that, while static compression inhibits protein production of chondrocytes in an explanted cartilage sample, dynamic compression promotes production of proteins in general and dynamic compression at various rates promotes production of specific proteins (68).

2.3.2. Modeling of Solid Deformation

Understanding the deformation of tissues requires constitutive models to describe the tissue's solid mechanics. Continuum models may be derived from a

molecular basis or strain-energy functions, and can be verified by comparing measurements taken from experiments subjecting tissues to well-defined stresses and strains.

Biological tissues are, in general, highly nonlinear and anisotropic. Many models exist that can cover a wide range of materials, for example, a model for any rubber-like or soft biological tissue (33). In other cases, a general constitutive model is modified and fit to a specific tissue. There are models specifically for mitral valve tissue (56) (Figure 14), aortic heart valves (51), skin (7), myo-

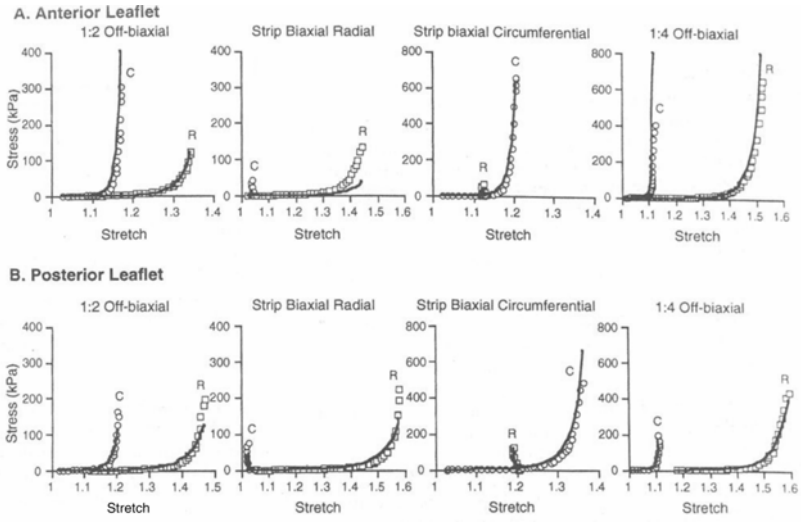


Figure 14. Experimental and theoretical curves for stretching mitral valve tissue in a variety of situations, showing a good match between experiment and model (56).

cardium (34), cartilage, tendons, blood vessel walls, and others. A common constitutive equation used for the representation of biological tissues is

$$t = -pI + 2W_1B + (W_\alpha/\alpha)F \cdot N \otimes N \cdot F^T, \qquad [1]$$

where t is the Cauchy stress, p is the Lagrange multiplier, B is the right Cauchy-Green deformation tensor, F is the deformation gradient tensor, W_1 and W_α are the derivatives of W with respect to α and the first invariant, respectively, and W is a strain-energy function that can be chosen depending on the tissue in question.

Constitutive models are matched to biological data by performing stress–strain experiments on excised tissue. Data may be taken for uniaxial tension (15), biaxial tension (35,52), compression, or any other arrangement the experi-

menter may choose. Biological tissues also often show time-dependent effects such as viscoelasticity (50). Predictions of in-vivo mechanical behavior can only be made using a model that has been verified in vitro.

2.3.3. *Effects of Fluid Mechanotransduction on Cells*

Just as stresses transmitted by the ECM can affect cell behavior, so can stresses transmitted by fluids flowing through an organ. Almost all physiological surfaces that face flow are coated with a layer of endothelial cells, so these are the only cells regularly affected by fluid mechanotransduction. The effect of flow on endothelial cells is a well-studied topic. Investigators have reported on hundreds of effects of flow mechanotransduction on endothelial cells (17,23,55) (Figure 15), so that we know the shears and pressures that give rise to different gene expressions, changes in intracellular makeup, physical deformation of the cell, electrical behavior of cells (4), and others. Importantly for tissue engineering, fluid mechanotransduction guides the formation of blood vessels and vasculature, and figures prominently in many vascular illnesses (23).

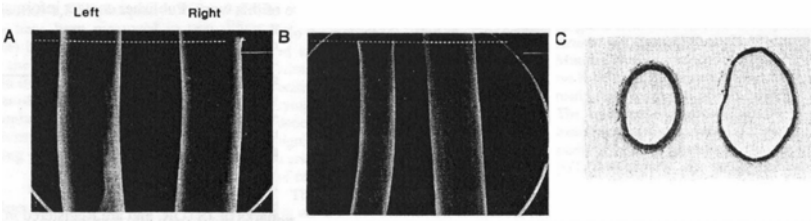


Figure 15. Response of vessels to decreased flow: (**A**) at the start of experiment, the two vessels are the same size; flow reduced in the left vessel; (**B**,**C**) two weeks later; note the reduction in size of the left vessel in response to flow reduction (46).

2.3.4. *Modeling of Blood Flow*

The modeling of blood flow through a single vessel can be complicated; while blood flow in almost all vessels is laminar, the highly non-Newtonian behavior of blood and distensibility of the vessel walls must be accounted for.

Blood's non-Newtonian behavior can be described by Casson's equation, which gives an apparent viscosity:

$$\mu = \left(k_1 + \frac{k_2}{\sqrt{U/d}} \right)^2, \qquad [2]$$

where μ is the apparent viscosity, U is the average velocity of the flow, d is the diameter of the vessel, and k_1 and k_2 are empirical coefficients that depend on the percentage of blood volume that is occupied by red blood cells, known as the hematocrit (69). This description of the apparent viscosity can be used in laminar flow correlations to predict blood flow behavior.

A single-vessel model for flow can be incorporated into a model for entire vasculature networks. In the context of design of microvascular networks for vascularized tissue engineering of vital organs, we (36) have developed a computational algorithm for simulation of blood flow in the microvascular networks. This algorithm takes into account the non-Newtonian blood rheology and its particulate nature, both of which are important in modeling the microcirculation. Pressure drop in each vessel is related to blood viscosity, which itself varies with vessel cross-sectional surface area and hematocrit (volume fraction of red blood cells).

In many cases, the distensibility of the vessel wall will also affect blood flow through a vessel. As with non-Newtonian behavior, wall deformation behavior can be predicted for a single vessel (20), and the single-vessel model can be incorporated into a network model. We have used such a model to model flows in distensible vasculature networks (77).

There are other widely varying models for flow through biological networks. Some models examine features general to all vasculature, such as the hypothesis that wall shear is constant throughout any biological vasculature system (39). Others develop models for specific tissues (31). The lungs are of particular interest, as the structure of the lungs and the flow through them is very different from other organs (20,44).

Modeling of network behavior requires understanding the structure of the networks. For decades researchers have injected polymers into an organ to obtain a cast of the vasculature, and taken measurements from these casts (33,40,59,86) (Figure 16).

Recent developments in imaging have allowed researchers to develop automated techniques for imaging of physiological flow networks (1) (Figure 17). These measurements of the vessel geometries and network connectivity can be tabulated in forms convenient for use in modeling of flows through the networks (Figure 18).

2.3.5. Modeling of Molecular Transport in Tissue

The transport of molecules through tissue occurs in three ways. Molecules are carried by the bulk flow of tissue in convective transport, the obvious example being blood carrying nutrients throughout the body (compare also with Part II, chapter 3, by Savage and West, this volume). Molecules are also transported by diffusion, such as nutrients diffusing from the blood, through blood vessel

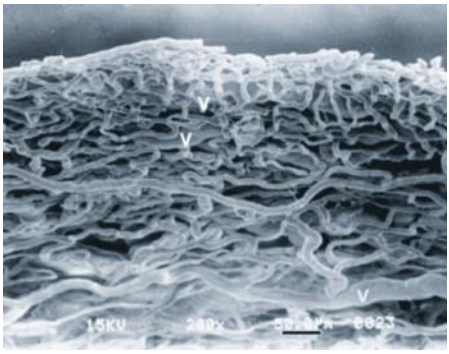


Figure 16. Corrosion cast of hamster eye vascular network (bar = 50 μm) (86).

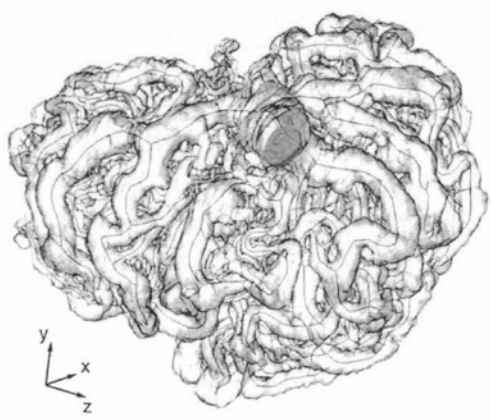


Figure 17. Three-dimensional regeneration of glomerulus capillary structure (1).

walls, into cells. Transport of molecules may also be affected by electrostatic interactions between the molecules and the extracellular matrix. To fully understand the transport of molecules in any system, the effects and interplay of these three factors in transport must be understood.

Analysis of molecular transport in physiological systems must include the effect of interactions between the molecules and cells. A simple case of nutrient transport by diffusion in the presence of cells can be modeled by considering each cell to be a point sink, and including a term for these point sink in the diffusion equation. In one dimension, the concentration of nutrient $L(x)$ is described by:

$$\frac{\partial L}{\partial t} = D\frac{\partial^2 L}{\partial x^2} - \rho\left(\frac{kL}{K_M + L}\right), \qquad [3]$$

where the first two terms constitute the standard diffusions equation and the third term represents the point sinks: ρ is cell density, x is the distance from the point sink, and K_M is the saturation constant and k the maximal uptake rate constant of a single cell (23).

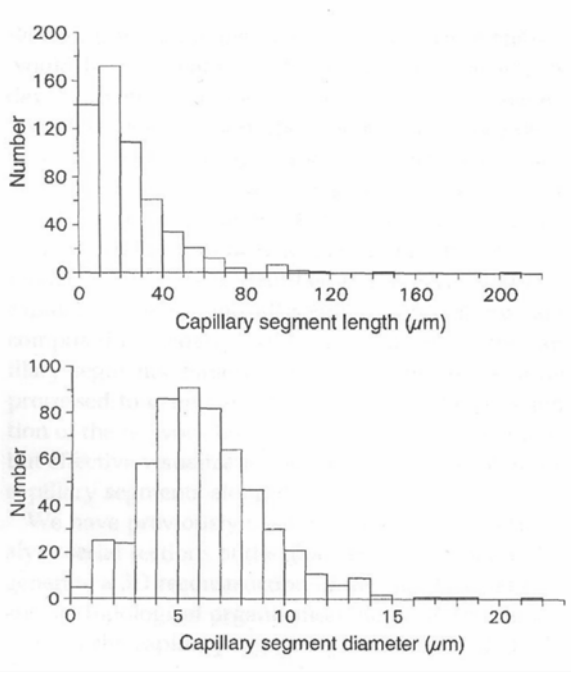


Figure 18. Capillary length and diameter data from glomerulus regeneration (1).

The distribution of signaling molecules is more difficult; where nutrients may be simply absorbed by the cell, growth factors and other regulatory molecules have more complicated interactions with the cell. In such cases the molecule–cell interaction can still be included in the diffusion equation, but a model must be provided for the molecular binding (25). Many models are available for describing molecular signaling interactions in terms of the number of binding sites available on a cell and the rate at which those sites will interact with an external molecule (48).

The effects of diffusion, convection, and electrostatic interactions can be combined by summing the fluxes due to each effect. For example, the one-dimensional flux N of a molecular solute through a tissue is given by

$$N = \Phi\left(-D\frac{\partial c}{\partial x} + \frac{z}{|z|}\mu cE\right) + WcU\,, \qquad [4]$$

where Φ is the porosity of the tissue, c is the concentration of the solute at position x, z is the valence of the solute (if the solute is charged), U is the average fluid velocity relative to the tissue, and E is the electric field within the matrix. D is the diffusivity of the solute, μ is the electrical mobility of the solute, and W is the hindrance factor for convection (25).

Oxygen is the most basic of nutrients, and investigations into oxygen transport are common. The response of cells to oxygen gradients can be finely measured (2), and numerical models can be used to design devices where oxygen transport to all cells is controlled (66).

2.4. Centimeter-Meter Scale

Biological phenomena typically encompass a range of time and length scales whose intrinsically complex interactions are critical to system function (37). For example, in the study of arterial disease, one needs to understand how the entire cardiovascular system responds to a variety of external factors that impact local flow characteristics. The fluid dynamic and solid stresses experienced by the vascular wall tissues lead to a cascade of critical biological events, which may contribute to disease progression. At the cellular level, these stresses produce deformations of the cytoskeletal network, the cell membrane, and the nuclear envelope, which lead, in turn, to conformational changes in individual proteins that elicit the biological response. Along with understanding what goes on within an organ, it is necessary to know how the organ interacts with the rest of the body.

2.5. Multiscaled Systems

To fabricate a functioning device, it is necessary to combine systems of different scales. This has been achieved in some experiments, most notably bioartificial liver systems. In such a system, hepatocytes are cultured on a scaffold and under continuous flow. The overall device provides liver function, and can be used to assist patients with liver disorders (Figure 19).

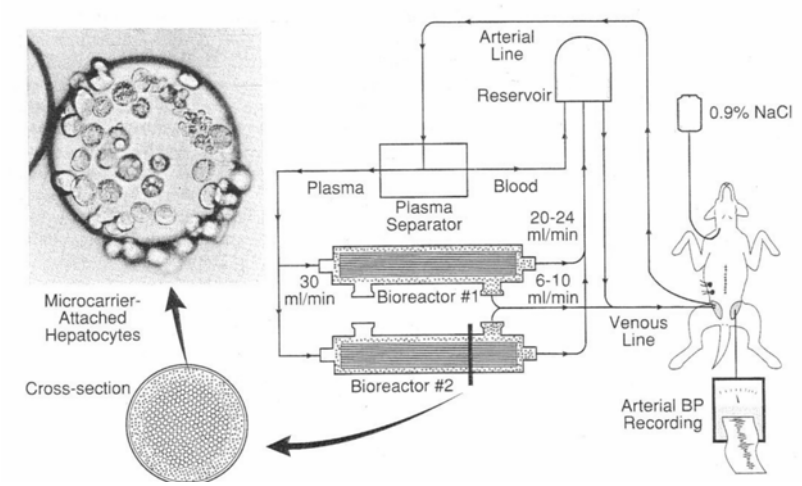


Figure 19. Schematic of liver assist device, animal experiment (67).

The success of these devices is possible only when the nutrient transport, cell attachment to scaffold, co-culture, and cell response to flow are controlled (64,65,67,75).

3. CONTINUING EFFORTS IN TISSUE ENGINEERING

Many groups continue to show success with engineering of simpler tissues such as skin, blood vessel walls, cartilage, and heart valves. These tissues are thin enough that they do not require vascularization, since diffusion can carry any necessary nutrients throughout the thickness of the tissue. Larger organs, though, are much thicker and require a vascular network to keep the cells alive and carry any products of the organ's activity. These systems are truly complex, and encompass the entire range of phenomena discussed here.

Two main approaches have been conceived for the fabrication of vascularized tissues. The first approach is to construct the entire vasculature. This is made possible by advances in micromachining (38) and the use of photolithography to create high-resolution patterns on a silicon wafer (21). Recently, methods have been developed for transferring the pattern from silicon to softer materials, such as polymers suitable for tissue engineering (78). Borenstein et al. have demonstrated the ability to construct complete blood vessel networks using these "soft lithography" techniques.

These vascular devices are fabricated by first creating the vascular pattern in a silicon wafer (Figure 20). Then polymer devices are replica-molded from the silicon master. The polymer devices are stacked to create three-dimensional

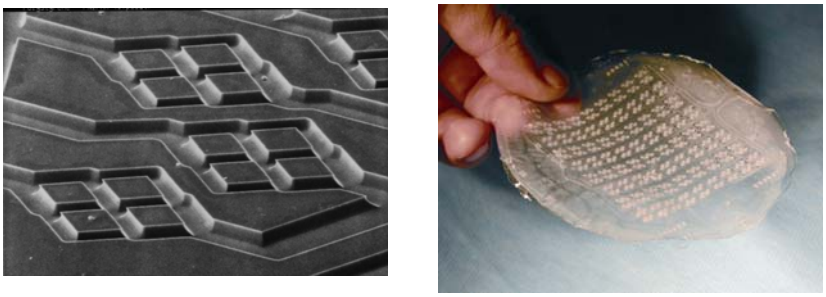


Figure 20. Silicon mold for a microfabricated capillary bed (left) and biodegradable vascular network (right) (10).

tissue and seeded with cells. Such a design can be implanted, so that when the polymer degrades the remaining cells form a completely natural tissue. Such devices have demonstrated organ function (49).

The design of a microfabricated network comes from numerical tools developed for the task (36,77). The flow of blood through a microfluidic device is simulated analogously to the calculation of currents in a resistor network. Each blood vessel has a fluidic resistance dependent on the vessel geometry and viscosity of the fluid, and the pressure drop from one end of the vessel to the other is related to the flow rate through the vessel by

$$\Delta P = QR, \qquad [5]$$

where ΔP is the pressure drop, Q is the flow rate, and R is the fluid resistance. The flow rates and pressures throughout the network are interdependent, so the equations for all the vessels are solved simultaneously as a matrix equation. Thus, if the geometry of a network is known, the flow behavior can be calculated. The geometry can be modified iteratively to create a network with the desired flow properties, or the system can be solved in reverse to determine the geometry from the flow conditions. If physiological flow conditions are specified, the resulting design will be a microfluidic network with physiological flow characteristics (see also this volume, Part IV, chapter 1, by Meinhart and Wereley).

The other approach is to create an experimental setting where the organ can self-assemble. In this approach, cell growth is controlled in a matrix material so that the cells form vasculature and organ features instead of designing and molding the features. Currently, there is much research aimed at understanding, controlling, and mimicking the natural process of vasculature formation, known as angiogenesis.

Numerous investigators have shown that angiogenesis is controlled by signaling proteins. For example, VEGF is known to promote recruitment of endothelial cells, while PDGF promotes vessel stabilization and maturity (81). Experiments show that including these and other growth factors in implanted matrices does promote vascularization of the matrix (82). As with micromachined approaches, progress continues towards the final goal of creating engineered vascularized tissue.

4. CONCLUSION

The field of tissue engineering has developed rapidly in the past decade. The generation of completely functional, implantable tissue-engineered organs in the near future will be an enormous achievement in science and medicine, but will require understanding and mastery of numerous processes affecting the cell over multiple length scales.

5. REFERENCES

1. Antiga L, Ene-Iordache B, Remuzzi G, Remuzzi A. 2001. Automatic generation of glomular capillary topological organization. *Microvasc Res* **62**:346–354.
2. Allen JW, Bhatia AN. 2003. Formation of steady-state oxygen gradients in vitro. *Biotechnol Bioeng* **82**:253–262.
3. Auger FA, Lopez Valle CA, Guignard R, Tremblay N, Noël B, Goulet F, Germain L. 1995. Skin equivalent produced with human collagen. *In Vitro Cell Dev Bio-Animal* **31**:432–439.
4. Barakat AI, Leaver EV, Pappone PA, Davies PF. 1999. A flow-activated chloride-selective membrane current in vascular endothelial cells. *Circ Res* **85**:820–828.
5. Bhatia SN, Balis UJ, Yarmush ML, Toner M. 1998. Microfabrication of hepatocyte/fibroblast co-cultures: role of homotypic cell interactions. *Biotechnol Prog* **14**:378–387.
6. Bhatia SN, Balis UJ, Yarmush ML, Toner M. 1999. Effect of cell–cell interactions in preservation of cellular phenotype. *FASEB J* **13**:1883–1900.
7. Bischoff JE, Arruda EM, Grosh K. 2000. Finite-element modeling of human skin using an isotropic nonlinear elastic constitutive model. *J Biomechan* **33**:645–652.
8. Bjerknes M, Cheng H, Ottaway CA. 1986. Dynamics of lymphocyte–endothelial interactions in vivo. *Science* **231**:402–405.
9 Block GD, Locker J, Bowen WC, Petersen BE, Katyal S, Strom SC, Riley T, Howard TA, Michalopoulos GK. 1996. Population expansion, clonal growth, and specific differentiation patterns in primary cultures of hepatocytes induced by HGF/SF, EGF and TGFalpha in a chemically defined (HGM) medium. *J Cell Biol* **132**(6):1133–1149.
10. Borenstein JT, Terai H, Terai H, King KR, Weinberg EJ, Kaazempur-Mofrad MR, Vacanti JP. 2002. Microfabrication technology for vascularized tissue engineering. *Biomed Microdevices* **4**(3):167–175.
11. Burke JF, Yannas IV, Bondoc CC, Quinby WC. 1974. Primary burn excision and immediate grafting: a method of shortening illness. *J Trauma* **14**:114–123.
12. Bustamante C, Macosko JC, Wuite GJL. 2000. Grabbing the cat by the tail: manipulating molecules one by one. *Nature* **1**:130–136.

13. Chapman GB, Durante W, Hellums D, Schafer AI. 2000. Physiological cyclic stretch causes cell cycle arrest in cultured vascular smooth muscle cells. *Am J Physiol Heart Circ Physiol* **278**:H748–H754.
14. Clark CB, McKnight NL, Frangos JA 2002. Strain and strain rate activation of G proteins in human endothelial cells. *Biochem Biophys Res Com* **299**:258–262.
15. Clark RE. 1973. Stress-strain characteristics of fresh and frozen human aortic and mitral leaflets and chordae tendineae. *J Thor Cardiovasc Surg* **2**(66):202–208.
16. Cross WR, Thomas DFM, Southgate J. 2003. Tissue engineering and stem cell research in urology. *JBU Intl* **92**:165–171.
17. Davies PF. 1995. Flow-mediated endothelial mechanotransduction. *Physiolog Rev* **75**(3):519–560.
18. Evans E, Ritchie K. 1997. Dynamic strength of molecular adhesion bonds. *Biophys J* **72**:1541–1544.
19. Friedmann LW. 1978. *The psychological rehabilitation of the amputee*. Charles C. Thomas, Springfield, IL
20. Fung YC. 1986. On pulmonary circulation [keynote lecture]. *J Biomech* **19**(6):465.
21. Gabriel KJ. 1998. Microelectrical systems (MEMS) tutorial. In *Proc. IEEE Test Conference*.
22 Gerlach JC, Schnoy N, Encke J, Smith MD, Muller C, Neuhaus P. 1995. Improved hepatocyte in vitro maintenance in a culture model with woven multicompartment capillary systems: electron microscopy studies. *Hepatol* **22**:546–552.
23. Gimbrone MA. 1999. Endothelial dysfunction, hemodynamic forces, and atherosclerosis. *Thromb Haemostasis* **82**(2):722–726.
24. Grinnel F, Ho C, Tamariz E, Lee DJ, Skuta G. 2003. Dendritic fibroblasts in three-dimensional collagen matrices. *Mol Biol Cell* **14**:384–395.
25. Grodzinsky AJ, Kamm RD, Lauffenburger DA. 2000. Quantitative aspects of tissue engineering: basic issues in kinetics, transport, and mechanics. In *Principles of tissue engineering*, pp. 193–207. Ed. RP Lanza, R Langer, J Vacanti. Academic Press, New York.
26. Hadlock T, Sundback C, Koka R, Hunter D, Cheney M, Vacanti J. 1999. A novel, biodegradable polymer conduit delivers neurotrophins and promotes nerve regeneration. *Laryngoscope* **109**(9):1412–1416.
27. Harimoto M, Yamato M, Okano T. 2003. Cell sheet engineering: intelligent polymer patterned surfaces for tissue engineered liver. *Macromol Symp* **195**:231–235.
28. Harlan DM, Karp CL, Matzinger P, Munn DH, Ransohoff RM, Metzger DW. 2002. Immunological concerns with bioengineering approaches. *Ann NY Acad Sci* **961**:323–330.
29. Hasirci V, Berthiaume F, Bondre SP, Gresser JD, Trantolo DJ, Toner M, Wise DL. 2001. Expression of liver-specific functions by rat hepatocytes seeded in treated PLGA biodegradable foams. *Tissue Eng* **7**(4):385–394.
30. Hoerstrup SP, Kadner A, Melnitchouk S, Trojan A, Eid K, Tracy J, Sodian R, Visjager J, Kolb S, Grunenfelder J, Zund G, Turina M. 2002. Tissue engineering of functional trileaflet heart valves from human marrow stromal cells. *Circulation* **106**[suppl 1]:I143–I150.
31. Hoffman JIE. 1995. Heterogeneity of myocardial blood flow. *Basic Res Cardiol* **90**:103–111.
32. Horgan CO, Saccomandi G. 2002. Constitutive modeling of rubber-like and biological materials with limiting chain extensibility. *Math Mech Solids* **7**:353–371.
33. Huang W, Yen RT, McLaurine M, Bledsoe G. 1996. Morphometry of the human pulmonary vasculature. *J Appl Physiol* **81(5)**:2123–2133.
34. Humphrey JD, Strumpf RK, Yin FCP. 1990. Determination of a constitutive relation for passive myocardium, I: a new functional form. *J Biomech Eng* **112**:333–339.
35. Humphrey JD, Yin FCP. 1988. Biaxial mechanical behavior of excised epicardium. *J Biomech Eng* **110**:349–351.
36. Kaazempur-Mofrad MR, Vacanti JP, Kamm RD. 2001. Computational modeling of blood flow and rheology in fractal microvascular networks. In *Computational fluid and solid mechanics*, Vol. 2, pp. 864–867. Ed. KJ Bathe. Elsevier Science, Oxford.

37. Kaazempur-Mofrad MR, Younis HF, Vacanti JP, Kamm RD. 2003. Biological simulations at all scales: from cardiovascular hemodynamics to protein molecular mechanics. In *Computational fluid and solid mechanics*, Vol. 2, pp. 8–12. Elsevier Science, Oxford.
38. Kaihara S, Borenstein JB, Koka R, Lalan S, Ochoa ER, Ravens M, Pien H, Cunningham B, Vacanti JP. 2000. Silicon micromachining to tissue engineer branched vascular channels for liver fabrication. *Tissue Eng* **6**(2):105–117.
39. Kassab GS, Fung YC. 1995. The pattern of coronary arteriolar bifurcations and the uniform shear hypothesis. *Ann Biomed Eng* **23**:13–20.
40. Kassab GS, Fung YC. 1994. Topology and dimensions of pig coronary capillary network. *Am J Physiol* **267**:H319–H325.
41. Kirkpatrick CJ, Unger RE, Krump-Konvalinkova V, Peters K, Schmidt H, Kamp G. 2003. Experimental approaches to study vascularization in tissue engineering and biomaterial applications. *J Mater Sci Mater Med* **14**:677–681.
42. Kojima K. 2003. A composite tissue-engineered trachea using sheep nasal chondrocyte and epithelial cells. *FASEB J* **17**:823–828.
43. Klein TJ, Schumacher BL, Schmidt TA, Li KW, Voegtline MS, Masuda K, Thonar EJ, Sah RL. 2003. Tissue engineering of stratified articular cartilage from chondrocyte subpopulations. *Osteoarthritis Cartilage* **11**(8):595–602.
44. Krenz GS, Dawson CA. 2002. Vessel distensibility and flow distribution in vascular trees. *J Math Biol* **44**:360–374.
45. Kwan MK, Woo SLY. 1989. A structural model to describe the nonlinear stress-strain behavior for parallel-fibered collagenous tissues. *J Biomech Eng* **111**:361–363.
46. Langille BL, O'Donnell F. 1986. Reductions in arterial diameter produced by chronic decreases in blood flow are endothelium-dependent. *Science* **231**(4736):405–407.
47. Lavine M, Roberts L, Smith O. 2002. If I only had a... *Science* **295**(8):995.
48. Lauffenburger DA, Linderman JJ. 1993. *Receptors: Models for binding, trafficking, and signaling*. Oxford UP, New York.
49. Lee H, Cusick RA, Browne F, Ho Kim T, Ma PX, Utsunomiya H, Langer R, Vacanti JP. 2002. Local delivery of basic fibroblast growth factor increases both angiogenesis and engraftment of hepatocytes in tissue-engineered polymer devices. *Transplantation* **73**(10):1589–1593.
50. Leeson-Dietrich J, Boughner D, Vesely I. 1995. Porcine pulmonary and aortic valves: a comparison of their tensile viscoelastic properties at physiological strain rates. *J Heart Valve Dis* **4**:88–94.
51. Li J, Lyo XY, Kuang ZB. 2002. A nonlinear anisotropic model for porcine aortic heart valves. *J Biomech* **34**:1279–1289.
52. Lo D, Vesely I. 1995. Biaxial strain analysis of the porcine aortic valve. *Ann Thorac Surg* **60**:S374–S378.
53. Marshall E. 2002. A space age vision advances in the clinic. *Science* **295**(8):1000–1001.
54. Maemura T, Shin M, Sato M, Mochizuki H, Vacanti JP. 2003. A tissue-engineered stomach as a replacement of the native stomach. *Transplantation* **76**(1):61–65.
55. Malek AM, Izumo S. 1996. Mechanism of endothelial cell shape change and cytoskeletal remodeling in response to fluid shear stress. *J Cell Sci* **109**(pt 4):713–726.
56. May-Newman K, Yin FCP. 1998. A constitutive law for mitral valve tissue. *J Biomech Eng* **120**(1):38–47.
57. Mcknight NL, Frangos JA. 2003. Strain rate mechanotransduction in aligned human vascular smooth muscle cells. *Ann Biomed Eng* **31**:239–249.
58. Millesi H, Meissl G, Berger G. 1972. The interfascicular nerve grafting of the median and ulnar nerves. *J Bone Joint Surg* **54A**:727–750.
59. Minnich B, Lametschwandtner A. 2000. Lengths measurements in microvascular corrosion castings: two-dimensionsal versus three-dimensional morphometry. *Scanning* **22**:173–177.

60. Mooney D, Hansen L, Vacanti J, Langer R, Farmer S, Ingber D. 1992. Switching from differentiation to growth in hepatocytes—control by extracellular matrix. *J Cell Physiol* **151**(3):497–505.
61. Ott K. 2002. The sum of its parts: an introduction to modern histories of prosthetics. In *Artificial parts, practical lives: modern histories of prosthetics*, pp. 1–23. Ed. K Ott, D Serlin, S Mihm. New York UP, New York.
62. Petruzelli L, Takami M, Humes D. 1999. Structure and function of cell adhesion molecules. *Am J Med* **106**:467–476.
63. Pierce SM, Skalak TC. 2003. Microvascular remodeling: a complex continuum spanning angiogenesis to arteiogenesis. *Microcirculation* **10**:99–111.
64. Powers MJ, Janigian DM, Wack KE, Baker CS, Beer Stolz D, Griffith LG. 2002. Functional behavior of primary rat liver cells in a three-dimensional perfused microarray bioreactor. *Tissue Eng* **8**:499–513.
65. Rivera DJ, Gores GJ, Misra SP, Hardin JA, Nyberg SL. 1999. Apoptosis by gel-entrapped hepatocytes in a bioartificial liver. *Transplantation Proc* **31**:671–673.
66. Roy P, Baskaran H, Tilles AW, Yarmush ML, Toner M. 2001. Analysis of oxygen transport to hepatocytes in a flat-plate microchannel bioreactor. *Ann Biomed Eng* **29**:947–955.
67. Rozga J, Williams F, Ro MS, Neuzil DF, Giorgio TD, Backfisch G, Moscioni AD, Hakim R, Demetriou AA. 1993. Development of a bioartificial liver: properties and function of a hollow-fiber module inoculated with liver cells. *Hepatology* **17**:258–265.
68. Sah RL, Kim YJ, Doong JY, Grodzinsky AJ, Plaas AH, Sandy JD. 1989. Biosynthetic response of cartilage explants to dynamic compression. *J Orthop Res* **7**:619–636.
69. Schmid-Schonbein GW. 1988. A theory of blood flow in skeletal muscle. *J Biomech Eng* **110**:20–26.
70. Shinoka T, Ma PX, Shum-Tim D, Breuer CK, Cusick RA, Zund G, Langer R, Vacanti JP, Mayer Jr JE. 1996. Tissue engineered heart valves: autologous valve leaflet replacement study in a lamb model. *Circulation* **94**(Suppl II):164–168.
71. Silvestre JS, Tamarat R, Ebrahimian TG, Le-Roux A, Clergue M, Emmanuel F, Duriez M, Schwartz B, Branellec D, Levy BI. 2003. Vascular endothelial growth factor-B promotes in vivo angiogenesis. *Circ Res* **93**(2):114–123.
72. Skalak TC. 2002. In vivo and in silico approaches for analysis and design of multisignal, multicomponent assembly processes in vascular systems. *Ann NY Acad Sci* **961**:243–245.
73. Sodian R, Hoerstrup SP, Sperling JS, Daebritz SH, Martin DP, Schoen FJ, Vacanti JP, Mayer Jr JE. 2000. Tissue engineering of heart valves: in vitro experiences. *Ann Thorac Surg* **70**:140–144.
74. Stock UA, Vacanti JP, Mayer Jr JE, Wahlers T. 2002. Tissue engineering of heart valves—current aspects. *Thorac Cardiov Surg* **50**:184–193.
75. Tilles AW, Baskaran H, Roy P, Yarmush ML, Toner M. 2001. Effects of oxygenation and flow on the viability and function of rat hepatocytes cocultured in a microchannel flat-plate bioreactor. *Biotechnol Bioeng* **73**:379–389.
76. Vacanti JP, Vacanti CA. 2000. The History and scope of tissue engineering. In *Principles of tissue engineering*, pp. 3–7. Eds. RP Lanza, R Langer, J Vacanti. Academics Press, New York.
77. Weinberg EJ, Kaazempur-Mofrad MR, Borenstein JT. 2003. Numerical model of flow in distensible microfluidic network. In *Computational fluid and solid mechanics*, Vol. 2, pp. 1569–1572. Ed. KJ Bathe. Elsevier Science, Oxford.
78. Whitesides GM, Strook AD. 2001. Flexible methods for microfluidics. *Phys Today* **54**(6):42–48.
79. Yannas IV. 2000. In vivo synthesis of tissues and organs. In *Principles of tissue engineering*, 167–178. Eds. RP Lanza, R Langer, J Vacanti. Academics Press, New York.
80. Yannas IV. 2001. *Tissue and organ regeneration in adults*. Springer Verlag, New York.

