NATO Science for Peace and Security Series - A:
Chemistry and Biology

Macromolecular Crystallography

Deciphering the Structure, Function and Dynamics of Biological Molecules

Edited by
Maria Arménia Carrondo
Paola Spadon

Springer

NATO OTAN

This publication is supported by:

The NATO Science for Peace and Security Programme

Macromolecular Crystallography

NATO Science for Peace and Security Series

This Series presents the results of scientific meetings supported under the NATO Programme: Science for Peace and Security (SPS).

The NATO SPS Programme supports meetings in the following Key Priority areas: (1) Defence Against Terrorism; (2) Countering other Threats to Security and (3) NATO, Partner and Mediterranean Dialogue Country Priorities. The types of meeting supported are generally “Advanced Study Institutes” and “Advanced Research Workshops”. The NATO SPS Series collects together the results of these meetings. The meetings are co-organized by scientists from NATO countries and scientists from NATO’s “Partner” or “Mediterranean Dialogue” countries. The observations and recommendations made at the meetings, as well as the contents of the volumes in the Series, reflect those of participants and contributors only; they should not necessarily be regarded as reflecting NATO views or policy.

Advanced Study Institutes (ASI) are high-level tutorial courses to convey the latest developments in a subject to an advanced-level audience

Advanced Research Workshops (ARW) are expert meetings where an intense but informal exchange of views at the frontiers of a subject aims at identifying directions for future action

Following a transformation of the programme in 2006 the Series has been re-named and re-organised. Recent volumes on topics not related to security, which result from meetings supported under the programme earlier, may be found in the NATO Science Series.

The Series is published by IOS Press, Amsterdam, and Springer, Dordrecht, in conjunction with the NATO Emerging Security Challenges Division.

Sub-Series

A.	Chemistry and Biology	Springer
B.	Physics and Biophysics	Springer
C.	Environmental Security	Springer
D.	Information and Communication Security	IOS Press
E.	Human and Societal Dynamics	IOS Press

http://www.nato.int/science
http://www.springer.com
http://www.iospress.nl

Series A: Chemistry and Biology

Macromolecular Crystallography

Deciphering the Structure, Function and Dynamics of Biological Molecules

edited by

Maria Arménia Carrondo
New University of Lisbon, Portugal

and

Paola Spadon
University of Padova, Italy

Published in Cooperation with NATO Emerging Security Challenges Division

Proceedings of the NATO Advanced Study Institute on
Structure and Function of Biomacromolecules as a Tool against CBRN Agents
Erice, Italy
3–13 June 2010

Library of Congress Control Number: 2011940793

ISBN 978-94-007-2532-4 (PB)
ISBN 978-94-007-2529-4 (HB)
ISBN 978-94-007-2530-0 (e-book)
DOI 10.1007/978-94-007-2530-0

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

All Rights Reserved
© Springer Science+Business Media B.V. 2012
No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Preface

The Advanced Study Institute "Macromolecular Crystallography: Deciphering the Structure, Function and Dynamics of Biological Molecules" was held in Erice, Italy at the Ettore Majorana Foundation and Centre for Scientific Culture from 3 to 13 of June 2010. The course counted with 137 total participants from various countries. The programme included lectures given by invited speakers and senior participants and workshops. Posters were also displayed, preceded by short oral presentations by the presenting author. The main objective of the Institute was to train the younger generation on advanced methods and techniques to discover relevant structural and dynamic aspects of biological macromolecules.

The Institute program focused on the role of macromolecular crystallography and other complementary techniques in studying spatial and dynamic nature of macromolecular assemblies and their roles in living organisms. The scientific programme was organized in several blocks according to a main topic: assemblies, membranes, imaging, cryo-electron microscopy, mass spectrometry, dynamic assemblies, viruses-large particles, small angle scattering, signalling and ribosomes, with the participation of the **three Nobel Laureates of 2009**. Twelve workshops on data bases and software developments were an important part of the school.

The quality of all lectures was of a very high standard, since the speakers were among the top leaders in the world on the corresponding subject. This volume comprises a good selection of papers presented in this school.

The course was financed by NATO as an ASI. Additional support was provided by the European Crystallography Association, the International Union of Biochemistry and Molecular Biology, the International Union of Crystallography, the University of Bologna, AstraZeneca, Bruker Axs, Douglas Instruments Ltd, Oxford Diffraction and Rigaku Americas & Rigaku Europe.

The NATO ASI Directors worked alongside and would like to specially thank Prof. Sir Tom Blundell, Director of the International School of Crystallography and the local organizers Dr. John Irwin and all the orange scarves. In particular the precious and unforgettable coordination of Prof. Lodovico Riva di Sanseverino on his last performance among the local organizers was of extreme value to the school.

Maria Arménia Carrondo
Paola Spadon

To Lodovico

This book is dedicated to Lodovico Riva di Sanseverino, the founder and the driving force behind the International School of Crystallography since its inception in 1974, who passed away unexpectedly after the conclusion of the 2010 Course. The Erice meeting is well known for its excellent scientific contents, but it was Lodovico's organizational skills, great humor and contagious enthusiasm that always made the school a truly unforgettable experience.

Scientific curiosity was what brought many people, old and young, professors and students, mentors and mentees, to the Erice School, but it was the unique combination of great science, exceptional atmosphere and an easy sense of community that brought them back year after year. Friendships and collaborations shaped in Erice and fostered by Lodovico's relentless passion for the School and for Crystallography have lasted many years after the conclusion of the courses.

In 2005 the International Union of Crystallography awarded Lodovico a special prize for his "Exceptional Service to Crystallography" officially recognizing his great work, and the success of the International School. This was a great moment of joy for Lodovico, but he would have been even happier for the words sent by Michel Rossmann, one of the first scientific directors In Erice, soon after the sad news of Lodovico's death: "Where else do the nights ring with songs led by Lodovico? Where else do you pay for your meals by signing a paper? Where else are the participants from every corner of the Earth? We have lost a very exceptional friend. Erice will surely remain an important stop on the itinerary of structural biologists, but, without Lodovico, it will be a new and different era".

Thanks Lodovico!

Contents

Contributors

Nenad Ban Institute of Molecular Biology and Biophysics, ETH Zurich, Schafmattstr. 20, 8093 Zurich, Switzerland, ban@mol.biol.ethz.ch

Tiago M. Bandeiras Instituto de Biologia Experimental e Tecnológica, Apartado 12, 2781-901 Oeiras, Portugal

David Barford Division of Structural Biology, Chester Beatty Laboratories, Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK, david.barford@icr.ac.uk

Wolfgang Baumeister Max-Planck-Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany, baumeist@biochem.mpg.de

Gregor Blaha Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA

Tom L. Blundell Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA, UK, tom@cryst.bioc.cam.ac.uk

Daniel Boehringer Institute of Molecular Biology and Biophysics, ETH Zurich, Schafmattstr. 20, 8093 Zurich, Switzerland

Victor M. Bolanos-Garcia Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA, UK

David Bulkley Department of Chemistry, Yale University, New Haven, CT, USA

Howard Hughes Medical Institute, New Haven, CT, USA

Maria Arménia Carrondo Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Apartado 127, 2781-901 Oeiras, Portugal, carrondo@itqb.unl.pt

Arnau Casañas Instituto de Biología Molecular de Barcelona (CSIC-Parc Cientific de Barcelona), Baldiri i Reixac 10, Barcelona 08028, Spain

Dimitri Y. Chirgadze Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA, UK

Esteban Domingo Centro de Biología Molecular Severo Ochoa (CSIC-UAM), Cantoblanco, Madrid 28049, Spain

Peter Donner Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Apartado 127, 2781-901 Oeiras, Portugal

Cristina Ferrer-Orta Instituto de Biología Molecular de Barcelona CSIC, Parc Cientific de Barcelona, Baldiri i Reixac 10, Barcelona 08028, Spain

Ignasi Fita Instituto de Biología Molecular de Barcelona (CSIC-Parc Cientific de Barcelona), Baldiri i Reixac 10, Barcelona 08028, Spain

Sabine Gorynia Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Apartado 127, 2781-901 Oeiras, Portugal

Lead Discovery Berlin – Protein Supply, Bayer Schering Pharma AG, 13353 Berlin, Germany

Department of Biological Chemistry, David Geffen School of Medicine, UCLA, 615 Charles E. Young Drive South, Box 951737, Los Angeles, CA 90095-1737, USA

Tobias Hainzl Department of Chemistry, Umeå University, SE-90187 Umeå, Sweden

Eran Hodis Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel

Kenneth C. Holmes Max Planck Institute for Medical Research, Heidelberg, Germany, holmes@mpimf-heidelberg.mpg.de

Amnon Horovitz Department of Structural Biology, Weizmann Institute of Science, 76100 Rehovot, Israel, amnon.horovitz@weizmann.ac.il

Shenghua Huang Department of Chemistry, Umeå University, SE-90187 Umeå, Sweden

C. Axel Innis Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA

Louise N. Johnson Laboratory of Molecular Biophysics, Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK

Diamond Light Source, Harwell Science and Innovation Campus, Didcot, Oxon OX11 0DE, UK, louise.johnson@Diamond.ac.uk

Brian Kelch Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA

Werner Kühlbrandt Max-Planck-Institute of Biophysics, Max-von-Laue-Str.3, 60438 Frankfurt am Main, Germany, Werner.Kuehlbrandt@mpibp-frankfurt.mpg.de

John Kuriyan Department of Molecular and Cell Biology and Chemistry, California Institute for Quantitative Biosciences, Howard Hughes Medical Institute, University of California, Berkeley

Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, kuriyan@berkeley.edu

Debora Makino Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA

Pedro M. Matias Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Apartado 127, 2781-901 Oeiras, Portugal

Colin E. McVey Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Apartado 127, 2781-901 Oeiras, Portugal

Takashi Ochi Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA, UK

Mike O'Donnell Laboratory of DNA Replication, Howard Hughes Medical Institute, The Rockefeller University, New York, USA

Filipa G. Pinho Instituto de Biologia Experimental e Tecnológica, Apartado 12, 2781-901 Oeiras, Portugal

Jaime Prilusky Bioinformatics Unit, Department of Biological Services, The Israel Structural Proteomics Center, Weizmann Institute of Science, Rehovot 76100, Israel

Jordi Querol Instituto de Biología Molecular de Barcelona (CSIC-Parc Cientific de Barcelona), Baldiri i Reixac 10, Barcelona 08028, Spain

Helen R. Saibil Crystallography, and Institute of Structural and Molecular Biology, Birkbeck College, University of London, Malet St, London WC1E 7HX, UK, h.saibil@mail.cryst.bbk.ac.uk

Elisabeth Sauer-Eriksson Department of Chemistry, Umeå University, SE-90187 Umeå, Sweden, elisabeth.sauer-eriksson@chem.umu.se

Lynn Sibanda Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA, UK

Kyle Simonetta Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA

Robin Evans Stanley Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA

NIDDK, National Institutes of Health, Bethesda, MD, USA

Thomas A. Steitz Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA

Department of Chemistry, Yale University, New Haven, CT, USA

Howard Hughes Medical Institute, New Haven, CT, USA, thomas.steitz@yale.edu

Joel L. Sussman Department of Structural Biology, The Israel Structural Proteomics Center, Weizmann Institute of Science, Rehovot 76100, Israel, joel.sussman@weizmann.ac.il

Dmitri I. Svergun European Molecular Biology Laboratory, Hamburg Outstation, Notkestraße 85, D-22603 Hamburg, Germany, Svergun@EMBL-Hamburg.DE

Núria Verdaguer Instituto de Biología Molecular de Barcelona CSIC Parc Cientific de Barcelona, Baldiri i Reixac 10, Barcelona 08028, Spain, nvmcri@ibmb.csic.es

Qian Wu Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA, UK

Ada Yonath Department of Structural Biology, Weizmann Institute of Science, 76100 Rehovot, Israel, ada.yonath@weizmann.ac.il

Chapter 1
Spatial and Temporal Organisation of Multiprotein Systems of Cell Regulation and Signalling: What Can We Learn from NHEJ System of Double-Strand Break Repair?

Qian Wu, Lynn Sibanda, Takashi Ochi, Victor M. Bolanos-Garcia, Tom L. Blundell, and Dimitri Y. Chirgadze

Abstract Multiprotein assemblies play major roles in most pathways involved in cell regulation and signaling. Weak binary interactions are transformed co-operatively into very specific systems, which achieve sensitivity, specificity and temporal control. Due to the complexity and transience of these regulatory and signaling systems, a combination of in vivo, cell, biochemical, biophysical, and structural approaches is needed to investigate their structures and dynamics. Here we describe the architecture and spatial organisation of the complexes mediating Non-Homologous End Joining (NHEJ), one of the two major pathways involved in DNA double-strand break repair. Our example illustrates the experimental challenges and conceptual questions that are raised by studying such complex systems. We discuss the potential of using knowledge of the spatial and temporal organization of multiprotein systems not only to give insights into the mechanisms of pathway regulation but also to help in the design of chemical tools and ultimately new therapeutic agents.

Keywords Multiprotein assemblies • DNA repair • Cell regulation • Cell signalling • Yeast two-hybrid • Crosslinking • Circular dichroism • AUC • SAXS • SPR • ITC • Nanospray MS • DLS • EM • Crystallisation • X-ray diffraction • Structural determination • NHEJ • Ku70/80 • DNA-PKcs • DNA ligase IV • XRCC4 • XLF

Abbreviations

EMSA	Electrophoretic mobility shift assays
CD	Circular dichroism
AUC	Analytical ultracentrifugation

Q. Wu • L. Sibanda • T. Ochi • V.M. Bolanos-Garcia • T.L. Blundell (✉) • D.Y. Chirgadze
Department of Biochemistry, University of Cambridge, Tennis Court Road,
Cambridge CB2 1GA, UK
e-mail: tom@cryst.bioc.cam.ac.uk

M.A. Carrondo and P. Spadon (eds.), *Macromolecular Crystallography*,
NATO Science for Peace and Security Series A: Chemistry and Biology,
DOI 10.1007/978-94-007-2530-0_1, © Springer Science+Business Media B.V. 2012

SV	Sedimentation Velocity
SE	Sedimentation Equilibrium
SAXS	Small angle X-ray scattering
SPR	Surface Plasmon Resonance
ITC	Isothermal titration calorimetry
DLS	Dynamic light scattering
EM	Electron microscopy
NHEJ	Non-homologous end joining
DSB	Double-strand break
DNA-PK	DNA-dependent protein kinase
DNA-PKcs	DNA-PK catalytic subunit
XRCC4	X-ray cross-complementation group 4
XLF	XRCC4-Like Factor
LigIV	DNA ligase IV

1.1 Why Are Multiprotein Systems Important in Cell Regulation?

Cell growth and multiplication are regulated by growth factors and other messengers, most of which are recognized by receptors at the cell surface. Signals are transduced inside the cell by second messengers, post-translational modification, gene activation and many other mechanisms. DNA damage activates similar signaling pathways within the cell. These must all have high signal-to-noise ratios, just like electrical circuits; indeed both living and man-made systems have switches, transducers, adaptors and so on.

But how do molecular systems in cells achieve the required sensitivity and specificity? The answer cannot be in terms of very tight, enduring molecular complexes, as the signals could not be turned off. On the other hand, weak binary complexes would lack specificity and give rise to a noisy system. The answer is to be found in multicomponent protein complexes, where the weak binary interactions are transformed through co-operativity into very specific control systems.

A good example is provided by the interactions of fibroblast growth factor receptor 2 (FGFR2) in complex with its ligand (FGF1) [56]. In solution this forms well defined 1:1 FGFR2:FGF1 complexes, which in the presence of heparin (a mimic of heparan sulphate) form stable 2:2:1 complexes (Fig. 1.1). The 2:2 complex (without heparin) is very unstable but can be seen in several crystal forms [3]. In the membrane the picture is likely to be even more complex with clustering of receptors mediated by heparan sulphate and protein-protein interactions (Fig. 1.2). The interactions are stabilised by the co-localisation of the components – FGFR and heparan sulphate – in the membrane, and the binding of other proteins to the intracellular regions, which include a juxta-membrane region and the receptor tyrosyl kinase. Indeed many signaling systems act through a cluster of co-localized components,

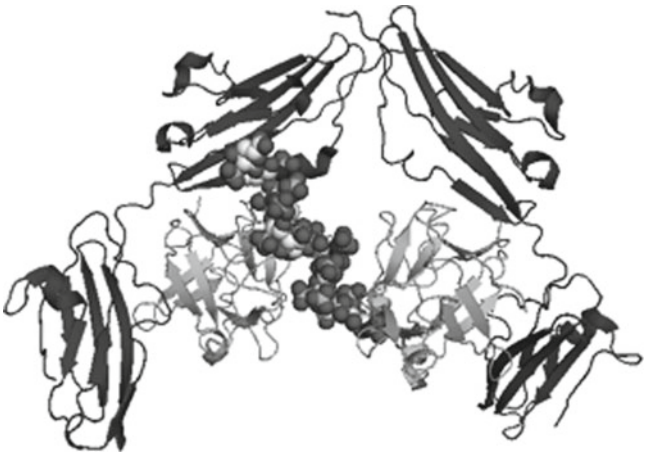


Fig. 1.1 Crystal structure of the ternary complex of fibroblast growth factor 1(FGF1) in complex with its receptor (FGFR2) and heparin (PDB: 1E0O). Heparin is an analogue of heparan sulphate, a secondary receptor involved in FGF signalling. The structure reveals a stable 2:2:1 complex, which is supported by nanospray spectrometry. Adapted from [56]

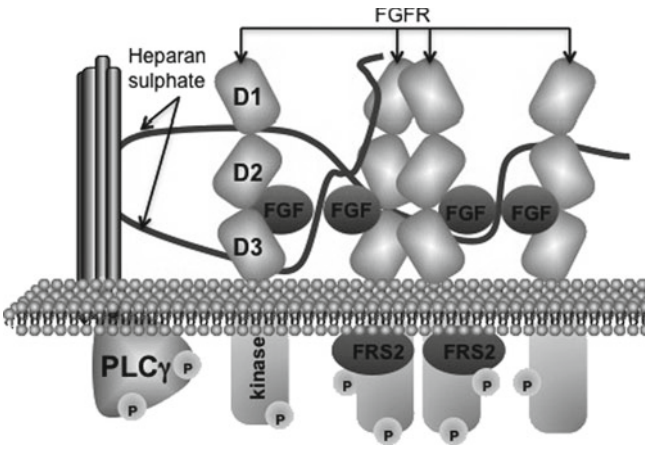


Fig. 1.2 Predicted clustering of FGFR receptor with growth factor FGF and secondary receptor heparan sulphate (HS) on membrane. Under the physiological conditions heparan sulphate plays an important role in the clustering of receptors. The interactions are further stabilised by the binding of other proteins to the intracellular regions, which includes a juxta-membrane region and the receptor tyrosine kinase

and these will signal back to the extracellular region, altering affinities of ligands and co-receptors.

In many cases proteins are disordered in the unbound state, but assemble into an ordered complex, sometimes through a zipper or Velcro mechanism. Probably the first examples to be studied of this kind were peptide-receptor interactions. An example is

Fig. 1.3 Crystal structure of BRC4 repeat in complex with Rad51 (PDB: 1N0W)

glucagon, the structure of which was defined in our laboratory in 1975 [69]. At that time it was shown by circular dichroism and NMR to have no secondary structure in aqueous solution but we hypothesized that it would adopt a helical structure in the receptor complex. Recent advances on this class of GPCRs indicate that such polypeptide hormones do fold into a helix at the receptor. Similar flexible peptides have been shown to be very common in signaling systems (see for example [22]). Recent examples of the same phenomenon from our laboratory are the DNA ligase IV peptide linker in complex with XRCC4 (see below) and the BRC4 repeat in complex with Rad51 (Fig. 1.3).

Our original hypothesis was that unbound flexible peptide hormones would be cleaved by proteases, thus removing them quickly from the circulation, but specificity would be enhanced by the need to nucleate folding during binding and therefore increasing intermolecular surface achieved in this process [69]. Thus, not only specificity but also transience might be achieved by such mechanism. A variation on this theme occurs with polysaccharides such as heparan sulphate chains, which probably bind in a similar way, with nucleation of the correct helical structure followed by recognition of sequence-specific sulphation patterns as the complex forms.

We have investigated the nature of the interactions in multicomponent signaling systems, which have structures defined at high resolution. Most involve interfaces of around 2,000 $Å^2$ of buried protein surface, usually comprised of mixed hydrophobic and polar patches, so that both the individual components and the supramolecular assemblies are stable.

Our current model [5, 7] then involves two kinds of complexes (see Fig. 1.4). The first of these has components that have preformed globular structures that often exhibit adaptive changes on assembly. The second involves one component that is

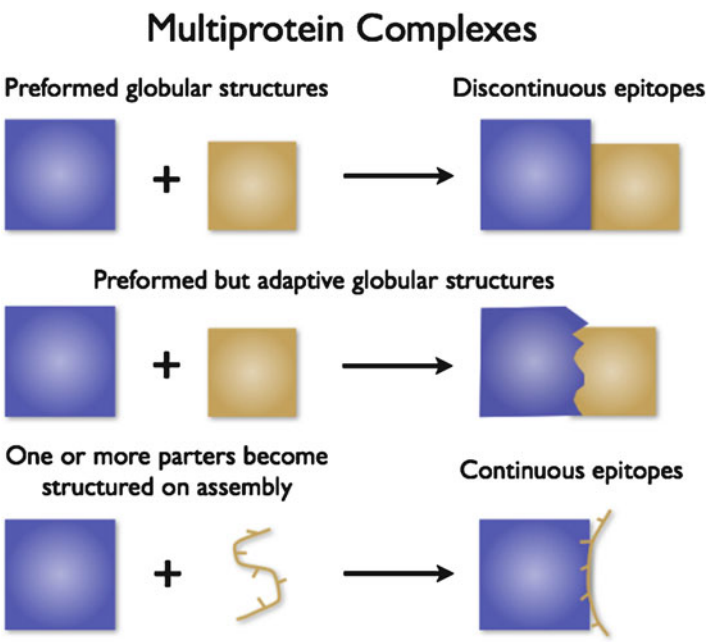


Fig. 1.4 Models of protein-protein interactions: Proteins of preformed globular structures can assemble with little change in conformation (*top*) but often exhibit adaptive changes following complex formation (*middle*). However, in many cases one component is flexible and disordered but becomes ordered in the complex (*bottom*)

flexible and disordered but becomes ordered in the final complex. These give selectivity and temporal control.

Multiprotein complexes have advantages simply by having multiple components. The chances of a binary complex arising opportunistically are relatively high. The greater the number of the components that have to be co-located the less chance that they occur in error, and so the higher the signal to noise. Furthermore, multiprotein systems, with components that may be adaptors, templates or scaffolds, can bring enzymes and receptors with their ligands together, thus decreasing entropy of the reaction and increasing specificity.

Because many diseases require therapeutic intervention at the level of multiprotein systems involved in cell regulation, the detailed knowledge of the architectures and organization of the complexes should provide an important basis for targeting these processes in drug discovery. The pharmaceutical and biotechnology industries are very aware of this problem. Although the large, flat and often flexible surfaces of the individual components are very difficult to modulate, some progress is being made. Most successful efforts seem to target sites where disordered peptides become ordered on interaction as helices or strands (see below). Nevertheless, most pharmaceutical companies continue to focus on individual proteins, and a large number are working on the ATP binding sites of protein kinases or the specificity pockets of regulatory proteases.

Thus for both understanding the basis of cell and whole organism biology, and for gaining selectivity of potential therapeutic agents, much more must be defined of the temporal and spatial organization of these regulatory systems. Unfortunately, individual structures usually give little indication of their nature and prediction of the poses of the complexes from docking individual components has proved to be challenging, probably more challenging than the so-called folding problem. The only solution is to define structures of complexes experimentally.

1.2 Methods to Study Multiprotein Assemblies

The complexity and transience of regulatory multiprotein systems demands the exploitation of sophisticated molecular, cell biology and microscopy techniques to study the structure and dynamics of protein-protein interactions in living cells. The combined use of *in vivo* and *in vitro* methods to confirm bona fide interactions may involve the following techniques:

1.2.1 In Vivo Approaches

1.2.1.1 Live-Cell Imaging

The monitoring of protein sub-cell localisation, stability and interactions using optical microscopy and related methodologies has been boosted by the use of fluorophores such as green fluorescent protein (GFP) fused to the gene encoding for the protein of interest. Nowadays, GFP and the ever-increasing number of its colour-shifted engineered derivatives are routinely used for monitoring the sub-cell localisation, activity and/or stability of protein molecules in living cells.

1.2.1.2 Förster Resonance Energy Transfer (FRET) Microscopy

FRET measures energy transfer between fluorescent probes in proteins, so providing valuable information about distances between them and thus intracellular molecular interactions and other spatial relationships.

1.2.1.3 Yeast Two-Hybrid

Yeast serves as an excellent organism to exploit the modular nature of eukaryotic transcription factors such as GAL4, which is composed of two physically separable domains: a sequence-specific DNA binding domain (BD) and a transcription activation domain (AD). The two domains need not be present in the same polypeptide to activate transcription. Thus, the interaction between two proteins can be determined

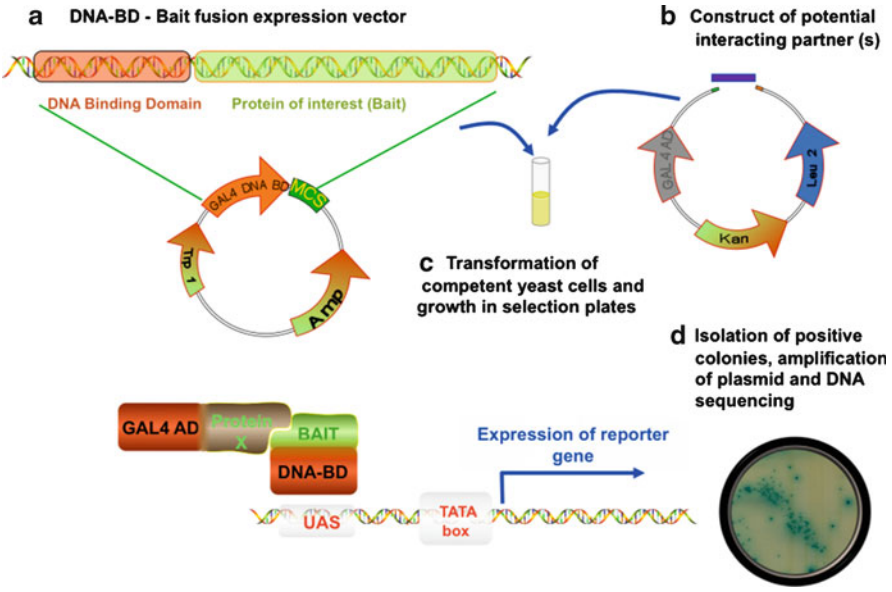


Fig. 1.5 Yeast two-hybrid. The method exploits the modular nature of eukaryotic transcription factors such as GAL4, which is composed of two physically separable domains: a sequence-specific DNA binding domain (BD) and a transcription activation domain (AD). One construct that couples the DNA sequence encoding the protein of interest is fused to the DNA (BD) to create the "bait" fusion. The interacting protein partners (or library) are expressed as fusions to the AD, creating "prey" fusion proteins. The association between BD and AD fusion proteins activates transcription factor that initiates the expression of one or more reporter genes (in the example, alpha-galactosidase)

by producing one construct that couples the DNA sequence encoding the protein of interest fused to the BD to create the "bait" fusion. The interacting protein partners (or library) are expressed as fusions to the AD, creating "prey" fusion proteins. Neither of these domains alone is able to activate the transcription machinery, but association between BD and AD fusion proteins reconstitutes an active transcription factor that initiates the expression of one or more reporter genes (see Fig. 1.5).

1.2.1.4 In Vivo Crosslinking

The method relies on the incorporation of photo-reactive amino acid analogues (for example analogues such as leucine and methionine derivatives with photoreactive diazirine groups) into the proteins of interest. Their similarity with the natural amino acids allows them to escape strict identity control mechanisms during protein synthesis and they are incorporated into proteins by the usual translation machinery. Diazirines groups are activated after exposure to ultraviolet radiation, allowing the monitoring of interacting proteins located within a few Ångstroms of the photo-reactive amino acid analogue [78].

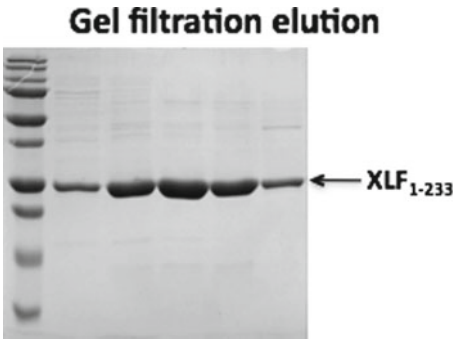


Fig. 1.6 XLF_{1-233} in denaturing gel

1.2.2 Biochemical Approaches

1.2.2.1 Co-immunoprecipitation and Immunoaffinity Chromatography

Co-immunoprecipitation (also known as a pull-down) is commonly used to verify interactions between suspected interacting partners in cell extracts. An antibody that binds specifically to a protein of interest is added and the antibody-protein complex is pelleted, often using protein-G sepharose which binds most antibodies. In immunoaffinity chromatography the specific antibody is used to immobilise the protein of interest to the column. Other interacting proteins are identified by Western blot or by sequencing a purified protein band. These approaches are suitable for testing direct interactions and for an initial screen for interacting proteins.

1.2.2.2 Denaturing and Native Gels

Denaturing and native gels are widely used to analyse protein samples. Denaturing gels (SDS-PAGE) (see Fig. 1.6) contain the denaturing detergent sodium dodecyl sulfate (SDS) and are used to check protein stability and sample purity after each step of purification. After heating to 100°C, negatively charged SDS binds to the denatured protein polypeptide in proportion to polypeptide size. The denatured proteins with uniform negatively charged density are then separated in denaturing gels under electrophoresis according to their individual molecular weights [70].

Native gels (Native-PAGE) do not contain protein-denaturing agent and therefore can be used for analyzing protein self-association or aggregation, as well as protein-protein and protein-DNA interactions in native conditions. Protein samples are separated according to their molecular mass, surface charge and molecular conformation. Protein samples run in a native gel can be recovered from the gel and analyzed further using denaturing gel. Both denaturing and native gels are essential biochemical tools for the study of multiprotein assemblies.

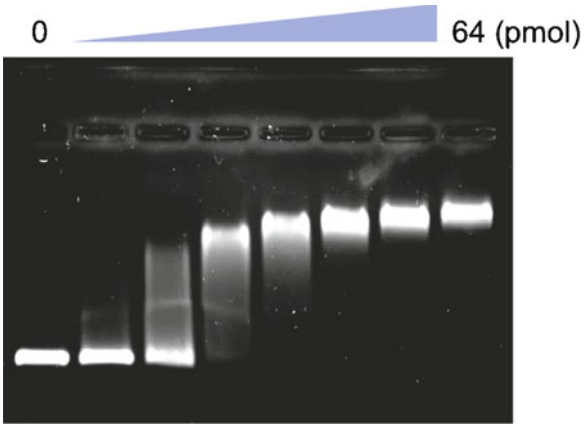


Fig. 1.7 EMSA of LigIV/XRCC4 with DNA. The figure shows the interaction of various quantities of LigIV/XRCC4 with 0.5 pmol of 350 bp DNA without 5' phosphates on 0.8%(w/v) TBE agarose gel. The quantity of LigIV/XRCC4 was doubled sequentially from 1 to 64 pmol. DNA was visualized with 0.5 μg/ml of EtBr

1.2.2.3 Gel Shifts for Nucleotide Interactions

Electrophoretic mobility shift assays (EMSA) (see Fig. 1.7) provide a simple and rapid way to study protein and DNA interactions [24]. Both acrylamide and agarose gels are used to study macromolecule/nucleotide interactions. Positions of nucleotides can be checked easily if they are labeled with radioactive or fluorescent markers. If not, gels can be stained with, for example, EtBr and SYBR Gold (Invitrogen). Macromolecule components of each band can be analyzed with SDS-page. A good example of the use of this method is Górna et al. [27], who describe an extensive EMSA study of protein/RNA complexes.

1.2.2.4 Protein Cross-linking

For highly dynamic multiprotein assemblies, protein cross-linking provides a powerful way to capture these protein assemblies. An example of a cross-linker is bis(sulfosuccinimidyl)suberate (BS3), which contains *N*-hydroxysulfosuccinimide (NHS) esters at each end and is a water-soluble version of disuccinimidyl suberate. Cross-linked proteins can then be identified and analyzed using SDS-PAGE gels and mass spectrometry. Such protein cross-linking information can be used to map the functional interaction partners in complex multiprotein assemblies.

1.2.2.5 Analytical Gel Filtration Chromatography

Evidence of oligomeric or multiprotein assemblies can be seen from gel filtration. Since gel filtration can separates molecules depending on molecular sizes and

shapes [61], the alternation of molecular volumes and shapes by forming complexes shifts elution volumes from gel filtration columns. A range of gel filtration matrices, which can separate various sizes of molecules, is available from different companies, e.g. Hagel [30], Table 8.3.4. Prepacked Superdex 75 or 200 5/150 GL columns (GE healthcare), which have become available recently, may be useful for structural studies of macromolecular complexes because a small volume of a sample can run in a very short time without diluting samples.

1.2.3 Biophysical Methods

1.2.3.1 Circular Dichroism

Circular dichroism (CD) arises from the difference of absorption of right- and left-handed polarized light. In protein solutions, far UV CD arises from amide bonds of the peptides and depends on the environment of the amides; thus each secondary structure has a characteristic CD spectrum and composition of secondary structures can be analyzed [29, 41]. The near UV CD can be used to investigate the environments of aromatic sidechains. Thus, CD can be used to check folding of macromolecules and to investigate its dependence on temperature, pH and chemical agents. CD can be used to investigate the effects of mutations, chemical modification and ligand binding on the conformations of proteins [41]. It can also be used to study DNA and RNA structures [63].

1.2.3.2 Analytical Ultracentrifugation

Analytical ultracentrifugation (AUC) is a powerful method for studying multiprotein assemblies in physiological solution conditions without labeling and chemical modification [71]. The protein sample distribution is monitored in real time by an optical detector during centrifugation. Two analytical ultracentrifugation experiments Sedimentation Velocity (SV) and Sedimentation Equilibrium (SE) analytical ultracentrifugation can be performed by analytical centrifuges (e.g. Optima XL-I Beckman) to study protein self-association, protein-protein, protein-DNA interaction (see Fig. 1.8).

Sedimentation velocity is normally used to as a hydrodynamic method to study the protein static association interaction, in which the dissociation process is relatively slow in the time scale of experiment. [36] The rate of protein complex sedimentation under high centrifugal force depends on its own molecular mass, density and shape. Therefore this real-time measured rate value is used to calculate sedimentation coefficient, which can lead to calculation of sample heterogeneity, protein complex molar mass, stoichiometry, low resolution complex shape and possible complex conformational change [18, 71].

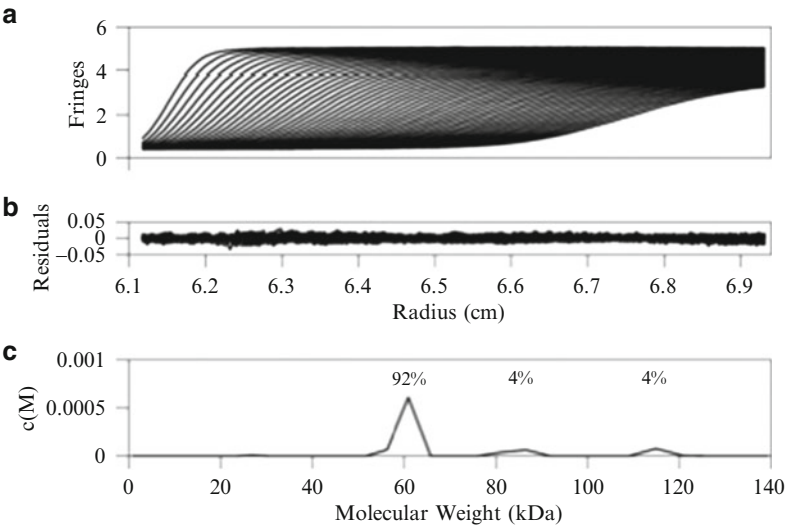


Fig. 1.8 AUC sedimentation velocity profile and residuals of XLF. The peak at 60 kDa (3) corresponds to the XLF dimer, which takes a relevant concentration of 92% [44]

For protein complexes with dynamic interaction, the association and dissociation process is too fast to be detected in the time scale of the experiment. In order to study the dynamic protein interaction, sedimentation equilibrium is more commonly used as a thermodynamic method to obtain information about the composition dependence of signal-average buoyant molar mass [36]. The sample is run at lower centrifugal force than sedimentation velocity in order to reach the sedimentation equilibrium. When the sedimentation transport force is balanced by the reverse-direction, protein molecular diffusion force leads to the establishment of concentration gradient. At sedimentation equilibrium, the concentration distribution only depends on molecular mass, not the shape. Sedimentation equilibrium is a powerful method for studying the protein self-association property under different concentrations and protein-protein interaction kinetics [71], and with a multiwavelength detection system can also be used to detect protein-DNA interactions.

1.2.3.3 SAXS

Small angle X-ray scattering (SAXS) can be used to study shapes, conformations and oligomeric states of macromolecules [79, 80], and to provide structural information of flexible and disordered macromolecules [4]. SAXS observes randomly orientated macromolecules in solution; therefore, in a way that differs from X-ray crystallography, the scattering intensity is averaged over orientation. Scattering profiles can be inverse Fourier transformed to the distance distribution function, shapes of which directly reflect the shapes of macromolecules [80]. To obtain characteristic parameters of macromolecules, a Guinier plot can be used to identify monodispersity

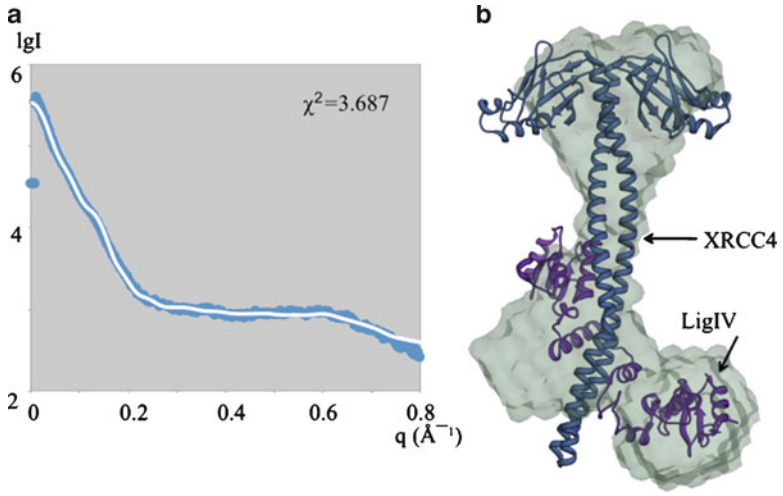


Fig. 1.9 SAXS study of BRCT domains of LigIV and a mutated XRCC4 (residue 1–213) (BmX4). (**a**) The figure shows the SAXS curve of BmX4 and a simulated SAXS curve from the corresponding crystallographic structure (PDB: 3II6) [86]. (**b**) The shape reconstruction envelop of BmX4 was produced using Gasbor [79]. The normalized spatial discrepancy from ten individual models was 1.86. The molecular envelope was produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081). The crystallographic model was superimposed by using Chimera [59]

and the slope gives the radius of gyration [80]. Folding states of molecules can be checked with a Kratky plot; a folded protein should give a parabolic peak [62].

It is also possible to obtain structural information of macromolecules. From simulated SAXS curves from atomic structures, we can extract structural information at up to 5 Å resolution, including shape and folding of macromolecules [79] (see Fig. 1.9). Structures of macromolecules are obtained by using *ab initio* modeling [80] as well as rigid-body modeling [58]. The modeling is now possible even if those molecules are flexible and contains disordered linkers [4, 55]. The SAXS data can be in incorporated into model building as special restraints [23]. In Europe, we can perform SAXS experiments at SAXS beamlines in synchrotron facilities such as ID14-3 in ESRF (Grenoble, France) and X33 in DESY (Hamburg, Germany). Recent hardware developments at the SIBYL beamline in the Advanced Light Source (Berkeley, U.S.A) allow one to perform SAXS experiments with only 12 µl of 1 mg/ml of sample [37].

1.2.3.4 Surface Plasmon Resonance

Biacore Surface Plasmon Resonance (SPR) technology is widely used to identify interactions, to measure binding and dissociation affinities, and to calculate dissociation constants and binding stoichiometries for protein-protein and protein-DNA

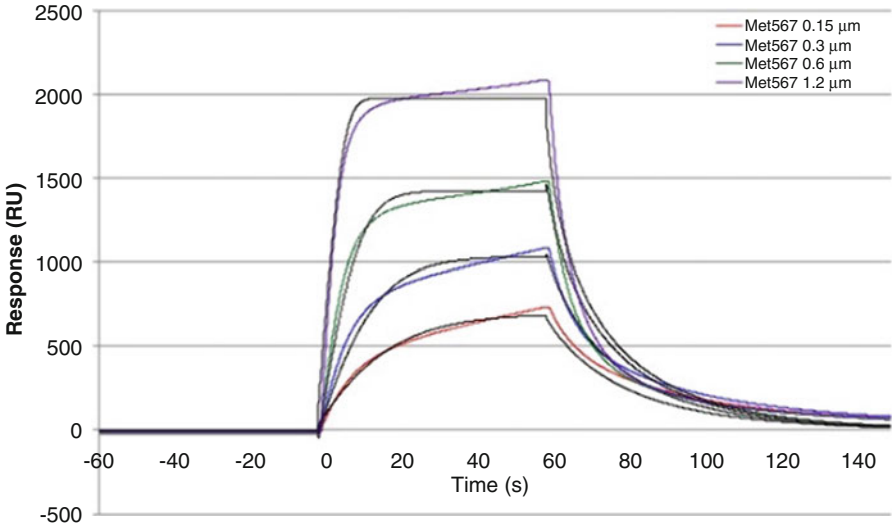


Fig. 1.10 SPR results for Met567 binding to NK1 with different initial concentrations (unpublished data from Anna Gudny Sigurdardottir, Department of Biochemistry, University of Cambridge)

interactions (Fig. 1.10). A Biacore sensor chip containing a thin layer of gold is used to immobilize one protein or DNA; other proteins can flow over the chip surface. If protein-protein or protein-DNA interaction occurs, the refractive index of the solvent near the protein-immobilized gold-layer side will change. SPR is an optical technique that can measure the refractive index differences by detecting the change of the reflection incidence angle of polarized light [38]. The output sensorgram is used to analyse the binding kinetics and interaction model.

The advantages of using Biacore SPR for protein-protein and protein-DNA interaction of complex multicomplex assemblies studies are (1) determination of real time association and dissociation constants; (2) label free for target proteins; (3) relative small quantities of protein and nucleic acids requirement; (4) automatic and high throughput.

1.2.3.5 Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) is currently the only biophysical method that can not only measure the interaction association constant (K_a), and also calculate the two thermodynamic values for the interaction process: the change of enthalpy ΔH and entropy ΔS [43].

In ITC, typically one protein solution is injected stepwise into the reaction cell containing another protein solution. The heat released or absorbed from protein-protein interaction process can be calculated by ITC instrument through measuring the energy needed to keep the reaction and reference cells in the same temperature level [43] (Fig. 1.11). Besides obtaining association constant (K_a), enthalpy change

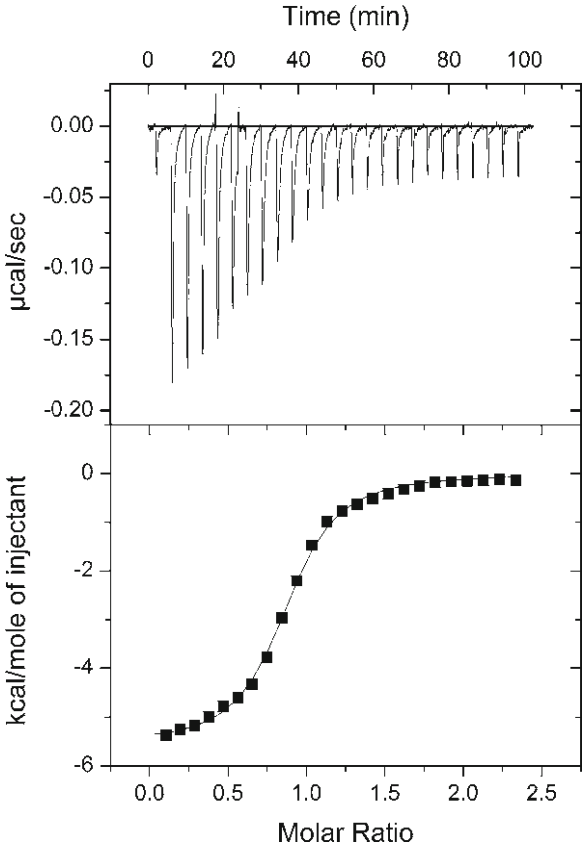


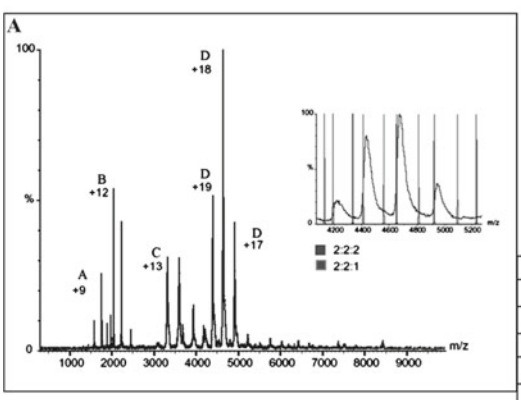
Fig. 1.11 ITC result for FGF1-heparin complex with FGFR2 (unpublished data from Alan Brown, Department of Biochemistry, University of Cambridge)

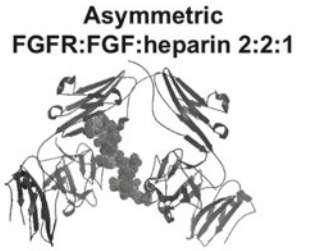
(ΔH) and entropy change (ΔS), further calculations can also lead to the free energy (ΔG), heat capacity of binding (ΔC_p) and complex stoichiometry. Together with structural information, ITC results can help to mapping the protein complex interaction region and energetic contribution [60].

1.2.3.6 Nano-electrospray Ionization MS

Nano-electrospray ionization mass spectrometry (Nano-ESI-MS) is a powerful method to study intact protein complexes. After buffer exchange with ammonium acetate, the sample is sprayed by nanoflow capillary. Sample droplets formed will desolvate and eventually turn into gas phase ions, entering into the vacuum stage mass analyser [32].

Nano-ESI-MS is used to (1) Identify the molecular weights of oligomeric proteins, protein complexes and protein-DNA complexes very accurately. Knowing the molecular mass of complex molecular and individual components, the complex stoichiometries can be identified (Fig. 1.12); (2) Study of the complex





Species	Theoretical Mass	Measured Mass (Da)
FGF (A)	15788	15787 ± 1
FGFR (B)	24492	24493 ± 6
Heparin	2881	not observed
2:2:1 (D)	83441	83452 ± 3
2:2:0	80560	81123 ± 42
2:2:1 Dimer	166882	167004 ± 31
2:2:1 Trimer	250323	261831 ± 30

Fig. 1.12 Nano-electrospray ionization mass spectrometry for FGF/FGFR/Heparin complexes reveals a predominant FGF2:FGFR:heparin 1:1:1 stoichiometry [33]

subunit binding strength through tandem MS. Protein complex subunits can dissociate from the original complex through increasing the voltage and pressure to the collision cell. By comparing the energy required to separate specific subunit interaction from the complex, the strength of protein assembled in the complex can be analyzed [35, 72].

1.2.3.7 Dynamic Light Scattering

For the study of multiprotein assembly structures, dynamic light scattering (DLS), also known as quasi elastic light scattering (Quels) and photon correlation spectroscopy (PCS), measures laser light scattered from soluble macromolecules or suspended particles. It is a useful tool for measuring the protein sample polydispersity and state of aggregation after purification, in different sample concentrations and during protein sample buffer screening prior to setting up protein crystallization trials. DLS can also determine the rough sizes and shapes of proteins in solution.

1.2.3.8 Fluorescence Spectroscopy

Fluorescence spectroscopy is used to study interactions of proteins and other molecules [10]. The modern approaches bring together classic fluorescence techniques and advances in laser excitation and detection capabilities with novel probes and chemistries to couple them to proteins and nucleic acids. For example, probes can be introduced by modifying cysteines at surface positions of proteins, where they do not affect aggregation. For nucleic acids 6-FAM tagged molecules can be purchased. Fluorescence is measured by luminescence spectrometry.

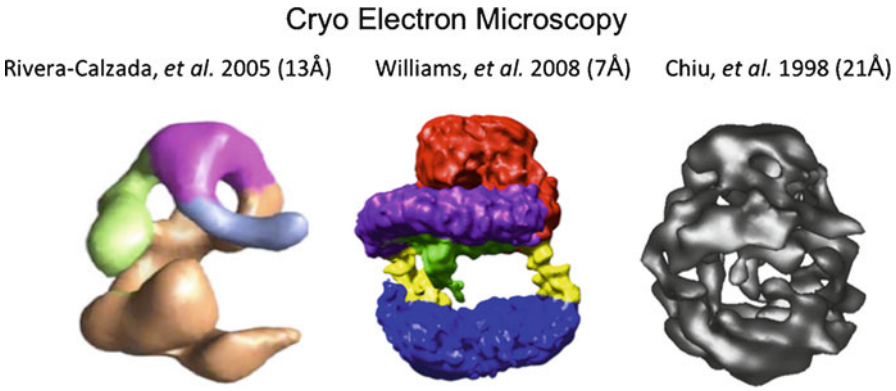


Fig. 1.13 Cryo EM results for DNA-PKcs. References to the publications and resolutions of the models are given above

1.2.4 Structure Determination

1.2.4.1 Electron Microscopy

Single particle reconstruction electron microscopy can be used to obtain low and medium resolution images of complexes. Negatively-stained images of macromolecules, for instance by uranium acetate, give molecular envelopes [9]. Although resolution is lower than cryo-EM, sample preparation is simple and stained molecules provide high contrast [9]. Cryo-EM provides internal structures of macromolecules [9] (Fig. 1.13). Recent developments of various aspects of cryo-EM pushed reconstructed images to near atomic resolution, e.g. 3.3 Å [89]. GraFix, ultracentrifugation in solutions with glycerol and fixation gradients is a recently developed method to provide homogeneous samples, and can improve the quality of EM images [40]. Model building into EM maps can be achieved with and without crystallographic structures [46, 68]. As a guide of the modeling, hexahistidine tags and DNA labeled with gold clusters (Nanoprobes) allow the location of the tags and DNA in complexes. In reverse, EM maps have been successfully used for phasing X-ray diffraction data by molecular replacement [51, 88].

1.2.4.2 Crystallisation

"State-of-the-art" equipment nowadays includes robotics for carrying out the crystallisation trials, preparation of crystallisation buffers and crystal growth monitoring. The search for crystallisation conditions typically involves trials of a number of commercial crystallisation screens (about 2,000 conditions). This can now be set up with sitting drops vapour-diffusion crystallisation trials, for example using Phoenix 96-channel crystallisation robot (Alpha Biotech) (Fig. 1.14). This robot can dispense 50 nL drops allowing preservation of protein material. Monitoring of crystal growth can be done with highly automated crystal imaging and monitoring

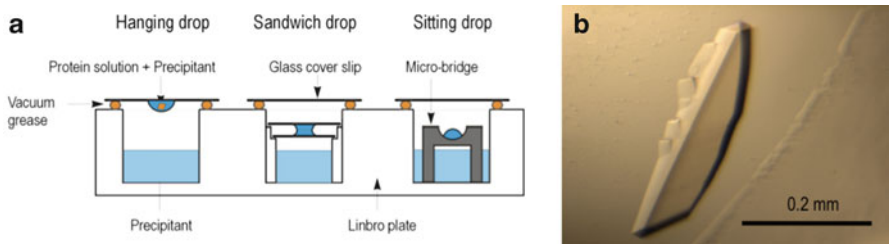


Fig. 1.14 Protein crystallisation using vapour-diffusion technique. (**a**) Schematic representations of protein crystallisation apparatus by achieving solution supersaturation of the protein through vapour-diffusion technique using hanging, sandwich or sitting drop methods. (**b**) Protein crystals – crystal of DNA-PKcs

system (for example: Minstrel, by Rigaku; Rock Imager by Formulatrix). Once the formulations of preliminary crystallisation conditions are found, they are refined by varying concentrations of precipitating agents and the protein, for which small factorial optimisation screens can be created using liquid handling robot (Firdom EVO by Tecan Ltd). The final crystallisation trials are often set up manually using the hanging drop vapour-diffusion technique.

1.2.4.3 X-ray Diffraction Data Collection

A high intensity X-ray generator equipped with a cryogenic device and a CCD area detector as an in-house source as well as access to synchrotron radiation sources such as Diamond Light Source (Oxford, UK) and European Synchrotron Radiation Facility (Grenoble, France) are optimally required for performing X-ray diffraction data collection experiments. Assessment of the diffraction quality of protein crystals and collection of preliminary sets of data are often carried out using an in–house source with the primary dataset collection performed at a synchrotron source. Crystals of large multicomponent complexes generally have limited diffracting abilities, therefore requiring the sole use of synchrotron facilities (Fig. 1.15). A nitrogen cryogenic device (like Cobra Cryostream by Oxford CryoSystems, Ltd) is routinely used to maintain cryogenic conditions for crystals during data collection.

1.2.4.4 Crystal Structure Determination

Crystal structure determinations of multiprotein complexes, where structures of individual components have been defined elsewhere, are performed using the Molecular Replacement (MR) method. In other cases the phase information is obtained by Multiwavelength Anomalous Dispersion (MAD), Single wavelength Anomalous Diffraction (SAD) or Multiple Isomorphous Replacement (MIR) methods using either selenomethionyl proteins or by incorporation of heavy metal ions. For very large multiprotein complexes, the use of heavy metal clusters, like tantalum bromide cluster, which was used in determination of DNA-PKcs crystal structure [74], proved to be more successful (Fig. 1.16).

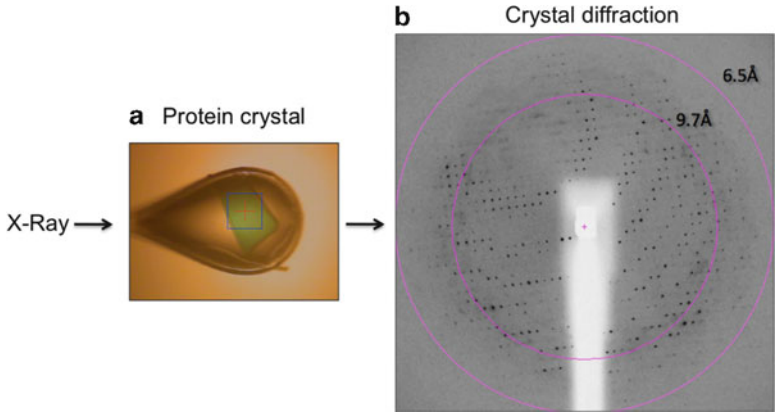


Fig. 1.15 X-ray diffraction data collection. (**a**) Mounted crystal of DNA-PKcs soaked with hexatantalum tetradecabromide ($Ta_6Br_{12}^{2+}$) are green in colour, in a loop prepared for cryogenic X-ray diffraction data collection. (**b**) Diffraction pattern extending to 6.6 Å resolution obtained from the above crystal of DNA-PKcs using synchrotron radiation source (beamline ID29, ESRF, France)

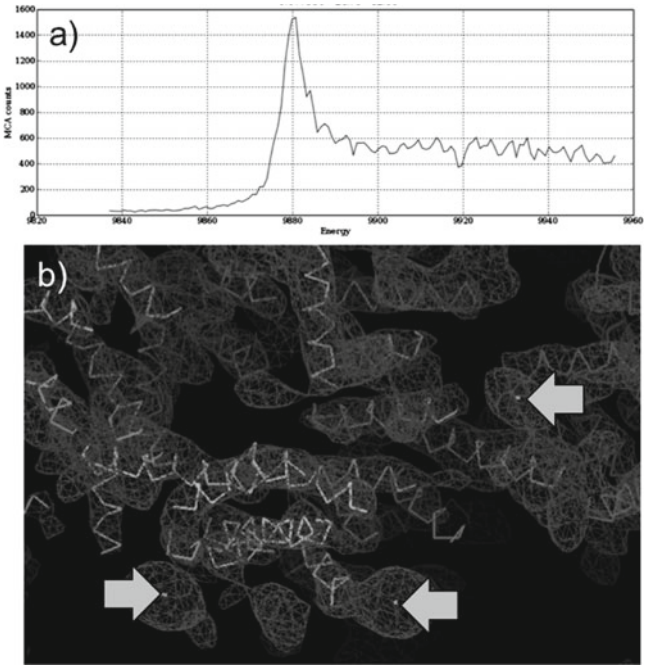


Fig. 1.16 Crystal structure determination (**a**) Experimentally measured tantalum atom edge fluorescence scan of DNA-PKcs crystals soaked with hexatantalum tetradecabromide ($Ta_6Br_{12}^{2+}$). (**b**) Electron density map calculated at 6.6 Å resolution using phases determined by the MAD method showing the location of $Ta_6Br_{12}^{2+}$ as well as manually built alpha-helices of DNA-PKcs

1.3 An Example in Depth: The Multiprotein System of Non-Homologous End Joining

We now illustrate the approach described above in the study of the spatial and temporal organisation of the Non-Homologous End Joining (NHEJ) complexes that are involved in DNA double-strand break repair.

1.3.1 Background

Non-Homologous End Joining (NHEJ) and Homologous Recombination (HR) are the two major pathways of DNA double-strand break (DSB) repair in human cells. HR functions mainly in cell cycle late S/G2 phases due to its requirement of a sister chromatid [81]. In contrast, NHEJ repairs DSBs directly without any DNA template and plays its role in G1/early S phases [52, 81]. DNA DSBs, the most severe DNA damages in eukaryotic cells, can be generated by ionizing radiation, reactive oxygen species and DNA replication across a nick [52]. A single DBS can potentially kill cells and erroneous repairing of DSBs may cause the loss and amplification of chromosomal materials or chromosomal translocations, which are carcinogenic [42]. NHEJ is also responsible for programmed DSBs in V(D)J recombination [26] and class switch recombination [13] during development of immune diversity.

The NHEJ process comprises *synapsis*, *end processing* and *ligation* [45]. During *synapsis* DNA-dependent protein kinase (DNA-PK), which consists of Ku70, Ku80, DNA-PK catalytic subunit (DNA-PKcs) and DNA. Ku70 and Ku80, assembles around the broken DNA ends and ring-shaped heterodimers maintain them in proximity [21, 82]. DNA-PKcs (phosphoinositide 3-kinase-related serine/threonine kinase), is recruited to DNA ends through interaction with the C-terminus of Ku80 [25, 34, 75]. Two DNA-PK complexes are probably required to hold the two DNA ends close together [77]. DNA-PKcs phosphorylates itself and various other proteins, including NHEJ components [76, 84]. The *end processing* involves nucleases such as Artemis [50], which exhibits 5′ to 3′ endonuclease activity after activation by DNA-PKcs phosphorylation [48, 50]. The final *ligation* step is mediated by DNA ligase IV (LigIV), in a stable complex with dimeric X-ray cross-complementation group 4 (XRCC4) [17, 28]. XLF/Cernunnos also interacts with XRCC4, and enhances the LigIV DNA ligation process [1, 12]. Figure 1.17 summarises our current knowledge of NHEJ protein interactions and phosphorylation by DNA-PKcs.

1.3.2 Structural Biology of Individual Components

1.3.2.1 Ku

The crystal structure of the Ku70/80, which does not include the C-terminal DNA-PKcs interaction domain of Ku80 (Ku80CTD), and its complex with a DNA

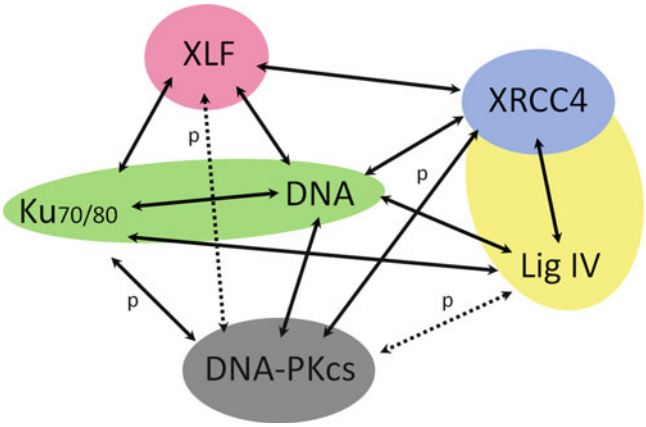


Fig. 1.17 Schematic diagram of interactions of NHEJ double-strand break DNA repair proteins. Colour filled shapes indicate proteins and complexes with known crystal structures. Letter "P" indicates phosphorylation by DNA-PKcs

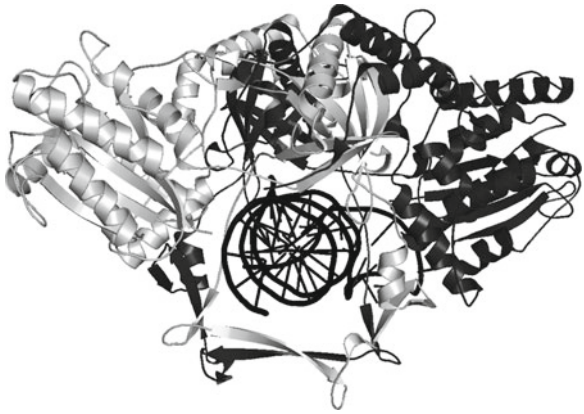


Fig. 1.18 Crystal structure of Ku70/80 bound with DNA (PDB: 1JEY)

fragment (Fig. 1.18) revealed a ring structure that encircles the duplex DNA [82]. No large conformational changes occur on binding of heterodimeric Ku except for the DNA binding C-terminal domain of Ku70. No contacts with DNA bases and only a few interactions with the sugar-phosphate backbone are made.

1.3.2.2 DNA-PKcs

Single particle electron microscopy reconstruction of DNA-PKcs Chiu et al. [14], Rivera-Calzada et al. [66]) and Williams et al. [85] gives a good impression of the overall structure. The crystal structure of DNA-PKcs in complex with the C-terminal domain of Ku80 (Fig. 1.19) at 6.6 Å resolution has been determined

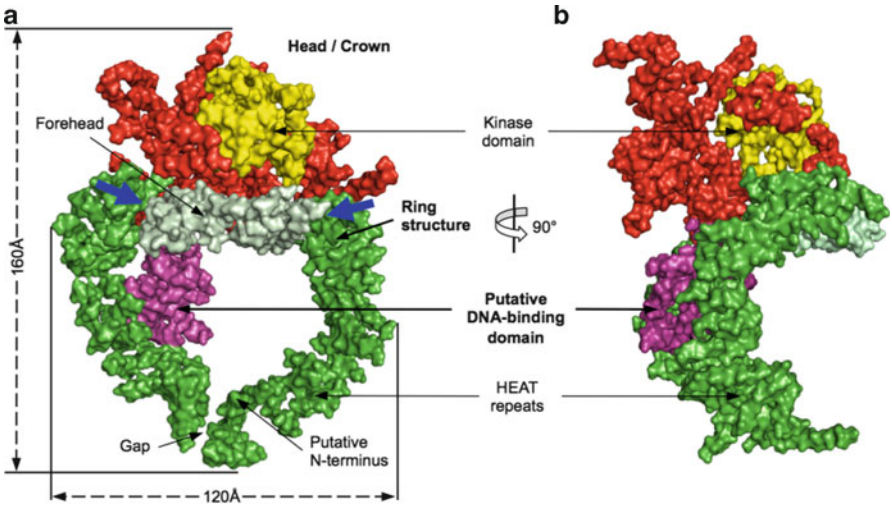


Fig. 1.19 Crystal structure of DNA-PKcs. Molecular surface of the DNA-PKcs structure showing (**a**) front and (**b**) side views. Also shown in (**a**) is the overall size of DNA-PKcs with the potential flexible sites indicated by arrows (figure adapted from [74])

using multi-wavelength anomalous dispersion method with $Ta_6Br_{12}^{2+}$ heavy metal cluster [74].

The DNA-PKcs tertiary structure (Fig. 1.3a) comprises repeats that form a hollow circular structure comprising ~66 α-helices. Within this circular structure, the regularity of the HEAT repeats breaks down, possibly indicating flexibility. The ring structure provides a platform for proteins that engage in the repair of broken DNA and which, together with Ku, holds in place the DNA while it is being repaired.

In the C-terminal region the chain forms the Head/Crown, which contains the FAT, kinase domain, FATC. The kinase structure was identified and modeled from PI(3)Kγ, one of the family members, which was superposed onto the Head/Crown region, resulting in a plausible fit to the N-lobe β-strands and the C-lobe α-helices.

1.3.2.3 DNA Ligase IV

Human LigIV is unstable by itself but it is stabilised by interaction with XRCC4 [11] and is pre-adenylated in human cells and ready for the ligation [67]. LigIV can be divided into catalytic and interaction regions. The catalytic region is conserved among other human DNA ligases and contains the DNA binding domain, (DBD), the nucleotidyltransferase domain (NTase) and the OB-fold domain (OBD) [54]. The region that interacts with XRCC4 [17] and Ku70/80 [16] consists of two BRCT domains connected by a flexible linker. Further work is required to understand the structural differences between the catalytic region of LigIV and that of the two other human DNA ligases and the spatial arrangement of the two BRCT domains in free form.

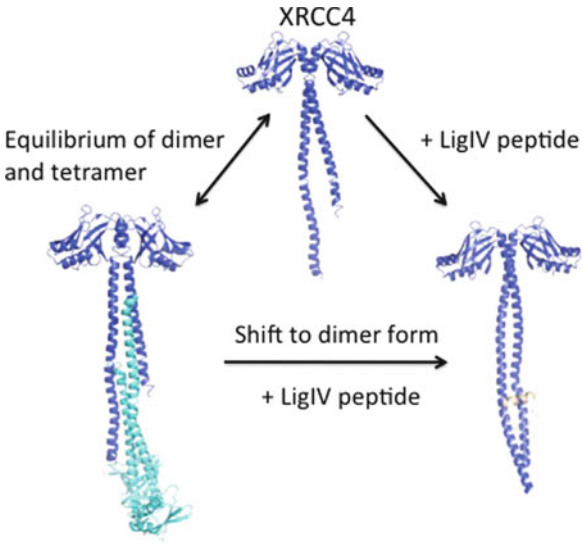


Fig. 1.20 The XRCC4 dimer and tetramer equilibrium. Complex of a dimer with the LigIV peptide shifts the equilibrium to dimer

1.3.2.4 XRCC4 and XLF

XRCC4 is a homodimer containing an N-terminal head domain and an elongated coiled coil at the C-terminal [39, 73]. The region of the XRCC4 C-terminal domain after residue 213 is not included in current solved XRCC4 crystal structures due to its predicted flexible structure. XRCC4 can exist as a salt-dependent equilibrium of dimers and tetramers in solution [49]. The dynamic of XRCC4 oligomers formation can be also shifted to dimmers through strong binding of LigIV to XRCC4 C-terminal helices [49]. The binding region between XRCC4 and LigIV overlaps with the XRCC4 tetramerisation region, which may explain why LigIV functions as a strong competitor to shift the equilibrium towards the XRCC4 dimer in solution [49] (Fig. 1.20).

XLF (XRCC4-Like Factor), which is conserved throughout a wide range of eukaryotes, localizes to the nucleus of human cells [1], consistent with the presence of a nuclear localization sequence (NLS) at the C-terminal. XLF is an obligate homodimer with a globular N-terminal head domain and extended coiled-coil helical tail, which is folded back around the coiled-coil (Fig. 1.21) [2, 44]. XLF and XRCC4 contain the similar head domains, which include seven-stranded antiparallel β-structure sandwiching a helix-turn-helix motif, but XLF contains an extra helix at the N-terminus. XLF contains distinct helices folding back at the C-terminus, which are absent in XRCC4. The structural differences between XLF and XRCC4 tail structures may explain why LigIV does not interact with XLF in the same way as XRCC4. XLF enhances the LigIV/XRCC4 DNA ligation process [65], but the exact functions and mechanisms of action of XLF in NHEJ are still not fully understood.

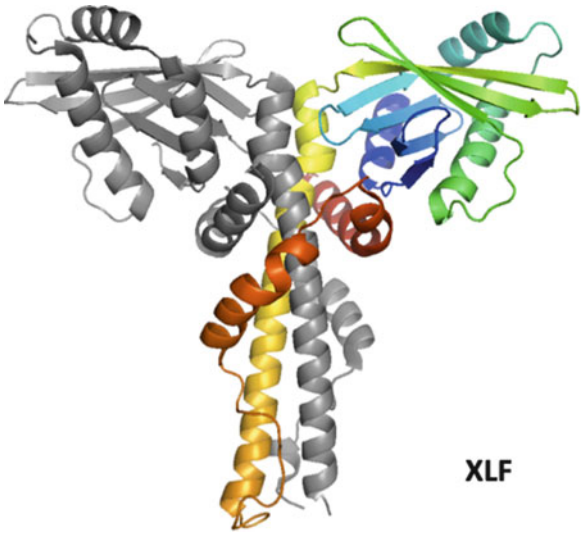


Fig. 1.21 Crystal structure of XLF1-233 homodimer (PDB:2QM4) (Modified figure from [44])

1.3.3 Structural Biology of Complexes

1.3.3.1 DNA-PKcs/Ku/DNA Ternary Complex (DNA-PK)

Ku80CTD, which is dispensable for the binding of Ku70/80 to DNA and is absent in the crystal structure of the Ku70/80 heterodimer, is an α-helical molecule required for DNA-PK recruitment to the sites of damaged DNA [25, 75]. The Ku heterodimer is required for binding double-stranded (ds) DNA ends and DNA binding leads to the recruitment of the of DNA-PKcs. X-ray crystallography [74], single-particle electron microscopy (EM) [66] and SAXS combined with live cell imaging [31] have not been successful in locating its position.

DNA-PKcs/Ku70/Ku80 holo-enzyme structures and possible synaptic complexes defined using cryo-electron microscopy [8] have provided evidence of conformational changes in human DNA-PKcs when double-stranded DNA binds, and suggested that this may correlate with the activation of the kinase. Spagnolo et al. [77] using single-particle electron microscopy at ~25 Å resolution of the holo-enzyme assembled on DNA found further evidence for conformational changes on binding of Ku and DNA to DNA-PKcs. SAXS studies of DNA-PK indicated two different modes of dimerisation [31] and have demonstrated that DNA-PK phosphorylation causes a large conformational change, sufficient to open the gap in the ring [74] and provide access to or release from DNA [57].

1.3.3.2 DNA Ligase IV/XRCC4

The tight complex of dimeric XRCC4 with BRCT domains of LigIV [17] is mediated by a well ordered linker [73] (Fig. 1.22a) which is likely to be unstructured in

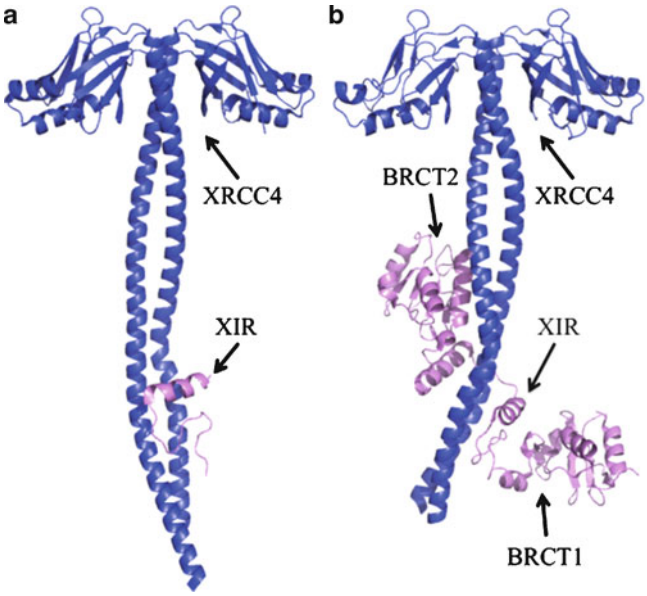


Fig. 1.22 Crystal structures of XRCC41-213 -LigIV peptide (PDB: 1IK9) and XRCC41-203-LigIV BRCT domains (PDB: 3II6)

isolation [73]. Additional contacts between LigIV and XRCC4 are made through the region following the linker and the second BRCT domain both in yeast and human (Fig. 1.22b) [20, 86]. The kink and right-handed undecad coiled-coil in one helix of the coiled-coil of human XRCC4 dimer in the linker complex is replaced by a bend in the opposite direction in the complex between XRCC4 and the BRCT domains [86]. The catalytic region of LigIV seems to be flexibly attached to the BRCT domains even though it forms complex with XRCC4 [53, 64]. This flexibility may be important for the catalytic activity of the complex in NHEJ [57].

1.3.3.3 XLF/XRCC4 Complexes

XLF was found to interact with XRCC4 through protein head domains in a yeast two-hybrid study of various mutants [1, 19]. The interaction between XLF and XRCC4 is salt sensitive and does not depend on DNA [65]. XLF that is bound to beads at its C-terminus is still able to pull down LigIV/XRCC4, implying that the C-terminal of XLF is not important for interaction with LigIV/XRCC4 [47]. The XLF-XRCC4 protein complex structure was recently solved at 8.5 Å, which revels that XLF and XRCC4 dimers interact through their head domains and form an alternating left-handed helical structure with polypeptide coiled coils and pseudo-dyads of individual XLF and XRCC4 dimers at right angles to the helical axis (Fig. 1.23).

Fig. 1.23 XLF-XRCC4 alternating helical complex structure [87]

1.3.3.4 Spatial Arrangement of Higher Order Complexes

In order to follow the formation of the LigIV/XRCC4/XLF/DNA complex, the order and dynamics of protein assembly must be determined. The interaction between XLF and XRCC4 is weak compared to the strong binding of XRCC4 and LigIV. It is not clear whether the XLF-dimer interactions with XRCC4-dimer are maintained when the ligase is recruited. Protein interaction assays have confirmed the XRCC4-independent XLF recruitment to DSBs ends through interaction with Ku only in the presence of DNA, so XLF may act independently of XRCC4.

1.4 Discussion

1.4.1 Multiprotein Systems and New Approaches to Structure Analysis

Our observations of NHEJ repair are typical of many cell regulation systems. Large numbers of protein, nucleic acid and other components tend to assemble and disassemble, requiring the ability to describe a series of transient complexes in both space and time. We have seen that for such dynamic systems, techniques like SAXS, EM, mass spectrometry and X-ray crystallography need to be combined. But new approaches for their study will be required.

One new opportunity is to exploit free-electron lasers (FEL), which produce X-ray pulses of very high intensity and short duration, during which the system will change very little. Fortunately, the molecule starts to decay as a result of the enormous forces generated by the strong incident light only after the X-ray flash has passed the sample and the image of the atomic structure has been collected. FELs have proven to be successful even for very small crystals of relatively bad quality. Current FEL facilities include the Free electron LASer in Hamburg (FLASH), the Linac Coherent Light Source (LCLS) at the SLAC National Accelerator Laboratory,

the European X-ray free-electron laser, the SPring-8 Compact SASE Source (SCSS) and the PSI SwissFEL. Up-to-date information on FEL research and applications can be found in The World Wide Web Virtual Library (URL: http://sbfel3.ucsb.edu/www/vl_fel.html).

An important aspect of the structural study of dynamic macromolecular complexes is their stabilisation and fixation. This might be achieved by phospho-mimicking mutations, truncation, post-translational modifications, such as phosphorylation and methylation, or the use of GraFix to stabilise macromolecule [40] for EM or X-ray studies.

Temporal aspects must be understood. FRET offers opportunities and of course is widely used in whole cell studies to study conformational changes, protein interactions and translocation. Also, time-resolved Raman scattering and X-ray scattering can be very powerful to study those macromolecular dynamics when they are combined with spatial techniques for structural biology. Indeed the challenge will be to extend the analysis of structure and dynamics of isolated complexes to direct studies of the spatial and temporal organisation in the cell.

1.4.2 Multiprotein Systems and Therapeutic Intervention

Because many diseases require therapeutic intervention at the level of multiprotein systems involved in cell regulation, the detailed knowledge of their architectures provides an important basis for targeting these processes in drug discovery.

The traditional approach would be to target well defined active sites of enzyme families, for example to improve the design of inhibitors that bind at the ATP site of DNA-PKcs or the active site of the ligase. In our view a more selective but also a more challenging approach would be to target allosteric sites, or template/adaptor binding sites in the complexes that are critical to the activation, co-localisation and/or specificity of the regulation of NHEJ – we call this *allotargetting*. This has advantages because regulatory interactions tend to be more specific to a member of a large family than active sites. But the disadvantage of this approach is the need to modulate protein-protein interfaces (see [83]). Our approach is to use fragment-based methods [6, 7, 15]. For DNA repair we might target protein-protein interactions critical to the spatial organisation and regulation. Likely targets would be the head-to-head interactions of XRCC4 and XLF, the interactions of BRCT domains, the interaction of Ku and DNA-PKcs and so on. Thus, knowledge of the structure in space and time for multiprotein systems will likely not only give insights into the mechanism of processes, but will also assist the design of chemical tools and ultimately the discovery of lead compounds for therapeutic intervention.

Acknowledgements We thank the Wellcome Trust for funding work in our laboratory on multiprotein systems involved in cell signalling, regulation and repair. We are grateful to Anna Gudny Sigurdardottir, Dr Anja Winter and Dr Alan Brown for preparing figures and for helpful discussion. TO was supported by an Overseas Research Studentship (ORS).

References

1. Ahnesorg P, Smith P, Jackson SP (2006) XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA Non-homologous End-joining. Cell 124(2):301–313
2. Andres S, Modesti M, Tsai CJ, Chu G, Junop MS (2007) Crystal structure of human XLF: a twist in nonhomologous DNA End-joining. Mol Cell 28(6):1093–1101
3. Beenken A, Mohammadi M (2010) The FGF family: biology, pathophysiology and therapy. Nat Rev Drug Discov 8:235–253
4. Bernadó P, Blanchard L, Timmins P, Marion D, Ruigrok RWH, Blackledge M (2005) A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. Proc Nat Acad Sci USA 102(47):17002–17007
5. Blundell TL, Bolanos-Garcia V, Chirgadze DY, Harmer NJ, Lo T, Pellegrini L, Sibanda BL (2002) Asymmetry in the multiprotein systems of molecular biology. Struct Chem 13:405–412
6. Blundell TL, Jhoti H, Abell C (2002) High-throughput crystallography for lead discovery in drug design. Nat Rev Drug Discov 1:45–54
7. Blundell TL, Sibanda BL, Montalvao RW, Brewerton S, Vijayalakshmi C, Worth CL, Harmer NJ, Davies O, Burke D (2006) Structural biology and bioinformatics in drug design: opportunities and challenges for target identification and lead discovery. Philos Trans R Soc B 361:413–423
8. Boskovic J, Rivera-Calzada A, Maman JD, Chacón P, Willison KR, Pearl LH, Llorca O (2003) Visualization of dna-induced conformational changes in the DNA repair kinase dna-pkcs. EMBO J 22(21):5875–5882
9. Bremer A, Henn C, Engel A, Baumeister W, Aebi U (1992) Has negative staining still a place in biomacromolecular electron microscopy? Ultramicroscopy 46(1–4):85–111
10. Brown M, Royer C (1997) Fluorescence spectroscopy as a tool to investigate protein interactions. Curr Opin Biotechnol 8(1):45–49
11. Bryans M, Valenzano MC, Stamato TD (1999) Absence of DNA ligase IV protein in xr-1 cells: evidence for stabilization by xrcc4. Mutat Res/DNA Repair 433(1):53–58
12. Buck D, Malivert L, de Chasseval R, Barraud A, Fondanèche MC, Sanal O, Plebani A, Stéphan JL, Hufnagel M, le Deist F, Fischer A, Durandy A, de Villartay JP, Revy P (2006) Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. Cell 124(2):287–299
13. Chaudhuri J, Alt FW (2004) Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. Nat Rev Immunol 4:541–552
14. Chiu CY, Cary RB, Chen DJ, Peterson SR, Stewart PL (1998) Cryo-EM imaging of the catalytic subunit of the DNA-dependent protein kinase. J Mol Biol 284:1075–1081
15. Congreve M, Murray CW, Blundell TL (2005) Structural biology and drug discovery. Drug Discov Today 10:895–907
16. Costantini S, Woodbine L, Andreoli L, Jeggo PA, Vindigni A (2007) Interaction of the ku heterodimer with the dna ligase iv/xrcc4 complex and its regulation by dna-pk. DNA Repair 6(6):712–722
17. Critchlow SE, Bowater RP, Jackson SP (1997) Mammalian dna double-strand break repair protein xrcc4 interacts with dna ligase iv. Curr Biol 7(8):588–598
18. Dam J, Schuck P (2004) Calculating sedimentation coefficient distributions by direct modeling of sedimentation velocity concentration profiles. Methods Enzymol 384:185–212. doi:10.1016/S0076-6879(04)84012-6
19. Deshpande R, Wilson T (2007) Modes of interaction among yeast Nej1, Lif1 and Dnl4 proteins and comparison to human XLF, XRCC4 and Lig4. DNA Repair 6(10):1507–1516
20. Doré AS, Furnham N, Davies OR, Sibanda BL, Chirgadze DY, Jackson SP, Pellegrini L, Blundell TL (2006) Structure of an xrcc4-dna ligase iv yeast ortholog complex reveals a novel brct interaction mode. DNA Repair 5(3):362–368
21. Downs JA, Jackson SP (2004) A means to a DNA end: the many roles of Ku. Nat Rev Mol Cell Biol 5(5):367–378

22. Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 6:197–208
23. Förster F, Webb B, Krukenberg KA, Tsuruta H, Agard DA, Sali A (2008) Integration of small-angle x-ray scattering data into structural modeling of proteins and their assemblies. J Mol Biol 382(4):1089–1106
24. Garner MM, Revzin A (1981) A gel electrophoresis method for quantifying the binding of proteins to specific dna regions: application to components of the escherichia coli lactose operon regulatory system. Nucleic Acids Res 9(13):3047–3060
25. Gell D, Jackson SP (1999) Mapping of protein-protein interactions within the dna-dependent protein kinase complex. Nucleic Acids Res 27(17):3494–3502
26. Gellert M (2002) V(d)j recombination: Rag proteins, repair factors, and regulation. Annu Rev Biochem 71:101–132
27. Górna MW, Pietras Z, Tsai Y-C, Callaghan AJ, Hernández H, Robinson CV, Luisi BF (2010) The regulatory protein rraa modulates rna-binding and helicase activities of the e. Coli rna degradosome. RNA 16(3):553–562
28. Grawunder U, Wilm M, Wu X, Kulesza P, Wilson TE, Mann M, Lieber MR (1997) Activity of dna ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. Nature 388(6641):492–495
29. Greenfield NJ (1996) Methods to estimate the conformation of proteins and polypeptides from circular dichroism data. Anal Biochem 235(1):1–10
30. Hagel L. (1998) Gel-Filtration Chromatography. Current Protocols in Protein Science 8.3.1-8.3.30
31. Hammel M, Yu Y, Mahaney BL, Cai B, Ye R, Phipps BM, Rambo RP, Hura GL, Pelikan M, So S, Abolfath RM, Chen DJ, Lees-Miller SP, Tainer JA (2010) Ku and dna-dependent protein kinase dynamic conformations and assembly regulate dna binding and the initial non-homologous end joining complex. J Biol Chem 285(2):1414–1423
32. Hanson CL, Robinson CV (2004) Protein-nucleic acid interactions and the expanding role of mass spectrometry. J Biol Chem 279(24):24907–24910. doi:10.1074/jbc.R300037200
33. Harmer NJ, Ilag LL, Mulloy B, Pellegrini L, Robinson CV, Blundell TL (2004) Towards a resolution of the stoichiometry of the fibroblast growth factor (FGF)-FGF receptor-heparin complex. J Mol Biol 339:821–834
34. Hartley KO, Gell D, Smith GCM, Zhang H, Divecha N, Connelly MA, Admon A, Lees-Miller SP, Anderson CW, Jackson SP (1995) Dna-dependent protein kinase catalytic subunit: a relative of phosphatidylinositol 3-kinase and the ataxia telangiectasia gene product. Cell 82(5):849–856
35. Hernandez H, Robinson C (2007) Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. Nat Protoc 2(3):715–726
36. Howlett GJ, Minton AP, Rivas G (2006) Analytical ultracentrifugation for the study of protein association and assembly. Curr Opin Chem Biol 10(5):430–436. doi:10.1016/j.cbpa.2006.08.017
37. Hura GL, Menon AL, Hammel M, Rambo RP, Poole Ii FL, Tsutakawa SE, Jenney FE Jr, Classen S, Frankel KA, Hopkins RC, Yang S-J, Scott JW, Dillard BD, Adams MWW, Tainer JA (2009) Robust, high-throughput solution structural analyses by small angle x-ray scattering (saxs). Nat Methods 6(8):606–612
38. Jönsson U, Fägerstam L, Ivarsson B, Johnsson B, Karlsson R, Lundh K et al (1991) Real-time biospecific interaction analysis using surface plasmon resonance and a sensor chip technology. Biotechniques 11(5):620–627
39. Junop MS, Modesti M, Guarné A, Ghirlando R, Gellert M, Yang W (2000) Crystal structure of the xrcc4 dna repair protein and implications for end joining. EMBO J 19(22):5962–5970
40. Kastner B, Fischer N, Golas MMM, Sander B, Dube P, Boehringer D, Hartmuth K, Deckert J, Hauer F, Wolf E, Uchtenhagen H, Urlaub H, Herzog F, Peters JMM, Poerschke D, Lührmann R, Stark H (2008) Grafix: sample preparation for single-particle electron cryomicroscopy. Nat Methods 5(1):53–55
41. Kelly S, Jess T, Price N (2005) How to study proteins by circular dichroism. Biochim Biophys Acta (BBA) - Proteins and Proteomics 1751(2):119–139
42. Khanna KK, Jackson SP (2001) Dna double-strand breaks: signaling, repair and the cancer connection. Nat Genet 27(3):247–254

43. Leavitt S (2001) Direct measurement of protein binding energetics by isothermal titration calorimetry. Curr Opin Struct Biol 11(5):560–566. doi:10.1016/S0959-440X(00)00248-7
44. Li Y, Chirgadze DY, Bolanos-Garcia V, Sibanda BL, Davies O, Ahnesorg P, Jackson S, Blundell T (2008) Crystal structure of human XLF/Cernunnos reveals unexpected differences from XRCC4 with implications for NHEJ. EMBO J 27(1):290–300
45. Lieber MR, Ma Y, Pannicke U, Schwarz K (2004) The mechanism of vertebrate nonhomologous dna end joining and its role in v(d)j recombination. DNA Repair 3(8–9):817–826
46. Lindert S, Stewart PL, Meiler J (2009) Hybrid approaches: applying computational methods in cryo-electron microscopy. Curr Opin Struct Biol 19(2):218–225
47. Lu H, Pannicke U, Schwarz K, Lieber M (2007) Length-dependent binding of human XLF to DNA and stimulation of XRCC4-DNA ligase IV activity. J Biol Chem 282(15):11155–11162
48. Ma Y, Pannicke U, Lu H, Niewolik D, Schwarz K, Lieber MR (2005) The DNA-dependent protein kinase catalytic subunit phosphorylation sites in human Artemis. J Biol Chem 280(40):33839–33846
49. Modesti M, Junop MS, Ghirlando R, van de Rakt M, Gellert M, Yang W, Kanaar R (2003) Tetramerization and dna ligase iv interaction of the dna double-strand break repair protein xrcc4 are mutually exclusive. J Mol Biol 334(2):215–228
50. Moshous D, Callebaut I, de Chasseval R, Corneo B, Cavazzana-Calvo M, Le Deist F, Tezcan I, Sanal O, Bertrand Y, Philippe N, Fischer A, de Villartay J-P (2001) Artemis, a novel dna double-strand break repair/v(d)j recombination protein, is mutated in human severe combined immune deficiency. Cell 105(2):177–186
51. Navaza J (2008) Combining X-ray and electron-microscopy data to solve crystal structures. Acta Crystallogr D 64(1):70–75
52. O'Driscoll M, Jeggo PA (2005) The role of double-strand break repair — insights from human genetics. Nat Rev Genet 7(1):45–54
53. Ochi T, Sibanda BL, Wu Q, Chirgadze DY, Bolanos-Garcia VM, Blundell TL (2010) Structural biology of DNA repair: spatial organisation of the multicomponent complexes of nonhomologous End joining, J Nucleic Acids 2010:1–19
54. Pascal JM, O'Brien PJ, Tomkinson AE, Ellenberger T (2004) Human DNA ligase I completely encircles and partially unwinds nicked DNA. Nature 432(7016):473–478
55. Pelikan M, Hura GL, Hammel M (2009) Structure and flexibility within proteins as identified through small angle x-ray scattering. Gen Physiol Biophys 28(2):174–189
56. Pellegrini L, Burke DF, von Delft F, Mulloy B, Blundell TL (2000) Crystal structure of fibroblast growth factor receptor ectodomain bound to ligand and heparin. Nature 407:1029–1034
57. Perry JJ, Cotner-gohara E, Ellenberger T, Tainer JA (2010) Structural dynamics in DNA damage signaling and repair. Curr Opin Struct Biol 1–12
58. Petoukhov MV, Svergun DI (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data. Biophys J 89(2):1237–1250
59. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF chimera–a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612
60. Pierce M, Raman C, Nall B (1999) Isothermal titration calorimetry of protein-protein interactions. Methods (San Diego, Calif) 19(2):213–221. doi:10.1006/meth.1999.0852
61. Potschka M (1987) Universal calibration of gel permeation chromatography and determination of molecular shape in solution. Anal Biochem 162(1):47–64
62. Putnam CD, Hammel M, Hura GL, Tainer JA (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Q Rev Biophys 40(03):191–285
63. Ranjbar B, Gill P (2009) Circular dichroism techniques: Biomolecular and nanostructural analyses- a review. Chem Biol Drug Des 74(2):101–120
64. Recuero-Checa MA, Doré AS, Arias-Palomo E, Rivera-Calzada A, Scheres SHW, Maman JD, Pearl LH, Llorca O (2009) Electron microscopy of xrcc4 and the dna ligase iv–xrcc4 dna repair complex. DNA Repair 8(12):1380–1389

65. Riballo E, Woodbine L, Stiff T, Walker S, Goodarzi A, Jeggo P (2009) XLF-Cernunnos promotes DNA ligase IV-XRCC4 re-adenylation following ligation. Nucleic Acids Res 37(2):482–492
66. Rivera-Calzada A, Maman JD, Spagnolo L, Pearl LH, Llorca O (2005) Three-dimensional structure and regulation of the DNA-dependent protein kinase catalytic subunit (DNA-PKcs). Structure 13:243–255
67. Robins P, Lindahl T (1996) Dna ligase iv from hela cell nuclei. J Biol Chem 271(39): 24257–24261
68. Rossmann M, Morais M, Leiman P, Zhang W (2005) Combining x-ray crystallography and electron microscopy. Structure 13(3):355–362
69. Sasaki K, Dockerill S, Adamiak DA, Tickle IJ, Blundell TL (1975) X-ray analysis of glucagon and its relationship to receptor binding. Nature 257:751–757
70. Schägger H, von Jagow G (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal Biochem 166:368–379
71. Scott D, Harding SE, Rowe A (2005) A brief introduction to the analytical ultracentrifugation of proteins for beginners. In: David RAS, Harding SE (eds) Analytical ultracentrifugation: technique and methods. The Royal Society of Chemistry, Cambridge. doi:10.1039/9781847552617
72. Sharon M, Taverner T, Ambroggio XI, Deshaies RJ, Robinson CV (2006) Structural organization of the 19 S proteasome lid: insights from MS of intact complexes. PLoS Biol 4(8):doi:10.1371/journal.pbio.0040267
73. Sibanda BL, Critchlow SE, Begun J, Pei XY, Jackson SP, Blundell TL, Pellegrini L (2001) Crystal structure of an xrcc4-dna ligase iv complex. Nat Struct Biol 8(12):1015–1019
74. Sibanda BL, Chirgadze D, Blundell TL (2010) Crystal structure of DNA-PKcs reveals a large open-ring cradle comprised of HEAT repeats. Nature 463(7277):118–121
75. Singleton BK, Torres-Arzayus MI, Rottinghaus ST, Taccioli GE, Jeggo PA (1999) The C-terminus of ku80 activates the dna-dependent protein kinase catalytic subunit. Mol Cell Biol 19(5):3267–3277
76. Smith GCM, Jackson SP (1999) The dna-dependent protein kinase. Genes Dev 13(8):916–934
77. Spagnolo L, Rivera-Calzada A, Pearl LH, Llorca O (2006) Three-dimensional structure of the human dna-pkcs/ku70/ku80 complex assembled on DNA and its implications for DNA dsb repair. Mol Cell 22(4):511–519
78. Suchanek M, Radzikowska A, Thiele C (2005) Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. Nat Methods 2(4):261–267. doi:10.1038/nmeth752
79. Svergun DI, Koch MHJ (2003) Small-angle scattering studies of biological macromolecules in solution. Rep Prog Phys 66(10):1735–1782
80. Svergun DI, Petoukhov MV, Koch MHJ (2001) Determination of domain structure of proteins from x-ray solution scattering. Biophys J 80(6):2946–2953
81. Takata M, Sasaki MS, Sonoda E, Morrison C, Hashimoto M, Utsumi H, Yamaguchi-Iwai Y, Shinohara A, Takeda S (1998) Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in verterbrate cells. EMBO J 17:5497–5508
82. Walker JR, Corpina RA, Goldberg J (2001) Structure of the ku heterodimer bound to dna and its implications for double-strand break repair. Nature 412(6847):607–614
83. Wells J, McClendon C (2007) Reaching for high-hanging fruit in drug discovery at protein,Äìprotein interfaces. Nature 450:1001–1009
84. Weterings E, Chen DJ (2008) The endless tale of non-homologous end-joining. Cell Res 18(1):114–124
85. Williams DR, Lee KJ, Chen DJ, Shi J, Stewart PL (2008) Cryo-EM structure of the DNA-dependent protein kinase catalytic subunit at subnanometer resolution reveals alpha helices and insight into DNA binding. Structure 16:468–477

86. Wu P-YY, Frit P, Meesala S, Dauvillier S, Modesti M, Andres SN, Huang Y, Sekiguchi J, Calsou P, Salles B, Junop MS (2009) Structural and functional interaction between the human dna repair proteins dna ligase iv and xrcc4. Mol Cell Biol 29(11):3163–3172
87. Wu Q, Ochi T, Matak-Vinkovic D, Robinson CV, Chirgadze DY, Blundell TL (2011) Non-homologous end-joining partners in a helical dance: structural studies of XLF-XRCC4 interactions. Biochem Soc Trans 39(5):1387–1392
88. Xiong Y (2008) From electron microscopy to X-ray crystallography: molecular-replacement case studies. Acta Crystallogr D 64(1):76–82
89. Zhang X, Jin L, Fang Q, Hui WH, Zhou ZH (2010) 3.3 Å cryo-em structure of a nonenveloped virus reveals a priming mechanism for cell entry. Cell 141(3):472–482

Chapter 2
Co-translational Protein Processing, Folding, Targeting, and Membrane Insertion of Newly Synthesized Proteins

Daniel Boehringer and Nenad Ban

Abstract Newly synthesized proteins leave the ribosome through the nascent polypeptide tunnel. Through the coordinated action of the ribosome associated chaperones, nascent chain processing enzymes, the signal recognition particle, and the protein insertion machinery newly synthesized proteins are brought into their native state and proper cellular localization. The interplay of these factors during ongoing synthesis requires spatial and temporal control of their interactions with the ribosome. We used electron microscopy in combination with crystallography and biochemical methods to study the structure of bacterial ribosomes and nascent chain interacting factors.

Keywords Ribosome • Co-translational processing • Folding • Targeting • Membrane insertion

2.1 The Ribosomal Tunnel

From the site of peptide bond formation, the peptidyl transferase active site of the ribosome, newly synthesized proteins have to diffuse through the ribosomal nascent polypeptide tunnel to reach the cytosol. During their progression through the tunnel initial folding events take place (Fig. 2.1).

The tunnel might actively promote the formation of secondary structure elements by providing a distinctive interaction surface for the nascent chain [28, 48]. However, these interactions might also delay folding and can lead to translational arrest by

D. Boehringer • N. Ban (✉)
Institute of Molecular Biology and Biophysics, ETH Zurich, Schafmattstr. 20, 8093 Zurich, Switzerland
e-mail: ban@mol.biol.ethz.ch

M.A. Carrondo and P. Spadon (eds.), *Macromolecular Crystallography*, NATO Science for Peace and Security Series A: Chemistry and Biology,
DOI 10.1007/978-94-007-2530-0_2, © Springer Science+Business Media B.V. 2012

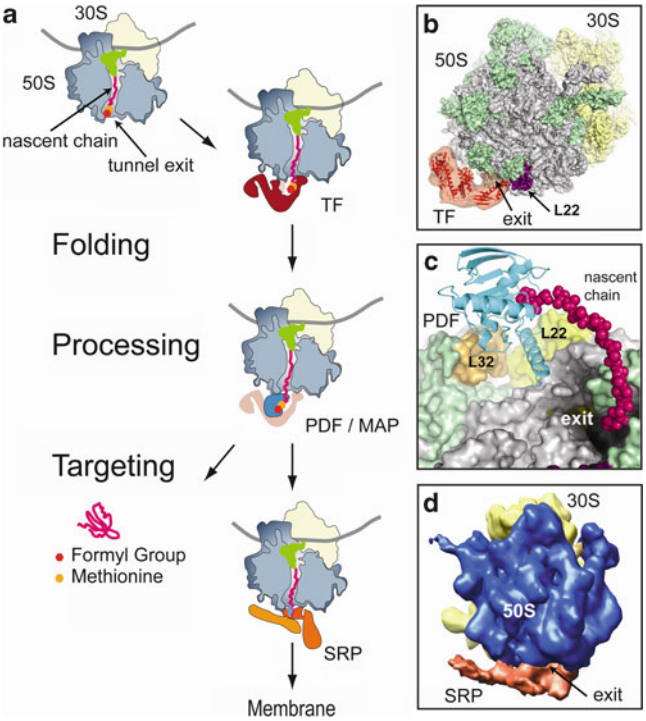


Fig. 2.1 Overview of co-translational protein folding and targeting in bacteria. (**a**) Schematic overview. (**b**) Structure of trigger factor (TF) bound on the ribosome. The crystal structure of trigger factor (pdb 2VRH, *red ribbon*) was fitted into the EM reconstruction (*transparent red*). The ribosomal subunits (pdb 2AW4) were fitted into the density and are shown in surface representation with the 50S rRNA in *grey* and the 50S ribosomal proteins in *green*. Protein L23 at the tunnel exit (*exit*) is colored *magenta*. The 30S ribosomal subunit is shown in *yellow*. (**c**) Peptide deformylase (PDF, *blue ribbon*) bound to the ribosomal tunnel exit (*exit*) on protein L32 (*orange*) and L22 (*yellow*) (50S rRNA in *grey*, 50S ribosomal proteins in *green*). The nascent chain interacting with the PDF is modeled (*red spheres*) (**d**) Signal recognition particle (SRP) bound on a translating ribosome. The EM-reconstructions shows SRP (*red*) bound to the tunnel exit of the ribosome with the 50S ribosomal subunit colored *blue* and the 30S ribosomal subunit colored *yellow*

stalling the ribosome [12, 37]. The tunnel wall is mainly formed by ribosomal RNA and extended loops of the proteins L4, L22 and L23 that reach from the ribosomal surface into the ribosomal RNA core [1, 47]. The newly synthesized protein leaves the ribosome through the funnel shaped exit of the ribosomal tunnel. At the rim of the tunnel exit several ribosomal proteins are located; L22, L23, L24 and L29 are universally conserved, whereas L32 and L17 are bacterial specific [38]. The proteins together with ribosomal RNA form the attachment site for folding, processing and targeting factors that help nascent polypeptides adopt their native state and reach the proper cellular localization.

2.2 The Ribosome Associated Chaperon Trigger Factor

The bacterial ribosome associated chaperone trigger factor is likely to be the first factor to interact with nascent chains at the tunnel exit [30]. The affinity of trigger factor for ribosomes and its cellular abundance makes it likely that it is stoichiometrically bound to all ribosomes in the cell [26]. Furthermore, many nascent chains emerging from the ribosomal tunnel exit can be cross-linked to trigger factor [15]. The affinity of trigger factor for ribosome nascent chain complexes increases with the length of the nascent chain and some newly synthesized proteins remain associated with trigger factor even after their release from the ribosome [19]. DnaK and Trigger factor act on overlapping pool of substrates in the cell and when either DnaK or trigger factor is deleted the other chaperone can take over its function [5]. The combined deletion of both chaperones causes aggregation of several hundred different newly synthesized proteins and synthetic lethality above 30°C in *Escherichia coli* [5, 43].

The crystal structure of trigger factor revealed the elongated shape of the factor with four protruding regions, tail, arms and head [9] (Fig. 2.1b). The four protrusions form a hydrophobic cradle that is critical for the chaperoning function of trigger factor. The N-terminal domain of trigger factor binds to the ribosome in the vicinity of the tunnel exit using ribosomal protein L23 as its docking site [25]. The head is formed by a peptidyl-prolyl isomerase (PPIase) domain, which is dispensable for the general chaperone function of trigger factor [24]. In the middle of the three dimensional structure of trigger factor arm like protrusions are formed from the C-terminal region of trigger factor that are critical for the chaperone function [34, 50]. The co-crystal structure of an N-terminal fragment of *E. coli* trigger factor and an archeal 50S ribosomal subunit suggested that trigger factor arches over the tunnel exit creating a protected folding space for the nascent chain [9]. This cradle for newly synthesized polypeptides appeared to be large enough to accommodate a small folded domain. In support of this model, the cryo EM structure of trigger factor bound to a *E. coli* ribosome nascent chain complex showed that trigger factor arches over the tunnel exit leaving enough space between the arms and the ribosomal surface for small domain to fold [33] (Fig. 2.1b). Comparing the conformations of trigger factor in complex with substrate protein and in its unbound state revealed only minor conformational changes. This is also supported by recent crystallographic experiments on *Thermotoga maritima* trigger factor in complex with a substrate protein [31]. The observed structural changes were limited to the position of the N-terminal tail and the PPIase domain, which somewhat influences the size of the cradle [31, 33]. This might allow nascent chains of different length to be accommodated inside trigger factor. The structural results agree with protease protection experiments on ribosome nascent chain complexes. Trigger factor can protect nascent chains of up to 100 amino acids against proteolytic degradation in an unfolded or folded state [16, 33, 44]. The interplay of DnaK and trigger factor in the folding of newly synthesized polypeptides is a matter of ongoing research and the exact role and the substrate pool of trigger factor in the cell still remains to be established.

2.3 Nascent Chain Processing

In bacteria translation is initiated with a formylmethionine [11]. The formyl group is assumed to prevent side reactions of the reactive amino group and thereby increases the efficiency of translation initiation [42]. After initiation, the formyl group is co-translationally cleaved away by peptide deformylase (PDF) followed by removal of the remaining methionine by methionine amino peptidase (MAP) [11]. Recent crystallographic studies showed that *E. coli* PDF binds to a grove between ribosomal proteins L22 and L32 near the tunnel exit on the large ribosomal subunit via its C-terminal helix [3] (Fig. 2.1c). In vivo studies showed that the truncating the C-terminal helix reduces the viability and growth rates of *E. coli* suggesting that ribosome binding is important for PDF function [3]. The binding site of PDF is close to the binding site of trigger factor but not overlapping. Modeling both enzymes on the ribosome indicates that both factors could potentially bind simultaneously [3, 33]. The lateral openings of the trigger factor cradle appear to be located such that PDF could access the nascent chain while trigger factor is bound. Therefore, trigger factor might act as a passive molecular router that presents the N-terminus of the growing nascent chain to PDF.

2.4 Targeting of Newly Synthesized Proteins to Membranes

Nascent membrane proteins need to be targeted to the membrane. In bacteria, this is achieved by the signal recognition particle (SRP) that co-translationally scans newly synthesized proteins for a signal sequence in their N-terminal region [20, 29]. The ribosome nascent chain complex is then targeted to the membrane. On the membrane SRP interacts with its receptor and, in a GTP dependent fashion, releases the nascent chain onto the translocon for insertion into the membrane.

Bacterial SRP is formed by the Ffh protein in complex with the 4.5S RNA [20]. The Ffh protein comprises of two highly conserved regions which undergo dramatic conformational rearrangements during protein targeting. One part is formed by two domains, a GTPase domain (G domain) and the ribosome binding N domain, whereas the other part of the protein is formed by the M domain. The two parts are connected via a long linker. The M domain is tightly associated with the 4.5S RNA and details of this interaction have been revealed by crystallography [2]. The M domain also recognizes the signal sequence with a hydrophobic pocket on its surface [18, 53].

SRP is bound to the ribosomes even in the absence of a nascent chain as the N domain can be cross-linked to protein L23 of empty ribosomes [13]. Cryo EM studies of this complex showed only a relatively week density above the L23 region for this complex indicating that SRP is flexibly bound to the ribosome [40]. This flexibility might allow SRP to bind to the ribosomes that already have trigger factor or processing enzymes bound to scan the nascent chain for the signal sequence. Once

the M domain detects a signal sequence, it binds to it and SRP adopts a particular conformation on the ribosome. The structure of SRP bound to a ribosome nascent chain complex with a signal sequence was investigated by cryo EM [14, 40] (Fig. 2.1d). In this conformation the M-domain binds the ribosome at the tunnel exit near L24, the 4.5S SRP RNA is stretched across the tunnel exit, and the NG domains of the Ffh protein protrude from the ribosomal surface [4].

SRP and ribosome nascent chain complex interact on the membrane with the SRP receptor [29]. Bacterial SRP receptor, FtsY, contains an N and G domain that are highly homologous to the NG domains of the SRP protein [36]. The release of the nascent chain from the SRP M domain onto the translocon is accompanied by reciprocal GTPase activation of the GTPase domains. The crystal structure of Ffh in complex with its receptor FtsY in the presence of non hydrolysable GTP analogs shows G and N domains arranged side by side [7, 10]. The ribosome accelerates complex formation between the SRP and the receptor [51]. The higher rate of complex formation depends on the presence of SRP RNA. Mutational analysis of FtsY and 4.5S RNA indicate that specific interactions are required for efficient complex assembly [17, 41]. This suggests that the ribosome bound conformation of SRP organizes the 4.5S RNA for efficient complex formation with the receptor [41].

2.5 Membrane Protein Insertion

Membrane proteins are inserted by a specialized insertion machinery into the inner membrane of bacteria [39, 49]. Co-translational protein insertion guarantees that the membrane segments will be inserted with the correct topology and prevents aggregation of the hydrophobic transmembrane regions in the cytoplasm. In contrast, most secreted proteins are transferred posttranslationally to the membrane for secretion.

Two pathways exist in bacteria for insertion of membrane proteins, the Sec and YidC pathway. The Sec pathway is used by most transmembrane proteins. The SecYEG complex is the core of this insertion machinery which associates with additional components SecDF, YajC and YidC. In this pathway YidC interacts with nascent transmembrane segments after their release from SecYEG. However there exists an alternative insertion pathway for a subset of transmembrane proteins that are inserted by YidC on its own [21] (Fig. 2.2a).

Structural insights into the binding mode of SecYEG on the ribosome were obtained by cryo-EM [35] (Fig. 2.2b). The structure of SecYEG bound to translating ribosomes showed that the translocon is located above the tunnel exit. The translocon is bound to the ribosome via three contact sites including protein Helix 59 of 50S rRNA, L24 and L23. Recent EM reconstructions of SecY in complex with non translating ribosomes showed SecY bound in a similar location contacting the same regions at the tunnel exit [32]. Thus, the translocon utilizes the same attachment sites on the ribosome as SRP. This indicates that SRP has to undergo dramatic conformational rearrangements when handing over the nascent chain to the translocon in a highly coordinated manner [23]. The crystal structure of archeal [46] and

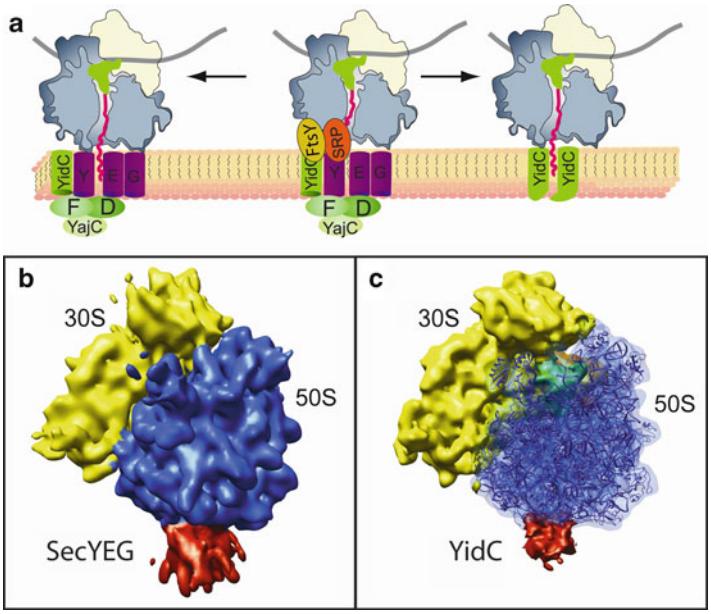


Fig. 2.2 Protein insertion by YidC and SecYEG. (**a**) Schematic overview. (**b**) Cryo-EM density of SecYEG bound to the tunnel exit of a translating ribosome. The 30S ribosomal subunit is shown in *yellow*, the 50S ribosomal subunit in *blue*, and the SecYEG density in *red*. (**c**) Cryo-EM density of YidC bound to a ribosome nascent chain complex. YidC (*red*) is bound at the tunnel exit on the 50S ribosomal subunit (*blue*). The crystal structure of the large ribosomal subunit is fitted into the density (pdb 2AW4). The 30S subunit is shown in yellow with the P-site (*green*) and E-site (*orange*) tRNA

bacterial SecY translocon complexes [45, 52] provided detailed insight into the architecture of the protein translocation pore. The SecY subunit can be divided into two inverted pseudo-symmetric halves with the channel in the center. Two pseudo-symmetric halves open at one side to release transmembrane segments of the nascent membrane protein into the membrane [6, 8, 46, 52].

In the YidC pathway, which is critical for a subset of membrane proteins, the membrane insertion is supported by YidC on its own. A recent cryo-EM structure of YidC bound to translating ribosomes provided first insights into the mechanism of this protein insertase on the ribosome [22] (Fig. 2.2c). The structure shows a dimer of YidC localized above the tunnel exit. The EM density of the ribosome-bound YidC is similar in size to a dimer of YidC as observed in a 10 Å projection map of YidC 2D crystals [27]. YidC is bound to the same regions on the ribosome as SecYEG including protein L23 and Helix 59 of 50S rRNA. The structural similarity and the mode of binding to the ribosome together with biochemical data indicates that YidC shares a common mechanism with the non homologous SecY complexes [22] (Fig. 2.2). It is therefore likely that the dimer of YidC forms a pore in the center with lateral openings through which nascent transmembrane segments are released into the membrane.

2.6 Conclusion and Outlook

The recent structures of nascent chain interacting factors on the ribosome provided important insight into folding, processing and targeting of newly synthesized proteins. Interdisciplinary experimental approaches were critical to unravel the binding modes of these factors to the ribosome and for providing first indications for the extensive conformational changes that must accompany these co-translational processes.

References

1. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science 289:905–920
2. Batey RT, Rambo RP, Lucast L, Rha B, Doudna JA (2000) Crystal structure of the ribonucleoprotein core of the signal recognition particle. Science 287:1232–1239
3. Bingel-Erlenmeyer R, Kohler R, Kramer G, Sandikci A, Antolic S, Maier T, Schaffitzel C, Wiedmann B, Bukau B, Ban N (2008) A peptide deformylase-ribosome complex reveals mechanism of nascent chain processing. Nature 452:108–111
4. Buskiewicz IA, Jockel J, Rodnina MV, Wintermeyer W (2009) Conformation of the signal recognition particle in ribosomal targeting complexes. RNA 15:44–54
5. Deuerling E, Schulze-Specking A, Tomoyasu T, Mogk A, Bukau B (1999) Trigger factor and DnaK cooperate in folding of newly synthesized proteins. Nature 400:693–696
6. du Plessis DJ, Berrelkamp G, Nouwen N, Driessen AJ (2009) The lateral gate of SecYEG opens during protein translocation. J Biol Chem 284:15805–15814
7. Egea PF, Shan SO, Napetschnig J, Savage DF, Walter P, Stroud RM (2004) Substrate twinning activates the signal recognition particle and its receptor. Nature 427:215–221
8. Egea PF, Stroud RM (2010) Lateral opening of a translocon upon entry of protein suggests the mechanism of insertion into membranes. Proc Natl Acad Sci USA 107:17182–17187
9. Ferbitz L, Maier T, Patzelt H, Bukau B, Deuerling E, Ban N (2004) Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. Nature 431:590–596
10. Focia PJ, Shepotinovskaya IV, Seidler JA, Freymann DM (2004) Heterodimeric GTPase core of the SRP targeting complex. Science 303:373–377
11. Giglione C, Boularot A, Meinnel T (2004) Protein N-terminal methionine excision. Cell Mol Life Sci 61:1455–1474
12. Gong F, Ito K, Nakamura Y, Yanofsky C (2001) The mechanism of tryptophan induction of tryptophanase operon expression: tryptophan inhibits release factor-mediated cleavage of TnaC-peptidyl-tRNA(Pro). Proc Natl Acad Sci USA 98:8997–9001
13. Gu SQ, Peske F, Wieden HJ, Rodnina MV, Wintermeyer W (2003) The signal recognition particle binds to protein L23 at the peptide exit of the Escherichia coli ribosome. RNA 9:566–573
14. Halic M, Blau M, Becker T, Mielke T, Pool MR, Wild K, Sinning I, Beckmann R (2006) Following the signal sequence from ribosomal tunnel exit to signal recognition particle. Nature 444:507–511
15. Hoffmann A, Bukau B, Kramer G (2010) Structure and function of the molecular chaperone trigger factor. Biochim Biophys Acta 1803:650–661
16. Hoffmann A, Merz F, Rutkowska A, Zachmann-Brand B, Deuerling E, Bukau B (2006) Trigger factor forms a protective shield for nascent polypeptides at the ribosome. J Biol Chem 281:6539–6545
17. Jagath JR, Matassova NB, de Leeuw E, Warnecke JM, Lentzen G, Rodnina MV, Luirink J, Wintermeyer W (2001) Important role of the tetraloop region of 4.5 S RNA in SRP binding to its receptor FtsY. RNA 7:293–301

18. Janda CY, Li J, Oubridge C, Hernandez H, Robinson CV, Nagai K (2010) Recognition of a signal peptide by the signal recognition particle. Nature 465:507–510
19. Kaiser CM, Chang HC, Agashe VR, Lakshmipathy SK, Etchells SA, Hayer-Hartl M, Hartl FU, Barral JM (2006) Real-time observation of trigger factor function on translating ribosomes. Nature 444:455–460
20. Keenan RJ, Freymann DM, Stroud RM, Walter P (2001) The signal recognition particle. Annu Rev Biochem 70:755–775
21. Kiefer D, Kuhn A (2007) YidC as an essential and multifunctional component in membrane protein assembly. Int Rev Cytol 259:113–138
22. Kohler R, Boehringer D, Greber B, Bingel-Erlenmeyer R, Collinson I, Schaffitzel C, Ban N (2009) YidC and Oxa1 form dimeric insertion pores on the translating ribosome. Mol Cell 34:344–353
23. Kramer G, Boehringer D, Ban N, Bukau B (2009) The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. Nat Struct Mol Biol 16:589–597
24. Kramer G, Patzelt H, Rauch T, Kurz TA, Vorderwulbecke S, Bukau B, Deuerling E (2004) Trigger factor peptidyl-prolyl cis/trans isomerase activity is not essential for the folding of cytosolic proteins in Escherichia coli. J Biol Chem 279:14165–14170
25. Kramer G, Ramachandiran V, Horowitz PM, Hardesty B (2002) The molecular chaperone DnaK is not recruited to translating ribosomes that lack trigger factor. Arch Biochem Biophys 403:63–70
26. Lill R, Crooke E, Guthrie B, Wickner W (1988) The "trigger factor cycle" includes ribosomes, presecretory proteins, and the plasma membrane. Cell 54:1013–1018
27. Lotz M, Haase W, Kuhlbrandt W, Collinson I (2008) Projection structure of yidC: a conserved mediator of membrane protein assembly. J Mol Biol 375:901–907
28. Lu J, Deutsch C (2005) Folding zones inside the ribosomal exit tunnel. Nat Struct Mol Biol 12:1123–1129
29. Luirink J, von Heijne G, Houben E, de Gier JW (2005) Biogenesis of inner membrane proteins in Escherichia coli. Annu Rev Microbiol 59:329–355
30. Maier T, Ferbitz L, Deuerling E, Ban N (2005) A cradle for new proteins: trigger factor at the ribosome. Curr Opin Struct Biol 15:204–212
31. Martinez-Hackert E, Hendrickson WA (2009) Promiscuous substrate recognition in folding and assembly activities of the trigger factor chaperone. Cell 138:923–934
32. Menetret JF, Schaletzky J, Clemons WM Jr, Osborne AR, Skanland SS, Denison C, Gygi SP, Kirkpatrick DS, Park E, Ludtke SJ et al (2007) Ribosome binding of a single copy of the SecY complex: implications for protein translocation. Mol Cell 28:1083–1092
33. Merz F, Boehringer D, Schaffitzel C, Preissler S, Hoffmann A, Maier T, Rutkowska A, Lozza J, Ban N, Bukau B et al (2008) Molecular mechanism and structure of trigger factor bound to the translating ribosome. EMBO J 27:1622–1632
34. Merz F, Hoffmann A, Rutkowska A, Zachmann-Brand B, Bukau B, Deuerling E (2006) The C-terminal domain of Escherichia coli trigger factor represents the central module of its chaperone activity. J Biol Chem 281:31963–31971
35. Mitra K, Schaffitzel C, Shaikh T, Tama F, Jenni S, Brooks CL 3rd, Ban N, Frank J (2005) Structure of the E. coli protein-conducting channel bound to a translating ribosome. Nature 438:318–324
36. Montoya G, Svensson C, Luirink J, Sinning I (1997) Crystal structure of the NG domain from the signal-recognition particle receptor FtsY. Nature 385:365–368
37. Nakatogawa H, Ito K (2002) The ribosomal exit tunnel functions as a discriminating gate. Cell 108:629–636
38. Nissen P, Hansen J, Ban N, Moore PB, Steitz TA (2000) The structural basis of ribosome activity in peptide bond synthesis. Science 289:920–930
39. Rapoport TA (2007) Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. Nature 450:663–669

40. Schaffitzel C, Oswald M, Berger I, Ishikawa T, Abrahams JP, Koerten HK, Koning RI, Ban N (2006) Structure of the E. coli signal recognition particle bound to a translating ribosome. Nature 444:503–506
41. Shen K, Shan SO (2010) Transient tether between the SRP RNA and SRP receptor ensures efficient cargo delivery during cotranslational protein targeting. Proc Natl Acad Sci USA 107:7698–7703
42. Solbiati J, Chapman-Smith A, Miller JL, Miller CG, Cronan JE Jr (1999) Processing of the N termini of nascent polypeptide chains requires deformylation prior to methionine removal. J Mol Biol 290:607–614
43. Teter SA, Houry WA, Ang D, Tradler T, Rockabrand D, Fischer G, Blum P, Georgopoulos C, Hartl FU (1999) Polypeptide flux through bacterial Hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains. Cell 97:755–765
44. Tomic S, Johnson AE, Hartl FU, Etchells SA (2006) Exploring the capacity of trigger factor to function as a shield for ribosome bound polypeptide chains. FEBS Lett 580:72–76
45. Tsukazaki T, Mori H, Fukai S, Ishitani R, Mori T, Dohmae N, Perederina A, Sugita Y, Vassylyev DG, Ito K et al (2008) Conformational transition of Sec machinery inferred from bacterial SecYE structures. Nature 455:988–991
46. Van den Berg B, Clemons WM Jr, Collinson I, Modis Y, Hartmann E, Harrison SC, Rapoport TA (2004) X-ray structure of a protein-conducting channel. Nature 427:36–44
47. Voss NR, Gerstein M, Steitz TA, Moore PB (2006) The geometry of the ribosomal polypeptide exit tunnel. J Mol Biol 360:893–906
48. Woolhead CA, McCormick PJ, Johnson AE (2004) Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins. Cell 116:725–736
49. Xie K, Dalbey RE (2008) Inserting proteins into the bacterial cytoplasmic membrane using the Sec and YidC translocases. Nat Rev Microbiol 6:234–244
50. Zeng LL, Yu L, Li ZY, Perrett S, Zhou JM (2006) Effect of C-terminal truncation on the molecular chaperone function and dimerization of Escherichia coli trigger factor. Biochimie 88:613–619
51. Zhang X, Kung S, Shan SO (2008) Demonstration of a multistep mechanism for assembly of the SRP x SRP receptor complex: implications for the catalytic role of SRP RNA. J Mol Biol 381:581–593
52. Zimmer J, Nam Y, Rapoport TA (2008) Structure of a complex of the ATPase SecA and the protein-translocation channel. Nature 455:936–943
53. Zopf D, Bernstein HD, Johnson AE, Walter P (1990) The methionine-rich domain of the 54 kd protein subunit of the signal recognition particle contains an RNA binding site and can be crosslinked to a signal sequence. EMBO J 9:4511–4517

Chapter 3
The Role of Multiple Sequence Repeat Motifs in the Assembly of Multi-protein Complexes

David Barford

Abstract Proteins incorporating multiple sequence repeats (for example ARM, HEAT, TPR, LRR and ankyrin) play critical roles in coordinating the assembly of multi-subunit complexes. This lecture will discuss the different types of protein architecture generated by successive copies of each repeat motif type and describe how these structures are suited to the formation of protein-protein interactions, allowing such proteins to function as scaffolding proteins in the assembly of multi-protein complexes.

3.1 Introduction

The specificity and regulation of events underlying cellular signalling and cell division processes are conferred by a variety of multi-subunit complexes. Such complexes are either constitutive or are assembled in response to extra- and intracellular signalling events. In the latter instance, formation of dynamic complexes is commonly promoted through post-translational modifications creating new recognition sites for protein-protein interfaces.

Proteins containing multiple repeats of relatively short sequence motifs (20–40 residues) play fundamental roles in mediating the assembly of multi-protein complexes. Such repeats include the tetratricopeptide repeat (TPR), HEAT repeat, armadillo repeat, ankyrin repeat and leucine rich repeat (LRR) [9]. This lecture will focus on the structure of such repeats and describe how their architecture allows the formation of protein-protein complexes and the assembly of multi-protein complexes.

D. Barford (✉)
Division of Structural Biology, Chester Beatty Laboratories,
Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK
e-mail: david.barford@icr.ac.uk

M.A. Carrondo and P. Spadon (eds.), *Macromolecular Crystallography*,
NATO Science for Peace and Security Series A: Chemistry and Biology,
DOI 10.1007/978-94-007-2530-0_3, © Springer Science+Business Media B.V. 2012

The lecture will focus on the TPR and HEAT repeats and describe the roles of these repeats for the generation of multi-protein complexes using PP2A and the anaphase promoting complex (APC/C) as examples.

3.2 Multiple Sequence Repeats

Armadillo, TPR and HEAT repeats belong to the class of helical repeat proteins. These proteins share the common property of being assembled from tandem repeats of an α-helical structural unit, creating extended superhelical structures that are ideally suited to create a protein recognition interface. Independently, these motifs lack a specific structure and function; however, they define a particular structural unit and function by virtue of their presence within a tandem array of between 3 and 25 repeating motifs.

3.2.1 The HEAT Repeat

The HEAT motif is a repetitive sequence that was first observed to be common to the huntingtin protein, an elongation factor required for protein synthesis, the PR65/A subunit of protein phosphatase 2A and the protein kinase TOR (target of rapamycin) and related PIKKs (PI3 kinase related kinases) such as DNA-PK [1]. The structure of a multi-HEAT motif protein was first defined in a subunit of protein phosphatase 2A (PP2A), namely the constant scaffolding PR65/A subunit [10]. Within the PR65/A subunit, the 15 consecutive HEAT motifs assemble to form a highly asymmetric structure consisting entirely of helices (predominantly α-helices, with some 3_{10} helices) and connecting loops, with an overall conformation that is reminiscent of a hook. In general each HEAT repeat motif is characterised by a pair of antiparallel helices (termed A and B) which are stacked in a consecutive array [10]. The repeats are usually stacked in parallel so that the A and B helices of each motif are parallel to their helix counterparts in a neighbouring repeat. This creates a structure composed of a double layer of α-helices with the A and B helices of the HEAT motif forming the outer (convex) and inner (concave) faces, respectively. The concave face of the molecule defines an arch with dimensions 65 Å × 45 Å × 35 Å. Variations of the HEAT repeat architecture caused by the non-parallel stacking of adjacent HEAT repeats are observed in the nuclear transport proteins of the β-importin family [5, 23, 26, 37]. When the relative rotations between repeats are constant, super-helical structures are generated. Recently the crystal structure of a giant HEAT motif protein, the DNA-PK catalytic subunit, incorporating over 40 HEAT repeats, was reported [31]. A variation of the HEAT repeat is the armadillo or ARM repeat, exemplified in structures of β-catenin [13] and α-importin-cargo complexes[6].

3.2.2 The TPR Motif

The TPR (tetratricopeptide) motif is a degenerate 34 amino acid sequence identified in a wide variety of proteins. It is present in tandem arrays of 3–16 motifs that form scaffolds in order to mediate protein-protein interactions and the assembly of multi-protein complexes. TPR motifs were first identified in subunits of the anaphase-promoting complex [12, 15, 19, 22, 32, 33, 36], and are present in a diverse range of proteins including the NADPH oxidase subunit $p67^{phox}$, Hsp90-binding immunophilins, transcription factors, and peroxisomal and mitochondrial import proteins. The structure of the three tandem TPR motifs in the protein serine/threonine phosphatase PP5 (protein phosphatase 5) revealed that each TPR motif consists of a pair of anti-parallel α-helices of equivalent length, termed A and B, associated with a packing angle of approximately 24° between the helix axes [8]. Adjacent TPR motifs are packed together in a parallel arrangement, such that sequentially adjacent α-helices are antiparallel in a manner that is reminiscent of a concertina. Within a tandem array of TPR motifs the packing of helix A against adjacent B helices is defined by the same spatial and angular parameters both within and between adjacent TPR motifs. Such a packing arrangement assembles an array of a single layer of antiparallel α-helices. The consequence of the uniform angular and spatial arrangement of neighbouring α-helices is the creation of a right-handed superhelical structure featuring an amphipathic channel. Multiple tandem repeats of TPR motifs, for example N-acetyl glucosamine transferase and synthetic TPR superhelices generated by Lynne Regan's laboratory adopt a super-helical right handed structure with a helical repeat of approximately seven TPR motifs, a pitch of 60 Å and a width of 42 Å [7, 17, 25]. The inside of the helix features a continuous groove for the formation of protein-protein interactions.

3.2.3 Leucine Rich Repeat (LRR)

The leucine rich repeat was first observed in the crystal structure of the ribonuclease inhibitor (Rl) which is composed of 15 tandem topics of an approximately 20 amino acid leucine-rich repeat (LRR) motif [20, 21]. Each repeat forms an α-β motif, with tandem repeats arranged consecutively and parallel to a common axis, so that the structure adopts a curved shape resembling a horseshoe, with α-helices lining the circumference and the β-strands forming a parallel β-sheet along the inner circumference.