81. Cassel OCS, Morrison WA, Bratt AM, Neufang B, Greene PL, Jackisch R, Hertting G, Will BE. 2001. The influence of extracellular matrix on the generation of vascularized, engineered, transplantable tissue. *Ann NY Acad Sci* **944**:429–442.
82. Richardson TP, Peters MC, Ennett AB, Mooney DJ. 2001. Polymeric system for dual growth factor delivery. *Nature* **19**:1029–34.
83. Vitali M. 1978. *Amputation and prosthesis*. Cassel & Co., London.
84. Edwards Lifesciences. 2004. www.edwards.com.
85. Otto Bock Healthcare. 2004. www.ottobockus.com.
86. Ninomiya H, Inomata T. 2005. Microvasculature of the hamster eye: scanning electron microscopy of vascular corrosion casts. *Veterinary Ophthalmol* **8**(1):7.

5

IMAGING THE NEURAL SYSTEMS FOR MOTIVATED BEHAVIOR AND THEIR DYSFUNCTION IN NEUROPSYCHIATRIC ILLNESS

Hans C. Breiter and Gregory P. Gasic

Departments of Radiology and Psychiatry, Massachusetts General Hospital and Harvard Medical School, Boston

Nikos Makris

Athinoula Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Massachusetts Institute of Technology, and Harvard Medical School, Boston

Tomographic imaging of the human brain has enabled neuroscientists to begin dissection of the complex distributed neural groups that make up the human brain. Of the available tomographic technologies, functional magnetic resonance imaging (fMRI) has emerged as an important tool for the systems neuroscience of cognitive and emotional functions. fMRI has also been an important technology in developing evidence for a generalized circuitry that processes reward/aversion information. Composed of an extended set of subcortical gray matter regions and the surrounding paralimbic girdle, this reward/aversion circuitry forms the core of an informational backbone for motivation (iBM) underlying behavior. Differential components of this iBM appear to be structurally or functionally affected in many neuropsychiatric illnesses. Should some of these structural and functional alterations in the iBM and connected systems be shown to be quantitative measures that are inherited versus state-dependent, they would likely group psychiatric illnesses on a more etiological basis than the diagnostic categories based on statistical clusters of behaviors and symptoms that are used in current psychiatric diagnosis. This chapter will explore how integrative systems biology approaches can bridge the distributed neural circuits responsible for the processing of reward/aversion

Address correspondence to: Hans C. Breiter, Athinoula Martinos Center for Biomedical Imaging, Building 149-2301, Thirteenth Street, Charlestown, MA 02129-2060 (hbreiter@partners.org); Gregory P. Gasic (ggasic@nmr.mgh.harvard.edu); Nikos Makris (nikos@cma.mgh.harvard.edu).

function and the networks of genes responsible for the development and maintenance of these neural circuits. A major challenge for future research will be to determine whether there is strong or weak scale invariance at the multiple spatiotemporal scales of brain organization. A combined genetics, genomics, and integrative neuroscience approach has the potential to redefine our conceptualization of neuropsychiatric illnesses with the implementation of objective quantitative measures that can then be translated into probabilistic diagnoses for these complex diseases of the human brain.

1. INTRODUCTION

The broad questions posed by astrophysics and by motivation neuroscience are conceptually similar but converse in their focus, one peering out at the universe and the other gazing into the brain. The former asks, "what is the nature of what we perceive?" and "how is something created from nothing?" The latter asks, "how do we perceive, control interpretation of perception, and exercise free will?" The principal question of motivation neuroscience asks, "why is there directed action?"

Motivation is the engine that allows organisms to make choices, direct their behavior, or plan their actions across time. Motivated behavior can be defined by goal-directed behavior that optimizes the fitness of an organism or social group. It depends on input from evaluative processes regarding internal homeostatic and socially acquired needs, potential goal-objects in the environment meeting these needs, remembered consequences of previous behavior toward goal-objects, and perceived needs in other cooperative or competitive organisms. The combined neural systems that produce this directed action constitute the neural basis of what we call motivation (263). These neural systems are composed of multiple subsystems that have evolved to allow an organism to assign a value to objects in the environment so that the organism works for "rewards" and avoids "sanctions" or aversive outcomes. Central to these neural systems are a set of subcortical gray matter regions [nucleus accumbens (NAc), caudate, putamen, hippocampus, amygdala, sublenticular extended amygdala of the basal forebrain (SLEA), hypothalamus, and thalamus] (117) and components of the paralimbic girdle [including the orbitofrontal cortex (GOb), insula, cingulate cortex, parahippocampus, and temporal pole] (174). Other networks across the prefrontal cortex also appear to be engaged in the evaluative and decision making components of motivated behavior. A number of these regions are modulated by dopaminergic neurons in the substantia nigra, the retrotuberal field, and the ventral tegmental area (henceforth jointly referred to as the ventral tegmentum: VT) (228). Less is understood regarding the roles of non-dopaminergic neuromodulators (noradrenergic, serotonergic, cholinergic, steroid hormones, and neuropeptides) that appear to alter the balance between excitatory and inhibitory synaptic neurotransmission during motivated behavior. In 1954, Olds and Milner (192) were the first to implicate a subset of these regions in reward-mediated behavior. In subsequent decades, pioneering studies by others contributed to our further understanding of these systems (89,98,148,265).

Over the last ten years, neuroimaging has allowed the study of these neural processes in humans and has started to dissect the contribution of individual brain regions to the processing of motivationally significant information. Recent structural and functional neuroimaging has implicated a number of these brain regions in psychiatric disease.

Since Aquinas, Spinoza, and Bentham, a central question has been how rewarding stimuli are experienced relative to aversive or painful events, and how the experience of reward is translated across categories of stimuli that reinforce behavior (9,22,240). Today, human neuroimaging studies (at the limits of their current resolution) have provided evidence for a generalized circuitry processing rewarding and aversive stimuli. Motivationally salient features of infused drugs of abuse, fruit juice consumption, perceived beautiful faces or music, monetary gains and losses, somatosensory pain, and cues of aversive events activate a common set of distributed neural circuits that process rewards and sanctions (3,19,24,27,35,37,38,88,137,189,200). Within some of these neural groups, separate local circuits have been shown during electrophysiological studies of mammals to selectively activate in response to distinct categories of rewarding input (51,52). A number of human neuroimaging studies have started dissecting the subcomponent processes of the functions processing reward/aversion information (33,38). The results of human studies together with those in phylogenetically lower species (132,147,215) point to the existence of an informational backbone (iBM) focused on processing reward/aversion information.

Circuits within this iBM have been reported to be functionally or structurally altered in a number of neuropsychiatric illnesses (36,63,103,126,161,181, 223,225,257). This body of research suggests that these illnesses can be characterized by distinct circuit-based alterations. If a subset of these circuit-based alterations were shown to be heritable, they might serve as endophenotypes for future genetic linkage studies. To be endophenotypes, or heritable/familial quantitative traits, these circuit-based alterations would need to correlate with individuals' risk of developing a disease but not be a sign of disease progression (4,106). Along with state-sensitive alterations, circuitry-based quantitative traits may serve as better diagnostic markers (121,186,241) than those currently used for psychiatric diagnosis based on statistical associations of behavioral signs and symptoms. A detailed characterization of neural circuits in affected individuals, their family members, and family-based matched controls (thereby producing a "systems biology map") may enable us to characterize the genetic and epigenetic factors that combine to produce these circuit-based alterations. Supporting these possibilities, recent studies involving presentation of motivationally salient stimuli allude to a potential correspondence between events at the molecular and brain circuitry levels (11,19,242). Correspondence of measures across scales of brain function suggests that similar principles of organization may be operative with extended molecular networks and with distributed neural groups (21; see Part II, chapters 4 (by Wuchty, Ravasz, and Barabási) and 5 (by Krakauer), and Part III, chapter 5.1 (by Reeke), this volume).

To develop these interlinked concepts, this chapter will be organized in five sections. The first section will describe one of the widely used approaches in brain mapping, namely, functional magnetic resonance imaging (fMRI), which has provided important insights into the neural systems in humans involved with emotion and the processing of reward/aversion information. This type of neuroimaging, in conjunction with developments in the experimental psychology of motivation described in the second section, has produced the data and approaches described in the third, fourth, and fifth sections. The second section will serve as a bridge between the first and third sections and describe a general model of motivation and the embedded systems for processing reward/aversion information (i.e., via an iBM) as well as those that give rise to emotion. The third section will describe converging evidence from human fMRI and other neuroimaging studies, as well as physiological studies in animals, for a common circuitry processing reward/aversion information and its component subprocesses. The fourth section will synthesize a body of human neuroimaging evidence that argues for a dysfunction in components of this reward/aversion circuitry in neuropsychiatric illnesses. The final section will describe how a dense mapping of the neural systems responsible for reward/aversion function combined with genetic and genotypic data should enable us to hone in on the networks of genes responsible for the development and maintenance of these neural circuits, in health and in neuropsychiatric illness. Through the use of objective quantitative measures, integrative neuroscience approaches have the potential to redefine our conceptualization of neuropsychiatric illness (32).

2. IN VIVO MEASUREMENT OF HUMAN BRAIN ACTIVITY USING fMRI

The majority of the data gathered over the past 15 years characterizing the neural substrates of human motivational function has been collected via tomographic and non-tomographic brain imaging techniques. Tomographic techniques that localize signal changes in three-dimensional space include: positron emission tomography (PET), single photon emission computed tomography (SPECT), magnetic resonance imaging (MRI), and optical imaging techniques. Non-tomographic techniques include electroencephalography (EEG) and magnetoencephalography (MEG). Each of these techniques has unique benefits that warrant its use for specific neuroscience questions (please see (251) for technical discussion of such considerations). Functional MRI (fMRI) has been the most widely used technique to study motivation in humans. In contrast to studies of normative reward circuitry function, dysfunction of these systems that contribute to neuropsychiatric illnesses has not yet been commonly studied with fMRI. Given the ease with which fMRI acquisitions can be combined with other forms of MRI: (1) high-resolution structural scanning for morphometric quantitative anatomy measures (see Figure 1a), (2) arterial spin-labeling scans for absolute resting perfusion, (3) diffusion tensor imaging for white matter tractography (see Figure 1b), or (4) spectroscopy for chemical signatures related to

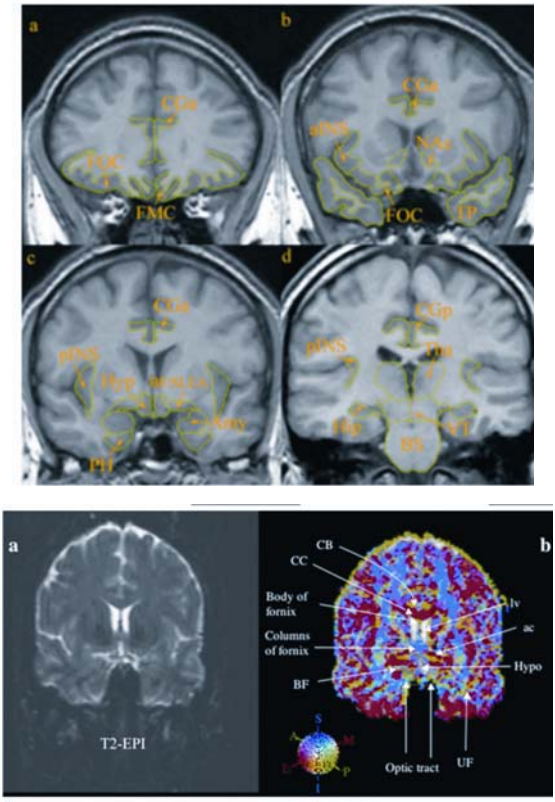


Figure 1. (**a**) Morphometric segmentation of T_1-weighted MRI data in coronal sections using the morphometric methodology of the Center for Morphometrc Analysis (54,91). a–d are four representative coronal sections of the human brain in the rostral-caudal dimension showing limbic and paralimbic structures. The colored sphere shows the color coding scheme applied for the visualization of the tensors: red stands for the medial-lateral orientation, green indicates the anterior-posterior orientation, and blue shows the superior-inferior orientation. Abbreviations: FOC = Frontal Orbital Cortex, FMC = Frontal Medial Cortex, CGa = Cingulate Gyrus (anterior), CGp = Cingulate Gyrus (posterior), NAc = Nucleus Accumbens, TP = Temporal Pole, INS = Insula, BF/SLEA = Basal Forebrain/Sublenticular Extended Amygdala, Hip = Hippocampus, PH = Parahippocampal Gyrus, VT = Ventral Tegmental Area, Tha = Thalamus, Hyp = Hypothalamus, BS = Brain Stem, Amy = Amygdala. (**b**) Diffusion tensor magnetic resonance (DT-MR) image of limbic fiber pathways. (b) shows the primary eigenvector map (PEM) of a coronal slice at the level of the anterior commissure; the T_2-EPI image of this slice is shown in (a). Dotted rectangles highlight regions that include major limbic pathways such as the fornix, cingulum bundle, medial forebrain bundle and ventral amygdalofugal projection (in the basal forebrain or BF). Abbreviations: CB = cingulum bundle, UF = uncinate fasciculus, ac = anterior commissure, CC = corpus callosum, lv = lateral ventricle. For further details on DT-MRI, please see Makris et al. (157,159).

neural integrity, fMRI is likely to become a technique more commonly used for researching neuropsychiatric illness, and potentially clinical diagnoses. What

follows is a brief overview of the development of fMRI and some of the research into what it actually samples of brain activity.

There are now a number of different fMRI techniques that can be used for making movies of focal changes in brain physiology related to one or a set of targeted mental functions. These techniques can be roughly categorized by the use of exogenous (injected) contrast agent versus methodologies based upon intrinsic contrast (natural contrast agent in the blood, or the effects of flow on MRI signal). Functional MRI with contrast agents was first demonstrated by Belliveau et al. (20), using echo-planar imaging in combination with the paramagnetic contrast agent gadolinium, bound to a chelating agent, DTPA. This general methodology works because the presence of Gd-DTPA within the parenchymal vasculature increases the decay rate of the MR signal ($1/T_2$) in a regionally specific fashion. This, in turn, changes the image contrast, and serial measurement of image intensity can be converted to regional cerebral blood volume. If injections of Gd-DTPA are made during different experimental conditions—for instance, the rest condition of no movement and the targeted condition of finger apposition—contrasting images acquired during each experimental condition can lead to a measure of cerebral blood volume change associated with the experimental perturbation to the system. The cerebral blood volume change can be evaluated statistically, and overlaid on a structural MRI to illustrate the anatomical localization. When using novel contrast agents with long blood half-lives, as can be employed in animal models, this technique becomes particularly powerful. Repeated injections are no longer required, and changes in blood volume can be assessed dynamically throughout the entire experiment. Moreover, the dose of injected agent can be dialed to the optimum value in order to produce the strongest of all possible fMRI signals. However, the gadolinium compounds used in humans have very short blood half-lives, and no suitable agents are approved at this time.

For human studies, the most widely utilized technique is based upon changes in an intrinsic contrast agent, deoxygenated hemoglobin. Its development followed from the classic work of Pauling and Coryell (197,198) on the diamagnetic versus paramagnetic state of oxyhemoglobin and deoxyhemoglobin, respectively. Subsequent work by Thulborn and colleagues (249) evaluated the in-vitro effect of oxygenation on the MRI signal. Independent groups led by Ogawa and Turner extended these observations to note similar changes alter T_2-weighted signals in vivo in mammals. Parallel work by Detre and colleagues (77) demonstrated how to use T_1-weighted signals to quantify perfusion. With important modifications, Kwong and colleagues (143) applied these developments to image oxygenation and flow changes associated with neural activity. The work of Kwong and colleagues was first presented to other scientists at the 10th Annual Meeting of the Society of Magnetic Resonance (August 1991) and were rapidly replicated and extended (10,143,190).

Like the method based upon injected contrast agents, intrinsic contrast is sensitive to T_2*-weighted signal changes and critically depends on the observations of Pauling and Coryell (197,198) that the magnetic properties of hemoglobin change from the oxy-state, which is diamagnetic, to the deoxy-state, which is paramagnetic. Because of this issue, the intrinsic contrast technique has been called blood oxygen-level-dependent contrast imaging or BOLD (190). Blood oxygen-level-dependent contrast results from a set of effects initiated by changes in local cellular activity. These effects include alterations in cerebral blood flow (CBF) and cerebral blood volume (CBV) that in general produce increased oxygen delivery beyond oxygen utilization, so that there is a relative decrease in local deoxyhemoglobin concentration. A relative decrease in deoxyhemoglobin concentration results in an increase in the relaxation time T_2*, or apparent T_2, leading to an increase in MR signal in brain regions with increased neural activity.

The mechanistic details behind this general model of intrinsic contrast fMRI continue to be a topic of active research, specifically around (a) the neural correlates of BOLD and other perfusion-weighted signals, (b) the coupling of neural activity with vascular responses, and (c) factors influencing the concentration of deoxyhemoglobin. What follows is a synopsis of research on the first and last of these topics, given their relevance to interpretation of fMRI studies of normative motivation, and altered motivational function in the form of neuropsychiatric illness.

In a number of circumstances, BOLD signal changes have been observed to be proportional to changes in neuronal spike rates (213). But other work involving stimulation of parallel fibers with neutralizing effects on measured spike rates has shown circumstances where spike rate and CBF diverge (166). Data have been further presented that local field potentials (LFPs) correlate better with CBF/BOLD effects than spike rate (156), suggesting that changes in BOLD signal reflect incoming synaptic activity and local synaptic processing. This relationship between LFPs and CBF/BOLD will vary with local neural architecture in that this relationship has been observed to be linear during climbing fiber stimulation, and nonlinear with parallel fiber stimulation (166).

When neural activity is altered, corresponding effects are observed in CBF, CBV, and oxygen consumption. The weight accorded to these effects and their impact on deoxyhemoglobin concentration has been the topic of intense investigation. Early in the development of fMRI, capillary perfusion studies in the rat brain demonstrated that essentially all cerebral capillaries are perfused in the basal state, and that increases in perfusion are accomplished primarily by increases in blood velocity within capillaries, as opposed to the capillary recruitment that is such a large factor in muscle (255). This observation was important in that it defines clear limits for the coupling ratios of blood flow to volume and to oxygen utilization. Activation-induced changes in oxygen utilization (CMRO2) are thought to be coupled to changes in blood flow by a diffusion

limitation on oxygen, so that relative changes in CBF always exceed those in CMRO2, leading to a positive change in blood oxygen (45). Subsequent research illustrated the complexity of the relationship between oxyhemoglobin and deoxyhemoglobin concentrations (153), and how this relationship was differentially impacted by oxygen consumption, CBF, and CBV over time after a controlled experimental stimulus (71,72,122,160). Recent work has begun to emphasize the importance of understanding how brain pathology and medications can alter the relative weightings of oxygen consumption and CBV effects on BOLD signal, so that they become a dominant factor (187). For example, altered neurovascular coupling has been observed with carotid occlusion (216), transient global ischemia (224), subarachnoid hemorrhage (80), and theophylline or scopolamine treatment (78,253). To date, altered neurovascular coupling in the basal state (as opposed to functionally induced changes in CBF or CBV) has not been demonstrated to be a consistent effect of Axis I neuropsychiatric illnesses, although medications, drugs of abuse, and changes in ventilation or heart rate will alter global parenchymal perfusion and make focal BOLD measures more difficult (105). It thus remains a defensible hypothesis among psychiatric neuroimagers using fMRI that illnesses such as major depressive disorder, generalized anxiety disorder, obsessive compulsive disorder, addiction, and schizophrenia have neurovascular coupling mechanisms that are similar to those in healthy controls. It also remains a defensible hypothesis that studies using rewarding stimuli, aversive stimuli, or stimuli with strong emotional content will not alter neurovascular coupling.

3. THEORETICAL MODEL OF MOTIVATION FUNCTION

To guide their behavior, humans integrate unconscious and conscious mental processes to produce judgments, select choices, and make decisions. These processes involve a systematic evaluation of: internal physiological, mental, and socially acquired needs; potential goal-objects in the environment meeting these needs; memories of outcomes from previous behaviors directed toward goal-objects; and the perceived needs of other cooperative or competitive organisms. To explain the intensity and direction of behavior, the integration of these evaluative processes has traditionally been referred to as a *drive* or *motivational state* (33). With evolutionary pressures toward the selection of fitness, motivational states seek to maximize personal fitness over time through choices of goal-objects and activities.

Motivational states necessitate planning over time, along with planning in parallel for alternative behaviors and choices. Drives that control behavior do not all have a well-defined temporal relationship to environmental events (i.e., curiosity) (33,142). Yet all drives select outcomes that produce variable arousal and satiation/relief. Motivational states allow an organism to intentionally moni-

tor its needs over time (as opposed to doing so in a stimulus response fashion), and to select environmental opportunities that fulfill these needs. A number of motivational states can be present concurrently (e.g., while running across a parking lot to catch a train you become hot, tired, and short-of-breath). Homeostatic and biological control of thermoregulation, oxygen saturation of hemoglobin, osmolality, or glucose level can be balanced against inter-organism and social objectives related to defense, shelter, procreation, hierarchical ordering, and curiosity. The collective neuronal and physiological processes that mediate drives based on such needs, and linked intentional activity that meets these needs via behavioral processes inclusive of decision-making, speech, and imagination, can be collectively referred to as motivation.

Current models of motivational function developed as a challenge to the model of behaviorism. The behaviorist model postulated that goal-objects in the environment have organizing effects on behavior through "stimulus-response" relationships. In this construct, "rewards" were goal-objects or stimuli that produced repeated approach behaviors or response repetitions. A rewarding stimulus could act via a memory or via salient sensory properties (i.e., a food odor) to be an *incentive* for approach behavior. In contrast, a rewarding stimulus that increased the probability that preceding behavioral responses would be repeated (i.e., drug self-administration) would *reinforce* previous behavior. This behavioral perspective had difficulties with concepts such as target detection by the brain (41). Likewise, its stimulus–response framework did not allow for symbolic manipulation as described by Chomskian linguistics (95), or inferential processes (178). These deficiencies instigated conceptual revolts in the form of cognitive neuroscience (113,139,165,176), neuro-computation (123,58,169,56), judgment and decision making (172,238), emotion neuroscience (69,146,196), behavioral ecology and neuroeconomics (102,140), and a more "pragmatist" perspective framed by nonlinear dynamics (95,96), and the effects of nonlinear processes on system information (109,110,267).

Synthesizing these viewpoints, a general schema for motivation functions is illustrated in Figure 2a. This schema for a Motivation Information Theoretic (MIT) model is generally consistent with recent neuro-computational evidence (75,96,243). In the MIT model, at least three fundamental operations can be ascribed to motivated behavior (33), which have precursors in models of animal cognitive physiology function and communication theory (234) (Figure 2b,c). These operations include a number of processes. One grouping of processes (A) includes evaluation of homeostatic and social needs, and selection of objectives to meet these needs. A second grouping of processes (B) includes sensory perception of potential goal-objects that may meet these objectives, assessment of potential reward/aversion outcomes related to these goal-objects, and comparison of these assessments against memory of prior outcomes. A third grouping (C) involves assessment, planning, and execution of action to obtain or avoid these outcomes (33,119,120,167,168,206,237,238). As these operations rely on

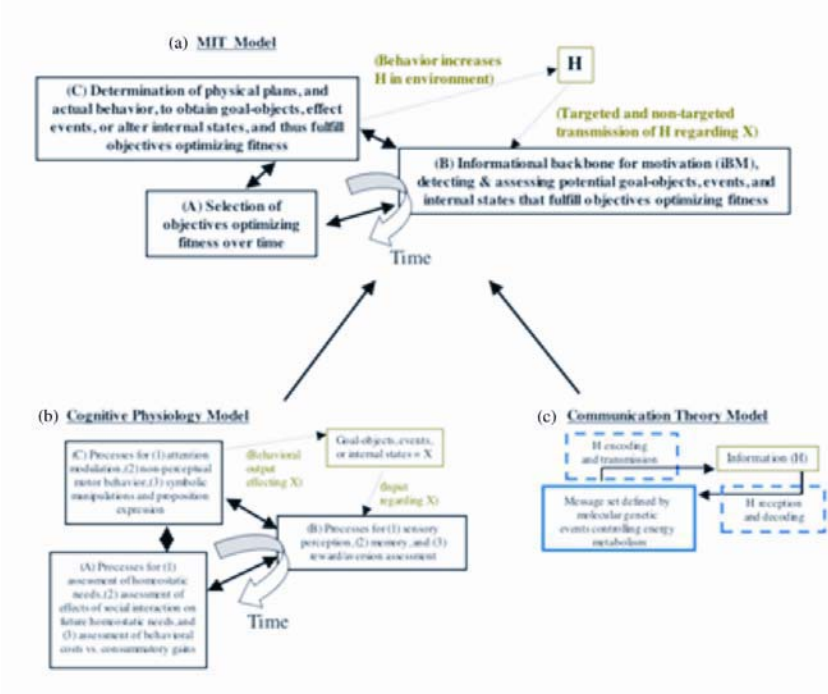


Figure 2. (**a**) The MIT (Motivation Information Theoretic) Model synthesizes the processes of (B) with those of (C) with input entry via an informational backbone for motivation (iBM). This information backbone assesses if potential goal-objects will fulfill organism objectives for fitness, and interfaces with the behavioral operation to obtain such goal-objects. Given the interdependence of these sets of brain processes on each other, they function as if they were all orthogonal to time. See text for further details. (**b**) A cognitive physiology model organized around three general operations for mediating directed action. One operation (A) is composed of processes that evaluate organism needs across multiple dimensions, and potential energy costs for fulfilling these needs by plans devised in (C). A second operation (B) includes processes for sensory perception, memory of previous outcomes and their contexts, and assessment of how rewarding or aversive potential goal-objects or events might be. (**c**) Information (H), as defined by Shannon and Weaver (234), is received and decoded during communication by processes that allow incoming information to be linked to the set of communicable messages. Messages, in turn, are encoded and transmitted in the form of behavior. Self-organizing organisms always generate entropy as an outcome, which acts as a force behind the development of complexity in coding/decoding systems such as the brain, and their evolution toward greater complexity (204). See text for definitions of abbreviations.

intricate feedback loops in their production of a behavioral trajectory, they are not necessarily sequential but orthogonal to time. The third operation, in humans, clusters a number of possible actions: (1) modulation of attention-based filtering of perceptual input, (2) organization of motor output to obtain goal-objects, (3) control of cognitive, logical, and internal imagery systems (and their symbolic output in the form of language) to increase the range of goal-objects that can be obtained, problems that can be solved, or events that can be experienced (139,177,238).

The outputs of this system, intentional behaviors, are a form of communication (234). These outputs also represent a means for modulating sensory inputs to the brain, as distinct from top–down adaptation of input once it is in modality-specific processing streams (97). The reference to communication theory (234) is given because it helped foster the revolt against behaviorism. This theory presented a schema for understanding communication in its most general sense, namely, how one mind or machine affects another. In it, Claude Shannon and colleagues focused on the technical constraints on communication, and did not address the "semantic problem" or the "effectiveness problem" of communication. Ideas from communication theory were integrated with neural systems biology only recently in domains such as: sensory representation and memory (56,58,123,203), reward prediction (229), serial response learning and novelty assessment (23), conditional probability computation (33), and nonlinear dynamics underpinning decision making (95,96). Recently, attempts have been made to address the "semantic problem" posed by inter-organism communication, and what constitutes meaning for a biological system. Two viable hypotheses have been advanced, which could be considered as two sides of the same coin. One hypothesis connects meaning in biological systems to the intersection of intentional behaviors between organisms (95). The alternative hypothesis places meaning within the context of organism optimization of fitness over time and tissue metabolic needs (33). For the latter hypothesis, communication between organisms utilizes message sets defined by genomic and epigenomic control of the bioenergetics of metabolism (32).

A number of the general operations and processes of the MIT model (Figure 2a) have been the target of experimental dissection. For instance, when an animal seeks and finds an object with motivational salience, a set of hypothetical informational subprocesses appear to be active (Figure 3) (15,33,66,76,96,114, 116,237,138,268). A partial listing of these subprocesses includes the following: (1) reception of input from the environment or internal milieu across multiple channels, (2) representation of this input by transient neuronal activity, (3) evaluation of input representations for sensory modality-specific characteristics such as color and motion, (4) combination of these representations across modality at theoretical convergence zones as potential percepts, (5) encoding of representations into memory and contrast with other stimulus memories, and (6) evaluation of representations for features (rate, delay, intensity, amount, category) that are important for the organization of behavior. Feature evaluation encompasses: (a) categorical identification of putative "rewards" or aversive stimuli, (b) extraction of rate and delay information from the object of worth, and (c) valuation of goal-object intensity (i.e., strength) and amount in the context of potential hedonic deficit states. The second of these feature evaluation subprocesses allows computation of a rate function to model temporal behavior (99), and of a probability function for possible outcomes (131,254) (Figure 3). The mechanism by which the output of these valuation and probability (i.e.,

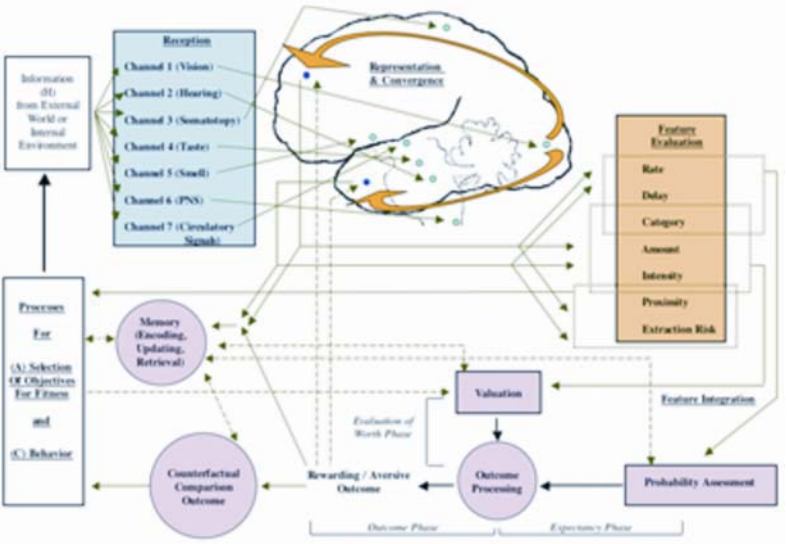


Figure 3. The informational backbone for processing reward-aversion input (iBM). Reflecting one interpretation of research from evolutionary ecology (102,237) and behavioral finance (131,238,254), the informational backbone is dissected into its theorized processes for sensory perception, memory, and reward/aversion assessment. Solid lines connect subprocesses as steps in the processing of information, whereas dashed lines represent subprocesses theorized to produce feedback leading to nonlinear system function (95,96). Postulated early subprocesses (blue box and brain cartoon) include: (1) information reception over discrete channels and representation (light blue points in cartoon of brain), (2) convergence of processed informational measures such as color, contrast and detected motion for vision (over ventral and dorsal processing streams as yellow arrows moving to dark blue points), and (3) convergence of represented information for construction of a percept from distinct receptive channels (15,66,76,114,268). Subsequent subprocesses are postulated to include the extraction of informational features (yellow box) with motivational relevance, including rate, delay, category, amount, and intensity information to be integrated during computation of probability functions and valuation functions (pink boxes), along with input regarding proximity and extraction risk (i.e., "cost" assessments) needed for general cost–benefit analyses (33,238). Features that are integrated in such computations are grouped with dotted lines. Determination of rewarding and aversive outcomes is theorized to evolve memory processes, particularly for evaluation of counterfactual comparisons (pink boxes) (170,171). See text for further details. Figure adapted with permission from Breiter and Gasic (32).

expectation) subprocesses is combined continues to be a domain of active inquiry (238).

For the MIT model in Figures 2 and 3, a "reward" is defined as the positive value that an animal attributes to a goal-object, an internal physical state, a behavioral act, or a cue associated with any of these. Rewards with a direct temporal connection to homeostatic regulation have a variable valuation dependent on the physiological state of the organism and the organism's previous experience with it (3,46). Although they are often referred to as "deficit states" connected to the physiological needs of the organism (e.g., glucose level, oxygen saturation, thermoregulation), they can be challenging to define in the case of social rewards (e.g., social or personal aspirations). The relationship of social rewards to

deficit states is not always apparent (3,46,48). But, social rewards may provide insurance over time for satiating some motivational states, or avoiding aversive outcomes (1,146) (Figure 4). Aversive events, in contrast, can be defined as deficit states whose reduction could be considered rewarding (19). Along with potential deficit states, rewarding and aversive outcomes also depend on valuations and probability assessments of alternative payoffs that do not occur (i.e., counterfactual comparisons involving memory). As an example of a counterfactual comparison (171), imagine that you and a friend saunter down a street. Both of you simultaneously find money on opposite sides of the walkway, but she finds a twenty dollar bill and you find a one dollar bill, resulting in you feeling you were not very fortunate (Figure 3).

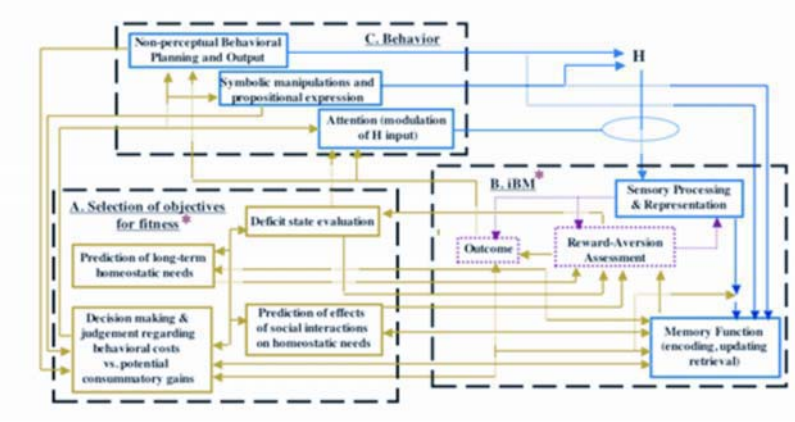


Figure 4. The schematic of the Motivation Information Theoretic model is shown with compartment lines represented by dashed black lines in bold. Compartments and connections in solid blue represent processes and their interactions for which substantial neuroscience data has accumulated. Compartments and interactions in light green are based on behavioral research and beginning neuroscience data (1,46,99). Substantially less is known for them than for processes in solid blue. Purple dashes represent processes and interactions based on a body of neuroscience data, which is still far from the level of knowledge currently available for the processes in solid blue. Stars are placed by the informational backbone for motivation (iBM), and the operation for selection of objectives that optimize fitness over time, to communicate a synthetic view that their processes comprise those that constitute the experience of emotion (65,67,129,146). See text for details. Figure adapted with permission from Breiter and Gasic (32).