3.2.4 The Ankyrin Repeat

Ankyrin repeats are observed in diverse proteins that share the common function to mediate protein-protein interactions. Ankyrin-repeat-containing proteins include the Notch membrane receptor, cyclin-dependent protein kinase inhibitors and the

inhibitory subunit (IκBα) of nuclear factor-κB (NF-κB). Most proteins contain tandem arrays of between 2–7 repeats; however, ankyrins have 24 repeats. Crystal and NMR structures of the ankyrin repeats present within p53BP2, $p16^{INK4}$, $p19^{INK4d}$ [4, 29], GABPβ and IkBα [14, 16] reveal a highly conserved structural motif. The ankyrin repeat is characterised by a pair of antiparallel α-helices that is linked to its neighbouring ankyrin repeat via an antiparallel β-loop, with the first β-strand of the β-loop contributed by the C-terminus of the repeat at position i and the second β-strand contributed by the N-terminus of the ankyrin repeat at position i+1. Tandem ankyrin repeats stack approximately in parallel, so that the helices in one repeat pack against their counterparts in adjacent repeats. The β-strands are roughly perpendicular to the axes of the helices, giving the stack an L-shaped cross-section. There is a left-handed twist to ankyrin repeats and the tandem repeat stack is slightly curved, creating concave and convex faces. In the IkBα structure, short helical segments following repeats 1 and 3 introduce local kinks, creating a crescent-shaped profile that appears to be important in forming a protein recognition interface with NF-κB.

The mode of protein-protein interactions that are mediated by ankyrin repeats is highly conserved. In all complexes contacts between the ankyrin repeats and target proteins involve the β-loop fingers of the ankyrin repeats. Additional contacts are provided by the surfaces of the inner α-helices of the ankyrin repeats within the NF-κB-IκBα and CDK6-$p19^{INK4d}$ complexes.

3.3 Protein Complexes

3.3.1 Anaphase Promoting Complex/Cyclosome (APC/C)

The APC/C is a multi-subunit cullin-RING E3 ubiquitin ligase that regulates progression through the mitotic phase of the cell cycle and controls entry into S phase by catalysing the ubiquitylation of cell cycle regulatory proteins such as cyclin B and securin. Selection of APC/C targets is achieved through recognition of short destruction motifs, predominantly the D-box and KEN-box. The temporal regulation of APC/C activity is achieved through a combination of two structurally related co-activator subunits, Cdc20 and Cdh1 coupled to protein phosphorylation, APC/C inhibitors, differential affinity for APC/C substrates, and auto-ubiquitylation of its cognate E2. The APC/C has been recently reviewed by [11, 27, 28, 34, 35].

In budding yeast, 13 genes encode core APC/C subunits, most of which are highly conserved in humans and other eukaryotes, and are essential for activity. The catalytic core of the APC/C is composed of the cullin subunit Apc2 and RING H2 domain subunit Apc11, analogous to the cullin and Rbx1 subunits, respectively of cullin-RING ligases (CRLs) of the SCF superfamily. The largest APC/C subunit, Apc1 is ~200 kDa in size. Interestingly, the C-terminal region of Apc1 is composed of between seven to eight tandem repeats of a 35–40 amino acid motif shared with the Rpn1 and Rpn2 subunits of the proteasome regulatory particle [24]. Such motifs,

termed PC (proteasome/cyclosome) repeats, for which no definitive structural data are available, are intriguing given the close functional relationship between the APC/C and proteasome. Suggestions that PC repeats adopt α/β architectures similar to leucine rich repeats [24], or α/α motifs typical of helical solenoids [18], thus far lack experimental evidence. The most abundant structural motif observed in APC/C subunits is the tetratricopeptide repeat (TPR) with four yeast and five vertebrate APC/C subunits composed almost exclusively of 12–15 copies of this 34-amino acid motif arranged in tandem. The TPR motif, originally discovered within proteins subsequently identified as components of the APC/C.

3.3.2 Protein Phosphatase 2A

The overall level of protein phosphorylation is controlled by the opposing and dynamic activities of protein kinases and phosphatases that catalyse protein phosphorylation and dephosphorylation events respectively. Insights into the regulation and specificity of the reactions catalysed by protein kinases and phosphatases are the key to understanding the process of protein phosphorylation. The protein phosphatases are encoded by three major families, which include the serine/threonine protein phosphatases of the PPP and PPM families and the protein tyrosine phosphatase (PTP) superfamily reviewed by [2, 3].

PP2A is a member of the PPP family of serine/threonine phosphatases (which also include PP1 and PP2B). Unlike the protein kinase family, where diversity is generated by the large number of genes encoding protein kinases (512 in humans), specificity within the PPP family of Ser/Thr phosphatases is determined by a diverse range of regulatory subunits. Diversity within the PP2A family is generated by the interaction of the core AC heterodimer (C catalytic subunit in complex with the PR65/A subunit) interacts with a diverse array of regulatory B subunits through the PR65/A subunit. The PR65/A subunit is composed entirely of 15 tandem HEAT motifs. We will discuss the structural basis for the ability of PR65/A to recognise the catalytic C subunit and a diverse array of regulatory B subunits (reviewed by Shi [30]).

References

1. Andrade MA, Bork P (1995) HEAT repeats in the Huntington's disease protein. Nat Genet 11(2):115–116
2. Barford D (1996) Molecular mechanisms of the protein serine/threonine phosphatases. Trends Biochem Sci 21(11):407–412
3. Barford D, Das AK, Egloff MP (1998) The structure and mechanism of protein phosphatases: insights into catalysis and regulation. Annu Rev Biophys Biomol Struct 27:133–164
4. Brotherton DH, Dhanaraj V, Wick S, Brizuela L, Domaille PJ, Volyanik E, Xu X, Parisini E, Smith BO, Archer SJ, Serrano M, Brenner SL, Blundell TL, Laue ED (1998) Crystal structure

of the complex of the cyclin D-dependent kinase Cdk6 bound to the cell-cycle inhibitor p19INK4d. Nature 395(6699):244–250
5. Chook YM, Blobel G (1999) Structure of the nuclear transport complex karyopherin-beta2-Ran x GppNHp. Nature 399(6733):230–237
6. Conti E, Uy M, Leighton L, Blobel G, Kuriyan J (1998) Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha. Cell 94(2):193–204
7. D'Andrea LD, Regan L (2003) TPR proteins: the versatile helix. Trends Biochem Sci 28(12):655–662
8. Das AK, Cohen PW, Barford D (1998) The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. EMBO J 17(5):1192–1199
9. Groves MR, Barford D (1999) Topological characteristics of helical repeat proteins. Curr Opin Struct Biol 9(3):383–389
10. Groves MR, Hanlon N, Turowski P, Hemmings BA, Barford D (1999) The structure of the protein phosphatase 2A PR65/a subunit reveals the conformation of its 15 tandemly repeated HEAT motifs. Cell 96(1):99–110
11. Harper JW, Burton JL, Solomon MJ (2002) The anaphase-promoting complex: it's not just for mitosis any more. Genes Dev 16(17):2179–2206
12. Hirano T, Kinoshita N, Morikawa K, Yanagida M (1990) Snap helix with knob and hole: essential repeats in S. pombe nuclear protein nuc2+. Cell 60(2):319–328
13. Huber AH, Nelson WJ, Weis WI (1997) Three-dimensional structure of the armadillo repeat region of beta-catenin. Cell 90(5):871–882
14. Huxford T, Huang DB, Malek S, Ghosh G (1998) The crystal structure of the IkappaBalpha/NF-kappaB complex reveals mechanisms of NF-kappaB inactivation. Cell 95(6):759–770
15. Irniger S, Piatti S, Michaelis C, Nasmyth K (1995) Genes involved in sister chromatid separation are needed for B-type cyclin proteolysis in budding yeast. Cell 81(2):269–278
16. Jacobs MD, Harrison SC (1998) Structure of an IkappaBalpha/NF-kappaB complex. Cell 95(6):749–758
17. Jinek M, Rehwinkel J, Lazarus BD, Izaurralde E, Hanover JA, Conti E (2004) The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin alpha. Nat Struct Mol Biol 11(10):1001–1007
18. Kajava AV (2002) What curves alpha-solenoids? Evidence for an alpha-helical toroid structure of Rpn1 and Rpn2 proteins of the 26 S proteasome. J Biol Chem 277(51):49791–49798
19. King RW, Peters JM, Tugendreich S, Rolfe M, Hieter P, Kirschner MW (1995) A 20 S complex containing CDC27 and CDC16 catalyzes the mitosis-specific conjugation of ubiquitin to cyclin B. Cell 81(2):279–288
20. Kobe B (1996) Leucines on a roll. Nat Struct Biol 3(12):977–980
21. Kobe B, Deisenhofer J (1993) Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. Nature 366(6457):751–756
22. Lamb JR, Michaud WA, Sikorski RS, Hieter PA (1994) Cdc16p, Cdc23p And Cdc27p form a complex essential for mitosis. EMBO J 13(18):4321–4328
23. Lee SJ, Matsuura Y, Liu SM, Stewart M (2005) Structural basis for nuclear import complex dissociation by RanGTP. Nature 435(7042):693–696
24. Lupas A, Baumeister W, Hofmann K (1997) A repetitive sequence in subunits of the 26 S proteasome and 20 S cyclosome (anaphase-promoting complex). Trends Biochem Sci 22(6):195–196
25. Main ER, Xiong Y, Cocco MJ, D'Andrea L, Regan L (2003) Design of stable alpha-helical arrays from an idealized TPR motif. Structure 11(5):497–508
26. Matsuura Y, Stewart M (2004) Structural basis for the assembly of a nuclear export complex. Nature 432(7019):872–877
27. Peters JM (2006) The anaphase promoting complex/cyclosome: a machine designed to destroy. Nat Rev Mol Cell Biol 7(9):644–656

28. Pines J (2006) Mitosis: a matter of getting rid of the right protein at the right time. Trends Cell Biol 16(1):55–63
29. Russo AA, Tong L, Lee JO, Jeffrey PD, Pavletich NP (1998) Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumour suppressor p16INK4a. Nature 395(6699):237–243
30. Shi Y (2009) Serine/threonine phosphatases: mechanism through structure. Cell 139(3):468–484
31. Sibanda BL, Chirgadze DY, Blundell TL (2010) Crystal structure of DNA-PKcs reveals a large open-ring cradle comprised of HEAT repeats. Nature 463(7277):118–121
32. Sikorski RS, Boguski MS, Goebl M, Hieter P (1990) A repeating amino acid motif in CDC23 defines a family of proteins and a new relationship among genes required for mitosis and RNA synthesis. Cell 60(2):307–317
33. Sudakin V, Ganoth D, Dahan A, Heller H, Hershko J, Luca FC, Ruderman JV, Hershko A (1995) The cyclosome, a large complex containing cyclin-selective ubiquitin ligase activity, targets cyclins for destruction at the end of mitosis. Mol Biol Cell 6(2):185–197
34. Sullivan M, Morgan DO (2007) Finishing mitosis, one step at a time. Nat Rev Mol Cell Biol 8(11):894–903
35. Thornton BR, Toczyski DP (2006) Precise destruction: an emerging picture of the APC. Genes Dev 20(22):3069–3078
36. Tugendreich S, Tomkiel J, Earnshaw W, Hieter P (1995) CDC27Hs colocalizes with CDC16Hs to the centrosome and mitotic spindle and is essential for the metaphase to anaphase transition. Cell 81(2):261–268
37. Vetter IR, Arndt A, Kutay U, Gorlich D, Wittinghofer A (1999) Structural view of the Ran-importin beta interaction at 2.3 A resolution. Cell 97(5):635–646

Chapter 4
Cryoelectron Tomography or Doing Structural Biology In Situ

Wolfgang Baumeister

Abstract Electron tomography enables the three-dimensional visualization of large and stochastically variable structures such as supramolecular assemblies, organelles or even cells. In conjunction with cryogenic techniques electron tomography avoids the artefacts that are notorious to conventional electron microscopy specimen preparation. At resolutions of a few (2–4) nanometers it provides unprecedented insights into the molecular organization of cellular landscapes and helps to bridge the divide that hitherto existed between molecular and cellular structural studies.

4.1 Principles and Limitations of Tomography

Electron tomography, like other tomography modalities, relies on the principle of recording images from different viewing angles, aligning the resulting projections to a common coordinate system and combining them computationally to form a 3D image. The practical realization of cryoelectron tomography had to reconcile two conflicting requirements: detailed reconstructions with minimal distortions require a large number of images covering as wide a tilt range as possible, yet, biological materials embedded in vitreous ice should not be exposed to high cumulative electron doses. The allowable total electron dose has to be fractionated over a maximal number of projections with the aid of automated data acquisition methods. Microscope control software for tomographic data acquisition incorporates the compensation of image shifts and focus changes resulting from the imperfect eucentricity of the tilting devices. The signal of the individual images must suffice to enable the precise tracking of the specimen during the acquisition of the tomograms [1, 2].

W. Baumeister (✉)
Max-Planck-Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany
e-mail: baumeist@biochem.mpg.de

M.A. Carrondo and P. Spadon (eds.), *Macromolecular Crystallography*,
NATO Science for Peace and Security Series A: Chemistry and Biology,
DOI 10.1007/978-94-007-2530-0_4, © Springer Science+Business Media B.V. 2012

The theoretical resolution, d, of a tomographic reconstruction from a cylindrical volume of diameter, D, generated from N equally spaced projections covering the full angular tilt range of ±90° can be described broadly by the relationship $d \sim \pi D/N$ [3]. Yet, the concept of resolution in electron tomography is not trivial and it is meaningless if one does not consider structural preservation. In addition to the limited number of projections, specimen geometry typically restricts the angular range to ±70° (see later). According to the projection theorem, the non-sampled region (here, ±20°), defines a 'missing wedge' in reciprocal space. The consequence is anisotropic resolution, giving rise to distortions in the reconstructed volume. More elaborate tilt geometries fill some of the missing wedge and, in the case of dual-axis tilting about orthogonal axes, the wedge is reduced to a 'pyramid' and the resolution becomes more isotropic. For a tilt range of ±45°, a single-axis tilt scheme samples only 50% of the information, whereas dual-axis tilting over this range samples 67% of the information. For the more optimistic tilt range of ±70°, the sampling completeness rises from 78% (single-axis) to 93% (dual-axis) [2].

The 'slab' geometry typical of vitreous thin films or sections results in a progressive increase in sample thickness at higher tilt angles: for a specimen with uniform thickness of 200 nm, tilting to 70° results in an effective pathlength of almost 600 nm. The mean free path for 300 keV electrons in ice is ~350 nm, implying that longer path lengths will be increasingly dominated by inelastic scattering events. The resultant blurring and degradation in image contrast can be alleviated with an energy filteroperating in 'zero-loss' mode. Of course, energy filters cannot alter the proportion of elastic to inelastic scattering events.

4.2 Molecular Interpretation of Tomograms

Cryoelectron tomograms of organelles and cells contain vast amounts of information that extends beyond cellular ultrastructure. Essentially, they are 3D representations of the entire proteome and they are snapshots of the interaction networks underlying cellular functions. However, retrieving this information is not a trivial task because the signal-to-noise ratio of the tomograms is low and individual macromolecules are difficult to recognize in an environment that is so crowded that they literally touch each other. There are two alternative (although not mutually exclusive) approaches for mapping macromolecules in cellular environments: (1) specific labelling with electron-densemarkers; or (2) computational strategies based on innate structural signatures and pattern recognition. Strategies based on labelling cannot detect more than a tiny fraction of the proteome simultaneously in any given cell. Computational methods, however, enable one to interrogate and interpret tomograms in a comprehensive manner. Methods based on pattern recognition are more demanding in terms of resolution and they require *a priori* knowledge of the macromolecular structures under scrutiny [4, 5].

The direct visualization of individual macromolecular complexes in the cytoplasm of an intact *Dictyosteliumdiscoideum* cell with a resolution good enough to identify them by their size and shape indicates that structural signature-based mapping could become a viable alternative to labelling [6]. The advantages are self-evident: a computational approach using pattern recognition methods is non-invasive and has the potential to map molecular landscapes inside cells in a comprehensive(i.e. proteome-wide) manner. Here, the task is to identify and locate *a priori* known structures (templates)in a cellular tomogram. One prerequisite is that the tomogram represents the spatial modulation of mass density inside the unadulterated cell and with good enough resolution, a requirement that can only be fulfilled by cryoelectrontomograms [7].

4.3 Correlative Imaging with Spatial and Temporal Resolution

The judicious use of multiple imaging techniques can provide complementary information concerning cell structure and function. These techniques should not only span several orders of magnitude in spatial resolution but they should also enable one to monitor cellular processes and capture them at crucial points in time. Vitrification provides a 'snapshot' with a temporal resolution in the range of milliseconds. An integrated approach must, therefore, be developed to observe specific cellular structures and processes in real time, using live cell imaging techniques, to subsequently immobilize such events without delay by vitrification and to record high-resolution information from precisely preselected sites. Recently, cryo- fluorescence microscopy has been developed as an adjunct to cryo- ET. Fluorescence microscopy can thus be exploited to navigate the cellular landscape under cryogenic conditions and to identify features of interest before zooming in on the area using the cryoelectron microscope. The fluorescence signal offers an independent and unambiguous confirmation of the identityof the feature and it provides an alternative to the tedious search for low-copy-number structures using the electron beam [8].

4.4 Overcoming the Specimen Thickness Problem

A critical limitation in cryoelectron tomography is specimen thickness. Only prokaryotic cells or peripheral regions or appendages of eukaryotic cells are thin enough to be investigated in their entirety. Larger cells or tissues must be sectioned prior to examining them in the EM. Cryosectioning with maintenance of the vitreous state has still more of an art than science. Mechanically generated slices of ice-embedded cells inevitably suffer from mechanical distortions. An alternative micromachining method for frozen-hydrated material is the emerging use of focused ion beam technology. First encouraging results using this technology will be shown.

4.5 Practical Applications

A number of examples will be presented which demonstrate the potential of cryoelectron tomography: (1) Studies of the nuclear pore complex in situ i.e. in intact nuclei from *Dictyostelium* [9, 10]. (2) Studies of the actin cytoskeleton in *Dictyostelium* [11]. (3) Studies of a variety of ribosomal configurations (polyribosomes, 100 S 'hibernating' ribosomes) both *in vitro* and *in situ* [12]. (4) Studies of the presynaptic cytomatrix [13].

References

1. Leis A, Rockel B, Andrees L, Baumeister W (2009) Visualizing cells at the nanoscale. Trends Biochem Sci 34:60–70
2. Lucic V, Förster F, Baumeister W (2005) Structural studies by electron tomography: from cells to molecules. Annu Rev Biochem 74:833–865
3. Crowther RA, DeRosier DJ, Klug A (1970) The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. Proc R Soc Lond A Math Phys Sci 317:319–340
4. Mercogliano CP, DeRosier DJ (2006) Gold nanocluster formation using metallothionein: mass spectrometry and electron microscopy. J Mol Biol 355:211–223
5. Frangakis AS, Böhm J, Förster F, Nickell S, Nicastro D, Typke D, Hegerl R, Baumeister W (2002) Identification of macromolecular complexes in cryoelectron tomograms of phantom cells. Proc Natl Acad Sci USA 99:14153–14158
6. Medalia O, Weber I, Frangakis AS, Nicastro D, Gerisch G, Baumeister W (2002) Macromolecular architecture in eukaryotic cells visualized by cryoelectron tomography. Science 298:1209–1213
7. Nickell S, Kofler C, Leis A, Baumeister W (2006) A visual approach to proteomics. Nat Rev Mol Cell Biol 7:225–230
8. Sartori A, Gatz R, Beck F, Rigort A, Baumeister W, Plitzko JM (2007) Correlative microscopy: bridging the gap between fluorescence light microscopy and cryo-electron tomography. J Struct Biol 160:135–145
9. Beck M, Förster F, Ecke M, Plitzko J, Melchior F, Gerisch G, Baumeister W, Medalia O (2004) Nuclear pore complex structure and dynamics revealed by cryoelectron tomography. Science 306:1387–1390
10. Beck M, Lucic V, Förster F, Baumeister W, Medalia O (2007) Snapshots of nuclear pore complexes in action captured by cryoelectron tomography. Nature 449:611–615
11. Medalia O, Beck M, Ecke M, Weber I, Neujahr R, Baumeister W, Gerisch G (2007) Organization of actin networks in intact filopodia. Curr Biol 17:79–84
12. Brandt F, Etchells SA, Ortiz JO, Elcock AH, Hartl FU, Baumeister W (2009) The native 3D organization of bacterial polysomes. Cell 136:261–271
13. Fernandez-Busnadiego R, Zuber B, Maurer UE, Cyrklaff M, Baumeister W, Lucic V (2010) Quantitative analysis of the native presynaptic cytomatrix by cryo-electron tomography. J Cell Biol 188:145–156

Chapter 5
RuvBL1 and RuvBL2 and Their Complex Proteins Implicated in Many Cellular Pathways

Sabine Gorynia*, Tiago M. Bandeiras, Pedro M. Matias, Filipa G. Pinho, Colin E. McVey, Peter Donner, and Maria Arménia Carrondo

Abstract RuvBL1 and its homolog RuvBL2 belong to the AAA$^+$ family of ATPases and play important roles in chromatin remodeling, in transcriptional regulation, in DNA repair and in the c-Myc and Wnt signaling pathways. Proteins involved in these pathways are often mutated in human cancers. Both RuvBL proteins form a complex and act alone or together in diverse cellular processes. The three-dimensional structures of human RuvBL1 refined using diffraction data to 2.2 Å resolution and of the human RuvBL1/RuvBL2 complex with a truncated domain II at 3 Å resolution are presented. The dodecameric RuvBL1/RuvBL2 complex structure differs from previously described models. It consists of two heterohexameric rings with alternating RuvBL1 and RuvBL2 monomers that interact with each other via domain II. ATPase and helicase activities of RuvBL1 and RuvBL2 were also tested. Interestingly, truncation of domain II resulted not only in a substantial increase of ATP consumption by the RuvBL proteins, but also in stimulation of helicase activity, which was not observed with the full-length proteins.

*Sabine Gorynia and Tiago M. Bandeiras contributed equally to this work.

S. Gorynia
Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa,
Apartado 127, 2781-901 Oeiras, Portugal

Lead Discovery Berlin – Protein Supply, Bayer Schering Pharma AG, 13353 Berlin, Germany

Department of Biological Chemistry, David Geffen School of Medicine, UCLA,
615 Charles E. Young Drive South, Box 951737, Los Angeles, CA 90095-1737, USA

T.M. Bandeiras • F.G. Pinho
Instituto de Biologia Experimental e Tecnológica, Apartado 12, 2781-901 Oeiras, Portugal

P.M. Matias • C.E. McVey • P. Donner • M.A. Carrondo (✉)
Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa,
Apartado 127, 2781-901 Oeiras, Portugal
e-mail: carrondo@itqb.unl.pt

M.A. Carrondo and P. Spadon (eds.), *Macromolecular Crystallography*,
NATO Science for Peace and Security Series A: Chemistry and Biology,
DOI 10.1007/978-94-007-2530-0_5, © Springer Science+Business Media B.V. 2012

Keywords AAA+ proteins • Helicase • X-ray crystallography • Chromatin remodeling • Transcriptional regulation

5.1 RuvBL1 and RuvBL2 Are Highly Conserved AAA+ Proteins

RuvBL1 and its homolog RuvBL2 are ubiquitously expressed proteins [1] that belong to the AAA+ family of ATPases (ATPases associated with diverse cellular activities) [2]. This class of ATPases includes nucleic acid processing enzymes, chaperones and proteases. AAA+ proteins often form hexameric ring structures and contain conserved motifs for ATP binding and hydrolysis such as the Walker A and Walker B boxes [3], the Arg-finger and sensor residues. All AAA+ proteins use ATP binding and hydrolysis to exert mechanical forces. ATP hydrolysis is clearly essential for the biological activity of RuvBL1 and RuvBL2 [4–6]. RuvBL1 and RuvBL2 share an homology of *ca.* 30% with the bacterial DNA-dependent ATPase and helicase RuvB [7, 8], which is the motor that drives branch migration of the Holliday junction in the presence of RuvA and RuvC during homologous recombination and recombinational repair of damaged DNA [9]. The ruvA, ruvB and ruvC genes are required for normal levels of cellular resistance to the effects of UV- or ionizing radiation [10]. Helicases, like RuvB, are molecular motor proteins which couple the energy of ATP hydrolysis to unwinding of the energetically stable duplex form of DNA or RNA and translocate along the nucleic acid in an ATP-dependent manner.

5.2 RuvBL1 and RuvBL2 Are Components of Chromatin Remodeling Complexes

RuvBL1 and RuvBL2 were found to be involved in chromatin remodeling by several groups. A fundamental regulatory step in transcription and other DNA-dependent processes in eukaryotes is the control of chromatin structure, which regulates access of proteins to DNA. In the eukaryotic nucleus, DNA is wrapped around an octamer of four core histones in approximately two superhelical turns to form the nucleosome, and arrays of nucleosomes are successively folded into higher-order structures that collectively define chromatin. Packaging of genes into chromatin represses basal transcription and several multisubunit complexes are needed to regulate gene expression by modulating the topology of the nucleosomes in a number of ways.

RuvBL1 and RuvBL2 were found to be part of diverse chromatin remodeling complexes. They are present in two related complexes containing p400, the p400 complex and the TIP60 complex. These complexes perform critical functions in a variety of cellular processes including transcriptional activation, as well as break

repair and apoptosis of double stranded DNA [11–13]. They display ATPase and helicase activities. It was shown that these functions are, at least in part, contributed by RuvBL1 and RuvBL2 [12].

Both proteins are also components of the INO80 complex which exists in yeast and higher eukaryotes. It catalyzes ATP-dependent sliding of nucleosomes along DNA and is involved in the repair of DNA double strand breaks and in transcriptional regulation [5, 14–16]. It was shown that RuvBL1 and RuvBL2 are essential for the structural and functional integrity of the INO80 chromatin remodeling complex [5]. RuvBL1 and RuvBL2 bound to ATP are in the correct conformation to associate with the INO80 complex and initiate the recruitment of the essential actin-like Arp5 subunit assembling the complete functional chromatin remodeling complex.

5.3 RuvBL1 and RuvBL2 Are Involved in Transcription

RuvBL1 and RuvBL2 regulate transcription not only via association with chromatin remodeling complexes, but also through interactions with diverse transcription factors and the RNA polymerase II holoenzyme complex. First, RuvBL1 and RuvBL2 were found to interact with the TATA-binding protein [17, 18] and the large RNA polymerase II holoenzyme complex [19], which contains over 50 components and is responsible for the transcription of protein-encoding genes. Later, RuvBL1/2 were also identified by their physical interaction with the transcription-associated protein β-catenin [1, 20], and with the transcription factors c-Myc [6], E2F1 (only RuvBL1 [21]) and ATF2 (only RuvBL2 [22]). Since then, the mammalian homologs have been implicated in at least two oncogenic pathways, one involving c-Myc and the other β-catenin. Among the transcription factors with oncogenic potential, c-Myc is one of the most frequent sites of mutation in human cancer [23]. The N-terminal portion of c-Myc contains two highly conserved regions, called Myc homology box I (MbI) and Myc homology box II (MbII). The MbII domain is necessary for virtually all c-Myc biological activities, including oncogenic transformation, apoptosis and the ability to block differentiation and stimulate cell proliferation [24–27]. This region was shown to bind to RuvBL1 and RuvBL2 [6]. A missense mutation in the RuvBL1 ATPase motif acts as a dominant inhibitor of c-Myc oncogenic activity but does not inhibit normal cell growth, indicating that functional RuvBL1 is an essential mediator of c-Myc oncogenic transformation [6, 21].

5.4 Crystal Structure and Activities of Human RuvBL1

The crystal structure of RuvBL1 was solved from data at 2.2 Å resolution, showing an overall hexameric molecule with a central channel of approximate diameter of 20 Å, where each monomer is complexed with one ADP molecule [28].

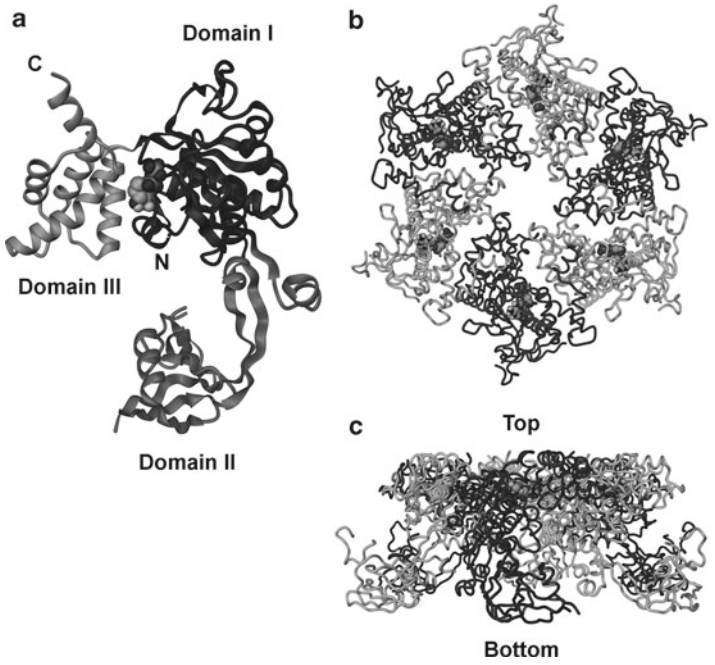


Fig. 5.1 Views of the RuvBL1 3D structure. (**a**) Ribbon diagram highlighting the domain structure; (**b**) Top view of the RuvBL1 hexamer; (**c**) Side view of the RuvBL1 hexamer. In (**b**) and (**c**) the C^{α} chain is represented in tube mode and the monomers are colored *light* and *dark gray* for clarity. In all views, the ADP molecules are drawn in space-filling mode

The monomers contain three domains, of which the first and the third are involved in ATP binding and hydrolysis. Structural analysis allowed the identification of four conserved motifs (Walker A, Walker B, sensor 1 from domain I and sensor 2 from domain III) in RuvBL1, likely to be important in nucleotide-driven conformational changes of the protein structure.

Structural homology and DNA-binding studies suggested that domain II, which is unique among AAA^{+} proteins and not present in the bacterial homolog RuvB, is a novel DNA/RNA-binding domain. The interface between adjacent subunits in the hexamer of RuvBL1 is made up entirely by domain I and domain III. The nucleotide binding pocket is located in this interface, and the hexamerization process seems to block access to this pocket, thus making impossible an exchange from ADP to ATP (Fig. 5.1).

Arg-357 from each monomer is near this interface and sufficiently close to the nucleotide binding site in the adjacent monomer to be able to act as an Arg finger, provided a suitable conformation change takes place.

We were able to demonstrate that RuvBL1 interacted with ssDNA/RNA and dsDNA. Because the central channel seems to be too small to accommodate dsDNA, we assumed that a region outside of the ring makes DNA contacts. Based on this

assumption and the similarity between domain II and DNA binding domains of other proteins we performed electrophoretic mobility shift assays experiments which showed that the new domain II acts as a nucleic acid binding domain.

Although it has been shown that ATPase activity of RuvBL1 is needed for several *in vivo* functions, we could only detect a marginal activity with the purified protein. However, the structure of the RuvBL1-ADP complex suggests that RuvBL1 has all the structural characteristics of a molecular motor, even of an ATP-driven helicase.

5.5 Crystal Structure of the Human RuvBL1/RuvBL2 Complex

The purified wild-type complex of RuvBL1 and RuvBL2 was used for crystallization trials in order to solve its three dimensional structure. Although thousands of conditions were tested, the wild-type complex never crystallized. Thus, deletion mutants of RuvBL1 and RuvBL2 with partially truncated domains II were generated for crystallization purposes [29].

The crystal structure of the RuvBL1ΔDII/RuvBL2ΔDII (R1ΔDII/R2ΔDII) complex was solved from a selenomethionine derivative at 3 Å resolution [30] by a combination of both molecular replacement and MAD methods. It is a double-heterohexameric ring structure with alternating R1ΔDII and R2ΔDII monomers, forming a dodecamer. The central channel (C^{α} – C^{α} distance) in the dodecameric R1ΔDII/R2ΔDII complex structure has an internal diameter similar to that of the RuvBL1 hexamer at both terminals but it is much wider in the central part. Contrary to what might be expected, the two heterohexameric rings in the R1ΔDII/R2ΔDII dodecamer do not interact through the flat ring surfaces, but rather via the retained section of the truncated domains II.

These results differ from those previously reported: our dodecamer is symmetrical as opposed to that proposed for the human [31] and for the yeast complexes [32], even though the hexamer arrangement in the dodecamer is similar. The dodecameric structure of the human RuvBL1/RuvBL2 complex was examined by negative stain electron microscopy at 20 Å resolution, and substantial differences were found between the top and bottom rings [31]. On the other hand, our heterohexameric arrangement of RuvBL1 and RuvBL2 monomers agrees with the yeast Rvb1/Rvb2 complex structure solved by EM [33], but our structure is clearly a dodecamer while they proposed isolated hexamers. In agreement with our results, the dodecameric yeast Rvb1/Rvb2 cryo-EM structure clearly shows that domain II constitutes the interaction site between the two hexameric rings [32]. However, their results suggest that each ring is composed of just one of the proteins, forming homo-oligomeric hexamers, whereas our crystallographic analysis shows two hetero-hexameric rings related by a crystallographic twofold axis forming the dodecamer.

Each of the R1ΔDII and R2ΔDII monomers in the dodecamer is complexed with a mixture of ADP and ATP molecule. The electron density for the the γ-phosphate is clearly weaker than for the other two phosphate groups, suggesting that it may

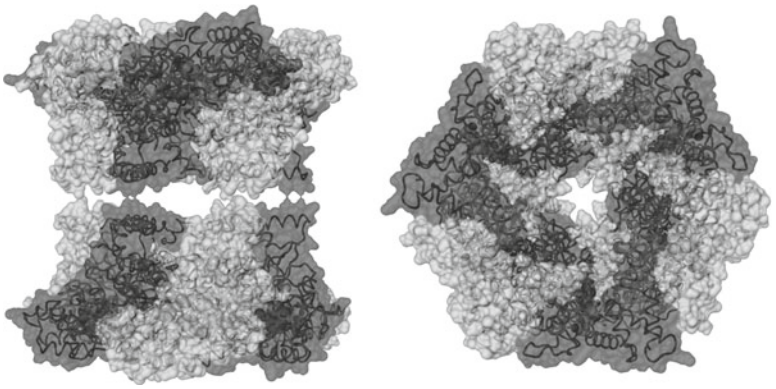


Fig. 5.2 Views of the RuvBL1ΔDII/RuvBL2ΔDII dodecamer – tube C^{α} diagrams with superimposed molecular surface. RuvBL1ΔDII monomers are *light-colored*, RuvBL2ΔDII monomers are *dark-colored*, nucleotide molecules are drawn in space-filling mode

have been partially hydrolized to ADP. For each monomer in the structure, the γ-phosphate is interacting with an aspartate residue in the same monomer located in the Walker B motif, essential for ATP hydrolysis, an arginine residue in the same monomer belonging to the sensor 2 region, and an aspartate from an adjacent monomer which precedes an arginine residue.

Given that both hexameric and dodecameric structures of the RuvBL1/RuvBL2 complex were reported as described above, and that RuvBL1 alone crystallized as a hexamer in its ADP-bound form [28], one can hypothesize that both hexamers and dodecamers of the complex exist *in vivo*. Once it becomes active, this AAA^{+} machine may switch between its hexameric and dodecameric state (Fig. 5.2).

Our crystallographic results suggest that there is a tight interaction between the truncated domains II of RuvBL1 and RuvBL2 forming a dodecameric structure. In general, AAA^{+} proteins undergo conformational changes upon hydrolysis of ATP to ADP. These changes may loosen the interactions between domains II of the two hexamers resulting in either a dodecamer with domains II in a different conformation, or two separate hexameric rings with bound ADP (as seen in the RuvBL1 crystal structure).

Therefore, the change between two different dodecameric conformations or between hexameric and dodecameric structures of the RuvBL1/RuvBL2 complex may be important for activity regulation.

5.6 Activity of the Human RuvBL1/RuvBL2 Complex

We only detected a weak ATPase activity of purified human RuvBL1 and RuvBL2. Important cellular processes of the higher eukaryotes, such as those dealing with DNA metabolism, are often regulated by large multi-protein complexes and it is

therefore likely that strong RuvBL1 or RuvBL2 enzymatic activity can only be seen in such an environment. Importantly, the wild-type RuvBL1/RuvBL2 complex exhibited a threefold to fourfold increase in ATP consumption compared to the single proteins. Surprisingly, the RuvBL1/RuvBL2 complexes with truncated domains II exhibited higher ATPase activity than the wild-type complex. Assembly of RuvBL1 and RuvBL2 into a dodecameric structure may have stimulated the ATPase activity. This synergistic effect on enzymatic activity clearly suggests communication between the two proteins coupled to conformational effects.

These findings, combined with our crystallographic results, indicate that, in vivo, domain II and dodecamerization may play a regulatory role to control ATP consumption. Given that the ATPase activity of RuvBL1 and RuvBL2 is needed for several in vivo functions [4, 6, 34], it can be speculated that cofactors regulate the ADP/ATP exchange by changing the position of domain II, thereby clearing the way to the nucleotide-binding pocket for a more efficient ADP/ATP exchange. As a result, the ATPase activity of RuvBL1 and RuvBL2 could be stimulated.

Even though RuvBL1 and RuvBL2 are the human homologs of the bacterial helicase RuvB, we and other groups did not detect helicase activity for recombinant human RuvBL1 and RuvBL2 [11, 19, 28]. Since the DII-truncated complex of RuvBL1 and RuvBL2 exhibited an unexpected increase in ATPase activity, it was possible that higher ATP consumption might allow the complex to exert helicase activity. Indeed, we observed helicase activity of the truncated constructs R1ΔDII, R2ΔDII, R1wt/R2ΔDII and R1ΔDII/R2ΔDII. Since the wild-type proteins exhibited no helicase activity in vitro, it can be concluded that cofactors binding to RuvBL1 and RuvBL2 in chromatin remodeling or transcription complexes alter the conformation of both proteins, most likely via domain II, and allow them to exert their helicase activity.

Small-angle X-ray scattering studies were performed with the wild-type single proteins and the wild-type and truncated complexes [30]. These studies reveal differences between the oligomerization states of RuvBL1 and RuvBL2 in solution. RuvBL2 forms mainly hexamers and dodecamers independently of its concentration, while RuvBL1 exists as a monomer at lower concentration and forms a hexamer at higher concentration, but is unable to form dodecamers. This diversity in oligomerization characteristics may explain why RuvBL1 and RuvBL2 are recruited into different complexes for their specialized functions. The results also show that truncation of domain II results in a complete dodecamerization of the RuvBL1/RuvBL2 complex, while the wild-type complex consists of a mixture of hexamers and dodecamers in solution, indicating that domain II is involved in dodecamerization and supporting the hypothesis that both oligomers could co-exist *in vivo*.

A detailed description and extensive discussion of the structural results as well as the SAXS studies will be presented elsewhere [30]. Correlations between the truncated dodecameric structure and the wild-type complex will be proposed along with a possible mechanism of how the highly conserved proteins RuvBL1 and RuvBL2 might exert their activities, a matter of long discussion in the literature.

References

1. Bauer A, Huber O, Kemler R (1998) Pontin52, an interaction partner of beta-catenin, binds to the TATA box binding protein. Proc Natl Acad Sci USA 95(25):14787–14792
2. Neuwald AF et al (1999) AAA+: a class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. Genome Res 9(1):27–43
3. Walker JE et al (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J 1(8):945–951
4. Feng Y, Lee N, Fearon ER (2003) TIP49 regulates beta-catenin-mediated neoplastic transformation and T-cell factor target gene induction via effects on chromatin remodeling. Cancer Res 63(24):8726–8734
5. Jonsson ZO et al (2004) Rvb1p/Rvb2p recruit Arp5p and assemble a functional Ino80 chromatin remodeling complex. Mol Cell 16(3):465–477
6. Wood MA, McMahon SB, Cole MD (2000) An ATPase/helicase complex is an essential cofactor for oncogenic transformation by c-Myc. Mol Cell 5(2):321–330
7. Putnam CD et al (2001) Structure and mechanism of the RuvB Holliday junction branch migration motor. J Mol Biol 311(2):297–310
8. Yamada K et al (2001) Crystal structure of the Holliday junction migration motor protein RuvB from Thermus thermophilus HB8. Proc Natl Acad Sci USA 98(4):1442–1447
9. Tsaneva IR, Muller B, West SC (1993) RuvA and RuvB proteins of Escherichia coli exhibit DNA helicase activity in vitro. Proc Natl Acad Sci USA 90(4):1315–1319
10. Mezard C et al (1999) Escherichia coli RuvBL268S: a mutant RuvB protein that exhibits wild-type activities in vitro but confers a UV-sensitive ruv phenotype in vivo. Nucleic Acids Res 27(5):1275–1282
11. Ikura T et al (2000) Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. Cell 102(4):463–473
12. Fuchs M et al (2001) The p400 complex is an essential E1A transformation target. Cell 106(3):297–307
13. Samuelson AV et al (2005) p400 is required for E1A to promote apoptosis. J Biol Chem 280(23):21915–21923
14. Jin J et al (2005) In and out: histone variant exchange in chromatin. Trends Biochem Sci 30(12):680–687
15. Jonsson ZO et al (2001) Rvb1p and Rvb2p are essential components of a chromatin remodeling complex that regulates transcription of over 5% of yeast genes. J Biol Chem 276(19):16279–16288
16. Shen X et al (2000) A chromatin remodelling complex involved in transcription and DNA processing. Nature 406(6795):541–544
17. Kanemaki M et al (1997) Molecular cloning of a rat 49-kDa TBP-interacting protein (TIP49) that is highly homologous to the bacterial RuvB. Biochem Biophys Res Commun 235(1):64–68
18. Kanemaki M et al (1999) TIP49b, a new RuvB-like DNA helicase, is included in a complex together with another RuvB-like DNA helicase, TIP49a. J Biol Chem 274(32):22437–22444
19. Qiu XB et al (1998) An eukaryotic RuvB-like protein (RUVBL1) essential for growth. J Biol Chem 273(43):27786–27793
20. Bauer A et al (2000) Pontin52 and reptin52 function as antagonistic regulators of beta-catenin signalling activity. EMBO J 19(22):6121–6130
21. Dugan KA, Wood MA, Cole MD (2002) TIP49, but not TRRAP, modulates c-Myc and E2F1 dependent apoptosis. Oncogene 21(38):5835–5843
22. Cho SG et al (2001) TIP49b, a regulator of activating transcription factor 2 response to stress and DNA damage. Mol Cell Biol 21(24):8398–8413
23. Cole MD (1986) The myc oncogene: its role in transformation and differentiation. Annu Rev Genet 20:361–384

24. Evan GI et al (1992) Induction of apoptosis in fibroblasts by c-myc protein. Cell 69(1):119–128
25. Li LH et al (1994) c-Myc represses transcription in vivo by a novel mechanism dependent on the initiator element and Myc box II. EMBO J 13(17):4070–4079
26. Penn LJ et al (1990) Negative autoregulation of c-myc transcription. EMBO J 9(4):1113–1121
27. Stone J et al (1987) Definition of regions in human c-myc that are involved in transformation and nuclear localization. Mol Cell Biol 7(5):1697–1709
28. Matias PM et al (2006) Crystal structure of the human AAA+ protein RuvBL1. J Biol Chem 281(50):38918–38929
29. Gorynia S et al (2008) Cloning, expression, purification, crystallization and preliminary X-ray analysis of the human RuvBL1-RuvBL2 complex. Acta Crystallogr Sect F Struct Biol Cryst Commun 64(Pt 9):840–846
30. Gorynia S et al (2010) Structural and functional insights into the dodecameric molecular machine – The RuvBL1/RuvBL2 complex. J Struct Biol 176 (2011), 279–291
31. Puri T et al (2007) Dodecameric structure and ATPase activity of the human TIP48/TIP49 complex. J Mol Biol 366(1):179–192
32. Torreira E et al (2008) Architecture of the pontin/reptin complex, essential in the assembly of several macromolecular complexes. Structure 16(10):1511–1520
33. Gribun A et al (2008) Yeast Rvb1 and Rvb2 are ATP-dependent DNA helicases that form a heterohexameric complex. J Mol Biol 376(5):1320–1333
34. Diop SB et al (2008) Reptin and pontin function antagonistically with PcG and TrxG complexes to mediate Hox gene control. EMBO Rep 9(3):260–266

Chapter 6
The Structural Biology of Muscle: Spatial and Temporal Aspects

Kenneth C. Holmes

Abstract Understanding muscle contraction has resulted from the synergy of a number of approaches for which structure has provided an integrating framework. Nearly 60 years ago interference and phase contrast light microscopy established the sliding filament model of muscle contraction. A little later H.E. Huxley exploited electron microscopy to visualize the macromolecular architecture of the sarcomere: the thick (myosin) and thin (actin) filaments with connecting myosin cross-bridges. These observations allowed him to outline a structural basis for muscle contraction: a rowing-like progression of the myosin cross-bridges along the actin filament. X-ray fibre diffraction from insect flight muscle first demonstrated that the cross-bridges could indeed take up two configurations that might represent the ends of an active stroke. Later intense X-ray synchrotron radiation allowed the recording of the movements of the cross-bridges during a contraction with high precision. In 1993 Rayments's group ushered in a much more detailed understanding of myosin function by solving the structure of the myosin cross-bridge by X-ray crystallography. It showed that the cross-bridge consists of a large catalytic domain, often called the motor domain, containing the ATP binding site and the actin binding site. At the C-terminus of the motor domain is a long lever arm. The catalytic mechanism is similar to the G-proteins: the active site contains a P-loop and switch 1 and switch 2 elements. The lever arm was later found in two different conformations (so called pre-power-stroke and post-rigor) showing how switch 2 movement is coupled to a swing of the lever arm. Further crystallographic studies coupled with high resolution em reconstructions of decorated actin (the rigor complex) showed how ATP binding sequesters switch 1 thereby opening the large cleft in the motor domain and breaking the strong binding to actin. Conversely, the strong binding to actin causes a movement of switch 1 with respect to the P-loop that destroys the nucleotide binding site, bringing about the release of ADP. There remains one unknown structure of seminal importance: the start the power stroke. During the cross-bridge cycle the cross-bridge in the pre-power

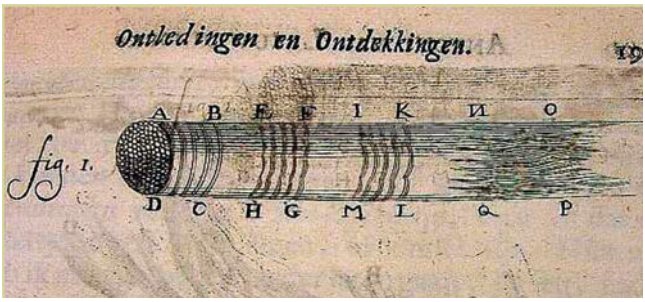
K.C. Holmes (✉)
Max Planck Institute for Medical Research, Heidelberg, Germany
e-mail: holmes@mpimf-heidelberg.mpg.de

M.A. Carrondo and P. Spadon (eds.), *Macromolecular Crystallography*,
NATO Science for Peace and Security Series A: Chemistry and Biology,
DOI 10.1007/978-94-007-2530-0_6, © Springer Science+Business Media B.V. 2012

stroke form loaded with ADP and phosphate rebinds to actin. The actin binding cleft must close without initially destroying the nucleotide binding site but in a way that enables phosphate release

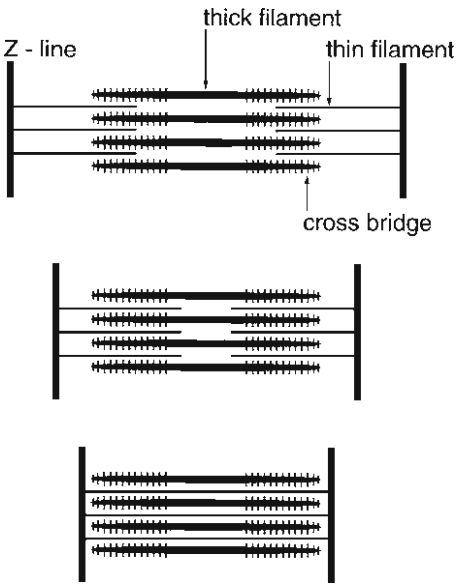
6.1 Early Microscopy Reveals Cross Striations in Muscle

In 1674 van Leeuwenhoek discovered the myofibrils and cross-striations in muscle fibres. His drawings showed the cross-striations which delineate the "globules" i.e., sarcomeres. In a typical 25 cm long human muscle, each muscle fibre may contain a hundred thousand sarcomeres arranged in series with each other.

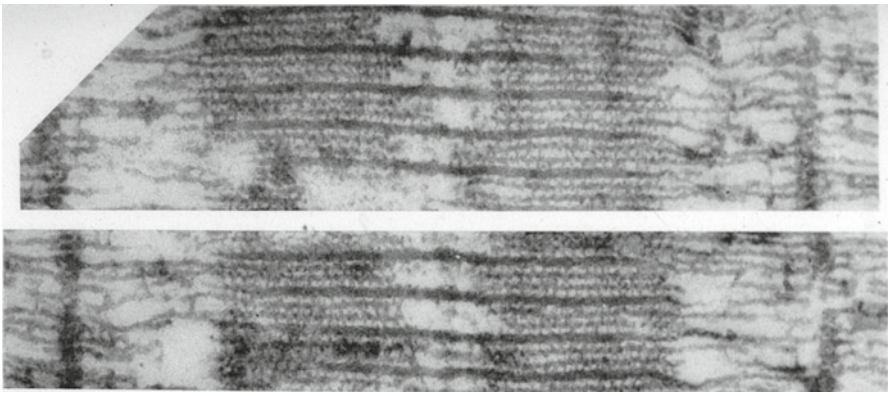


6.2 Sliding Filament Hypothesis

The sarcomeres shorten by sliding the thin and thick filaments past each other. The sliding filament hypothesis Huxley [12, 14].



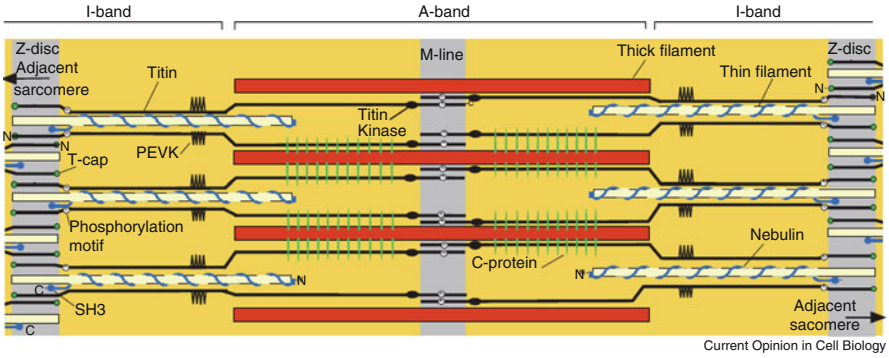
6.3 Cross Bridges (N-Terminal Domains of Myosin) Cause the Shortening



Very thin sections taken transversely show the cross-bridges that in rigor (absence of ATP) connect the thick and thin filaments [13]. Because of the geometry of the lattice and the section chosen two actin filaments can be seen between each of the thick myosin filaments.

6.4 The Sarcomere

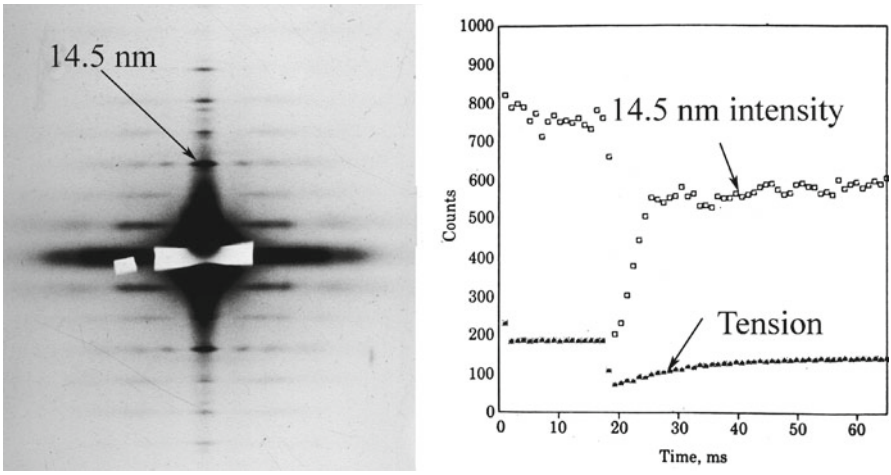
The thin filaments are made of actin: the thick filaments are made of myosin



A diagram of a skeletal muscle sarcomere showing the *A-bands* that contain myosin thick filaments and I-bands that contain actin thin filaments. Overlapping thin filaments from opposite sarcomeres are anchored in the Z-lines, whereas thick filaments

from opposite half-sarcometers are anchored in the M-lines. The two giant protein filament systems (titin and nebulin filaments) are also shown [8].

6.5 X-Ray Fibre Diffraction Patterns Can See the Cross-Bridges Move



Frog muscles give well ordered low angle X-ray diffraction patterns (fiber axis vertical). The 14.5 nm meridional peak arises from cross-bridges attached to actin. The muscles can be activated by electrical stimulation and the diffraction pattern monitored during a contraction. Using intense synchrotron X-ray sources the strength of the 14.5 nm peak can be monitored following a quick release of the muscle [15]. A large drop of intensity of the 14.5 nm meridional peak follows quick release. The initial drop in 14.5 nm intensity can be reversed if the release (or stretch) is reversed. These changes provide direct evidence that contraction is produced by swinging of attached cross-bridges. These measurements were repeated using a much more intense synchrotron X-ray source with higher time resolution [16, 17].

6.6 X-Ray Interference Gives the Cross-Bridge Positions with High Accuracy

The high spatial resolution available from synchrotron radiation X-ray sources shows that the 14.5 nm peak is actually split into two peaks. The splitting is an interference effect arising from fringes caused by the in-phase scattering of the cross-bridges in the two half sarcomeres. As these move towards each other during a contraction so the fringes move. By measuring the fringe positions one can get

very accurate values for the distance between the cross-bridges on both sides of the M-line [19, 26] and hence measure their movement.

6.7 Structure of Skeletal Muscle Myosin (Myosin II)

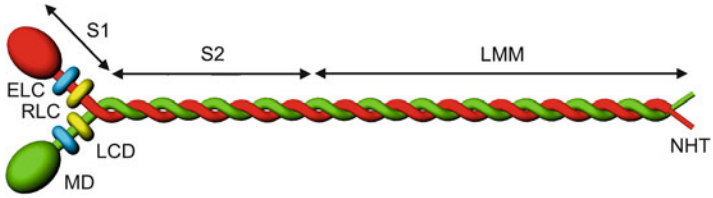
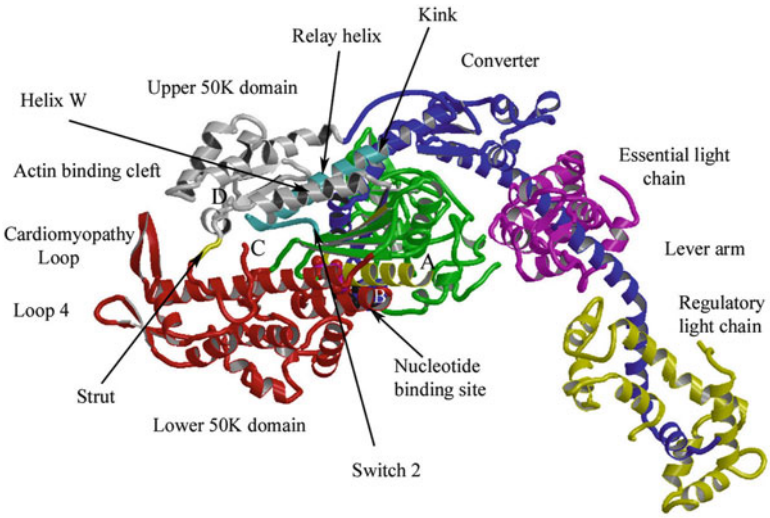


Diagram of myosin II (muscle myosin) [3] showing coiled-coil tail (consisting of C-terminal half of each heavy chain), and two heads, each comprising the N-terminal fraction of the heavy chain plus an essential light chain (ELC) and regulatory light chain (RLC). LCD, light chain domain (lever arm); MD, motor domain; NHT, non-helical tailpiece). The myosin molecule can be split into two fragments light meromyosin (LMM) and heavy meromyosin (the fragments S1 and S2) by proteolytic cleavage [27]. Heavy meromyosin is soluble at low salt concentrations, which makes it amenable for biochemical studies. Further proteolytic cleavage [21] results in subfragments 1 and 2 (S1 and S2). S1 contains the ATPase, the actin binding site and is identical with the head or morphological cross-bridge. S2 is an α-helical coiled-coil that links to the LMM, which is part of the thick filament There is a flexible region at the S2 LMM junction. The myosin molecules assemble into the thick filament. The first molecules assemble anti parallel and then grow by parallel assembly into bipolar filaments.

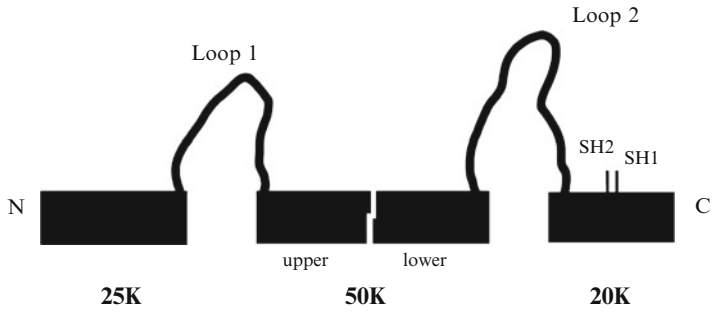
6.8 Structure of the Cross-Bridge (Myosin S1)

The post-rigor structure of the myosin motor domain [25]. The structure of the myosin cross-bridge shown as a ribbon diagram in the orientation it would take on binding to actin [24] viewed from the pointed (–) end of the actin filament. The N-terminus is shown green and the nucleotide binding P-loop and adjoining helix are shown yellow; the upper 50 K is red; the lower 50 K domain is grey. Note the cleft separating the upper and lower 50 K domains. The lower 50 K domain appears to be the primary actin-binding site. The N-terminal boundary of the upper 50 K domain comprises the disordered loop 1 (between the points marked A and B). The upper and lower 50 K domains are also connected by a disordered loop (loop 2 between C and D). The C-terminal long helix (dark blue) carries two calmodulin-like light chains and joins onto the thick filament. The C-terminal helix forms a

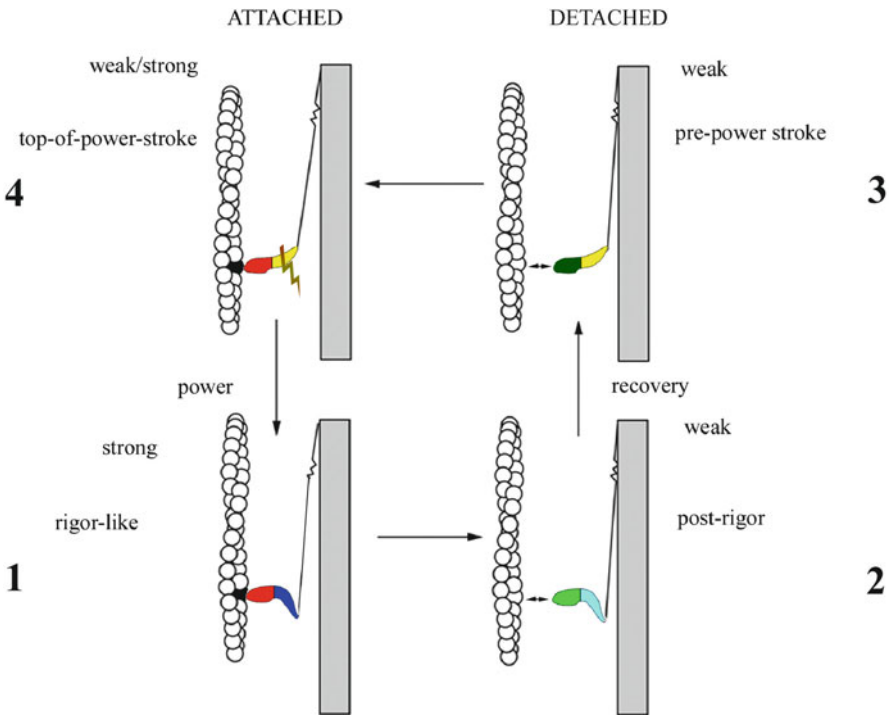
lever arm that rotates during contraction to move the actin filament past the myosin filament. The relay helix and converter domain are shown. In this conformation of the cross-bridge (post-rigor state) the lever arm is in the post-power stroke position or "DOWN", as in the rigor state. The colouring corresponds with sub-domain boundaries. The proximal end of the relay helix is shown light blue. The distal end (beyond the kink) is firmly attached to the converter domain. In the post-rigor structure the relay helix is straight (no kink). This figure is from [7].



Treatment with trypsin breaks myosin S1 into three "domains". The crystal structure shows that the "domains" are produced by cleavage of unstructured loops. The "50 K" domain is actually split into two sub-domains (called upper and lower) by a deep cleft extending from the actin binding site to the ATP binding site.



6.9 The Cross-Bridge Cycle



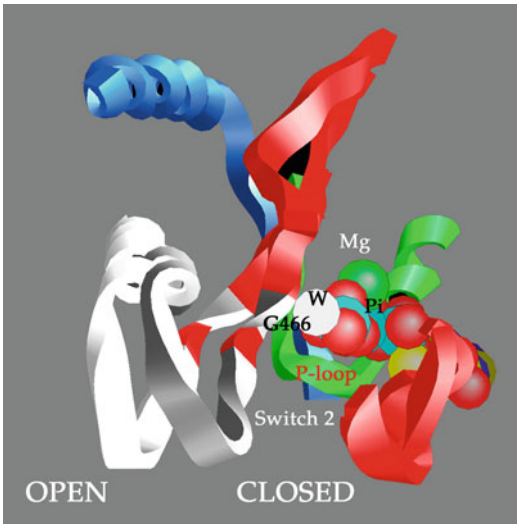
The cross-bridge cycle [20] is shown incorporating the swinging lever arm. The binding of ATP to the actin-myosin complex (**1**) leads to rapid dissociation of the cross-bridge from actin but without hydrolysis of ATP. (**2**). The cross-bridge then undergoes a conformational change (recovery stroke) that puts the lever arm in the pre-power-stroke conformation. (**3**). This form is the ATPase. Subsequent rebinding to actin (**4**) leads to product release and the moving "DOWN" of the lever arm (power stroke). Species 1,2 and 3 and their properties are derived from 3 crystal structures (discussed below). No crystal structure analogous to species 4 has been defined to date and its form is inferred from the neighbouring states (from [7]). A movie of the Lymn-Taylor cycle can be seen at http://valelab.ucsf.edu/images/movies/mov-muscmyosinmotrev6.mov.

6.10 Energetics

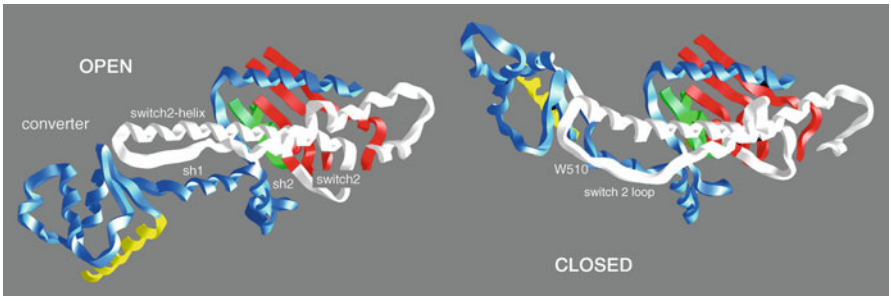
Each time the cross bridge goes through a cycle of contraction it hydrolyses one molecule of ATP. The free energy of ATP hydrolysis under physiological conditions is 85 zJ/molecule ($zJ = zeptoJ = 10^{-21}\,J$). The efficiency of muscle is about 45% at

high load (it decreases at high velocities of shortening) giving an available free energy of 38 zJ. The observed values of the size of the working stroke (6–7 nm) and force generated (6 pN) give a value of 36–42 zJ, which is close to 38 zJ [23].

6.11 The Recovery Stroke



In going from post rigor 'Open' (white) to pre-power stroke 'Closed' (grey) switch 2 moves in to form a hydrogen bond with the gamma phosphate [6]. The ATP binding site is shown: the P-loop, switch 1 (SW1) switch 2 (SW2), the relay helix and converter domain in a. the near-rigor state (**2**) and b. the pre-power stroke state (**3**).

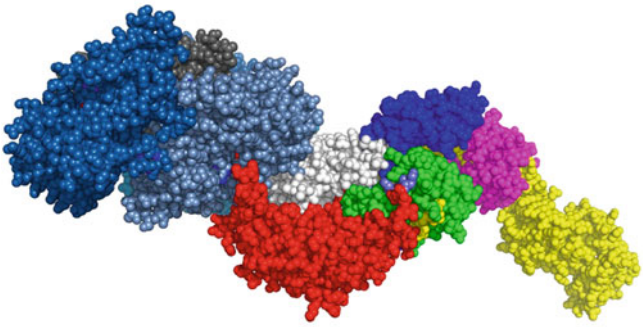


The recovery stroke is shown looking at right angles to the actin filament axis (from [6]). Note the rotation of the converter (blue) and lever arm (yellow) accompanying

the movement of switch 2. The reaction pathway between these two end states has been calculated [4]): http://www.pnas.org/content/vol0/issue2005/images/data/0408784102/DC1/08784Movie1.mpg, http://www.pnas.org/content/vol0/issue2005/images/data/0408784102/DC1/08784Movie3.mpg.

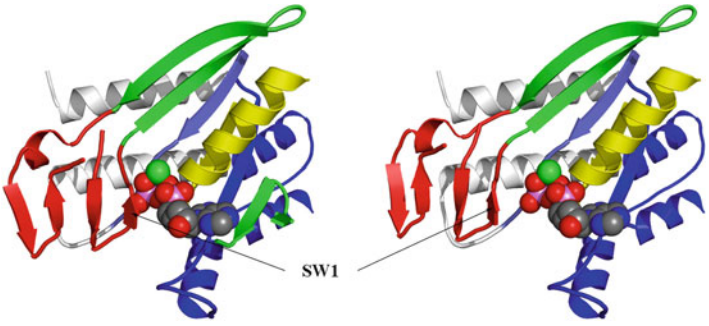
6.12 Weak and Strong Binding to Actin

ATP binding causes weak actin binding; actin binding leads to a closing of the actin binding cleft and release of products (γ-Phosphate and ADP) from myosin.



The strong binding of the myosin cross-bridge to actin is shown looking along the actin helix (space filling models). The actin helix is to the left (blue). One cross-bridge is shown (right). The lever arm is to the right attached to the converter domain (dark blue). Note the proximity of the upper (red) and lower (white) 50 K domains in strong binding (see [1, 2, 11]).

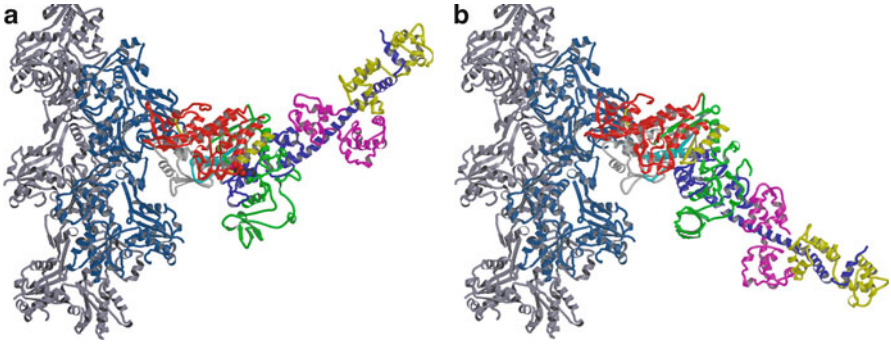
6.13 Strong Binding: β-Sheet Twists and Moves Switch 1 Away from the γ-Phosphate



Left shows the pre-power-stroke; right the strongly bound top of power stroke. The strong binding to actin twists the β-sheet and rotates the upper 50 K domain clockwise about an axis passing through the sheet. SW1 (switch 1) is fixed to the upper 50 K domain and on strong binding to actin rotates (moves down into the plane of the paper) away from the γ-phosphate, which opens the nucleotide binding pocket.

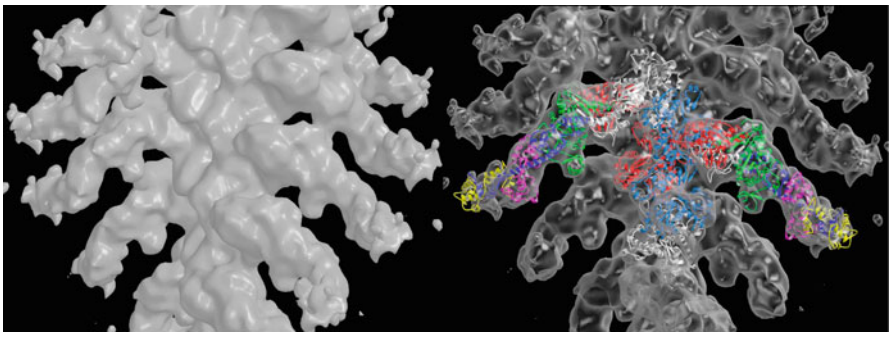
6.14 The Power Stroke

At present we can only get an approximate model the power stroke by looking at the recovery stroke and imagine it running backwards. The actual power stroke will be different because it occurs while bound to actin, which probably means that the β -sheet stays in the twisted form. The position of the converter domain depends on whether the relay helix has a kink near its middle point or not. Removing the kink causes the lever arm to rotate by 60°, which is very likely to be the elementary structural event in the power stroke. This is brought about by binding to actin. The position of the converter domain is stabilised by the interaction of the outer end of the relay helix with the SH1 helix. These two helices rotate together as the converter domain rotates. The axis of rotation runs between the two helices. (See movie for the recovery stroke: http://www.pnas.org/content/vol0/issue2005/images/data/0408784102/DC1/08784Movie3.mpg).



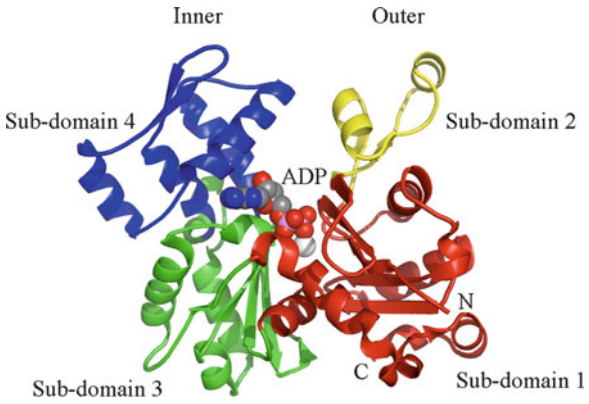
The power stroke is shown: **a.** A model of the strongly bound pre-power stroke state. **b** shows the rigor state. The actin is shown in blue/grey. From [7].

6.15 Decorated Actin Gives a View of the Actin-Myosin Interface in Strong Binding



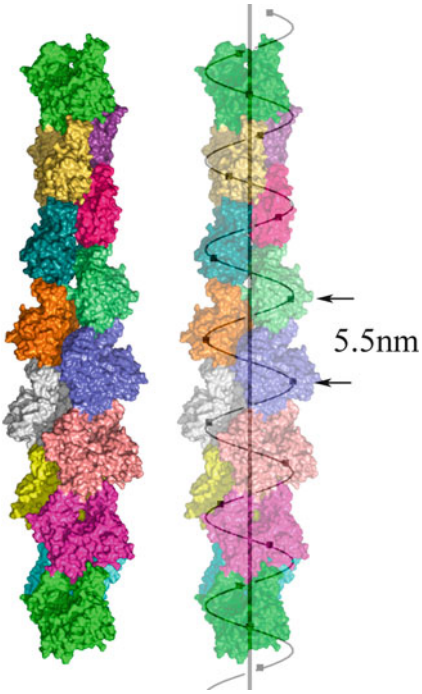
A space filling model is shown calculated from atomic coordinates which has been fitted to a high resolution electron microscope reconstruction of decorated actin [9, 11]- see also ref 24. Two cross-bridges (subfragment-1) on opposite sided of actin are shown.

6.16 Structure of Actin

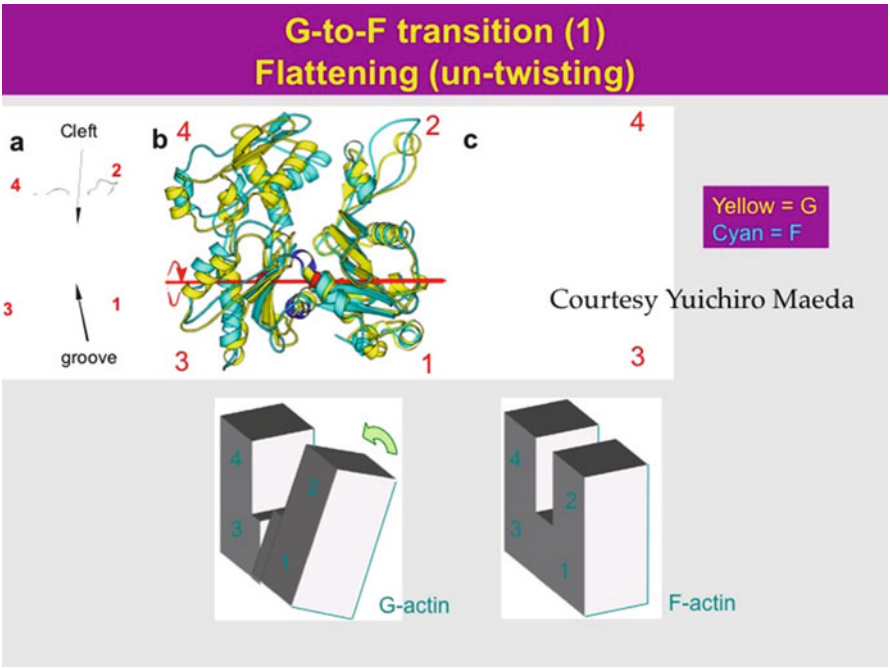


Actin comes in two forms: G-actin (monomeric) and F-actin (filamentous). The diagram shows a representative structure of G-actin (data from [22], PDB: 1J6Z). Sub-domains 1 and 2 constitute the "outer" domain and sub-domains 3 and 4 the "inner" domain. These names refer to their positions in the F-actin helix – the inner domain is close to the helix axis. Sub-domain 1 contains the binding site for myosin. Many capping proteins bind in the cleft between sub-domains 1 and 3. The nucleotide (ATP or ADP) binds between the outer and inner domains. The binding pocket for the base is on

the inner domain between sub-domains 3 and 4. The structures of sub-domain 1 and sub-domain 3 are similar and appear to have arisen by gene duplication. Both the N- and C-termini are in sub-domain 1. The first four residues of the N-terminus are disordered.



The helical structure of F-actin is shown [10]. This was derived from the X-ray fiber diffraction pattern and the crystal structure of G-actin. The molecules are arranged on a single helix with 13 molecules repeating in 6 left-handed turns. The rise per molecule is 2.75 nm and the twist per molecule is −166.67º. Since this value is close to 180º the helical structure takes on the appearance of a two-stranded right-handed screw. The rise per molecule along the long-pitch helix is 5.5 nm. The molecules used to make this diagram are actually G-actin [18]. A recent EM study [5] has given atomic coordinates for F-actin. The main difference between G and F actin is a propeller-like twist of the molecule. The F-actin structure is flatter.



Upper: superposition of the F- and G- structures.

Lower: a diagrammatic representation of the differences between F- and G- actin Note the 17º rotation of the outer domain with respect to the inner domain in passing from G- to F-.

References

1. Coureux PD, Sweeney HL, Houdusse A (2004) Three myosin V structures delineate essential features of chemo-mechanical transduction. EMBO J 23:4527–4537
2. Coureux PD, Wells AL, Menetrey J, Yengo CM, Morris CA, Sweeney HL, Houdusse A (2003) A structural state of the myosin V motor without bound nucleotide. Nature 425:419–423
3. Craig R, Woodhead JL (2006) Structure and function of myosin filaments. Curr Opin Struct Biol 16:204–212
4. Fischer S, Windshugel B, Horak D, Holmes KC, Smith JC (2005) Structural mechanism of the recovery stroke in the myosin molecular motor. Proc Natl Acad Sci USA 102:6873–6878
5. Fujii T, Iwane A, Yanagida T, Namba K (2010) Direct visualisation of secondary structrues of F-actin by electron cryomicroscopy. Nature 467:724–728
6. Geeves MA, Holmes KC (1999) Structural mechanism of muscle contraction. Annu Rev Biochem 68:687–728
7. Geeves MA, Holmes KC (2005) The molecular mechanism of muscle contraction. Adv Protein Chem 71:161–193
8. Gregorio CC, Granzier H, Sorimachi H, Labeit S (1999) Muscle assembly: a titanic achievement? Curr Opin Cell Biol 11:18–25

9. Holmes KC, Angert I, Kull FJ, Jahn W, Schroder RR (2003) Electron cryo-microscopy shows how strong binding of myosin to actin releases nucleotide. Nature 425:423–427
10. Holmes KC, Popp D, Gebhard W, Kabsch W (1990) Atomic model of the actin filament. Nature 347:44–49
11. Holmes KC, Schroder RR, Sweeney HL, Houdusse A (2004) The structure of the rigor complex and its implications for the power stroke. Philos Trans R Soc Lond B Biol Sci 359:1819–1828
12. Huxley AF, Niedergerke, RM (1954) Structural changes in muscle during contraction. Interference microscopy of living muscle fibres. Nature 173:71–973
13. Huxley HE (1957) The double array of filaments in cross-striated muscle. J Biophys Biochem Cytol 3:631–648
14. Huxley H, Hanson J (1954) Changes in the cross-striations of muscle during contraction and stretch and their structural interpretation. Nature 173:973–976
15. Huxley HE, Simmons RM, Faruqi AR, Kress M, Bordas J, Koch MH (1981) Millisecond time-resolved changes in x-ray reflections from contracting muscle during rapid mechanical transients, recorded using synchrotron radiation. Proc Natl Acad Sci USA 78:2297–2301
16. Irving M, Lombardi V, Piazzesi G, Ferenczi MA (1992) Myosin head movements are synchronous with the elementary force-generating process in muscle. Nature 357:156–158
17. Irving M, Piazzesi G, Lucii L, Sun YB, Harford JJ, Dobbie IM, Ferenczi MA, Reconditi M, Lombardi V (2000) Conformation of the myosin motor during force generation in skeletal muscle. Nat Struct Biol 7:482–485
18. Kabsch W, Mannherz HG, Suck D, Pai EF, Holmes KC (1990) Atomic structure of the actin:DNase I complex. Nature 347:37–44
19. Linari M, Piazzesi G, Dobbie I, Koubassova N, Reconditi M, Narayanan T, Diat O, Irving M, Lombardi V (2000) Interference fine structure and sarcomere length dependence of the axial x-ray pattern from active single muscle fibers. Proc Natl Acad Sci USA 97:7226–7231
20. Lymn RW, Taylor EW (1971) Mechanism of adenosine triphosphate hydrolysis by actomyosin. Biochemistry 10:4617–4624
21. Margossian SS, Lowey S (1982) Preparation of myosin and its subfragments from rabbit skeletal muscle. Methods Enzymol 85:55–71
22. Otterbein LR, Graceffa P, Dominguez R (2001) The crystal structure of uncomplexed actin in the ADP state. Science 293:708–711
23. Piazzesi G, Reconditi M, Linari M, Lucii L, Bianco P, Brunello E, Decostre V, Stewart A, Gore DB, Irving TC, Irving M, Lombardi V (2007) Skeletal muscle performance determined by modulation of number of myosin motors rather than motor force or stroke size. Cell 131:784–795
24. Rayment I, Holden HM, Whittaker M, Yohn CB, Lorenz M, Holmes KC, Milligan RA (1993) Structure of the actin-myosin complex and its implications for muscle contraction. Science 261:58–65
25. Rayment I, Rypniewski WR, Schmidt-Base K, Smith R, Tomchick DR, Benning MM, Winkelmann DA, Wesenberg G, Holden HM (1993) Three-dimensional structure of myosin subfragment-1: a molecular motor. Science 261:50–58
26. Reconditi M (2006) Recent improvements in small angle x-ray diffraction for the study of muscle physiology. Rep Prog Phys 69:2709–2759
27. Szent-Györgyi AG (1953) Meromyosins, the subunits of myosin. Arch Biochem Biophys 42:305–320

Chapter 7
Molecular Basis of Allosteric Transitions: GroEL

Amnon Horovitz

Abstract Chaperonins such as GroEL from *Escherichia coli* are molecular machines that facilitate protein folding by undergoing energy (ATP)-dependent movements that are coordinated in time and space owing to complex allosteric regulation. Here, we describe some of the various functional (allosteric) states of GroEL, the pathways by which they inter-convert and the coupling between allosteric transitions and protein folding reactions.

7.1 Introduction

The activity of many enzymes (or other proteins) is controlled by the binding of effector molecules at sites that are distinct from the proteins' active sites. Such control, which reflects communication between active (e.g. catalytic) and regulatory sites that are often distant from each other, has been termed allosteric regulation. The term allostery comes from the Greek words 'allos' and 'stereos' that respectively mean "other" and "solid (object)" in reference to the fact that the regulatory and active sites of allosteric proteins are physically distinct. Effectors that enhance a protein's activity are referred to as allosteric activators whereas those that decrease a protein's activity are called allosteric inhibitors. Allosteric regulation was initially discovered in the context of metabolic control loops, such as feedforward from upstream substrates or feedback from downstream products. In recent years, it has become apparent, however, that allosteric regulation is also crucial for the proper functioning of many biomolecular machines for reasons that will be elaborated below.

A. Horovitz (✉)
Department of Structural Biology, Weizmann Institute of Science, 76100 Rehovot, Israel
e-mail: amnon.horovitz@weizmann.ac.il

M.A. Carrondo and P. Spadon (eds.), *Macromolecular Crystallography*,
NATO Science for Peace and Security Series A: Chemistry and Biology,
DOI 10.1007/978-94-007-2530-0_7, © Springer Science+Business Media B.V. 2012

Allosteric regulation is often achieved *via* ligand-mediated changes in the conformations of multimeric proteins that are manifested in cooperative ligand binding. Such cooperativity may be either positive, when ligand binding to one site enhances the affinity of the ligand to other sites, or negative when the opposite occurs. Several models, in particular the Monod-Wyman-Changeux (MWC) [15] and the Koshland-Némethy-Filmer (KNF) [12] models have been developed to describe cooperativity in ligand binding by oligomeric proteins. In both models, cooperative ligand binding is due to conformational changes in the protein that may be either concerted [15], sequential [12] or a combination of both [3]. In the case of the MWC model, cooperativity in ligand binding is due to an equilibrium between two unligated states: a tense (**T**) state with relatively low affinity for the ligand, which is the predominant form in the absence of ligand, and a relaxed (**R**) state with relatively high affinity for the ligand. In this model, the extent of cooperativity is determined by the equilibrium constant L (=[T]/[R]) and by the relative affinities of the ligand for the **T** and **R** states ($c = K_R/K_T$). The MWC model is, therefore, relatively simple since it involves only two parameters, L and c, but it has the limitation that it cannot account for negative cooperativity that the more complex KNF model is able to explain.

7.2 Nested Allostery in the Chaperonin GroEL

The MWC and KNF models were initially developed with relatively small oligomeric proteins such as hemoglobin in mind. Large macromolecular assemblies can, however, display allosteric properties that are not observed in smaller proteins. These unusual allosteric properties arise owing to interactions between allosteric units that form the large assembly. Nested allosteric models, which were first developed to describe certain linkage phenomena in hemoglobin [22], are particularly useful for describing allostery in large systems with hierarchical structure. The hierarchical structure in such systems suggests that a corresponding hierarchy in allosteric interactions may also exist. The chaperonins, which consist of two back-to-back stacked oligomeric rings with a cavity at each end where protein folding takes place (for recent review see [7]), are striking examples for large multimeric assemblies with hierarchical structures that display unusual allosteric properties. Steady-state kinetic measurements of initial rates of ATP hydrolysis by the *Escherichia coli* chaperonin GroEL at different concentrations of ATP (Fig. 7.1a) showed that it undergoes two ATP-promoted allosteric transitions: one with a midpoint at relatively low ATP concentrations and the second with a midpoint at higher concentrations of ATP [26]. Each of the allosteric transitions is reflected in intra-ring positive cooperativity in ATP binding and hydrolysis by GroEL, with respect to ATP and K^+ (for review see [6]). The higher ATP concentration required to effect the second allosteric transition reflects inter-ring negative cooperativity in ATP binding. A nested allosteric model (Fig. 7.2) for cooperativity in ATP binding by GroEL that accounts for these findings was put forward [26] in which, in accordance

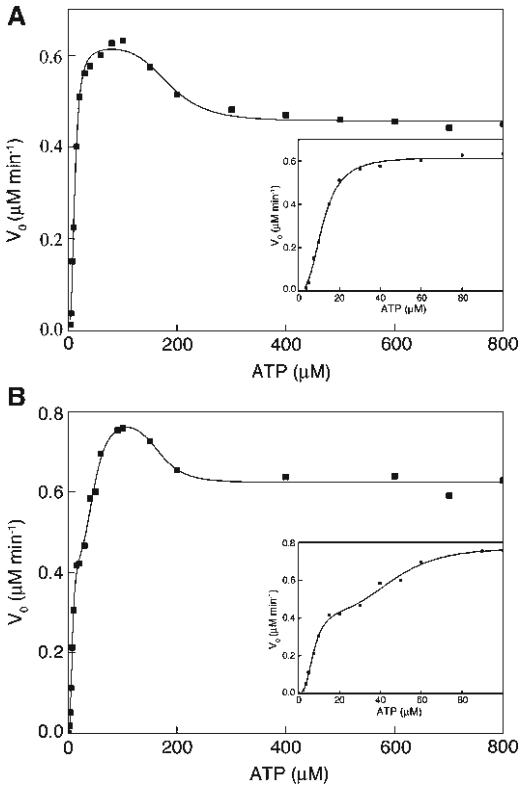


Fig. 7.1 Initial velocities of ATP hydrolysis by a wild-type version of GroEL (**a**) and by the Asp155→Ala mutant (**b**) as a function of ATP concentration. The data were fitted to the Hill equation for two (**a**) or three (**b**) allosteric transitions. For more details, see Danziger et al. [2]

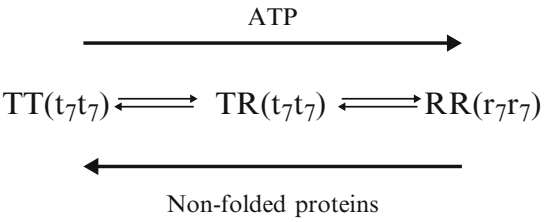


Fig. 7.2 Scheme for the different allosteric states of GroEL according to the nested model. In the absence of ligands, GroEL is predominantly in the **TT** (t_7t_7) state with low affinity for ATP and high affinity for non-folded substrates. In the presence of ATP, the equilibrium is shifted toward the **TR** (t_7r_7) and **RR** (r_7r_7) states. Rings in the **R** state have high affinity for ATP and low affinity for non-folded substrates

with the MWC representation, each ring is in equilibrium between tense (**T**) and relaxed (**R**) states with relatively low and high affinities for ATP, respectively. In the presence of increasing concentrations of ATP, the GroEL double-ring switches in a sequential manner from the **TT** state (both rings are in the **T** state) via the **TR** state to the **RR** state in accordance with the KNF model. In this model [26], MWC-type

allosteric interactions that lead to intra-ring positive cooperativity in ATP binding by GroEL are nested in KNF-type allosteric interactions that lead to inter-ring negative cooperativity in ATP binding. Plots of the observed rate constant of the **T** → **R** transition as a function of ATP concentration for double-ring GroEL variants (in which the mutations F44W [28], Y485W [1] or R231W [20] were introduced to facilitate the following of ATP-promoted conformational changes by monitoring time-resolved changes in fluorescence) have also been found to be bi-sigmoidal. By contrast, in the case of single-ring versions of GroEL (SR1), plots of initial rates of ATP hydrolysis [8] or fractional saturation [4] at different concentrations of ATP were found to be mono-sigmoidal, thereby indicating that SR1 undergoes only one allosteric transition as predicted by the nested model.

What are the relative affinities of the **T** and **R** states of GroEL for protein substrates? Folding reactions are usually initiated by rapid transfer of the unfolded protein from unfolding to refolding conditions and, therefore, are far from equilibrium (or steady-state with respect to the protein that is folding). Hence, in order to study the effects of non-folded proteins on allostery in GroEL, it was necessary to find a protein that is unfolded and soluble under the folding conditions in which the GroEL ATPase reactions are studied. α-Lactalbumin was found to be suitable for this purpose since it unfolds after reduction and removal of Ca^{++} ions but remains soluble. The extent of cooperativity (measured by the Hill coefficient) in the **TT** → **TR** transition was found to first increase and then decrease as a function of α-lactalbumin concentration [27]. These results indicated that (1) the **T** and **R** states have relatively high and low affinities for non-folded protein substrates, respectively; and that (2) binding of ATP does not occur exclusively to the **R** state. These results showed, therefore, that the **T** and **R** states serve as protein substrate acceptor and release-states, respectively. Structural analyses (for review see, for example, [13]) have shown that the lining of the cavity changes from hydrophobic to more hydrophilic upon the **T** to **R** transition, thus explaining the functional switch from a protein substrate-acceptor state to a protein substrate release state. The change in the lining of the cavity from hydrophobic to hydrophilic is most pronounced when a GroEL ring is bound to GroES, a heptameric ring-shaped [23] co-factor required for the folding and release of so-called stringent substrates [7].

7.3 Pathways of Allosteric Transitions of GroEL

Allosteric regulation is responsible for the repeated cycling between different functional states that is characteristic of molecular machines. Hence, a preliminary understanding of how chaperonins function as machines requires knowing their main allosteric states and associated functional properties. A deeper question concerns the nature of the transitions between the relatively stable different allosteric states. In other words, do populated kinetic intermediates exist (e.g. are the transitions concerted or sequential)? Are there single or parallel pathways between states [5]? Such questions are often ignored (and might indeed be of little importance) when

dealing with allosteric regulation in the context of metabolic control (e.g. oxygen uptake and release by hemoglobin) but they are of considerable potential significance for molecular machines since their efficiency may be path-dependent.

Support for the assumption in the nested model regarding the concerted nature of the ATP-promoted intra-ring allosteric transitions of GroEL has come from kinetic experiments [29] and simulations [14] that showed that steric repulsion would arise if one subunit switched from the **t** to **r** conformation (**t** and **r** stand for the respective conformations of a subunit in the low (**T**) and high (**R**) affinity states of a ring for ATP) while its neighbors have not. In addition to steric hindrance, salt-bridge switching involving R197, in particular, plays a key role in the allosteric transition of GroEL [17, 25]. Interestingly, the mutation D155A in GroEL that breaks an intra-subunit salt-bridge of D155 with R395 was found to convert its intra-ring allosteric transitions from concerted to sequential [2], thereby demonstrating that cooperativity in this system is indeed due to coupled tertiary conformational changes [14]. The break-in-symmetry in this mutant is reflected in the observation that a plot of its initial rates of ATP hydrolysis as a function of ATP concentration is not bi-sigmoidal as in the case of wild-type GroEL (Fig. 7.1a) but tri-phasic (Fig. 7.1b). The break-in-symmetry was visualized by electron microscopy analysis of SR1 containing the D155A mutation that showed that this mutant undergoes sequential allosteric transitions such as $\mathbf{t}_7 \rightarrow \mathbf{t}_4\mathbf{r}_3 \rightarrow \mathbf{r}_7$ or $\mathbf{t}_7 \rightarrow \mathbf{t}_5\mathbf{r}_2 \rightarrow \mathbf{r}_7$ as opposed to the $\mathbf{t}_7 \rightarrow \mathbf{r}_7$ allosteric transition of wild-type GroEL. The D155A mutant provides, therefore, a powerful tool to test the question of how concerted vs. sequential allosteric transitions impact GroEL's folding function.

7.4 Coupling Between Allosteric Transitions in GroEL and Protein Folding

Electron-microscopy [18] and kinetic [19] data indicate that the intra-ring ATP-promoted allosteric transitions of the eukaryotic chaperonin CCT (chaperonin-containing t-complex polypeptide 1 (TCP-1)) are sequential, in contrast with the concerted transitions of the prokaryotic chaperonin GroEL. Consequently, we hypothesized that ATP-induced sequential conformational changes in the eukaryotic chaperonin CCT facilitate sequential protein substrate release and, as a result, domain-by-domain folding [18]. By contrast, we suggested that ATP-induced concerted conformational changes in the prokaryotic chaperonin GroEL facilitate all-or-none substrate release. The difference in these release mechanisms was proposed to reflect the need for a more efficient folding mechanism of multi-domain proteins that are more abundant in eukaryotic cells and prone to misfold. Results of lattice model simulations [9] were found to support this proposal. Hence, we decided to test it experimentally [11, 16] by comparing assisted folding by wild-type GroEL with that by the D155A GroEL mutant that undergoes sequential transitions.

Two types of model multi-domain substrates were used in these experiments. In the first type, we took advantage of the fact that GroEL-assisted folding of

so-called stringent substrates requires GroES [7, 23]. We fused a stringent (i.e. GroES-dependent) substrate, rhodanese, to the non-stringent substrates mouse dihydrofolate reductase (mDHFR) or enhanced green fluorescent protein (EGFP) and showed that folding of the non-stringent part of the chimeras occurs in a step-wise fashion, with respect to ATP concentration, when folding is assisted by the D155A mutant but not when it is assisted by the wild-type variant [11]. In these experiments, only the non-stringent component of the chimeras was able to fold to the native state while the stringent part remained bound and unfolded. The second type of substrate we employed was a chimera comprising the cyan and yellow fluorescent proteins, CyPet and YPet, which are both non-stringent substrates. The advantage in using this substrate is that the folding yield of each domain can be determined separately by measuring its respective intrinsic fluorescence and that of the chimera (with both domains folded) by measuring the Förster resonance energy transfer (FRET) between the two domains. Hence, by using this chimera as a substrate, it was possible to determine whether assisted folding by GroEL (either the wild-type variant or the D155A mutant) of one domain tends to be accompanied by folding of the other domain (i.e. a 'concerted' mechanism) or whether release and folding of the two domains is independent. It was found that a concerted ATP-induced allosteric switch results in a more concerted release of substrate domains [16], thereby linking the allosteric properties of the GroEL machine with its folding function.

7.5 Conclusions

Considerable progress has been made in the structural characterization of the different relatively stable allosteric states of GroEL [17, 23] and their corresponding functional properties. However, there is little experimental data regarding the pathways by which these allosteric states inter-convert owing to (1) the transient nature of the intermediates; (2) the potential multiplicity of pathways; and (3) the difficulties in measuring networks of energetic connectivity in large systems. Computational methods, mainly elastic network models [21, 24] and correlated mutation analysis [10], have been used extensively to identify pathways of allosteric communication in GroEL but the results of these studies remain to be verified experimentally.

References

1. Cliff MJ, Kad NM, Hay N, Lund PA, Webb MR, Burston SG, Clarke AR (1999) A kinetic analysis of the nucleotide-induced allosteric transitions of GroEL. J Mol Biol 293:667–684
2. Danziger O, Rivenzon-Segal D, Wolf SG, Horovitz A (2003) Conversion of the allosteric transition of GroEL from concerted to sequential by the single mutation Asp-155 → Ala. Proc Natl Acad Sci USA 100:13797–13802
3. Eigen M (1967) Kinetics of reaction control and information transfer in enzymes and nucleic acids. Nobel Symp 5:333–369

4. Frank GA, Kipnis Y, Smolensky E, Daube SS, Horovitz A, Haran G (2008) Design of an optical switch for studying conformational dynamics in individual molecules of GroEL. Bioconjug Chem 19:1339–1341
5. Horovitz A, Amir A, Danziger O, Kafri G (2002) φ value analysis of heterogeneity in pathways of allosteric transitions: Evidence for parallel pathways of ATP-induced conformational changes in a GroEL ring. Proc Natl Acad Sci USA 99:14095–14097
6. Horovitz A, Willison KR (2005) Allosteric regulation of chaperonins. Curr Opin Struct Biol 15:646–651
7. Horwich AL, Fenton WA (2009) Chaperonin-mediated protein folding: using a central cavity to kinetically assist polypeptide chain folding. Q Rev Biophys 42:83–116
8. Inobe T, Makio T, Takasu-Ishikawa E, Terada TP, Kuwajima K (2001) Nucleotide binding to the chaperonin GroEL: non-cooperative binding of ATP analogs and ADP, and cooperative effect of ATP. Biochim Biophys Acta 1545:160–173
9. Jacob E, Horovitz A, Unger R (2007) Different mechanistic requirements for prokaryotic and eukaryotic chaperonins: a lattice study. Bioinformatics 23:i240–i248
10. Kass I, Horovitz A (2002) Mapping pathways of allosteric communication in GroEL by analysis of correlated mutations. Proteins: Struct Funct Genet 48:611–617
11. Kipnis Y, Papo N, Haran G, Horovitz A (2007) Concerted ATP-induced allosteric transitions in GroEL facilitate release of protein substrate domains in an all-or-none manner. Proc Natl Acad Sci USA 104:3119–3124
12. Koshland DE Jr, Némethy G, Filmer D (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits. Biochemistry 5:365–385
13. Lin Z, Rye HS (2006) GroEL-mediated protein folding: making the impossible, possible. Crit Rev Biochem Mol Biol 41:211–239
14. Ma J, Sigler PB, Xu Z, Karplus M (2000) A dynamic model for the allosteric mechanism of GroEL. J Mol Biol 302:303–313
15. Monod J, Wyman J, Changeux JP (1965) On the nature of allosteric transitions: a plausible model. J Mol Biol 12:88–118
16. Papo N, Kipnis Y, Haran G, Horovitz A (2008) Concerted release by ATP of individual domains of a protein substrate of GroEL is demonstrated with FRET. J Mol Biol 380:717–725
17. Ranson NA, Farr GW, Roseman AM, Gowen B, Fenton WA, Horwich AL, Saibil HR (2001) ATP-bound states of GroEL captured by cryo-electron microscopy. Cell 107:869–879
18. Rivenzon-Segal D, Wolf SG, Shimon L, Willison KR, Horovitz A (2005) Sequential ATP-induced allosteric transitions of the cytoplasmic chaperonin containing TCP-1 revealed by EM analysis. Nat Struct Mol Biol 12:233–237
19. Shimon L, Hynes GM, McCormack EA, Willison KR, Horovitz A (2008) ATP-induced allostery in the eukaryotic chaperonin CCT is abolished by the mutation G345D in CCT4 that renders yeast temperature-sensitive for growth. J Mol Biol 377:469–477
20. Taniguchi M, Yoshimi T, Hongo K, Mizobata T, Kawata Y (2004) Stopped-flow fluorescence analysis of the conformational changes in the GroEL apical domain: relationships between movements in the apical domain and the quaternary structure of GroEL. J Biol Chem 279:16368–16376
21. Tehver R, Chen J, Thirumalai D (2009) Allostery wiring diagrams in the transitions that drive the GroEL reaction cycle. J Mol Biol 387:390–406
22. Wyman J (1967) Allosteric linkage. J Am Chem Soc 89:2202–2218
23. Xu Z, Horwich AL, Sigler PB (1997) The crystal structure of the asymmetric GroEL-GroES-$(ADP)_7$ chaperonin complex. Nature 388:741–750
24. Yang Z, Májek P, Bahar I (2009) Allosteric transitions of supramolecular systems explored by network models: application to chaperonin GroEL. PLoS Comput Biol 5:e1000360
25. Yifrach O, Horovitz A (1994) Two lines of allosteric communication in the oligomeric chaperonin GroEL are revealed by the single mutation Arg196 → Ala. J Mol Biol 243:397–401
26. Yifrach O, Horovitz A (1995) Nested cooperativity in the ATPase activity of the oligomeric chaperonin GroEL. Biochemistry 34:5303–5308

27. Yifrach O, Horovitz A (1996) Allosteric control by ATP of non-folded protein binding to GroEL. J Mol Biol 255:356–361
28. Yifrach O, Horovitz A (1998) Transient kinetic analysis of adenosine 5'-triphosphate binding-induced conformational changes in the allosteric chaperonin GroEL. Biochemistry 37:7083–7088
29. Yifrach O, Horovitz A (1998) Mapping the transition state of the allosteric pathway of GroEL by protein engineering. J Am Chem Soc 120:13262–13263

Chapter 8
Cell Signalling Through Covalent Modification and Allostery

Louise N. Johnson

Abstract Phosphorylation plays essential roles in nearly every aspect of cell life. Protein kinases catalyze the transfer of the γ-phosphate of ATP to a serine, threonine or tyrosine residue in protein substrates. This covalent modification allows activation or inhibition of enzyme activity, creates recognition sites for other proteins and promotes order/disorder or disorder/order transitions. These properties regulate signalling pathways and cellular processes that mediate metabolism, transcription, cell cycle progression, differentiation, cytoskeleton arrangement and cell movement, apoptosis, intercellular communication, and neuronal and immunological functions. In this lecture I shall review the structural consequences of protein phosphorylation using our work on glycogen phosphorylase and the cell cycle cyclin dependent protein kinases as illustrations. Regulation of protein phosphorylation may be disrupted in the diseased state and protein kinases have become high profile targets for drug development. To date there are 11 compounds that have been approved for clinical use in the treatment of cancer.

8.1 The Discovery of Protein Kinases and the Number of Protein Kinases Encoded in the Human Genome

The discovery of phosphorylation as a regulatory physiological mechanism arose from the work in 1955 of Eddie Fischer and Ed Krebs who showed that the activation of glycogen phosphorylase b to glycogen phosphorylase a in response to

L.N. Johnson (✉)
Laboratory of Molecular Biophysics, Department of Biochemistry,
University of Oxford, Oxford OX1 3QU, UK

Diamond Light Source, Harwell Science and Innovation Campus,
Didcot, Oxon OX11 0DE, UK
e-mail: louise.johnson@Diamond.ac.uk

M.A. Carrondo and P. Spadon (eds.), *Macromolecular Crystallography*,
NATO Science for Peace and Security Series A: Chemistry and Biology,
DOI 10.1007/978-94-007-2530-0_8, © Springer Science+Business Media B.V. 2012

adrenaline was dependent on a protein kinase action together with ATP. The second enzyme that was discovered to be controlled by phosphorylation was glycogen synthase and for many years control by phosphorylation was considered to be an idiosyncrasy of glycogen metabolism. The isolation and purification of cyclic AMP dependent protein kinase (PKA) in 1968 that showed protein phosphorylation was a much more ubiquitous phenomena. Until 1979, protein phosphorylation in eukaryotes appeared to be confined to serine and threonine residues, but in that year Tony Hunter and colleagues identified the third phospho-amino acid, phospho-tyrosine, as the product of a protein kinase activity in immunoprecipitates of a viral oncoprotein. In humans, phosphorylation on serine, threonine and tyrosine is approximately 86%, 12% and 2%, respectively. Although less numerous than serine or threonine phosphorylation, tyrosine phosphorylation is crucially important in health and disease. The discoveries that many oncoproteins possess tyrosine kinase activity and that this activity is associated with their transforming ability provided exciting advances into understanding the regulation of normal and abnormal cellular processes.

The completion of the human genome sequence allowed identification of the total number of human protein kinases [1]. This landmark paper produced the iconic diagram of the protein kinome in which most of the eukaryotic protein kinases (ePKs) could be assigned to seven major groups defined by their related functions. Manning et al. identified a total of 518 protein kinases in the human kinome comprising 478 ePKs and 40 aPKs (atypical protein kinases, proteins reported to have biochemical kinase activity, but which lack sequence similarity to the ePK domain).

8.2 Structural Consequences of Phosphorylation

Phosphorylation can have profound effects on the function of the target protein. The phosphoryl group with a pK_a ~6.7 is likely to be predominantly dianionic at physiological pH. The property of a double negative charge (a property not carried by any of the naturally occurring amino acids) and the capacity for the phosphoryl oxygens to form hydrogen bond networks confers special characteristics. Two types of interaction predominate. Firstly at tight binding sites used to stabilise a conformational state, the phosphate group frequently interacts with the side chain of one or more arginine residues. The guanidinium group of an arginine is well suited for such interactions because of its planar structure and its ability to form multiple hydrogen bonds. The guanidinium group ($pK_a > 12$) is a poor proton donor and cannot function as a general acid in the hydrolysis of phosphorylated amino acids. Theoretical calculations on the strengths of hydrogen bonds have shown that the bidentate interactions available to arginine with phosphate provide much stronger interactions than those that can be formed with ammonium ions as in lysine side chains. Secondly an interaction often observed at less tight phosphoryl group binding sites involves the interaction of the phosphate group with the main chain nitrogens at the start of

an α-helix, utilizing the partial positive charge of the helix dipole. In addition a number of other polar residues many also be involved in contacts including lysine, histidine, tyrosine, serine and threonine.

Phosphorylation can result in a number of different responses in a protein target [2]. Phosphorylation can activate enzyme activity through allosteric conformational changes as observed for glycogen phosphorylase and many protein kinases that rely on phosphorylation by upstream kinases for activity. Phosphorylation can inhibit enzyme activity as observed in isocitrate dehydrogenase, where the phosphate group acts as a steric blocking agent and does not promote any conformational change, and in the cyclin dependent protein kinase CDK2 where phosphorylation on Tyr15 impedes protein substrate recognition. Phosphorylation can lead to recognition sites for other protein molecules such as in the phospho-tyrosine recognition SH2 domains important for regulation of kinases such as Src, ZAP70, Fes and Abl protein. Less extensive but also important are the regulatory domains that recognise phospho-serine or phospho-threonine such as the 14-3-3 proteins or the polo-box domain of polo-like kinase where a phospho-serine site is recognised by the polo-box domain, which then targets the polo-like kinase to its substrate. In a further variation of specific site recognition, some protein kinases require hierarchical substrate phosphorylation where the phosphorylation of one site is necessary to create a recognition site to allow subsequent phosphorylation as occurs in the phosphorylation of the substrate APC (adenomatous polyposis coli) protein by CK1 and GSK3 as part of the pathway for β-catenin degradation in Wnt signalling. Phosphorylation may also cause an order to disorder transition as in the K^+ channel inactivation domain, or it may cause a disorder to order transition as in the KIX/pKID CBP co-activator protein/CREB transactivation domain system, these transitions being demonstrated in NMR structural studies. Phosphorylation can promote conformational changes that lead to protein association as in ERK2 or STAT proteins and entry to the nucleus. Phosphorylation may also cause protein/protein disassociation as in the CDK mediated phosphorylation of pRb that promotes dissociation of pRb from the transcription factor E2F/DP1 [3]. For almost all of these systems, structures have explained the molecular basis for these phenomena. ([4] and references therein).

8.3 Glycogen Phosphorylase; an Example of the Structural Consequences of Phosphorylation

Glycogen phosphorylase was the first phospho-protein structure to be understood in its non-phospho and phospho states [5]. According to the Monod-Wyman Changeux theory for allosteric activation it is proposed that the enzyme exists in two (or at least two) functional states, the T (tense) state, which is less active and is characteristic of non-phospho glycogen phosphorylase b (GPb), and the R state (relaxed) characteristic of the active state phosphorylated glycogen phosphorylase a (GPa). The response to stimulation by adrenaline that leads to activation of phosphorylase kinase and phosphorylation of GPb to GPa results in nearly instantaneous activation that is

needed for a fight or flight response. The activation of phosphorylase by phosphorylation is the end result of one of the best-understood signalling pathways. We now understand (almost) all the structures of the players involved in this pathway from the β-adrenergic receptor signalling through the heterotrimeric G proteins, to adenylyl cyclase to produce the second messenger, cyclic-AMP leading to activation of PKA that phosphorylates phosphorylase kinase, which activates phosphorylase.

Glycogen phosphorylase catalyzes the first step in the breakdown of glycogen. The reaction involves the phosphorylysis of the α-1,4-glycosidic bond of the terminal sugar of glycogen to yield glucose-1-phosphate. Phosphorylation by phosphorylase kinase takes place on serine 14, near the N-terminus of this large polypeptide chain (842 residues). The question is how does phosphorylation of a single serine residue activate such a large molecule? Glycogen phosphorylase is a dimer in which the two subunits are related by a two-fold axis of symmetry. The X-ray crystallography studies showed that phosphorylation produced conformational changes that signalled a change in quaternary structure that in turn led to the correct conformation of residues at the catalytic site. The catalytic site is at the centre of each subunit and well away from the subunit/subunit interface.

Thus the allosteric mechanism of phosphorylase allows binding events at sites that are over 45 Å from the catalytic site to activate the enzyme based on the intimate connection between the tertiary structure and the quaternary structure as the structure goes from the less active T state (GPb) to the more active R state (GPa).

8.4 Cyclin Dependent Protein Kinases Regulate the Cell Cycle

The cell cycle refers to the process by which a cell grows and divides [6]. Each of the main transitions of the cell cycle is controlled by protein kinases known as the cyclin dependent protein kinases (Cdks) because they are dependent of an additional subunit, the cyclin, for activity. Cdk2/cyclin A, is the kinase that regulates the transitions through S phase, where the cell replicates its DNA prior to cell division, and is the best understood Cdk. Like many protein kinases, most but not all Cdks are dependent on phosphorylation for activity. Phosphorylation takes place on a threonine residue (Thr160 in human Cdk2) in a region known as the activation segment. The catalytic core of all protein kinases comprises two lobes, the N-terminal lobe and the C-terminal lobe with a hinge region between them. ATP binding takes place in the region between the lobes. Protein substrate recognition takes place in the C-terminal lobe and involves the activation segment.

Cyclin binding to inactive Cdk2 brings about large changes in the N-terminal lobe that result in the correct location of ATP so that its γ-phosphate group is aligned for phosphoryl transfer. However the enzyme is inactive until it is phosphorylated on Thr160. This phosphorylation aligns the activation segment so that the protein substrate is correctly presented to the ATP [7, 8]. Thus in Cdk2/cyclin A phosphorylation is acting as an allosteric activator within a monomeric protein. Phosphorylation brings about conformational changes that are at a distance from the catalytic site but which are essential for catalysis.

Cdk2 is also controlled by phosphorylation-mediated inhibition. Phosphorylation at Tyr15 in the glycine rich loop of the N-terminal lobe inhibits kinase activity, a process that is used by the cell as a checkpoint to ensure that S-phase has been completed successfully before the cell moves on to the next stage of the cell cycle. Although this modification is close to the ATP binding site, it does not prevent ATP binding but rather prevents substrate binding [9].

In response to anti-proliferation signals, cells produce a number of Cdk inhibitor proteins, of which $p27^{Kip1}$ is most important during the early stages of the cell cycle. p27 inhibits Cdk2/cyclin A by exploiting a tight binding site on the cyclin and by wrapping around the kinase to distort the catalytic site [10]. p27 itself is controlled by phosphorylation. It is phosphorylated at one site (Thr187 towards its C-terminus) by Cdk2/cyclin A and this phosphorylation targets the inhibitor for ubiquitylation by the Skp1/Cullin/F-box complex (SCF). Ubiquitylation marks the protein for destruction by the proteasome. But how does Cdk2/cyclin A become active in order to destroy its own inhibitor? Recently it was discovered that p27 can be phosphorylated on two tyrosine residues in the region of p27 that locates to the Cdk2 catalytic site and these phosphorylations relieve inhibition by p27 [11]. The upstream kinases have not been definitively established but kinases such as Src and Lck, which are up-regulated in response to growth signals, are possible candidates.

8.5 Protein Phosphatases

Dephosphorylation by protein phosphatases is of equal importance in these signalling pathways. Protein kinases phosphorylate specific substrates often recognising a short sequence epitope around the site of phosphorylation to generate specificity. The protein phosphatases in contrast have broad recognition at their catalytic sites but utilize additional subunits to generate specificity at remote sites. An example of this is with the cyclin dependent protein kinase associated phosphatase KAP. KAP catalyzes the dephosphorylation of Cdk2 thus returning the kinase to its inactive state. It recognizes a structure in the C-terminal region of the kinase and uses this to direct the phospho-Thr160 to the catalytic site [12].

8.6 Protein Kinase Inhibitors

Protein kinase inhibitors are important targets in the fight against unwanted cell proliferation in cancer. Structure has provided knowledge that both explains the inhibitory properties and guides improvements to lead compounds [13–18]. An outstanding example has been the discovery of Glivec (imatinib) that inhibits the tyrosine kinase Abl, a kinase that becomes uncontrollably activated in chronic myeloid leukemia. Glivec binds preferentially to an inactivate conformation of Abl, a confirmation that is not readily accessible to other protein kinases and provides high specificity of Glivec for Abl [19, 20].

References

1. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S (2002) The protein kinase complement of the human genome. Science (New York) 298:1912–1934
2. Johnson LN, Lewis RJ (2001) Structural basis for control by phosphorylation. Chem Rev 101:2209–2242
3. Rubin SM, Gall AL, Zheng N, Pavletich NP (2005) Structure of the Rb C-terminal domain bound to E2F1-DP1: a mechanism for phosphorylation-induced E2F release. Cell 123:1093–1106
4. Johnson LN (2009) The regulation of protein phosphorylation. Biochem Soc Trans 37:627–641
5. Barford D, Hu SH, Johnson LN (1991) Structural mechanism for glycogen phosphorylase control by phosphorylation and AMP. J Mol Biol 218:233–260
6. Morgan DO (2007) The cell cycle: principles of control. New Science Press, London
7. Brown NR, Noble ME, Endicott JA, Johnson LN (1999) The structural basis for specificity of substrate and recruitment peptides for cyclin-dependent kinases. Nat Cell Biol 1:438–443
8. Pavletich NP (1999) Mechanisms of cyclin-dependent kinase regulation: structures of Cdks, their cyclin activators, and Cip and INK4 inhibitors. J Mol Biol 287:821–828
9. Welburn JP, Tucker JA, Johnson T, Lindert L, Morgan M, Willis A, Noble ME, Endicott JA (2007) How tyrosine 15 phosphorylation inhibits the activity of cyclin-dependent kinase 2-cyclin A. J Biol Chem 282:3173–3181
10. Russo AA, Jeffrey PD, Patten AK, Massague J, Pavletich NP (1996) Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A-Cdk2 complex. Nature 382:325–331
11. Ray A, James MK, Larochelle S, Fisher RP, Blain SW (2009) p27Kip1 inhibits cyclin D-cyclin-dependent kinase 4 by two independent modes. Mol Cell Biol 29:986–999
12. Song H, Hanlon N, Brown NR, Noble ME, Johnson LN, Barford D (2001) Phosphoprotein-protein interactions revealed by the crystal structure of kinase-associated phosphatase in complex with phosphoCDK2. Mol Cell 7:615–626
13. Johnson LN (2009) Protein kinase inhibitors: contributions from structure to clinical compounds. Q Rev Biophys 42:1–40
14. Liao JJ (2007) Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors. J Med Chem 50:409–424
15. Liu Y, Gray NS (2006) Rational design of inhibitors that bind to inactive kinase conformations. Nat Chem Biol 2:358–364
16. Noble ME, Endicott JA, Johnson LN (2004) Protein kinase inhibitors: insights into drug design from structure. Science (New York) 303:1800–1805
17. Sebolt-Leopold JS, English JM (2006) Mechanisms of drug inhibition of signalling molecules. Nature 441:457–462
18. Tintelnot-Blomley M, Lewis RA (2006) A critical appraisal of structure-based drug design. IDrugs 9:114–118
19. Nagar B, Bornmann WG, Pellicena P, Schindler T, Veach DR, Miller WT, Clarkson B, Kuriyan J (2002) Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571). Cancer Res 62:4236–4243
20. Schindler T, Bornmann W, Pellicena P, Miller WT, Clarkson B, Kuriyan J (2000) Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. Science (New York) 289:1938–1942

Chapter 9
Combining Cryo-EM and X-ray Crystallography to Study Membrane Protein Structure and Function

Werner Kühlbrandt

Abstract Membrane proteins perform a wide range of essential functions in all cells of all living organisms, ranging from the sensing, processing and propagation of extrinsic signals, passive or active transport of ions and solutes, creating or utilizing a membrane potential, to the import or secretion of entire proteins. In spite of intense, and increasingly successful efforts in determining membrane protein structures, they still present a formidable challenge in structural biology.

Only about 0.5% of all structures deposited in the pdb are membrane proteins, and this proportion has not been increasing rapidly in recent years. Most of the known membrane protein structures are prokaryotic, because these are easier to handle than their eukaryotic counterparts, often less complex, more stable, and more easily produced in the amounts required for structural studies. However, in terms of structure and molecular mechanisms they are often similar to those of eukaryotic organisms, and much can, and has, been learned from prokaryotic homologues.

X-ray crystallography is unsurpassed for providing detailed information about the three-dimensional structure, and mechanisms by which membrane proteins work at the atomic level (Fig. 9.1). Even more than 25 years after the first membrane protein structure [5], each new one is a triumph that takes years, or even decades, of dedicated, painstaking work to achieve. The most recent examples include a bacterial respiratory Complex-I [6], and our two structures of a bacterial carnitine transporter CaiT [20]. Structure determination takes a long time mainly because it remains

W. Kühlbrandt (✉)
Max-Planck-Institute of Biophysics, Max-von-Laue-Str.3, 60438
Frankfurt am Main, Germany
e-mail: Werner.Kuehlbrandt@mpibp-frankfurt.mpg.de

M.A. Carrondo and P. Spadon (eds.), *Macromolecular Crystallography*,
NATO Science for Peace and Security Series A: Chemistry and Biology,
DOI 10.1007/978-94-007-2530-0_9, © Springer Science+Business Media B.V. 2012

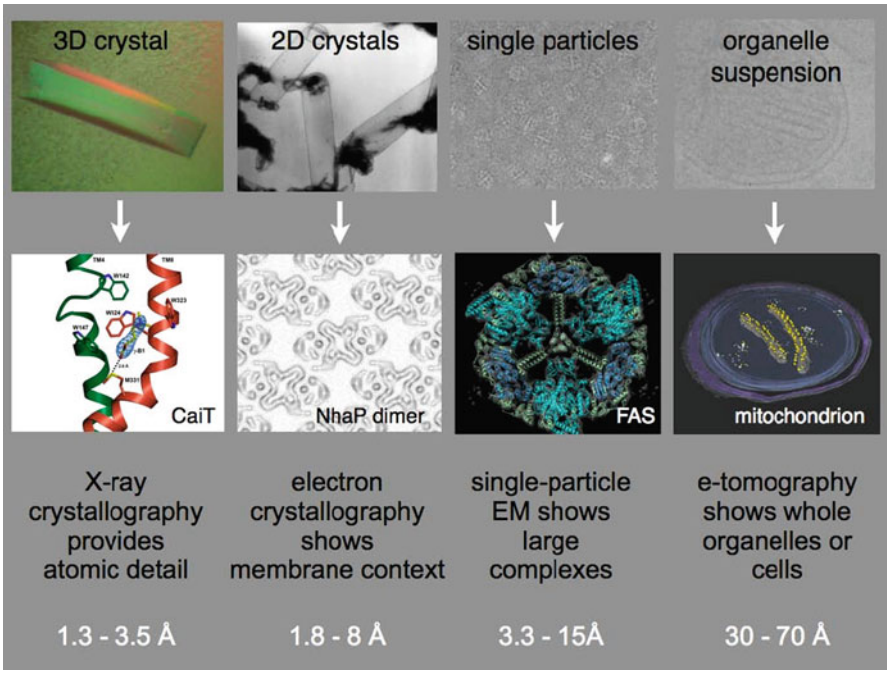


Fig. 9.1 Summary of methods used for investigating membrane protein structure, with resolution ranges (best to typical) given below

difficult to produce membrane proteins in sufficient quantities, to isolate and purify them, and to grow well-diffracting crystals.

For crystallization in 3D, membrane proteins have to be solubilised with detergents, which takes them out of their native membrane context. For functional studies it is often necessary to reconstitute the isolated proteins into a lipid bilayer. By similar methods, many membrane proteins form two-dimensional (2D) crystals, which can then be studied by **electron crystallography**. 2D crystals often grow more easily than 3D crystals, and because the protein is in a quasi-native lipid environment, it tends to be more stable. However, 2D crystals are rarely suitable for high-resolution structure determination, because of limited size and intrinsic disorder. To date, only very few membrane protein structures at better than 4 Å resolution have been determined by electron crystallography (Fig. 9.2, [9, 10, 13]).

The method is however well suited to studying membrane proteins under different conditions. Examples are pH-induced conformational changes in sodium-proton antiporters NhaA and NhaP [2, 22], which would be difficult with 3D crystals. 3D maps at 6–8 Å resolution can usually be obtained by crystallographic image processing. At this resolution, membrane- spanning alpha-helices are clearly visible,

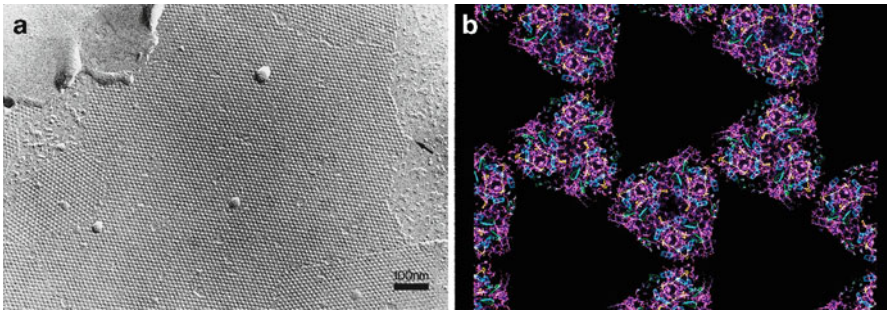


Fig. 9.2 Electron crystallography of 2D crystals. (**a**) 2D crystal of plant light-harvesting complex LHC-II on carbon film, shadowed with Pt/C. (**b**) 3.4 Å resolution structure of LHC-II as determined by electron crystallography [13]

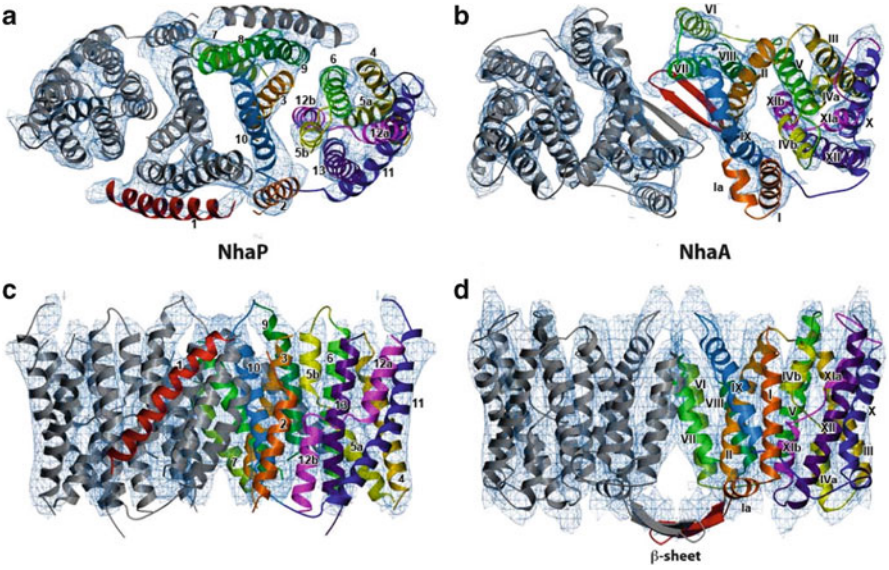


Fig. 9.3 Electron crystallographic structures of the sodium/proton antiporters NhaP1 from *Methanococcus jannaschii* (**a, b**) and NhaA from *E. coli* (**c**, **d**). (**a**, **c**) cytoplasmic view; (**b**, **d**) side view. The X-ray structure of the 12-helix NhaA [11] was fitted to the 6 Å map [24] of the NhaA dimer. A homology model of the 13-helix NhaP1 was built and fitted to the 7 Å 3D map of the NhaP1 dimer [8]

and if the X-ray structure of a related protein is known, atomic models can be built with confidence, as we have done in the case of NhaP1 (Fig. 9.3, [8]).

High levels of membrane proteins are often not tolerated by the expressing cells, or the proteins are not processed or folded correctly. Moreover, many membrane proteins assemble into large, hetero-oligomeric complexes that cannot be over- expressed with current technology. Such large and often fragile

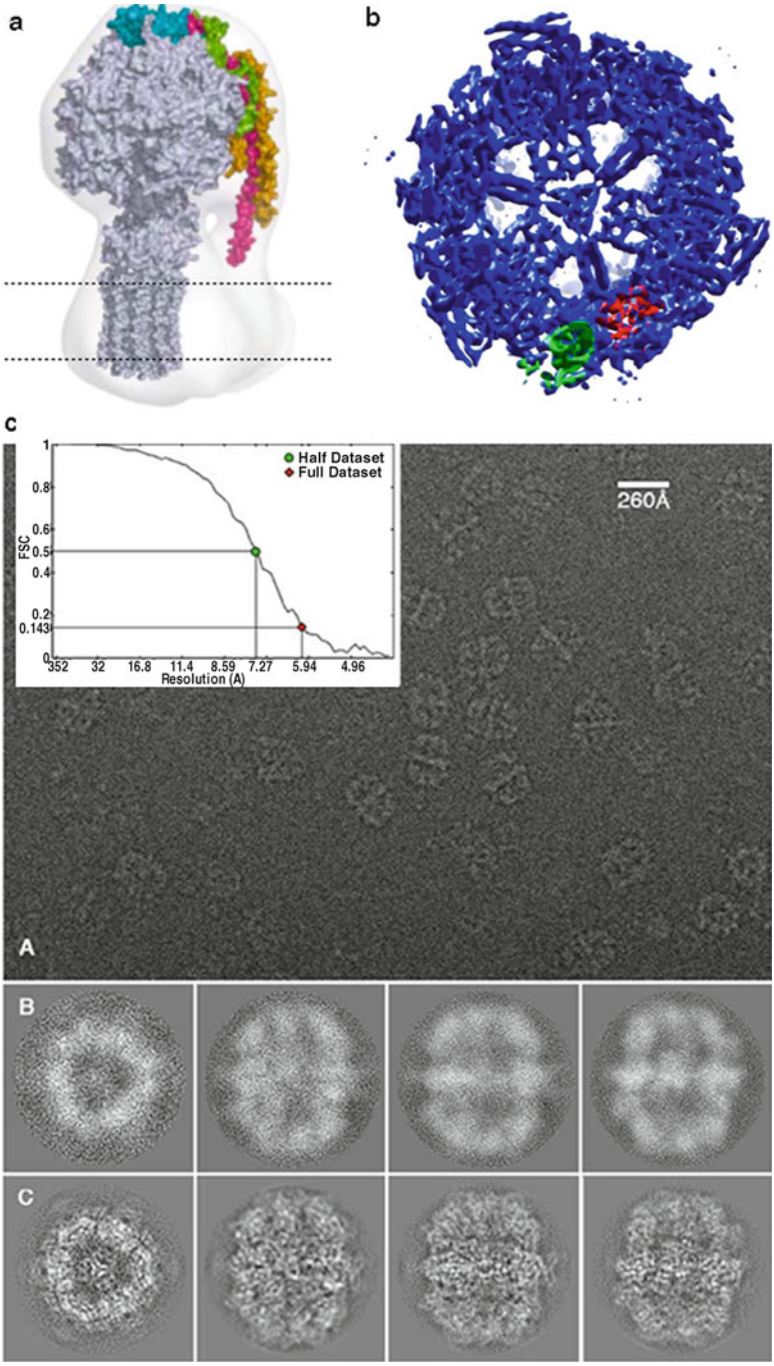


Fig. 9.4 Cryo-EM and image processing of large, non crystalline single particles. (**a**) 3D map of mitochondrial ATP synthase [18] with known X-ray structures of subunits fitted. (**b**) 5.9 Å map of the yeast fatty acid synthase equatorial wheel [7]. (**c**) Cryo-EM image of FAS (*top*) with Fourier shell correlation indicating the resolution in the inset. Selected class averages and re-projections below (From Gipson et al. [7])

macromolecular complexes are best studied by **single-particle cryo-EM** of non-crystalline material. Examples include the mitochondrial ATP synthase (Fig. 9.4a, [18]), the V-type ATP ases [23], respiratory chain supercomplexes [19], and the protein translocase of the outer mitochondrial membrane (TOM, [16]). As a further example, our recent 5.9 Å map of yeast fatty acid synthase (Fig. 9.4b, c, [7]) is included in this lecture. This is of course not a membrane protein, although it does produce the fatty acid chains of membrane lipids. The FAS map demonstrates the great potential of single-particle cryo-EM for studying structures at a resolution that, until recently, was thought to be the exclusive domain of crystallography. With non-enveloped spherical viruses, resolutions of 3.5 or better have been achieved [25], although with membrane protein complexes typical resolutions have been in the 10–30 Å range, probably due to inherent flexibility or conformational heterogeneity.