Objectives for optimizing fitness (Figures 2 and 3) focus on satiating both short-term homeostatic needs and projected long-term needs through the insurance provided by social interaction and planning (1). They represent multiple motivational states, whose differing temporal demands produce complex dynamics between competing behavioral incentives. Darwin first recognized this idea (67), and hypothesized that motivational states form the basis for emotion.

Extrapolating from observation of facial expression, he theorized that the expression of emotion communicates internal motivational states between organisms. Experimental evidence supporting such a thesis of internal sources for emotion was produced by Cannon (50) and others (49,73,144,147). An alternative theory on emotion from James and Lange (129,145) posited that sensory inputs regarding bodily function were central to emotional experience. The James-Lange thesis has also been supported by experimental data (65). These perspectives on emotional function are both represented by processes within the general schema for motivation in Figure 4. By this view, emotion represents an interaction between processes for (1) evaluation of potential deficit states, (2) prediction of future needs, (3) processing of sensory input about the condition of the body and others' bodies, (4) assessment of the presence of potential goal-objects or aversive events that might alter particular deficit states, and (5) retrieval and updating of memories regarding (a) the outcome of prior deficit states, (b) social interactions plus conversations, and (c) contexts with particular goal-objects or aversive events. A view of this sort potentially allows for the intrapsychic complexity of human psychology (177). It conceptualizes emotion within the schema of motivation, potentially permitting linkage to processes that have been a strong focus of cognitive neuroscience research, and synthesizing the original perspectives of Darwin and James.

In Figure 4, processes shown in solid blue are supported by behavioral and neuroscience data. Processes determining input and output to the organism appear to be readily observed via experimentation. In contrast, the processes shown in light green are supported by emerging behavioral data, although much remains to be known about their systems biology. Relatively recently, the processes indicated with purple dashes (also see subprocesses at the bottom of Figure 3) have been associated with neural activity in a distributed set of deep brain regions, suggesting that they are part of an informational backbone for motivation processing rewarding and aversive events.

4. NEUROIMAGING OF THE GENERAL REWARD/AVERSION SYSTEM UNDERLYING MOTIVATED BEHAVIOR

Animal studies implicate the many projection fields of the VT dopamine neurons—such as the nucleus accumbens (NAc), hypothalamus, amygdala, sublenticular extended amygdala (SLEA) of the basal forebrain, and multiple fields in the paralimbic girdle (117,154,174,263) (Figure 5)—as components of the neural system that selects rewarding goal-objects and avoids their obverse. Over the last 8 years, functional neuroimaging studies of humans have identified homologous systems to be processing reward/aversion information, and have begun to dissect their contributions to motivated behavior.

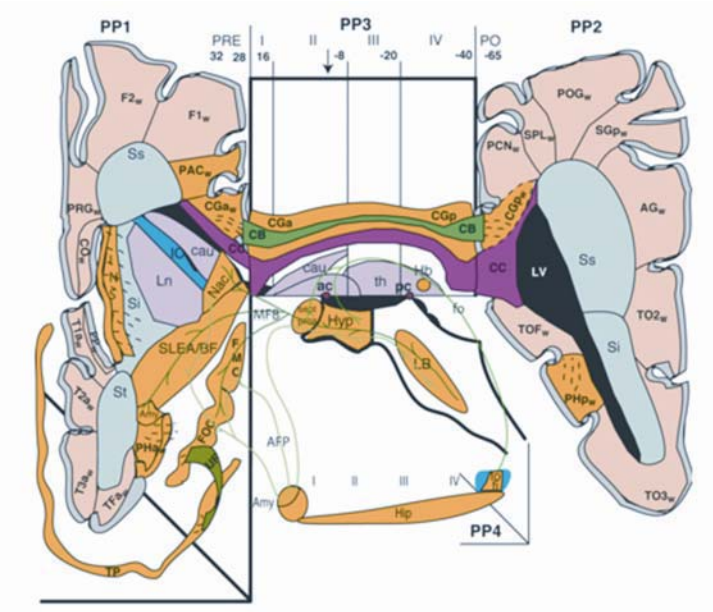


Figure 5. Diagrammatic representation of topologic relationships of brain stem and parcellated forebrain structures (158). Limbic and paralimbic structures (in yellow) and some of their main connections are shown in a schematic fashion. The diagram is flanked on the left and right by coronal projection planes (PP1, PP2). The paracallosal coronal slabs I–IV are distinguished by verticals in the interval, displayed on the midsagittal plane (PP3) of the hemisphere. The numbers aligned along the top of this plane correspond to the *y*-axis (anterior-posterior) coordinates of the Talairach stereotactic system (246) for the standard brain used to develop this system. PP1 corresponds to a composite coronal plane (temporal lobe is forced to be more anteriorly located) projected at the Talairach coordinate, indicated by the vertical black arrow in paracallosal slab II. PP2 corresponds to the immediately posterior callosal coronal plane at Talairach coordinate –40. The projection of these slabs within the temporal lobe is indicated by step PP4. Whereas the amygdalo-hippocampal junction will have the approximate lateral projection of PP4, the hippocampus and fornix will actually curve medially and approximate the plane of PP3. Its representation in PP4 is a schematic emphasis of the anterior-posterior projection of the structure. The decussation of anterior (ac) and posterior (pc) commissures is indicated by brown squares at PP3/II and IV respectively. Ventricular system (LV) (black) is projected topologically within the 3-dimensional representation. Cortex: neocortex (gray), limbic cortex (yellow). nuclei: thalamus (th), caudate (cau), putamen-pallidum (Ln), amygdala (amy) (pink). White matter: radiata (beige); corpus callosum (cc) (red); internal capsule (IC). Cortical paralimbic structures: parahippocampal gyrus (PH); temporal pole (TP); frontoorbital cortex (FOC); frontomedial cortex (FMC). Gray limbic structures: limbic brain stem (LB); hypothalamus (Hyp); hippocampus (Hip); amygdala (Amy); septal area (sept); preoptic area (proa); nucleus accumbens septi (NAc); sublenticular extended amygdala and basal forebrain (SLEA/BF); insula (INS); habenula (Hb). Abbreviations: White matter: superior sagittal stratum (Ss), inferior sagittal stratum (Si), temporal sagittal stratum (St). Limbic fascicles shown in this figure: uncinate fasciculus (UF); cingulum bundle (CB); dorsal hippocampal commissure (dhc), fornix (fo), fimbria (fi); medial forebrain bundle (MFB); amygdalofugal projection (AFP).

Some of the first neuroimaging studies to identify neural activity in a subset of these brain regions during the processing of rewarding stimuli used monetary or drug rewards (23,37,250). In the double-blind cocaine vs. saline infusion

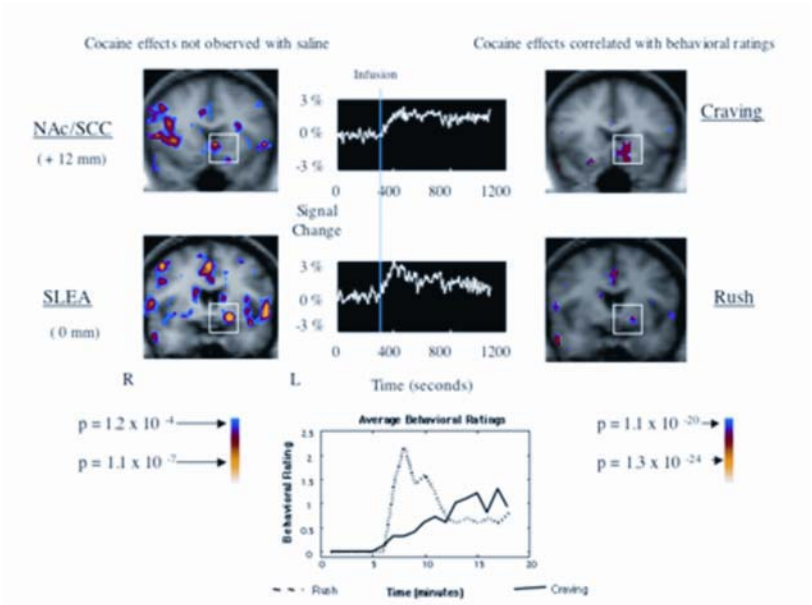


Figure 6. Nonparametric statistical maps in pseudocolor (with *p*-value coding bar), showing functional magnetic resonance imaging "activation," are juxtaposed on structural images (figure adapted with permission from Breiter et al. (37)). Activation represents brain signal related to blood flow and volume changes that are linked to changes in neural local field potentials. Images on the left show significant signal change in the nucleus accumbens/subcallosal cortex (NAc/SCC) and sublenticular extended amygdala (SLEA) to infusion of cocaine and not saline. These images are brain slices in the same orientation as the human face, and are 12 mm (NAc/SCC) and 0 mm (SLEA) anterior to a brain landmark, the anterior commissure. Signal time courses from the NAc/SCC and SLEA are graphed in the middle of the figure as percent signal change during the cocaine pre- and post-infusion intervals (infusion onset shown with a blue line). These signals were correlated with the average behavioral ratings for rush (euphoria and physiological experience of initial cocaine effects) and craving (motivational drive to obtain more cocaine) shown in a graph at the bottom of the figure. The statistical correlations of the behavioral ratings with the brain signal responses to cocaine are shown as statistical maps on the right of the figure.

study (37), multiple projection fields of the VT dopamine neurons were specifically targeted and visualized (Figure 6). As the study involved chronic cocaine-dependent subjects, the results correlating subjective reports of euphoria and craving (i.e., a mono-focused motivational state) to activation in reward circuitry, could not be separated from neuro-adaptations to subject drug abuse. A number of follow-up studies in healthy controls confirmed the initial findings with cocaine, using monetary reward, social reward in the form of beautiful faces, and thermal aversive stimuli. Together, this series of studies provided strong evidence for a generalized circuitry that processes stimuli with motivational salience (3,19,31,37,38). Modeling a game of chance, the monetary reward study incorporated principles from Kahneman and Tversky's prospect

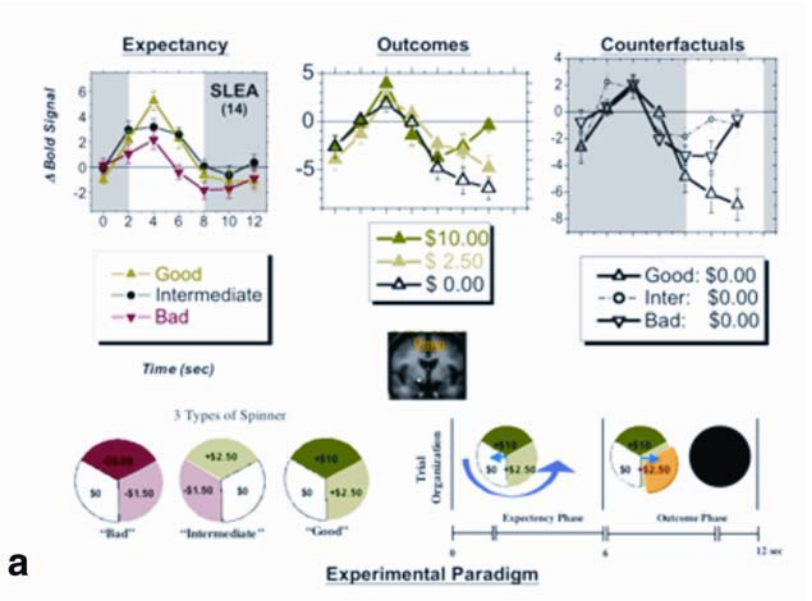


Figure 7a. An experimental design that applies the principles of prospect theory and decision affect theory (figure adapted with permission from (32,38)) utilizing three spinners in a game of chance. Organized as a series of single trials, the trial sequence started with presentation of one of these spinners, and continued with an arrow rotating on it. After six seconds, this rotating arrow would abruptly stop and the sector upon which it had landed would flash for 5.5 seconds, showing that the subject had won or lost that amount of money. With three spinners, each with three outcomes, this experiment asked which reward/aversion regions in the brain would process differential expectancy and/or outcome effects. With one outcome ($0) shared across spinners, it could explicitly also evaluate counterfactual comparison effects (171,170). The graphs at top display differential expectancy effects (left), differential outcome effects (middle), and counterfactual comparisons (right) from a region of signal change in the sublenticular extended amygdala (SLEA). The *x*-axis displays time in seconds, while the *y*-axis displays normalized the fMRI signal.

theory as well as Mellor's decision affect theory (131,170,171,254). Its results temporally segregated expectancy effects from outcomes in a number of reward regions (Figure 7a), and observed rank ordering of signal responses in a set of brain reward regions that reflected the differential expectancy conditions. Rank ordering of fMRI signal responses to differential monetary outcomes was also observed. Strong effects of expectancy on subsequent outcomes, or counterfactual comparisons (170,171), were measured for an outcome shared across expectancy conditions. The observation of effects from counterfactual comparisons is a fundamental experimental control supporting the reliability of the expectancy measures. At the high spatial resolution of this 3T fMRI study, some reward regions were activated solely by expectancy or outcome effects. Intriguingly, a few regions were involved with multiple functions and processed differential expectancies, outcomes, and counterfactual comparisons. Later studies using

Figure 7b. Overlap in NAc for expectancy responses. Expectancy of a cocaine infusion is shown in the upper panel from a study of double-blind, randomized, cocaine vs. saline infusions in cocaine dependent subjects (37,33). Expectancy of a monetary gain is shown in the lower panel from a study involving a game of chance in healthy controls (38). Note the close anatomic proximity for NAc signal changes during positive expectancy in the context of uncertainty for both experiments. Results are shown in the radiological orientation as pseudocolor statistical maps juxtaposed on coronal group structural images in gray tone.

monetary reward have significantly extended these observations of differential expectancy responses in some reward regions (137) and differential outcome effects (88). Other important studies with categorical rewards have shown segregation of expectancy and outcome effects, without counterfactual comparison effects (24,189). There appears to be a concordance of findings between different research groups, and overlap across studies within the same research group (e.g., such as similar expectancy findings in the NAc) for a monetary stimulus (38) and, retrospectively, for a cocaine infusion (33) (Figure 7b). This argues that different categories of reward are evaluated by a distributed set of reward regions, which can be functionally dissected.

Complementing the monetary reward studies, the study of social stimuli in the form of beautiful versus average faces addressed the issue of valuation and relative preference (see Figure 3). Incorporating a keypress paradigm, this study objectively quantified the reinforcement value of each stimulus by measuring the effort that experimental subjects expended to increase or decrease their viewing time of each face relative to a default viewing time (Figure 8a). For this procedure quantifying relative preferences, these face stimuli could be considered to be items in an economic bag of goods. The output of this keypress procedure was a representation of the utility function for each individual toward this set of items (Figure 8b). The results of this study argued that judgements of relative preference or utility were related to activity in a subset of reward regions.

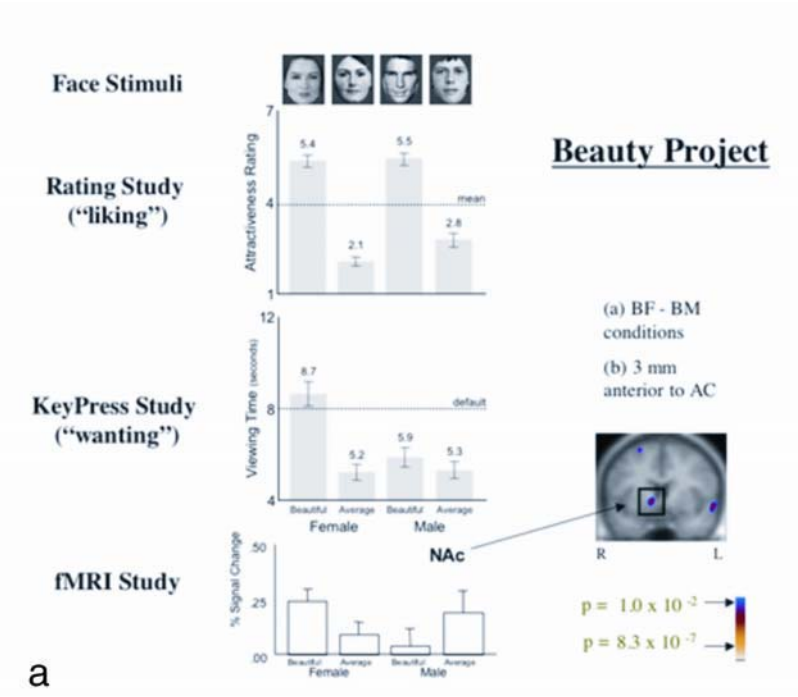


Figure 8a. Beauty Project. Behavioral and fMRI results regarding the viewing of beautiful vs. average faces adapted with permission from Aharon et al. (3). A sample of the four picture types used in these tasks (from left to right) is shown at top: beautiful female (BF), average female (AF), beautiful male (BM), and average male (AM). In the graph just below these sample face pictures, rating responses are shown for eight heterosexual males who rated picture attractiveness on a 1–7 scale for a randomized sequence of these pictures. The responses grouped themselves with tight standard deviations in the four categories illustrated at top. This process was interpreted as a "liking" response, whereas the keypress procedure (whose results are shown as the second graph down from the top) was interpreted as a "wanting" response (25). For the keypress procedure, a separate cohort of 15 heterosexual males performed a task where picture viewing time was a function of the number of their keypresses. Within each gender, the faces were always presented in a new random order, with beautiful and average faces intermixed (3). On the lowest graph, percent BOLD signal from the NAc for a third cohort of heterosexual males is shown for each face category relative to a fixation point baseline. Significant by a random effects analysis, the fMRI results in the NAc were driven by the response to the beautiful female and the beautiful male faces, and more closely approximated the "wanting" response rather than the "liking" response. On the right, a pseudocolor statistical map of signal collected during the beautiful female condition vs. the beautiful male condition (with *p*-value coding bar) is juxtaposed on a coronal group structural image in gray tone.

As a quantitative measure of relative preference, non-rewarding stimuli produced different regional signal profiles to rewarding stimuli (3), an observation that was further supported by a study using thermal pain (19) (Figure 9). This study of judgments of relative preference or utility along with the monetary study based on prospect theory (38) are early examples of the developing field

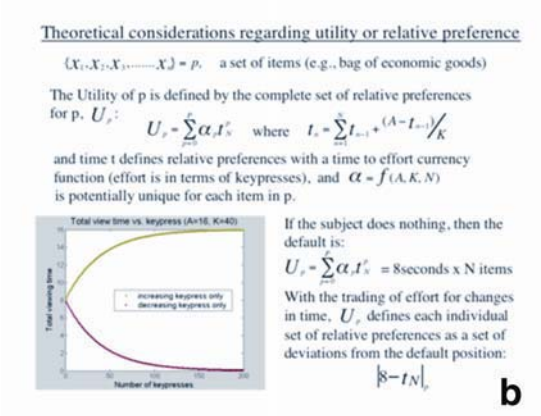


Figure 8b. Theoretical considerations regarding utility or relative preference. The output of this keypress procedure was a representation of the utility function for each individual toward this set of items. The objective was to determine their relative preferences for these items compared to the default position of not expending any effort to change a set viewing time. The keypress procedure quantified both the valence and the amount of the value that each item or face picture had relative to the default position of 8 seconds of viewing time.

of neuroeconomics (102). Overall, this set of studies in healthy controls observed that "classic" reward circuitry (including the NAc, SLEA, amygdala, VT, and GOb) processes a continuum between rewarding and aversive stimuli, with salient similarities and differences in regional activation (Figure 10a).

The segregation of neural systems that process aversive stimuli from those that process rewarding stimuli might be an artificial distinction (132). Comprised of subcortical gray structures, the "classic reward system" is activated by aversive stimuli such as thermal pain, expectancies of bad outcomes, and social stimuli that are not wanted (3,19,38). Comprised of paralimbic cortical and thalamic structures, "classic pain circuitry" (18,19,57,62,71,72,201,207,208,221, 247) is also activated by rewarding stimuli (18,19,57,62,71,72,201,207,208, 221,247) (Figure 10b). This commonality of activation patterns produced in healthy humans by stimuli with positive and negative outcomes (Figure 10b) argues that an extended set of subcortical gray matter and paralimbic cortical regions processes both rewarding and aversive information, and could be considered a generalized system (31).

A metaanalysis or general survey of neuroimaging studies presenting rewarding stimuli to humans suggests that an extended set of reward/aversion regions responds across multiple categories of rewarding stimuli (3,12,23,24, 27,28,33,38,44,83,87,88,133,137,155,188,189,239,250,257,270,272–274,277,279, 280) (Figure 11). This observation is complemented by animal data indicating

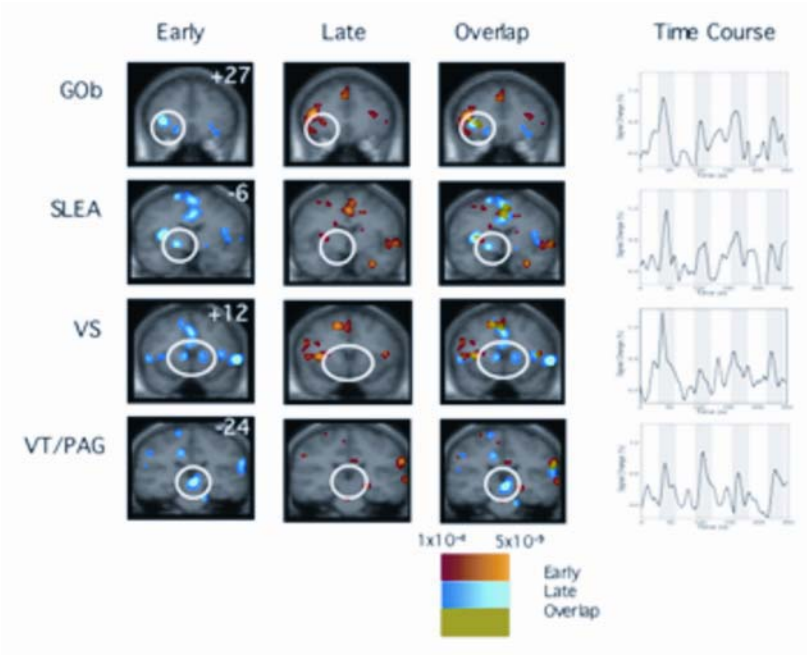


Figure 9. Pain study of reward circuitry in healthy control subjects. In this figure, adapted with permission from Becerra et al. (19), representative coronal slices containing "classic" reward circuitry [GOb, SLEA, ventral striatum (VS), ventral tegmentum/periacqueductal gray area (VT/PAG)] are segregated into early and late phases of BOLD signal change following the 46°C stimulus (left and middle columns). As in Figures 6, 7b, and 8a, the statistical maps are overlaid in pseudocolor on gray scale average structural maps. The right column of statistical maps shows the overlap (red) of early (yellow) and late (blue) phase activation. Time courses of % signal change vs. time are shown in the column at far right for each structure. To aid anatomic localization, the anterior-posterior coordinate in mm from the anterior commissure for each slice is shown in the far left column. For this figure, activated pixels are thresholded at $p < 5 \times 10^{-4}$.

that subsets of neurons may respond to one type of rewarding stimulus but not another (51,52). As some of these neuroimaging studies (Figure 11) only focused on select brain regions, such as the GOb, anterior cingulate cortex, or amygdala, or did not have the spatial resolution to observe a subset of subcortical regions, the relative prevalence of documented brain activity in some brain regions is overweighted. Although the majority of these studies involved a motor component for the experimental task, the bulk of activations reported did not involve regions associated with some aspect of motor control (i.e., dorsal caudate, putamen, globus pallidus, posterior cingulate gyrus, and thalamus). Furthermore, activation patterns in subcortical gray matter and paralimbic cortex were similar between experiments using the passive presentation of social/aesthetic stimuli, appetitive stimuli, and drug stimuli, and those using tasks that included motor performance. The experimental results summarized in Figure 11

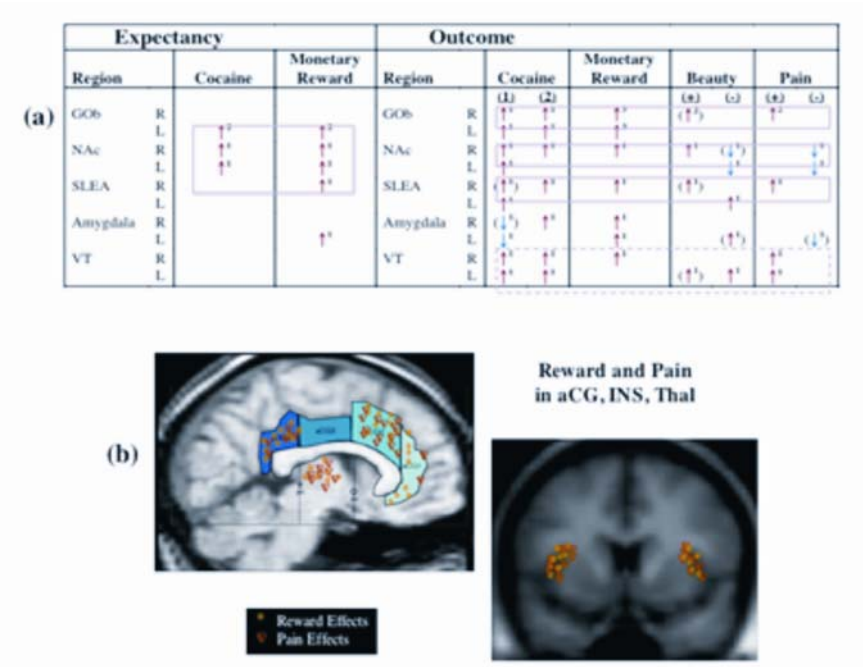


Figure 10. (**a**) Consolidated data from studies at one lab show common and divergent patterns of activation (3,19,33,37,38) regarding the analysis of expectancy, or of outcome. Two separate cocaine infusion studies are listed, as are positive and negative valuation results for the beautiful faces experiment and the thermal pain experiment. Red up-arrows symbolize positive signal changes while blue down-arrows stand for negative signal change. Numeric notation that is raised indicates more than one focus of signal change in that region, whereas brackets indicate the signal change was statistically subthreshold for that study. The right GOb, right NAc, right SLEA, and potentially the left VT, are observed during the outcome conditions for most of the experiments, while bilateral NAc and left GOb are observed in both studies with expectancy conditions. Table adapted with permission from Breiter and Gasic (32). (**b**) The gray tone structural images in the sagittal orientation (left), and in the coronal orientation (right; +6 mm anterior of the anterior commissure), juxtapose published human neuroimaging results in "classic" pain regions from rewarding and from painful stimuli (19). Other regions illustrated include the thalamus (Thal = left image between ac and pc), the cingulate cortex, and the anterior insula (INS = right image). The cingulate cortex is segmented into four units following the standardized methods of the MGH Center for Morphometric Analysis (158,175), and includes aCG1 and aCG2. Note the approximation of reported activation from stimuli of opposite valance. Figure adapted with permission from Breiter and Gasic (32).

indicate that "classic" reward circuitry (i.e., NAc, SLEA, amygdala, VT, GOb) is not uniquely involved with reward functions, given a proportionate number of activations are also reported in the rest of the paralimbic girdle.

Some minor caveats need to be considered when evaluating the literature around a generalized circuitry for the assessment of reward/aversion information. One caveat is that none of these reward studies controlled for relative valuations across stimulus categories, or controlled for the presence of a deficit state, with the exception of a study using chocolate stimuli and a study that

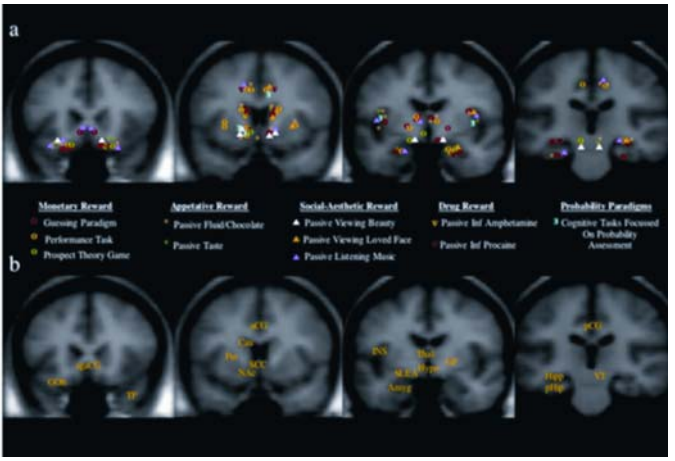


Figure 11. The bottom row of images indicates the anatomy of subcortical gray matter regions and paralimbic cortex. The top row of images consolidates reports of significant signal change for a number of distinct categories of rewarding stimuli in these regions. The colored symbols on the brain slices in the top row consolidate activation surveyed from 26 studies of reward function in healthy controls. These include four studies focused on appetitive reward with fruit juice, chocolate, or pleasant tastes. Ten studies utilized monetary reward (five with a guessing paradigm determining compensation, four with a performance task determining compensation, and one with a prospect theory based game of chance). Five studies focused on some aspect of social reward (two with beautiful faces, one with passive viewing of a loved face, and two with music stimuli). Five studies involved amphetamine or procaine reward, and two studies focused on a probabilistic paradigm. The gray tone structural images in the bottom row are coronal slices taken (left to right) +18, +6, –6, and –21 mm relative to the anterior commissure. In this diagram, subcortical gray matter implicated in the processing of reward and aversion input include the NAc (nucleus accumbens), Put (putamen), Cau (caudate), SCC (subcallosal cortex), Amyg (amygdala), SLEA (sublenticular extended amygdala), Hypo (hypothalamus), GP (globus pallidus), Thal (thalamus), Hipp (hippocampus), VT (ventral tegmentum). Components of the paralimbic girdle include: sgaCG (subgenual anterior cingulate gyrus), GOb (orbitofrontal cortex), aCG (anterior cingulate gyrus), pCG (posterior cingulate gyrus), INS (insula), pHip (parahippocampus), and TP (temporal pole). Abbreviations for anatomy follow the schema adapted from the Massachusetts General Hospital Center for Morphometric Analysis (37,38,158,175). Figure adapted with permission from Breiter and Gasic (32).

evaluated relative preferences with multiple categories of beautiful and average faces. At issue is the question of how to gauge the relative reward value of stimuli. Another issue is that all of the drugs infused into healthy controls have known global effects along with purported regional effects, making the association of regional activation to subjective reports of euphoria less certain. Lastly, most studies compiled in Figure 11 involved monetary reward, which is theorized to substitute for most other categories of reward (47,172). Despite the saliency of expectancy effects, most of these reports did not control for expectancy, and thus the results reflect a combination of expectancy and outcome effects (Figures 3 and 7). With these considerations in mind, there appears to be strong convergence between animal and human studies on the neural basis for reward/aversion (3,19,38,132).

The subcortical gray matter and paralimbic cortices that assess reward/aversion information as part of a theoretical iBM (Figures 2 and 3) also appear to be integrated into other processes for: (1) selection of objectives for fitness and (2) behavior. Perceptual inputs from multiple channels are processed through successive stages in unimodal association regions of the frontal, temporal and parietal cortices. This processing achieves more complex discrimination of features relevant for organizing behavior (Figure 3). Channel-specific information is conveyed to multimodal areas for intermodal integration in the neocortical heteromodal association areas. In turn, information relayed to paralimbic and limbic structures such as the cingulate gyrus, the insula, the orbitofrontal, frontomedial, parahippocampal and temporopolar cortices, as well as the amygdala, sublenticular extended amygdala, and hippocampus, is used for feature extraction and encoding. Feature extraction and integration for probability determination further involves the orbitofrontal cortex, amygdala, and cingulate gyrus, with the ventral tegmentum and nucleus accumbens septi. Contingent probability assessments require extensive working memory and attentional resources, and thus will be integrated with activity in limbic brain stem structures, and heteromodal frontal, parietal, and limbic cortices in a modality- and domain-specific way. These probability assessments are necessary for making predictions about future homeostatic needs or potential deficit states, and, accordingly, will be involved in selecting objectives for fitness over time (Figures 2–4). They can also be important for prediction of effects of social interactions on homeostatic needs, and for focusing the majority of behavioral output on social functions. Valuation and probability computations around possible goal-objects and events, their combination as outcomes, and subsequent counterfactual comparisons are further integrated with information regarding the costs of changing body position in space, potential risks to action and inaction, and discounted benefits of other consummatory opportunities. Subcortical gray matter and paralimbic cortices function in concert with multiple corticothalamic circuits for determination of physical plans and actual behavior. For instance, the cingulate gyrus has extensive involvement with the alteration of attention for motivational state. Other paralimbic cortices and ventral striatal regions interact with the supplementary motor and premotor frontal areas in preparation for executive behavior and directed action appropriate to environmental and internal factors (173,180,184,195).

The circuits that process reward/aversion information as an iBM, and interact with other brain regions to produce behavior and to determine objectives optimizing fitness, are fundamental for normal emotion function, as well as its malfunction. The systems biology of reward/aversion assessment, in line with the systems biology of other subprocesses such as attention or memory, represents an interface across which the genome, epigenome, and environment interact. The interaction of the genome, epigenome, and environment across this interface (see Figure 14) determines the set of all possible behaviors. This inter-

face can be densely sampled with neuroimaging procedures to characterize a particular behavioral function in a set of individuals (Figure 10a), resulting in a quantitative representation of the neural processes necessary for that behavioral function (e.g., the utility function relating to a set of economic goods). If multiple samplings are obtained in each individual, covering defined subprocesses, these functional (and structural) measures will define a complex set of physiological and/or mechanistic interrelationships. These interrelationships can be grouped into functionally related clusters, or systems biology maps, as is done with cardiovascular function to produce vascular, heart, renal, endocrine, and morphometric clusters. The physiological and/or mechanistic relationships within such clusters, or systems biology maps, can be defined as quantitative phenotypes. These quantitative phenotypes can be subdivided into sets of phenotypes with different contingent probabilities for susceptibility to illness/malfunction, or resistance to illness/malfunction.