An important subset of membrane proteins is devoted to biological energy conversion in specialized membrane organelles, the mitochondria of animal, fungal and plant cells, and the chloroplasts of green plants. These organelles are particularly rich sources of membrane protein complexes, as they contain large copy numbers, and the proteins are comparatively stable. As a result we have now an almost complete picture of how individual complexes in respiration and photosynthesis work. However, to understand these and other membrane processes fully, we need to know how the complexes interact, how many of them there are in a given area or volume, and how they are organized in the membrane. At present, this information is most easily obtained by **electron cryo-tomography** (cryo-ET), which bridges the gap between crystallography and light microscopy. Although the resolution of cryo-ET is limited to ~30–70 Å, this comparatively new technique offers fascinating insights into the molecular makeup of membranes, organelles or even whole cells, especially when combined with higher-resolution information from X-ray crystallography or single- particle cryo-EM. Examples are the molecular arrangement of the F_1-F_o ATP synthase in the mitochondrial inner membrane, and of the chloroplast ATP synthase and photosystem-II in plant thylakoids.

Using electron cryo-tomography of small, whole mitochondria or mitochondrial membranes, we found that the ATP synthase is arranged in ~1 μm long rows of dimers, which are always located at the position of highest curvature of the inner membrane cristae (Fig. 9.5, [21]). This striking molecular arrangement is lost in yeast mutants lacking the ATP synthase subunits thought to be responsible for dimer formation, and in the mitochondria of ageing cells that are about to undergo apoptosis (unpublished).

In marked contrast to mitochondria, the chloroplast ATP synthase is monomeric, and confined to membrane regions of minimal curvature [4]. We believe this to reflect differences in membrane potential and pH gradient in the two organelles (Fig. 9.6). Visualizing the photosystem-II dimer in chloroplast membranes helps us to understand the molecular basis of lateral segregation into stacked and unstacked membrane regions (Fig. 9.7, [4]).

Electron microscopy in biology covers an enormous range from near-atomic resolution of individual membrane proteins [9, 10, 13], molecular assemblies at

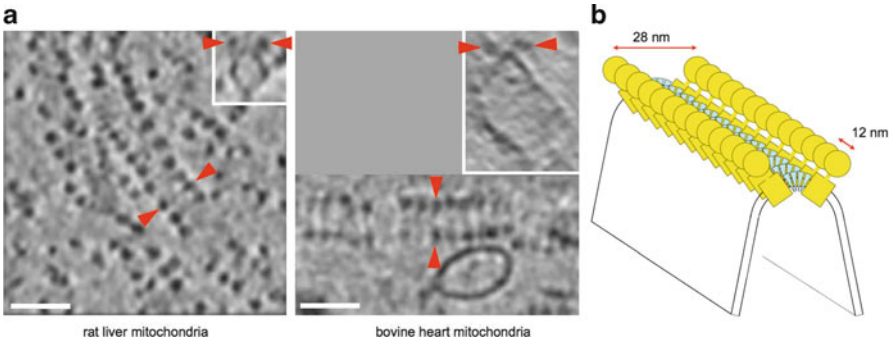


Fig. 9.5 Electron cryo-tomography of mitochondrial membranes. (**a**) Slabs of tomographic volumes showing dimer rows of ATP synthase in top and side views [21]. (**b**) Schematic of ATP synthase dimers along the highly curved edges of cristae membranes

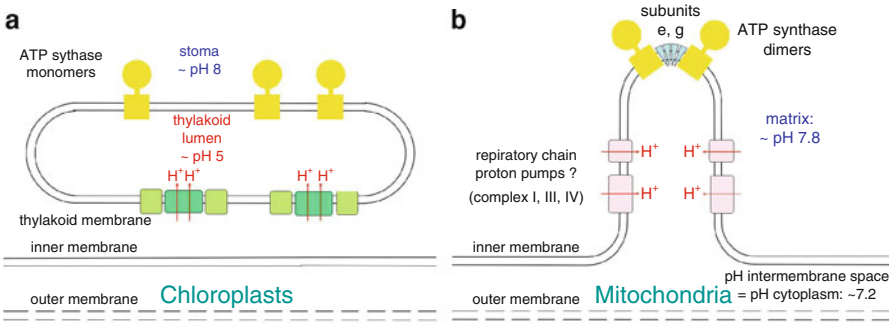


Fig. 9.6 Schematic arrangement of membrane protein complexes in chloroplast thylakoids (**a**) and mitochondria cristae (**b**)

increasingly higher resolution ([7, 12, 16, 19, 23, 25]), to the organellar or cellular context [1, 4, 21]. Whereas X-ray protein crystallography is a mature and comparatively straightforward method that has yielded tens of thousands of high-resolution structures, this is not yet the case for electron microscopy, where important fundamental improvements in instrument design are likely to be implemented within the next decade. These include the development of near-perfect solid-state electron detectors [15], and of phase plates to maximize the inherently low image contrast (Fig. 9.8, [3, 14]), which is limited by the electron dose tolerated by the radiation-sensitive biological objects.

We are developing a prototype instrument for in-focus phase contrast electron microscopy that will allow us to record tomograms and single particle images with higher contrast and, hopefully, resolution. Once these devices become routine, cryo-EM will be an even better complement for X-ray crystallography, and may even rival its success.

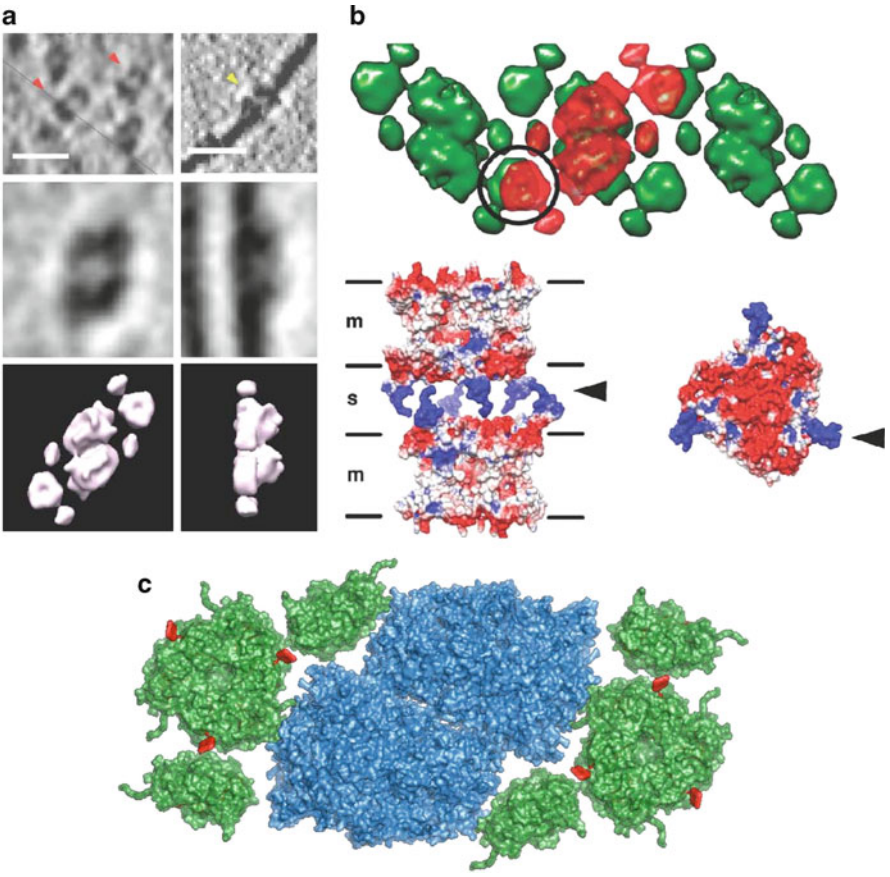


Fig. 9.7 Electron cryo-tomography of chloroplast membranes. (**a**) Photosystem-II dimers in the thylakoid membrane (*top*); averaged volumes top and side view (*centre*); PS-II supercomplex dimer map at 30 Å resolution (**b**) Arrangement of PS-II supercomplex dimers in crystalline PS-II arrays (*top*), with modelled LHC-II trimer interactions across the stromal gap (*below*) (From Daum et al. [4]). (**c**) Atomic model of photosystem-II supercomplex [17]

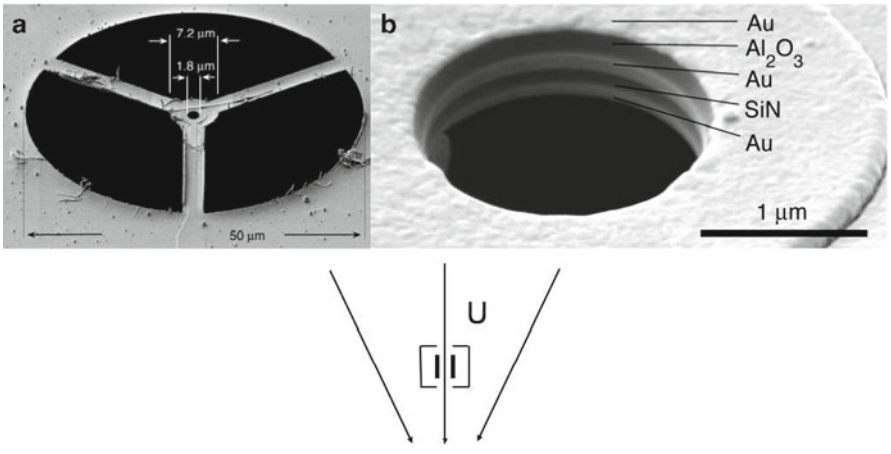


Fig. 9.8 Electrostatic Boersch phase plate (From Majorovits et al. [14])

References

1. Al-Amoudi A, Díez DC, Betts MJ, Frangakis AS (2007) The molecular architecture of cadherins in native epidermal desmosomes. Nature 450:832–837
2. Appel M, Hizlan D, Vinothkumar KR, Ziegler C, Kühlbrandt W (2009) Conformations of NhaA, the Na^+/H^+ exchanger from *Escherichia coli*, in the pH- activated and ion-translocating states. J Mol Biol 388:659–672
3. Danev R, Nagayama K (2008) Single particle analysis based on Zernike phase contrast transmission electron microscopy. J Struct Biol 161:211–218
4. Daum B, Nicastro D, Austin J II, McIntosh JR, Kühlbrandt W (2010) Arrangement of photosystem-II and ATP synthase in chloroplast membranes of spinach and pea. Plant Cell 22:1299–1312
5. Deisenhofer J, Epp O, Miki K, Huber R, Michel H (1984) X-ray structure analysis of a membrane protein complex electron-density map 3 Å resolution and a model of the chromophores of the photosynthetic reaction centre from *Rhodopseudomonas viridis*. J Mol Biol 180:385–398
6. Efremov RG, Baradaran R, Sazanov LA (2010) The architecture of respiratory complex I. Nature 465: 441–445, 27 May 2010, doi:10.1038/nature09066
7. Gipson P, Mills D, Wouts R, Grininger M, Vonck J, Kühlbrandt W (2010) Direct structural insight into the substrate-shuttling mechanism of yeast fatty acid synthase by electron cryomicroscopy. Proc Natl Acad Sci 107:9164–9169
8. Goswami P, Paulino C, Hizlan D, Vonck J, Yildiz Ö, Kühlbrandt W (2011) Structure of the archaeal Na+/H+ antiporter NhaP1 and functional role of transmembrane helix 1. EMBO J 30:439–449
9. Grigorieff N, Ceska TA, Downing KH, Baldwin JM, Henderson R (1996) Electron-crystallographic refinement of the structure of bacteriorhodopsin. J Mol Biol 259:393–421
10. Hite RK, Li Z, Walz T (2010) Principles of membrane protein interactions with annular lipids deduced from aquaporin-0 2D crystals. EMBO J 29:1652–1658
11. Hunte C, Screpanti E, Venturi M, Rimon A, Padan E, Michel H (2005) Structure of a Na^+/H^+ antiporter and insights into mechanism of action and regulation by pH. Nature 435: 1197–1202
12. Jiang W, Baker ML, Jakana J, Weigele PR, King J, Chiu W (2008) Backbone structure of the infectious e15 virus capsid revealed by electron cryomicroscopy. Nature 451:1130–1135
13. Kühlbrandt W, Wang DN, Fujiyoshi Y (1994) Atomic model of plant light-harvesting complex by electron crystallography. Nature 367:614–621
14. Majorovits E, Barton B, Schultheiss K, Pérez-Willard F, Gerthsen D, Schröder RR (2007) Optimizing phase contrast in transmission electron microscopy with an electrostatic (Boersch) phase plate. Ultramicroscopy 107:213–226
15. McMullan G, Faruqi AR, Henderson R, Guerrini N, Turchetta R, Jacobs A, van Hoften G (2009) Experimental observation of the improvement in MTF from backthinning a CMOS direct electron detector. Ultramicroscopy 109:1144–1147
16. Model K, Meisinger C, Kühlbrandt W (2008) Cryo-electron microscopy structure of a yeast mitochondrial preprotein translocase. J Mol Biol 383:1049–1057
17. Nield J, Barber J (2006) Refinement of the structural model for the Photosystem II supercomplex of higher plants. Biochim Biophys Acta 1757:353–361
18. Rubinstein JL, Walker JE, Henderson R (2003) Structure of the mitochondrial ATP synthase by electron cryomicroscopy. EMBO J 22:6182–6192
19. Schäfer E, Dencher NA, Vonck J, Parcej DN (2007) Three-dimensional structure of the respiratory chain supercomplex $I_1III_2IV_1$ from bovine heart mitochondria. Biochemistry 46:12579–12585
20. Schulze S, Köster S, Geldmacher U, Terwisscha van Scheltinga AC, Kühlbrandt W (2010) Structural basis of Na^+-independent and cooperative substrate/product antiport in CaiT. Nature 467:233–236

21. Strauss M, Hofhaus G, Schröder RR, Kühlbrandt W (2008) Dimer ribbons of ATP synthase shape the inner mitochondrial membrane. EMBO J 27:1154–1160
22. Vinothkumar KR, Smits SHJ, Kühlbrandt W (2005) pH-induced structural change in a sodium/proton antiporter from *Methanococcus jannaschii*. EMBO J 24:2720–2729
23. Vonck J, Pisa KY, Morgner N, Brutschy B, Müller V (2009) Three-dimensional structure of A_1A_0 ATP synthase from the hyperthermophilic archaeon *Pyrococcus furiosus* by electron microscopy. J Biol Chem 284:10110–10119
24. Williams KA (2000) Three-dimensional structure of the ion-coupled transport protein NhaA. Nature 403:112–115
25. Zhang X, Jin L, Fang Q, Hui WH, Zhou ZH (2010) 3.3 Å cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry. Cell 141:472–482

Chapter 10
Molecular Mechanisms of DNA Polymerase Clamp Loaders

Brian Kelch, Debora Makino, Kyle Simonetta, Mike O'Donnell, and John Kuriyan

Abstract Clamp loaders are ATP-driven multiprotein machines that couple ATP hydrolysis to the opening and closing of a circular protein ring around DNA. This ring-shaped clamp slides along DNA, and interacts with numerous proteins involved in DNA replication, DNA repair and cell cycle control. Recently determined structures of clamp loader complexes from prokaryotic and eukaryotic DNA polymerases have revealed exciting new details of how these complex AAA+ machines perform this essential clamp loading function. This review serves as background to John Kuriyan's lecture at the 2010 Erice School, and is not meant as a comprehensive review of the contributions of the many scientists who have advanced this field. These lecture notes are derived from recent reviews and research papers from our groups.

10.1 Background

Cellular chromosomal replicases are multiprotein DNA polymerase assemblies that replicate DNA with very high processivity during cell division (reviewed in [1, 2]). Their tight grip on DNA derives from a **sliding clamp** protein, a ring-shaped

B. Kelch • D. Makino • K. Simonetta
Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA

M. O'Donnell
Laboratory of DNA Replication, Howard Hughes Medical Institute,
The Rockefeller University, New York, USA

J. Kuriyan (✉)
Department of Molecular and Cell Biology and Chemistry, California Institute for Quantitative Biosciences, Howard Hughes Medical Institute,
University of California, Berkeley

Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA
e-mail: kuriyan@berkeley.edu

M.A. Carrondo and P. Spadon (eds.), *Macromolecular Crystallography*,
NATO Science for Peace and Security Series A: Chemistry and Biology,
DOI 10.1007/978-94-007-2530-0_10, © Springer Science+Business Media B.V. 2012

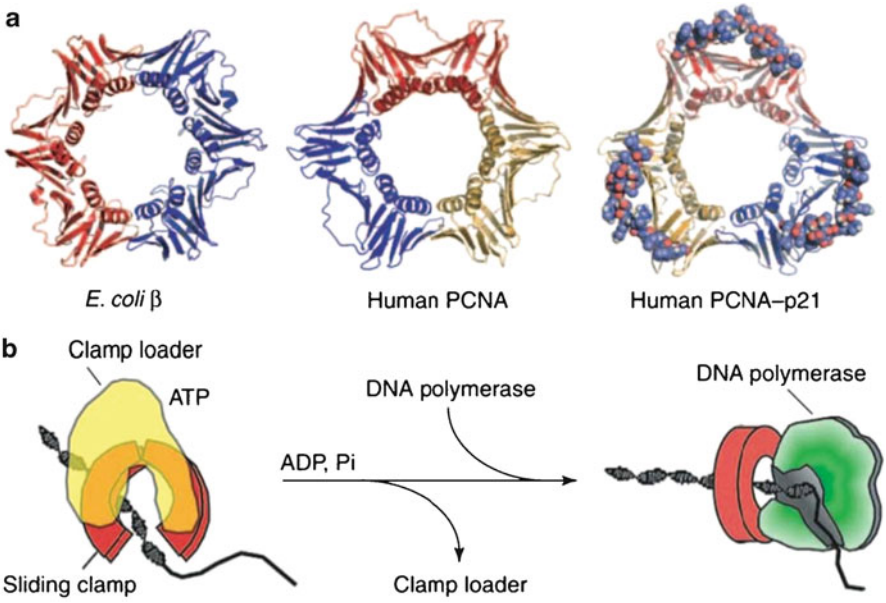


Fig. 10.1 Sliding clamps and clamp loaders. (**a**) *E. coli* β (*left*; PDB code 2POL), human PCNA with bound p21 removed (*middle*) and human PCNA bound to a C-terminal peptide of p21 (*right*; PDB code 1AXC for both PCNA structures shown). p21 binds to PCNA and serves as the prototype for how proteins bind to sliding clamps (Adapted from [3, 4]). (**b**) The clamp loader uses ATP to load the sliding clamp on DNA (Adapted from [5])

homooligomer that encircles DNA and slides freely along the duplex. Representative structures of these clamps are shown in Fig. 10.1a for *Escherichia coli* β and human proliferating cell nuclear antigen (PCNA) [3, 4]. Whereas the prokaryotic clamp is a homodimer, eukaryotic PCNA is a homotrimer. The clamps have a similar appearance, as their architecture is based on a domain of common structure.

The sliding clamps are required at primer–template junctions, where they bind to the polymerase and continuously hold it to the DNA during chain extension (Fig. 10.1b). Clamps are repeatedly placed at new RNA-primed sites on the lagging strand, where they recruit the polymerase subunits for numerous initiation events during discontinuous replication [6]. The task of placing the clamp at a primed site is accomplished by a multi-protein **clamp loader complex**, which recognizes the junction of single-stranded (ss) and double-stranded (ds) DNA in a structure-specific fashion. Despite the symmetrical appearance of the clamp, the two flat faces of the ring are structurally distinct. The polymerase associates with the face from which the C-termini protrude in both prokaryotic and eukaryotic systems [7, 8]. Therefore, the clamp loader must correctly orient the clamp on the 3′ terminus of the primed site for interaction with the DNA polymerase. Interestingly, clamp loaders from both prokaryotic and eukaryotic systems interact with the same face of their respective

clamp as the DNA polymerase; thus, the clamp loader must depart from the clamp for the polymerase to function [9, 10].

The first detailed view of how proteins bind clamps was provided by the structure of PCNA bound to a peptide of the p21 cell cycle regulator [4]. p21 binds the C-terminal face of PCNA in a hydrophobic pocket between the two domains (Fig. 10.1a). It is now apparent that many different polymerases, regulators and repair proteins bind PCNA and use the same binding pocket [11–14]. An analogous binding pocket exists in β [15–17] and gp45, the T4 clamp [18].

This review provides an overview of clamp loader structure and function as it is currently understood in both prokaryotes and eukaryotes. Overall, they are strikingly similar and lessons learned from one generally apply to the other.

10.2 The *E. coli* Clamp Loader

Clamp loader subunits, both prokaryotic and eukaryotic, are members of the **AAA+** (triple A+) family [19]. It turns out that the first crystal structure of a AAA+ protein was that of the *E. coli* δ′ clamp loader subunit, although we did not know that it was a AAA+ protein when we determined that structure (Fig. 10.2a) [22–24]. This particular subunit is not an active ATPase and the details of ATP binding by AAA+ proteins was first revealed by structures of the NSF/p97 family from the groups of Bill Weis and Axel Brünger. John Kuriyan first realized that clamp loaders were AAA+ proteins when he was shown the structure of NSF by Bill Weis in a coffee shop on Powell Street in San Francisco, and he was amazed to see a protein fold in NSF very similar to that determined for the δ′ clamp loader subunit. Andy Neuwald [19] established the relationship between AAA+ proteins and clamp loaders by using sequence comparisons.

Structures of intact clamp loaders reveal five core subunits arranged in a circle, as illustrated in Fig. 10.2b for the *E. coli* $\gamma_3\delta\delta'$ (**γ complex**) clamp loader [21]. The subunits of the γ complex are designated A–E in Fig. 10.2c for convenience in comparing it to the eukaryotic RFC clamp loader [25]. The intact *E. coli* clamp loader consists of the γ complex and also two additional small subunits (χ and ψ), which are not required for clamp loading. The structure of the χ-ψ subassembly is unrelated to that of the other subunits [26]. Each of the γ, δ and δ′ subunits consists of three domains: The ATP-binding AAA+ region of homology is located in the two N-terminal domains, whereas the C-terminal domain forms the major intersubunit connections that hold the pentamer together (Fig. 10.2a). Oligomers of AAA+ proteins typically consist of six identical subunits that form a symmetrical circular disk [27, 28]. The circular pentameric clamp loaders are unique because they contain a gap instead of a sixth subunit (Fig. 10.2b). The C-terminal domains form a tight uninterrupted ‘collar’, but there is a prominent gap between the AAA+ domains of δ and δ′ (subunits A and E). Recent experiments indicate that the gap

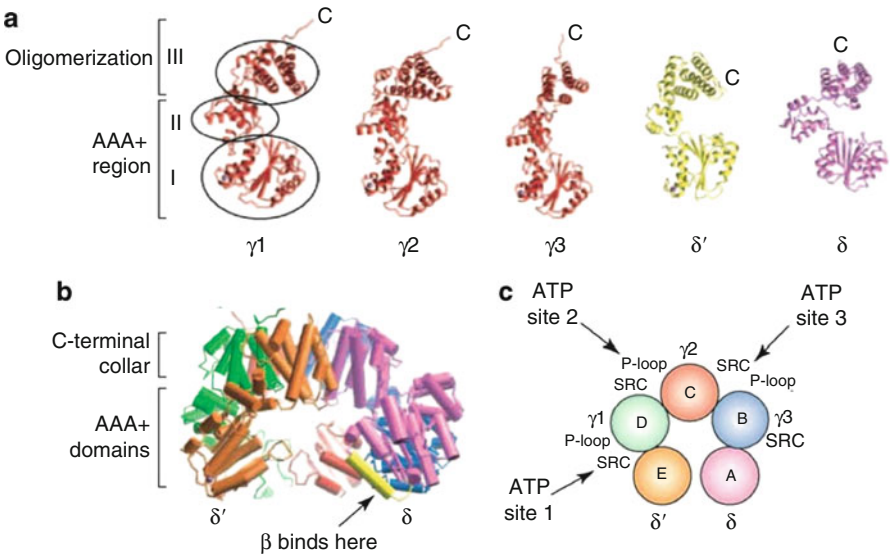


Fig. 10.2 The *E. coli* γ complex clamp loader. (**a**) Individual subunits of $\gamma_3\delta\delta'$ are displayed side by side [20]. (**b**) The $\gamma_3\delta\delta'$ minimal clamp loader structure (side view; PDB code 1JR3) (Adapted from [21]). The *yellow-colored helix* in δ interacts with β [15]. (**c**) The A–E nomenclature of the subunits. ATP binds only the three γ subunits, at sites located at subunit interfaces

remains open throughout the reaction and is present to allow DNA entry into the open clamp (described later) [29].

Most if not all of the γ complex subunits interact with β [30]. However, the tightest contact to β occurs through the δ subunit, which acts as a wrench to destabilize the β dimer interface [15, 31]. To bind the clamp, the γ complex must first bind ATP, which promotes a conformational change [32]. The γ complex crystal structure lacks nucleotide and thus is in the inactive conformation for clamp binding. Only the γ subunits bind ATP, as the ATP sites of the δ and δ′ subunits are degenerate. Typical of AAA+ oligomers, the ATP sites are positioned at subunit interfaces [27, 28]. In particular, an arginine finger residue located in a conserved SRC motif reaches across the interface to participate in the ATP site of the adjacent subunit. The strategic location of ATP sites at subunit interfaces of AAA+ oligomers may enable global conformational changes in response to ATP binding and hydrolysis [27, 28]. Site-specific mutants of the arginine finger of γ and δ′ confirm its importance in ATP sensing and hydrolysis [33, 34].

ATP sites 1 and 3 in the γ complex appear open and accessible to ATP binding, but ATP site 2 is 'squeezed' shut by amino acid sidechains from subunit B that occlude the ATP-binding pocket in subunit C. Interestingly, the structure of γ complex in the presence of the nonhydrolyzable ATP analog ATPγS shows that only sites 1 and 3 are occupied; ATP site 2 remains unfilled and the complex retains nearly the same conformation as the unliganded structure, suggesting that β may be required to assist the binding of the third ATP [35].

10.3 The Eukaryotic Clamp Loader

The eukaryotic clamp loader, replication factor C (RFC), was first identified by its requirement in SV40 replication in vitro [36]. RFC consists of five non-identical subunits, four of which have a consensus P-loop [37]. Four of the RFC subunits are similar in size to the γ3δδ′ subunits, whereas the A subunit (RFC1) is over twice as large and contains N- and C-terminal extensions. The structure of *Saccharomyces cerevisiae* RFC in complex with PCNA and ATPγS (Fig. 10.3a) yields insight into the nature of the ATP-induced conformational change and reveals how the clamp loader recognizes a primer–template junction [25, 38].

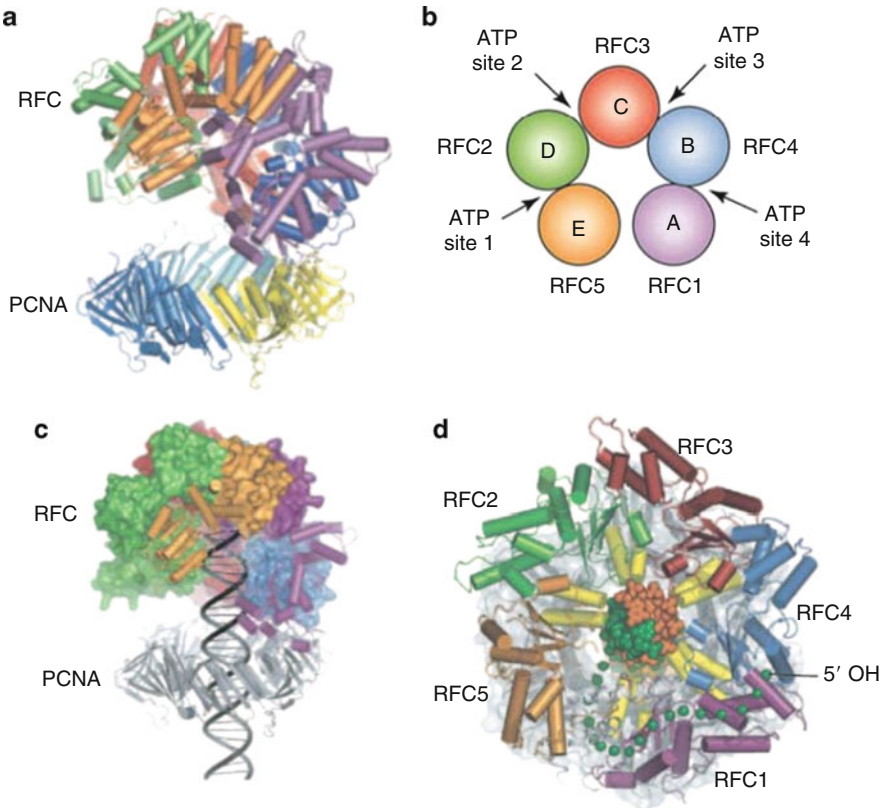


Fig. 10.3 The S. cerevisiae RFC clamp loader. (**a**) Structure of yeast RFC bound to PCNA and ATPγS (PDB code 1SXJ). (**b**) Cartoon of the subunit arrangement and ATP sites of the RFC pentamer. (**c**) Model of DNA posi-tioned within the RFC–PCNA structure. (**d**) Top view of the RFC–PCNA–DNA model, with the C-terminal collar. PCNA is in *gray*, and DNA is in *green* and *orange*. The helices that track DNA are highlighted in *yellow*. *Green spheres* indicate a possible exit path for template ssDNA from the central chamber (Adapted from [25])

The RFC pentamer contains four ATP sites (Fig. 10.3b). The additional ATP site in RFC (relative to $\gamma_3\delta\delta'$) is in RFC1 (subunit A). Biochemical studies of RFC P-loop mutants indicate that the RFC1 ATP site is not required for clamp loading and its role remains unclear [39]. The RFC structure unexpectedly shows five bound nucleotides, even though RFC5 (in position E, corresponding to *E. coli* δ′) lacks a consensus P-loop. The gap between subunits A and E is present, and the C-terminal extension of RFC1 folds back and packs against subunit E.

The PCNA clamp is attached to the A, B and C subunits; the D and E subunits lift off the plane of the ring and the PCNA ring is closed (Fig. 10.3a). In common with the *E. coli* γ complex, biochemical studies have demonstrated that ATPγS induces RFC to bind PCNA and open its ring for DNA binding; hydrolysis is required for RFC to eject from PCNA, thereby enabling DNA polymerases to function with the sliding clamp [40]. Hence, it is somewhat surprising that the PCNA ring is closed in the RFC– PCNA–ATPγS structure. This dilemma may be explained by the fact that the structural study utilized an RFC mutant in which the arginine fingers in the SRC motifs of subunits A–D were mutated to SQC to prevent the hydrolysis of ATPγS, which could occur during the time-frame of crystal growth. The SRC to SQC mutations in subunits A–D of the RFC–PCNA–ATPγS structure may tilt the equilibrium towards a form in which PCNA is closed, thus severing the connection between PCNA and subunits D and E. This form is proposed to represent an intermediate in which PCNA is closed around DNA, just upstream of the ATP hydrolysis and RFC ejection step [25]. Hydrolysis of ATP would complete the detachment of RFC from PCNA and DNA (explained further in the section below and Fig. 10.5).

The orientations of the five ATPγS in the RFC–PCNA structure are related by a set of similar screw operations about a central axis; this results in a spiral architecture with a pitch that closely matches B-form DNA [25, 38]. Furthermore, the RFC AAA+ domains define a central cavity into which duplex DNA may snugly fit and the size of the footprint corresponds to earlier footprinting studies of the RFC–DNA complex [41] (Fig. 10.3c). Hence, it was hypothesized that the conformation induced by nucleotide binding, perhaps assisted by the clamp, brings the subunits into a helical arrangement to fit around DNA and track the duplex (Fig. 10.3d). A helical disposition of subunits within an oligomer that encircles DNA is also observed in certain proteins based on the RecA ATPase fold, including T7 helicase, Rho and RecA [42–44].

Figure 10.3d shows a cut-away view looking down on the top of an RFC–PCNA complex in which the C-terminal domains that form the clamp loader collar have been removed and the DNA has been modeled through PCNA and into the central chamber of RFC. Each subunit contains a pair of helices for which the positive dipole is directed toward the central chamber. Minor groove interactions have been demonstrated to mediate the interaction of protein with DNA in non-sequence-specific fashion [45]; when DNA is modeled inside the RFC chamber, it can be positioned such that these helices track the minor groove (yellow helices in Fig. 10.3d). Conserved positively charged and polar residues exist on and near these helices, several of which are also conserved in prokaryotic

clamp loader subunits. As a test for DNA binding inside the clamp loader, advantage has been taken of the prokaryotic clamp loader, which contains three identical copies of γ [46]. Thus, any given mutation in γ is repeated three times in γ complex. Biochemical studies of γ complex with mutated putative DNA-binding residues in γ and/or δ′ support the central chamber as the locus of DNA binding [46].

10.4 Primed Template Recognition and Insertion of DNA into the Clamp

Clamp loaders must open the clamp, recognize a primed template and then insert the DNA through the center of the clamp. Studies in the T4 system indicate that the gp45 clamp may open to form a left-handed helix [47]. Recent molecular simulation studies of yeast PCNA indicate that, upon destabilizing the clamp interface, the PCNA ring springs open into a right-handed spiral [48]. EM reconstruction studies of an archaeal RFC–PCNA–DNA–ATPγS complex by Kosuke Moriakawa and colleagues indicate that PCNA has an open lock-washer appearance, enabling it to extensively dock on the helical surface of the RFC pentamer [49]. These observations are summarized in Fig. 10.5a, wherein the open clamp docks underneath the clamp loader and adopts a complementary spiral to that of the AAA+ domains of the clamp loader subunits. This helical open form allows the clamp to interact with all the clamp loader subunits.

A recent structure of the *E.coli* clamp loader in complex with primer template DNA has illuminated the mechanism of ATP-dependent DNA recognition by the clamp loader [50]. DNA enters the central chamber of the clamp loader through the gap between subunits A and E (e.g. Fig. 10.4a). The tight pentameric contacts comprising the collar provide no opening, thus presenting an obstacle to the fit of a rigid duplex straight through the clamp–clamp loader complex [50]. A conserved Tyr residue stacks on the final base in the primer strand, resulting in a sharp bend of the template strand as it exits the clamp loader in a manner reminiscent of the ‘separation pin’ of UvrB helicase [51]. The ssDNA portion of a primed template bends out of the gap between subunits A and E and is bound into a surface groove along the A subunit (δ).

Additionally, ATP binding causes a symmetric arrangement of B, C, D, and E subunits around the longitudinal axis of the DNA [50] (Fig. 10.4b), thus tracking the helical spiral of DNA, as hypothesized from the RFC-PCNA structure [25]. Surprisingly, the clamp loader only contacts the template strand, which allows both RNA and DNA hybrids to be accommodated in the central chamber. Additionally, the non-hydrolyzable ATP analogs (ADP•BeF_3) in the structure seem to be positioned for hydrolysis and subsequent release of the DNA. Indeed, recent biochemical studies have shown that the clamp loader can bind either DNA with a 3′ or 5′ recessed end, but that efficient ATP hydrolysis and loading of the clamp is only facilitated with DNA with the proper 3′ recessed end [52]. Therefore,

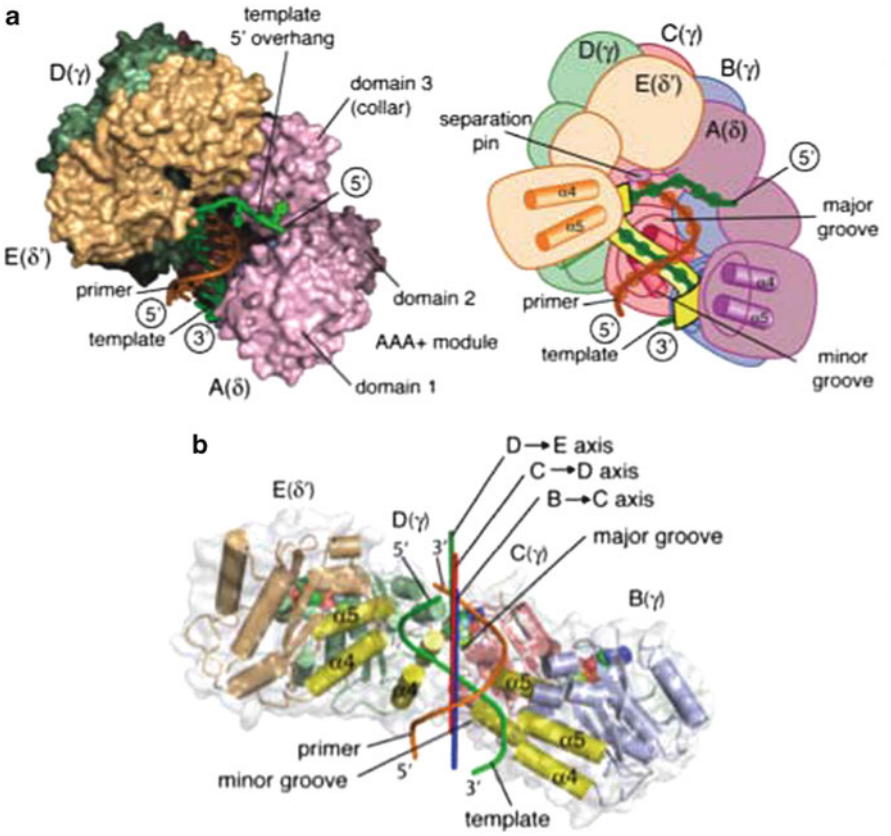


Fig. 10.4 DNA recognition by the *E.coli* clamp loader (**a**) γ complex bound to primer-template DNA (PDB code: 3GLF). (**b**) Diagram showing the ATPase subunits of the clamp loader and DNA duplex. The DNA-interacting helices are shown in yellow. The three rotation axes that relate the B subunit to the C subunit, the C subunit to the D subunit, and the D subunit to the E subunit are shown in *blue*, *red*, and *green*, respectively. The three axes are nearly coincident with each other and with the axis of the DNA duplex (not shown) (Adapted from [50])

discrimination of the minor groove, and thus proper DNA architecture, must be achieved at the ATP hydrolysis step.

Recent work has also highlighted the role of accessory factors in the clamp loading process. Structures of $\gamma_3\delta\delta'$ in complex with a peptide from the ψ subunit show that this peptide increases affinity of clamp loader for DNA via a rearrangement of the clamp loader that facilitates its binding to DNA [50]. In addition, recent structural and biochemical studies of both β and PCNA have demonstrated that the clamp itself assists in the clamp loading process and, once loaded, sits in a tilted orientation relative to the DNA [52–54]. The tilted orientation suggests that DNA may switch between various factors that are bound to the clamp simultaneously.

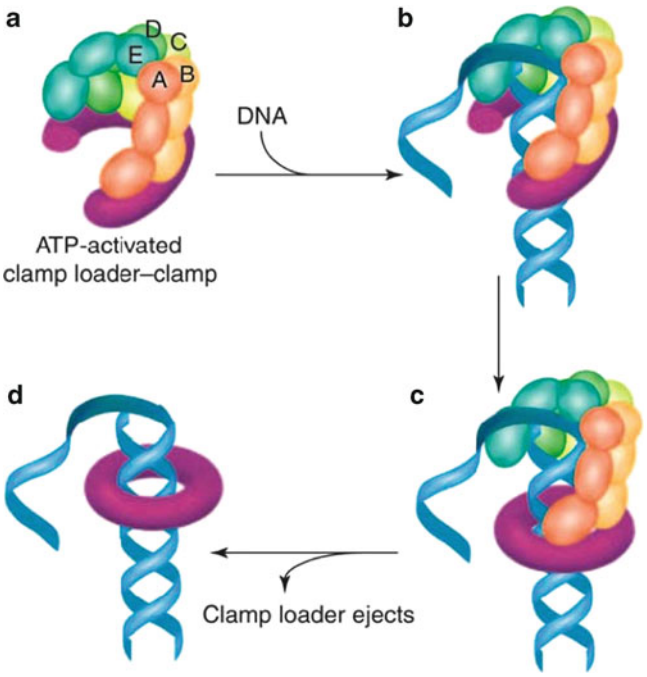


Fig. 10.5 Clamp opening and DNA recognition. (**a**) The open clamp is illustrated as a right-handed lock-washer that fits on the helical surface defined by the five AAA+ domains of the ATP-activated clamp loader. (**b**) Primed DNA enters through the gap in the open ring and the slot between the A and E clamp loader subunits. Template ssDNA provides a flexible joint for DNA exit from the central chamber through the slot in the side of the clamp loader. (**c**) The ring snaps shut around DNA, severing the tie between the clamp and the D and E clamp loader subunits. (**d**) ATP hydrolysis ejects the clamp loader from the clamp–DNA complex, leaving the clamp topologically linked to the DNA (Adapted from [2])

When the clamp closes around DNA, it will become a planar ring and therefore may detach from some subunits of the clamp loader spiral (Fig. 10.5). This proposed intermediate may correspond to the structure captured in the RFC–PCNA–ATPγS crystal, in which subunits A, B and C interact with the closed clamp, while the ATP sites are poised for hydrolysis [25]. Hydrolysis ejects the clamp loader, thus enabling DNA polymerase to access the clamp (Fig. 10.5).

10.5 Concluding Remarks

Important new questions arise from recent studies on prokaryotic, eukaryotic and archaeal clamp loaders. For example, the precise coordination of DNA and clamp binding with ATP-hydrolysis remains uncertain. Is the clamp actively 'cracked open' by the clamp loader? Alternatively, the clamp may spontaneously breathe and

the clamp loader may trap the clamp in the open form. Also, there exist alternative RFC complexes in which subunit A (RFC1) is replaced by another subunit. For example, RFC1 is replaced by a Rad protein (Rad24 in *S. cerevisiae*, Rad17 in humans) in response to DNA damage [55]. Rad–RFC loads the clamp (the 911 clamp, which is different from PCNA) onto the 5′ end of a primed site and thus recognizes the opposite side of a single-stranded/double-stranded junction compared to RFC–PCNA [56]. How does this alternate A subunit cause the clamp loader to recognize DNA of opposite polarity?

Although not covered in this review, both prokaryotic and eukaryotic clamps interact with a wide variety of DNA polymerases, repair factors and cell cycle regulators. How these various factors coordinate their action on the clamp can be added to the list of the many exciting questions that remain for future studies.

References

1. Waga S, Stillman B (1998) The DNA replication fork in eukaryotic cells. Annu Rev Biochem 67:721–751
2. Johnson A, O'Donnell M (2005) Cellular DNA replicases: components and dynamics at the replication fork. Annu Rev Biochem 74:283–315
3. Kong XP et al (1992) Three-dimensional structure of the beta subunit of *E. coli* DNA polymerase III holoenzyme: a sliding DNA clamp. Cell 69:425–437
4. Gulbis JM et al (1996) Structure of the C-terminal region of p21(WAF1/CIP1) complexed with human PCNA. Cell 87:297–306
5. Lopez de Saro F (2004) Protein trafficking on sliding clamps. Philos Trans R Soc Lond B Biol Sci 359:25–30
6. Stukenberg PT et al (1994) An explanation for lagging strand replication: Polymerase hopping among DNA sliding clamps. Cell 78:877–887
7. Naktinis V et al (1996) A molecular switch in a replication machine defined by an internal competition for protein rings. Cell 84:137–145
8. Johansson E et al (2004) The Pol32 subunit of DNA polymerase delta contains separable domains for processive replication and profliferating cell nuclear antigen (PCNA) binding. J Biol Chem 279:1907–1915
9. Ason B et al (2003) Mechanism of loading the *Escherichia coli* DNA polymerase III β sliding clamp on DNA: *bona fide* primer/templates preferentially trigger the γ complex to hydrolyze ATP and load the clamp. J Biol Chem 278:10033–10040
10. Gomes XV et al (2001) ATP utilization by yeast replication factor C: II. Multiple stepwise ATP binding events are required to load proliferating cell nuclear antigen onto primed DNA. J Biol Chem 276:34776–34783
11. Sakurai S et al (2005) Structural basis for recruitment of human flap endonuclease 1 to PCNA. EMBO J 24:683–693
12. Chapados BR et al (2004) Structural basis for FEN-1 substrate specificity and PCNA-mediated activation in DNA replication and repair. Cell 116:39–50
13. Warbrick E (1998) PCNA binding through a conserved motif. Bioessays 20:195–199
14. Matsumiya S et al (2001) Crystal structure of an archaeal DNA sliding clamp: proliferating cell nuclear antigen from *Pyrococcus furiosus*. Protein Sci 10:17–23
15. Jeruzalmi D et al (2001) Mechanism of processivity clamp opening by the delta subunit wrench of the clamp loader complex of *E. coli* DNA polymerase III. Cell 106:417–428
16. Bunting KA et al (2003) Structural basis for recruitment of translesion DNA polymerase Pol IV/DinB to the β-clamp. EMBO J 22:5883–5892

17. Burnouf DY et al (2004) Structural and biochemical analysis of sliding clamp/ligand interactions suggest a competition between replicative and translesion DNA polymerases. J Mol Biol 335:1187–1197
18. Shamoo Y, Steitz TA (1999) Building a replisome from interacting pieces: sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. Cell 99: 155–166
19. Neuwald AF et al (1999) AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. Genome Res 9:27–43
20. Davey MJ et al (2002) Motors and switches: AAA+ machines within the replisome. Nat Rev Mol Cell Biol 3:826–835
21. Jeruzalmi D et al (2001) Crystal structure of the processivity clamp loader gamma (g) complex of *E. coli* DNA polymerase III. Cell 106:429–441
22. Guenther B et al (1997) Crystal structure of the δ′ subunit of the clamp-loader complex of *E. coli* DNA polymerase III. Cell 91:335–345
23. Oyama T et al (2001) Atomic structure of the clamp loader small subunit from Pyrococcus furiosus. Mol Cell 8:455–463
24. Podobnik M et al (2003) Nucleotide-induced conformational changes in an isolated *Escherichia coli* DNA polymerase III clamp loader subunit. Structure (Camb) 11:253–263
25. Bowman GD et al (2004) Structural analysis of a eukaryotic sliding DNA clamp–clamp loader complex. Nature 429:724–730
26. Gulbis JM et al (2004) Crystal structure of the chi:psi sub-assembly of the *Escherichia coli* DNA polymerase clamp-loader complex. Eur J Biochem 271:439–449
27. Yu RC et al (1998) Structure of the ATP-dependent oligomerization domain of N-ethylmaleimide sensitive factor complexed with ATP. Nat Struct Biol 5:803–811
28. Lenzen CU et al (1998) Crystal structure of the hexamerization domain of N-ethylmaleimide-sensitive fusion protein. Cell 94:525–536
29. Goedken ER et al (2004) Fluorescence measurements on the *E. coli* DNA polymerase clamp loader: implications for conformational changes during ATP and clamp binding. J Mol Biol 336:1047–1059
30. Leu FP, O'Donnell M (2001) Interplay of clamp loader subunits in opening the beta sliding clamp *Escherichia coli* DNA polymerase III holoenzyme. J Biol Chem 276:47185–47194
31. Stewart J et al (2001) Mechanism of β clamp opening by the δ subunit of *Escherichia coli* DNA polymerase III holoenzyme. J Biol Chem 276:19182–19189
32. Naktinis V et al (1995) Assembly of a chromosomal replication machine: two DNA polymerases, a clamp loader, and sliding clamps in one holoenzyme particle. J Biol Chem 270: 13358–13365
33. Johnson A, O'Donnell M (2003) Ordered ATP hydrolysis in the γ complex clamp loader AAA+ machine. J Biol Chem 278:14406–14413
34. Snyder AK et al (2004) Mechanism of loading the *Escherichia coli* DNA polymerase III sliding clamp: II. Uncoupling the β and DNA binding activities of the γ complex. J Biol Chem 279:4386–4394
35. Kazmirski SL et al (2004) Structural analysis of the inactive state of the *Escherichia coli* DNA polymerase clamp-loader complex. Proc Natl Acad Sci USA 101:16750–16755
36. Fien K, Stillman B (1992) Identification of replication factor C from Saccharomyces cerevisiae: a component of the leading-strand DNA replication complex. Mol Cell Biol 12:155–163
37. Cullmann G et al (1995) Characterization of the five replication factor C genes of Saccharomyces cerevisiae. Mol Cell Biol 15:4661–4671
38. Bowman G et al (2005) DNA polymerase clamp loaders and DNA recognition. FEBS Lett 579:863–867
39. Schmidt SL et al (2001) ATP utilization by yeast replication factor C: III. The ATP-binding domains of Rfc2, Rfc3, and Rfc4 are essential for DNA recognition and clamp loading. J Biol Chem 276:34784–34791
40. Tsurimoto T, Stillman B (1990) Functions of replication factor C and proliferating cell nuclear antigen: functional similarity of DNA polymerase accessory proteins from human cells and bacteriophage T4. Proc Natl Acad Sci USA 87:1023–1027

41. Tsurimoto T, Stillman B (1991) Replication factors required for SV40 DNA replication in vitro. I. DNA structure-specific recognition of a primer-template junction by eukaryotic DNA polymerases and their accessory proteins. J Biol Chem 266:1950–1960
42. Story RM et al (1992) The structure of the *E. coli recA* protein monomer and polymer. Nature 355:318–325
43. Singleton MR et al (2000) Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. Cell 101:589–600
44. Skordalakes E, Berger JM (2003) Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading. Cell 114:135–146
45. Murphy FV IV, Churchill ME (2000) Nonsequence-specific DNA recognition: a structural perspective. Struct Fold Des 8:R83–R89
46. Goedken E et al (2005) Mapping the interaction of DNA with the *Escherichia coli* DNA polymerase clamp loader complex. Nat Struct Mol Biol 12:183–190
47. Trakselis MA et al (2001) Creating a dynamic picture of the sliding clamp during T4 DNA polymerase holoenzyme assembly by using fluorescence resonance energy transfer. Proc Natl Acad Sci USA 98:8368–8375
48. Kazmirski S et al (2005) Out-of-plane motions in open sliding clamps: Molecular dynamics simulations of eukaryotic and archaeal proliferating cell nuclear antigen. Proc Natl Acad Sci USA 102:13801–13806
49. Miyata T et al (2005) Open clamp structure in the clamp-loading complex visualized by electron microscopic image analysis. Proc Natl Acad Sci USA 102:13795–13800
50. Simonetta K et al (2009) The mechanism of ATP-dependent primer-template recognition by a clamp loader complex. Cell 137:659–671
51. Lee JY, Yang W (2006) UvrD helicase unwinds DNA one base pair at a time by a two-part power stroke. Cell 127:1349–1360
52. Park M, O'Donnell M (2009) The clamp loader assembles the beta clamp onto either a 3′ or 5′ primer terminus: the underlying basis favoring 3′ loading. J Biol Chem 284:31473–31483
53. Georgescu R et al (2008) Structure of a sliding clamp on DNA. Cell 132:43–54
54. McNally R et al (2010) Analysis of the role of PCNA-DNA contacts during clamp loading. BMC Struct Biol 10:3. doi:10.1186/1472-6807-10-3
55. Parrilla-Castellar ER et al (2004) Dial 9-1-1 for DNA damage: the Rad9-Hus1-Rad1 (9–1-1) clamp complex. DNA Repair (Amst) 3:1009–1014
56. Ellison V, Stillman B (2003) Biochemical characterization of DNA damage checkpoint complexes: clamp loader and clamp complexes with specificity for 5′ recessed DNA. PLoS Biol 1:E33

Chapter 11
Electron Microscopy of Macromolecular Machines

Helen R. Saibil

Abstract The field of molecular electron microscopy is developing powerful tools for visualizing the dynamics of macromolecular machines in action. This chapter briefly reviews the basics of macromolecular structure determination by electron microscopy and its application to dynamic machines.

Keywords Cryo-electron microscopy • Three-dimensional reconstruction • Structural biology

11.1 Three-Dimensional Electron Microscopy (3D EM) in Structural Biology

Since the early developments in this field in the late 1960s, 3D EM has become a major field of structural biology. Experimental and computational advances have created powerful tools for 3D structural analysis over the full range from macromolecules to cells. In the lower resolution range, EM is highly complementary to cellular imaging by optical microscopy. "Super-resolution" fluorescence microscopy methods allow the precise 3D mapping of fluorophores in living cells. In contrast, cryo-EM samples are trapped by rapid freezing, but the full 3D density is recorded. At the high resolution end of the scale, single particle EM is now capable of resolving near atomic resolution structures of complexes in solution provided they are sufficiently large (~1 MDa) and rigid [12]. X-ray crystallography covers a large size range from atoms to megaDalton complexes (for objects that can be

H.R. Saibil (✉)
Crystallography, and Institute of Structural and Molecular Biology, Birkbeck College, University of London, Malet St, London WC1E 7HX, UK
e-mail: h.saibil@mail.cryst.bbk.ac.uk

M.A. Carrondo and P. Spadon (eds.), *Macromolecular Crystallography*, NATO Science for Peace and Security Series A: Chemistry and Biology,
DOI 10.1007/978-94-007-2530-0_11, © Springer Science+Business Media B.V. 2012

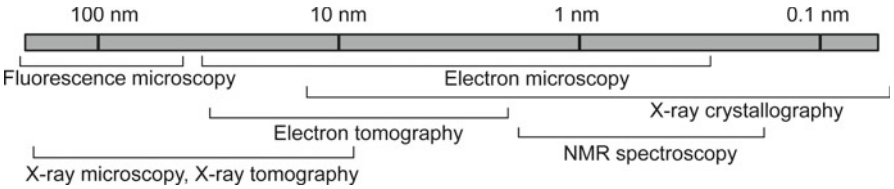


Fig. 11.1 Multiscale imaging. Approximate length scales covered by 3D structure determination methods used in biology

crystallized), but confinement of the molecules to a crystal lattice restricts their conformations. NMR spectroscopy can provide high resolution structures, mainly in solution, but its application is restricted to relatively small complexes or to selected sites in larger structures. An overview of scale lengths probed by different structural biology methods is shown in Fig. 11.1. Although molecular EM and electron tomography cover a wide range, these methods impose a limit on sample thickness – thick samples such as cells and tissues must be sectioned for EM imaging and tomography, and the preparation of sections can introduce chemical or mechanical artefacts. A developing area that overlaps with the low resolution range of EM is that of X-ray microscopy and X-ray tomography. The greater penetration power of X-rays means that whole cells can be imaged, albeit at lower resolution. Again, the full 3D density is visualized, so that this approach is also highly complementary to fluorescence imaging.

In this chapter I will briefly review the main features of 3D EM in the study of macromolecular machines.

11.2 Sample Preparation

For macromolecules, there are two main methods of sample preparation – the quicker negative stain method that gives a high contrast, but indirect view of the molecular surface, and the more demanding cryo-EM method that can ultimately provide a full density map of the atomic structure. In the negative stain approach, the solution of macromolecules is applied to an EM grid coated with a thin film of carbon, and then a solution of heavy metal salt, typically uranyl acetate, is added. The liquid is blotted to a thin layer and allowed to dry, resulting in a shell of electron-dense stain coating the molecules, which are then seen in negative contrast. This procedure often works quite well and gives a good idea of the shape and symmetry of the object. However, some samples are damaged by the heavy metal salt or collapse when dehydrated, so it is not always reliable. Cryo-EM, on the other hand, gives a low contrast, but more accurate view of the structure. The solution is applied to an EM grid with a carbon support film (which can be perforated to give regions free of support film for improved contrast), blotted and rapidly plunged into a cryogen, usually liquid ethane, for vitrification. The frozen grid

must then be kept at liquid nitrogen temperature and transferred into a cryo specimen holder for loading into the microscope, avoiding surface contamination by exposure to ambient humidity.

Cryo-EM has two important advantages. Firstly, the rapid freezing step brings the sample to the vitreous state in which the water solidifies without rearranging into ice crystals, and therefore avoids dehydration or damage to the macromolecules [5]. Vitrified samples can then be imaged in their native, hydrated state in a microscope with a cryogen-cooled specimen holder. The second major advantage of cryo-EM is that the effects of radiation damage by the electron beam are greatly slowed at low temperature. Radiation damage determines the practical limit on the resolution of biological macromolecules, as will be explained in the following section.

Unlike X-ray crystallography, EM of isolated complexes in solution, "single particles", has a lower, rather than an upper, size limit, as long as the sample thickness along the beam direction is within a few 100 nm. Large complexes are easier to analyze, since they scatter more strongly, facilitating the process of orientation determination. Cryo-EM becomes very difficult for particles less than several hundred kDa. Therefore, negative stain is normally used for smaller particles, down to 50–100 kDa.

The high symmetry of ordered assemblies such as 2D crystals or helical arrays facilitates EM analysis and makes it easier to obtain high resolution structures, using crystallographic approaches. But even for these assemblies, it is often advantageous to apply single particle methods to correct for small distortions of the lattice.

11.3 Principles of EM Image Formation and 3D Reconstruction from Projections

Transmission EM images are formed by scattering or absorption of the incident electrons by the specimen. Electrons can be scattered elastically, without energy change, or inelastically, with energy transfer between the electron and the specimen atoms. Inelastic scattering causes damage to the specimen and produces an incoherent background since the change in electron energy is equivalent to a change in wavelength, so that these electrons are focused in different planes. For thin specimens composed of light elements, such as those found in biological macromolecules, the scattering is weak and there is very little absorption. A low contrast image is formed, mainly resulting from interference between unscattered and elastically scattered electrons. This contrast is enhanced by underfocussing the objective lens, because the combination of underfocus and spherical aberration of the objective lens results in a phase shift between scattered and unscattered waves. The scattered waves pass through the outer part of the lens whereas the unscattered beam passes through the center of the lens, which has a different focal length owing to spherical aberration. The resulting phase difference enhances the image contrast.

Phase plates for EM are currently being developed (e.g. [4]), in direct analogy to those used for Zernike phase contrast in optical microscopy (now more commonly replaced by the related differential interference contrast). EM phase plates offer the potential of greatly improved contrast, especially important in cryo-electron tomography (see below).

The images recorded in a transmission electron microscope are 2D projections of the 3D density of the sample, modified by the optical characteristics of the microscope and the focus setting used. The 3D structure can be restored from a set of projections (views) of the object in different orientations relative to the incident beam, with corrections for the optical distortions. The principle of 3D reconstruction from projections is the same as in medical tomography. In medical X-ray tomography the patient is surrounded by a ring of X-ray sources and detectors, so that images are recorded at a series of angles all around the patient. In transmission EM, the beam direction remains fixed, but the specimen can be tilted to different angles to give the different views. However, EM specimens are thin and planar, so that they cannot be tilted beyond about 70° – the beam path through a highly tilted specimen becomes too long for the electrons to penetrate, and the beam is blocked by the specimen holder. Because of this limitation, electron tomograms have a missing region of information and consequently lower resolution along the beam direction than in the plane of the specimen. Another problem that limits the overall resolution is the need for many exposures of the object to collect the tilt series. For good 3D resolution, the angular step size should be as fine as possible, but for cryo-EM specimens, the accumulated radiation damage is limiting. Therefore, cryo-electron tomography data collection is a compromise between these two conflicting requirements. The limitation on electron dose means that each exposure is extremely weak, with a very low signal to noise ratio. Despite these difficulties, it is currently possible to achieve a resolution of ~4 nm in cryo-electron tomography, and this is improved if subregions of the tomogram can be extracted, aligned and averaged.

The problems of angular spacing and radiation damage are greatly reduced by the experimental strategy used for studying isolated macromolecular complexes – single particle analysis. In this approach, the complex, such as a virus particle, ribosome, or other biochemically defined assembly, is imaged in a thin layer of vitrified solution. In general the particles adopt random orientations in the vitrified layer. Thus, images of many (thousands) of particles can be combined and reconstructed in favourable cases to near atomic (~3 Å) resolution (e.g. [26]).

Once a data set of single particle images has been collected, the main challenge is to determine the relative orientations of all the particles. The main tool for alignment is cross-correlation, to give a measure of similarity between pairs of images. After in-plane alignment, the images can be classified by statistical analysis for sorting into similar subsets. Averaging within classes improves the signal to noise ratio and thus the visibility of structural features. To determine the 3D structure, the out-of-plane angular relationships between the views (Euler angles) must be deduced. There are two general approaches for Euler angle determination in the absence of an initial model, the use of common line relationships, or collection of tilted and untilted views of the same object. The common line approach is based on the

principle that any two 2D projections of the same 3D object will have at least one 1D (line) projection in common, and the angle at which the common lines are found between pairs of images defines their relative Euler angles [3, 24]. In the tilt pair approach, two views are recorded of each object with a known tilt angle. With known angular relationships, the 3D structure can be determined by the random conical tilt method [21]. Once an initial map is available, the structure can be refined by cycles of cross correlation between the input images and reprojections of the current map.

For radiation-insensitive materials such as metals, a higher electron dose can be used, giving a more favourable signal to noise ratio. The short wavelength of high voltage electrons (~0.02 Å) means that features smaller than 1 Å can be resolved, despite imperfections in the electromagnetic lenses.

11.4 Methods for Analysis of Dynamic and Heterogeneous Complexes

As the experimental and computational methods have advanced, it has become increasingly evident that single particle analysis is often limited by sample heterogeneity, arising from causes such as flexibility of the molecules in solution or variable occupancy of ligands. Because of the low signal to noise ratio of cryo-EM images, it is not usually possible to interpret the image features directly, and certainly not to detect small structural variations. Statistical analysis of large data sets of single particle images is a powerful tool for extracting detailed information about the structure and its variations. Principal component analysis and related methods are used to detect significant variations and to sort the particles into homogeneous subsets for separate reconstruction. The problem is complicated because variations in orientation must be distinguished from structural variations. Several approaches have been developed [14, 19]. The ability to sort particle populations is a very powerful feature of image analysis, making it a quasi “single molecule” method, well suited to the analysis of dynamics of macromolecular machines such as GroEL (Fig. 11.2, shown in the process of folding a substrate protein, [2]) or ribosomes [6]. It can be used to determine a series of structures from a sample consisting of multiple structural states.

11.5 Resolution, Reliability and Information Obtainable

How accurate is the structural information in EM maps? For single particle and tomography structures, it is not completely straightforward to measure the resolution of the map. With diffraction data, well-defined diffraction peaks are clearly greater than the surrounding background noise, and the signal to noise ratio can be estimated as a function of spatial frequency to define the resolution of the structural

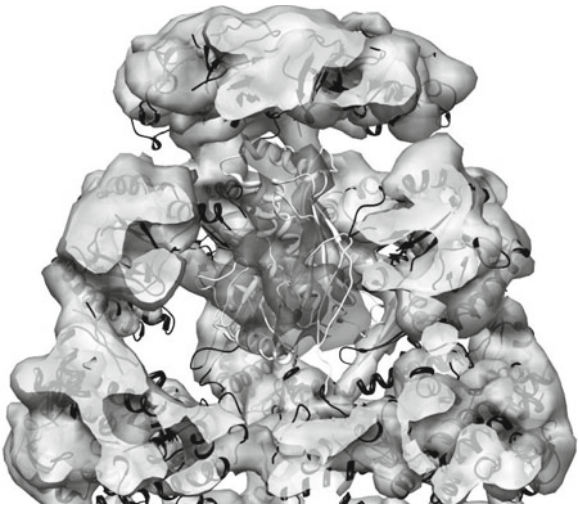


Fig. 11.2 A folding protein trapped inside the chaperonin cage. The bacteriophage T4 capsid protein gp23 requires the chaperonin complex to fold, but is near the maximum size that can be accommodated inside the folding chamber. Multivariate statistical analysis was used to sort a large set of cryo-EM images of a mixture of folding intermediates in different states into more homogeneous classes for 3D reconstruction. The map reveals an almost native, large domain of gp23 (*darker grey density*) inside the chaperonin folding chamber (*light grey density*). The atomic structures of the chaperonin domains (*black ribbons*) and of viral gp24 (a close homologue of gp23; *white ribbons*) are docked into the EM density. The front surface of the map is cut away to reveal the contents of the folding chamber. The figure was created from the EM structure (EM databank EMD-1548, [2]) and PDB entries 1OEL and 1G31

information. However, with single particle or tomography data, the signal from the structural information is continuously distributed and not easily distinguished from background noise. Therefore, a common method for resolution estimation is to divide the data set into two equivalent halves, calculate two separate reconstructions and then compare them in resolution shells. The resolution of the map is defined as the spatial frequency at which the correlation between the two reconstructions falls to a specified threshold, e.g. 50% of the maximum correlation at low resolution. There are pitfalls in map and resolution calculation, and sometimes the resolution estimate is over-optimistic because noise has been correlated with signal during the alignment procedure, or other systematic errors are present. It is necessary that the raw data and averaged views must be consistent with reprojections of the reconstruction, but this does not prove the validity of the reconstruction. Ultimately, the structure must make sense biologically and be tested against other types of experimental data, e.g. chemical labeling or mutagenesis. Often there is atomic structure information for at least part of the structure. If rigid body structures can be recognized by fitting atomic models into the EM map, this can provide validation.

Fitting of known or closely related atomic structures of components in a complex is the major tool for interpretation of low and intermediate resolution EM maps. If domain shapes or secondary structures are resolved, flexible fitting approaches can be used to allow for hinge movements (e.g. [23]). Conversely, EM maps can

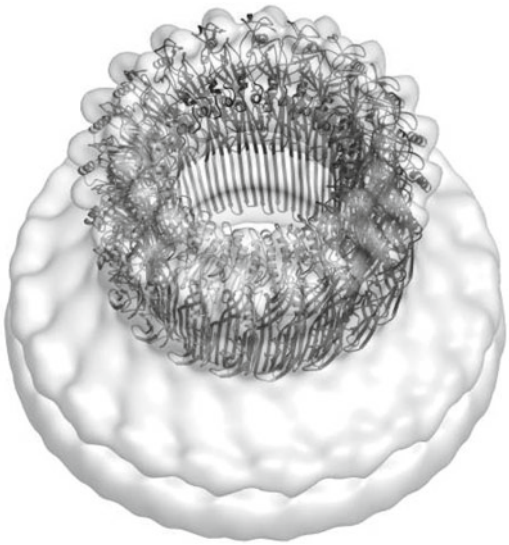


Fig. 11.3 3D reconstruction of a membrane pore formed by the immune system protein perforin. The pore contains 20 subunits, which are modeled with β-hairpins forming the pore wall, and is surrounded by a patch of the membrane bilayer. The figure was created from map EMD-1769 and atomic modeling based on the perforin crystal structure [16]. In this example, the data set was limited by the variable number of subunits in the assembly, and by the small number of isolated views of the pores distributed on liposomes. The map resolution is around 29 Å

sometimes be used as molecular replacement models to phase X-ray crystallographic data [18].

What can we expect to see in different resolution ranges? At 20–30 Å, the overall shape and perhaps domains might be resolved (Fig. 11.3 [16]). The map does not become much more interpretable until the resolution becomes better than 8–9 Å, when α-helices begin to be resolved as tubes of density. For structures rich in α-helices, reaching this threshold makes a big difference to the accuracy and reliability of atomic structure fitting (e.g. filamentous actin at 6.6 Å, Fig. 11.4 [10]), and once the resolution is better than 4.5 Å, most of the secondary structure is resolved. For the best single particle maps, in the 3–4 Å resolution range, an atomic model can be built (e.g. for a rigid icosahedral virus, Fig. 11.5, [26]).

11.6 Further Reading

Many useful review articles have appeared on this subject. Examples include those by Frank [7], van Heel et al. [25], Lucic et al. [17] and Cheng and Walz [1]. A comprehensive series of articles in three volumes of Methods in Enzymology have recently appeared [13–15] containing a substantial body of reference material on the principles, methods and examples of cryo-EM. In addition, there are excellent books on electron optics [11, 22], single particle analysis [8] and tomography [9].

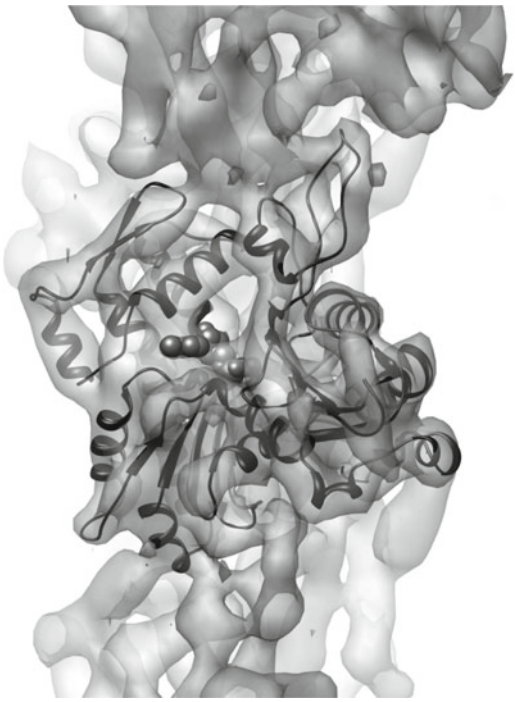


Fig. 11.4 The structure of an actin filament determined by cryo-EM and atomic structure fitting. Although the actin filament is quite flexible, a 6.6 Å resolution map was obtained by optimizing data collection and contrast, and the actin backbone was largely resolved (*black ribbons*). An ADP molecule (*spheres*) is bound in the central cleft. The figure was prepared from the map EMD-5168 and PDB entry 3MFP [10]

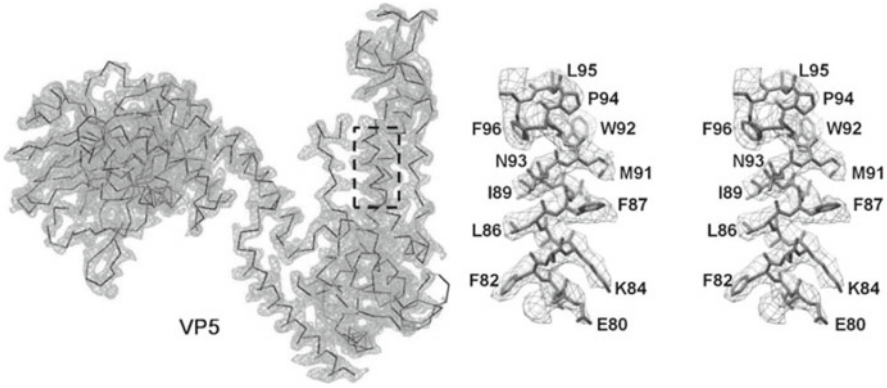


Fig. 11.5 The structure of aquareovirus determined by single particle cryo-EM at 3.3 Å resolution. The map and backbone are shown for viral subunit VP5 (*left*), and an enlarged, stereo view of the boxed region shows the resolution of side chains (*right*). Figure modified from Zhang et al. [26], with permission. The map is EMD-5160 and the atomic structure is PDB 3IYL. Copyright Elsevier 2010

References

1. Cheng Y, Walz T (2009) The advent of near atomic resolution in single particle electron microscopy. Annu Rev Biochem 78:723–742
2. Clare DK, Bakkes PJ, van Heerikhuizen H, van der Vies SM, Saibil HR (2009) Chaperonin complex with a newly folded protein encapsulated in the folding chamber. Nature 457:107–110
3. Crowther RA (1971) Procedures for three-dimensional reconstruction spherical viruses by Fourier synthesis from electron micrographs. Philos Trans R Soc Lond B Biol Sci 261:221–230
4. Danev R, Kanamaru S, Marko M, Nagayama K (2010) Zernike phase contrast cryo-electron tomography. J Struct Biol 171:174–181
5. Dubochet J, Adrian M, Chang JJ, Homo JC, Lepault J, McDowell AW, Schultz P (1988) Cryo-electron microscopy of vitrified specimens. Q Rev Biophys 21:129–228
6. Fischer N, Konevega AL, Wintermeyer W, Rodnina MV, Stark H (2010) Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. Nature 466:329–333
7. Frank J (2002) Single-particle imaging of macromolecules by cryo-electron microscopy. Annu Rev Biophys Biomol Struct 31:303–319
8. Frank J (2006) Three-dimensional electron microscopy of macromolecular assemblies. Oxford University Press, New York, USA
9. Frank J (ed) (2006) Electron tomography, 2nd edn. Springer, New York, USA
10. Fujii T, Iwane AH, Yanagida T, Namba K (2010) Direct visualization of secondary structures of F-actin by electron cryomicroscopy. Nature 467:724–728
11. Hawkes P, Valdre U (1990) Biophysical electron microscopy. Academic, Oxford, UK
12. Henderson R (2004) Realizing the potential of electron cryo-microscopy. Q Rev Biophys 37:3–13
13. Jensen GJ (ed) (2010) Cryo-EM, Part A: sample preparation and data collection, vol 481, Methods in enzymology. Elsevier, San Diego
14. Jensen GJ (ed) (2010) Cryo-EM, Part B: 3-D reconstruction, vol 482, Methods in enzymology. Academic Press, Amsterdam, Boston
15. Jensen GJ (ed) (2010) Cryo-EM, Part C: analyses, interpretation, and case studies, vol 483, Methods in Enzymology. Academic Press, Amsterdam, Boston
16. Law RHP, Lukoyanova N, Voskoboinik I, Caradoc-Davies TT, Baran K, Dunstone MA, D'Angelo ME, Orlova EV, Coulibaly F, Verschoor S, Browne KA, Ciccone A, Kuiper MJ, Bird PI, Trapani JA, Saibil HR, Whisstock JC (2010) The structural basis for membrane binding and pore formation by lymphocyte perforin. Nature 468:447–451. doi:10.1038/nature09518
17. Lucic V, Leis A, Baumeister W (2008) Cryo-electron tomography of cells: connecting structure and function. Histochem Cell Biol 130:185–196
18. Navaza J (2008) Combining X-ray and electron-microscopy data to solve crystal structures. Acta Crystallogr D Biol Crystallogr 64:70–75
19. Orlova EV, Saibil HR (2010) Methods for three-dimensional reconstruction of heterogeneous assemblies. Methods Enzymol 482:321–341
20. Orlova EV, Saibil HR (2011) Structural analysis of macromolecular assemblies by electron microscopy. Chem Rev (in press). http://pubs.acs.org/doi/full/10.1021/cr100353t
21. Radermacher M (1988) Three-dimensional reconstruction of single particles from random and nonrandom tilt series. J Electron Microsc Tech 9:359–394
22. Reimer L (1997) Transmission electron microscopy – physics of image formation and microanalysis. Springer, Berlin
23. Topf M, Lasker K, Webb B, Wolfson H, Chiu W, Sali A (2008) Protein structure fitting and refinement guided by cryo-EM density. Structure 16:295–307
24. van Heel M (1987) Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. Ultramicroscopy 21:111–124
25. van Heel M, Gowen B, Matadeen R, Orlova EV, Finn R, Pape T, Cohen D, Stark H, Schmidt R, Schatz M, Patwardhan A (2000) Single-particle electron cryo-microscopy: towards atomic resolution. Q Rev Biophys 33:307–369
26. Zhang X, Jin L, Fang Q, Hui WH, Zhou ZH (2010) 3.3 Å cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry. Cell 141:472–482

Chapter 12
Assembly and Function of the Signal Recognition Particle from Archaea

Elisabeth Sauer-Eriksson, Shenghua Huang, and Tobias Hainzl

Abstract The signal recognition particle (SRP) is a protein-RNA complex that associates with ribosomes to mediate co-translational targeting of membrane and secretory proteins to biological membranes. The universally conserved core of SRP consists of SRP RNA and the SRP54 protein, and plays the key role in signal-sequence recognition and binding to the SRP receptor. Critical for SRP function is communication between the two conserved SRP54 domains, the GTPase- and the M-domain, so that signal-sequence binding at the M domain directs receptor binding at the GTPase domain. The structural basis for signal-sequence binding by SRP and subsequent signaling is still poorly understood. By studying the structures of the SRP RNA in its free form as well as in complex with its different protein partners, we have made steady progress towards the elucidation of structural states of the SRP, using the archaeon *Methanococcus jannaschii* as model system. Together with other structures of SRP proteins and RNA-protein complexes, these structures provide new insights into the mechanisms of SRP-mediated protein targeting.

Keywords Protein transport • Signal recognition particle • X-ray structure • *Methanococcus jannaschii* • Signal sequence

12.1 Introduction

Approximately 30% of all proteins that are synthesized in the cell cytosol need to be transported to their final destination for proper function. They must be transferred not only to other cell compartments, like the Golgi, the endoplasmic reticulum (ER),

E. Sauer-Eriksson (✉) • S. Huang • T. Hainzl
Department of Chemistry, Umeå University, SE-90187 Umeå, Sweden
e-mail: elisabeth.sauer-eriksson@chem.umu.se

M.A. Carrondo and P. Spadon (eds.), *Macromolecular Crystallography*,
NATO Science for Peace and Security Series A: Chemistry and Biology,
DOI 10.1007/978-94-007-2530-0_12, © Springer Science+Business Media B.V. 2012

or the mitochondria in eukaryotic cells, but also to the external environment outside the cell. The highly hydrophobic character of membranes is not conducive to protein transport, and the cell has developed sophisticated solutions to overcome this problem. Divergent pathways exist for sending proteins into various compartments or out of cells. These protein translocation systems consist of complex molecular machines that are highly specialized for the specific compartments. One of these protein transport systems is the signal recognition particle (SRP), which functions as an assisting molecule in co-translational protein targeting. Newly synthesized proteins destined for transfer go directly from the ribosome to the membrane-bound pore-forming protein, referred to as the translocon, through which channel the secreted protein should cross the membrane. The role of SRP in this mechanism is to guide the translating ribosome to the translocon.

12.2 Signal Recognition Particle

12.2.1 Background

SRP is a multi-subunit ribonucleoprotein (RNP) found in all three kingdoms of life, and is essential for normal cell function (reviewed in [6, 8, 11, 24, 30]). The SRP recognizes the N-terminal signal sequences of nascent polypeptide chains as they emerge from the ribosome, and targets the SRP-nascent chain-ribosome complex to the endoplasmic reticulum (ER) in the case of eukaryotic cells or to the plasma membrane in prokaryotes via interactions with the SRP receptor. At the membrane, SRP release is mediated by mutual GTP hydrolysis by the SRP and its receptor, leaving the nascent chain directed through the translocon or integrated into the membrane (Fig. 12.1).

Extensive biochemical analysis of the SRP, in combination with structural studies of the RNA subdomains, individual protein components, and RNA-protein complexes, has highlighted the main structural features that govern assembly and function. More recently, cryo-EM work on SRP-ribosome-nascent chains has also revealed the interaction with the ribosome [16, 18, 31].

12.2.2 SRP Composition

Mammalian SRP is the most complex of all SRPs. It consists of a single 7S RNA (also termed SRP RNA, ~300 nucleotides (nt)) and six protein components named SRP9, SRP14, SRP19, SRP54, SRP68, and SRP72, according to their mass in kDa (Fig. 12.2). All proteins except SRP72 bind directly to 7S RNA, which can be structurally and functionally divided into two domains, the S-domain and the *Alu*-domain. The S-domain (7S.S RNA) binds to SRP19, SRP54, and the SRP68/72 heterodimer, of which SRP54 plays the key role in recognizing the signal sequences and interacting

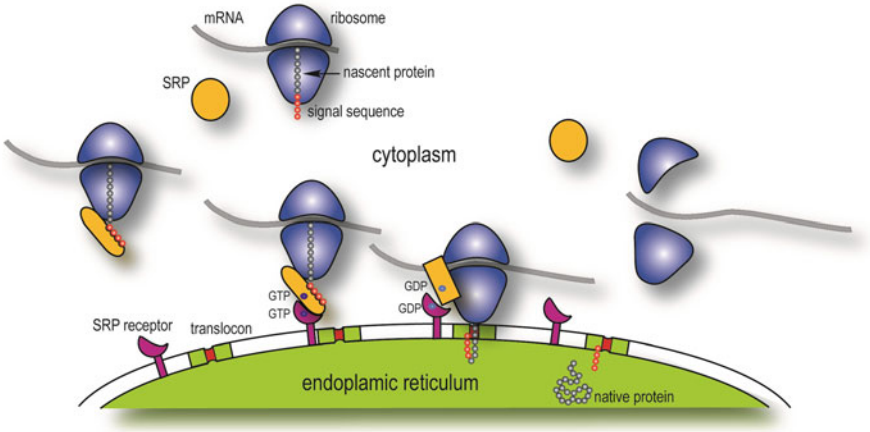


Fig. 12.1 The SRP cycle

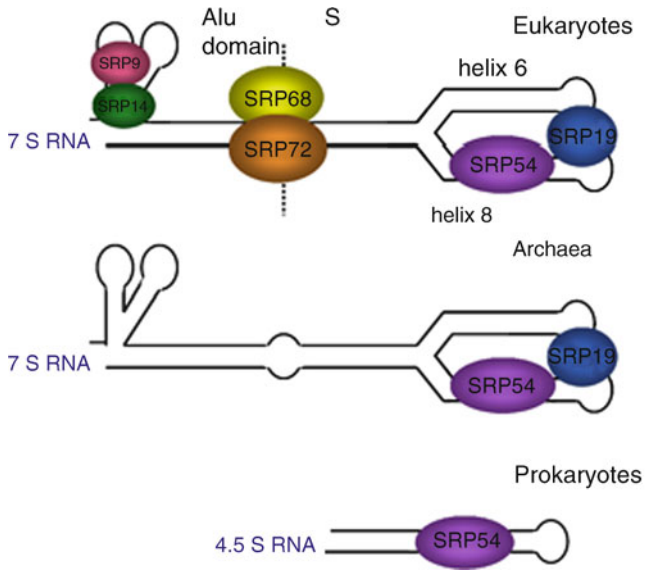


Fig. 12.2 Schematic representation of SRP from the three kingdoms of life. Protein components bound to SRP RNA are shown as spheres

with the SRP receptor. The secondary structures of archaeal 7S RNAs are highly similar to their mammalian counterparts. Genome sequence analyses, however, indicate that archaeal species code for only two homologues of the six eukaryotic SRP proteins, namely SRP19 and SRP54 [1]. In eubacteria, a minimal set of SRP subunits exists composed of 4.5S RNA, a single RNA helix (domain IV, homologue to helix 8), and the SRP54 homologue Ffh (54 homologue) [27, 28].

12.2.3 The Domain Structure of SRP54

SRP54, or Ffh as it is referred to in bacteria, is the protein that is most highly conserved among the different species and consists of three structural domains, N, G, and M, which mediate different functions [2, 10]. The M-domain (methionine-rich domain) recognizes the signal sequences and binds the SRP RNA, whereas the NG-domain is a ras-type GTPase that interacts with the SRP receptor [7, 9, 23]. The secondary structure of 7S.S RNA is approximately Y-shaped with helices 6 and 8 forming the two apical stems capped by tetraloops. Helix 8 consists of three short regular helical segments connected by a symmetric and an asymmetric internal loop. Sequence and secondary structures of the two loops show strong similarities in all organisms, and they comprise the binding site for the SRP54 M domain. 7S RNA does not function solely as a rigid scaffold for the SRP protein components. Rather, it functions in an integrated way with SRP54 and appears to have an important role in both the signal-sequence binding and the SRP-receptor interactions [3, 4, 19, 26, 32, 34].

12.2.4 Function of SRP19

We use SRP from the thermophilic archaeon *Methanococcus jannaschii* as model system for SRP assembly and function. In 2002 we and others published the crystal structure of SRP19 bound to 7S.S RNA [12, 25]. The structures revealed that SRP19 binds to the tetraloops of helices 6 and 8, clamps them together, and induces extensive interactions between them. In particular, looped-out adenosines in helix 6 of *M. jannaschii* SRP RNA seemed to stabilize three nucleotides in the asymmetric loop of helix 8 in an SRP54-binding competent state. SRP biogenesis in higher eukaryotes involves sequential binding of SRP19 and SRP54 proteins to 7S RNA. To better understand the mechanism by which SRP19 promotes SRP54 binding to 7S.S RNA, the knowledge of the free 7S.S RNA structure seemed indispensable. We therefore determined the crystal structure of the protein-free S-domain of SRP RNA from *M. jannaschii* [13] (Fig. 12.3). Interestingly, in the free form of SRP RNA the three nucleotides of the asymmetric loop, shown to be important for M domain binding, are directed towards the RNA's helical axis as predicted from analysis of the binary structure. The structure revealed the nature of long-range SRP19-induced conformational changes, which promote subsequent SRP54 attachment. To conclude, the main function of SRP19 seems to be to alter the RNA interaction geometries and, consequently, cause an asymmetric loop conformation that resembles the one in the SRP54-bound form.

12.2.5 The Structure of the Complete SRP S Domain from Archaea

Biochemical studies as well as X-ray and EM structures of free SRP54, and SRP54 in complex with SRP RNA and the ribosome, show major structural flexibility

Fig. 12.3 (*Left*) The 2.6 Å crystal structure of the free S domain from SRP RNA (pdb code 1z43 [13]). (*Right*) The 2.3 Å crystal structure of the SRP19 bound SRP RNA (pdb code 1lng, [12]). The *arrows* point at the three unpaired bases 195-ACC in the asymmetric loop, which are pointing toward the interior of the RNA helix in the free RNA structure. Both structures are from *M. jannaschii*

within SRP54. It is suggested that these conformational changes in SRP54 are fundamental to the coordinated binding and release of the ribosome, the signal peptide, the SRP receptor, and the translocon. We characterized the structure of the SRP54–SRP19–7S.S RNA complex from *M. jannaschii* in its free state [14], and showed that in the ligand-free SRP core the SRP54 G-domain has no direct contact with the M-domain and that a movement of the NG-domain is restricted by direct interactions with the RNA (Fig. 12.4, right). The NG domain is associates lengthwise with the 7S RNA, burying more than ~2,000 Å^2 of solvent-accessible surface area. This makes the NG domain–RNA interface considerably larger than the ~1,300 Å^2 M domain–RNA interface. Still, the M domain binds to the RNA by formation of an intricate intermolecular interface, including seven base-specific interactions. In contrast, the NG domain makes relatively few direct contacts and those that are formed are exclusively with the sugar–phosphate backbone. The absence of most of the G domain binding site in the 45-mer RNA used for crystallization of the *S. solfataricus* SRP54–helix 8 complex may explain the observed open non-RNA-bound structure in this case [29] (Fig. 12.4, left). However, the different NG domain orientations observed in *S. solfataricus* and *M. jannaschii* SRP, could represent examples from a dynamic range or discrete states of conformations that occur in free SRP.

The 25–30 amino-acid long GM-linker constitutes the key structural element that allows the reported large domain rearrangements within SRP54 linking binding of external ligands by SRP to the acquisition of proper NG–M configurations [5, 10, 16, 29, 32]. In the open structure of the *S. solfataricus* SRP core, this GM-linker forms a well defined long α-helix that places the N domain in contact with the fingerloop in the M domain [29]. In the crystal of the *M. jannaschii* SRP

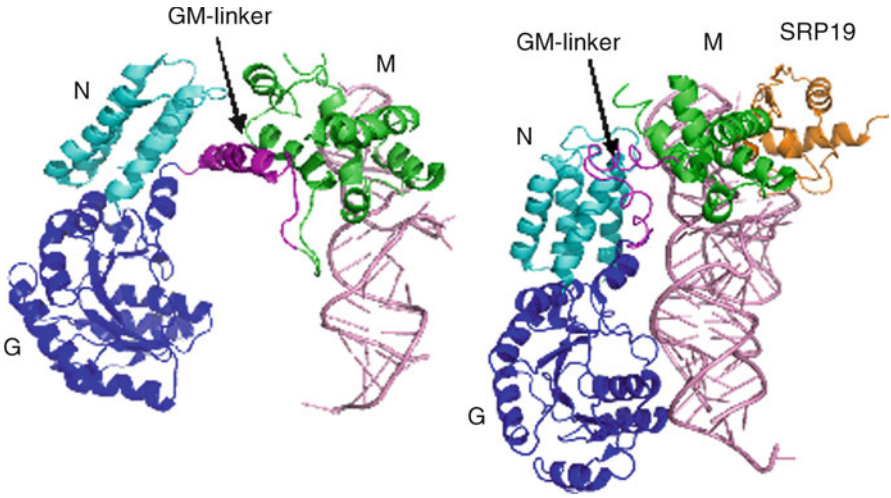


Fig. 12.4 (*Right*) The crystal structure at 2.5 Å resolution of the SRP54–SRP19–SRP RNA complex of *M. jannaschii* SRP (pdb code 2v3 [14]). (*Left*) The crystal structure at 4.0 Å resolution of the SRP54–helix 8 RNA complex of *Sulfolobus solfataricus* SRP (pdb code 1qzw [29]). The position of the GM-linker and SRP54 domains are indicated. The structures are shown in the same orientation based on the M domains

core, the GM-linker forms a long but irregular helical-like structure that bends sharply and contacts both the N domain tip and the RNA. It seems likely that the more open structure of SRP, as seen for example in *S. solfataricus* [29], is coupled to an unbending of the GM-linker. In other words, the GM-linker might act as a flexible spring between the M and NG domains, to allow for the large ribosome-induced conformational rearrangement within the SRP core revealed by the cryo-EM structures. Upon signal-sequence binding, the hinge regions at the GM-linker's N and C termini may then allow for further adjustment of the NG–M interface.

The SRP54 domain arrangement on the ribosome is strikingly different than that observed in the free S domain of *M. jannaschii.* It is plausible that in the absence of a ligand, i.e. the signal sequence, the locked SRP54 domain arrangement on the SRP RNA keeps the NG and the M domains spatially well separated. In this arrangement, the direct NG–M communications that are thought to be critical for the binding of signal sequences and subsequent signal transduction are unlikely to occur.

12.2.6 Signal-Sequence Binding to the SRP

Even though the known structures of SRP determined by us and others have revealed a wealth of information about SRP assembly and function, the molecular mechanism of signal-sequence binding is poorly understood. Currently unanswered is the

question of how SRP recognizes and productively binds almost any such hydrophobic signal sequence. Critical for SRP function is also the communication between the two conserved SRP54 domains, the GTPase- and the M-domain, so that signal-sequence binding at the M domain directs receptor binding at the GTPase domain. Biochemical and structural data of a signal-sequence bound SRP core of *M. jannaschii* suggest how SRP recognizes a signal sequence, and describe signal-sequence induced structural changes within the SRP core.

N-terminal signal sequences can be divided into three regions: a basic N-terminal region, a central hydrophobic region of 7–20 amino acids, and a more polar C-terminal region that contains the cleavage site for signal peptidase [33]. However, neither the N-terminal signal sequences nor the transmembrane segments recognized by SRP show primary sequence conservation. The unifying characteristic of all signal sequences, and apparently the only essential feature for binding to SRP, is an uninterrupted stretch of at least seven hydrophobic amino acids surpassing a threshold level of hydrophobicity [22]. A typical signal sequence comprises 9–12 hydrophobic residues in a row that adopt an α-helical conformation. Recent data show that the function of *E. coli* SRP RNA, with respect to accelerating SRP-SR complex formation, requires the presence of a signal sequence [3].

From the first structures of SRP54, the signal-sequence binding site was identified as a hydrophobic groove present in the M domain [2, 21]. Cryo-EM studies on SRP-ribosome complexes showed additional electron density in this area of the M domain, which was attributed to the signal sequence [17, 18, 31]. More recently, the same binding site was also identified in the 3.9 Å structure of SRP54 with a bound signal peptide [18]. We have determined the crystal structure of the *M. jannaschii* SRP core bound to a signal sequence, providing us with new insights into the mechanism of SRP activity [15]. Due to their strong hydrophobic character, isolated signal sequences are poorly soluble in biological buffers. By fusing the signal sequence to the C-terminus of α-M5 in the SRP54 M domain, separated by a flexible glycine/serine linker this problem could be overcome. A similar approach was also used in the free SRP54 structure [20]. For the *M. jannaschii* SRP core, biochemical data including GTPase activities of SRP-SR complexes verified that the fused signal sequence was functional, accelerating SRP-SR interaction in an SRP RNA-dependent manner [15]. In *M. jannaschii*, the signal sequence binds at the M domain in a similar but not identical conformation to the one found previously [17, 18, 20]. Comparing the *M. jannaschii* SRP core structures with and without the signal sequence provides the basis for a model that describes how the signal-sequence may stimulate SRP-SR interaction [15].

The structures of the *M. jannaschii* SRP provide evidence for the important role SRP RNA plays in the regulation of the SRP54 activities allowing for an ordered sequence of events during protein targeting. By regulating the activity of the receptor, the SRP RNA coordinates signal-sequence binding to interaction with the translocon. To verify this hypothesis, further structural evidence that elucidate the interactions between SRP RNA and the receptor is needed.

References

1. Andersen ES, Rosenblad MA, Larsen N, Westergaard JC, Burks J, Wower IK, Wower J, Gorodkin J, Samuelsson T, Zwieb C (2006) The tmRDB and SRPDB resources. Nucleic Acids Res 34((Database issue)):D163–D168
2. Batey RT, Rambo RP, Lucast L, Rha B, Doudna JA (2000) Crystal structure of the ribonucleoprotein core of the signal recognition particle. Science 287(5456):1232–1239
3. Bradshaw N, Neher SB, Booth DS, Walter P (2009) Signal sequences activate the catalytic switch of SRP RNA. Science 323(5910):127–130
4. Bradshaw N, Walter P (2007) The signal recognition particle (SRP) RNA links conformational changes in the SRP to protein targeting. Mol Biol Cell 18(7):2728–2734
5. Chu F, Shan SO, Moustakas DT, Alber F, Egea PF, Stroud RM, Walter P, Burlingame AL (2004) Unraveling the interface of signal recognition particle and its receptor by using chemical cross-linking and tandem mass spectrometry. Proc Natl Acad Sci USA 101(47):16454–16459
6. Doudna JA, Batey RT (2004) Structural insights into the signal recognition particle. Annu Rev Biochem 73:539–557
7. Egea PF, Shan SO, Napetschnig J, Savage DF, Walter P, Stroud RM (2004) Substrate twinning activates the signal recognition particle and its receptor. Nature 427(6971):215–221
8. Egea PF, Stroud RM, Walter P (2005) Targeting proteins to membranes: structure of the signal recognition particle. Curr Opin Struct Biol 15(2):213–220
9. Focia PJ, Shepotinovskaya IV, Seidler JA, Freymann DM (2004) Heterodimeric GTPase core of the SRP targeting complex. Science 303(5656):373–377
10. Freymann DM, Keenan RJ, Stroud RM, Walter P (1997) Structure of the conserved GTPase domain of the signal recognition particle. Nature 385(6614):361–364
11. Grudnik P, Bange G, Sinning I (2009) Protein targeting by the signal recognition particle. Biol Chem 390(8):775–782
12. Hainzl T, Huang S, Sauer-Eriksson AE (2002) Structure of the SRP19 RNA complex and implications for signal recognition particle assembly. Nature 417(6890):767–771
13. Hainzl T, Huang S, Sauer-Eriksson AE (2005) Structural insights into SRP RNA: an induced fit mechanism for SRP assembly. RNA 11(7):1043–1050
14. Hainzl T, Huang S, Sauer-Eriksson AE (2007) Interaction of signal-recognition particle 54 GTPase domain and signal-recognition particle RNA in the free signal-recognition particle. Proc Natl Acad Sci USA 104(38):14911–14916
15. Hainzl T, Huang S, Merilainen G, Brannstrom K, Sauer-Eriksson AE (2011) Structural basis of signal-sequence recognition by the signal recognition particle. Nat Struct Mol Biol 18(3):389–391
16. Halic M, Becker T, Pool MR, Spahn CM, Grassucci RA, Frank J, Beckmann R (2004) Structure of the signal recognition particle interacting with the elongation-arrested ribosome. Nature 427(6977):808–814
17. Halic M, Blau M, Becker T, Mielke T, Pool MR, Wild K, Sinning I, Beckmann R (2006) Following the signal sequence from ribosomal tunnel exit to signal recognition particle. Nature 444(7118):507–511
18. Halic M, Gartmann M, Schlenker O, Mielke T, Pool MR, Sinning I, Beckmann R (2006) Signal recognition particle receptor exposes the ribosomal translocon binding site. Science 312(5774):745–747
19. Jagath JR, Matassova NB, de Leeuw E, Warnecke JM, Lentzen G, Rodnina MV, Luirink J, Wintermeyer W (2001) Important role of the tetraloop region of 4.5S RNA in SRP binding to its receptor FtsY. RNA 7(2):293–301
20. Janda CY, Li J, Oubridge C, Hernandez H, Robinson CV, Nagai K (2010) Recognition of a signal peptide by the signal recognition particle. Nature 465(7297):507–510
21. Keenan RJ, Freymann DM, Walter P, Stroud RM (1998) Crystal structure of the signal sequence binding subunit of the signal recognition particle. Cell 94(2):181–191

22. Lee HC, Bernstein HD (2001) The targeting pathway of Escherichia coli presecretory and integral membrane proteins is specified by the hydrophobicity of the targeting signal. Proc Natl Acad Sci USA 98(6):3471–3476
23. Montoya G, Svensson C, Luirink J, Sinning I (1997) Crystal structure of the NG domain from the signal-recognition particle receptor FtsY. Nature 385(6614):365–368
24. Nagai K, Oubridge C, Kuglstatter A, Menichelli E, Isel C, Jovine L (2003) Structure, function and evolution of the signal recognition particle. EMBO J 22(14):3479–3485
25. Oubridge C, Kuglstatter A, Jovine L, Nagai K (2002) Crystal structure of SRP19 in complex with the S domain of SRP RNA and its implication for the assembly of the signal recognition particle. Mol Cell 9(6):1251–1261
26. Peluso P, Shan SO, Nock S, Herschlag D, Walter P (2001) Role of SRP RNA in the GTPase cycles of Ffh and FtsY. Biochemistry 40(50):15224–15233
27. Poritz MA, Bernstein HD, Strub K, Zopf D, Wilhelm H, Walter P (1990) An E. coli ribonucleoprotein containing 4.5S RNA resembles mammalian signal recognition particle. Science 250(4984):1111–1117
28. Ribes V, Romisch K, Giner A, Dobberstein B, Tollervey D (1990) E. coli 4.5S RNA is part of a ribonucleoprotein particle that has properties related to signal recognition particle. Cell 63(3):591–600
29. Rosendal KR, Wild K, Montoya G, Sinning I (2003) Crystal structure of the complete core of archaeal signal recognition particle and implications for interdomain communication. Proc Natl Acad Sci USA 100(25):14701–14706
30. Sauer-Eriksson AE, Hainzl T (2003) S-domain assembly of the signal recognition particle. Curr Opin Struct Biol 13(1):64–70
31. Schaffitzel C, Oswald M, Berger I, Ishikawa T, Abrahams JP, Koerten HK, Koning RI, Ban N (2006) Structure of the E. coli signal recognition particle bound to a translating ribosome. Nature 444(7118):503–506
32. Spanggord RJ, Siu F, Ke A, Doudna JA (2005) RNA-mediated interaction between the peptide-binding and GTPase domains of the signal recognition particle. Nat Struct Mol Biol 12(12):1116–1122
33. von Heijne G (1985) Signal sequences. The limits of variation. J Mol Biol 184(1):99–105
34. Zhang X, Rashid R, Wang K, Shan SO (2010) Sequential checkpoints govern substrate selection during cotranslational protein targeting. Science 328(5979):757–760

Chapter 13
Structural Studies of the Functional Complexes of the 50S and 70S Ribosome, a Major Antibiotic Target

Thomas A. Steitz, Gregor Blaha, C. Axel Innis, Robin Evans Stanley, and David Bulkley

Abstract Our crystal structure of the *Haloarcula marismortui (H.ma.)* 50S ribosomal subunit and its complexes with substrates and antibiotics have illuminated the mechanism of peptide bond formation and its inhibition by antibiotics. Our structures of the *Thermus thermophilus (T.th.)* 70S ribosome complexed with tRNAs, protein factor EF-P or antibiotics have also provided insights into their mechanisms of action. We conclude that the CCA ends of the A- and P-site tRNAs bind to the 70S ribosome as the CCA fragments bind to the 50S subunit; macrolide antibiotics bind to the T.th. 70S ribosome as they bind to the H.ma. 50S subunit; EF-P binds to the 70S ribosome adjacent to and interacting with a P-site tRNA, and cryoEM maps of a 70S ribosome bound to a peptidyl-tRNA containing an arresting sequence shows an extended polypeptide interacting with the tunnel wall.

T.A. Steitz (✉)
Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA

Department of Chemistry, Yale University, New Haven, CT, USA

Howard Hughes Medical Institute, New Haven, CT, USA
e-mail: thomas.steitz@yale.edu

G. Blaha • C.A. Innis
Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA

R.E. Stanley
Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA

NIDDK, National Institutes of Health, Bethesda, MD, USA

D. Bulkley
Department of Chemistry, Yale University, New Haven, CT, USA

Howard Hughes Medical Institute, New Haven, CT, USA

M.A. Carrondo and P. Spadon (eds.), *Macromolecular Crystallography*,
NATO Science for Peace and Security Series A: Chemistry and Biology,
DOI 10.1007/978-94-007-2530-0_13, © Springer Science+Business Media B.V. 2012

In all organisms messenger-directed protein synthesis is catalyzed by ribonucleoprotein particles called ribosomes. Bacterial ribosomes sediment at about 70S, and consist of two non-equivalent subunits – a small subunit that sediments at 30S and a large, 50S subunit. The latter has a molecular weight of about 1.6×10^6 Da and catalyzes peptide bond formation, functioning as a ribozyme [9, 31]. The 50S subunit contains a 2,900 nt 23S rRNA and a 120 nt 5S rRNA as well as about 33 different proteins [63]. The high resolution atomic structures of the large [8, 24] and small [62] ribosomal subunits along with their substrate complexes as well as the model of the 70S ribosome derived initially from a 5.5 Å resolution map [66] and more recently from maps calculated at 3.5 Å, 3.7 Å and 2.8 Å resolution [28, 42, 44] have yielded extensive insights into the structural basis of protein synthesis.

13.1 Determination of the Ribosome Structure

Ada Yonath and coworkers made a seminal contribution to the quest for a high resolution structure of the ribosome by obtaining the first crystals of ribosomal subunits, and by discovering crystals of the *Haloarcula marismortui* (*H.ma.*) large subunit that diffract to 3 Å resolution [20, 45, 57, 65]. This accomplishment established that, in principle, an atomic level structure of the ribosome could be determined using x-ray crystallography. However, the first published electron density maps of the *H.ma.* 50S subunit from the Yonath lab, which were based on 7 Å resolution x-ray diffraction data from crystals derivatized by using heavy atom cluster compounds, did not show the continuous density features corresponding to duplex RNA expected to be seen at this resolution, suggesting the possibility that the heavy atoms may not have been correctly located [36]. For this reason, we decided to initiate our crystallographic studies of the *H.ma.* 50S ribosomal subunit and pursued a strategy of solving heavy atom difference Patterson maps at very low (20 Å) resolution where the scatter from the cluster compounds is almost the square of the total number of electrons in the cluster, and thus extremely high. The first successful derivative was that of a W_{18} cluster compound, which produced one major peak in each of the three orthorhombic Harker sections. In order to verify that we had indeed correctly located the W_{18}, we then used cryo-E.M. maps produced by Frank and coworkers to phase the x-ray diffraction data and calculate a difference Fourier map which showed the same location of the heavy atom cluster compound that we had derived from the difference Patterson map. Using this approach, we were successful in solving other heavy atom derivatives and in producing the first x-ray crystallographically derived electron density map at 9 Å resolution. It showed convincingly recognizable molecular features of duplex RNA and the same overall shape of the 50S subunit observed in the cryo-EM maps [6]. We then extended the resolution to 5 Å at which point known r-protein structures and some rRNA could be fitted into the map [7].

In the summer of 2000 we published the first atomic structure of a ribosomal subunit derived from a 2.4 Å resolution map of the *H.ma.* 50S subunit [8] as well as that of a complex with a substrate intermediate analogue [31]. Shortly thereafter, the Ramakrishnan group published the structure of the 30S subunit derived from a 3 Å

resolution map [62] and subsequently obtained the structures of the 30S subunit complexed with a short mRNA and two anticodon stem loops that provided many insights into the mechanism of decoding [33]. Yusupov et al. [66] were able to use the coordinates of the separate 30S and 50S subunits to build a model of the whole ribosome including tRNA molecules bound to the A, P and E sites fitted to a 5.5 Å resolution map. While this model clearly showed the relationships between the two subunits and the tRNA molecules, accurate modeling of specific interactions was not possible.

Subsequent further refinement of the *H.ma.* 50S subunit structure at 2.4 Å resolution [27], as well as additional analyses of the structure led to important conclusions about the architectural principles underlying the structure of the large ribonuclear protein machine [26, 27, 32]. Klein and Schmeing discovered a new secondary structure motif they called the kink-turn or K-turn [26]. They found six examples in 23S rRNA and two in 16S rRNA showing a kink in the phosphodiester backbone that results in a sharp turn in the RNA helix. Nissen et al. [32], found a new motif that stabilizes RNA tertiary structure which they termed the "A-minor motif" because it involves the insertion of adenines from one RNA secondary structure into the minor groove of a neighboring RNA helix. They found 186 examples in the 23S rRNA, and now examples in other systems abound. The A minor motif may be the most abundant and significant tertiary structure interaction in RNA structures. Klein et al. [27] identified 116 Mg^{++} ions and 88 monovalent cations bound to the large subunit, interacting mainly with the RNA backbone, presumably functioning to stabilize RNA secondary and tertiary structures.

13.2 The Mechanism of Peptide Bond Formation

Our structural studies of the *H.ma.* 50S subunit complexes with substrate analogues as well as mutational, biochemical and kinetic studies by others [19, 41, 60] have led to a detailed understanding of the mechanism by which the ribosome, functioning as a ribozyme, is able to catalyze peptide bond formation [39, 40, 52]. In the absence of an A-site substrate the ribosome protects the peptidyl-CCA from hydrolysis by preventing the access of water to the peptidyl ester link. Binding of a CC-hydroxypuromycin substrate analogue to the A site produces a conformational change in the peptidyl transferase center (PTC) that repositions and deprotects the carbonyl carbon of the ester-linked P site substrate, thereby enabling a nucleophilic attack by the alpha-amino group. The 2' OH of A76 of the P-site tRNA forms a hydrogen bond with the attacking alpha-amino group and acts as a proton shuttle to remove a proton from the attacking alpha-amino group and transfer a proton to the leaving 3' OH. Structures of transition state analogues show that catalysis is also enhanced by interactions of the oxyanion with a ribosome-positioned water molecule, thereby stabilizing the transition state [39, 40].

The suggestion has been made that since these studies were carried out using lower than physiological salt concentrations, our crystals therefore consisted of inactive 50S subunits [5]. However, assays of the catalytic activity of crystals of *H.ma.* 50S subunit showed that the crystals are nearly as active as the subunit in

solution and that the activity does not vary between 1.6 M (where the crystals are grown) and 3 M NaCl concentrations [41]. Another criticism of our structural studies of the mechanism of peptide bond formation was that we used fragment substrate analogues on the isolated 50S subunit. Noller and colleagues concluded from a model they built of the 70S ribosome complex with an A-site tRNA derived from a 3.7 Å resolution map [28] that the orientations and interactions of the CCA when a full tRNA and 70S ribosome were used were different from those observed when CCA analogues were bound to the 50S subunit. However, subsequent re-refinement using averaging of the Ramakrishnan and Noller 70S ribosome data [50] as well as two 3.1 Å resolution structures of the 70S ribosome complexed with tRNAs in the A and P sites from our lab [51] and from the Ramakrishnan lab [44, 59] showed that the structures of the PTC and the CCA ends bound to the A and P sites were identical to those seen in the *H.ma.* studies.

13.3 Antibiotic Complexes

The structures of many different families of antibiotics bound to either the large or small ribosomal subunit or to the 70S ribosome have been determined (reviewed in Wilson [61]), including those of more than a dozen antibiotics bound in and around the peptidyl transferase center of the *H.ma.* 50S ribosomal subunit [12, 22, 23, 55]. These studies have not only illuminated how these antibiotics inhibit protein synthesis, they have also provided an important starting point for a new biotech company, Rib-X Pharmaceuticals, Inc., that has enabled the structure based drug design of ribosome-targeting antibiotics. By chemically linking fragments of antibiotics that bind to adjacent locations on the ribosome, Rib-X is successfully creating novel antibiotics, one of which has now finished phase II clinical trials and should move to phase III soon.

In spite of *H.ma.* being an archaeon and its ribosome being closer to that of eukaryotes than that of eubacteria, several macrolide antibiotics, including azithromycin, could be bound to the *H.ma.* 50S subunit using high antibiotic concentrations [22]. These macrolides were all seen to bind at the top of the polypeptide exit tunnel, just below the PTC. They appear to inhibit protein synthesis by blocking the egress of the nascent polypeptide chain. While, the orientations of their macrolide rings were nearly identical to each other in our structures, they differed dramatically from that proposed for erythromycin bound to crystals of the *D. radiodurans* (*D.ra.*) 50S ribosomal subunit [37]. The question then arose – was this difference due to incorrect modeling of erythromycin into modest resolution *D.ra.* electron density maps, or was it due to species-related differences between eubacteria and archaea? To address this issue, we mutated G2099 to A (2058 in E. coli) in the *H.ma.* subunit, which enabled erythromycin to bind with approximately 10^4 fold higher affinity [55]. In the structure of the complex, the lactone ring of erythromycin was seen to be oriented exactly as was observed in the *H.ma.* subunit complexes with other macrolides, but nearly orthogonal to the *D.ra.* subunit model [55]. These results were consistent with our view that the differences between our antibiotic results and those

from the *D.ra. s*tudies were a consequence of the insufficient resolution of the maps in the latter studies, but doubts persisted.

It was suggested that the structures of macrolides bound to the G2099A mutant 50S subunit may not be representative of macrolide binding in eubacteria in general and that the discrepancies between the two sets of experiments may have arisen from differences in the mode of binding of macrolide antibiotics to ribosomes from different kingdoms of life [64].

In order to address these concerns, we generated complexes between these macrolide antibiotics and the 70S ribosome from the eubacterium *T.thermophilus (T.th.)* [15]. The structures of the complexes that we obtained are entirely consistent with those obtained with *H.ma.* large subunit complexes (Fig. 13.1c, d). Most significantly, the lactone rings of erythromycin, azithromycin and telithromycin are all oriented equivalently in complex with either the 70S ribosome of *T.th.* or the 50S ribosomal subunit of *H.ma.* (Fig. 13.2). Minor differences were observed for the antibiotic telithromycin, whose alkyl-aryl moiety has distinct modes of interaction with ribosomes from each of the three species studied (Fig. 13.2k).

We also reinvestigated the mode of binding of chloramphenicol, which directly competes with the aminoacyl moiety of an incoming t-RNA for binding to the A-site crevice of the PTC [30], by generating a complex with the *T.th.* 70S ribosome. Only one structural model of chloramphenicol bound to a bacterial ribosome was available [37]. In order to evaluate the accuracy of this model, which again was built into a map of modest resolution, we solved the structure of chloramphenicol in complex with the 70S bacterial ribosome from *T. th.* and discovered that the antibiotic binds in a radically different way from what had been proposed [15]. An unbiased F_o-F_c difference electron density map calculated using phases that did not include chloramphenicol showed that chloramphenicol is rotated essentially 180° relative to its orientation in the earlier model. Further, the interactions between the ribosomal RNA and chloramphenicol that we observed are entirely different from what had been proposed earlier, and a single, well-characterized potassium ion (rather than the two putative magnesium ions from the previous model) is seen stabilizing the antibiotic in the ribosomal A-site crevice (Fig. 13.1b). Interestingly, the position and orientation of the chloramphenicol moiety bound to the eubacterial ribosome that we observe is very similar to that of the eukaryotic-specific antibiotic anisomycin complexed with the archaeal ribosome from *H.ma.* [23]. The structure of the complex between 70S *T.th*. ribosome and chloramphenicol explains its species specificity for eubacterial rather than archaeal ribosomes, and importantly, this corrected structure of the chloramphenicol complex will be enabling for the design of new antibiotics.

We have determined the co-crystal structures of the 70S ribosome from *T.th.* complexed with three full-length tRNAs and a short piece of mRNA as well as one of two of the tuberactinomyins: viomycin or capreomycin [51], which are among the most effective antibiotics in use today for the treatment of multidrug-resistant tuberculosis. The three tRNA molecules are bound in the classical A, P and E sites in this complex with the 70S ribosome. As observed earlier [33], the bases of A1492 and A1493 make stabilizing A-minor interactions with the two base-pairs between the codon and anti-codon of the A-site tRNA and mRNA. Both drugs bind to a site

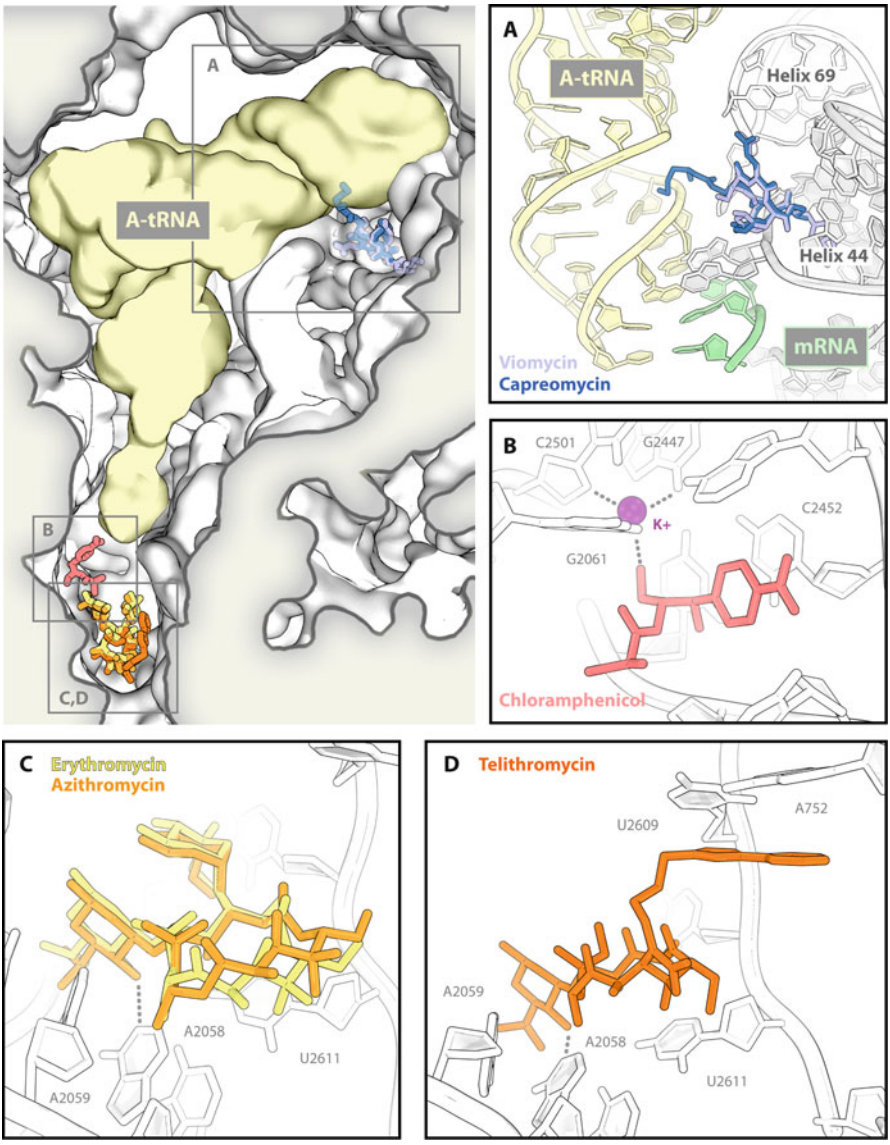


Fig. 13.1 The structures of several antibiotics in complex with a bacterial 70S ribosome. The *top left panel* shows a cross section of the 70S ribosome from *T.thermophilus* with an A-site tRNA [51] and various antibiotics bound near the decoding center (*panel A*), in the A-site crevice (*panel B*) or near the constriction of the exit tunnel (panel C,D). (**a**) Close-up view of the viomycin (PDB codes: 3KNH, 3KNI, 3KNJ and 3KNK) and capreomycin (PDB codes: 3KNL, 3KNM, 3KNN and 3KNO) binding site [51]. The codon-anticodon base pairing between the A-site tRNA and the mRNA can be seen in the *lower part of the panel*. (**b**) Close-up view of the chloramphenicol binding pocket in the *T.th.* 70S (PDB codes: 3OGE, 3OGY, 3OH5 and 3OH7), with a key potassium ion highlighted [15]. (**c**) Interactions of Erythromycin (PDB codes: 3OHC, 3OHD, 3OHJ and 3OHK) and Azithromycin (PDB codes: 3OHY, 3OHZ, 3OI0 and 3OI1) with the *T.th.* ribosome [15]. (**d**) Interactions between Telithromycin (PDB codes: 3OI2, 3OI3, 3OI4 and 3OI5) and the *T.th.* ribosome. The alkyl-aryl moiety of the drug can be seen stacking with the bases of U2609 and A752 [15]

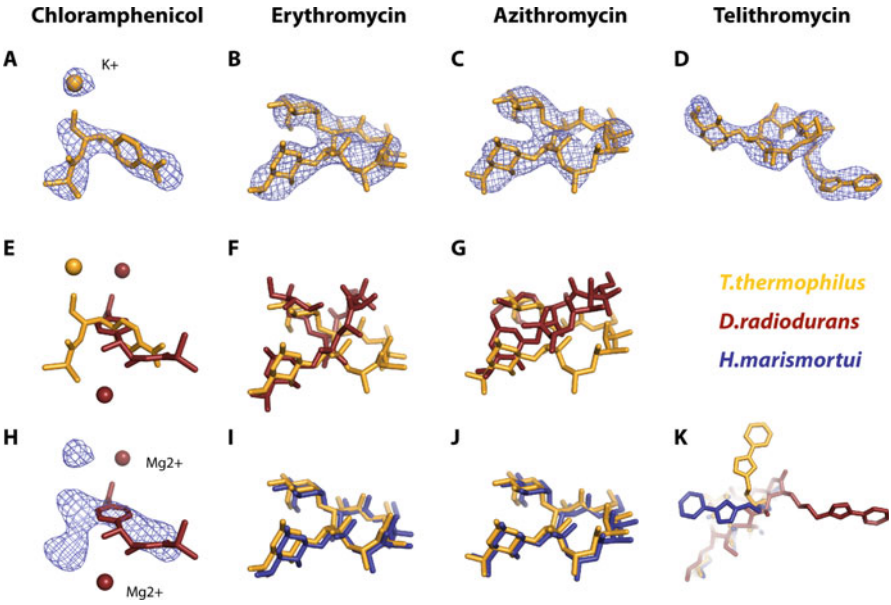


Fig. 13.2 Comparison of the *T.th.*, *D.ra.* and *H.ma.* antibiotic complex structures. Structures of the antibiotics chloramphenicol (**a**, **e**, **h**), erythromycin (**b**, **f**, **i**), azithromycin (**c**, **g**, **j**) and telithromycin (**d**, **k**) in complex with 70S ribosomes from *T.thermophilus* (*light gray*) or isolated 50S subunits from *D.radiodurans* (*dark gray*) and *H.marismortui* (*black*). Unbiased F_o-F_c difference electron density maps contoured at +3σ calculated using amplitudes from *T.th.* 70S ribosome crystals soaked with the different antibiotics are shown as a blue mesh (**a**, **b**, **c**, **d**, **h**). Overlays of the various antibiotic models were obtained by superimposing phosphate atoms from equivalent 23S rRNA residues of the *T.th./D.ra.* (**e**, **f**, **g**, **k**) and *T.th./H.ma.* (**h**, **i**, **j**, **k**) structures. The PDB codes corresponding to the structures shown are as follows: 1K01 (*D.ra.* 50S with chloramphenicol [37]), 1JZY (*D.ra.* 50S with erythromycin [37]), 1NWY (*D.ra.* 50S with azithromycin [38]), 1P9X (*D. ra.* 50S with telithromycin [11]), 1YI2 (*H.ma.* 50S with erythromycin [56]), 1M1K (*H.ma.* 50S with azithromycin [22]), 1YIJ (*H.ma.* 50S with telithromycin [56])

which lies between the large and small subunit in a cleft formed between helix 44 of the 16S rRNA (h44) and the tip of Helix 69 of the 23S rRNA (H69) (Fig. 13.1a).

The macrocycle of both drugs lies within hydrogen bonding distance of the ribose-phosphate backbone of the binding pocket, which is formed by residues of the 23S rRNA and of the 16S rRNA. The orientation of the macrocycle is determined by stacking interactions with bases from the 16S rRNA, which bring the six-membered ring of the drug into the vicinity of bases of A1492 and A1493 from the same subunit. The guanidinium moiety of this six-membered ring effectively locks the antibiotics into place by making a salt bridge to the backbone phosphate of A1493.

Classes of antibiotics that affect translocation of the tRNAs on the ribosome, such as the tuberactinomycins or aminoglycosides (Reviewed in [47]), either stabilize the tRNA in the pre-translocation state or interfere with the conformational changes of the ribosome or tRNA required for translocation [34]. Paromomycin is

an aminoglycoside that binds into the major groove of h44 and stabilizes A1492 and A1493 of the 16S rRNA in the flipped-out conformation adopted upon readout of the mRNA [33]. Because paromomycin stabilizes the binding of both cognate and near-cognate tRNAs to the A site, it not only inhibits translocation but also promotes miscoding [16]. Interestingly, although hygromycin B also binds to h44, it causes only the base of A1493 to adopt a unique orientation, in which the base interferes with the movement of the tRNA from the A site to the P site [14].

Viomycin and capreomycin not only affect the positioning of the bases of A1492 and A1493, but they also appear to affect the position of A1913 of the 23S rRNA. The base of A1913 adopts a position in which it forms hydrogen bonds with the A-site tRNA. Consistent with our results, biochemical experiments show that viomycin increases the affinity of tRNA to the A site by 1,000 fold [34] and promotes the back translocation of the complex of tRNA with mRNA on the ribosome [53]. Therefore we propose that the tuberactinomycins inhibit translocation by stabilizing the tRNA in the A-site.

Importantly, the positions of both paromomycin and hygromycin B when bound to the ribosome are very close to the position observed for the tuberactinomycins, suggesting that it might be possible to create new antibiotics that are effective against drug resistant strains of TB by chemically linking capreomycin to either paromomycin or hygromycin B.

13.4 EF-P and the First Peptide Bond

EF-P is conserved in all eubacteria and though it is not required in a minimal *in vitro* translation system [46], it has been shown to have a stimulatory effect *in vitro* on the formation of the first peptide bond [3]. EF-P binds stoichiometrically to the 70S ribosome. Consistent with its involvement in the initial stages of protein synthesis, the ratio of bound EF-P declines with the increasing size of polysomes [4]. One copy of EF-P is present in the cell per ten ribosomes, a ratio comparable to that observed for the translational initiation factors [2, 17].

In order to illuminate the mechanism by which EF-P facilitates the formation of the first peptide bond, we determined the crystal structure of EF-P bound to the complex of the *T.th.* 70S ribosome, mRNA, and initiator tRNA, at a resolution of 3.5 Å [13]. In our model, EF-P spans both ribosomal subunits and binds between the P and E tRNA binding sites making contact with the initiator tRNA near its anticodon stem-loop on the 30S subunit, and its D-loop and acceptor stem on the 50S subunit (Fig. 13.3a, b). The L1 stalk also undergoes a major conformational change that positions ribosomal protein L1 in the E site to interact with EF-P (Fig. 13.3c). The initiator tRNA bound to the P site displays the same conformation as the P-site tRNA of the 2.8 Å-resolution structure of the *T.th.* 70S ribosome with bound mRNA and tRNAs [44].

A loop of domain I of EF-P makes numerous interactions with the CCA end of the initiator tRNA, allowing the conserved arginine/lysine (R32) at its tip to

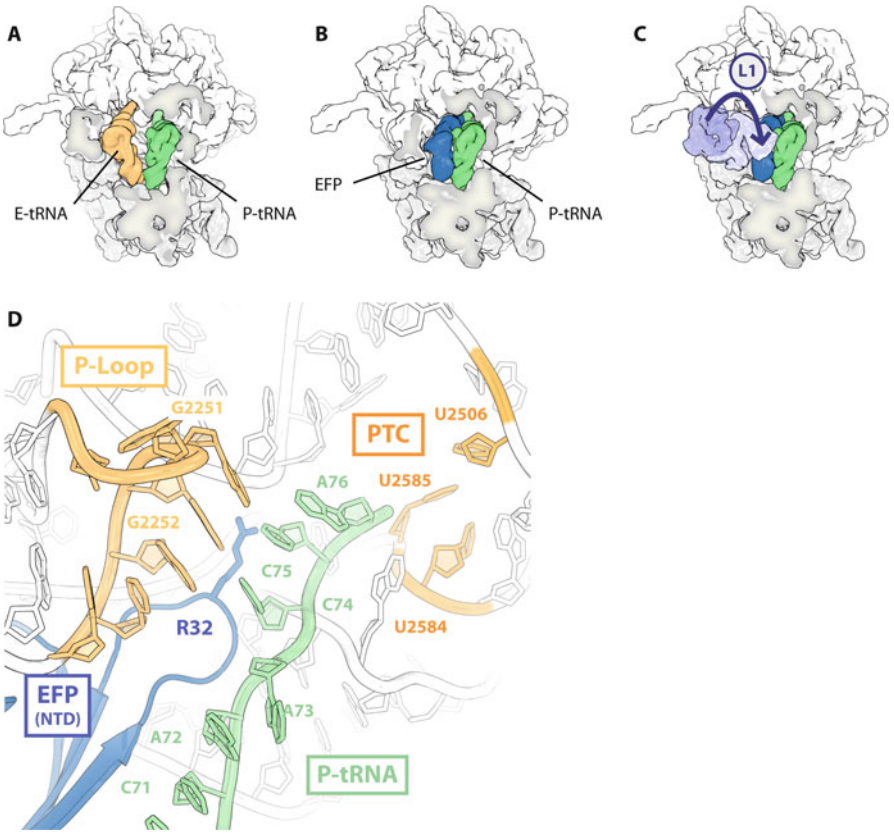


Fig. 13.3 The structure of EF-P bound to the ribosome. Schematic overview of (**a**) E- and P-site tRNAs bound to the 70S ribosome ([44]; PDB codes: 2J00, 2J01, 2J02 and 2J03), and of (**b**) EF-P and a P-site tRNA in complex with the *T.th.*70S ([13]; PDB codes: 3HUW, 3HUX, 3HUY and 3HUZ). (**c**) Same as in B, but with the movement of the L1 stalk observed upon binding of EF-P indicated by an arrow. The stalk of a 70S ribosome occupied with all three tRNAs is shown in dark gray, while the stalk from the EF-P complex is depicted in light gray. (**d**) Interactions of the N-terminal domain of EF-P near the PTC of the large ribosomal subunit

reach into the PTC, though not deep enough to directly participate in peptide bond formation (Fig. 13.3d). The modification of the corresponding Lys residue with 4-amino-2-hydroxybutyl to hypusine in eukaryotes and with lysyl in *E.coli* would extend the side chain, reaching even closer to the active site. It appears that EF-P's role may be to correctly position or stabilize the initiator tRNA in the P site.