5. IMPLICATIONS OF REWARD/AVERSION NEUROIMAGING FOR PSYCHIATRIC ILLNESS

Traditionally, major psychiatric disorders have been categorized by clusters of patient-based reports of symptoms and behaviors observed in patients. This phenomenological description of categorical outward signs produced the nosology of illness based on exophenotypes that is the American Psychiatric Associations Diagnostic Statistical Manual (DSM) (5). Neuroscientists have recently begun to suggest approaches to replace current symptom-based characterizations of illness, or exophenotypes, using a nosology based on genes, molecules, neuronal organelles, and specific neural systems (60,61). Such a nosology would potentially develop a unitary basis for psychiatric and neurological illnesses. A nosology based on descriptions of brain structure and function would also have to consider the impact of time, as many of these neuropsychiatric diseases appear to have a neurodevelopmental and/or neurodegenerative component (34,150,245). In this section, we will examine the current evidence for a neural systems approach, focusing on alterations in reward/aversion function, that might objectively categorize the major (i.e., Axis I disorders per DSM) neuropsychiatric illnesses.

Over the past decade, studies of neuropsychiatric illness with positron emission tomography, single photon emission computed tomography, magnetoencephalography, magnetic resonance spectroscopy, morphometric MRI, and fMRI have begun to suggest that neuropsychiatric illnesses might be distinguished by alterations in circuitry structure and function (42,68). So far, no research has focused on classifying the major categories of psychiatric illness on the basis of their patterns of circuitry function or structural differences using a unitary set of experimental paradigms or structural imaging protocols. Meta-

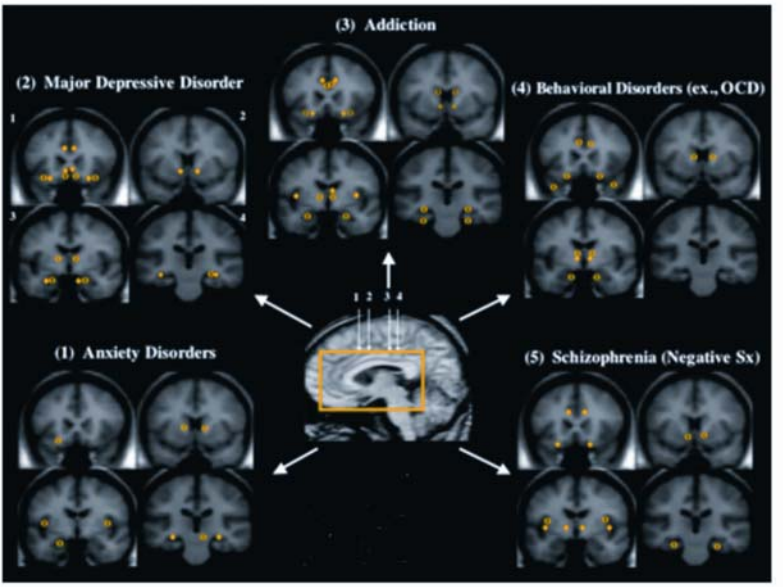


Figure 12. The same structural scans shown in Figure 11 are displayed here, grouped two-by-two, and numbered to correspond with the anterior-to-posterior orientation. Groupings (1)–(5) are placed around a central sagittal slice indicating the general location of each coronal slice relative to a yellow rectangle around brain regions hypothesized to mediate reward/aversion functions. These groupings represent partial consolidations of reports in the neuroimaging literature comparing patient groups to healthy controls. Neuroimaging measures include (a) structural differences, (b) blood flow, receptor binding, or resting metabolism, (c) blood flow or metabolic responses to normative stimuli (i.e., pictures of masked emotional faces presented to subjects with post-traumatic stress disorder), or (d) magnetic resonance spectroscopy measures. As in Figure 11, regions with functional differences (b,c above) between subjects and healthy controls are noted with an "O" symbol. Regions with differences in regional morphology, volume, or spectroscopy signal from healthy baselines (a,d above) are noted with a diamond. Regions with an asterisk indicate that a set of studies found differences between patients and healthy controls for a large region, but a more recent study with improved spatial resolution noted an effect for the same experimental paradigm (i.e., drug infusions) localized to a specific subregion (i.e., the NAc vs. the basal ganglia). This consolidation of psychiatric neuroimaging findings suggests that neuropsychiatric illness may in the future be objectively diagnosed by use of circuitry-based measures. Figure adapted with permission from Breiter and Gasic (32).

analysis of the neuropsychiatric imaging literature, with a focus on brain systems that collectively process reward/aversion information (Figure 12), suggests that such an approach could segregate major categories such as the anxiety disorders, major depressive disorder, addiction, the behavioral disorders, and schizophrenia. The metaanalysis in Figure 12 compiled studies comparing patients to unaffected controls for each of the major categories of neuropsychiatric illnesses on the basis of (a) patterns of resting brain metabolism, blood flow, or

receptor binding; (b) functional differences in responses to normative stimuli (i.e., pictures of emotional faces that are rapidly masked in an effort to present them subconsciously); (c) volumes of brain structure,; or (d) quantifiable chemical signatures of neuronal integrity. In Figure 12, anxiety disorders (1), major depressive disorder (2), and addiction (3) are displayed along a potential continuum along the left side of the figure. Behavioral disorders (4) and schizophrenia (5) are segregated on the right side of the figure. The studies compiled for anxiety disorder (1) focused on post-traumatic stress disorder, social phobia, and simple phobia; this compilation included symptom provocation studies that were compared across illness category but not to a healthy control group (211,212, 225,226). The studies compiled for major depression (2) were focused on (a) recurrent depression with strong familial loading (i.e., familial pure depressive disorder), (b) primary depression with and without obsessive-compulsive disorder and without manifested familial connections, and (c) primary and secondary depression in older subjects studied postmortem (30,39,40,81–84,90,141,161, 181,193,209,210,219,223). These studies are compiled together in Figure 12, and segregated for 1–3 in Figure 13. Multiple stimulant addictions were grouped for addiction (3) (55,94,100,104,107,151,256–259). The studies compiled for behavioral disorders (4) were grouped following more recent suggestions that place obsessive-compulsive disorder (OCD) on a continuum with tics (Tourette's), attention-deficit hyperactivity disorder, and other behavioral problems such as conduct disorder, oppositional behavior, and learning disabilities (130). For this particular metaanalysis, the focus was on studies of OCD (13,14,36,92,108,185,199,222,223,244). Lastly, the studies compiled for schizophrenia (5) used subjects who were not actively psychotic, and included studies with relevance to negative symptomatology such as amotivation, avolition, and anhedonia (8,63,103,115,162,163).

In sum, the studies compiled for anxiety disorder, major depressive disorder, addiction, behavioral disorders, and schizophrenia reveal differences between patient and control groups primarily in the subcortical gray matter and paralimbic cortices illustrated in Figure 11. These brain regions mediate subprocesses such as reward/aversion assessment, which are fundamental to emotional function and the generation of motivated behavior (1,7,23,33,38,73, 147,174,215,227,238,263,266). Dysfunction of these brain regions has been previously hypothesized to be responsible for a variety of psychiatric symptoms such as olfactory or gustatory hallucinations, autonomic discharges, episodic amnesias, depersonalizations, avolition (or lack of motivation), abulia (or lack of will), anaffectiveness (or affective flattening), asociality, as well as delusion, hallucinations, thought disorder, and bizarre or disorganized behavior (276,277).

A circuitry-based nosology for neuropsychiatric illnesses would facilitate the identification of endophenotypes for these disorders (4,81,101,149,161), particularly as morphometric MRI studies allude to the heritability of structural alterations (179,248). Disorders such as depression have been hypothesized to

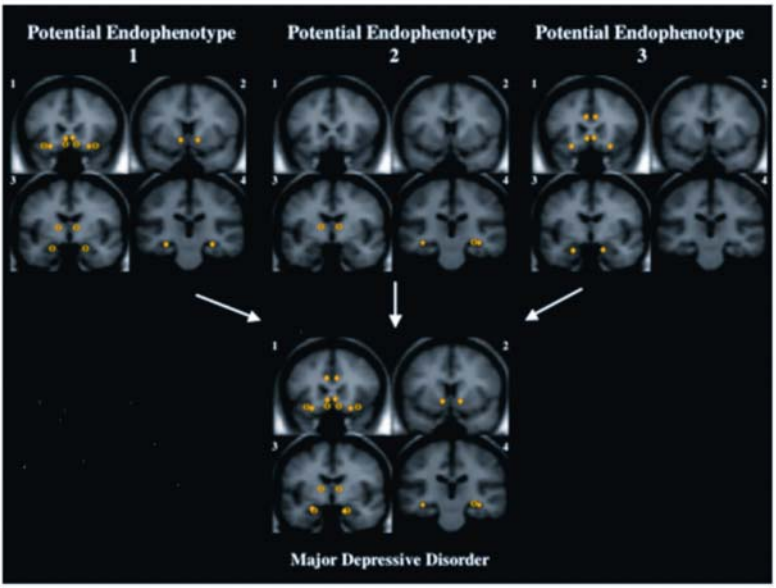


Figure 13. The groupings of structural images in gray tone are the same as in Figure 12, and display changes in the structure, function, or morphology of subcortical gray matter and paralimbic cortices for three putative categories of depression. "Putative endophenotype variation a" grouped studies focused on recurrent depression with strong familiality. "Putative endophenotype variation b" grouped studies of primary depression with and without obsessive-compulsive features and without manifested familial connections. "Putative endophenotype variation c" grouped studies of primary and secondary depression in older subjects who were studied postmortem (see text for references). Regions with differences in resting brain metabolism from healthy baselines are noted with an "O" symbol. Regions with differences in regional morphology or volume from healthy baselines are noted with a diamond. These studies are suggestive of the potential for circuitry-based endophenotypes for major depressive disorder, and point to a strong focus on the generalized reward/aversion system for circuitry-based alterations characterizing major depressive disorder. Such circuitry-based subtypes could facilitate treatment planning in the future. Figure adapted with permission from Breiter and Gasic (32).

have circuitry-based phenotypes, on the basis of evidence that subtyping of depressed patients appears to be important for reducing variability in the patterns of regional activity observed with functional imaging (81,83). At least three putative phenotypes are observed when findings are grouped from structural and functional imaging studies of depression with large cohort sizes or replicated findings (Figure 13). Alluding to the existence of circuit-based phenotypes, which may be heritable markers or endophenotypes, a number of neuroimaging studies comparing individuals with neuropsychiatric disorders to unaffected controls have documented qualitative differences (presence or absence of a regional signal), and quantitative differences (numeric alterations in the mean or median signal) (36,43,82,93,115,162,163). The presence of circuitry-based

endophenotypes would have implications for the nested genetic, molecular, subcellular, and cellular mechanisms that produce them (53,86,112,125,161).

To develop a unitary nosology for neuropsychiatric illness based on functional and structural circuitry measures, it is likely that more than one or two motivation subprocesses will need to be studied. Strong evidence exists that disorders such as schizophrenia involve abnormalities in multiple cortical regions and functional domains separate from the circuitry implicated in the subprocess of reward/aversion assessment (220,235). The classification of neuropsychiatric illnesses using abnormalities in the circuitry for reward/aversion assessment may be one dimension of a multidimensional schema for circuitry-based (or systems-based) characterization of neuropsychiatric illnesses. Per Figure 4, other dimensions might include processes involved with sensory perception (29), or processes for attention and memory (79,231,261). At the systems biology interface between genome, epigenome, and environment (Figure 14), a substantial combination of brain subprocesses involved with motivated behavior may be dysfunctional in concert with that of reward/aversion assessment to produce neuropsychiatric signs and symptoms.

6. LINKING THE DISTRIBUTED NEURAL GROUPS PROCESSING REWARD/AVERSION INFORMATION TO THE GENE NETWORKS THAT ESTABLISH AND MODULATE THEIR FUNCTION

An organism arises from a complex set of interactions between its genome-epigenome and its environment. Disease states can be conceptualized as an organism's failure to adapt effectively to its environment. A new nosology for psychiatric illness could consider these adaptation failures at multiple spatio-temporal scales of brain function beginning with common circuitry alterations, which represent adaptation failures on an immediate timescale (Figures 12 and 13). Alterations in genome-epigenome, molecular machinery, and cellular function, which represent adaptation failures on a broader timescale, would also be relevant in this classification system. Although deleterious to the individual, these changes, which appear at the genetic, molecular, and organelle levels, can be viewed as the byproduct of the "capacitors" and "gain-controls" responsible for species-wide adaptations to a changing environment over time (17,135, 205,217,252,260). Although circuitry and molecular genetic functions are interrelated, systems-level descriptors (e.g., the reward/aversion systems described above) and molecular genetic-level descriptors will both be essential components for the characterization of all neuropsychiatric illnesses.

A view of how future characterization of psychiatric diseases might incorporate molecular-genetic descriptors can be obtained from characterizations of other less prevalent neuropsychiatric diseases with neurodegenerative (2) or

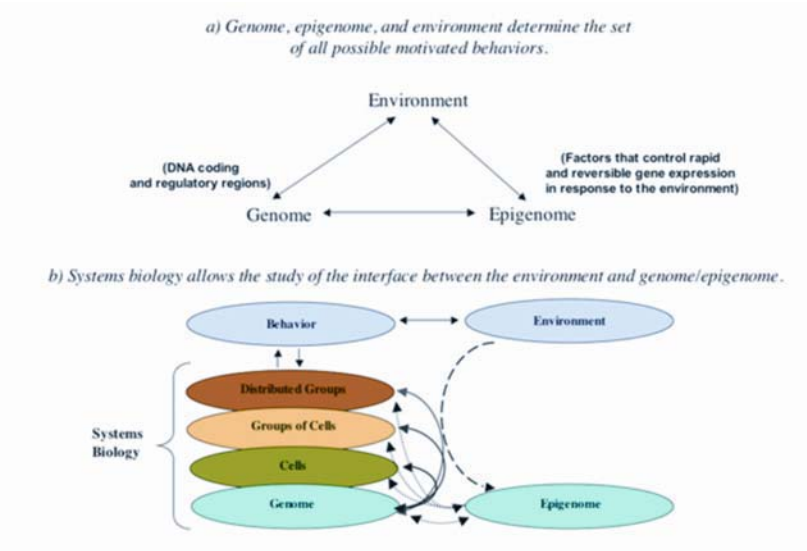


Figure 14. The top diagram emphasizes the tripartite division of influences that shape an organism, namely the genome, epigenome, and environment. The set of all possible behaviors for an organism (i.e., communication) is determined by these three influences, although the specific sequence of output is not. The internal environment produced by the genome/epigenome (bottom) produces the putative spatiotemporal scales of brain function. In this case, activity at the level of distributed groups of cells, local networks or groups of cells, and individual neurons modulate the function of the genome/epigenome, and activity at the level of the genome/epigenome significantly modulates the function of each of the spatiotemporal scales of function that embed it. The linked spatiotemporal scales of brain function are again distinct from observed behavior in the outside world (i.e., exophenotype) and will have a stronger connection, as endophenotypes observable with neuroimaging and other measurement systems of brain function, with the genome/epigenome. The scale of distributed groups of cells produces behavior, and accordingly serves as an interface between the environment and genome/epigenome.

neurodevelopmental etiologies (34,150,245). For instance, prion diseases and many of the neurodegenerative diseases with patterns of mixed Mendelian and/or non-Mendelian inheritance (i.e., Alzheimer's disease, Parkinson's disease, frontotemporal dementias, Huntington's disease) have a strong component of their etiology from two processes. One process involves the dysfunction and/or cell death of a subset of brain neurons/glia that express an aberrant gene product, whereas a second process involves the non cell-autonomous consequences (e.g., altered homeostasis) of this cellular vulnerability. These diseases result from an inability to maintain mutant proteins: (a) in a properly folded and/or functional state, (b) in their proper subcellular organelles, or (c) at appropriate steady-state levels to prevent their gain-of-function role (59,64,85,111,118,124,183,191,

232,262). Ultimately, the energy state of the cell and/or mitochondrial function may become impaired and normal transport processes may likewise be affected.

In model organisms, protection against degenerative disease can be conferred by over-expression of some members of a family of heat-shock proteins that keep proteins in a folded state, and are upregulated during cellular stress conditions (152,194,236). Aging causes these cellular defense proteins to decline, possibly heralding the onset of neurodegenerative disease whose prevalence increases with age (152,194,236; see also this volume, Part III, chapter 7.3, by Lipsitz). Molecular systems that keep proteins in a folded state serve as "capacitors" for cellular evolution (135,205,217,252).

Heritable alterations in gene expression that do not rely on coding or regulatory polymorphisms in DNA sequences (e.g., methylation of DNA bases) but on (a) parental origin of the DNA (epigenetic modifications such as imprinting), (b) allele specific gene expression not dependent on obvious imprinting, and (c) variations in gene copy numbers (127,128,136,230) may also contribute to human variation and neuropsychiatric disease. For example, Down's syndrome, Turner's syndrome, and Praeder-Willi and Angelman syndromes are neuropsychiatric diseases that can be caused by alterations in gene dosage and/or imprinting rather than by mutations in the DNA itself (17,182,218,245). Such observations have led to a "rheostat" model for gene expression (17), which acts as a gain-control to allow rapid and reversible attenuation of gene expression (over generations and during development). Together, these mechanisms regulating gene expression may contribute to the spectrum nature of psychiatric diseases (e.g., Autism-Asperger's syndrome).

The molecules that serve as the putative gain-control and capacitors for producing adaptive phenotypic variation function in a substantial and stepwise fashion rather than an incremental and progressive one. Variations in both systems may be present in neuropsychiatric diseases such as Rett syndrome (6,232,262,269). The molecular genetic basis of such neuropsychiatric diseases may be the outcome of evolutionary events that strike a delicate balance between minimizing deleterious mutations while allowing phenotypic variations that are adaptive to a species in a changing environment (17,135,205,217, 233,252,260).

These genetic variations, which may be adaptive or maladaptive in a changing environment, have the potential to influence brain function at a number of spatiotemporal scales (Figure 15). The parsimonious description of scales of brain function and their embedding remains a topic of active discussion (see 56,95). There is also an open question of whether the dynamic principles governing information processing at one level of organization are applicable to other levels of organization (i.e., neural scale invariance) (242). For at least one brain region, the NAc, a qualitative similarity is noted between reports of transcription factor cAMP response element-binding protein (CREB) phosphorylation in response to aversive and rewarding stimuli (11,278), and the signal

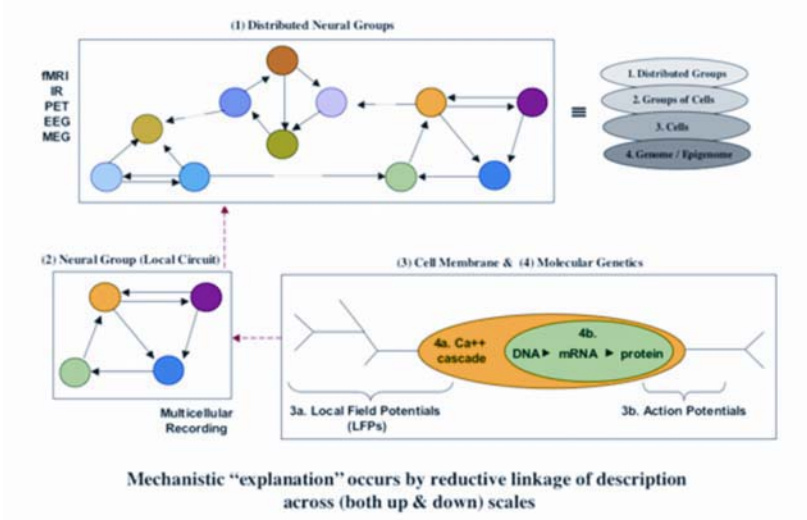


Figure 15. Systems biology and nested scales of function. Brain processes can be analogized to a set of nested scales of function as an approximation. The genome/epigenome is nested in cells (neural and glial). These cells are nested in neural groups as local circuits. These neural groups are nested in sets of interconnected groups distributed across the brain, and modulated by monoaminergic and hormonal systems. The scale of distributed neural groups can be sampled using tomographic imaging modalities such as fMRI and PET. Multicellular recording techniques can be used to sample local circuits or neural groups, comprised of excitatory and inhibitory synapses, axonal and dendro-dendritic circuits. The individual cell, with its intracellular signaling and surface receptors, can be characterized by measures of local field potential and sequences of action potentials. Across these scales of organization, reductionistic explanation of experimental observation has to occur both from "top–down" and "bottom–up" to be self-sufficient. Given the nesting of scales, dense sampling of one scale of brain function will reflect processes at the other scales (see Figure 10). Figure adapted with permission from Breiter and Gasic (32).

representing distributed group function observed with fMRI to similar aversive and rewarding stimuli (19,37,38). Reverse-engineering how activity is linked across levels of brain organization will have implications for reductive understanding of health and disease. Circuitry-based endophenotypes that have larger effect sizes, and are more reliable than behavioral phenotypes, should enable researchers to constrain future genetic association and linkage studies for neuropsychiatric illnesses (Figure 16). Such an integrative neuroscience approach ("top–down"; systems to genes) was utilized to find an EEG-based endophenotype in individuals susceptible to alcohol dependence, forging an association with a locus that contains a subunit of the $GABA_A$ receptor (202,214,264).

The linkage of systems level measures to molecular genetic level descriptors assumes that the probability of illness manifestation will be related to (a) the probability associated with having a specific allele(s) at a particular locus (loci), (b) the probability of having a particular endophenotype, and (c) the probability

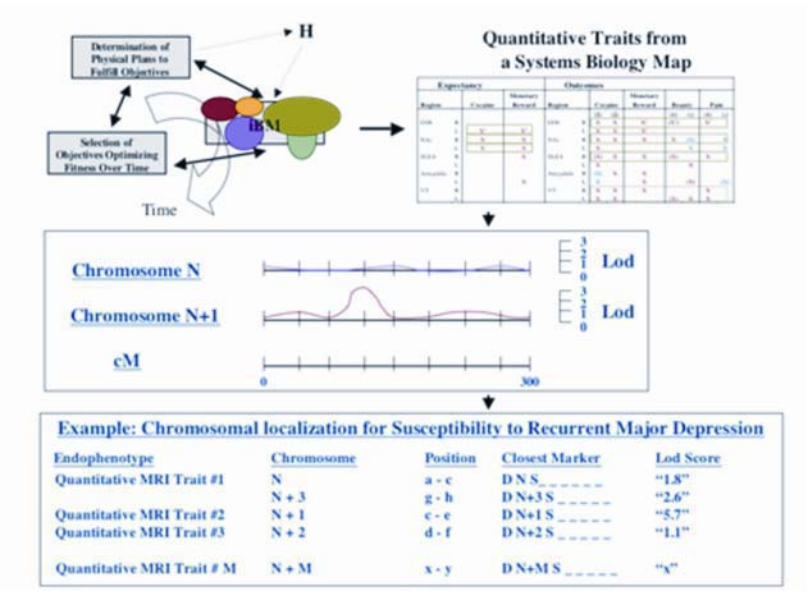


Figure 16. This schematic illustrates a possible "top–down" approach for identifying genes associated with a susceptibility or resistance to major depressive disorder. Overlapping sampling of circuitry processing reward/aversion input (cartoon in top left) from families with depression, could be used to produce a systems biology map (cartoon top right). Disease susceptibility would be defined by continuous quantitative traits measured from systems biology (as with MRI), and could be used to perform a total genome scan and a multipoint linkage analysis using a variance component approach (for quantitative and potential qualitative traits). Gene identification could then be achieved via analysis of microsatellite repeats and SNP markers. Figure adapted with permission from Breiter and Gasic (36).

of having a particular set of epigenetic elements (e.g., this might be expressed as P (illness) $\approx$ P (allele, locus) $\times$ P (Endophenotype N, t) $\times$ P (Epigenome, t)). Epigenetic elements appear to be species-specific (260), and may explain significant differences in phenotypes between species that otherwise have 99% sequence similarity (271). Until specific genetic variations are identified, phylogenetically lower animal species may only serve as course models for the complex neuropsychiatric diseases.

Epigenetic contributions may partly explain the difficulties underlying genetic linkage and association studies for these illnesses. Coarse clinical phenotypes and dichotomous behavioral distinctions rather than quantitative markers (endophenotypes) to cluster subjects is another confounding aspect for these studies. Activity in a variable number of distinct, distributed neural groups may yield multiple endophenotypes for an illness, yet produce indistinguishable symptom/sign clusters. The spatiotemporal scale of distributed cell groups is

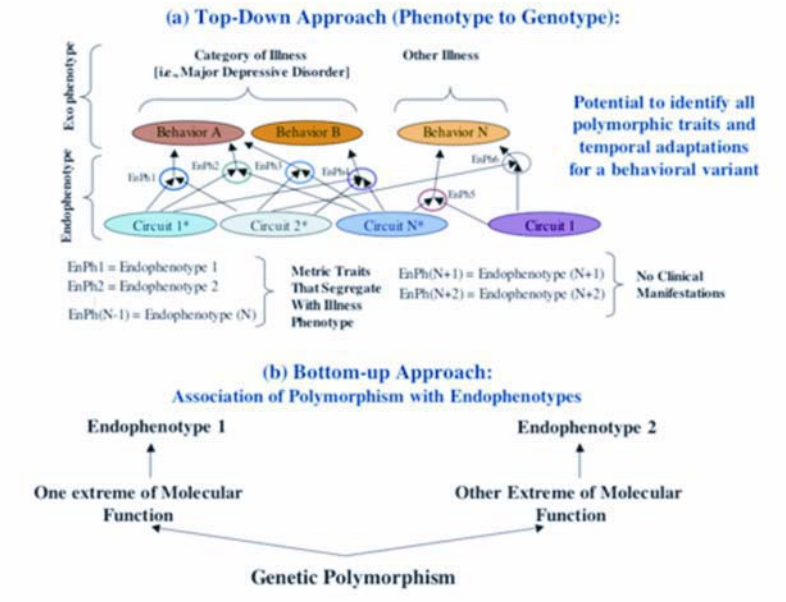


Figure 17. (**a**) Attempts at a "top–down" approach to integrative neuroscience have frequently started from the delineation of behaviorally defined exophenotypes, which are theoretically related to circuitry-based phenotypes (endophenotypes). Illness category can stand in for any number of American Psychiatric Associations Diagnostic Statistical Manual Axis I neuropsychiatric disorders, such as subtypes of major depressive disorder, or cocaine abuse and dependence. Altered function in a distributed set of neural groups (referred to in the figure as "circuits") is symbolized by an asterisk after the circuit number. This altered function may include diminished or increased circuitry activity, or substitution of an alternative circuitry to fulfill a functional deficit. There may be a number of altered functions or metric traits, determined by altered circuitry performance, which determine a particular neuropsychiatric disorder. This is highly likely given the use of multiple signs and symptoms currently used to define neuropsychiatric exophenotypes using the American Psychiatric Associations Diagnostic Statistical Manual (5). Given the embedding of scales of brain function, "top–down" approaches starting from continuous quantitative measures of systems biology, would, with the appropriate subject sample size, have the potential to identify all polymorphic traits and temporal adaptations for a behavioral variant. (**b**) "Bottom–up" approaches evaluate one genetic polymorphism at a time to determine how it leads to an altered profile of circuitry function.

responsible for generating behavior, but it is heavily dependent on genetic/epigenetic function. Weinberger and colleagues have demonstrated with fMRI that genetic variations in COMT and 5HT transporter are correlated with fMRI signal changes in human amygdala and prefrontal cortex, respectively (86,112). These types of studies represent a "bottom–up" approach to complement findings from the "top–down" approach (Figure 17). In some diseases, such as Huntington's, a single major disease locus may be enough to produce the endophenotypes and exophenotypes that characterize the illness. In contrast, oligogenetic and polygenic diseases (16), such as Parkinson's disease and most

neuropsychiatric illnesses, appear to involve more than one genetic locus. In such cases, future genome-wide association studies using circuit-based endophenotypes will have to demonstrate that variations at multiple loci (when quantitative trait loci become quantitative trait nucleotides), which produce alterations in gene dosage or allele specific expression, are both necessary and sufficient to produce the alterations in the functional subprocesses and their mediating neurocircuitry.

7. ACKNOWLEDGMENTS

Work on this was supported by funding from the National Institute of Drug Abuse (grants 14118 and 09467), the Office of National Drug Control Policy, the Counterdrug Technology Assessment Center (ONDCP-CTAC), and the Massachusetts General Hospital Department of Radiology. This work was further supported, in part, by the National Center for Research Resources (P41RR14075), the Mental Illness and Neuroscience Discovery (MIND) Institute, and the Division on Addictions, Harvard Medical School. We thank Byoung Woo Kim and Joe Mandeville for their technical consultation, and James Howard for his assistance and editorial input.

8. REFERENCES

1. Adolphs R. 2003. Cognitive neuroscience of human social behaviour. *Nature Rev Neurosci* **4**:165–178.
2. Aguzzi A, Haass C. 2003. Games played by rogue proteins in prion disorders and Alzheimer's disease. *Science* **302**:814–818.
3. Aharon I, Etcoff N, Ariely D, Chabris CF, O'Connor E, Breiter HC. 2001. Beautiful faces have variable reward value: fMRI and behavioral evidence. *Neuron* **32**:537–551.
4. Almasy L, Blangero J. 2001. Endophenotypes as quantitative risk factors for psychiatric disease: rationale and study design. *Am J Med Genet* **105**:42–44.
5. American Psychiatric Association. 1994. *DSM-IV: diagnostic and statistical manual of mental disorders*, 4th ed. rev. American Psychiatric Association, Washington, DC.
6. Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY. 1999. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature Genet* **23**:185–188.
7. Anderson AK, Phelps EA. 2002. Is the human amygdala critical for the subjective experience of emotion? Evidence of intact dispositional affect in patients with amygdala lesions. *J Cogn Neurosci* **14**:709–720.
8. Andreasen NC, Arndt S, Swayze II V, Cizadlo T, Flaum M, O'Leary D, Ehrhardt JC, Yuh WT. 1994. Thalamic abnormalities in schizophrenia visualized through magnetic resonance image averaging. *Science* **266**:294–298.
9. Aquinas T. 1993. Summa theologiae: feelings. In *Aquinas: selected philosophical writings*. pp. 156–168. Ed. T McDermott. Oxford UP, New York.
10. Bandettini PA, Wong EC, Hinks RS, Tikofsky RS, Hyde JS. 1992. Time course EPI of human brain function during task activation. *Magn Reson Med* **25**:390–397.

11. Barrot M, Olivier JDA, Perrotti LI, DiLeone RJ, Berton O, Eisch AJ, Impey S, Storm DR, Neve RL, Yin JC, Zachariou V, Nestler EJ. 2002. CREB activity in the nucleus accumbens shell controls gating of behavioral responses to emotional stimuli. *PNAS* **99**:11435–11440.
12. Bartels A, Zeki S. 2000. The neural basis of romantic love. *NeuroReport* **11**:3829–3834.
13. Baxter Jr LR, Phelps ME, Mazziotta JC, Guze BH, Schwartz JM, Selin CE. 1987. Local cerebral glucose metabolic rates in obsessive-compulsive disorder: a comparison with rates in unipolar depression and in normal controls. *Arch Gen Psychiatry* **44**:211–218.
14. Baxter Jr LR, Schwartz JM, Mazziotta JC, Phelps ME, Pahl JJ, Guze BH, Fairbanks L. 1988. Cerebral glucose metabolic rates in nondepressed patients with obsessive-compulsive disorder. *Am J Psychiatry* **145**:1560–1563.
15. Beauchamp MS, Lee KE, Haxby JV, Martin A. 2002. Parallel visual motion processing streams for manipulable objects and human movements. *Neuron* **34**:149–159.
16. Beaudet AL, Scriver CR, Sly WS, Valle D. 2001. Genetics, biochemistry, and molecular bases of variant human phenotypes. In *The metabolic and molecular basis of inherited disease*, pp. 3–44. Ed. CR Scriver, AR Beaudet, WS Sly, D Valle. McGraw-Hill, New York.
17. Beaudet AL, Jiang YH. 2002. A rheostat model for a rapid and reversible form of imprinting-dependent evolution. *Am J Hum Genet* **70**:1389–1397.
18. Becerra LR, Breiter HC, Stojanovic M, Fishman S, Edwards A, Comite AR, Gonzalez RG, Borsook D. 1999. Human brain activation under controlled thermal stimulation and habituation to noxious heat: an fMRI study. *Magn Reson Med* **41**:1044–1057.
19. Becerra L, Breiter HC, Wise R, Gonzalez RG, Borsook D. 2001. Reward circuitry activation by noxious thermal stimuli. *Neuron* **32**:927–946.
20. Belliveau JW, Kennedy DN, McKinstry RC, Buchbinder BR, Weisskoff RM, Cohen MS, Vevea JM, Brady TJ, Rosen BR. 1991. Functional mapping of the human visual cortex by magnetic resonance imaging. *Science* **254**:716–719.
21. Ben-Shahar Y, Robichon A, Sokolowski MB, Robinson GE. 2002. Influence of gene action across different time scales on behavior. *Science* **296**:741–744.
22. Bentham J. 1996. *An introduction to the principles of morals and legislation*. Ed. JH Burns, HLA Hart. Clarendon Press, Oxford.
23. Berns GS, Cohen JD, Mintun MA. 1997. Brain regions responsive to novelty in the absence of awareness. *Science* **276**:1272–1275.
24. Berns GS, McClure SM, Pagnoni G, Montague PR. 2001. Predictability modulates human brain response to reward. *J Neurosci* **21**:2793–2798.
25. Berridge KC. 1996. Food reward: brain substrates of wanting and liking. *Neurosci Behav Revs* **20**(1):1–25.
26. Blazquez PM, Fujii N, Kojima J, Graybiel AM. 2002. A network representation of response probability in the striatum. *Neuron* **33**:973–982.
27. Blood AJ, Zatorre RJ. 2001. Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proc Natl Acad Sci USA* **98**:11818–11823.
28. Blood J, Zatorre RJ, Bermudez P, Evans AE. 1999. Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. *Nature Neurosci* **2**:382–387.
29. Bonhomme V, Fiset P, Meuret P, Backman S, Plourde G, Paus T, Bushnell MC, Evans AC. 2001. Propofol anesthesia and cerebral blood flow changes elicited by vibrotactile stimulation: a positron emission tomography study. *J Neurophysiol* **85**:1299–1308.
30. Botteron KN, Raichle ME, Drevets WC, Heath AC, Todd RD. 2002. Volumetric reduction in left subgenual prefrontal cortex in early onset depression. *Biol Psychiatry* **51**:342–344.
31. Breiter HC. 1999. The generalized human circuitry of brain reward and its functional dissection. In *Proc. Workshop on the neural mechanisms of addiction*, Vol. 103, pp. 51–52. The Juan March Foundation Center for International Meetings on Biology, Madrid, December.
32. Breiter HC, Gasic GP. 2004. A general circuitry processing reward/aversion information and its implications for neuropsychiatric illness. In *The cognitive neurosciences*, 3rd ed, pp. 1043–1065. Ed. M Gazzaniga, editor. MIT Press, Cambridge.

33. Breiter HC, Rosen BR. 1999. Functional magnetic resonance imaging of brain reward circuitry in the human. *Ann NY Acad Sci* **877**:523–547.
34. Breiter HC, Filipek PA, Kennedy DN, Baer L, Pitcher DA, Olivares MJ, Renshaw PF, Caviness Jr VS. 1994. Retrocallosal white matter abnormalities in patients with obsessive-compulsive disorder. *Arch Gen Psychiatry* **51**:663–664.
35. Breiter HC, Rauch SL, Kwong KK, Baker JR, Weisskoff RM, Kennedy DN, Kendrick AD, Davis TL, Jiang A, Cohen MS, Stern CE, Belliveau JW, Baer L, O'Sullivan RL, Savage CR, Jenike MA, Rosen BR. 1996. Functional magnetic resonance imaging of symptom provocation in obsessive-compulsive disorder. *Arch Gen Psychiatry* **53**:595–606.
36. Breiter HC, Etcoff NF, Whalen PJ, Kennedy WA, Scott LR, Buckner RL, Strauss MM, Hyman SE, Rosen BR. 1996. Response and habituation of the human amygdala during visual processing of facial expression. *Neuron* **17**:875–887.
37. Breiter HC, Gollub RL, Weisskoff RM, Kennedy DN, Makris N, Berke JD, Goodman JM, Kantor HL, Gastfriend DR, Riorden JP, Mathew RT, Rosen BR, Hyman SE. 1997. Acute effects of cocaine on human brain activity and emotion. *Neuron* **19**:591–611.
38. Breiter HC, Aharon I, Kahneman D, Dale A, Shizgal P. 2001. Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron* **30**:619–639.
39. Bremner JD, Narayan M, Anderson ER, Staib LH, Miller HL, Charney DS. 2000. Hippocampal volume reduction in major depression. *Am J Psychiatry* **157**:115–118.
40. Bremner JD, Vythilingam M, Vermetten E, Nazeer A, Adil J, Khan S, Staib LH, Charney DS. 2002. Reduced volume of orbitofrontal cortex in major depression. *Biol Psychiatry* **51**:273–279.
41. Brown H, Kosslyn SM, Breiter HC, Baer L, Jenike MC. 1994. Can patients with obsessive-compulsive disorder discriminate between percepts and mental images? A signal detection analysis. *J Abnorm Psych* **103**:445–454.
42. Buchbinder BR, Cosgrove GR. 1998. Cortical activation MR studies in brain disorders. *Magn Reson Imaging Clin N Am* **6**:67–93.
43. Bush G, Frazier JA, Rauch SL, Seidman LJ, Whalen PJ, Jenike MA, Rosen BR, Biederman J. 1999. Anterior cingulate cortex dysfunction in attention-deficit/hyperactivity disorder revealed by fMRI and the Counting Stroop. *Biol Psychiatry* **45**:1542–1552.
44. Bush G, Vogt BA, Holmes J, Dale AM, Greve D, Jenike MA, Rosen BR. 2002. Dorsal anterior cingulate cortex: a role in reward-based decision making. *Proc Natl Acad Sci USA* **99**:523–528.
45. Buxton RB, Frank LR. 1997. A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. *J Cereb Blood Flow Metab* **17**:64–72
46. Cabanac M. 1971. Physiological role of pleasure. *Science* **173**:1103–1107.
47. Cabanac M. 1992. Pleasure: the common currency. *J Theor Biol* **155**:173–200.
48. Cabanac M, Duclaux R, Spector, NH. 1971. Sensory feedback in regulation of body weight: is there a ponderostat? *Nature* 229:125–127.
49. Cain ME, Kapp BS, Puryear CB. 2002. The contribution of the amygdala to conditioned thalamic arousal. *J Neurosci* **22**:11026–11034.
50. Cannon WD, Britton SW. 1925. Pseudoaffective meduliadrenal secretion. *Am J Physiol* **72**:283–294.
51. Carelli RM. 2002. Nucleus accumbens cell firing during goal-directed behaviors for cocaine vs. "natural" reinforcement. *Physiol Behav* **76**:379–387.
52. Carelli RM, James SG, Crumling AJ. 2000. Evidence that separate neural circuits in the nucleus accumbens encode cocaine versus "natural" (water and food) reward. *J Neurosci* **20**:4255–4266.
53. Caspi A, Sugden K, Moffitt TE, Taylor A, Craig IW, Harrington H, McClay J, Mill J, Martin J, Braithwaite A, Poulton R. 2003. Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science* **301**:386–389.

54. Caviness Jr VS, Kennedy DN, Richelme C, Rademacher J, Filipek PA. 1996. The human brain age 7–11 years: a volumetric analysis based on magnetic resonance images. *Cerebral Cortex* **6**(5):726–736.
55. Childress AR, Mozley PD, McElgin W, Fitzgerald J, Reivich M, O'Brien CP. 1999. Limbic activation during cue-induced cocaine craving. *Am J Psychiatry* **156**:11–18.
56. Churchland P, Sejnowski T. 1992. *The computational brain*. MIT Press, Cambridge.
57. Coghill RC, Talbot JD, Evans AC, Meyer E, Gjedde A, Bushnell MC, Duncan GH. 1994. Distributed processing of pain and vibration by the human brain. *J Neurosci* **14**:4095–4108.
58. Cohen M, Grossberg S. 1983. Absolute stability of global pattern formation and parallel memory storage by competitive neural network. *IEEE Trans SMC* **13**:815–826.
59. Collinge J. 2001. Prion diseases of humans and animals: their causes and molecular basis. *Annu Rev Neurosci* **24**:519–550.
60. Cowan WM, Harter DH, Kandel ER. 2000. The emergence of modern neuroscience: some implications for neurology and psychiatry. *Annu Rev Neurosci* **23**:343–391.
61. Cowan WM, Kopnisky KL, Hyman SE. 2002. The human genome project and its impact on psychiatry. *Annu Rev Neurosci* **25**:1–50.
62. Craig AD, Reiman EM, Evans A, Bushnell MC. 1996. Functional imaging of an illusion of pain. *Nature* **384**:258–260.
63. Crespo-Facorro B, Paradiso S, Andreasen NC, O'Leary DS, Watkins GL, Ponto LL, Hichwa RD. 2001. Neural mechanisms of anhedonia in schizophrenia: a PET study of response to unpleasant and pleasant odors. *JAMA* **286**:427–435.
64. Cummings CJ, Zoghbi HY. 2000. Trinucleotide repeats: mechanisms and pathophysiology. *Annu Rev Genomics Hum Genet* **1**:281–328.
65. Damasio AR. 1999. *The feeling of what happens*. William Heinemann, London.
66. Damasio AR. 1999. How the brain creates the mind. *Sci Am* **281**:112–117.
67. Darwin C. 1872. *The expression of the emotions in man and animals*. John Murray, London.
68. David A, Blamire A, Breiter H. 1994. Functional magnetic resonance imaging: a new technique with implications for psychology and psychiatry. *Br J Psychiatry* **164**:2–7.
69. Davidson RJ, Lewis DA, Alloy LB, Amaral DG, Bush G, Cohen JD, Drevets WC, Farah MJ, Kagan J, McClelland JL, Nolen-Hoeksema S, Peterson BS. 2002. Neural and behavioral substrates of mood and mood regulation. *Biol Psychiatry* **52**:478–502.
70. Davidson RJ, Pizzagalli D, Nitschke JB, Putnam K. 2002. Depression: perspectives from affective neuroscience. *Annu Rev Psychol* **53**:545–574.
71. Davis KD, Kiss ZH, Luo RR, Tasker RR, Lozano AM, Dostrovsky JO. 1998. Phantom sensations generated by thalamic microstimulation. *Nature* **391**:385–387.
72. Davis KD, Kwan C, Crawley AP, Mikulis DJ. 1998. Event-related fMRI of **pain**: entering a new era in imaging pain. *NeuroReport* 9:3019–3023.
73. Davis M, Whalen PJ. 2001. The amygdala: vigilance and emotion. *Mol Psychiatry* **6**:13–34.
74. Davis TL, Kwong KK, Weisskoff RM, Rosen BR. 1998. Calibrated functional MRI: mapping the dynamics of oxidative metabolism. *Proc Natl Acad Sci USA* **95**:1834–1839.
75. Dayan P, Kakade S, Montague PR. 2000. Learning and selective attention. *Nature Neurosci* **3**(suppl):1218–1223.
76. Dayan P, Balleine BW. 2002. Reward, motivation, and reinforcement learning. *Neuron* **36**:285–298.
77. Detre JA, Leigh JS, Koretsky AP. 1992. Perfusion imaging. *Magn Reson Med* **23**:265–270.
78. Dirnagl U, Niwa K, Lindauer U, Villringer A. 1994. Coupling of cerebral blood flow to neuronal **activation**: role of adenosine and nitric oxide. *Am J Physiol* 267:H296–H301.
79. Donaldson DI, Petersen SE, Buckner RL. 2001. Dissociating memory retrieval processes using **fMRI**: evidence that priming does not support recognition memory. *Neuron* 31:1047–1059.
80. Dreier JP, Ebert N, Priller J, Megow D, Lindauer U, Klee R, Reuter U, Imai Y, Einhaupl KM, Victorov I, Dirnagl U. 2000. *J Neurosurg* **93**:658–666.

81. Drevets WC. 1998. Functional neuroimaging studies of depression: the anatomy of melancholia. *Annu Rev Med* **49**:341–361.
82. Drevets WC. 2000. Functional anatomical abnormalities in limbic and prefrontal cortical structures in major depression. *Prog Brain Res* **126**:413–431.
83. Drevets WC. 2001. Neuroimaging and neuropathological studies of depression: implications for the cognitive-emotional features of mood disorders. *Curr Opin Neurobiol* **11**:240–249.
84. Drevets WC, Frank E, Price JC, Kupfer DJ, Holt D, Greer PJ, Huang Y, Gautier C, Mathis C. 1999. PET imaging of serotonin 1A receptor binding in depression. *Biol Psychiatry* **46**:1375–1387.
85. Dunah AW, Jeong H, Griffin A, Kim YM, Standaert DG, Hersch SM, Mouradian M, Young AB, Tanese N, Krainc D. 2002. Sp1 and TAFII130 transcriptional activity disrupted in early Huntington's disease. *Science* **296**:2238–2243.
86. Egan MF, Goldberg TE, Kolachana BS, Callicott JH, Mazzanti CM, Straub RE, Goldman D, Weinberger DR. 2001. Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. *Proc Natl Acad Sci USA* **98**:6917–622.
87. Elliott R, Friston KJ, Dolan RJ. 2000. Dissociable neural responses in human reward systems. *J Neurosci* **20**:6159–6165.
88. Elliott R, Newman JL, Longe OA, Deakin JF. 2003. Differential response patterns in the striatum and orbitofrontal cortex to financial reward in humans: a parametric functional magnetic resonance imaging study. *J Neurosci* **23**:303–307.
89. Everitt BJ 1978. Monoamines and sexual behaviour in non-human primates. *Ciba Found Symp* **62**:329–358.
90. Fava M, Kendler KS. 2000. Major depressive disorder. *Neuron* **28**:335–341.
91. Filipek PA, Richelme C, Kennedy DN, Caviness Jr VS. 1994. The young adult human brain: an MRI-based morphometric analysis. *Cerebral Cortex* **4**(4):344–360.
92. Fitzgerald KD, Moore GJ, Paulson LA, Stewart CM, Rosenberg DR. 2000. Proton spectroscopic imaging of the thalamus in treatment-naive pediatric obsessive-compulsive disorder. *Biol Psychiatry* **47**:174–182.
93. Fowler JS, Volkow ND, Wang GJ, Pappas N, Logan J, MacGregor R, Alexoff D, Shea C, Schlyer D, Wolf AP, D. Warner, I. Zezulkova and R. Cilento. 1996. Inhibition of monoamine oxidase B in the brains of smokers. *Nature* **379**:733–736.
94. Franklin TR, Acton PD, Maldjian JA, Gray JD, Croft JR, Dackis CA, O'Brien CP, Childress AR. 2002. Decreased gray matter concentration in the insular, orbitofrontal, cingulate, and temporal cortices of cocaine patients. *Biol Psychiatry* **51**:134–142.
95. Freeman WJ. 2001. *How brains make up their minds*. Columbia UP, New York.
96. Freeman WJ, Barrie JM. 1994. Chaotic oscillations and the genesis of meaning in cerebral cortex. In *Temporal coding in the brain*, pp. 13–37. Ed. G Buzsaki, R Llinas, W Singer, A Berthoz, Y Christen. Springer, Berlin.
97. Friston K. 2002. Beyond phrenology: what can neuroimaging tell us about distributed circuitry? *Annu Rev Neurosci* **25**:221–250.
98. Gallistel CR. 1978. Self-stimulation in the rat: quantitative characteristics of the reward pathway. *J Comp Physiol Psychol* **92**:977–998.
99. Gallistel CR. 1990. *The organization of learning*. MIT Press, Cambridge.
100. Garavan H, Pankiewicz J, Bloom A, Cho JK, Sperry L, Ross TJ, Salmeron BJ, Risinger R, Kelley D, Stein EA. 2000. Cue-induced cocaine craving: neuroanatomical specificity for drug users and drug stimuli. *Am J Psychiatry* **157**:1789–1798.
101. Gershon ES, Kelsoe JR, Kendler KS, Watson JD. 2001. A scientific opportunity. *Science* **294**:957.
102. Glimcher P. 2003. *Decisions, uncertainty, and the brain: the science of neuroeconomics*. MIT Press, Boston.

103. Goldstein JM, Goodman JM, Seidman LJ, Kennedy DN, Makris N, Lee H, Tourville J, Caviness Jr VS, Faraone SV, Tsuang MT. 1999. Cortical abnormalities in schizophrenia identified by structural magnetic resonance imaging. *Arch Gen Psychiatry* **56**:537–547.
104. Goldstein RZ, Volkow ND, Wang GJ, Fowler JS, Rajaram S. 2001. Addiction changes orbitofrontal gyrus **function**: involvement in response inhibition. *NeuroReport* 12:2595–2599.
105. Gollub RL, Breiter HC, Kantor H, Kennedy D, Gastfriend D, Mathew RT, Makris N, Guimaraes A, Riorden J, Campbell T, Foley M, Hyman SE, Rosen B, Weisskoff R. 1998. Cocaine decreases cortical cerebral blood flow, but does not obscure regional activation in functional magnetic resonance imaging in human subjects. *J Cerebral Blood Flow Metab* **18**:724–734.
106. Gottesman II, Shields J. 1973. Genetic theorizing and schizophrenia. *Br J Psychiatry* **122**:15–30.
107. Grant S, London ED, Newlin DB, Villemagne VL, Liu X, Contoreggi C, Phillips RL, Kimes AS, Margolin A. 1996. Activation of memory circuits during cue-elicited cocaine craving. *Proc Natl Acad Sci USA* **93**:12040–12045.
108. Graybiel AM, Rauch SL. 2000. Toward a neurobiology of obsessive-compulsive disorder. *Neuron* **28**:343–347.
109. Grebogi C, Ott E, Yorke JA. 1983. Fractal basin boundaries, long–lived chaotic transients and stable-unstable pair bifurcation. *Phys Rev Lett* **50**:935–938.
110. Grebogi C, Ott E, Yorke JA. 1986. Critical exponent of chaotic transients in nonlinear dynamical systems. *Phys Rev Lett* **57**:1284–1287.
111. Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE, Watkins PC, Ottina K, Wallace MR, Sakaguchi AY, Sakaguchi AY. 1983. A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* **306**:234–248.
112. Hariri AR, Mattay VS, Tessitore A, Kolachana B, Fera F, Goldman D, Egan MF, Weinberger DR. 2002. Serotonin transporter genetic variation and the response of the human amygdala. *Science* **297**:400–403.
113. Hauser MD, Chomsk N, Fitch WT. 2002. The faculty of language: what is it, who has it, and how did it evolve? *Science* **298**:1569–1579.
114. Haxby JV, Grady CL, Horwitz B, Ungerleider LG, Mishkin M, Carson RE, Herscovitch P, Schapiro MB, Rapoport SI. 1991. Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proc Natl Acad Sci USA* **88**:1621–1625.
115. Heckers S, Curran T, Goff D, Rauch SL, Fischman AJ, Alpert NM, Schacter DL. 2000. Abnormalities in the thalamus and prefrontal cortex during episodic object recognition in schizophrenia. *Biol Psychiatry* **48**:651–657.
116. Heckers S, Weiss AP, Alpert NM, Schacter DL. 2002. Hippocampal and brain stem activation during word retrieval after repeated and semantic encoding. *Cereb Cortex* **12**:900–907.
117. Heimer L, Harlan RE, Alheid GF, Garcia MM, de Olmos J. 1997. Substantia innominata: a notion which impedes clinical-anatomical correlations in neuropsychiatric disorders. *Neuroscience* **76**:957–1006.
118. Heppner FL, Musahl C, Arrighi I, Klein MA, Rulicke T, Oesch B, Zinkernagel RM, Kalinke U, Aguzzi A. 2001. Prevention of scrapie pathogenesis by transgenic expression of anti-prion protein antibodies. *Science* **294**:178–182.
119. Herrnstein RJ. 1971. Quantitative hedonism. *J Psychiatr Res* **8**:399–412.
120. Herrnstein RJ, Loveland DH. 1974. Hunger and contrast in a multiple schedule. *J Exp Anal Behav* **21**:511–517.
121. Hill AV. 2001. The genomics and genetics of human infectious disease susceptibility. *Annu Rev Genomics Hum Genet* **2**:373–400.
122. Hoge RD, Atkinson J, Gill B, Crelier GR, Marrett S, Pike GB. (1999). Investigation of BOLD signal dependence on cerebral blood flow and oxygen consumption: the deoxyhemoglobin dilution model. *Magn Reson Med* **42**:849–863.
123. Hopfield JJ. 1982. Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci USA* **79**:2554–2558.

124. The Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. 1993. *Cell* **72**:971–983.
125. Hyman SE. 2002. Neuroscience, genetics, and the future of psychiatric diagnosis. *Psychopathology* **35**:139–144.
126. Hyman SE, Malenka RC. 2001. Addiction and the brain: the neurobiology of compulsion and its persistence. *Nature Rev Neurosci* **2**:695–703.
127. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. 2004. Detection of large-scale variation in the human genome. *Nature Genet* **36**:949–951.
128. Jaenisch R, Bird A. 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genet* **33**(suppl):245–254.
129. James W. 1884. What is an emotion? *Mind* **9**:188–205.
130. Jankovic J. 2001. Tourette's syndrome. *N Engl J Med* **345**:1184–1192.
131. Kahneman D, Tversky A. 1979. Prospect theory: an analysis of decision under risk. *Econometrica* **47**:263–291.
132. Kelley AE, Berridge KC. 2002. The neuroscience of natural rewards: relevance to addictive drugs. *J Neurosci* **22**:3306–3311.
133. Ketter TA, Andreason PJ, George MS, Lee C, Gill DS, Parekh PI, Willis MW, Herscovitch P, Post RM. 1996. Anterior paralimbic mediation of procaine-induced emotional and psychosensory experiences. *Arch Gen Psychiatry* **53**:59–69.
134. Ketter TA, Kimbrell TA, George MS, Dunn RT, Speer AM, Benson BE, Willis MW, Danielson A, Frye MA, Herscovitch P, Post RM. 2001. Effects of mood and subtype on cerebral glucose metabolism in treatment-resistant bipolar disorder. *Biol Psychiatry* **49**:97–109.
135. Kirschner M, Gerhart J. 1998. Evolvability. *Proc Natl Acad Sci USA* **95**:8420–8427.
136. Knight JC. 2004. Allele-specific gene expression uncovered. *Trends Genet* **20**:113–116.
137. Knutson B, Adams CM, Fong GW. Hommer D. 2001. Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *J Neurosci* **21**:RC159.
138. Knutson B, Fong GW, Adams CM, Varne JL. Hommer D. 2001. Dissociation of reward anticipation and outcome with event-related fMRI. *NeuroReport* **12**:3683–3687.
139. Kosslyn SM, Shin LM. 1992. The status of cognitive neuroscience. *Curr Opin Neurobiol* **2**:146–149.
140. Krebs J, Davies N. 1997. *Behavioural ecology: an evolutionary perspective*. Blackwell Science, Oxford.
141. Krishnan KR, Doraiswamy PM, Figiel GS, Husain MM, Shah SA, Na C, Boyko OB, McDonald WM, Nemerof CB, Ellinwood Jr EH. 1991. Hippocampal abnormalities in depression. *J Neuropsychiatry Clin Neurosci* **3**:387–391.
142. Kupferman I, Kandel ER, Iversen S. 2000. Motivational and addictive states. In *Principles of neural science*, pp. 998–1013.Ed. T Jessell, ER Kandel, JH Schwartz. McGraw-Hill, New York.
143. Kwong KK, Belliveau JW, Chesler D, Goldberg IE, Weisskoff RM, Poncelet BP, Kennedy DN, Hoppel BE, Cohen MS, Turner R. 1992. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc Natl Acad Sci USA* **89**:5675–5679.
144. LaBar KS, Gatenby JC, Gore JC, LeDoux JE, Phelps EA. 1998. Human amygdala activation during conditioned fear acquisition and extinction: a mixed-trial fMRI study. *Neuron* **20**:937–945.
145. Lange CG. 1985. *Om sindsbevaegelser et psyko. fysiolog. studie*. Kromar, Copenhagen.
146. LeDoux, J. 1996. *The emotional brain*. Simon and Shuster, New York.
147. LeDoux JE. 2000. Emotion circuits in the brain. *Annu Rev Neurosci* **23**:155–184.
148. LeDoux JE, Sakaguchi A, Iwata J, Reis DJ. 1985. Auditory emotional memories: establishment by projections from the medial geniculate nucleus to the posterior neostriatum and/or dorsal amygdala. *Ann NY Acad Sci* **444**:463–464.

149. Lenox RH, Gould TD, Manji HK. 2002. Endophenotypes in bipolar disorder. *Am J Med Genet* **114**:391–406.
150. Lewis DA, Levitt P. 2002. Schizophrenia as a disorder of neurodevelopment. *Annu Rev Neurosci* **25**:409–432.
151. Li SJ, Wang Y, Pankiewicz J, Stein EA. 1999. Neurochemical adaptation to cocaine abuse: reduction of N-acetyl aspartate in thalamus of human cocaine abusers. *Biol Psychiatry* **45**:1481–1447.
152. Li YJ, Scott WK, Hedges DJ, Zhang F, Gaskell PC, Nance MA, Watts RL, Hubble JP, Koller WC, Pahwa R, Stern MB, Hiner BC, Jankovic J, Allen FA Jr, Goetz CG, Mastaglia F, Stajich JM, Gibson RA, Middleton LT, Saunders AM, Scott BL, Small GW, Nicodemus KK, Reed AD, Schmechel DE, Welsh-Bohmer KA, Conneally PM, Roses AD, Gilbert JR, Vance JM, Haines JL, Pericak-Vance MA. 2002. Age at onset in two common neurodegenerative diseases is genetically controlled. *Am J Hum Genet* **70**:985–993.
153. Lindauer U, Royl G, Leithner C, Kuhl M, Gold L, Gethmann J, Kohl-Bereis M, Villringer A, Dirnagl U. 2001. *NeuroImage* **13**:988–1001.
154. Lindvall O, Bjorklund A. 1974. The organization of the ascending catecholamine neuron systems in the rat brain as revealed by the glyoxylic acid fluorescence method. *Acta Physiol Scand Suppl* **412**:1–48.
155. Liu Y, Gao JH, Liu HL, Fox PT. 2000. The temporal response of the brain after eating revealed by functional MRI. *Nature* **405**:1058–1062.
156. Logothetis MK, Pauls J, Augath M, Trinath T, Oeltermann A. 2001. Neurophysiological investigation of the basis of the fMRI signal. *Nature* **412**:150–157.
157. Makris N, Worth AJ, Sorensen AG, Papadimitriou GM, Wu O, Reese TG, Wedeen VJ, Davis TL, Stakes JW, Caviness VS, Kaplan E, Rosen BR, Pandya DN, Kennedy DN. 1997. Morphometry of in vivo human white matter association pathways with diffusion-weighted magnetic resonance imaging. *Ann Neurol* **42**(6):951–962.
158. Makris N, Meyer JW, Bates JF, Yeterian EH, Kennedy DN, Caviness VS. 1999. MRI-based topographic parcellation of human cerebral white matter and nuclei, II: rationale and applications with systematics of cerebral connectivity. *NeuroImage* **9**:18–45.
159. Makris N, Pandya DN, Normandin JJ, Papadimitriou GM, Rauch SL, Caviness VS, Kennedy DN. 2002. Diffusion tensor MRI investigations of the human cingulum bundle. *CNS Spectrums* **7**(7):522–528.
160. Mandeville JB, Marota JJ, Kosofsky DE, Keltner JR, Weissleder R, Rosen BR, Weisskoff RM. 1998. Dynamic functional imaging of relative cerebral blood volume during rat forepaw stimulation. *Magn Reson Med* **39**:615–624.
161. Manji HK, Drevets WC, Charney DS. 2001. The cellular neurobiology of depression. *Nature Med* **7**:541–547.
162. Manoach DS, Gollub RL, Benson ES, Searl MM, Goff DC, Halpern E, Saper CB, Rauch SL. 2000. Schizophrenic subjects show aberrant fMRI activation of dorsolateral prefrontal cortex and basal ganglia during working memory performance. *Biol Psychiatry* **48**:99–109.
163. Manoach DS, Halpern EF, Kramer TS, Chang Y, Goff DC, Rauch SL, Kennedy DN, Gollub RL. 2001. Test-retest reliability of a functional MRI working memory paradigm in normal and schizophrenic subjects. *Am J Psychiatry* **158**:955–958.
164. Manoach DS, Lindgren KA, Cherkasova MV, Goff DC, Halpern EF, Intriligator J, Barton JJ. 2002. Schizophrenic subjects show deficient inhibition but intact task switching on saccadic tasks. *Biol Psychiatry* **51**:816–826.
165. Marr D. 1982. *A computational investigation into the human representation and processing of visual information*. Freeman, San Francisco.
166. Mathiesen C, Caesar K, Akgoren N, Lauritzen M. 1998. Modification of activity-dependent increases of cerebral blood flow by excitatory synaptic activity and spikes in rat cerebellar cortex. *J Physiol* **512**:555–566.

167. Mazur JE. 1994. Effects of intertrial reinforcers on self-control choice. *J Exp Anal Behav* **61**:83–96.
168. Mazur JE, Vaughan Jr W. 1987. Molar optimization versus delayed reinforcement as explanations of choice between fixed-ratio and progressive-ratio schedules. *J Exp Anal Behav* **48**:251–261.
169. McClelland JL, Rumelhart DE. 1985. Distributed memory and the representation of general and specific information. *J Exp Psychol Gen* **114**:159–197.
170. Mellers BA. 2000. Choice and the relative pleasure of consequences. *Psychol Bull* **126**:910–924.
171. Mellers BA, Schwartz A, Ho K, Ritov I. 1997. Decision affect theory: how we feel about risky options. *Psych Sci* **8**:423–429.
172. Mellers BA, Schwartz A, Cooke AD. 1998. Judgment and decision making. *Annu Rev Psychol* **49**:447–477.
173. Mesulam MM. 1998. From sensation to cognition. *Brain* **121**:1013–1052.
174. Mesulam M. 2000. *Principles of behavioral and cognitive neurology*. Oxford UP, London.
175. Meyer JW, Makris N, Bates JF, Caviness VS, Kennedy DN. 1999. MRI-based topographic parcellation of human cerebral white matter. *NeuroImage* **9**:1–17.
176. Miller GA. 2003. The cognitive revolution: a historical perspective. *Trends Cogn Sci* **7**:141–144.
177. Modell AH. 2003. *Imagination and the meaningful brain*. MIT Press, Cambridge.
178. Montague PR, Dayan P, Person C, Sejnowski TJ. 1995. Bee foraging in uncertain environments using predictive hebbian learning. *Nature* **377**:725–728.
179. Narr PL, Cannon TD, Woods RP, Thompson PM, Kim S, Asunction D, van Erp TG, Poutanen VP, Huttunen M, Lonnqvist J, Standerksjold-Nordenstam CG, Kaprio J, Mazziotta JC, Toga AW. 2002. Genetic contributions to altered callosal morphology in schizophrenia. *J Neurosci* **22**:3720–3729.
180. Nauta WJH, Feirtag M. 1986. *Fundamental neuroanatomy*. W.H. Freeman, New York.
181. Nestler EJ, Barrot M, DiLeone RJ, Eisch AJ, Gold SJ, Monteggia LM. 2002. Neurobiology of depression. *Neuron* **34**:13–25.
182. Nicholls RD, Knepper JL. 2001. Genome organization, function, and imprinting in Prader-Willi and Angelman syndromes. *Annu Rev Genomics Hum Genet* **2**:153–175.
183. Nicotera P. 2001. A route for prion neuroinvasion. *Neuron* **31**:345–348.
184. Nieuwnhuys R, Voogd J, van Huizen C. 1988. *The human central nervous system*. Springer, New York.
185. Nordahl TE, Benkelfat C, Semple WE, Gross M, King AC, Cohen RM. 1989. Cerebral glucose metabolic rates in obsessive compulsive disorder. *Neuropsychopharmacology* **2**:23–28.
186. O'Brien SJ, Nelson GW, Winkler CA, Smith MW. 2000. Polygenic and multifactorial disease gene association in man: lessons from AIDS. *Annu Rev Genet* **34**:563–591.
187. Obrig H, Israel H, Kohl-Bareis M, Uludag K, Wenzel R, Muller B, Arnold G, Villringer A. 2002. Habituation of the visually evoked potential and its vascular response: implications for neurovascular coupling in the healthy adult. *NeuroImage* **17**:1–18.
188. O'Doherty J, Rolls ET, Francis S, Bowtell R, McGlone F. 2001. Representation of pleasant and aversive taste in the human brain. *J Neurophysiol* **85**:1315–1321.
189. O'Doherty JP, Deichmann R, Critchley HD, Dolan RJ. 2002. Neural responses during anticipation of a primary taste reward. *Neuron* **33**:815–826.
190. Ogawa S, Tank DW, Menon R, Ellermann JM, Kim SG, Merkle H, Ugurbil K. 1992. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc Natl Acad Sci USA* **89**:5951–5955.
191. Okazawa H, Rich T, Chang A, Lin X, Waragai M, Kajikawa M, Enokido Y, Komuro A, Kato S, Shibatay M, Hatanaka H, Mouradian MM, Sudol M, Kanazawa I. 2002. Interaction between mutant ataxin-1 and PQBP-1 affects transcription and cell death. *Neuron* **34**:701–713.

192. Olds J, Milner PM. 1954. Positive reinforcement produced by electrical stimulation of the septal area and other regions of the rat brain. *J Comp Physiol Psychol* **47**:419–427.
193. Ongur D, Drevets WC, Price JL. 1998. Glial reduction in the subgenual prefrontal cortex in mood disorders. *Proc Natl Acad Sci USA* **95**:13290–13325.
194. Opal P, Zoghbi HY. 2002. The role of chaperones in polyglutamine disease. *Trends Mol Med* **8**:232–236.
195. Pandya DN, Yeterian EH. 1985. Architecture and connections of cortical association. . In *Cerebral cortex*, Vol. 4. pp. 3–61. Ed A Peters, EG Jones. Plenum, London.
196. Panksepp J. 1998. *Affective Neuroscience*. Oxford UP, New York.
197. Pauling L, Coryell C. 1936. The magnetic properties and structure of the hemochromogen and related substances. *Proc Natl Acad Sci USA* **22**:159.
198. Pauling L, Coryell C. 1936. The magnetic properties and structure of hemoglobin, oxyhemoglobin, and carbonmonoxyhemoglobin. *Proc Natl Acad Sci USA* **22**:210.
199. Perani D, Colombo C, Bressi S, Bonfanti A, Grassi F, Scarone S, Bellodi L, Smeraldi E, Fazio F. 1995. [18F]FDG PET study in obsessive-compulsive disorder. A clinical/metabolic correlation study after treatment. *Br J Psychiatry* **166**:244–250.
200. Phelps EA, O'Connor KJ, Gatenby JC, Gore JC, Grillon C, Davis M. 2001. Activation of the left amygdala to a cognitive representation of fear. *Nature Neurosci* **4**:437–441.
201. Ploghaus A, Tracey I, Gati JS, Clare S, Menon RS, Matthews PM, Rawlins JN. 1999. Dissociating pain from its anticipation in the human brain. *Science* **284**:1979–1981.
202. Porjesz B, Almasy L, Edenberg HJ, Wang K, Chorlian DB, Foroud T, Goate A, Rice JP, O'Connor SJ, Rohrbaugh J, J, Kuperman S, Bauer LO, Crowe RR, Schuckit MA, Hesselbrock V, Conneally PM, Tischfield JA, Li TK, Reich T, Begleiter H. 2002. Linkage disequilibrium between the beta frequency of the human EEG and a GABAA receptor gene locus. *Proc Natl Acad Sci USA* **99**:3729–3733.
203. Pouget A, Dayan P, Zemel R. 2003. Inference and computation with population codes. *Annu Rev Neurosci* **26**:381–410.
204. Prigogine I. 1985. The rediscovery of time. In *Science and complexity*, pp. 11–25. Ed. S Nash. Science Reviews, Northwood, Middlesex.
205. Queitsch C, Sangster TA, Lindquist S. 2002. Hsp90 as a capacitor of phenotypic variation. *Nature* **417**:618–624.
206. Rachlin H, Raineri A, Cross D. 1991. Subjective probability and delay. *J Exp Anal Behav* **55**:233–244.
207. Rainville P, Duncan GH, Price DD, Carrier B, Bushnell MC. 1997. Pain affect encoded in human anterior cingulate but not somatosensory cortex. *Science* **277**:968–971.
208. Rainville P, Bushnell MC, Duncan GH. 2001. Representation of acute and persistent pain in the human **CNS**: potential implications for chemical intolerance. *Ann NY Acad Sci* 933:130–141.
209. Rajkowska G, Miguel-Hidalgo JJ, Wei J, Dilley G, Pittman SD, Meltzer HY, Overholser JC, Roth BL, Stockmeier CA. 1999. Morphometric evidence for neuronal and glial prefrontal cell pathology in major depression. *Biol Psychiatry* **45**:1085–1098.
210. Rajkowska G, Halaris A, Selemon LD. 2001. Reductions in neuronal and glial density characterize the dorsolateral prefrontal cortex in bipolar disorder. *Biol Psychiatry* **49**:741–752.
211. Rauch SL, Savage CR, Alpert NM, Fischman AJ, Jenike MA. 1997. The functional neuroanatomy of anxiety: a study of three disorders using positron emission tomography and symptom provocation. *Biol Psychiatry* **42**:446–452.
212. Rauch SL, Whalen PJ, Shin LM, McInerney SC, Macklin ML, Lasko NB, Orr SP, Pitman RK. 2000. Exaggerated amygdala response to masked facial stimuli in post-traumatic stress disorder: a functional MRI study. *Biol Psychiatry* **47**:769–776.
213. Rees G, Friston K, Koch C. 2000. A direct quantitative relationship between the functional properties of human and macque V5. *Nature Neurosci* **3**:716–723.