Premature movement of the initiator tRNA to the E site may also be prevented by domain III of EF-P which stabilizes the A minor interactions between two G·C base pairs in the anticodon stem loop of the initiator tRNA and the bases of residues A1339 and G1338 of the 16S rRNA. A1339 and G1338 have been proposed to function as a "gate" between the P and E sites, because their A-minor interactions with the P-site

tRNA have to be broken during translocation [44]. By stabilizing these interactions EF-P may strengthen this gate and stabilize the fMet-tRNA$_i^{fMet}$ in the P-site.

Initiation of translation in eubacteria is a multistep process that involves the formation of several intermediate complexes with different compositions and conformations [48]. Structures of initiation complexes derived from cryo-EM studies have revealed that during the process of initiation the fMet-tRNA$_i^{fMet}$ adopts several different conformations on both the 30S subunit and the 70S ribosome before finally reaching its proper position in the P/P state [1, 48, 49]. By stabilizing the P/P state of the initiator tRNA EF-P could shift initiation towards the first elongation step of protein translation.

The structure of the EF-P complex with the 70S ribosome reveals the detailed interactions between EF-P and the 70S ribosome, initiator tRNA and the ribosomal protein L1. The essential role of EF-P in the cell may be to correctly position the fMet-tRNA$_i^{fMet}$ in the P site for the first step of peptide bond formation by making several interactions with the backbone of the tRNA.

13.5 The TnaC Leader Peptide as a Regulator of Protein Synthesis

Some nascent peptides can influence their own rate of translation by interacting with components of the 23S rRNA and ribosomal proteins lining the walls of the tunnel [10, 25, 29, 35, 54]. In these few instances, translation can come to a complete halt, and cells have evolved to exploit the unusual properties of these peptides as a means of regulating a variety of physiological processes. Although several arresting peptides have been relatively well characterized from a biochemical perspective (e.g., [18]), biophysical techniques have so far failed to provide us with a detailed mechanistic understanding of the underlying structural processes. We have begun to investigate the way in which specific nascent peptide sequences can stall translation in *cis*, with our long term goal being to dissect the molecular details governing this process using X-ray crystallography.

The nascent chain that we initially focused on is TnaC, a 24-residue peptide encoded by the leader region of the tryptophanase (*tna*) operon of *E.coli* [21]. Stalling of ribosomes on the *tna* transcript occurs in the presence of inducing levels of soluble tryptophan. This in turn leads to the expression of the downstream *tnaA* and *tnaB* genes and to the clearance of cytoplasmic tryptophan until non-inducing levels are restored, thus making TnaC a *de facto* tryptophan sensor for certain species of bacteria. By designing a peptide sequence with calmodulin binding peptide linked to the N-terminus of TnaC via a flexible linker, stalled TnaC-70S ribosome complexes generated using our cell-free protein synthesis system could then be purified to near homogeneity by affinity chromatography on a calmodulin sepharose matrix. Yields of up to 2 mg of pure complex could be obtained from 10 mL of translation reaction, making it both possible and practical to perform crystallization trials on this complex [43].

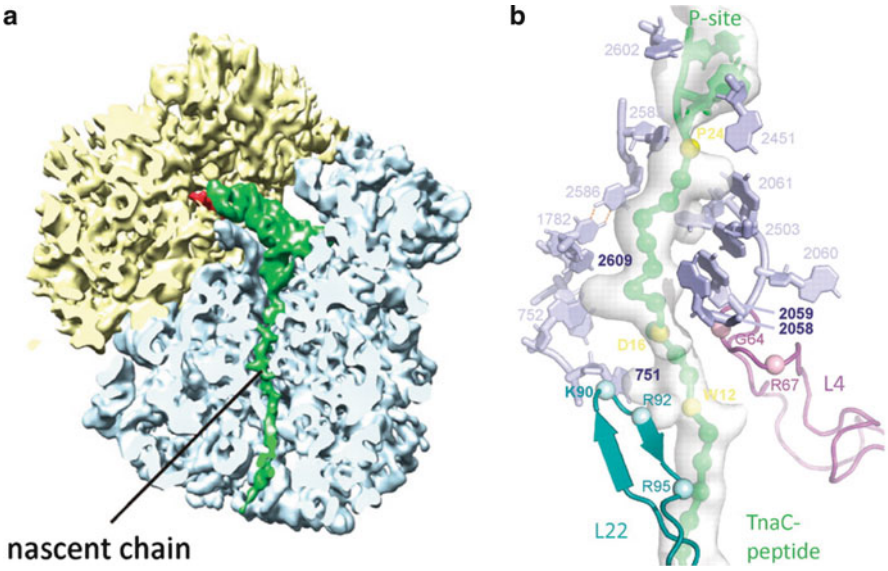


Fig. 13.4 Translational stalling mediated by the TnaC peptide. (**a**) Cryo-EM reconstruction of the TnaC-70S complex at a resolution of 5.8 Å (FSC = 0.5) ([43]; EMDB accession code: 1657). Density corresponding to the P-site tRNA and to the TnaC nascent chain is shown in dark gray. (**b**, **c**) Relay model for PTC silencing. Ribosomal elements that are likely to play a role in a relay mechanism to inactivate the PCT are shown, with residues important for stalling highlighted in bold. Density for the nascent chain is shown as a transparent surface in **b**. Potential relay pathways from Trp12 of TnaC to the PTC (R1-R3) are indicated in **c** (Adapted from Seidelt et al. 2009)

While crystallization experiments were under way, we began to collaborate with the group of Roland Beckmann at the University of Munich, and surprisingly the first set of cryo-electron micrographs collected for the stalled TnaC-70S complexes rapidly yielded a 5.8 Å-resolution single particle reconstruction, in which not only a P-site tRNA was visible, but an extended nascent peptide could be seen spanning the entire length of the ribosomal exit tunnel (Fig. 13.4a) [43]. The high quality of the map revealed distinct points of contact between the nascent chain and ribosomal components of the exit tunnel. At the peptidyl-transferase center of the ribosome, the universally conserved A2602 and U2585 were shown to adopt conformations that restrict the access of the catalytic GGQ loop of the termination release factors to the PTC. A model was proposed to explain how the coordinated jamming of the peptide inside the ribosomal exit tunnel ultimately leads to the complete inactivation of the PTC and to the shutdown of translation (Fig. 13.4b, c).

In the future, we seek to obtain a full atomic model of a stalled RNC complex by X-ray crystallography in order to establish the structural mechanism by which the stalling of polypeptide synthesis is achieved. Nascent chains of interest include TnaC and more recently ErmCL, an arresting peptide that leads to ribosome stalling in the presence of the macrolide erythromycin [58]. When work on the ribosome nascent chain project was initiated in our laboratory, it was not known whether

nascent peptides could be visualized inside a translating ribosome. By providing the first direct evidence that this is indeed possible, our latest findings suggest that a complete molecular understanding of this process is now within reach.

Acknowledgments This work was supported by the U.S. National Institutes of Health grant GM 022778 to T.A.S.

References

1. Allen GS, Zavialov A, Gursky R, et al (2005) The cryo-EM structure of a translation initiation complex from Escherichia coli. Cell 121:703–712
2. An G, Glick BR, Friesen JD et al (1980) Identification and quantitation of elongation factor EF-P in Escherichia coli cell-free extracts. Can J Biochem 58:1312–1314
3. Aoki H, Dekany K, Adams SL et al (1997) The gene encoding the elongation factor P protein is essential for viability and is required for protein synthesis. J Biol Chem 272:32254–32259
4. Aoki H, Xu J, Emili A et al (2008) Interactions of elongation factor EF-P with the Escherichia coli ribosome. FEBS J 275:671–681
5. Bashan A, Agmon I, Zarivach R, Schluenzen F, Harms J, Pioletti M, Bartels H, Gluehmann M, Hansen H, Auerbach T et al (2001) High-resolution structures of ribosomal subunits: initiation, inhibition, and conformational variability. Cold Spring Harb Symp Quant Biol 66:43–56
6. Ban N, Freeborn B, Nissen P et al (1998) A 9 Å resolution X-ray crystallographic map of the large ribosomal subunit. Cell 93:1105–1115
7. Ban N, Nissen P, Hansen J et al (1999) Placement of protein and RNA structures into a 5 Å resolution map of the 50S ribosomal subunit. Nature 400:841–847
8. Ban N, Nissen P, Hansen J et al (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. Science 289:905–920
9. Barta A, Steiner G, Brosius J et al (1984) Identification of a site on 23S ribosomal RNA located at the peptidyl transferase center. Proc Natl Acad Sci USA 81:3607–3611
10. Beringer M (2001) Modulating the activity of the peptidyl transferase center of the ribosome. RNA 14:795–801
11. Berisio R, Harms J, Schluenzen F et al (2003) Structural insights into the antibiotic action of telithromycin against resistant mutants. J Bacteriol 185(14):4276–4279
12. Blaha G, Gurel G, Schroeder SJ et al (2008) rRNA mutations far from antibiotic binding sites can make ribosomes drug-resistant. J Mol Biol 379:505–519
13. Blaha G, Stanley RE, Steitz TA (2009) Formation of the first peptide bond: the structure of EF-P bound to the 70S ribosome. Science 325: 966–970
14. Borovinskaya MA, Shoji S, Fredrick K, Cate JH (2008) Structural basis for hygromycin B inhibition of protein biosynthesis. RNA 14(8): 1590–1599
15. Bulkley D, Innis CA, Blaha G, Steitz TA (2010) Revisiting the structures of several antibiotics bound to the bacterial ribosome. Proc Natl Acad Sci USA 107(40):17158–17163
16. Carter AP, Clemons WM, Bordensen DE et al (2000) Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. Nature 407:340–348
17. Cole JR, Olsson CL, Hershey JW, Grunberg-Manago M, Nomura M (1987) Feedback regulation of rRNA synthesis in Escherichia coli. Requirement for initiation factor IF2. J Mol Biol 198:383–392
18. Cruz-Vera LR and Yanofsky C (2008) Conserved residues Asp16 and Pro24 of TnaC-tRNAPro participate in tryptophan induction of Tna operon expression. J Bacteriol 190: 4791–4797
19. Dorner S, Panuschka C, Schmid W, Barta A (2003) Mononucleotide derivatives as ribosomal P-site substrates reveal an important contribution of the 2'-OH to activity. Nucleic Acids Res 31:6536–6542
20. Franceschi F, Weistein S, Evers U et al (1993) Towards atomic resolution of prokaryotic ribosomes: crystallographic, genetic and biochemical studies. In: Nierhaus KH, Franceschi

F, Subramanian AR, Erdman VA, Wittmann-Liebold B (eds) The translational apparatus. Plenum Press, New York, pp 397–410
21. Gong F, Yanofsky C (2002) Instruction of translating ribosome by nascent peptide. Science 297:1864–1867
22. Hansen JL, Ippolito JA, Ban N et al (2002) The structures of four macrolide antibiotics bound to the large ribosomal subunit. Mol Cell 10:117–128
23. Hansen JL, Moore PB, Steitz TA (2003) Structures of five antibiotics bound at the peptidyl transferase center of the large ribosomal subunit. J Mol Biol 330:1061–1075
24. Harms J, Schluenzen F, Zarivach R et al (2001) High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. Cell 107:679–688
25. Ito K, Chiba S, Pogliano K (2010) Divergent stalling sequences sense and control cellular physiology. Biochem Biophys Res Commun 393:1–5
26. Klein DJ, Schmeing TM, Moore PB, Steitz TA (2001) The Kink-turn: a new RNA secondary structure motif. EMBO J 20:4214–4221
27. Klein DJ, Moore PB, Steitz TA (2004) The roles of ribosomal proteins in the structure, assembly and evolution of the large ribosomal subunit. J Mol Biol 340:141–177
28. Korostolev A, Trakhanov SK, Laurberg M, Noller HF (2006) Crystal structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements. Cell 126:1066–1077
29. Lovett PS and Rogers EJ (1996) Ribosome regulation by the nascent peptide. Microbiol Rev 60:366–385
30. Mozad D, Noller HF (1987) Chloramphenicol, erythromycin, carbomycin and vernamycin B protect overlapping sites in the pepticyl transferase region of 23S ribosomal RNA. Biochimie 69(8):879–884
31. Nissen P, Hansen J, Ban N et al (2000) The structural basis of ribosome activity in peptide bond synthesis. Science 289:920–930
32. Nissen P, Ippolito JA, Ban N et al (2001) RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. Proc Natl Acad Sci USA 98:4899–4903
33. Ogle JM, Brodersen DE, Clemons WM Jr et al (2001) Recognition of cognate transfer RNA by the 30S ribosomal subunit. Science 292:897–902
34. Peske F, Savelsbergh A, Katunin VI et al (2004) Conformational changes of the small ribosomal subunit during elongation factor G-dependent tRNA-mRNA translocation. J Mol Biol 343:1183–1194
35. Ramu H, Mankin A, Vazquez-Laslop N (2009) Programmed drug-dependent ribosome stalling. Mol Microbiol 71:811–824
36. Schluenzen F, Hansen HAS, Thygesen J et al (1995) A milestone in ribosomal crystallography: the construction of preliminary electron density maps at intermediate resolution. Biochem Cell Biol 73:739–749
37. Schluenzen F, Zarivach R, Harms J et al (2001) Structural basis for the interaction of antibiotics with the peptidyl transferase center in eubacteria. Nature 413:814–821
38. Schlunzen F, Harms JM, Franceschi F, et al (2003) Structural basis for the antibiotic activity of ketolides and azalides. Structure 11(3):329–338
39. Schmeing TM, Huang KS, Kitchen DE et al (2005) Structural insights into the roles of water and the 2' hydroxyl of the P site tRNA in the peptidyl transferase reaction. Mol Cell 20:437–448
40. Schmeing TM, Huang KS, Strobel SA, Steitz TA (2005) An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA. Nature 438:520–524
41. Schmeing TM, Seila AC, Hansen JL et al (2002) A pre-translocational intermediate in protein synthesis observed in crystals of enzymatically active 50S subunits. Nat Struct Biol 9:225–230
42. Schuwirth BS et al (2005) Structures of the bacterial ribosome at 3.5 Å resolution. Science 310:827–834
43. Seidelt B, Innis CA, Wilson DN et al (2009) Structural insight into nascent polypeptide chain-mediated translational stalling. Science 326:1412–1415
44. Selmer M et al (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. Science 313:1935–1942
45. Shevack A, Gewitz HS, Hennemann B et al (1985) Characterization and crystallization of ribosomal particles from Haloarcula marismortui. FEBS Lett 184:68–71

46. Shimizu Y, Inoue A, Tomari Y et al (2001) Cell-free translation reconstituted with purified components. Nat Biotechnol 19:751–755
47. Shoji S, Walker SE, Fredrick K (2009) Ribosomal translocation: one step closer to the molecular mechanism. ACS Chem Biol 4:93–107
48. Simonetti A, Marzi S, Jenner L et al (2009) A structural view of translation initiation in bacteria. Cell Mol Life Sci 66:423–436
49. Simonetti A, Marzi S, Myasnikov AG et al (2008) Structure of the 30S translation initiation complex. Nature 455:416–420
50. Simonovic M, Steitz TA (2008) Cross-crystal averaging reveals that the structure of the peptidyl-transferase center is the same in the 70S ribosome and the 50S subunit. Proc Natl Acad Sci USA 105:500–505
51. Stanley RE, Blaha G, Grodzicki RL et al (2010) The structures of the anti-tuberculosis antibiotics viomycin and capreomycin bound to the 70S ribosome. Nat Struct Mol Biol 17:289–293
52. Steitz TA (2008) A structural understanding of the dynamic ribosome machine. Nat Rev Mol Cell Biol 9:243–253
53. Szaflarski W, Vesper O, Teraoka Y et al (2008) New features of the ribosome and ribosomal inhibitors: non-enzymatic recycling, misreading and back-translocation. J Mol Biol 380:193–205
54. Tenson T, Ehrenberg M (2002) Regulatory nascent peptides in the ribosomal tunnel. Cell 108:591–594
55. Tu D, Blaha G, Moore PB, Steitz TA (2005) Gene replacement in Haloarcula marismortui: construction of a strain with two of its three chromosomal rRNA operons deleted. Extremophiles 9:427–435
56. Tu D, Blaha G, Moore PB, Steitz TA (2005) Structures of MLSBK antibiotics bound to mutated large ribosomal subunits provide a structural explanation for resistance. Cell 121:257–270
57. van Bohlen K, Makowski I, Hansen HAS et al (1991) Characterization and preliminary attempts for derivatization of crystals of large ribosomal subunits from Haloarcula marisumortui diffracting to 3 Å resolution. J Mol Biol 222:11–15
58. Vazquez-Laslop N, Thum C, Mankin AS (2008) Molecular mechanism of drug-dependent ribosome stalling. Mol Cell 30:190–202
59. Voorhees RM, Weixlbaumer A, Loakes D et al (2009) Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. Nat Struct Mol Biol 16:528–533
60. Weinger JS, Parnell KM, Dorner S et al (2004) Substrate-assisted catalysis of peptide bond formation by the ribosome. Nat Struct Mol Biol 11:1101–1106
61. Wilson DN (2009) The A-Z of bacterial translation inhibitors. Crit Rev Biochem Mol Biol 44:393–433
62. Wimberly BT, Brodersen DE, Clemons WM Jr et al (2000) Structure of the 30S ribosomal subunit. Nature 407:327–339
63. Wittmann-Liebold B (1986) Ribosomal proteins: their structure and evolution. In: Hardesty B, Karmer G (eds) Structure, function and genetics of ribosome. Springer-Verla, New York, pp 324–361
64. Yonath A (2005) Antibiotics targeting ribosomes: resistance, selectivity, symergism and cellular regulation. Annu Rev Biochem 74:649–679
65. Yonath A, Mussig J, Tesche B et al (1980) Crystallization of the large ribosomal subunits from Bacillus stearothermophilus. Biochem Int 1:428–435
66. Yusupov MM, Yusupova GZ, Baucom A et al (2001) Crystal structure of the ribosome at 5.5 Å resolution. Science 292:883–8986

Chapter 14
Proteopedia: Exciting Advances in the 3D Encyclopedia of Biomolecular Structure

Jaime Prilusky, Eran Hodis, and Joel L. Sussman

Abstract *Proteopedia* is a collaborative, 3D web-encyclopedia of protein, nucleic acid and other structures. *Proteopedia* (http://www.proteopedia.org) presents 3D biomolecule structures in a broadly accessible manner to a diverse scientific audience through easy-to-use molecular visualization tools integrated into a wiki environment that anyone with a user account can edit. We describe recent advances in the web resource in the areas of content and software. In terms of content, we describe a large growth in user-added content as well as improvements in automatically-generated content for all PDB entry pages in the resource. In terms of software, we describe new features ranging from the capability to create pages hidden from public view to the capability to export pages for offline viewing. New software features also include an improved file-handling system and availability of biological assemblies of protein structures alongside their asymmetric units.

Keywords Structural biology • Communication • 3D • Dissemination • Wiki • *Proteopedia* • Education • Instruction • Collaboration • Encyclopedia • Web resource • Jmol • PDB

J. Prilusky
Bioinformatics Unit, Department of Biological Services, The Israel Structural Proteomics Center, Weizmann Institute of Science, Rehovot 76100, Israel

E. Hodis
Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel

J.L. Sussman (✉)
Department of Structural Biology, The Israel Structural Proteomics Center, Weizmann Institute of Science, Rehovot 76100, Israel
e-mail: joel.sussman@weizmann.ac.il

M.A. Carrondo and P. Spadon (eds.), *Macromolecular Crystallography*,
NATO Science for Peace and Security Series A: Chemistry and Biology,
DOI 10.1007/978-94-007-2530-0_14, © Springer Science+Business Media B.V. 2012

14.1 Introduction

Proteopedia [1] (http://www.proteopedia.org) is a collaborative, 3D web-encyclopedia of biomolecular structures. While structures of proteins, nucleic acids and other biomolecules are often conveyed using two dimensional images in journal articles, textbooks and even in lectures, such images can often be difficult to comprehend because they flatten a 3D structure into two dimensions. Structural biologists and chemists utilize a wide array of molecular visualization programs to navigate biomolecular structures in 3D, but these programs are generally inaccessible to those scientists and students without a background in structural biology. *Proteopedia* presents 3D biomolecular structures in a broadly accessible manner to a diverse scientific audience through easy-to-use molecular visualization tools (mostly Jmol [2], some Kinemage [3]) integrated into a wiki environment that anyone with a user account can edit. Structural information is intuitively communicated through rotatable and zoomable 3D structures displayed on *Proteopedia* pages adjacent to descriptive text containing links called 'scene-links' that when clicked elicit a change in the view, colors, and representation of the 3D structure, highlighting a point made in the text (Fig. 14.1).

Proteopedia's strength lies in the ease with which a user can both explore a protein structure as well as contribute his or her own description and structural annotation using *Proteopedia*'s "Scene Authoring Tools" to create 'scene-links'. Additional features of the website complement this strength. For instance, in order to appeal to the scientific community, pages in *Proteopedia* are editable only by registered users, who register for free user accounts using their full real names and with biographical informational about their scientific background. At the bottom of every *Proteopedia* page is an automatically generated list of the users who have edited that page, giving both credit and responsibility to *Proteopedia* users, as well as providing a measure of the reliability of the page. Each user has an area where he or she can create pages that only he or she can edit, to allow for the creation of tutorials for the classroom or of material for projection during a lecture, over which the user would understandably need complete control. While the focus is on user-added pages, each entry in the PDB (over 65,000) has a *Proteopedia* page automatically generated for it, seeded with relevant information and primed for expansion by *Proteopedia* users. Such automatically seeded PDB entry pages typically contain a rotatable, zoomable 3D structure next to the abstract from the publication describing the structure, along with 'scene-links' highlighting ligands or functional sites, structural and functional annotation aggregated from various resources, links to view the PDB entry in various useful resources and links to download the PDB entry. These and other *Proteopedia* features have been described in detail previously [1], but *Proteopedia* is constantly evolving with pages added and edited daily, as well as frequent improvements to the website's software. In this chapter we discuss several advances in *Proteopedia* content and software.

In terms of content, *Proteopedia* has advanced both in the number and quality of user-added pages as well as in the automatically generated information provided on

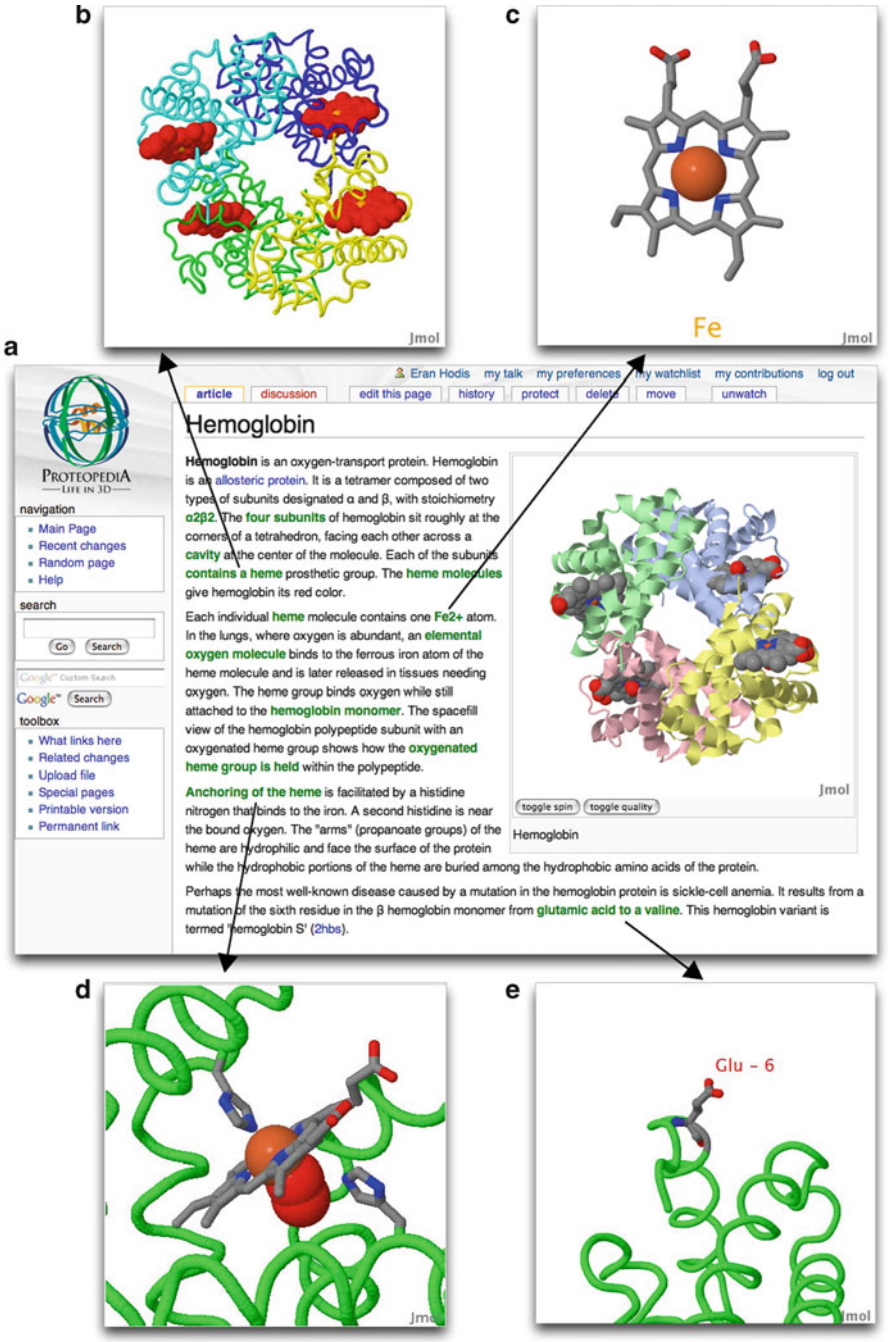


Fig. 14.1 Green links change from one easily-authored molecular scene to another Thus, text discussing and describing the structure and function is reinforced by immediate and significant visual input (Reproduced from Hodis et al. [1])

each page describing a PDB entry. When *Proteopedia* was first presented to the international scientific community at the 40th Crystallographic Meeting at Erice "From Molecules to Medicine" (2008), the resource's informational value stemmed mostly from the automatically generated pages for each entry in the PDB, rather than from the tens of user-added pages that existed at that point.

Since then, *Proteopedia* user-added pages have gone from tens of pages to well over two hundred pages. While some of these pages are of higher quality than others, the growth in the number of contributions and in the use of the site are promising and inviting. The types of pages that have been added span from encyclopedic pages on particular protein structures, through tutorial pages for use in classroom settings to supplementary material pages accompanying publications in scientific journals. *Proteopedia's* automatically generated pages for each entry in the PDB have benefitted from a collaboration with the ConSurf Team, and now each such *Proteopedia* page offers a view of the 3D structure with each residue colored by its level of evolutionary conservation [4, 5].

In terms of software, *Proteopedia* has added many new capabilities: A 'Workbench' area now allows users to create pages that are hidden from public view or shared with a select group of users. An 'Export' feature allows a user to save a page to his or her computer for offline viewing. 'Scene-links' have now been optimized and protected against future PDB remediations that have in the past, have been disruptive. The *Proteopedia* Scene Authoring Tools (SAT) now include "Undo" and "Redo" buttons and tooltips for several of the SAT buttons, as well as the option to color a protein's residues based on their level of evolutionary conservation according to ConSurf [4, 5]. Citing of scientific publications within the body of a *Proteopedia* page has been simplified and users can simply indicate a publication's PubMed ID to have the full reference appear in the page's 'References' section. Visualization advances include a 'Pop-up' button to allow for expansion of any *Proteopedia* Jmol applet and the beginnings of color-keys associated with 'scene-links'. The addition of a semi-automatic mechanism for users to reserve blocks of many 'Sandbox' pages aids educators in cordoning off pages for use in workshops or for class projects. Biological assemblies are more functionally relevant than asymmetric units, and they are now available in *Proteopedia* [6, 7]. Finally, a translator from PyMOL [8] to Jmol allows a user to create 'scene-links' in *Proteopedia* using PyMOL ".pse" session files (in preparation for publication).

14.2 Advances in Content

14.2.1 User-Added Content

The over 800 registered *Proteopedia* users have created a variety of pages on the web resource, from pages describing a particular protein structure to pages serving as tutorials for university courses. Many of these pages have been organized into

the *Proteopedia* Table of Contents [9], accessible via a link on the left-hand side of any *Proteopedia* page. Although because the Table of Contents is manually curated, and *Proteopedia* is a growing resource, new pages may have yet to be added to the Table of Contents.

A page describing the ribosome is currently featured on the Main Page of *Proteopedia* [10, 11] and previously featured articles include 'Swine Flu, Neuraminidase & Tamiflu' and 'Poly-A Polymerase' [12]. Other new additions include pages on DNA [13], Lac Repressor [14], Mechanosensitive Channels [15] and HIV-1 Protease [16] and others (see What's New [17] and Topic Pages [18]).

Tutorials created by one educator are often useful for other educators, and *Proteopedia* users are encouraged to share and adapt tutorials on the site. Examples of new tutorials include a tutorial on Serine Proteases [19] and a tutorial on Structural Templates [20]. Other tutorials and pages created by educators for teaching are listed on the site as well [21].

Educators have also assigned class projects involving the creation of *Proteopedia* pages. Courses and programs making use of *Proteopedia* range from the high-school level [22] to the undergraduate level (see GFP [23] and Proteins involved in cancer [24], in particular the pages linked to from that page are all created by undergraduate students) to the graduate level (see Triose_Phosphate_Isomerase [25] and CBI_ Molecules [26]).

Structural biologists have also created pages in *Proteopedia* as supplementary material for publications in scientific journals. The page '3btp' describes the crystal structure of *Agrobacterium tumefaciens* VirE2 in complex with its chaperone VirE1 [27] and serves as supplementary material for an article in *PNAS* [28]. Another page describes engineered mutants of *B. cereus* HlyIIR, a member of the TetR family of dimeric transcriptional regulators [29], and serves as supplementary material for a recent article in *Proteins* [30].

14.2.2 Coloring by Evolutionary Conservation for all PDB Entries (ConSurf)

Of tantamount importance to a biologist examining a solved protein structure is how that protein's function relates to its structure. Biochemical and molecular biology studies best illuminate a protein's structure-function relationship, but computational analysis can also be informative, especially when the more laborious experimental work has yet to be carried out.

Evolutionary conservation of specific protein residues can often indicate functional importance. If certain residues have been protected from change over millions of years of evolution, there is a good chance that they are crucial for protein folding or difficult to comprehend function or protein folding. Almost all *Proteopedia* pages titled for a PDB entry (for example '1h88') now have a link underneath the 3D structure titled "Evolutionary conservation [show]". Clicking on "[show]" colors the residues of the rotating 3D structure by their level of evolutionary conservation

according to ConSurfDB [4, 5] (see: http://consurfdb.tau.ac.il), on a scale from 1 (Variable, teal-colored) to 5 (Average, white-colored) to 9 (Conserved, magenta-colored), with a color key explaining the coloring scheme. Briefly, ConSurfDB assigns conservation levels to a protein sequence by multiple sequence alignment to similar sequences followed by Bayesian analysis of the level of conservation of each residue (for a more detailed explanation see ConSurf-DB Process [31] or ConSurf [4, 5]). Residues for which there are insufficient data are colored yellow, and residues for which there are no data are colored grey. Clicking on "[show]" also expands a previously hidden "Evolutionary conservation display area" containing buttons for toggling the conservation colors on and off for different chains of the structure, allowing separate examination of individual chains separately. It should be noted that in *Proteopedia*'s coloring by evolutionary conservation, all the chains in the molecule are colored as such, but this can be misleading since ConSurf calculates evolutionary conservation independently for sequence-different chains, thus the scale of conservation may be different between chains, although the colors are the same (ConSurfDB avoids this problem by only displaying one chain at a time). Links in the "Evolutionary conservation display area" offer the user a page describing this and other caveats [32], a page with a detailed explanation of ConSurf in *Proteopedia* [33], and a page at ConSurfDB with complete results for the PDB entry being viewed (for example [34]).

14.3 Advances in Software and Development

14.3.1 Pages Hidden from Public View

While previously a user could create a page that only he or she could edit, that page would still be publicly visible. Public visibility of page content is problematic in certain cases, especially when developing a page for use as supplementary material to accompany a journal article submission.

A new feature in *Proteopedia* called the "Workbench" feature allows users to create hidden pages that they can either keep private or share with selected users. A Workbench page is created as a sub-page of a user's userpage. Each user has a userpage created in his or her name when he or she joins the site, for example 'User:Eran Hodis', and he or she can create subpages of that userpage such as 'User:Eran Hodis/Hemoglobin'. The Workbench area is such a subpage: 'User:Eran Hodis/Workbench', and any subpage of the Workbench area is similarly protected: 'User:Eran Hodis/Workbench/SupplementaryPNAS'. A page in the Workbench can be shared with selected users by clicking on the 'Workbench' tab near the top of the page and entering the usernames to whom the page should be also visible (Fig. 14.2). Different subpages in the Workbench area can be assigned to be visible to different users. Files, including images or PDB files, can be uploaded with the prefix 'Workbench' in order to similarly protect them from public viewing.

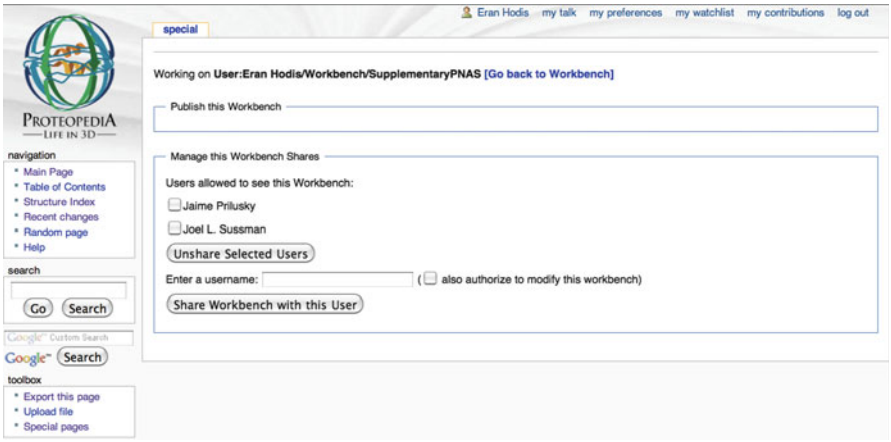


Fig. 14.2 The "Workbench" interface allowing a user to choose with whom to share a Workbench page. In this case the user "Eran Hodis" has shared the page "User:Eran Hodis/Workbench/SupplementaryPNAS" with the users "Jaime Prilusky" and "Joel L. Sussman"

14.3.2 Save Pages for Offline Viewing

Online resources such as *Proteopedia* depend on the user having a working Internet connection in order to access content. Simple web pages that are composed only of text and HTML can easily be saved locally to a user's computer for offline viewing, but *Proteopedia* pages contain complex content like Java applets used to run Jmol or Kinemage to visualize 3D structures on the page.

Proteopedia's 'Export' feature allows a user to save a *Proteopedia* page, including visualizations of 3D structures, to his or her computer for offline viewing. Clicking on the "Export this page" link visible on the left hand side of any *Proteopedia* page (Fig. 14.3) brings up a dialog asking the user where he or she wants to save the page on his or her computer. The page is then scanned for all relevant Java applets, Javascript files, Jmol scripts, and other files necessary for reliable offline viewing, and these files are packaged and downloaded to the user's computer into a folder with the same name as the *Proteopedia* page, along with the textual content of the page itself. Inside the downloaded folder, opening the file "index.html" provides an offline replica of the exported *Proteopedia* page.

This 'Export' feature allows *Proteopedia* users to store pages offline, whether for perusal during times when an Internet connection is unavailable or for a backup in case of Internet failure when using *Proteopedia* pages in a lecture or presentation.

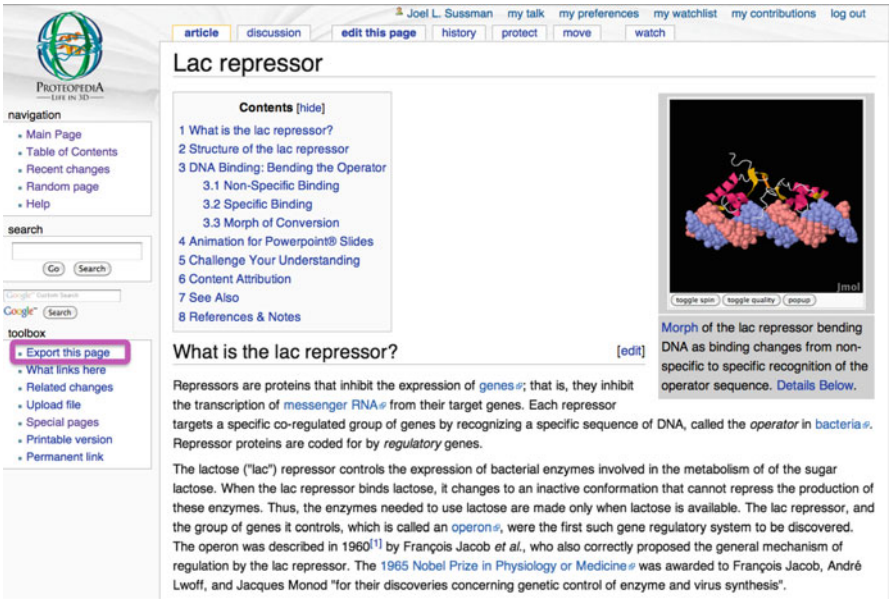


Fig. 14.3 View of a page with the Export this page highlighted for the Lac repressor page in *Proteopedia* [14]

14.3.3 Scenes Optimized and Protected Against Changes in PDB Files

The PDB remediation, released on March 17, 2009, caused problems across many 'scene-links' on many *Proteopedia* pages. The root of the problem was that at the time *Proteopedia* 'scene-links' did not save the version of the PDB file that was used to create them. This meant that a 'scene-link' that loaded the nucleosome structure in PDB entry 1aoi simply loaded the latest version of PDB entry 1aoi before recalling the view, coloring, and representations stored in the scene script. When the latest version of a PDB entry changed in the remediation, records such as chain names might have changed in that file, but many commands in the scene script are chain-specific, and were not changed. Thus, the large-scale PDB remediation resulted in many 'scene-links' failing to properly recall their saved scene script.

Proteopedia's solution to this problem, and to future remediations, is to save the files (PDB or otherwise) that each 'scene-link' loads, at the time that the scene is saved. This "freezes" the files in the state they were in when the scene was created, thus avoiding any future problems due to remediation. Before "freezing" a new file, *Proteopedia* first checks whether that same file has already been frozen, avoiding duplicates and ensuring that scenes that load the same file are aware of that fact. A smooth transition between 'scene-links' is crucial for *Proteopedia's* intuitive feel, and by making sure that two scenes that load the same structure are aware of

this commonality, we prevent reloading of the same structure when transitioning between these two scenes, which would result in a choppy transition.

14.3.4 Improvements to the Scene Authoring Tools

The *Proteopedia* Scene Authoring Tools (SAT) are central to *Proteopedia's* collaborative features – they provide users with the ability to easily create, edit, and add 'scene-links' to any *Proteopedia* page in order to describe and annotate 3D protein structures. Several recent additions to the SAT merit mention: "Undo/Redo" buttons, explanatory "Tooltips", and the option to choose an evolutionary conservation color scheme, according to ConSurf.

The new "Undo" and "Redo" buttons in the SAT allow a user to undo and redo any changes they may have made to the scene they are currently creating using the SAT. In choosing the perfect view, colors, representations and labels for a particular scene, sometimes mistakes are made. Before the existence of an "Undo" button, mistakes may have sometimes been frustrating to fix. Now, all changes made to the scene displayed in the SAT Jmol applet are recorded and are undoable, regardless of whether they have been made using the SAT or using the Jmol console.

The SAT offers many options for creating a revealing scene of a biomolecule structure, but with a wealth of options comes confusion. Beginning with many of the buttons visible on the "colors" tab that allow the user to choose different coloring schemes, buttons and options in the SAT now can include explanatory tooltips that appear when the user hovers over them. These tooltips will slowly spread from the "colors" tab to other areas of the SAT and are editable by *Proteopedia* users, to allow for user improvement of SAT help features.

One of the coloring schemes with a new tooltip of its own is the "evolutionary conservation" coloring scheme button offered on the SAT "colors" tab. Clicking on this button colors the residues of the current selection in the Jmol applet by their level of evolutionary conservation, according to ConSurf, on a scale from 1 (Variable, teal-colored) to 5 (Average, white-colored) to 9 (Conserved, magenta-colored). It is recommended that users include a color key in their *Proteopedia* page when including scenes with residues colored by evolutionary conservation.

14.3.5 Simple Citations Using PubMed ID Numbers

Proper references are important to any scientific writing, but adding a reference longhand is a laborious process prone to error. A new mechanism in *Proteopedia* allows users to add references using their PubMed ID alone. For instance, to cite the 2005 ConSurf paper [5] in a *Proteopedia* page, whereas previously a user would have to type out the entire reference between reference tags like so:

```
<ref>Landau M., Mayrose I., Rosenberg Y., Glaser F., Martz E., Pupko T. and Ben-Tal N. 2005. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucl. Acids Res. 33:W299-W302.</ref>
```

Now using the PubMed ID in the following format suffices to add the same reference to the page:

```
<ref>PMID:15980475</ref>
```

Such references can be placed anywhere when editing a *Proteopedia* page, and they will show up properly numbered and aggregated at the bottom of the *Proteopedia* page, wherever the user places the wikitext "</references>". Additional information on citing references in *Proteopedia* is available at Help:Editing#Citing_Literature_References [35].

14.3.6 Visualization Advances

Additional visualization advances include a "Pop-up" button underneath every *Proteopedia* Jmol applet, and a preliminary working version of "Color Keys" in *Proteopedia*.

The "Pop-up" button appears underneath every Jmol applet in *Proteopedia* and when clicked results in a large Jmol applet pop-up window. It provides a much larger view of the displayed structure and can be re-sized. Clicking on "refresh model" in the pop-up window results in the pop-up Jmol applet refreshing to emulate the display in the original Jmol applet.

"Color Keys" in *Proteopedia* are a way to provide a legend for each 'scene-link' since often each scene-link can highlight many complex structural elements with several different distinguishing colors. Users have been working around the lack of "Color Keys" in *Proteopedia* by using colored text next to the scene-link (for example Lac_Repressor [14]), which is acceptable, but not ideal as too much colored text can confuse the reader. "Color Keys" will allow the user to specify a unique color key for each 'scene-link' at the time the user saves the 'scene-link' using the SAT. Then, as each 'scene-link' is clicked by the viewing user, the appropriate color key appears underneath the Jmol applet as the scene is loaded. This feature is still in development, but a preliminary test has been created publicly, here: for Gramicidin Channel in Lipid Bilayer [36].

14.3.7 Semi-Automatic Reservation of Sandbox Pages for Courses

Users can now semi-automatically reserve a block of continuous 'Sandbox' pages (e.g. 'Sandbox 50' through 'Sandbox 100') in *Proteopedia*. Sandbox pages are typically used as a place to experiment and create content without committing to its reliability. Educators and lecturers often need to reserve such a block of Sandbox pages for use by their students in workshops and class projects, but previous reservation was done by hand (*i.e.*, an educator would have to find out which pages are not being used, and edit these pages with a note indicating their now-reserved status). The new mechanism allows a registered user to reserve Sandbox pages by filling out a form including general information on the course, the number of Sandbox pages needed and the information that they would like to have appear on each Sandbox page. The process is referred to as semi-automatic because a *Proteopedia* Administrator must review and approve the request before it is filled. Future improvements would allow automatic approval of Sandbox reservation for a select group of trusted educators.

14.3.8 Biological Assemblies vs Asymmetric Units

The biological assembly of a protein represent its functional form, whereas the asymmetric unit of a protein crystal structure does not. The asymmetric unit may not be relevant at all to a non-structural biologist. With this in mind, *Proteopedia* has added the option to load or display a PDB entry's biological assembly as well as its asymmetric unit [6, 7]. This option appears in the *Proteopedia* SAT for creating 'scene-links' using biological assemblies and will soon appear by default on all automatically seeded pages for PDB entries.

14.4 Conclusions

The advances in both content and software described here help *Proteopedia* to achieve its goal of making both structural information and its relationship to functional information accessible and understandable to a broad scientific audience. Future developments and improvements to the web resource will similarly be judged on their ability to help strive toward this goal.

Acknowledgements We acknowledge financial support by the Israel Structural Proteomics Center, the Divadol Foundation, the Nalvyco Foundation, the Bruce Rosen Foundation, the Jean and Jula Goldwurm Memorial Foundation, the Kimmelman Center for Biomolecular Structure and Assembly, the Neuman Foundation, the European Commission Sixth Framework Research and Technological Development Programme 'SPINE2-COMPLEXES' Project under contract

No. 031220 and the 'Teach-SG' Project under contract No. ISSG- CT-2007-037198. JLS is the incumbent of the Morton and Gladys Pickman Chair of Structural Biology. We thank Eric Martz, John Moult, Israel Silman, Roni Gordon, Adityha Sagar, Judith Voet, Don Voet and Warren DeLano - the creator of PyMOL – for useful suggestions.

References

1. Hodis E, Prilusky J, Martz E, Silman I, Moult J, Sussman JL (2008) Proteopedia – a scientific 'wiki' bridging the rift between 3D structure and function of biomacromolecules. Genome Biol 9:R121
2. Jmol: an open-source Java viewer for chemical structures in 3D. http://www.jmol.org. Accessed 9 Nov 2011
3. Richardson DC, Richardson JS (1992) The kinemage: a tool for scientific communication. Protein Sci 1:3–9
4. Glaser F, Pupko T, Paz I, Bell RE, Bechor-Shental D, Martz E, Ben-Tal N (2003) ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics 19:163–164
5. Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T, Ben-Tal N (2005) ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucleic Acids Res 33:W299–W302
6. Berman H, Henrick K, Nakamura H, Markley JL (2007) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. Nucleic Acids Res 35:D301–D303
7. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. J Mol Biol 372:774–797
8. DeLano WL (2002) The PyMOL molecular graphics system. DeLano Scientific, Palo Alto. http://www.pymol.org. Accessed 9 Nov 2011
9. *Proteopedia*: Table of contents. http://www.proteopedia.org/wiki/index.php/Table_of_Contents. Accessed 9 Nov 2011
10. *Proteopedia*: Main page. http://www.proteopedia.org. Accessed 9 Nov 2011
11. *Proteopedia*: Ribosome. http://www.proteopedia.org/wiki/index.php/Ribosome. Accessed 9 Nov 2011
12. *Proteopedia*: Featured article archives. http://www.proteopedia.org/wiki/index.php/Proteopedia:Featured_article_archives. Accessed 9 Nov 2011
13. *Proteopedia*: DNA. http://www.proteopedia.org/wiki/index.php/DNA. Accessed 9 Nov 2011
14. *Proteopedia*: Lac repressor. http://www.proteopedia.org/wiki/index.php/Lac_repressor. Accessed 9 Nov 2011
15. *Proteopedia*: Mechanosensitive channels: opening and closing. http://www.proteopedia.org/wiki/index.php/Mechanosensitive_channels:_opening_and_closing. Accessed 9 Nov 2011
16. *Proteopedia*: HIV-1 protease. http://www.proteopedia.org/wiki/index.php/HIV-1_protease. Accessed 9 Nov 2011
17. *Proteopedia*: What's new. http://www.proteopedia.org/wiki/index.php/Proteopedia:What's_New. Accessed 9 Nov 2011
18. *Proteopedia*: Topic pages. http://www.proteopedia.org/wiki/index.php/Proteopedia:Topic_Pages. Accessed 9 Nov 2011
19. *Proteopedia*: A tutorial of chymotrypsin, trypsin and elastase. http://www.proteopedia.org/wiki/index.php/Serine_Proteases:_A_Tutorial_of_Chymotrypsin,_Trypsin_and_Elastase. Accessed 9 Nov 2011
20. *Proteopedia*: Structural templates. http://www.proteopedia.org/wiki/index.php/Structural_templates. Accessed 9 Nov 2011
21. *Proteopedia*: Teaching scenes, tutorials, and educators' pages. http://www.proteopedia.org/wiki/index.php/Teaching_Scenes,_Tutorials,_and_Educators'_Pages. Accessed 9 Nov 2011

22. *Proteopedia*: 2010 Pingry SMART team. http://www.proteopedia.org/wiki/index.php/2010_Pingry_SMART_Team. Accessed 9 Nov 2011
23. *Proteopedia*: Green fluorescent protein. http://www.proteopedia.org/wiki/index.php/GFP. Accessed 9 Nov 2011
24. *Proteopedia*: Proteins_involved_in_cancer. http://www.proteopedia.org/wiki/index.php/Proteins_involved_in_cancer. Accessed 9 Nov 2011
25. *Proteopedia*: Triose phosphate isomerase. http://www.proteopedia.org/wiki/index.php/Triose_Phosphate_Isomerase. Accessed 9 Nov 2011
26. *Proteopedia*: CBI molecules. http://www.proteopedia.org/wiki/index.php/CBI_Molecules. Accessed 9 Nov 2011
27. *Proteopedia*: 3btp. http://www.proteopedia.org/wiki/index.php/3btp. Accessed 9 Nov 2011
28. Dym O, Albeck S, Unger T, Jacobovitch J, Branzburg A, Michael Y, Frenkiel-Krispin D, Wolf SG, Elbaum M (2008) Crystal structure of the Agrobacterium virulence complex VirE1-VirE2 reveals a flexible protein that can accommodate different partners. Proc Natl Acad Sci USA 105:11170–11175
29. *Proteopedia*: User:Oleg Kovalevskiy/Engineered mutants of HlyIIR. http://www.proteopedia.org/wiki/index.php/User:Oleg_Kovalevskiy/Engineered_mutants_of_HlyIIR. Accessed 9 Nov 2011
30. Kovalevskiy OV, Solonin AS, Antson AA (2010) Structural investigation of transcriptional regulator HlyIIR: influence of a disordered region on protein fold and dimerization. Proteins 78:1870–1877
31. *Proteopedia*: ConSurfDB_vs._ConSurf#ConSurf-DB_Process. http://www.proteopedia.org/wiki/index.php/ConSurfDB_vs._ConSurf#ConSurf-DB_Process. Accessed 9 Nov 2011
32. *Proteopedia*: Conservation, evolutionary: caveats. http://www.proteopedia.org/wiki/index.php/Conservation,_Evolutionary#Caveats. Accessed 9 Nov 2011
33. *Proteopedia*: Conservation, evolutionary. http://www.proteopedia.org/wiki/index.php/Conservation,_Evolutionary. Accessed 9 Nov 2011
34. ConSurf: Complete results: 1h88. http://consurfdb.tau.ac.il/chain_selection.php?pdb_ID=1h88. Accessed 9 Nov 2011
35. *Proteopedia*: Help:Editing#Citing_Literature_References http://www.proteopedia.org/wiki/index.php/Help:Editing#Citing_Literature_References. Accessed 9 Nov 2011
36. *Proteopedia*: Gramicidin channel in Lipid Bilayer. http://www.proteopedia.org/wiki/index.php/Gramicidin_Channel_in_Lipid_Bilayer. Accessed 9 Nov 2011

Chapter 15
Structure Analysis of Biological Macromolecules by Small-Angle X-ray Scattering

Dmitri I. Svergun

Abstract Small-angle X-ray scattering (SAXS) is a low resolution (1–2 nm) structural method, which is applicable to macromolecules in solution providing information about the overall structure and structural transitions. The method covers an extremely broad range of sizes (from a few kDa to hundreds MDa) and experimental conditions (temperature, pH, salinity, ligand addition *etc.*). Recent progress in instrumentation and novel data analysis methods significantly enhanced resolution and reliability of structural models provided by the technique and made SAXS a useful complementary tool to high resolution methods. In particular, SAXS allows for rapid validation of high resolution crystallographic or theoretically predicted models, identification of biologically active oligomers and visualization of missing fragments. Quaternary structure of complexes can be analyzed by rigid body movements/rotations of high resolution models of the individual subunits of domains. The basics of SAXS will be presented and illustrated by advanced applications to macromolecular solutions.

Keywords Solution structure • *Ab initio* methods • Rigid body modeling • Oligomeric mixtures • Flexible proteins

15.1 Introduction

The structural genomics initiatives aiming at large-scale expression and purification for subsequent structure determination using X-ray crystallography (MX) and NMR spectroscopy [10, 12] have already yielded unprecedented numbers of high

D.I. Svergun (✉)
European Molecular Biology Laboratory, Hamburg Outstation, Notkestraße 85,
D-22603 Hamburg, Germany
e-mail: Svergun@EMBL-Hamburg.DE

M.A. Carrondo and P. Spadon (eds.), *Macromolecular Crystallography*,
NATO Science for Peace and Security Series A: Chemistry and Biology,
DOI 10.1007/978-94-007-2530-0_15, © Springer Science+Business Media B.V. 2012

resolution models for isolated proteins and/or their domains These numbers are expected to grow rapidly in the coming years [24]. The focus of modern structural biology has largely shifted towards the study of macromolecular machines accomplishing most important cellular functions [1]. The macromolecular complexes are usually too large for the structural NMR studies, and they often possess inherent structural flexibility making them difficult to crystallize.

The structural analysis approach to macromolecular complexes includes new crystallographic initiatives complemented by the use of methods yielding structural information in solution at lower resolution. In particular, Cryo-EM allows one to obtain excellent results in many cases [39], but it is usually limited to relatively large macromolecular aggregates (starting from 200–300 kDa).

Small-angle X-ray scattering (SAXS) is a rapid method to characterize low resolution structures of individual macromolecules and complexes in solution and to analyse structural changes in response to variation of external conditions. For establishing the three-dimensional structural models this technique needs monodisperse solutions of purified macromolecules but does not require special sample treatment (growth of crystals, isotopic labelling, cryo-freezing *etc.*). SAXS is applicable to a broad range of conditions and sizes (from a few kDa to hundreds MDa). Unlike most other structural methods, SAXS is able to quantitatively characterize equilibrium and non-equilibrium mixtures and monitor kinetic processes such as (dis)assembly and (un)folding.

Recently, the power of SAXS has been boosted by the significant improvements in instrumentation (most notably, by the high brilliance synchrotron radiation sources) accompanied by the development of novel data analysis methods. These developments made it possible to significantly improve resolution and reliability of the structural models constructed from the SAXS data. Here, the main aspects of SAXS including data processing and interpretation procedures and some applications will be briefly reviewed.

15.2 Basics of a SAXS Experiment

This section will briefly describe the basic theoretical and experimental aspects of SAXS to understand the main principles of the technique as applied to solutions of biological macromolecules. The reader is referred to textbooks [11, 14] or to recent reviews [22, 37, 44] for more detailed description information.

Conceptually, a SAXS experiment is rather simple, as illustrated in Fig. 15.1. The samples are exposed to a collimated monochromatic X-ray or neutron beam with the wave vector $k = 2\pi/\lambda$ where λ is the radiation wavelength (Fig. 15.1). The isotropic scattered intensity I is recorded as a function of the momentum transfer $s = 4\pi \sin\theta/\lambda$, where 2θ is the angle between the incident and scattered beam. The scattering from the solvent is measured separately and subtracted to remove the solvent and parasitic background signals.

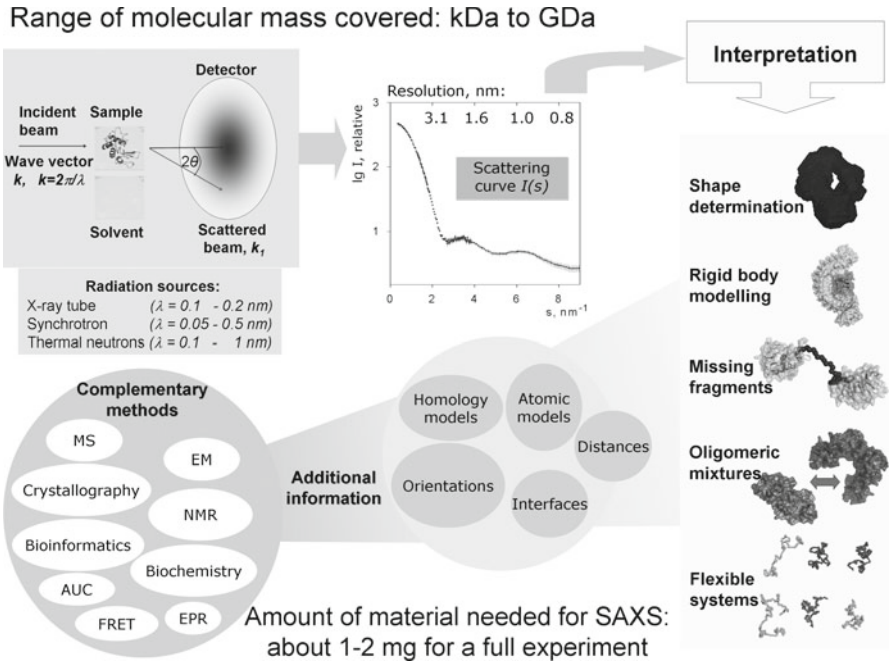


Fig. 15.1 A general scheme of a SAXS/SANS experiment, structural tasks addressed by the technique and its synergistic use with other methods. The nominal resolution of the data in the scattering pattern is indicated as $d = 2\pi/s$. *MS* mass spectroscopy, *AUC* analytical untracentrifugation, *FRET* fluorescence resonance energy transfer, *EM* electron microscopy, *NMR* nuclear magnetic resonance, *EPR* electron paramagnetic resonance

The SAXS experiments are usually performed at synchrotrons, and the experimental stations offering biological SAXS are available at all synchrotron major sites. Laboratory SAXS cameras, (available from various producers) yield much lower beam brilliance but may be useful at least for preliminary analysis. For structure analysis (shape, quaternary structure), the samples with monodispersity better than 90% are required, which must be verified by other methods (native gel filtration, dynamic light scattering, analytical ultracentrifugation) prior to the synchrotron SAXS experiment. Typical concentrations required are in the range of 0.5–10 mg/ml, and a concentration series is usually measured to get rid of interparticle interference effects. The sample volume per measurement on modern stations is about 10–50 μl so that about 1–2 mg of purified material is usually required for a complete study. The upcoming microfluidic devices [48] will allow one to work on high brilliance sources with nanoliter volumes and μg sample amounts.

One should also mention that neutrons are also employed for small-angle scattering (SANS) analysis of biological macromolecules. SANS (which is performed on research reactors or spallation sources) is sensitive to isotopic H/D exchange. This property is exploited for contrast variation involving measurements in different

H_2O/D_2O mixtures and/or specific perdeuteration of subunits, providing unique information about complex particles [51]. The basic equations and the analysis methods are similar for SAXS and SANS.

15.3 Basics of SAXS Data Analysis

The net SAXS intensity after solvent subtraction contains, generally speaking, two contributions. The so-called form factor *I(s)* emerges from the scattering from individual particles in solution and is employed to extract the structural information. The "structure factor" *S(s)* is due to interference effects between the different particles and yields information about the interparticle interactions (see e.g. [22] for a review).

Purified dilute solutions of macromolecules at concentrations in mM range are usually employed in SAXS to get rid of the interference effects and perform the structural studies assuming that I(s) contains only "form-factor" contrinution. Two important cases are distinguished: (1) monodisperse systems, when all the particles are identical and (2) polydisperse systems, when they are different in size and/or shape.

15.4 Monodisperse Systems

For monodisperse solutions, the net intensity *I(s)* is proportional to the scattering from a single particle averaged over all orientations. This allows one to immediately determine the overall geometrical and weight parameters e.g. radius of gyration R_g [16], volume of the hydrated particle V_p [36], and the molecular mass of the particle MM [27]. The Fourier transformation of the scattering intensity provides a characteristic function (averaged Patterson function), which also yields the maximum particle diameter D_{max} [13, 25, 41], Moreover, the low resolution macromolecular shape can be obtained *ab initio* (i.e. without information from other methods). Several approaches have been proposed [2, 7, 8, 18, 43, 46], and *ab initio* shape determination belongs nowadays to routine analysis of the SAXS data. Usually, the shape analysis programs are ran several times and analysed to obtain the most probable and an averaged model [50].

Calculation of the SAXS profiles from atomic models [45] is used to validate theoretically predicted models and verify the structural similarity between macromolecules in crystals and in solution. Moreover, if high resolution models of individual fragments or subunits in a complex are available from crystallography or NMR, rigid body refinement can be employed to model the quaternary structure of the complex. Automated and semi-automated procedures based on screening randomly or systematically generated models were employed by different authors [19, 21, 28]. A comprehensive rigid body modelling program suite is based on the use of spherical harmonics formalism [31, 32, 40, 42].

SAXS is also very useful for the cases when loops or entire domains are missing in high resolution models (e.g. because of flexibility). The missing portions are

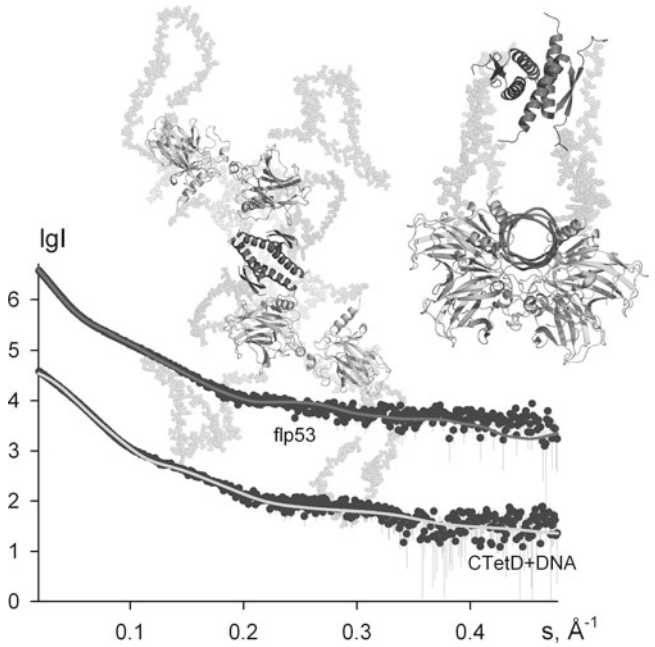


Fig. 15.2 The models of free and DNA-bound tumour suppressor p53 generated by combining SAXS with MX, NMR and EM data. The available high resolution structures of the domains employed in the modelling are displayed as ribbons, the flexible portions of p53 as semi-transparent beads (dummy residues). In the free form (*left panel*) the four domains that recognize DNA (peripheral domains) are arranged in two separate dimers, forming a relatively "flat" structure, with the tetramerization domains in the center. Upon binding to DNA (*right panel*) these core domains wrap around the latter trapping it into a cleft. The models are obtained by rigid body modeling against the SAXS curves displayed as intensity versus momentum transfer in the bottom panel. The data from the free and bound p53 are appropriately displaced along the abscissa axis for better visualization (*dots*, experimental data; *solid lines*, computed patterns from the models)

represented as chains of the so-called dummy residues [34], and the known domains/subunits can be translated and rotated as rigid bodies while simultaneously changing the local conformation of the chains representing the unknown fragments [23]. Numerous applications of rigid body modelling are reported to build structural models of complicated objects in solution (see e.g. references in [33, 37]).

The structural modeling based on the SAXS data is especially effective if one simultaneously incorporates information provided by other methods. An example of such a successful multipronge approach is given by a study of a human tumour suppressor p53 [47], which is a homotetrameric transcription factor (four times 393 residues) playing a central role in the cell cycle. The protein contains a folded core and tetramerization domains, linked and flanked by intrinsically disordered segments. *Ab initio* and rigid body SAXS modelling accounting for NMR-derived interfaces revealed an extended cross-shaped structure with tetrameric contacts and a pair of loosely coupled core domain dimers at the ends (Fig. 15.2). In contrast, the calculated

scattering from a previously published rather compact cryo-EM structure of murine p53 [29] showing dissociated tetramerization domains did not fit the experimental SAXS data. The structure of the complex of p53 with 24 bp DNA independently determined by SAXS and negative stain EM displays a compact complex with the core domains closing around DNA (Fig. 15.2). Interestingly, negatively stained EM analysis of the conformationally mobile, unbound p53 selected a minor compact conformation (less than 20% of the adsorbed particles). The study underlines the significance of the synergistic use of different techniques together with SAXS, in particular, for the structural characterization of a rapidly growing number of proteins with inherent disorder.