214. Reich T, Hinrichs A, Culverhouse R, Bierut L. 1999. Genetic studies of alcoholism and substance dependence. *Am J Hum Genet* **65**:599–605.
215. Robbins TW, Everitt BJ. 2002. Limbic-striatal memory systems and drug addiction. *Neurobiol Learn Mem* **78**:625–636.
216. Rother J, Knab R, Hamzei F, Fiehler J, Reichenback JR, Buchel C, Weiller C. 2002. Negative dip in BOLD fMRI is caused by blood flow–oxygen consumption uncoupling in humans. *NeuroImage* **15**:98–102.
217. Rutherford SL, Lindquist S. 1998. Hsp90 as a capacitor for morphological evolution. *Nature* **396**:336–342.
218. Sapienza C, Hall JG. 2001. Genome imprinting in human disease. In *The metabolic and molecular bases of inherited disease*, pp. 437–458. Ed CR Scriver, AL Beaudet, WS Sly, D Valle. McGraw-Hill/Medical Publishing, New York.
219. Sapolsky RM. 2001. Depression, antidepressants, and the shrinking hippocampus. *Proc Natl Acad Sci USA* **98**:12320–12322.
220. Sawa A, Snyder SH. 2002. Schizophrenia: diverse approaches to a complex disease. *Science* **296**:692–695.
221. Sawamoto N, Honda M, Okada T, Hanakawa T, Kanda M, Fukuyama H, Konishi J, Shibasaki H. 2000. Expectation of pain enhances responses to nonpainful somatosensory stimulation in the anterior cingulate cortex and parietal operculum/posterior insula: an event-related functional magnetic resonance imaging study. *J Neurosci* **20**:7438–7445.
222. Sawle GV, Hymas NF, Lees AJ, Frackowiak RS. 1991. Obsessional slowness: functional studies with positron emission tomography. *Brain* **114**(pt 5):2191–2202.
223. Saxena S, Brody AL, Ho ML, Alborzian S, Ho MK, Maidment KM, Huang SC, Wu HM, Au C, Baxter Jr LR. 2001. Cerebral metabolism in major depression and obsessive-compulsive disorder occurring separately and concurrently. *Biol Psychiatry* **50**:159–170.
224. Schmitz B, Bock C, Hoehn-Berlage M, Kerskens CM, Bottiger BW, Hossmann KA. 1998. Recovery of the rodent brain after cardiac arrest: a functional MRI study. *Magn Reson Med* **39**:783–788.
225. Schneier FR, Liebowitz MR, Abi-Dargham A, Zea-Ponce Y, Lin SH, Laruelle M. 2000. Low dopamine D(2) receptor binding potential in social phobia. *Am J Psychiatry* **157**:457–459.
226. Schuff N, Neylan TC, Lenoci MA, Du AT, Weiss DS, Marmar CR, Weiner MW. 2001. Decreased hippocampal N-acetylaspartate in the absence of atrophy in post-traumatic stress disorder. *Biol Psychiatry* **50**:952–959.
227. Schultz W. 2000. Multiple reward signals in the brain. *Nature Rev Neurosci* **1**:199–207.
228. Schultz W. 2002. Getting formal with dopamine and reward. *Neuron* **36**:241–263.
229. Schultz W, Dayan P, Montague PR. 1997. A neural substrate of prediction and reward. *Science* **275**:1593–1599.
230. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M. (2004). Large-scale copy number polymorphism in the human genome. *Science* **305**:525–528.
231. Seidman LJ, Breiter HC, Goodman JM, Goldstein JM, Woodruff PW, O'Craven K, Savoy R, Tsuang MT, Rosen BR. 1998. A functional magnetic resonance imaging study of auditory vigilance with low and high information processing demands. *Neuropsychology* **12**:505–518.
232. Shahbazian MD, Antalffy B, Armstrong DL, Zoghbi HY. 2002. Insight into Rett syndrome: MeCP2 levels display tissue- and cell-specific differences and correlate with neuronal maturation. *Hum Mol Genet* **11**:115–124.
233. Shahbazian MD, Zoghbi HY. 2002. Rett syndrome and MeCP2: linking epigenetics and neuronal function. *Am J Hum Genet* **71**:1259–1272.
234. Shannon CE, Weaver W. 1949. *The mathematical theory of communication*. U Illinois P, Urbana.

235. Shenton ME, Dickey CC, Frumin M, McCarley RW. 2001. A review of MRI findings in schizophrenia. *Schizophr Res* **49**:1–52.
236. Sherman MY, Goldberg AL. 2001. Cellular defenses against unfolded proteins: a cell biologist thinks about neurodegenerative diseases. *Neuron* **29**:15–32.
237. Shizgal P. 1997. Neural basis of utility estimation. *Curr Opin Neurobiol* **7**:198–208.
238. Shizgal P. 1999. On the neural computation of utility: implications from studies of brain stimulation reward. In *Well-being: the foundations of hedonic psychology*, pp. 502–506. Ed. D Kahneman, E Diener, N Schwarz. Russell Sage, New York.
239. Small DM, Zatorre RJ, Dagher A, Evans AC, Jones-Gotman M. 2001. Changes in brain activity related to eating chocolate: from pleasure to aversion. *Brain* **124**:1720–1733.
240. Spinoza B. 1883. On the origin and nature of emotions. In *The ethics*, pp. 1–132. Ed RHM Elwes. Princeton UP, Princeton.
241. Stoll M, Cowley Jr AW, Tonellato PJ, Greene AS, Kaldunski ML, Roman RJ, Dumas P, Schork NJ, Wang Z, Jacob HJ. 2001. A genomic-systems biology map for cardiovascular function. *Science* **294**:1723–1726.
242. Sutton J, Breiter HC. 1994. Neural scale invariance: An integrative model with implications for neuropathology. *World Cong Neural Netw* **4**:667–672.
243. Sutton RS, Barto AG. 1981. Toward a modern theory of adaptive **networks**: expectation and prediction. *Psychol Rev* 88:135–170.
244. Swedo SE, Schapiro MB, Grady CL, Cheslow DL, Leonard HL, Kumar A, Friedland R, Rapoport SI, Rapoport JL. 1989. Cerebral glucose metabolism in childhood-onset obsessive-compulsive disorder. *Arch Gen Psychiatry* **46**:518–523.
245. Tager-Flusberg H. 1999. *Neurodevelopmental disorders*. MIT Press, Cambridge.
246. Talairach J, Tournoux P. 1988. *Co-planar stereotaxic atlas of the human brain*. Thieme Medical, New York.
247. Talbot JD, Marrett S, Evans AC, Meyer E, Bushnell MC, Duncan GH. 1991. Multiple representations of pain in human cerebral cortex. *Science* **251**:1355–1358.
248. Thompson PM, Cannon TD, Narr KL, van Erp T, Poutanen VP, Huttunen M, Lonnqvist J, Standertskjold-Nordenstam CG, Kaprio J, Khaledy M, Dail R, Zoumalan CI, Toga AW. 2001. Genetic influences on brain structure. *Nature Neurosci* **4**:1253–1258.
249. Thulborn KR, Waterton JC, Matthews PM, Radda GK. 1982. Oxygenation dependence of the transverse relaxation time of water protons in whole blood at high field. *Biochim Biophys Acta* **714**:265–270.
250. Thut G, Schultz W, Roelcke U, Nienhusmeier M, Missimer J, Maguire RP, Leenders KL. 1997. Activation of the human brain by monetary reward. *NeuroReport* **8**:1225–1228.
251. Toga AW, Mazziotta JC. 2002. *Brain mapping: the methods*, 2nd ed. Elsevier Science, San Diego.
252. True HL, Lindquist SL. 2000. A yeast prion provides a mechanism for genetic variation and phenotypic diversity. *Nature* **407**:477–483.
253. Tsukada H, Kakiuchi T, Shizuno H, Nishiyama S. 1998. Interactions of cholinergic and glutamatergic neuronal systems in the functional activation of cerebral blood flow response: a PET study in unanesthetized monkeys. *Brain Res* **796**:82–90.
254. Tversky A, Kahneman D. 1992. Advances in prospect theory: cumulative representation of uncertainty. *J Risk Uncertainty* **5**:297–323.
255. Villringer A, Them A, Lindauer U, Einhaupl K, Dirnagl U. 1994. Capillary perfusion of the rat brain cortex: an in vivo confocal microscopy study. *Circ Res* **75**:55–62.
256. Volkow ND, Fowler JS, Wolf AP, Hitzemann R, Dewey S, Bendriem B, Alpert R, Hoff A. 1991. Changes in brain glucose metabolism in cocaine dependence and withdrawal. *Am J Psychiatry* **148**:621–626.
257. Volkow ND, Wang GJ, Fischman MW, Foltin RW, Fowler JS, Abumrad NN, Vitkun S, Logan J, Gatley SJ, Pappas N, Hitzemann R, Shea CE. 1997. Relationship between subjective effects of cocaine and dopamine transporter occupancy. *Nature* **386**:827–830.

258. Volkow ND, Wang GJ, Fowler JS, Franceschi D, Thanos PK, Wong C, Gatley SJ, Ding YS, Molina P, Schlyer D, Alexoff D, Hitzemann R, Pappas N. 2000. Cocaine abusers show a blunted response to alcohol intoxication in limbic brain regions. *Life Sci* **66**:PL161–PL167.
259. Volkow ND, Chang L, Wang GJ, Fowler JS, Ding YS, Sedler M, Logan J, Franceschi D, Gatley J, Hitzemann R, Gifford A, Wong C, Pappas N. 2001. Low level of brain dopamine D2 receptors in methamphetamine abusers: association with metabolism in the orbitofrontal cortex. *Am J Psychiatry* **158**:2015–2021.
260. Vrana PB, A. Fossella J, Matteson P, del Rio T, O'Neill MJ, Tilghman SM. 2000. Genetic and epigenetic incompatibilities underlie hybrid dysgenesis in Peromyscus. *Nature Genet* **25**:120–124.
261. Wagner AD, Schacter DL, Rotte M, Koutstaal W, Maril A, Dale AM, Rosen BR, Buckner RL. 1998. Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. *Science* **281**:1188–1191.
262. Watase K, Weeber EJ, Xu B, Antalffy B, Yuva-Paylor L, Hashimoto K, Kano M, Atkinson R, Sun Y, Armstrong DL, Sweatt JD, Orr HT, Paylor R, Zoghbi HY. 2002. A long CAG repeat in the mouse Sca1 locus replicates SCA1 features and reveals the impact of protein solubility on selective neurodegeneration. *Neuron* **34**:905–919.
263. Watts AG, Swanson LW. 2002. Anatomy of motivation. In *Stevens' handbook of experimental psychology: learning, motivation, and emotion*, Vol. 3. pp. 213–236. Ed H Pasher, CR Gallistel. John Wiley & Sons, New York.
264. Williams JT, Begleiter H, Porjesz B, Edenberg HJ, Foroud T, Reich T, Goate A, Van Eerdewegh P, Almasy L, Blangero J. 1999. Joint multipoint linkage analysis of multivariate qualitative and quantitative traits, II: alcoholism and event-related potentials. *Am J Hum Genet* **65**:1148–1160.
265. Wise RA. 1978. Catecholamine theories of reward: a critical review. *Brain Res* **152**:215–247.
266. Wise RA. 2002. Brain reward circuitry: insights from unsensed incentives. *Neuron* **36**:229–240.
267. Yorke JA, Yorke ED. 1979. Metastable chaos: The transition to sustained chaotic oscillations in a model of Lorentz. *J Stat Phys* **21**: 263–277.
268. Zeki S. 2001. Localization and globalization in conscious vision. *Annu Rev Neurosci* **24**:57–86.
269. Zoghbi HY. 2001. Rett syndrome. In *The metabolic and molecular bases of inherited disease*, pp. 1291–1326. Ed. CR Scriver, AL Beaudet, WS Sly, D. Valle. McGraw-Hill/Medical Publishing, New York.
270. Delgado MR, Nystrom LE, Fissell C, Noll DC, Fiez JA. 2000. Tracking the hemodynamic responses to reward and punishment in the striatum. *J Neurophysiol* **84**(6):3072–3077.
271. Enard W, Khaitovich P, Klose J, Zollner S, Heissig F, Giavalisco P, Nieselt-Struwe K, Muchmore E, Varki A, Ravid R, Doxiadis GM, Bontrop RE, Paabo S. 2002. Intra- and interspecific variation in primate gene expression patterns. *Science* **296**(5566):340–343.
272. Kahn I, Yeshurun Y, Rotshtein P, Fried I, Ben-Bashat D, Hendler T. 2002. The role of the amygdala in signaling prospective outcome of choice. *Neuron* **33**(6):983–994.
273. Kampe KK, Frith CD, Dolan RJ, Frith U. 2001. Reward value of attractiveness and gaze. *Nature* **413**(6856):589.
274. Knutson B, Adams CM, Fong GW, Hommer D. 2001. Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *J Neurosci* **21**(16):RC159.
275. Mesulam MM, Geschwind N. 1976. Disordered mental states in the postoperative period. *Urol Clin North Am* **3**(2):199–215.
276. Mesulam MM, Mufson EJ. 1985. The insula of Reil in man and monkey. In *Cerebral cortex*, Vol. 4, pp. 179–226. Ed. A Peters, EG Jones. Plenum Press, New York.
277. O'Doherty J, Kringelbach ML, Rolls ET, Hornak J, Andrews C. 2001. Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature Neurosci* **4**(1):95–102.
278. Pliakas AM, Carlson RR, Neve RL, Konradi C, Nestler EJ, Carlezon Jr WA. 2001. Altered responsiveness to cocaine and increased immobility in the forced swim test associated with

elevated cAMP response element-binding protein expression in nucleus accumbens. *J Neurosci* **21**(18):7397–7403.

279. Volkow ND, Ding YS, Fowler JS, Wang GJ, Logan J, Gatley JS, Dewey S, Ashby C, Lieberman J, Hitzemann R. 1995. Is methylphenidate like cocaine? Studies on their pharmacokinetics and distribution in the human brain. *Arch Gen Psychiatry* **52**(6):456–463.
280. Volkow ND, Wang GJ, Fowler JS, Logan J, Angrist B, Hitzemann R, Lieberman J, Pappas N. 1997. Effects of methylphenidate on regional brain glucose metabolism in humans: relationship to dopamine D2 receptors. *Am J Psychiatry* **154**(1):50–55.

6

A NEUROMORPHIC SYSTEM

David P. M. Northmore
Department of Psychology, University of Delaware, Newark

John Moses
Biophysics Group, Los Alamos National Laboratory, Los Alamos, New Mexico

John G. Elias
Department of Computer and Electrical Engineering, University of Delaware, Newark

The essential functions of neurons can be emulated electronically on silicon chips. We describe such a neuron analogue, or neuromorph, that is compact and low power, with sufficient flexibility that it could perform as a general-purpose unit in networks for controlling robots or for use as implantable neural prostheses. We illustrate some possible applications by a dynamical network that recognizes spatiotemporal patterns and by a network that uses a biologically inspired learning rule to develop sensory-guided behavior in a moving robot. Finally, design requirements for neuromorphic systems are discussed.

1. INTRODUCTION: ARTIFICIAL NERVOUS SYSTEMS

In the quest for alternative forms of computing, especially computing that generates useful behavior in the real world, one naturally looks to biology for inspiration. While brains are economical in size and energy consumption, they depend for their computing power and speed on very large numbers of process-

Address correspondence to: John G. Elias, Department of Computer and Electrical Engineering, University of Delaware, 105 Evans Hall, Newark, DE 19716 (elias@ee.udel.edu).

ing elements richly interconnected. At the present state of technological development, the best way of emulating brains, and the behavior they generate is by constructing "neuromorphs" (14), mimics of neurons fabricated in silicon with VLSIs (very large scale integrated circuits), and interconnecting them with a richness which approaches that of the central nervous system. Using current technology, neuromorphic systems could be the brains of smart structures—small devices that sense the real world and behave adaptively in it, or implanted in humans to repair or extend their capabilities.

Part of the impetus for building artificial nervous systems is to gather insights into how real nervous systems work (see Part III, section 5, this volume). We argue that neuromorphic systems could foster understanding in ways that are not easily achieved, if at all, by conventional digital computer simulations. A common view is that in trying to understand intelligent systems a concern for the details of implementation (i.e., the hardware) is unnecessary. We do not subscribe to this view, but think that the nature of the neural machinery is closely bound up with the solutions that have evolved to perform perceptual, cognitive, and motor tasks. The kinds of computing operations these most naturally support are therefore very different from those supported by a von Neumann architecture. Neuromorphs, such as ours, patterned after biological neurons, depend upon spike processing of information, giving them powerful signal processing capabilities, indeed more powerful than the typical sigmoidal units used in artificial neural networks (10). Rather than review the field of neuromorphic engineering generally, we refer the reader to some recent review articles and collected papers (2,11,21), and instead focus on our neuromorphic system and the approaches taken in our laboratories to some promising applications.

At least initially, neuromorphic systems are likely to be used where compactness and low power consumption are at a premium, for example, as the brains of autonomous vehicles and for neural prosthetics that interface between the nervous system and artificial effectors or a patient's own musculature. Therefore, in thinking about what kinds of networks to investigate, an underlying consideration is that small systems must make good use of their resources. Dynamic networks are of interest because modest numbers of interconnected neuromorphs should be able to store and process large amounts of information in the transition of states they undergo. In the control of autonomous vehicles, a fast, compact system is required that can efficiently utilize information about the environment. We are therefore exploring ways in which neuromorphic networks can learn to behave adaptively through sensorimotor experience.

2. THE NEURON AND THE NEUROMORPH

The design of our artificial dendritic tree (ADT) neuromorph is based on what one might call the classical neuron as conceptualized in the mid 1950s (3). The input structures of this neuron are the branched processes forming the dendritic tree, and to a lesser extent the cell body or soma from which the dendrites

extend. The neuron receives input in the form of impulses or spikes at numerous synaptic sites over these surfaces. The output structure is a spike-firing fiber or axon that may extend for considerable distances in the nervous system.

An input spike activating a synapse generates a brief postsynaptic potential within the dendrite by opening ion channels in the cell membrane. The potential may be of either polarity depending on which ion channels are involved. If it is positive-going, the effect is excitatory because it will tend to increase the spike-firing frequency of the axon; if negative it is inhibitory, tending to reducing spike firing. The dendritic branches are considered to be passive cables over which the postsynaptic potentials mingle and diffuse. Their function, then, is to collect many synaptic inputs, delaying, attenuating and summing the synaptic potentials generated. The net potential change that accrues at the junction of soma and axon determines the rate of firing of spikes emitted as output along the axon.

We now know that dendrites, which vary greatly in form across cell types, vary also in function, and are not usually passive, boost the transmission of potentials along them with voltage-sensitive ion channels. Nevertheless, theoretical analyses show that even passive dendrites are able to perform useful spatiotemporal filtering, allowing discrimination of different input spike patterns (19). Experiments with our ADT neuromorphs, which are analogues of the classical neuron, showed that they could be connected so as to respond selectively to patterns of input spikes, for example, to specific frequencies of an input spike train or to specific temporal orderings of spikes (17). In designing the neuromorph, we saw the role of the dendritic tree as very important to the computing power of the device, as we now know it to be in neurons (15). While the function of dendrites in the latter is complex, depending as it does on a multiplicity of molecular and ionic mechanisms, the simple passive dendrites that have been modeled form a starting point for exploiting the ingenuity of neuronal architecture. Fortunately, dendrites in a variety of spatial configurations can be readily fabricated in a VLSI. Neuromorphic modeling is, after all, very much the art of the possible.

The dendrites' filtering properties are strongly influenced by their dynamics—the resistances and capacitances that determine the time course of the postsynaptic potentials. Being able to control dynamics enables one to lengthen or shorten branches of the artificial dendrite (5), providing a way to vary the functional properties of different pools of units. Another parameter of great importance to control is a unit's excitability or spike-firing threshold. It is desirable to be able to modulate the spike-firing threshold with the spiking activity of other units (6). A capability for learning by altering the efficacy of synaptic inputs is important, but at this stage of hardware development it is probably best done off-chip, where different learning rules can be tested. In sum, our aim is to produce a neuron analog with sufficient flexibility that it could perform as a general-purpose unit, adaptable to many uses in a central neuromorphic system.

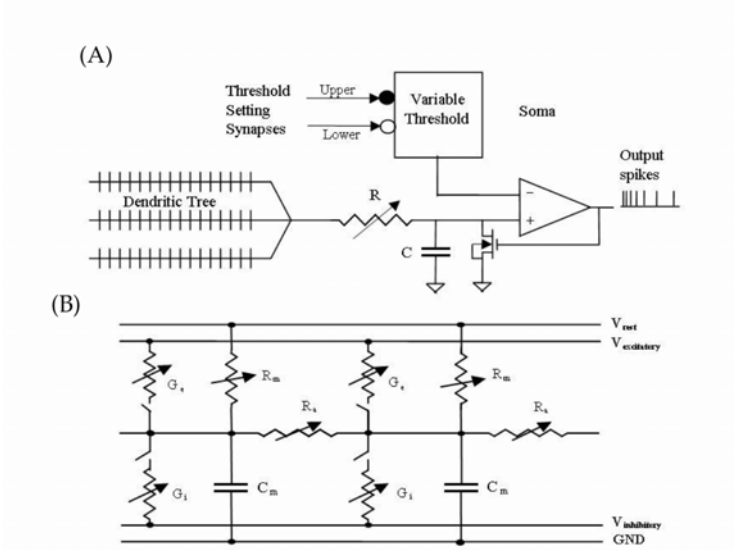


Figure 1. (**A**) Schematic of ADT neuromorph. Synapses are located at the cross points on the 3-branch dendritic tree and at the soma. Activating soma synapses sets the spike-firing threshold for the integrate-and-fire spike generator, whose integration time constant is determined by the programmable resistor, R, and fixed capacitor, C, which is discharged whenever a spike is generated. (**B**) A two-compartment section of dendrite. Each compartment contains a membrane capacitance (C_m), a membrane resistance (R_m) that connects to V_{rest}, and axial resistances (R_a) that connect to adjacent compartments. The excitatory (G_e) and inhibitory (G_i) synaptic conductances, which turn on momentarily when a synapse is activated, pull the compartment capacitor voltage towards $V_{excitatory}$ and $V_{inhibitory}$, respectively. The label GND corresponds to the common voltage node to which all other nodes are referenced.

3. HARDWARE SYSTEM

In our ADT neuromorphs fabricated by VLSI, the dendritic branches are modeled electrically by arrays of capacitors and resistors (Figure 1B) (4,5). The dendrites are composed of multiple compartments, each with a capacitor, C_m, representing a membrane capacitance, and two programmable resistors, R_m and R_a, representing a membrane resistance and an axial or cytoplasmic resistance. In most of our neuromorph chips (Figure 2), each dendritic branch has sixteen compartments, with 3–8 branches connected together to form a tree like that shown in Figure 1A. Every compartment has an excitatory and an inhibitory synapse formed by MOS (metal oxide semiconductor) field-effect transistors that enable brief currents into or out of the compartment. Applying an impulse to the transistor gate turns on a synapse. The resultant "transmembrane" current depends upon the potential difference across the transistor, and upon the synaptic weight, which can be controlled (a) by the conductance of the transistor in

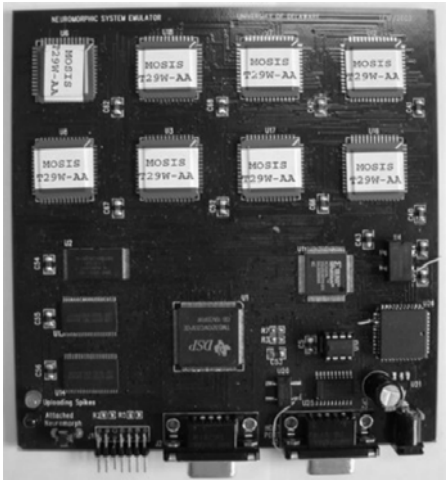


Figure 2. Domain board. Dimensions: 14.5 cm^2. The top half of the board is occupied by 8 chips (0.5-μm CMOS (complementary metal oxide semiconductor)), each containing 16 neuromorphs, each of which containing 4 dendritic branches. Most of the 2.2 × 2.2 mm chip area is filled by dendrites. The lower half of the board contains the spike-routing system, connection memories, and connectors for a host computer.

the activated or on state, and (b) by the duration of the impulse supplied to the transistor's gate. Activation of excitatory and inhibitory synapses at different sites set up potentials that diffuse over the dendritic tree. The potential at the point where the branches converge (Figure 1A) affects the rate of output spike firing. The dynamics, which determine the dendritic delays, are set by the programmable membrane and axial resistances (R_m) and (R_a), and by the fixed-value compartmental capacitance (5). Whether the neuromorph fires or not and its rate of firing depend upon whether and how much the branch-point potential exceeds a firing threshold implemented by a comparator in the "soma" (Figure 1A). For this threshold potential to be influenced by the activity of other neuromorphs in a network, it is set by the ratio of spike frequency applied to two special synapses, the "upper" and "lower" in Figure 1A (6).

To build a network, one must provide a means for sending the output spikes of any neuromorph to the synapses of any other. In addition, the spikes must arrive at their destinations in a time that is short compared to network dynamics. To this end, we developed a multiplexing scheme that we call "Virtual Wires" (4). As with Mahowald's (12) method of connecting neuron outputs to synapses, addresses are multiplexed in time over a few dedicated wires rather than through a much larger number of direct connections. In our scheme, a spike generated by a neuromorph activates destination synapses with delays that are programmable for each connection. The number of connections each neuromorph may have is

limited only by the overall bandwidth (i.e., the multiplexing rate) of the system. Connections to synapses can be programmed to have one of 30 different weights, 15 inhibitory and 15 excitatory (24). During activation of a synapse, its conductance changes from an essentially nonconducting state to one of the 30 conductance values for a duration of 50 ns. Weights for each synapse are stored off-chip in a connection list along with the synapse address, thereby allowing multiple connections with independent weights to the same compartment.

The present system shown in Figure 2 holds 128 neuromorphs. However, with higher-density neuromorph chips it is capable of simultaneously interconnecting and running over 1000 neuromorphs. The entire network connectivity can be changed in a fraction of a second, allowing rapid evaluation of a large number of different architectures. For this purpose, the spiking activity from a selected set of neuromorphs can be sampled at any time.

4. NEUROMORPHS IN A WINNERLESS COMPETITION NETWORK

The coordinated firing of neurons in a network is likely to be a key feature of neural computation. Interconnected neurons with steady input behave as coupled oscillators, and such systems can display a range of coordinated behaviors. The phenomenon of synchronous firing of groups of neurons (8) is probably the best known example of this. Other, more complex patterns of spikes can result from the interaction between neurons—the behavior depends on the details of the coupling. Such networks are strong candidates for computational mechanisms that biological systems may actually use for representation and recognition of patterns (16). One that has attractive features for application to small networks of neuromorphs is the winnerless competition network of Rabinovich et al. (23). In the absence of a stimulus, these networks are quiescent or display unpatterned activity, but when stimulated they generate cyclic patterns of spiking activity that are distinct and characteristic of the stimulus. These networks achieve spatiotemporal coding by executing heteroclinic orbits around saddles in a space that changes when the pattern of stimuli changes. The behavior, which captures some features observed of olfactory processing in the locust antennal lobe, was generated in a simulation with nine model neurons (FitzHugh-Nagumo) with strong asymmetric inhibitory interconnections.

We have produced similar behavior in a network of neuromorphs interconnected and stimulated similarly to the model neurons in the simulation of Rabinovich et al. The excitability of each neuromorph in the network is controlled by (a) supplying a 100-Hz spike train to the "lower" threshold synapse (Figure 1A) and (b) feeding back the neuromorph's output spikes to the "upper" threshold synapse. The ratio of "upper" to "lower" spikes determines the spike-firing threshold; the negative feedback limits the firing rate, giving the desired background behavior (see (6)). The temporal pattern of spiking of the neuromorphic

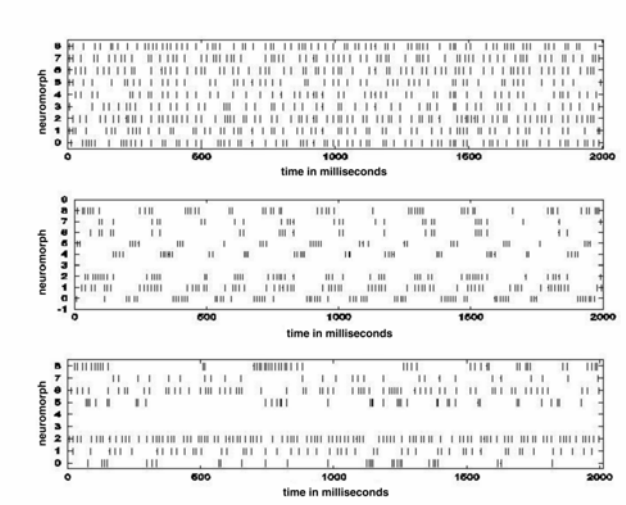


Figure 3. Spike rasters of nine neuromorphs in a winnerless competition network. **Top**: Spiking in the absence of stimulus and interconnections. **Middle**: Neuromorphs connected via asymmetric inhibitory synapses and driven by a fixed pattern of stimuli. Periodic behavior is evident. **Bottom**: Same network, but with the stimuli rotated by one neuromorph. Note a different, weaker cyclic behavior.

network is shown in the series of raster plots in Figure 3. The top plot illustrates quiescent behavior in the absence of any coupling or stimulus—the units are firing fairly steadily and independently of each other. When the neuromorphs are interconnected and driven with spike patterns corresponding to the inputs used by Rabinovich et al. (23), a cyclic pattern of firing results, shown in the middle plot of Figure 3. The neuromorphs fire in a pattern that repeats with a period just under 200 milliseconds. It is easily recognized by eye, and presumably could be easily recognized by a downstream neural network. Changing the stimuli by doubling or halving their spike frequencies compresses or expands the pattern, but does not change it qualitatively. Changing the stimuli by shifting the connections by one or two neuromorphs changes the pattern, as shown in the bottom plot of the Figure 3. The network producing this pattern of spikes is interconnected as before, but the stimuli are displaced cyclically by one unit. The effect is dramatic. There is a weak appearance of periodic firing by several of the neuromorphs, but two or three seem to be firing at a roughly steady rate, and two are not firing at all. The network strongly selects between the two stimuli, and thus serves to identify the inputs, especially for the conditions of the middle plot.

The number of patterns that can be discriminated in this way is related to the number of combinations of neurons that can be formed by the network, and is far larger than the number of fixed points in a static network of the same size (23). For computations with small networks—and with neuromorphs we are limited to networks that are small compared to biological networks—a dynamical network of this type is capable of much greater capacity as a memory or for pattern recognition. We may expect that networks exploiting this kind of dynamical behavior will play a significant role as applications are realized.

5. SENSORIMOTOR DEVELOPMENT IN A NEUROMORPHIC NETWORK

In this section we present an experimental system of simulated neuromorphs that develops sensorimotor capabilities. It was studied to find ways of using neuromorphs for controlling the behavior of an autonomous vehicle, but the same principles could be applied in a neuromorphic prosthetic where arbitrary patterns of spikes need to be interpreted and used to control artificial effectors or the patient's own muscles.

We have been investigating ways in which development could be performed in neuromorphic circuits using principles that are neurobiologically plausible. In animal nervous systems, the patterns of neural activity that emerge during development, influenced both endogenously and exogenously, play a vital role in establishing normal processing. While there is no compelling reason to make the development of a neuromorphic "brain" or prosthetic neurobiologically plausible, using mechanisms evolved over eons would appear to be a good strategy.

If one can point to a single principle underlying both the development of neural connections and the subsequent adjustment of connections during learning, it would be the rule ascribed to the psychologist, D.O. Hebb. In essence, the rule requires the strengthening of the connection between a sending, or presynaptic neuron and a receiving, or postsynaptic neuron if the two fire together close together in time. Otherwise, weakening of the connection may result, depending on which of several formulations of the rule are applied. Recent neurobiological work has shown that the relative timing of pre- and postsynaptic spiking is indeed responsible for potentiation and depression of synaptic efficacy or weight (13,25).

To illustrate the neuromorphic approach, we present results on a simulation of a network of spiking neurons that approximate the behavior of our silicon neuromorphs. Previous theoretical work showed that the application of Hebb rules can lead to the development of adaptive connections and realistic sensory receptive fields (9). Although the intent here is to develop a controller for an autonomous vehicle, the results demonstrate for a system of spiking neurons how inputs from sources with very different characteristics can combine automatically to generate order and useful functionality—in this case directional

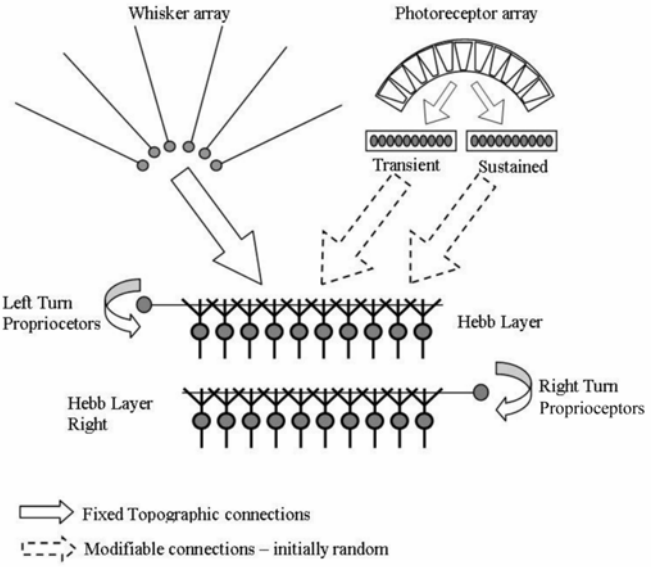


Figure 4. Schematic of sensory system. Both whisker and photoreceptor arrays sample the frontal 120° of space. The spiking whisker units project topographically to both Hebb layers. The photoreceptors drive the transient and sustained spiking layers that are connected to the Hebb layers via modifiable, initially random connections. Each Hebb layer is excited by proprioceptive spikes signaling left or right eye/head turns.

selectivity. In those sensory systems that have topographic arrangements (i.e., visual, somatosensory, auditory), neurons are found that respond selectively to one direction of motion, whether it be in space or in some other dimension like sound frequency. Such selectivity is important for higher neural processing of all kinds, for example, our ability to use vision to guide ourselves through our environment, or to recognize speech sounds.

6. SIMULATED NETWORK

The simulated vehicle is equipped with two kinds of exteroceptors—whiskers and photoreceptors—and proprioceptors that signal the vehicle's own motion. Figure 4 shows the arrangement of sensors and the network of spiking neurons whose activity can be used to guide movements of the vehicle as it roams around an environment with obstacles.

The six whiskers and the 10–20 photoreceptors are fixed to the vehicle and arrayed to sense the frontal 120° in the horizontal plane. Each whisker, when it

contacts an obstacle, generates a spike train at a frequency inversely proportional to the distance of intersection along the whisker length. The photoreceptors, which respond proportionately to light intensity integrated over a receptive field with Gaussian profile, drive two kinds of spiking units, the sustained and transient units, analogous to the main types of ganglion cells in the vertebrate retina. The receptive fields of these spiking units are about 10° wide.

The whisker and the photoreceptor spike signals connect to two layers of simulated neuromorphs, Hebb Layer Left and Hebb Layer Right, which, by a Hebbian learning rule, are to acquire directionally selective properties to stimuli moving left or right respectively. These layers are composed of 8 units, each of which is based on the spike–response (S–R) model of Gerstner and Kistler (8). This model provides an approximation to the functional properties of a silicon neuromorph limited to one operative synaptic site. A spike input (from unit *i*) to one of these S–R units (unit *j*) generates a postsynaptic potential that is positive-going (excitatory) or negative-going (inhibitory), the amplitude and sign depending on a numeric weight (w_{ij}) between –1.0 and +1.0. The rise and fall times of the postsynaptic potentials are fixed, unlike those in the ADT neuromorph, in which they depend on the site of synaptic activation along the dendrite. Postsynaptic potentials are summed and fed to an integrate-and-fire spike generator, the firing threshold of which is controlled by a bias input. Refractoriness after spike firing is implemented by immediately raising the spike-firing threshold and allowing it to decay to the bias threshold over a fixed time course.

The connections from the whisker array to both Hebb layers are preset and not modifiable. The whiskers are mapped topographically across the Hebb layers so that stimulation of the leftmost whisker, for example, activates units at the leftmost end of the Hebb layers (see Figure 6A). Hebb layer units that are not related topographically to a whisker are inhibited by that whisker. The other preset and unmodifiable connections are the proprioceptive afferents to the Hebb layers. The left-turn proprioceptors fire when the vehicle rotates leftward and tends to excite all the units in Hebb Layer Left. Similarly the right-turn proprioceptors excite Hebb Layer Right. The transient and sustained units of the "retina" are projected in a fully connected pattern with initially zero weights onto both Hebb layers. The latter connections are modifiable according to a Hebb rule akin to the mechanism that strengthens and weakens synapses in mammalian cortex (13) and amphibian tectum (25). Synaptic weight modification depends upon the relative timing of pre- and postsynaptic spikes: if the presynaptic spike occurs in a time window before the postsynaptic unit fires, the weight of the presynaptic synapse is increased; if it occurs after the postsynaptic spike, it is decreased; otherwise, no change occurs (Figure 5). The effect is that the synapses from any visual units that fire consistently just before the firing of a Hebb layer unit will have their connections strengthened or made more excitatory. If the visual units fire just after, their connections are weakened and eventually may become inhibitory. Visual unit firing that is unrelated to the activity of the

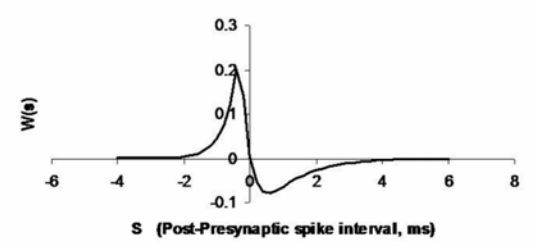


Figure 5. Hebbian window function. Synaptic weight is increased in proportion to $W(s)$. When a presynaptic spike precedes a postsynaptic spike by less than about 2 ms, $W(s)$ is positive; when a presynaptic spike follows a postsynaptic spike by up to about 4 ms, $W(s)$ is negative; it is otherwise vanishingly small (see (8)).

Hebb layer units results in no net change of weight if the positive and negative phases of the window function (Figure 5) integrate to zero.

Weights onto unit i that have fired recently are updated by an amount Δw_{ij} according to

$$\Delta w_{ij} = \varepsilon(W(s_{ij}) - w_{ij}), \qquad [1]$$

where ε is the learning rate, s_{ij} is the time interval between the firing of unit i and j, and $W(s)$ is the function shown in Figure 5. Subtraction of the existing weight w_{ij}, ensures that weights are soft bounded, i.e., they do not grow without limits (cf. the CPCA learning rule (18)). The repeated application of this rule leads to stable weights but an excess of excitatory drive to the Hebb layer units. To maintain an equitable balance between excitation and inhibition, weight normalization is applied to make all the weights onto each unit sum to 1.

In the first phase of development, the vehicle traverses the field populated by obstacles of various sizes, advancing continuously in space by an equal distance every simulation time step. For the present demonstration, the vehicle's steering is disabled, so that it moves in a straight line, brushing against randomly placed obstacles with its whiskers while its "retina" generates visual impulses that have no effect on the Hebb layers. At this stage, we need only consider the sustained units and their connections. The connections from the whiskers, which are topographically mapped onto the Hebb layers (Figure 6A), are sufficiently strong to fire the Hebb layer units reliably. The application of the learning rule results in strengthening of the connections from those visual units that were active shortly before the Hebb layer units spiked. Weights from the visual units are continuously updated so that the sustained visual units come to map topographically over both Hebb layers. This is shown in the weight matrices of Figure 6B by the diagonal of excitatory weights. The effect of weight normalization is to

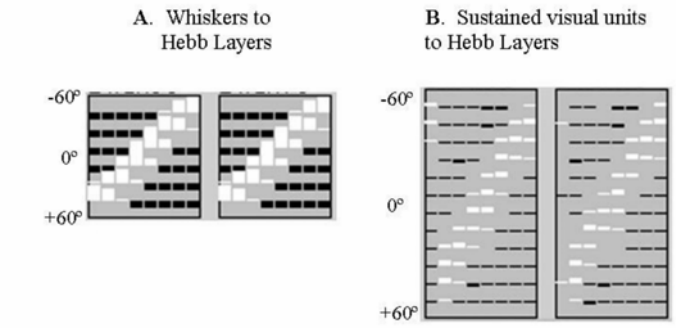


Figure 6. Weight matrices connecting (**A**) whiskers and (**B**) visual sustained units to the left and right Hebb Layers. Excitatory weights are shown in white, inhibitory in black. Height of bars represent weight values. Inputs in the whiskers and visual space (–60 to +60°) are represented along rows; the different Hebb layer units along columns.

form an inhibitory surround. If the spiking activity of the Hebb layers is now connected to the vehicle's steering, the vehicle avoids obstacles visually, without contact with the whiskers.

Negotiating a complex environment requires that information other than just luminance is brought to bear; judging distances, reacting to impending collisions and guiding locomotion require the use of visual motion cues. The network of Figure 4 and the application of Hebbian mechanisms during sensorimotor experience results in the development of motion selective units. For this, we added transient visual units that fire to changes in luminance. Because the extraction of velocity information depends on correlations between stimulus positions at different times, the spikes from the transient units are connected to the Hebb layers via different time delays. For the sake of illustration, the transient unit array is connected to the Hebb layers via four weight matrices, one each for delays of 0, 0.5, 1.0, and 1.5 ms (Figure 7A,B). In a neuromorphic network, these delays could be implemented by the facility for programmable axonal delays in the Virtual Wires system, or by dendritic delays that depend on the site of synaptic activation along a dendrite. Under appropriate conditions, the application of the Hebb rule in the presence of a moving stimulus image could selectively strengthen inputs that come from the four positions in space occupied by the image at times t, $t - 0.5$, $t - 1.0$, and $t - 1.5$ ms. For Hebb layer units to come to respond to the motion, they must be firing at time t. Therefore, they require an "instructive" input generated during image movement to bias them on so that they fire. In this example, the image motion is self-produced, as by eye or head

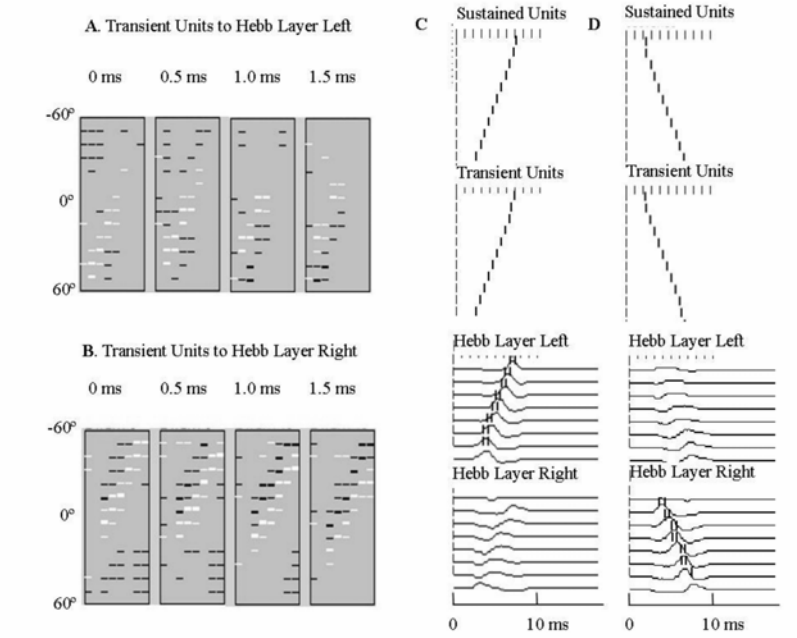


Figure 7. (**A,B**) Weights of the transient units onto the Hebb Layers with delays of 0, 0.5, 1.0, and 1.5 ms. White indicates excitatory weights; black inhibitory. Inputs in visual space (–60 to 60°) are represented along rows; the different Hebb layer units along columns. (**C,D**) Responses as a function of time. Spikes of sustained and transient visual units, and the two Hebb Layers in response to a stimulus rotating from right to left (C) and from left to right (D) through the entire visual field. Waves in the lower half show the unit's internal potentials.

movements while looking at stimulus objects in space. The sources of the instructive inputs are proprioceptors that signal the vehicle's own turning movements. The instructive inputs could also be derived from motor command signals. In the network of Figure 4, proprioceptors for leftward or rightward movement fire a burst of spikes to all the units of their corresponding Hebb layer. Thus, the convergence of transient and proprioceptive information onto Hebb layer units should result in their acquiring selectivity to one direction of motion, but without regard to position in the visual field. Extracting depth from parallax, however, likely requires that the spatial arrangement of different velocity vectors be preserved. As we have seen, the topography was established in mapping of the sustained units over the Hebb layers, so if the sustained inputs also bias the Hebb units to fire, along with the proprioceptive inputs, directionally selective maps with topography should be formed.

Figure 7 shows examples of how weights from the transient units to the Hebb layer units developed. The four weight matrices at increasing delays show progressive shifts in their receptive fields, thereby generating a skewed spatiotemporal receptive field required for directional selectivity (1). To test directional selectivity, a single visible object was rotated at a constant rate through the visual field. Figure 7C,D (lower half) shows that Hebb Layer Left responded with spikes to leftward motion but not to rightward motion, and vice versa for Hebb Layer Right. Note also that the units have restricted receptive fields organized topographically, a first step towards analyzing motion flow fields. We can now envision some further steps that would be required to use this information for behavioral guidance: the development of selectivities for different velocities and different patterns of motion as exhibited by neurons in the visual areas of the cerebral cortex.

7. NEUROMORPHS IN NEURAL PROSTHETICS

Neuromorphic systems of the type we are developing may have a special advantage for neural prosthetics in that neuromorphs naturally deal with spike signals: they accept spike inputs and they generate spike outputs. A very small neuromorphic network implanted in a paralyzed patient could interpret spike signals from an array of electrodes embedded in the motor cortex (20). The network would extract the relevant information contained in the frequency and timing of the cortical spikes and would generate output spike trains patterned so as to activate neurons, fiber tracts, or muscles, restoring lost function to the patient. Neuromorphic prosthetics will certainly require the ability to adapt and learn because the input spikes are largely arbitrary in nature—they are whatever the implanted electrodes can pick up from the CNS—and they must be associated with patterns of spike outputs for controlling behavior (22). While our neuromorphic development system (domain board) can implement arbitrary learning schemes, it currently requires the assistance of an external computer.

8. CONCLUSIONS

The challenge to building neuromorphic systems is deciding what features to incorporate into the neuromorphic units and how to connect them to perform usefully. A process of design from biophysical principles is an estimable approach but difficult (7). Faced with the great variety of potential mechanisms that neurobiological research is revealing and the complex, nonlinear interactions between them, we prefer to seek neuromorphic systems that are relatively simple to make and able to self-organize and adapt dynamically. The results presented here show that valuable capabilities can emerge in networks through

interaction with the environment, in effect, on-the-job training. To be sure, quite a bit of structure has to be imposed for this development to happen. How to come up with the appropriate structures is an interesting problem. One strategy is to borrow from what is known of structure in brains; an alternative is to evolve structure through evolutionary methods. Another problem is how to regulate excitabilities so that development can occur in an adaptive fashion. As we have seen, for anything sensible to be learned by Hebbian mechanisms, units must be biased so that they are driven to fire when learning is appropriate and inhibited when it is not. Brains seem to employ diffuse neuronal projections and humoral controls for this purpose and these could well be emulated in neuromorphic systems. Among the technical problems that need to be solved for neuromorphic systems to gain independence in small-scale applications is how to store and modify connections locally or "on chip" so that learning can occur without the intervention of external, digital computers. Notwithstanding these problems and unknowns, the hardware system described achieves, to greater or lesser extents, the following desirable features of a neuromorphic system:

- Parallelism of processing elements for fast processing of large amounts of information
- Distributed processing whereby many units simultaneously represent a sensory stimulus or execute a movement
- Robustness in the face of variation and faults in individual components, tolerance to damage and noise, and performance degrading gradually, not catastrophically
- Real-time processing with wide time-scale dynamics that can be matched to events in the real world, and to the dynamics of limbs and other body systems
- Flexible, rapidly reprogrammable architecture and connections; capable of ontogenetic and evolutionary adaptation
- Learning based on short-, medium-, and long-term changes in synaptic efficacy, unit excitability, and persistent patterns of activity
- Efficient, noise-immune signaling by impulses with rich temporal coding possibilities
- Powerful analog computation distributed throughout a dendritic tree and soma
- Compact size and low power consumption, essential for mobility and portability

9. REFERENCES

1. Adelson EH, Bergen JR. 1985. Spatiotemporal energy models for the perception of motion. *J Opt So Am [A]* **2**:284–299.
2. Douglas R, Rasche C. 2002. Silicon neurons. In *The handbook of brain theory and neural networks*, pp. 128–145. Ed. M Arbib. MIT Press, Cambridge.
3. Eccles JC. 1957. *The physiology of nerve cells*. Johns Hopkins UP, Baltimore.
4. Elias JG. 1993. Artificial dendritic trees. *Neural Comput* **5**:648–664.
5. Elias JG, Northmore DPM. 1995. Switched-capacitor neuromorphs with wide-range variable dynamics. *IEEE Trans Neural Netw* **6**:1542–1548.
6. Elias JG, Northmore DPM, Westerman W. 1997. An analog memory circuit for spiking silicon neurons. *Neural Comput* **9**:419–440.
7. Elias C, Anderson CH. 2002. *Neural engineering*. MIT Press, Cambridge.
8. Gerstner W, Kistler W. 2002. *Spiking neuron models*. Cambridge UP, Cambridge.
9. Linsker R. 1990. Perceptual neural organization: some approaches based on network models and information theory. *Annu Rev Neurosci* **13**:257–281.
10. Maass W. 1997. Fast sigmoidal networks via spiking neurons. *Neural Comput* **9**:279–304.
11. Maass W, Bishop CM. 1999. *Pulsed neural networks*. MIT Press, Cambridge.
12. Mahowald MA. 1992. Evolving analog VLSI neurons. In *Single neuron computation*, pp. 413–435. Ed T McKenna, J Davis, SF Zornetzer. Academic Press, San Diego.
13. Markram H, Lubke M, Frotscher B, Sakmann B. 1997. Regulation of synaptic efficacy by coincidence of postsynaptic Aps and EPSPs. *Science* **275**:213–215.
14. Mead C. 1989. *Analog VLSI and neural systems*. Addison-Wesley, Reading, MA.
15. Mel BW. 1994. Information processing in dendritic trees. *Neural Comput* **6**:1031–1085.
16. Natschläger T, Maass W, Markram H. 2002. The "liquid computer": a novel strategy for real-time computing on time series. *TELEMATIK* **8**:39–43.
17. Northmore DPM, Elias JG. 1996. Spike train processing by a silicon neuromorph: the role of sublinear summation in dendrites. *Neural Comput* **8**:1245–1265.
18. O'Reilly RC, Munakata Y. 2000. *Computational explorations in cognitive neuroscience*. MIT Press, Cambridge.
19. Rall W. 1964. Theoretical significance of dendritic trees for neuronal input–output relations. In *Neural theory and modeling*, pp. 73–79. Ed. RF Reiss. Stanford UP, Stanford.
20. Serruya MD, Hatsopoulos NG, Paninski L, Fellows MR, Donoghue JP. 2002. Instant neural control of a movement signal. *Nature* **416**:141–142.
21. Smith LS, Hamilton A. 1998. *Neuromorphic systems: engineering silicon from neurobiology*. World Scientific, Singapore.
22. Taylor DM, Tillery SIH, Schwartz AB. 2002. Direct cortical control of 3D neuroprosthetic devices. *Science* **296**:1829–1832.
23. Rabinovich M, Volkovskii A, Lecanda P, Huerta R, Abarbanel HD, Laurent G. 2001. Dynamical encoding by networks of competing neuron groups: winnerless competition. *Phys Rev Lett* **87**:068102–068104
24. Westerman WC, Northmore DPM, Elias JG. 1997. Neuromorphic synapses for artificial dendrites. *Analog Integrated Circ Signal Proc* **13**:167–184.
25. Zhang LI, Tao HW, Holt CE, Harris WA, Poo, M. 1998. A critical window for cooperation and competition among developing retinotectal synapses. *Nature* **395**:37–44.

7

A BIOLOGICALLY INSPIRED APPROACH TOWARD AUTONOMOUS REAL-WORLD ROBOTS

Frank Kirchner and Dirk Spenneberg

Department of Computer Science,
University of Bremen, Bremen, Germany

We present an approach inspired by biological principles to design the control system for an eight-legged walking robot. The approach is based on two biological control primitives: central pattern generators and coupled reflexes. By using these mechanisms we can achieve omnidirectional walking and smooth gait transitions in a high-degree-of-freedom (14) walking machine. Additionally, the approach allows us to freely mix rhythmic activity with posture changes of the robot without reducing forward speed. This approach has proved to be extremely successful on rough terrain and has been evaluated in real-world tests over a variety of different substrates,

1. INTRODUCTION

The focus of our approach was to develop an eight-legged robot (see Figure 1), which should be capable of robust locomotion on extremely difficult terrains (rock, sand, inclinations, and combinations thereof).

The approach features a hierarchical control architecture that combines biologically inspired principles of locomotion control, such as central pattern generators and reflex mechanisms with a quasi-parallel behavior-based approach (for biological background on the acquisition of motor behavior, see Part III, chapter 5.3, by Kolb and Timmann, this volume). The system also scales upward

Address correspondence to: Frank Kirchner, Department of Computer Science, University of Bremen, Building MZH3100, Bibliothekstrasse 1, 28350 Bremen, Germany (frank.kirchner@ informatik.uni-bremen.de)

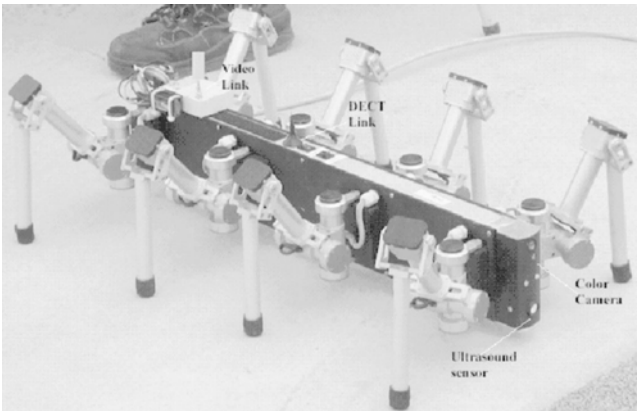


Figure 1. The eight-legged robot "Scorpion."

to include higher-level principles of robot control such as planning and scheduling algorithms. This combination of biologically robust low-level control with high-level planning algorithms provides an interesting basis for many biomedical applications, from "Neuromorphic Systems" (see preceding chapter 6, by Northmore, Moses, and Elias) to "Intraoperative Navigation" (see following chapter 8, by Heilbrun).

2. MECHATRONICS

The most challenging parts of a walking system are the legs (Figure 2). The leg design presented here provides 3 degrees of freedom, which is the minimum needed for a robust, outdoor walking robot, i.e., it provides the possibility for omnidirectional walking even in narrow, difficult environments. The legs consist of a thoracic joint for protraction and retraction, a basilar joint for elevation and depression, and a distal joint for extension and flexion of the leg (Figure 2). The joints are actuated by standard 6-W/24-V DC motors with a high gear transmission ratio for sufficient lifting capacity.

Outdoor capability was an important constraint in the development process. Therefore, a good tradeoff between lightweight design—to achieve maximum lift capacity—and shielding against environmental influences (like dust and water) had to be found. We achieved a weight of 950 grams and a weight:lift capacity ratio of 1:8. This proved to be sufficient for the system to walk up steep inclinations and/or walk over obstacles higher than the robot itself. It also proved to be important to integrate compliant elements in the mechanical design in order to make the robot sufficiently robust to withstand the mechanical stress of an outdoor terrain.

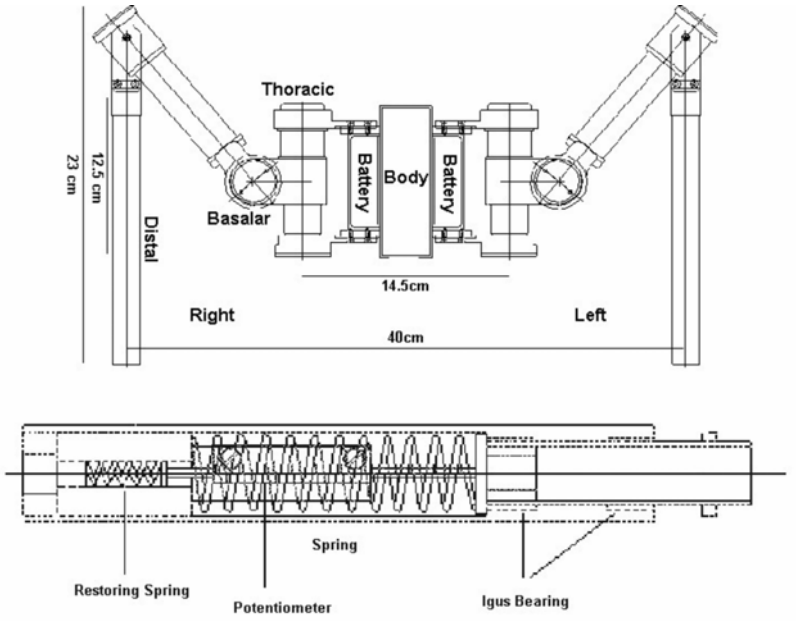


Figure 2. The mechanical design of the Scorpion legs. This front view of the robot shows the left and right side legs with the body in the center. Each leg consists of three joints: (i) thoracic joint, (ii) basilar joint, and (iii) distal joint (right). The distal segment contains a spring-damped compliant element with a built-in potentiometer to measure contact and load on individual legs. The most energy-absorbing part in our design is a spring element integrated in the distal segment of the leg. The distal spring element is also used to measure the ground contact force by means of an integrated linear potentiometer. From this the robot can compute the load on each leg. This enables us to make use of the principle of "Early Retraction Acceleration" (described in (2,13)).

2.1. The Sensors

The robot is equipped with the following sensors that sense the internal state, the so-called proprioceptive sensors:

- Motor encoders for each motor to measure the relative joint angle
- Hall-effect motor current sensors for each motor
- The analogue load/pressure sensor in each foot tip
- Power-management sensors, providing current battery voltage and current power drain
- Three-dimensional inclinometers (pitch, roll, and yaw)

The following sensors to obtain information from the environment, so-called exteroceptive sensors, are integrated:

- An ultrasound distance sensor for obstacle avoidance
- A compass sensor for heading control
- Contact/pressure sensors at foot tips

It is important to note that the legs themselves can be used as exteroceptive sensors. One can use the current sensors of the joint motors as a tactile sensor during movement, e.g., whether a leg presses against an object can be sensed by an increasing current in the corresponding joint motor. This issue is currently under further investigation. In an ongoing project, we are analyzing the robot's capabilities to use its front legs in order to determine shape information of obstacles, which it previously detected with ultrasound sensors.

In order to allow an operator to communicate with the robot or to take data samples during a test run, the robot is equipped with a wireless 28K-baud bidirectional communication link and a PAL CCD camera with a 5-GHz video/audio link for video transmission. It is thus possible to use the robot as a semiautonomous system. The operator can control it via high-level commands like "walk forward," "left," "right," "go up," "go down," "move sideways," and "turn." To supervise the system, all relevant sensor data are sent back from the robot to the operator.

2.2. The Processing Hardware

A network of Infineon C167 and C164 microcontroller derivates are used, containing one master controller (the C167), which functionally contains the higher behavioral level, the communication to the operator, and executes processing of data from the exteroceptive sensors. The master controller is connected via a CAN–bus network with leg controllers (the C164). The leg controller executes local controller functions (via the central pattern generator (CPG)) and local reflex control functions, processes the proprioceptive sensor data, and controls the DC motors. In a new version of the hardware we will use an Motorola MPC555 master controller for higher-level control and an FPGA for local control of the legs.

3. AMBULATION CONTROL

Our architecture (14) is based on two approaches to robust and flexible real-world locomotion in biological systems. These are the central pattern generator (CPG) model and the coupled reflex approach (1,6,7).

A CPG is able to produce a rhythmic motor pattern even in the complete absence of sensory feedback. The general model of a CPG has been identified in

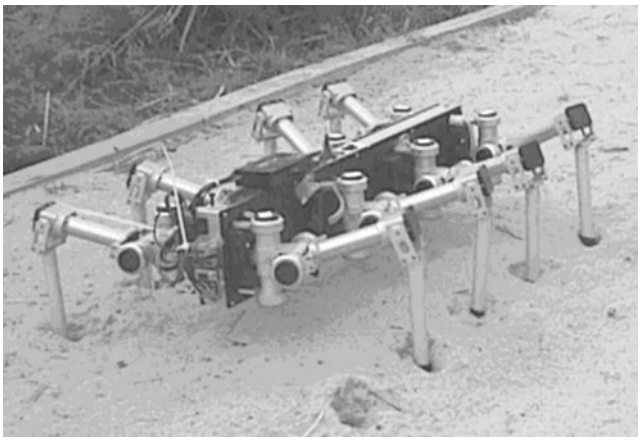


Figure 3. The scorpion robot during autonomous exploration into a sand bed. The beach-like sand bed was 3 m wide and 9 m long. The robot's feet penetrated the sand by ≈3–5 cm. A reflex mechanism helped overcome obstacles.

nearly every species, even though the specific instantiations vary among species to reflect individual kinematical characteristics in the animals.

The idea seems to be very promising as a concept to stabilize locomotion in kinematically complex robotic systems (see Figure 3), as it resembles the divide-and-conquer strategies reflected in nearly all solutions to complex control problems (4).

Another model for support of robust locomotion is provided by evolution in the animal kingdom. This is the concept of reflex-based control (7). A reflex can be viewed as a closed loop control system with fixed input/output characteristics. In some animals, like the locust, this concept is said to actually perform all locomotion control, and no further levels of control, like a CPG, are involved (6).

Whether or not complex motion control can be achieved via only reflex systems is subject to further discussion; however, the concept of a set of fixed, wired reactions to sensory stimuli is of high interest to roboticists who aim to gain stability in a system's locomotion.

The design of the control architecture described here was thus driven by these two concepts. The CPG approach appeared to be interesting in generating rhythmic walking patterns, which can be implemented in a computationally efficient manner, while the reflex-driven approach seemed to provide a simple way to stabilize these walking patterns by providing: (a) a set of fixed situation-reaction rules to external disturbances, and (b) a way to bias leg coordination among multiple independent legs (6).

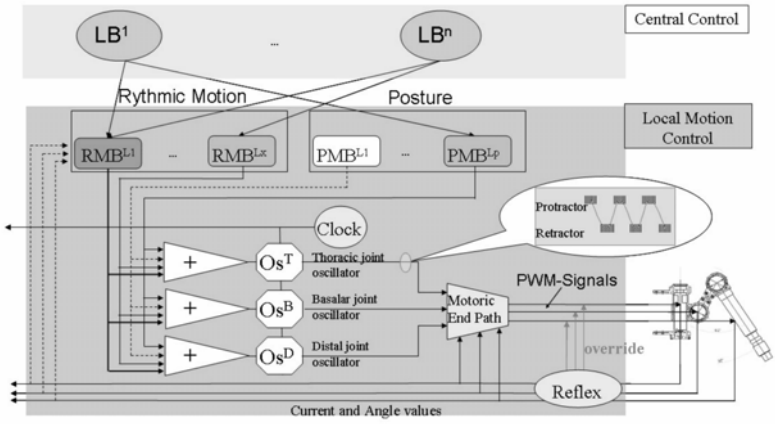


Figure 4. The overall architecture for low-level actuation.

Figure 4 outlines the general idea. On the global level (light gray area), we have implemented locomotion behaviors (LBs): typically forward, backward, and lateral locomotion. These global behaviors are connected to all local leg controllers and activate (with continuous strength) the local (single-leg) motion behaviors. At the same time, they implement the inter-leg phase relation by setting/resetting the local clocks. The local level (dark gray area) implements rhythmic motion behaviors (RMBs) and postural motion behaviors (PMBs). These behaviors simultaneously influence the amplitude and frequency (see Figures 4 and 5) parameters of three oscillating networks: OS^T, OS^B, and OS^D. The oscillators are connected to a common clock, which is used for local and global (in relation to other legs) synchronization. The oscillator output is a rhythmic, alternating flexor and extensor stimulation signal (see callout box, Figure 4), which is implemented as sine waves. This activation signal represents the desired behavioral pattern, which is translated into pulse-width-modulated (PWM) signals via the motor end path. In line with the output of the motor end path is a set of perturbation-specific reflexes, which are implemented as "watchdogs." These reflexes override the signals on the motor end path with precompiled activation signals if the sensor information from the physical joints meets a set of defined criteria. For example, if the current of a thoracic joint is above a certain threshold during the swing phase, a reflex is triggered to move the leg higher (for more details, see §4.1).

4. RESULTS

This approach was implemented using inter-leg coordination data as observed in real scorpions (3) and successfully tested on our "SCORPION" robot

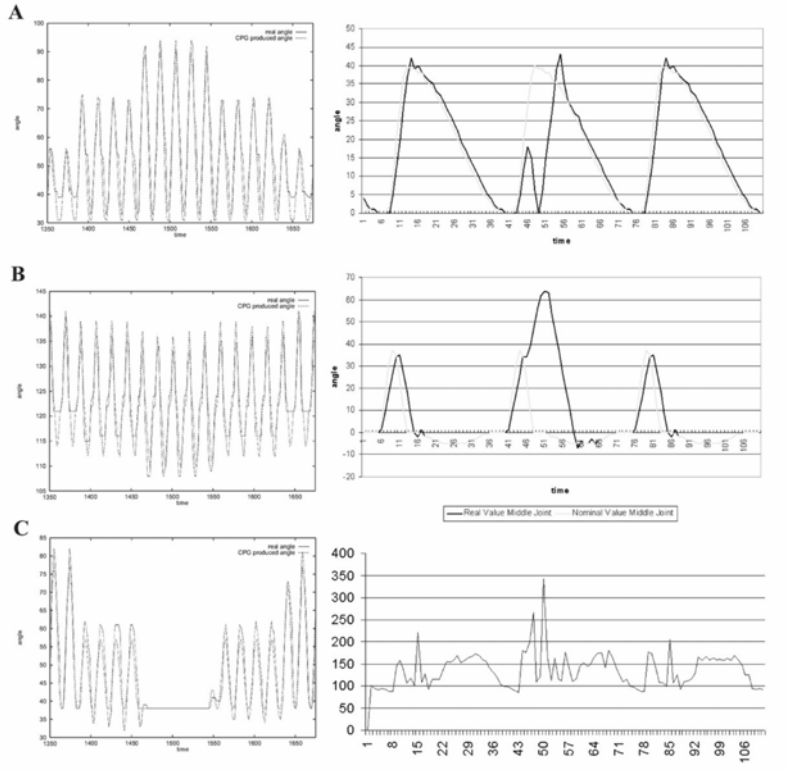


Figure 5 (left side): Traces A–C show the movements of the thoracic (**A**), basilar (**B**), and distal (**C**) joints during transition from forward walking (pure FORWARD activation), to diagonal walking (equal activation of FORWARD and LATERAL), to lateral walking (pure LATERAL activation) and back. See Figure 4 for the activation pathways. (right side): A reflex initiated in a leg during a course through a rock bed. The current in the thoracic joint (trace C) increased as a result of the obstacle blocking the way. At the same time the angular displacement error (trace A) in the thoracic joint increased, indicating an exception in the regular swing cycle. As a result of these factors the basilar joint controller (trace B) initiated the reflex that elevated the leg further, thereby overcoming the obstacle.

study (see Figure 5). The data for the performance of one leg is shown in Figures 5A–C. The solid line is the real angle of the leg, measured with motor encoders. The angle for the distal and basilar joints increases during elevation, while the angle for the thoracic joint increases during protraction. The frequency was set to 1.3 Hz (19 time units on the *x*-axis). The data were taken every 1/25 s, and the curves are directly computed from the raw data. The mean starting position is at 37° for the thoracic joint, 121° for the basilar joint, and ≈30° for the

distal joint. At first, only a local FORWARD behavior is stimulated (until t = 1375); then the LATERAL behavior is activated simultaneously. Because of the equal strength of activation, the system now tries to walk forward as well as laterally, which results in diagonal walking.

Approximately at time t = 1460, activation of the FORWARD behavior is set to 0, which leaves only the LATERAL behavior to influence the oscillator networks (see Figures 4 and 5A–C). Thus, the system walks laterally, which can be observed from the data as the amplitude of the thoracic joint at 0 while the basilar and especially distal joint perform large-amplitude oscillations. Subsequently, this process is reversed.

4.1. Local Reflexes

The approach described here for generation of rhythmic motion deals very well with plain surfaces without obstacles. However, in the case of uneven ground poor results are to be expected. Our approach to deal with uneven terrain was to implement a set of reflexes in parallel with the motor end path (see Figure 5, right side), which override, for a short and predefined period of time, the rhythmic activity of the oscillators. For example (see Figure 5, right side), if the current values of the thoracic joint increase steeply and a significant angular displacement error is detected at the same time, it is assumed that the "planed" trajectory is blocked. This triggers a reflex, which moves the leg backward and upward (via joint activity in the thoracic and basilar joints) and then forward at maximum speed. This reflex is illustrated in Figure 5 (right side; start point at t = 46, stop at t = 58). The reaction time of the reflexes can be as fast as 1/100 s, because they are directly in line with the motor control signals. The three pictures also illustrate how fast the motor controller returns to the pattern given by the oscillator, after the reflex is no longer active. It is important to notice that the action of the reflex does in fact sit right on top of the ongoing rhythmic activity. As can be seen in Figure 5 (right-hand traces) the oscillatory activity is always present in the background (light gray lines); as soon as the reflex is terminated, the locomotion returns to the oscillation.

5. DISCUSSION AND OUTLOOOK

Recently, the psychological point of view that grants the body a more significant role in cognition has also gained attention in spatial cognition theory. Proponents of this approach would claim that instead of a "mind that works on abstract problems" we have to deal with and understand "a body that needs a mind to make it function" (16).

These ideas differ quite radically from the traditional approach that describes a cognitive process as an abstract information-processing task where the real physical connections to the outside world are of only subcritical importance, sometimes discarded as mere "informational encapsulated plug-ins" (17). Thus, most theories in cognitive psychology have tried to describe the process of human thinking in terms of propositional knowledge. At the same time, artificial intelligence research has been dominated by methods of abstract symbolic processing, even if researchers often used robotic systems to implement them (18).

Ignoring sensorimotor influences on cognitive ability is in sharp contrast to research by William James (19) and others (see (20) for a review) that describe theories of cognition based on motor acts, or a theory of cognitive function emerging from seminal research on sensor-motor abilities by Jean Piaget (see (16)) and the theory of affordances by Gibson (21). In the 1980s the linguist Lakoff and the philosopher Johnson (22) put forward the idea of abstract concepts based on metaphors for bodily, physical concepts; around the same time, Brooks made a major impact on artificial intelligence research by his concepts of behavior-based robotics and interaction with the environment without internal representation instead of the sense–reason–act cycle. This approach has gained wide attention ever since, and there appears to be a growing sense of commitment to the idea that cognitive ability in a system (natural or artificial) has to be studied in the context of its relation to a "kinematically competent" physical body.

In the future of this project, we will focus our attention on experimenting with the robot to tackle two main questions. First, can "cognitive workload" be offloaded to the environment by tacking into account combined, sensorimotor representations of a robot with a kinematically complex body that have a high disposition to manipulate and interact with the world? Second, can cognitive mechanisms (perception/memory etc.) be better understood (or modeled in technical systems) if they are studied in the light of their ultimate contribution to appropriate, or goal-achieving behavior of a kinematically complex body (robot) that acts in a complex, dynamic environment.

6. REFERENCES

1. Ayers J, Kashin S, Blidberg DR, Massa D. 1994. Biologically-based underwater robots. *Unmanned Syst* **12**:30–36.
2. Blickhan R, Full RJ. 1998. Similarity in multilegged locomotion, bouncing like a monopode. *J Comp Physiol* **A173**:509–517.
3. Bowerman RF. 1975. The control of walking in the scorpion, I: leg movement during normal walking. *J Comp Physiol* **100**:83–196.
4. Brooks RA. 1986. A robust layered control system for a mobile robot. *IEEE J Robotics Automat* **2**:14–23.
5. Calvitti A, Beer RD. 2000. Analysis of a distributed model of leg coordination, I: individual coordination mechanisms. *Biol Cybern* **82**:197–192.
6. Cruse H, Kindermann T, Schumm M, Dean J, Schmitz J. 1998. Walknet—a biologically inspired network to control six-legged walking. *Neural Netw* **11**:1435–1447.

7. Delcomyn F. 1980. Neural basis of rhythmic behavior in animals. *Science* **210**:492–498.
8. Fearing RS, Chiang KH, Dickinson M, Pick DL, Sitti M, Yan J. 2000. Wing transmission for a micromechanical flying insect. In *Proceedings of the IEEE international conference on robotics and automation*, pp. 1509–1516. San Francisco, April.
9. Hebb D. 1949. *The organization of behavior: a neuropsychological theory*. Wiley, New York.
10. Kirchner F. 1995. Automatic decomposition of reinforcement learning tasks. In *Proceedings of the AAAI95 fall symposium series on active learning*, pp. 56–59. MIT, Cambridge.
11. Kirchner F. 1998. Q-learning of complex behaviors on a six-legged walking Machine. *J Robotics Auton Syst* **25**:256–263.
12. Kupferman I, Weiss KR. 1978. The command neuron concept. *Behav Brain Sci* **1**:3–39.
13. Stein PSG. 1978. Motor systems with specific reference to the control of locomotion. *Ann Rev Neurosci* **1**:61–81.
14. Spenneberg D, Kirchner F. 2000. Omnidirectional walking in an eight legged robot. *Proceedings of the international symposium of robotics and automation* (ISRA2000), Monterrey, Mexico.
15. Whitehead S. 1990. Active perception and reinforcement learning. In Proceedings of the *7th international conference on machine learning*, pp. 162–169. Ed. B Porter, RJ Mooney. Morgann Kaufmann, San Francisco.
16. Wilson M. 2002. Six views of embodied cognition. *Psychon Bull Rev* **9**:625–636.
17. Fodor JA. 1983. *The modularity of mind*. Cambridge: MIT Press.
18. Nilsson NJ. 1984. *Shakey the robot*. SRI Technical Note 323. Menlo Park, CA: AI Center, SRI International.
19. James W. 1890. *The principles of psychology*. New York: Henry Holt.
20. Prinz W. 1987. Ideomotor action. In *Perspectives on perception and action*, p. 47–76. Ed H Heuer, AF Sanders. Hillsdale, NJ: Lawrence Erlbaum Associates.
21. Gibson JJ. 1977. The theory of affordances. In *Perceiving, acting, and knowing: toward and ecological psychology*, pp. 62–82. Ed R Shaw, J Brandsford. Hillsdale, NJ: Lawrence Erlbaum Associates.
22. Lakoff G, Johnson M. 1980. *Metaphors we live by*. Chicago: U Chicago P.
23. Brooks RA. 1986. A robust layered control system for a mobile robot. *IEEE J Rob Automation* **2**(1):14–23.

8

VIRTUAL REALITY, INTRAOPERATIVE NAVIGATION, AND TELEPRESENCE SURGERY

M. Peter Heilbrun

Department of Neurosurgery, Stanford University Medical Center, Stanford, California

A part of the expertise of the operating surgeon is the practiced development of motor skills. These skills are attained by practice in multiple environments, including the animal laboratory and the operating room. The surgeon starts as a surgical assistant and gradually is granted increased responsibilities in performing the critical portions of operative procedures. Building on many years of experience using types of radiographic images of the brain and spine for intraoperative navigation, neurosurgeons, working with bioengineers and computer scientists, have developed methods of image-guided computer-assisted and computer-directed operative procedures using anatomic and pathologic structures identified in volumetric three dimension reformatted brain and spine images co-registered to the physical operative workspace using a variety of three-dimensional digitizers. With the computation power available today, such image sets can be used to create a virtual environment within which a surgeon could realistically both practice skills and attain new skills. This can now be accomplished with partial immersion. It is realistic to contemplate in the near future a total immersion environment realistically simulating all of the sensations and forces associated with an actual operative field. We have termed this development a "surgical holodeck." This chapter reviews the development and application of these methods, which are the foundation of a simulation environment close to the real operative suite.

Address correspondence to: M. Peter Heilbrun, Professor of Neurosurgery, Department of Neurosurgery, Stanford University Medical Center, 300 Pasteur Drive, Room 200, Stanford CA 94305-5327 (peter.heilbrun@stanford.edu).

1. INTRODUCTION

The goal of medical surgery has not changed since the time of Hippocrates. That is, every surgeon seeks to relieve a patient's diagnosed complaint without causing harm. In this endeavor, both daring and caution are required—first to plan, and then to execute the most precise surgical procedure that will result in removing the underlying pathology or interrupting its pathways. Once planned, the success of each operation depends upon a range of external sensory inputs as interpreted through stored experiential memories to produce finely honed hand motor skills and extended periods of focused attention. In this context, the predominant senses involve vision and touch, but sound and smell are also significant. The equanimity to perform these acts of delicate balance comes from rigorous training and continued practice, awareness of scientific advances, not only in medicine but also in other disciplines, and the testing and incorporation of appropriate new technologies for the particular surgical procedures. In recent years surgeons have worked closely with bioengineers and computer scientists to design virtual-reality environments that simulate the actual surgical workspace in order to create and test realistic training and practice platforms.

How far are we from the development of an actual "surgical holodeck?"

2. BIOMEDICAL BACKGROUND

2.1. Perspective

This review is biased to neurosurgery because the foundations for creating a "surgical holodeck" are based initially on advances in brain imaging coupled with hundred-year-old techniques of stereotactic brain surgery, i.e., localization of the physical three-dimensional position of structures making up the brain. In the past few decades, the scientific and practical advances of neurosurgical stereotaxy have resulted in an enormous wealth and breadth of structural and functional data about the brain. With each incremental advance, these image databases have promoted increasingly more sophisticated techniques of surgical navigation, leading to "fail-safe" surgical simulation environments similar to those created for training and assessing the competence of airline pilots. Using computer-generated and stored databases, today's neurosurgeon routinely navigates within a Cartesian coordinate system that targets structures in the human brain to both define and alter function.

2.2. Historical Context

In most parts of the body, routine surgical operations were traditionally performed through openings that allowed direct visualization of the abnormality to be removed or altered. Obviously, however, the cranial vault is not conducive to wide, exploratory openings. Therefore, from the outset, neurosurgeons necessarily relied on physical and neurological examination as correlated with a range of

imaging modalities to localize nervous system abnormalities. The first methods for targeted navigation to plan efficient approaches in the brain were conceived over one hundred years ago, starting with Horsley and Clark (13). From that conception, neurosurgeons began to use landmarks identified on plain x-rays of the skull and spine to guide them to brain and spinal cord regions. In 1918, Dandy introduced the pneumoencephalogram, the first advance to allow identification of non-bony structures, by measuring the shift that resulted from introducing air into the cavities of the brain (10). In 1926, Moniz invented cerebral angiography, i.e., x-ray visualization of injected radiopaque contrast material into the carotid arteries supplying blood to the brain (20). The next major imaging advance occurred in the 1960s, when neurosurgeons began to use contrast encephalograms and myelograms, which also involved the injection of radiopaque contrast material. Unlike angiography, these studies involved direct injections into brain and spine cavities in order to demonstrate a measurable displacement and distortion of specific brain and spinal cord structures.

During this entire period, neurologists and surgeons developed extensive maps of the three-dimensional relationships of focal brain structures and their functions and connections. By the 1940s the measurement and localization of the in-vivo and in-vitro electrical and chemical states of the brain and spinal cord became essential tools for this type of mapping. Through such brain explorations, numerous stereotactic tools evolved to increase the precision, accuracy, and safety of brain explorations. Starting in the late 1940s, neurosurgeons Spiegel and Wycis first used these tools to destroy focal areas in the human brain to alter function (26). Thereafter, the field of stereotactic brain surgery, using two-dimensional radiographic air contrast images that identified the skull and the brain ventricular system landmarks to map and locate three-dimensional points in the physical human brain, quickly advanced. These remarkable technical achievements rapidly led to effective ablative lesions to relieve the debilitating symptoms of many chronic brain disorders, particularly Parkinson's disease.

2.3. The Modern Era

Beginning in the 1970s and through the 1990s, the invention of computerized axial tomographic (CAT) scanning, positron emission tomography (PET), magnetic resonance imaging (MRI), and functional MRI (fMR) (see also this volume, Part IV, chapter 5, by Breiter, Gasic, and Makris), provided direct imaging and functional information about the brain and spinal cord. Consequently, an enormous structural and functional database about the nervous system, founded upon neurosurgeons' embrace and support of advanced computer programming and graphics, emerged. How this imaging database was translated and co-registered into multiple three-dimensional coordinate systems that could be co-registered to real physical space is one of the remarkable neurosurgical technical advances of the past 40 years.

Thus, soon after the introduction of CT scanning, existing stereotactic systems (e.g., Leksell, Reichert-Mundinger, and Todd-Wells) were modified so that CT images could be used for co-registration (19). However, these older, pre-CT systems were constrained by their initial design purpose, i.e., to access deep midline brain structures, and the first modifications did not take advantage of the potential for navigation throughout the cranial vault as imaged by CT.

In the late 1970s, the Brown-Roberts-Wells stereotactic system was invented based on a whole new concept for both co-registering image and physical space and accessing that space, particularly the surface and periphery of the brain (6,7). During this period, Kelly demonstrated that these stereotactic systems could be combined with computer graphic techniques to define and target not only single points, but also multiple points representing the three-dimensional volume and position of lesions within the brain. That is, stereotactic techniques could be used for volumetric craniotomies to navigate to specific lesions within the brain for biopsy and ablation (16,17). These advances led to more precise and smaller cranial skull openings to approach lesions, which decreased the morbidity of many brain operations.

At the same time, operating microscopes, which provide exquisite lighting and magnification, as well as endoscopic visual techniques using both monocular and binocular video images directed either to both eyes or formatted stereoscopically on a computer console, became routine tools to enhance the surgeon's operating precision. Additionally, the development of approaches through and below the skull and in all quadrants around the spine to lesions in the brain and the spinal cord decreased potential operative damage (11).

In the mid to late 1980s, several neurosurgeons—including Roberts, Watanabe, and Bucholz—introduced the concept of frameless stereotactic localization by mating three-dimensional digitizers designed to be used in the industrial workplace with advanced computer graphics (8,12,22,25,27). These devices were introduced into neurosurgical practice in the early 1990s, tested, and commercialized, and have become an accepted standard of care for surgical approaches to most intracranial lesions and many spinal lesions.

Neurosurgeons practicing at the end of the twentieth century could rely on highly defined CT, MRI, functional MRI, and PET images of actual brain lesions and structures to plan a surgical procedure, and they had established operative magnification and unique frameless approaches from all quadrants of the cranial vault and spine to manage ablation and resection of those lesions while protecting surrounding normal structures.

2.4. Robots

The next frontier involved mating these technologies with robots (see preceding chapter 7 by Kirchner and Spenneberg). During the mid-1980s, several neurosurgical investigators, including Kwoh, Young, Drake, Benebid, and

Figure 1. Upper images showing ceiling and floor stand-mounted robotic microscope used for image-guided navigation, with the lower image of a neurosurgeon and his assistant in the operating suite planning and resecting the tumor seen in Figure 2.

Kelly, began exploring how robots could be adapted to accomplish both computer-assisted and computer-directed image-guided operations (5,14,15,18).

2.4.1. Surgical Navigation with an Operating Microscope

Among the early developments was a method of attaching a surgical operating microscope to a robot so that the microscope could be used as both a navigation tool and a vision enhancing tool. By the early 1990s, prototype systems were developed in which the operating microscope head was connected to six-degree-of-freedom robots. Both floor-based and ceiling-mounted robotic systems were developed that were able to use microscope focal points as pointers for navigation. However, because both systems required significant support for day-to-day function and were sometimes cumbersome to maneuver, they were installed and used only in a few centers in the United States and Europe (Figures 1 and 2). Nevertheless, the lessons learned from these robotic

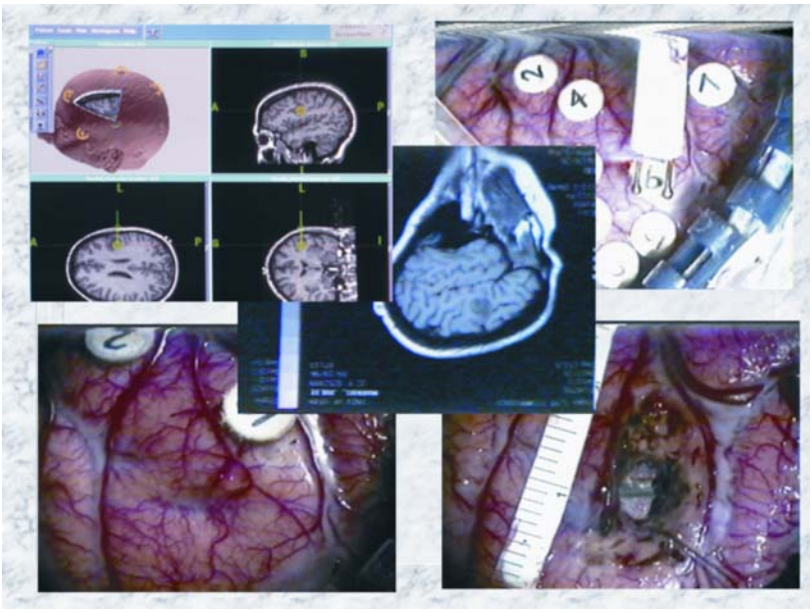


Figure 2. An interactive image-guided neurosurgical tumor resection demonstrating the steps of the cranial procedure of (i) identification of the location of the subcortical tumor (lower left) using depiction of its location on the images on the graphics display (upper left); (ii) electrocortical stimulation to identify the cortical areas of face and tongue motor function and their proximity to the tumor location (upper right); and (iii) view of the tumor bed post-guided microsurgical resection (lower right).

microscope manipulators led to better mating of optical digitizer systems with standard microscopes. Thus, neurosurgeons could use the microscope focal point as a navigation device and have continuous co-registration of their surgical view through the microscope with planning CT and MR imaging studies. These freehand operating microscope navigation systems have proved very useful (28).

2.4.2. Radiosurgery

The standard linear accelerator delivers beams with two degrees of freedom. The beam points to a focal point and can be rotated through approximately 270 degrees to a central point. The linear accelerator table also rotates. These two rotations are robotic in that they are servo-motor controlled. For focused radiation treatment, the target volume in the body must be positioned in the central point.

The Accuray (Sunnyvale, CA) CyberKnife, invented at Stanford University, mates a lightweight linear accelerator with a six-degree-of-freedom movement industrial robot (1,9). This increased range allows radiation beams to be positioned in an infinite number of positions around a target. Using inverse planning algorithms, the surgeon can now generate a non-isocentric collection of beam paths that provides a focused dose of radiation which is highly conformal to the three-dimensional shape of a target volume, yet has rapid fall-off, so that surrounding tissue is protected from radiation damage. Additionally, the CyberKnife system uses stereotactic radiographic image tracking to monitor target volume motion and has the capability of adjusting the beam trajectories in a few seconds if the patient moves during treatment. This system is used routinely for radiosurgical treatment of lesions of the cranium and spinal column. It has recently been coupled with optical digitizers to track chest-wall and abdominal-wall movement so that moving target volumes in organs such as the lung and pancreas can also be treated with radiosurgery (Figures 3 and 4).

2.4.3. Robotic Micromanipulators

While neurosurgeons focused on efficient systems for use within the unique parameters of the brain, other medical specialists were developing robotic micromanipulators. These manipulators not only enhance the surgeon's manual dexterity, but also allow indirect visualization through endoscopic ports, thereby permitting operations through smaller openings than were previously possible. This development soon led to designs for manual robotic manipulation of standard surgical tools coupled with remote vision, allowing certain surgeries to be performed remotely. Communication between the site where the surgeon is manipulating the remote tools and the actual surgical site can be accomplished either by direct wiring or long-distance broadband radio waves.

Systems are now commercially available that use telepresence hand manipulators to permit remote manipulation for inserting instruments through endoscopic channels. Such systems are manufactured by Intuitive Surgical (Mountain View, CA) and Computer Motion (Goleta, CA) and are starting to be used by cardiac surgeons (3,4). These systems have a unique method for manipulating a forceps, with simulated wrist movements that are moved deep into the operative field, thus permitting dissection and more complex surgical maneuvers such as suturing to be performed through small access ports and at depth. This capability has been applied to coronary artery bypass. Similar miniaturized systems are being developed for neurosurgical procedures.

3. THE FUTURE

3.1. Teleremote Surgery

With appropriate wired and wireless computer communication, robotic systems similar to the telepresence systems have been developed so that a surgeon

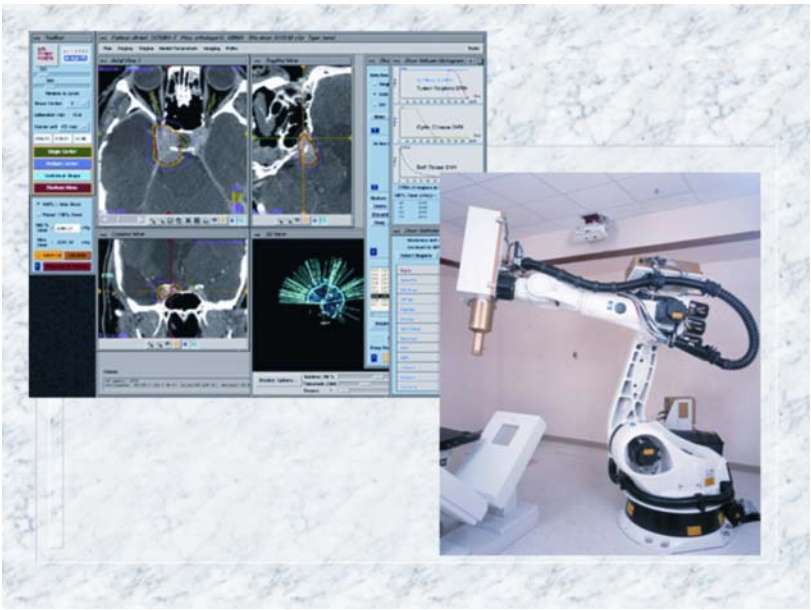


Figure 3. Depiction of the CyberKnife miniature linear accelerator mounted on the robotic manipulator with the two orthogonal amorphous silicon x-ray cameras used for image guidance (right) along with the computer display of the treatment plan for a cavernous sinus meningioma showing the radiation isodose curves in the axial, coronal, and sagittal views, and three-dimensional depiction of the selected beam trajectories that define the robot's sequential vectors and positions in physical space.

could be positioned remotely, for example, from a battlefield operating room or possibly a space station, and perform critical portions of an operation (2,23). U.S. Army laboratories have demonstrated the feasibility of such remote systems in large animal experiments in which a trained paramedic would make a surgical opening into an abdomen to expose the organ requiring surgical correction. After manual positioning of monocular or binocular stereoscopic video cameras by the paramedic, the surgeon would use similar micromanipulators to operate remotely with standard surgical tools. At the present time, if the distances are quite long and the surgical objects have some motion, the response time can be too slow for remote adjustment of the position of the surgical tools being manipulated. However, motion tracking and position prediction may allow compensation for signal delay.

3.2. Advanced Volume Rendering

Computer graphic techniques depicting a simulated surgical field are now available and are used routinely by neurosurgeons for surgical planning. At this

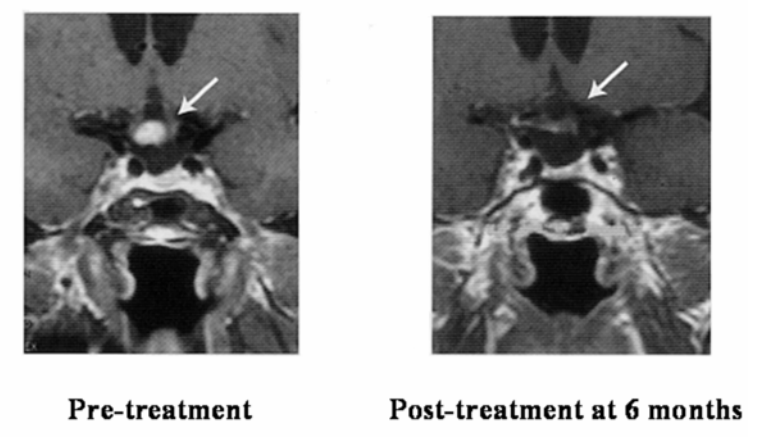


Figure 4. Pretreatment and 6-month-posttreatment coronal MR images of a patient treated with staged CyberKnife radiosurgery.

time, several developers are working to adapt this technology to perspective volume rendering of three-dimensional image sets. The goal is to use these modes of rendering virtual reality, with both partial and total immersion, for actual surgical navigation as well as simulation (for modeling and simulation of brain tumors, see, e.g., this volume, Part III, chapter 6.3, by Mansury and Deisboeck). A current problem in using these systems during actual surgery is the inability to comfortably track the movement of a surgeon's head and eyes and to adjust the simulated image position. Presently available heads-up helmets and eyeglass designs are uncomfortable after a prolonged period of usage. However, using large high-resolution projection or flat video screens may eliminate the necessity for cumbersome head tracking, as these large screens can encompass both the central and peripheral visual fields and provide the surgeon with a comfortable mode of virtual reality through partial immersion (24).

3.3. Haptics

Another area that is a work in progress involves methods for simulating the surgeon's sense of touch. Surgeons depend on tactile feedback from the body tissues they are manipulating during a surgical procedure. All tissues have varying degrees of rigidity, elasticity, tensile strength, and deformability (see also this volume, Part IV, chapter 4, by Kaazempur-Mofrad, Weinberg, Borenstein, and Vacanti). A surgeon's innate and conscious sense of these tissue conditions through touch with a gloved finger or indirectly through surgical instruments is a key factor in the precision by which different tissues are separated to create cor-

ridors to a lesion in order to remove it from normal tissue. The multiple methods for tissue dissection utilize a variety of sharp and blunt as well as cold and heated tools.

Force-feedback algorithms have been incorporated into a few surgical training and computerized simulation systems related to blood vessel anastomoses (21). However, systems that simulate tissue conditions which require tactile experience and feedback so that a trainee could determine whether it were better to separate tissue planes sharply with a scissors or knife or bluntly with non-sharp or rounded instruments have not been developed as of yet. Additionally, it is not presently realistic to simulate the adverse effects of improper dissection techniques, such as blood vessel disruption, resultant bleeding, bleeding control, and tissue disruption,

4. DISCUSSION AND CONCLUSIONS

The field of image-guided surgery today depends on the three-dimensional co-registration of a variety of high-resolution structural and functional image sets with the actual physical surgical workspace. Advances in several areas of medical computer graphics and image processing mated to a variety of three-dimensional digitizers including robotic manipulators have provided surgeons with sophisticated tools for precision navigation during operative procedures. Many of these state-of-the-art systems are used routinely in today's operating suites.

Most image-guidance systems are based on the co-registration or fusion of preoperative visual images. A surgeon's view can be enhanced by the improved lighting and magnification provided by an operating microscope. In operations executed deep within the body, indirect visual input and substitution can be accomplished with video cameras attached to both endoscopes and microscopes. Such indirect video views of the operative field, if they have high enough resolution, can be enlarged and projected to a flat video screen and provide a partial immersive virtual-reality environment. These advances alone provide significant benefits to surgeons in the actual precise and safe performance of many standard operative procedures. With lightweight head-tracking devices, video projection, and perspective volume rendering and navigation, this partial immersive environment can be made even more realistic.

These systems are important for training neophyte surgeons in both operative planning and execution. Similarly, they can assist trained surgeons in practicing and maintaining their skills. However, until more sophisticated haptic systems are available to substitute for the surgeon's sense of touch, these devices cannot substitute for actual tissue manipulation either in the surgical training laboratory with animals or in the operating suite as surgical assistants during operations on human beings. Nevertheless, with modern advancements in bio-

engineering and computer science, soon, fail-safe, virtually realistic environments will be able to provide the full range of sensory and experiential input for the surgeon to learn, practice, and develop the fine motor skills to perform surgical procedures with expected competency and excellence. This environment will approach that of a "surgical holodeck."

5. REFERENCES

1. Adler Jr JR, Chang SD, Murphy MJ, Doty J, Geis P, Hancock SL. 1997. The CyberKnife: a frameless robotic system for radiosurgery. *Stereotact Funct Neurosurg* **69**(1–4 pt 2):124–128.
2. Angood PB, Satava R, Doarn C, Merrell R; E3 Group. 2000. Telemedicine at the top of the world: the 1998 and 1999 Everest extreme expeditions. *Telemed JE Health* **6**(3):315–325.
3. Argenziano M. 2003. Totally endoscopic, robotic cardiac surgery. *Heart Surg Forum* **6**(2):104.
4. Argenziano M, Oz MC, DeRose Jr JJ, Ashton Jr RC, Beck J, Wang F, Chitwood WR, Nifong LW, Dimitui J, Rose EA, Smith Jr CR. 2002. Totally endoscopic atrial septal defect repair with robotic assistance. *Heart Surg Forum* **5**(3):294–300.
5. Benebid AL, Cinquin P, Lavalle S, Le Bas JF, Demongeot J, de Rougemont J. 1987. Computer-driven robot for stereotactic surgery connected to CT scan and magnetic resonance imaging: technological design and preliminary results. *Appl Neurophysiol* **50**:153–154.
6. Brown RA. 1979. A computed tomography-computer graphics approach to stereotactic localization. *J Neurosurg* **50**:715–720.
7. Brown RA, Roberts TS, Osborne AG. 1980. Stereotaxic frame and computer software for CT-directed neurosurgical localization. *Invest Radiol* **15**(4):308–312.
8. Bucholz RD, Greco DJ. 1996. Image-guided surgical techniques for infections and trauma of the central nervous system. *Neurosurg Clin N Am* **7**(2):187–200.
9. Chang SD, Main W, Martin DP, Gibbs IC, Heilbrun MP. 2003. An analysis of the accuracy of the CyberKnife: a robotic frameless stereotactic radiosurgical system. *Neurosurg* **52**(1):146–147.
10. Dandy WE. 1918. Ventriculography following injection of air into the cerebral ventricles. *Ann Surg* **68**:1–5.
11. Donaghy RM, Yasargil G. 1968. Microangeional surgery and its techniques. *Prog Brain Res* **30**:263–267.
12. Friets EM, Strohbehn JW, Hatch JF, Roberts DW. 1989. A frameless stereotaxic operating microscope for neurosurgery. *IEEE Trans Biomed Eng* **36**(6):608–617.
13. Horsley V, Clarke LH. 1908. The structure and functions of the cerebellum examined by a new method. *Brain* **31**:45–124.
14. Kall BA, Kelly PJ, Goerss SJ. 1985. Interactive stereotactic surgical system for the removal of intracranial tumors utilizing the CO_2 laser and CT-derived database. *IEEE Trans Biomed Eng* **32**(2):112–116.
15. Kelly PJ. 1987. Future possibilities in stereotactic surgery: where are we going? *Appl Neurophysiol* **50**(1–6):1–8.
16. Kelly PJ. 1991. Computer-assisted volumetric stereotactic resection of superficial and deep seated intra-axial brain mass lesions. *Acta Neurochir Suppl (Wien)* **52**:26–29.
17. Kelly PJ. 1992. Stereotactic resection and its limitations in glial neoplasms. *Stereotact Funct Neurosurg* **59**(1–4):84–91.
18. Kwoh YS, Young RF, eds. 1992. *Robotic-aided surgery: computers in stereotactic neurosurgery*. Blackwell Scientific Publications, Oxford.
19. Leksell L, Jernberg B. 1980. Stereotaxis and tomography: a technical note. *Acta Neurochir (Wien)* **52**(1–2):1–7.

20. Moniz E. 1927. L'encephalographie arterielle: sons importance dans la localisation des tumeurs cerebrates. *Rev Neurol* **34**:72–90.
21. Ottensmeyer MP, Ben-Ur E, Salisbury JK. 2000. Input and output for surgical simulation: devices to measure tissue properties in vivo and a haptic interface for laparoscopy simulators. *Stud Health Technol Inform* **70**:236–242.
22. Roberts DW, Strohbehn JW, Friets EM, Kettenberger J, Hartov A. 1989. The stereotactic operating microscope: accuracy refinement and clinical experience. *Acta Neurochir Suppl (Wien)* **46**:112–114.
23. Satava RM. 1999. Emerging technologies for surgery in the 21st century. *Arch Surg* **134**(11):1197–1202.
24. Shahidi R, Clarke L, Bucholz RD, Fuchs H, Kikinis R, Robb RA, Vannier MW. 2001. White paper: challenges and opportunities in computer-assisted interventions January 2001. *Comput-Aided Surg* **6**(3):176–181.
25. Smith KR, Frank KJ, Bucholz RD. 1994. The NeuroStation: a highly accurate, minimally invasive solution to frameless stereotactic neurosurgery. *Comput Med Imaging Graph* **18**(4):247–256.
26. Spiegel EA, Wycis HT, Marks M, Lee AJ. 1947. Stereotactic apparatus for operations on the human brain. *Science* **106**:349–350.
27. Watanabe E, Watanabe T, Manaka S, Mayanagi Y, Takakura K. 1987. Three-dimensional digitizer (neuronavigator): new equipment for computed tomography-guided stereotaxic surgery. *Surg Neurol* **27**(6):543–547.
28. Westermann B, Trippel M, Reinhardt H. 1995. Optically navigable operating microscope for image-guided surgery. *Minim Invas Neurosurg* **38**:112–116.

INDEX