15.5 Poyldisperse Systems and Mixtures

For polydisperse systems consisting of different types of non-interacting particles, the measured scattering pattern can be written as a linear combination

$$I(s) = \sum_{k=1}^{K} \nu_k I_k(s) \tag{15.1}$$

where $\nu_k > 0$ and $I_k(s)$ are the volume fraction and the scattering intensity from the k-th type of particle (component), respectively, and K is the number of components.

When neither the number nor intensities of the components are known *a priori*, but multiple data sets are recorded from the system with varying volume fractions of the components, the number of components can be determined extracted by model-independent analysis using singular value decomposition (SVD [15]). If the number of components and their scattering intensities are known, the volume fractions can be readily found by a linear least-squares fit to the experimental data. Numerous applications of these approaches encompass e.g. the analysis of intermediates during folding and assembly processes and quantitative description of oligomeric equilibria [9, 17, 49, 52].

SAXS belongs to very few structural methods able to quantitatively characterize flexible macromolecules, and the method was traditionally used to monitor the processes of protein folding/unfolding [30]. For flexible systems, SAXS data reflect conformational average over the entire ensemble and the scattering patterns are to be interpreted accounting for this average instead of searching for a single model. This has recently become possible with a general approach called ‘ensemble optimization method’ (EOM) allowing for coexistence of multiple conformations [4]. Given a pool of (random) conformers, EOM selects sub-ensembles from them, which, taken as mixtures, fit the experimental profile using Eq. 15.1. The EOM is already actively used to characterize flexible proteins and complexes [5, 26] and it is expected to find broad applications, in particular, in combination with NMR to provide information about both structure and dynamics of the system [3, 6].

15.6 Conclusions

During the last decade, biological SAXS has become increasingly popular in molecular biology revealing low resolution structures of macromolecule in close to native conditions. SAXS can be readily and usefully combined with other computational and experimental techniques to yield comprehensive description of complex objects and processes. The advanced analysis methods are well established by now and are publicly available e.g. in the program package ATSAS (http://www.embl-hamburg.de/biosaxs/). Automated sample changers and pipelines are being developed for high throughput SAXS on synchrotrons [20, 35, 38]. All these developments taken together make the technique readily available for a broad scope of tasks and a broad community of scientists in structural biology.

References

1. Aloy P, Russell RB (2006) Structural systems biology: modelling protein interactions. Nat Rev Mol Cell Biol 7(3):188–197
2. Bada M, Walther D, Arcangioli B, Doniach S, Delarue M (2000) Solution structural studies and low-resolution model of the Schizosaccharomyces pombe sap1 protein. J Mol Biol 300(3):563–574
3. Bernado P, Blanchard L, Timmins P, Marion D, Ruigrok RW, Blackledge M (2005) A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. Proc Natl Acad Sci USA 102(47):17002–17007
4. Bernado P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. J Am Chem Soc 129(17):5656–5664
5. Bertini I, Calderone V, Fragai M, Jaiswal R, Luchinat C, Melikian M, Mylonas E, Svergun DI (2008) Evidence of reciprocal reorientation of the catalytic and hemopexin-like domains of full-length MMP-12. J Am Chem Soc 130(22):7011–7021
6. Blobel J, Bernado P, Svergun DI, Tauler R, Pons M (2009) Low-resolution structures of transient protein-protein complexes using small-angle X-ray scattering. J Am Chem Soc 131(12):4378–4386
7. Chacon P, Moran F, Diaz JF, Pantos E, Andreu JM (1998) Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm. Biophys J 74(6):2760–2775
8. Chacon P, Diaz JF, Moran F, Andreu JM (2000) Reconstruction of protein form with X-ray solution scattering and a genetic algorithm. J Mol Biol 299(5):1289–1302
9. Doniach S (2001) Changes in biomolecular conformation seen by small angle X-ray scattering. Chem Rev 101(6):1763–1778
10. Edwards AM, Arrowsmith CH, Christendat D, Dharamsi A, Friesen JD, Greenblatt JF, Vedadi M (2000) Protein production: feeding the crystallographers and NMR spectroscopists. Nat Struct Biol 7(Suppl):970–972
11. Feigin LA, Svergun DI (1987) Structure analysis by small-angle x-ray and neutron scattering. Plenum Press, New York, pp. xiii, 335
12. Gerstein M, Edwards A, Arrowsmith CH, Montelione GT (2003) Structural genomics: current progress. Science 299(5613):1663
13. Glatter O (1977) A new method for the evaluation of small-angle scattering data. J Appl Crystallogr 10:415–421

14. Glatter O, Kratky O (1982) Small Angle X-ray Scattering. Academic, London, p 515
15. Golub G, Reinsh C (1970) Singular value decomposition and least squares solution. Numerische mathematik 14(5):403–420
16. Guinier A (1939) La diffraction des rayons X aux tres petits angles; application a l'etude de phenomenes ultramicroscopiques. Ann Phys (Paris) 12:161–237
17. Hamiaux C, Perez J, Prange T, Veesler S, Ries-Kautt M, Vachette P (2000) The BPTI decamer observed in acidic pH crystal forms pre-exists as a stable species in solution. J Mol Biol 297(3):697–712
18. Heller WT, Abusamhadneh E, Finley N, Rosevear PR, Trewhella J (2002) The solution structure of a cardiac troponin C-troponin I-troponin T complex shows a somewhat compact troponin C interacting with an extended troponin I-troponin T component. Biochemistry 41(52):15654–15663
19. Heller WT, Finley NL, Dong WJ, Timmins P, Cheung HC, Rosevear PR, Trewhella J (2003) Small-angle neutron scattering with contrast variation reveals spatial relationships between the three subunits in the ternary cardiac troponin complex and the effects of troponin I phosphorylation. Biochemistry 42(25):7790–7800
20. Hura GL, Menon AL, Hammel M, Rambo RP, Poole FL 2nd, Tsutakawa SE, Jenney FE Jr, Classen S, Frankel KA, Hopkins RC, Yang SJ, Scott JW, Dillard BD, Adams MW, Tainer JA (2009) Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). Nat Methods 6(8):606–612
21. King WA, Stone DB, Timmins PA, Narayanan T, von Brasch AA, Mendelson RA, Curmi PM (2005) Solution structure of the chicken skeletal muscle troponin complex via small-angle neutron and X-ray scattering. J Mol Biol 345(4):797–815
22. Koch MH, Vachette P, Svergun DI (2003) Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. Q Rev Biophys 36(2):147–227
23. Konarev PV, Petoukhov MV, Volkov VV, Svergun DI (2006) ATSAS 2.1, a program package for small-angle scattering data analysis. J Appl Crystallogr 39:277–286
24. Levitt M (2007) Growth of novel protein structural data. Proc Natl Acad Sci USA 104(9):3183–3188
25. Moore PB (1980) Small-angle scattering: Information content and error analysis. J Appl Crystallogr 13:168–175
26. Mylonas E, Hascher A, Bernado P, Blackledge M, Mandelkow E, Svergun DI (2008) Domain conformation of tau protein studied by solution small-angle X-ray scattering. Biochemistry 47(39):10345–10353
27. Mylonas E, Svergun DI (2007) Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. J Appl Crystallogr 40:s245–s249
28. Nollmann M, He J, Byron O, Stark WM (2004) Solution structure of the Tn3 resolvase-crossover site synaptic complex. Mol Cell 16(1):127–137
29. Okorokov AL, Sherman MB, Plisson C, Grinkevich V, Sigmundsson K, Selivanova G, Milner J, Orlova EV (2006) The structure of p53 tumour suppressor protein reveals the basis for its functional plasticity. EMBO J 25(21):5191–5200
30. Perez J, Vachette P, Russo D, Desmadril M, Durand D (2001) Heat-induced unfolding of neocarzinostatin, a small all-beta protein investigated by small-angle X-ray scattering. J Mol Biol 308(4):721–743
31. Petoukhov MV, Svergun DI (2005) Global rigid body modelling of macromolecular complexes against small-angle scattering data. Biophys J 89(2):1237–1250
32. Petoukhov MV, Svergun DI (2006) Joint use of small-angle X-ray and neutron scattering to study biological macromolecules in solution. Eur Biophys J 35:567–576
33. Petoukhov MV, Svergun DI (2007) Analysis of X-ray and neutron scattering from biomacromolecular solutions. Curr Opin Struct Biol 17(5):562–571
34. Petoukhov MV, Eady NA, Brown KA, Svergun DI (2002) Addition of missing loops and domains to protein models by x-ray solution scattering. Biophys J 83(6):3113–3125

35. Petoukhov MV, Konarev PV, Kikhney AG, Svergun DI (2007) ATSAS 2.1 – towards automated and web-supported small-angle scattering data analysis. J Appl Crystallogr 40(s1):s223–s228
36. Porod G (1982) General theory. In: Glatter O, Kratky O (eds) Small-angle X-ray scattering. Academic, London, pp 17–51
37. Putnam CD, Hammel M, Hura GL, Tainer JA (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Q Rev Biophys 40(3):191–285
38. Round AR, Franke D, Moritz S, Huchler R, Fritsche M, Malthan D, Klaering R, Svergun DI, Roessle M (2008) Automated sample-changing robot for solution scattering experiments at the EMBL Hamburg SAXS station X33. J Appl Crystallogr 41:913–917
39. Sali A, Glaeser R, Earnest T, Baumeister W (2003) From words to literature in structural proteomics. Nature 422(6928):216–225
40. Svergun DI (1991) Mathematical methods in small-angle scattering data analysis. J Appl Crystallogr 24:485–492
41. Svergun DI (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J Appl Crystallogr 25:495–503
42. Svergun DI (1994) Solution scattering from biopolymers: advanced contrast variation data analysis. Acta Crystallogr A50:391–402
43. Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys J 76(6):2879–2886
44. Svergun DI, Koch MHJ (2003) Small angle scattering studies of biological macromolecules in solution. Rep Prog Phys 66:1735–1782
45. Svergun DI, Barberato C, Koch MHJ (1995) CRYSOL – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J Appl Crystallogr 28:768–773
46. Svergun DI, Petoukhov MV, Koch MHJ (2001) Determination of domain structure of proteins from X-ray solution scattering. Biophys J 80(6):2946–2953
47. Tidow H, Melero R, Mylonas E, Freund SM, Grossmann JG, Carazo JM, Svergun DI, Valle M, Fersht AR (2007) From the Cover: Quaternary structures of tumor suppressor p53 and a specific p53 DNA complex. Proc Natl Acad Sci USA 104(30):12324–12329
48. Toft KN, Vestergaard B, Nielsen SS, Snakenborg D, Jeppesen MG, Jacobsen JK, Arleth L, Kutter JP (2008) High-throughput Small Angle X-ray Scattering from proteins in solution using a microfluidic front-end. Anal Chem 80(10):3648–3654
49. Vestergaard B, Groenning M, Roessle M, Kastrup JS, van de Weert M, Flink JM, Frokjaer S, Gajhede M, Svergun DI (2007) A helical structural nucleus is the primary elongating unit of insulin amyloid fibrils. PLoS Biol 5(5):e134
50. Volkov VV, Svergun DI (2003) Uniqueness of ab initio shape determination in small angle scattering. J Appl Crystallogr 36:860–864
51. Whitten AE, Trewhella J (2009) Small-angle scattering and neutron contrast variation for studying bio-molecular complexes. Methods Mol Biol 544:307–323
52. Xu XF, Reinle WG, Hannemann F, Konarev PV, Svergun DI, Bernhardt R, Ubbink M (2008) Dynamics in a pure encounter complex of two proteins studied by solution scattering and paramagnetic NMR spectroscopy. J Am Chem Soc 130(20):6395–6403

Chapter 16
Structural Dynamics of the Vault Ribonucleoprotein Particle

Arnau Casañas, Jordi Querol, Ignasi Fita, and Núria Verdaguer

Abstract Vaults are ubiquitous, highly conserved, 13 MDa ribonucleoprotein particles, involved in a diversity of cellular processes, including multidrug resistance, transport mechanisms and signal transmission. There are between 10^4 and 10^6 vault particles per mammalian cell and they do not trigger autoimmunity. The vault particle shows a hollow barrel-shaped structure organized in two identical moieties, each consisting of 39 copies of the major vault protein (MVP). Other data indicated that vault halves can dissociate at acidic pH. The high resolution, crystal structure of the of the seven N-terminal domains (R1–R7) of MVP, forming the central vault barrel, together with that of the native vault particle (solved at 8 Å resolution), revealed the interactions governing vault association and suggested a pH-dependent mechanism for a reversible dissociation induced by low pH. Vault particles posses many features making them very promising vehicles for the delivery of therapeutic agents including self-assembly, 100 nm size range, emerging atomic-level structural information, natural presence in humans ensuring biocompability, recombinant production system, existing features for targeting species to the large lumen and a dynamic structure that may be controlled for manipulation of drug release kinetics. All these attributes provide vaults with enormous potential as a drug/gene delivery platform.

Keywords Major vault protein • Nanocontainer • Ribonucleoprotein complex • Vault particle • X-ray structure

A. Casañas • J. Querol • I. Fita • N. Verdaguer (✉)
Instituto de Biología Molecular de Barcelona (CSIC-Parc Cientific de Barcelona),
Baldiri i Reixac 10, Barcelona 08028, Spain
e-mail: nvmcri@ibmb.csic.es

M.A. Carrondo and P. Spadon (eds.), *Macromolecular Crystallography*,
NATO Science for Peace and Security Series A: Chemistry and Biology,
DOI 10.1007/978-94-007-2530-0_16, © Springer Science+Business Media B.V. 2012

16.1 Introduction

Major challenges for treating a variety of diseases are the lack of suitable methods to deliver therapeutic agents (e.g. drugs, proteins, or nucleic acids) to specific host tissues. Many chemotherapeutic agents are small hydrophobic molecules with limited aqueous solubility that requires the use of potentially toxic organic solvents for effective delivery. Encapsulation of these agents into nanoparticles is a promising approach that could increase delivery to specific tissues and reduce systemic side effects by altering the pharmacokinetics of the encapsulated compound [2]. Similar benefits of encapsulation are expected for protein and peptide-based anticancer agents [24].

Vaults are highly conserved, barrel-shaped, ribonucleoprotein particles, found in phylogeny as diverse as mammals, avians, amphibians, echinaderms, kinetoplasts, and amoebas [22]. With an average of 10^4–10^6 vault particles per human cell, they appear to be one of the most abundant particles in the body. However its precise function remains unknown. Their large structure and size, coupled with the potential to encompass hundreds of proteins, have led to the proposal that they could be utilized as natural nanocapsules for drug, nucleic acid, or protein delivery [12]. Vaults do not present some of the major drawbacks found in common nanoscale drug/gene delivery systems such as viral capsids, liposomes, and polymer particles. Although viruses have proven useful as vaccines, their well-documented immunogenicity can produce life-threatening immune responses when used as gene or drug delivery vehicles [15]. During circulation, liposomes commonly exhibit low physical and chemical stability. An ideal drug delivery vehicle should be biocompatible and capable of protecting the encapsulated drug from premature degradation, releasing its contents only upon reaching the target tissues.

Native vaults are formed from 78 copies of an ~100 kDa protein, termed the major vault protein MVP [23] and two vault-associated proteins; vault poly-ADP-ribose polymerase VPARP, 193 kDa [10]and the 290 kDa telomerase associated protein one TEP1 [11] and a small untranslated RNA, vRNA [9]. All of the information for vault particle assembly is inherent in the MVP protein sequence [21] When insect cells are infected with a baculovirus containing an MVP cDNA, large quantities of recombinant vaults are produced; demonstrating that multimerization of this single protein is sufficient to form the exterior shell of the particle. Cryo-electron microscopy (cryoEM) and image reconstruction of recombinant vaults, formed from modified MVPs, revealed that they form empty capsules with an overall structure virtually identical to native vaults. The N-terminal region of MVP forms the particle waist and accounts for the non-covalent interface at the vault midsection [16], whereas the C-terminus builds the cap as well as the cap/barrel junction

The fact that the murine MVP was found to be orthologous to the earlier described human lung resistance-related protein, known to be overexpressed in multiple chemotherapy resistance models immediately associated vaults with intrinsic drug resistance [20]. This particle has also been implicated in the regulation of several cellular processes including transport mechanisms, signal transmission and immune

responses [1]. Another function for vaults was proposed by Herlevsen et al. [8] that found that MVP knockdown disrupted the lysosomal compartment. Interestingly, an independent set of experiments demonstrated vault dissociation at low pH [5, 6]. The former authors proposed that the acidic nature of the lysosomes may serve as an excellent microenviroment with which to trigger vault dissociation. Vault lability was also observed by Poderycki et al. [17] that managed to incorporate vault-associated proteins into preformed MVP-only recombinant vaults proving that they are not rigid, impenetrable boxes, but more a fluctuating dynamic structure presenting substantial flexibility. Their capsular structure and the occasional occurrence of a mass in the inner hollow cavity [13] led to the hypothesis that they might represent transport vehicles. However, both the cellular signals responsible for vault opening and the nature of this cargo have still to be determined.

For vaults to be effective in such applications, efficient means to load the interiors of the capsules must be demonstrated. A passive encapsulation of a semiconducting polymer into Vaults was described confirming that Vaults are dynamic structures that allow facile encapsulation of macromolecules from their environment in the time scale of seconds to minutes. The exact molecular interactions were no described but as the polymer was an anionic polyelectrolyte, a significant number of positively charged residues were predicted to the inside of the Vault. This was confirmed by the solution of the 3D structure of the Vault particle, where a poly basic region was detected on the inside of the Vault Cage [19].

A specific vault-targeting peptide sequence was also identified at the C-terminus of the vault interacting protein VPARP (aa 1563–1724). This sequence was designated, mINT, for the minimal interaction domain. When heterologous proteins were fused to mINT, they were directed to the inside of recombinant vaults, and these packaged proteins retain their native properties [12]. Difference mapping showed that the mINT domain binds to the inside of recombinant vaults at two locations, above and below the waist of the particle. The exterior shell of the vault acted as a protective barrier, as either fluorescent or enzymatic mINT containing fusion proteins were quenched with slower kinetics when packaged inside of vaults. Recent work describes the use of a domain derived from a vault lumen-associated protein as a carrier to target both gold nanoclusters and heterologous His-tagged proteins to specific binding sites on the vault interior wall [7].

Knowing that Recombinant vault particles enter cells via macropinocytosis or phagocytosis but lack demonstrable membrane penetrating activity other Vault modifications have also been developed. For example, in order to improve vault penetration into target cells, the vault-targeting peptide mINT was fused to the membrane lytic domain of adenovirus protein VI (pVI) Not only was the fusion protein incorporated into Vault particles but the membrane lytic activity of the pVI domain was also retained. These findings indicate that vault particles can be modified to enhance cell transfer of selected biomolecules [14].

In the view of all encouraging results, it is not surprising that the first medical applications are appearing. For example, the use of a recombinant Vault nanoparticle was described as a vaccine delivery platform against mucosal infections [4]. The fact that Vault can reversibly separate into two symmetrical halves has also

been the focus of intense research. The possibility of closing and opening vaults in a controllable manner would be an attractive goal.

Very recently, the structure of the rat vault at 3.5 Å resolution was published [23]. It showed that the vault shell is organized in two identical moieties, each consisting of 39 copies of MVP. The MVP monomers are folded into 12 distinct domains: nine repeat domains, a shoulder domain, a cap-helix and a cap ring. Very shortly after we reported the high resolution structures of a recombinant MVP fragment, containing the seven N-terminal domain repeats (R1–R7), in three different crystal forms and of the intact vaults from rat liver determined at 8 Å resolution [19]. The comparison between the structures of R1–R7 and the equivalent region in the reported model [23] showed fundamental discrepancies in the tracing of domains R1 and R2. The quality of the R1–R7 data (2.1-Å resolution) allows unequivocal tracing of domains R1 and R2 that form the rims between the two vault halves. The positioning of the R1–R7 structure into the 8-Å map of the entire vault reveals the interactions stabilizing vault association and provides an explanation for a reversible dissociation induced by low pH.

Here we report the processes of structure determination of the MVP domains R1–R7 and of the complete vault particle. We will also discuss our actual experiments aimed to alter the vault stability in order to finely tune their dissociation increasing the efficiency of drug delivery into the specific cell type.

16.2 Results and Discussion

16.2.1 Solving the Structures of the MVP R1–R7 Modules and of the Complete Vault Particle

The N-terminal fragment of MVP (amino acid residues from Met1 to Asp383) containing the first seven domain repeats R1–R7 was crystallized in three different crystal forms: triclinic P1 (a = 29.39 Å, b = 50.79 Å, c = 76.85 Å, α = 104.32º, β = 92.44º γ = 99.75º, with one R1–R7 molecule per a.u), diffracting to 2.1 Å resolution and two monoclinic crystal forms, $P2_1A$ (2.5 Å; a = 36.5 Å, b = 98.8 Å, c = 140.6 Å, β = 97.2°; two molecules in the a.u) and $P2_1B$ (3.0 Å; a = 58.6 Å, b59.7 Å, c = 68.3 Å β = 95.5°; one molecule in the a.u). The structure was determined by multiwavelength anomalous dispersion of seleniomethionated protein from the monoclinic $P2_1A$ crystals. The P1 and $P2_1B$ structures were solved by molecular replacement, using the coordinates of the $P2_1A$ crystals as a starting model. The seven MVP repeats share a similar structure, each consisting of five antiparallel β-strands connected by loops ([19]; Fig. 16.1a). They are organized in an extended conformation, showing a characteristic bend between repeats R2 and R3. In the P1 space group, the R1–R7 repeats are packed in an almost parallel arrangement, generating a flat wall of a ~18 Å thickness ([19]; Fig. 16.1b) in good agreement with the dimensions estimated by cryo EM [13, 16].

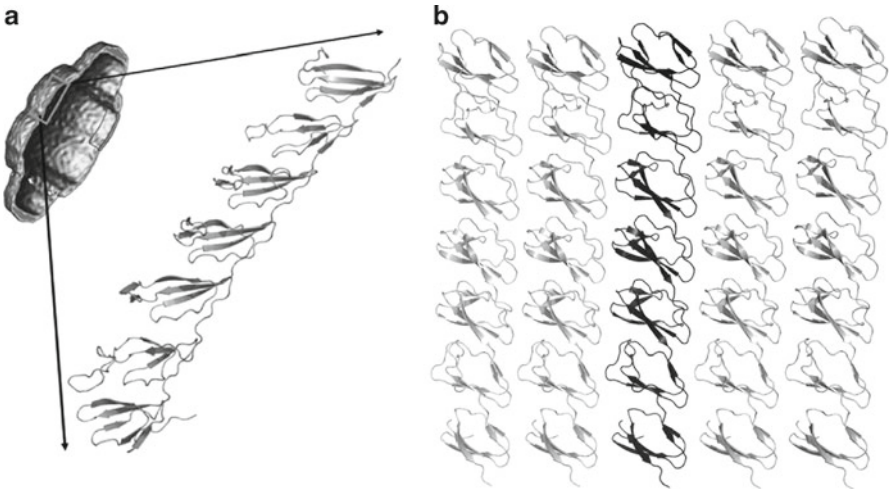


Fig. 16.1 The Structure of the R1–R7 fragment of MVP and, packing contacts in the P1 crystal. (**a**) 39 copies of R1–R7 form the central barrel of each half-vault moiety. The seven MVP repeats share a similar structure, each consisting of five antiparallel β-strands connected by long loops. (**b**) In the P1 space group, the R1–R7 repeats are packed in an almost parallel arrangement, generating a flat wall of a ~18 Å thickness [19]. Five R1–R7 fragments are represented in the picture

3D crystals of native vault particles purified from rat liver were obtained and crystallized as described in [18]. Different crystal forms were obtained and two of them were analysed by X-ray crystallography. The first crystals were monoclinic, space group C2, with unit cell parameters a = 726.2 Å, b = 391.4 Å, c = 607.6 Å and β = 124.1° and diffracted to a maximum of about 10 Å resolution. The second type of crystals, that in a few cases diffracted well beyond 7 Å resolution, were also monoclinic, space group $P2_1$, with unit cell parameters a = 601.1 Å, b = 386.6 Å, c = 627.1 Å and β = 108.6°.

Both, $P2_1$ and C2, unit cells have similar volumes and their cell axis can be related as: $\bar{\mathbf{a}}_C \approx \bar{\mathbf{a}}_P + \mathbf{c}_P$; $\mathbf{b}_C \approx \mathbf{b}_P$ and, $\mathbf{c}_C \approx -\mathbf{a}_P$.

The self-rotation functions, calculated at 10 Å resolution with the program MOLREP, showed for both crystal forms the same, very unique, features (Fig. 16.2): (1) A sharp and strong peak in the a c plane, which remains dominant for essentially all κ values (named henceforth the vaults rotational symmetry peak along the longest vault axis) and (2) A prominent continuous streak, in the κ = 180° section, that is located in the plane perpendicular to the strong peak and, consequently, also to the a c plane. The similarities of these characteristic features indicate closely related packing of the vault particles in both crystal forms, even despite the different crystallographic parameters. For the C2 crystals, a half vault particle can be placed in the asymmetric unit with the two halves of a vault related by a crystallographic two fold, as showed in the structure reported by [23]. For the $P2_1$ crystals, with a strong peak in the native Patterson function at (0.5, 0.5, 0.5), the asymmetric unit

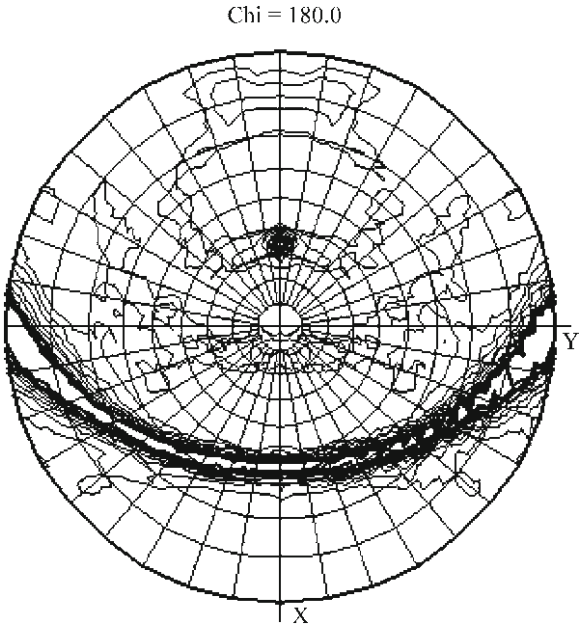


Fig. 16.2 Self-rotation function, calculated from the 8 Å diffraction data of the of the P21 vault crystals. The function was calculated with program MOLREP [3] from the $P2_1$ diffraction data, 180° section

can contain a whole vault particle, with its centre in (0.25, Y, 0.25), that would be related to a second vault particle in the unit cell by the crystal 2_1 symmetry. The translation between the centres of both particles corresponds to the Patterson peak. Therefore, the compact packing of vault particles is almost identical in both crystal forms In fact, in the $P2_1$ cell only small deviation from the C2 packing should be expected because reflexions with h + k + l odd are absent or very weak (pseudo-body centred) till about 15 Å resolution. These observations could explain, at least in part, the easiness of the interconvertibility observed among different crystals forms.

Data from the $P2_1$ crystals were used to solve the structure of the whole vault particle to 8 Å resolution by molecular replacement. Initial models were generated as rings of the R1–R7 structure with different rotational symmetries. The interactions observed in the triclinic crystal packing served as a restraint for building the central barrel of the vault. The bending between domains R2 and R3 could explain the invagination in the central part of the vault particle, with the R1 module located in this central part, as previously suggested by cryo-EM and mutational studies [16]. The positioning of the R1–R7 rings in both the C2 and $P2_1$ vault crystals indicate that only models with rotational symmetry of 39 allow intermolecular interactions between neighbour vault particles without introducing steric problems (Fig. 16.3). Density averaging and solvent flattening with DM [3] were applied for the $P2_1$ crystals, using the independent positioning of the two R1–R7 rings as starting phases and rotational symmetry 39. The final averaged maps at 8 Å resolution showed unambiguous density, fitting the complete MVP protein. In particular, two additional β-domains (R8 and R9) and the long helical domain, for which no

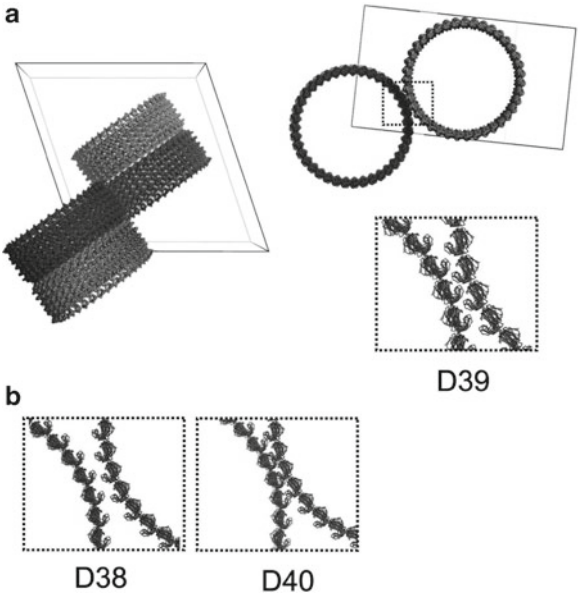


Fig. 16.3 Crystal packing of the vault particles in the C2 unit cell. (**a**) Packing is represented in the ac (*left*) and ab (*right*) planes. The box shows a close-up of the intermolecular contacts established between R domains, considering a rotational symmetry of 39. This symmetry allows optimal interactions between neighbor vault particles. (**b**) Details of the intermolecular contacts, considering rigs of different numbers of copies of the R1R7 fragment. No direct interactions are found when 38 copies are considered (*left*) and models with rotational symmetries of 40 or higher result in steric clashes (*right*)

information was included in the initial model were clearly defined [19], rendering the same overall structure to that reported by [23].

Accurate superimpositions of the individualized domains between the R1–R7 structures with the same region in the intact vault particle, determined by [23], revealed close similarities in repeats R3 to R7. However, critical differences were observed when repeats R1 and R2 were compared [19]. Repeat R1 showed well defined secondary structural elements (strands R1β1–R1β5) in all three crystal forms determined for the R1–R7 fragment (Fig. 16.1a). In the 3.5 Å vault structure, most of the secondary structural elements of R1 were lost and the main chain was out of register from residue Ile10 to Gln88. The largest discrepancies concentrate in the β2–β3 loops of both, R1 and R2 domains [19] and affect most of the R1–R2 and R1–R1 interdomain interactions, including those responsible for the contacts between the two vault halves. The correct positioning of the 2.1 Å structure of R1–R7 into the 8 Å map of the complete particle, reveals important charge complementarity at the interface between the two vault halves (Fig. 16.4). R1–R1 interactions involve two R1 subunits in each half vault moiety (Fig. 16.4a). Among these contacts, Asp39 seems to be the key residue. Its side chain contacts the main chain N atoms of Ala21 and Gln22 and forms a salt bridge with Arg42. The charged

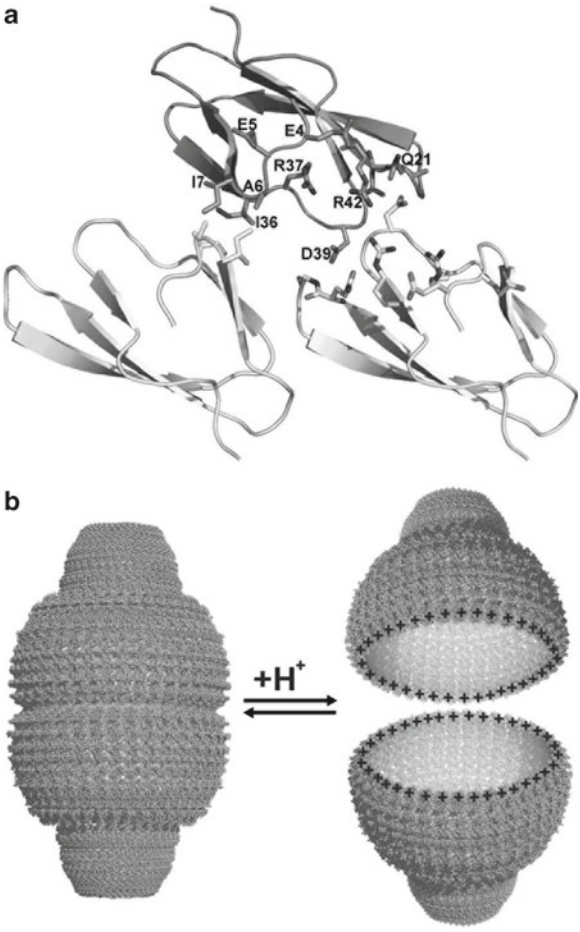


Fig. 16.4 Interdomain interactions at the interface between vault halves. (**a**) Ribbon diagram of the R1–R1 contacts at the half vault interface. The reference R1 domain contacts two consecutive R1 molecules through the molecular two-fold axis. Interacting residues are shown in sticks and labeled. (**b**) Schematic drawing that shows the mechanism of vault opening

amino acids Glu4, Glu5 and Arg37 also appear at the interface. Furthermore, a cluster of hydrophobic residues (Ala6, Ile7 and Ile36), interacting through the two-fold axis of the particle also determine the contact interface (Fig. 16.4a). The interactions observed let us propose a reversible mechanism of dissociation of the vault particle induced by a pH change. At low pH, the 312 acidic residues at the vault interface would become neutral, leaving a highly electropositive charge and inducing the disassembly of the vault particle by charge repulsion (Fig. 16.4b). At higher pH, the aspartate and glutamate residues, would recover their acidic state and re-establish the electrostatic interactions, allowing the re-association between the two vault halves. Subsequently, the hydrophobic interactions would contribute to stabilize the locked conformation of the particle.

16.2.2 The Structure of the Vault Complex as a Starting Point for Future Research

The application of nanotechnology to medicine has emerged as an area of intense interest, particularly the creation of nanosystems for drug delivery. Vault particles posses a large number of features that make them promising vehicles for drug delivery. These include: self-assembly, large size, growing body of high resolution structural information, natural presence in humans ensuring biocompability, recombinant production systems, existing elements for targeting species to the large lumen and a dynamic structure that may be controlled for manipulation of drug release kinetics. These properties provide vaults with enormous potential as nanocontainers for drug/gene transport and delivery. Modified Vaults can already encapsulate specific therapeutic agents and have been directed to specific cell tissues. Our actual work aims to take this molecular engineering a step further by altering the size and length of the Vault particle, and also to mutate the key residues for the association of the two vault halves. The overall objectives would be to design cargo-specific Vaults (for example larger quantities or larger therapeutic agents) and also to modify the stability of Vaults in order to finely tune their dissociation increasing the efficiency of drug delivery into the specific cell type. Once the methodology of engineering modified vaults will be available, the possibilities and potential application are numerous. An exciting option would be, taking into account that the encapsulation of polyanionic compounds has already been described, applications for gene therapy where large DNAs or RNAs carrying genetic information would constitute the Vault cargo.

The extensive structural knowledge that we have obtained from our high resolution crystallographic data and experience with Vault particles put us in a very advantageous position to accomplish these objectives.

References

1. Berger W, Steiner E, Grusch M, Elbling L, Micksche M (2009) Vaults and the major vault protein: novel roles in signal pathway regulation and immunity. Cell Mol Life Sci 66:43–61
2. Brigger I, Dubernet C, Couvreur P (2002) Nanoparticles in cancer therapy and diagnosis. Adv Drug Deliv Rev 54:631–651
3. CCP4 (1994) The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 50:760–763
4. Champion CI, Kickhoefer VA, Liu G, Moniz RJ, Freed AS, Bergmann LL, Vaccari D, Raval-Fernandes S, Chan AM, Rome LH, Kelly KA (2009) A vault nanoparticle vaccine induces protective mucosal immunity. PLoS One 4(4):e5409
5. Esfandiary R, Kickhoefer VA, Rome LH, Joshi SB, Middaugh CR (2009) Structural stability of vault particles. J Pharm Sci 98:1376–1386
6. Goldsmith LE, Yu M, Rome LH, Monbouquette HG (2007) Vault nanocapsule dissociation into halves triggered at low pH. Biochemistry 46:2865–2875
7. Goldsmith LE, Pupols M, Kickhoefer VA, Rome LH, Monbouquette HG (2009) Utilization of a protein “shuttle” to load vault nanocapsules with gold probes and proteins. ACS Nano 3(10):3175–3183

8. Herlevsen M, Oxford G, Owens CR, Conaway M, Theodorescu D (2007) Depletion of major vault protein increases doxorubicin sensitivity and nuclear accumulation and disrupts its sequestration in lysosomes. Mol Cancer Ther 6(6):1804–1813
9. Kickhoefer VA, Searles RP, Kedersha NL, Garber ME, Johnson DL, Rome LH (1993) Vault ribonucleoprotein particles from rat and bullfrog contain a related small RNA that is transcribed by RNA polymerase III. J Biol Chem 268:7868–7873
10. Kickhoefer VA, Siva AC, Kedersha NL, Inman EM, Ruland C, Streuli M, Rome LH (1999) The 193-kD vault protein, VPARP, is a novel poly(ADP-ribose) polymerase. J Cell Biol 146:917–928
11. Kickhoefer VA, Stephen AG, Harrington L, Robinson MO, Rome LH (1999) Vaults and telomerase share a common subunit, TEP1. J Biol Chem 274:32712–32717
12. Kickhoefer VA, Garcia Y, Mikyas Y, Johansson E, Zhou JC, Raval-Fernandes S, Minoofar P, Zink JI, Dunn B, Stewart PL, Rome LH (2005) Engineering of vault nanocapsules with enzymatic and fluorescent properties. Proc Natl Acad Sci USA 102:4348–4352
13. Kong LB, Siva AC, Rome LH, Stewart PL (1999) Structure of the vault, a ubiquitous celular component. Structure 7:371–379
14. Lai CY, Wiethoff CM, Kickhoefer VA, Rome LH, Nemerow GR (2009) Vault nanoparticles containing an adenovirus-derived membrane lytic protein facilitate toxin and gene transfer. ACS Nano 3:691–699
15. Mastrobattista E, van der Aa MA, Hennink WE, Crommelin DJ (2006) Artificial viruses: a nanotechnological approach to gene delivery. Nat Rev Drug Discov 2:115–121
16. Mikyas Y, Makabi M, Raval-Fernandes S, Harrington L, Kickhoefer VA, Rome LH, Stewart PL (2004) Cryoelectron microscopy imaging of recombinant and tissue derived vaults: localization of the MVP N termini and VPARP. J Mol Biol 344:91–105
17. Poderycki MJ, Kickhoefer VA, Kaddis CS, Raval-Fernandes S, Johansson E, Zink JI, Loo JA, Rome LH (2006) The vault exterior shell is a dynamic structure that allows incorporation of vault-associated proteins into its interior. Biochemistry 45:12184–12193
18. Querol-Audi J, Perez-Luque R, Fita I, Lopez-Iglesias C, Caston JR, Carrascosa JL, Verdaguer N (2005) Preliminary analysis of two and three dimensional crystals of vault ribonucleoprotein particles. J Struct Biol 151:111–115
19. Querol-Audí J, Casañas A, Usón I, Luque D, Castón JR, Fita I, Verdaguer N (2009) The mechanism of vault opening from the high resolution structure of the N-terminal repeats of MVP. EMBO J 28(21):3450–3457
20. Scheffer GL, Wijngaard PL, Flens MJ, Izquierdo MA, Slovak ML, Pinedo HM, Meijer CJ, Clevers HC, Scheper RJ (1995) The drug resistance-related protein LRP is the human major vault protein. Nat Med 1:578–582
21. Stephen AG, Raval-Fernandes S, Huynh T, Torres M, Kickhoefer VA, Rome LH (2001) Assembly of vault-like particles in insect cells expressing only the major vault protein. J Biol Chem 276:23217–23220
22. Suprenant KA (2002) Vault ribonucleoprotein particles: sarcophagi, gondolas, or safety deposit boxes? Biochemistry 41:14447–14454
23. Tanaka H, Kato K, Yamashita E, Sumizawa T, Zhou Y, Yao M, Iwasaki K, Yoshimura M, Tsukihara T (2009) The structure of rat liver vault at 3.5 angstrom resolution. Science 323:384–388
24. Torchilin VP, Lukyanov AN (2003) Peptide and protein drug delivery to and into tumors: challenges and solutions. Drug Discov Today 8:259–266

Chapter 17
Structural Dynamics of Picornaviral RdRP Complexes. Implications for the Design of Antivirals

Núria Verdaguer, Cristina Ferrer-Orta, and Esteban Domingo

Abstract Genome replication in picornavirus is catalyzed by a virally encoded RNA dependent RNA polymerase, termed 3D. These viruses also use a small protein primer, named VPg to initiate RNA replication. Polymerase 3D also catalyzes the covalent linkage of UMP to a N-terminal tyrosine on VPg. Seven different crystal structures of foot-and-mouth disease virus (FMDV) 3D catalytic complexes have enhanced our understanding of template and primer recognition, VPg uridylylation and rNTP binding and catalysis. In addition, the biochemical and structural analyses of six different FMDV 3D ribavirin resistant mutants provided evidences of three different mechanisms of resistance to this mutagenic nucleoside analogue. Such structural information is providing new insights into the fidelity of RNA replication, and for the design of antiviral compounds.

Keywords Picornavirus • RNA-dependent RNA polymerase • Replication fidelity • Ribavirin

17.1 Introduction

Picornaviruses are associated with important and diverse diseases, including poliomyelitis, hepatitis A, the common cold, febrile and respiratory illness in humans, and foot-and-mouth disease (FMD), the economically most important animal viral

N. Verdaguer (✉) • C. Ferrer-Orta
Instituto de Biología Molecular de Barcelona CSIC, Parc Científic de Barcelona, Baldiri i Reixac10, Barcelona 08028, Spain
e-mail: nvmcri@ibmb.csic.es

E. Domingo
Centro de Biología Molecular Severo Ochoa (CSIC-UAM), Cantoblanco, Madrid 28049, Spain

M.A. Carrondo and P. Spadon (eds.), *Macromolecular Crystallography*, NATO Science for Peace and Security Series A: Chemistry and Biology,
DOI 10.1007/978-94-007-2530-0_17, © Springer Science+Business Media B.V. 2012

disease. These viruses have also been instrumental in establishing the relevance of high mutation rates for RNA viruses.

RNA-dependent RNA polymerases (RdRPs) synthesize RNA using an RNA template and are normally associated with other virus- or/and host-encoded proteins that modulate RNA polymerization activity and template specificity. An important feature of RNA directed RNA replication is the high error frequency compared to DNA directed replication. This is due, at least in part, to the low fidelity of RdRPs and the absence of error-repair mechanisms in RNA viruses. A new strategy against RNA viruses consists in using mutagenic nucleotides. The objective is to provoke an excessive number of mutations, to deteriorate the viral functions to the point that the virus can not survive. This strategy is called lethal mutagenesis (reviewed in Domingo et al. [8]). One of the mutagens used in research on lethal mutagenesis is a purine analogue termed ribavirin (1-(β-D-ribofuranosyl)-1 *H*-1,2,4-triazole-3-carboxamide), extensively employed in clinical practice. Unfortunately, viral mutants that are resistant to ribavirin have been selected, thus facilitating viral escape from lethal mutagenesis.

As reviewed here, the structural studies with different picornaviruses allowed to interpret the mutagenic activity of nucleoside analogues and helped to elucidate the molecular mechanisms of viral resistance. Among them, FMDV is the picornavirus with the most complete structural information on replication complexes available (reviewed in Ferrer-Orta et al. [14]). Several high resolution structures of the FMDV polymerase (3D) have unveiled key interactions of the enzyme with template-primer RNA, the protein primer VPg, and nucleotide substrates [10, 11, 13]. Of relevance to lethal mutagenesis are the differences in the interactions established by FMDV 3D with standard nucleoside-triphosphates and with the mutagenic analogues 5-fluorouridine-triphosphate (FUTP) and ribavirin-triphosphate (RTP) [13]. Such differences have provided new information on the molecular basis of nucleotide recognition by 3D, and the mechanism of mutagenesis by nucleotide analogues. A number of R-resistant mutants of FMDV have been selected and studied. The structure of their mutant polymerases has unveiled a complex network of interactions within the 3D molecule that affect template binding and RTP recognition [1, 14].

17.2 Structures of Picornavirus RdRPs and RdRP Complexes

17.2.1 Overall Architecture of RdRPs

A large number of three-dimensional structures of viral RdRPs are known, including four different members of the *Picornaviridae* family [11, 12, 16]. All these enzymes share a closed “right hand” conformation, which is accomplished by interconnecting the fingers and thumb domains through the N-terminal portion of the protein and several loops protruding from fingers that largely restrict the inter-

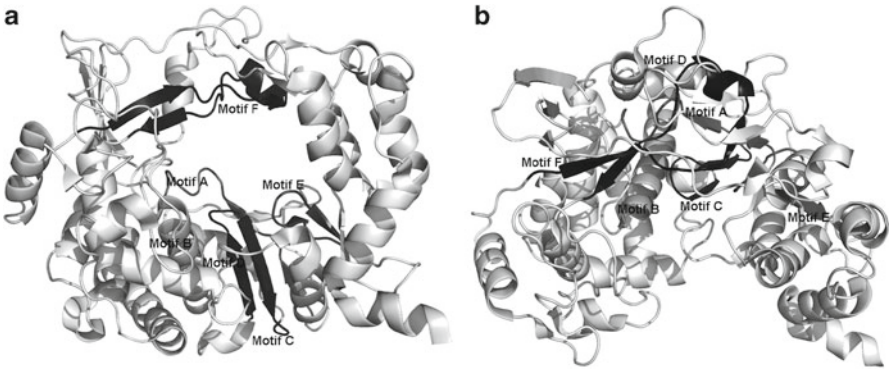


Fig. 17.1 Overall structure of the FMDV RdRP with the conserved structural motifs highlighted in different gray tones and labeled. Two different views "*front*" (**a**) and "*top down*" (**b**) views are represented for the structure

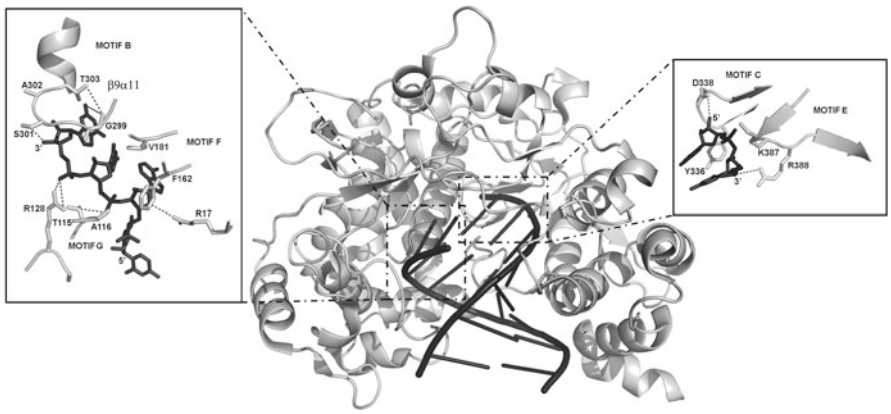


Fig. 17.2 Interactions between the FMDV 3D polymerase and the template-primer RNA (PDB id:1WNE). The polymerase is shown in grey ribbons and RNA molecule is depicted in black sticks. The 3D residues contacting the template-primer RNA are shown in grey sticks and labelled. The 5' overhang region of the template binds the template channel (*left side inset*), where the different residues of the N-terminal region, motif G and motif F drive the ssRNA to the active site cavity. In the active site, the position of the template acceptor base is stabilized by different interactions mediated by residues of motif B and loop β9α11. The primer strand (*right side inset*) interacts with motifs C and E of the palm sub-domain and, with different residues in the thumb sub-domain. Based in Ferrer-Orta et al. [10, 11, 13]

domain mobility (Fig. 17.1). This architecture encircles seven conserved structural motifs, termed A to G, involved in key functions such as the interaction with template RNA, positioning of the 3'-hydroxyl group of the primer RNA, binding of the incoming substrates and ions, as well as catalyzing the nucleotidyltransferase reaction (Figs. 17.1 and 17.2).

17.2.2 The FMDV RdRP in Initiation and Elongation Complexes with Standard Nucleoside Triphosphates and Nucleoside Analogues

Picornavirus replication is initiated by the successive attachment of two UMP molecules to the hydroxyl group of a tyrosine (Tyr3) in the primer protein VPg (a small virus-coded protein which is found covalently attached to the 5'end of the viral RNA). VPg priming is catalyzed by 3D using an internal genomic stem-loop (cis-acting replication element, cre) as template, through a slide-back mechanism [17]. The X-ray structures of two complexes between FMDV 3D and VPg, depicted both, uridylylated and non-uridylylated forms of VPg, and revealed that VPg occupied the central cleft of the polymerase (Fig. 17.3). The hydroxyl group of Tyr3 was positioned as a molecular mimic of the free 3' OH group of a nucleic acid primer at the active site for uridylylation, thereby initiating replication [11]. Several amino acid contacts between 3D and VPg, predicted to be important for initiation of RNA synthesis, were confirmed by site-directed mutagenesis of 3D polymerase and by using chemically synthesized mutant versions of VPg. Twelve out of 16 3D residues that were identified as interacting with VPg, are strictly conserved among different picornaviral polymerases, suggesting that they may play an important role during the critical VPg-uridylylation step [11]. In the 3D-VPg-pU complex, the hydroxyl group of the Tyr-3 side chain was found covalently attached to the α-phosphate moiety of the uridine-monophosphate (UMP) molecule. Two divalent cations, together with the catalytic aspartic acid residues of motifs A and C, participate in the uridylylation reaction, following a similar mechanism to that described for the nucleotidyl transfer reaction in other polymerases [21]. A cluster of positively charged residues of motif F also participate in the uridylylation process, stabilizing Tyr3 and UMP in a proper conformation for the catalytic reaction (Fig. 17.3; [11]).

Insights into rNTP binding, incorporation and RNA elongation have been obtained from the analysis of two 3D elongation complexes [13]. One is the structure of FMDV 3D in complex with a tempate-primer RNA and the substrate ATP, showing the formation of a new base pair, translocation of the RNA product and the release of pyrophosphate (PPi). The other structure is the 3D-RNA template-primer and ATP/UTP complex that illustrates the translocation of the RNA product and the positioning of the new incoming substrate UTP to the active site [13]. Finally, eight additional structures of 3D elongation complexes, using either wild type 3D and mutant 3D polymerases, in complex with RNA and the nucleotide analogs: RTP and FUTP shed new light into the molecular basis of the low fidelity of copy of this enzyme and the mutagenic activities displayed by rNTP analogues during viral replication ([1, 13, 15], Fig. 17.2). In all complexes analysed, the single stranded 5'-end of the template extends across the face of the fingers domain towards the active site cleft, contacting different residues in motif G, which drive the template chain towards the active site cavity. The acceptor base

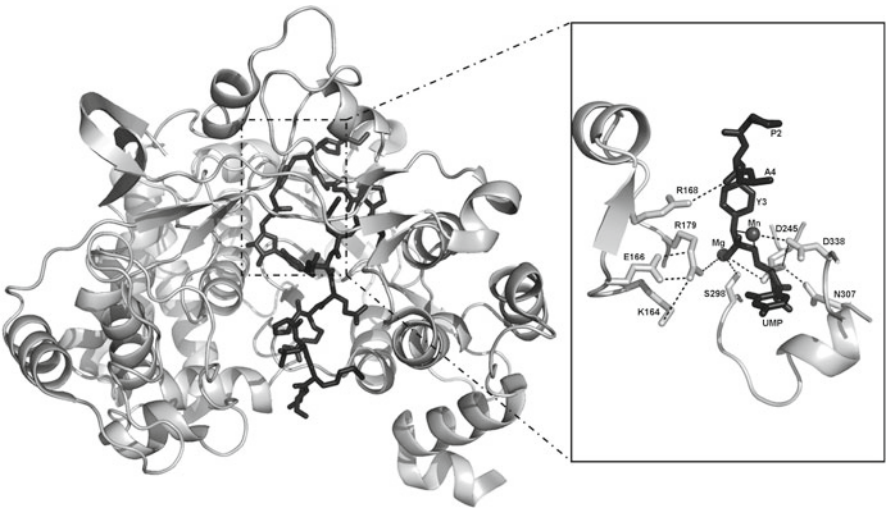


Fig. 17.3 Structure and interactions of the FMDV 3D-VPg-UMP complex (PDB id: 2F8E). The FMDV polymerase is shown in light grey, and the primer protein VPg and the UMP molecule in dark grey. VPg lines the RNA binding cleft of the 3D polymerase, positioning its Tyr3 hydroxyl group as a molecular mimic of the free 3' hydroxyl group of a nucleic acid primer at the active site for nucleotidylylation. The inset in the right side shows a close-up of the interactions established between VPg and different polymerase residues. In the active site, the hydroxyl group of Tyr3 side-chain was found covalently attached to an UMP molecule by a phosphodiester linkage. Two metal ions (*grey spheres*) participate in the uridylylation reaction. Metal 1 bridges the catalytic aspartate, Asp338 of motif C and the O^- of tyrosine side-chain, now covalently bound to phosphate α of UMP. Metal 2 coordinates the carboxyl group of Asp245 of motif A, the O1 oxygen of phosphate α and the hydroxyl group of Ser298 within loop β9-α11, next to motif B. The conserved Tyr336 of motif C and the positively charged residues K164 and R179 of motif F also participate in the uridylylation process. Based in Ferrer-Orta et al. [12, 14]

of the template is located adjacent to the nucleotide binding pocket, being accessible to the incoming nucleotide (Fig. 17.2). In the active site, the 3'-hydroxyl group of the primer strand interacts directly with the catalytic aspartic acid of motif C. Only one metal ion was observed, in contact with the triphosphate moiety of the incoming nucleotide (Fig. 17.4). The structures of the ternary complexes 3D-RNA-rNTP, where the incoming nucleotides were trapped close to or at the polymerase active site, showed the central role of amino acid D245 of motif A, N307 of motif B, and S298, G299 and T303, of a loop preceding motif B (loop β9-α11), in nucleotide recognition and correct positioning of the sugar in the ribose-binding pocket (Fig. 17.4). Structural comparisons of all elongation complexes revealed that the loop β9-α11 is able to adopt different conformations in response to different template and incoming nucleotides, being the most flexible element of the active site of the FMDV polymerase ([13, 15], Fig. 17.4).

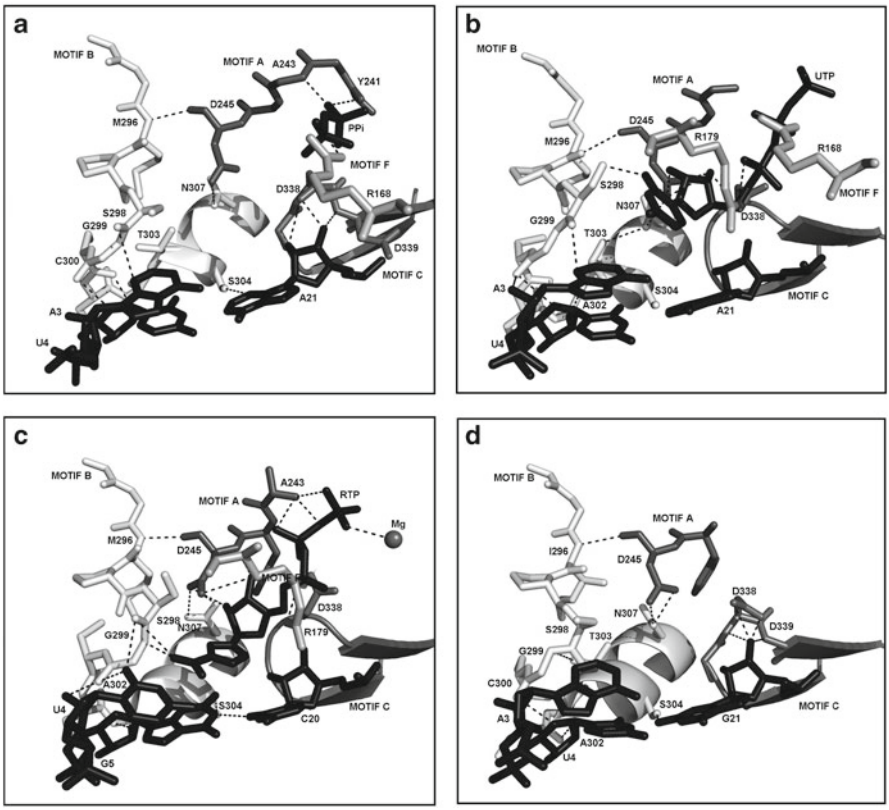


Fig. 17.4 Conformation and interactions in the polymerase active site in different complexes. The polymerase residues in the active site are shown as sticks in grey and explicitly labelled. The RNA template/primers and incoming nucleotides in black with only the first base pairs represented. (**a**) The 3D polymerase active site on AMP incorporation and PPi release (PDB id:2ECO). The newly incorporated nucleotide contacts the catalytic aspartic acid of motif C and motif B. The pyrophosphate product interacts with motifs A and F in the rNTP tunnel. (**b**) The 3D polymerase active site after AMP incorporation and positioning of the new incoming substrate UTP (cyan) close to the active site, as seen in the structure of the FMDV 3D-RNA-ATP/UTP complex (PDB id:2E9Z). (**c**) The structure of the 3D-RNA-RTP complex (PDB id:2E9R). The incoming ribavirin is located at the active site, adjacent to the 3' terminus of the primer, and base paired to the template acceptor base. The ribavirin pseudo-base establishes a number of specific contacts with residues Ser298 and Gly299, within the loop β9-α11 of 3D. This β9-α11 loop changes its conformation to accommodate the nucleoside analog into the cavity. The side chains of residues Asp245 of motif A and Asn307 of motif B have also changed their rotamer conformations to facilitate the interactions with the ribose moiety of the mutagenic nucleotide. The triphosphate moiety is hydrogen bonded to different residues of motifs A and F and interacts with one metal ion. (**d**) Structure and interactions of the mutated polymerase 3D(M296I)-RNA-GTP complex (PDB id:3KOA)**.** Substitution M296I seems to prevent the conformational changes in the loop β9-α11 as well as the side chain rearrangements in residues Asp245 and Asn307 required to interact with ribavirin. However this mutated polymerase is able to misincorporate GMP into the nascent RNA. Based in Ferrer-Orta et al. [13, 15]

17.3 RdRPs as Targets for Antiviral Therapy

RNA dependent RNA polymerases synthesize RNA using an RNA template. This biochemical activity is not present in mammalian cells offering the opportunity to identify very selective inhibitors of this viral enzyme. HCV is a clear example of how much effort has been directed towards developing drugs that inhibit viral replication [6]. The rational search of substrate analogs against the HCV RdRP led to the identification of several nucleoside analogs, among them, ribavirin is a classic antiviral agent used in clinical practice. Ribavirin displays several mechanisms of action, and it is used to treat several important viral diseases, notably chronic hepatitis C virus infections, administered in combination with interferon alpha or some of its derivatives [9]. Remarkably, Crotty et al. [4] documented that ribavirin acts as a mutagen for poliovirus, and its triphosphate form RTP is incorporated by the poliovirus RdRP. Actually there is a growing body of information available, providing evidence that ribavirin can be mutagenic for a number of RNA viruses, contributing to virus extinction through lethal mutagenesis, reviewed in [7]. RTP can be incorporated into the nascent viral RNA, acting either as an adenylate or guanidylate analog, inducing base transitions.

Selection of mutagen-resistant viruses may be a problem for the efficacy of antiviral treatments based on lethal mutagenesis. Mutant PV and FMDV with decreased sensitivity to R have been isolated, and the resistance phenotype maps in the polymerase 3D [18, 20]. Interestingly, the amino acid substitutions that confer R-resistance are different for the two viruses: G64S in PV and M296I in FMDV. G64S is located in the fingers domain, away from the active site of the enzyme, while M296I is located in loop β9-α11, close to the active site of 3D (Fig. 17.4d). Biochemically, the two mutant enzymes show remarkable differences. The purified mutant polymerase 3D(G64S) displays a 5-fold increase in copying fidelity relative to wild type 3D [3, 5]. The increased copying fidelity gave rise to PV populations that were less adaptable than wild type populations to a complex environment, represented by PV-susceptible mice [19]. In contrast to G64S in PV, M296I, the substitution in 3D that decreased the sensitivity of FMDV to Ribavirin, affected only minimally the copying fidelity of the enzyme, despite M296 being located close to the active site of 3D. Biochemical studies showed that recombinant FMDV 3D polymerase with substitution M296I incorporated ribavirin-monophosphate less efficiently than wild type polymerase [20]. Kinetic analyses for single nucleotide addition catalyzed by wild type and mutant (M296I) FMDV 3D polymerases, confirmed that the mutant 3D incorporated ribavirin less efficiently than wild type polymerase. The study revealed that the mutant enzyme missincorporated guanosine-monophosphate more efficiently than wild type FMDV 3D, resulting in a low-fidelity variant with a specific defect for ribavirin utilization [2]. Therefore, restriction of ribavirin incorporation by 3D with M296I was not the result of a general increase of template-copying fidelity.

The structure of the elongation complex wild 3D-RNA-RTP showed the RTP acting as an adenylate analog paired with the uridine in the template chain. Its position in the polymerase active site was further stabilized by interactions involving the ribavirin

pseudo-base and the loop β9-α11 (Fig. 17.4d). The residues from the loop that interact with RTP were moved more than 2 Å from their position in the unbound form. This displacement favored the accommodation of the pseudo-base, and the contacts with the template acceptor nucleotide [13]. The crystal structure of the M296I 3D polymerase determined very recently suggests that the M296I mutation confers resistance to incorporation of ribavirin due to a steric conflict produced in the polymerase active site [15]. The combination of the kinetic and structural data led the conclusion that the FMDV 3D active is able to acquire different conformations in response to correct and incorrect nucleotide incorporation and that the flexibility of the loop β9-α11 plays a central role in these conformational changes. The search for ligands with the capacity to modulate the conformation of this loop (or adjacent residues) to decrease the copying fidelity of the enzyme would be important for lethal mutagenesis-based antiviral approaches.

The difference between PV 3D(G64S) and FMDV 3D(M296I) ribavirin resistant mutants suggests that there are at least two possible mechanisms of R-resistance in picornaviruses. One mechanism is based on a general increase of copying fidelity that includes discrimination against RTP as substrate, and the other is based on the partial exclusion of RTP (and perhaps other specific nucleotide analogues) without a general increase in copying fidelity. Very recently a new mechanism of resistance to ribavirin has been described in FMDV [1]. Populations replicated in the presence of high concentrations of ribavirin (ranging from 800 μM to 5,000 μM), accumulated, in addition to M296I, two new amino acid substitutions in 3D in a step-wise fashion (P44S and P169S). The main biological effect of these substitutions is to attenuate the consequences of the mutagenic activity of ribavirin. The mutated polymerase is able to modulate the types of ribavirin-induced misincorporations during RNA synthesis by an alteration of the pairing preference of ribavirin opposite C and U, avoiding the production of an excess of some types of mutations. Thus, this new mechanism of resistance to ribavirin is based not as much in limiting the number of mutations in the virus genetic material but in ensuring an equilibrium among different types of mutations that favors viral survival [1].

Studies of polymerization activity by the purified polymerases suggest that a single amino acid substitution in a loop of the fingers domain (P44S) is the alteration chiefly responsible of the altered mutational pattern. The crystal structures of the substituted polymerases in complex with RNA show a conformational change in the template entry channel of the polymerase that may affect the binding of the ssRNA template to 3D, mainly at the base of the template which is immediately downstream of the position that receives the incoming nucleotide ([1], Fig. 17.5). Alteration of the position of the template RNA at the active site of the enzyme may affect nucleotide recognition and modify the transition mutation pattern in the presence of ribavirin.

superimposed in grey sticks. Comparison of the three-dimensional structure of wild type and mutant polymerases suggests that the amino acid substitutions alter the position of the template RNA in the entry channel of the enzyme thereby affecting nucleotide recognition. Residues of the loop β2-α2, containing S44, the amino acids contacting with this loop and, those contacting with the template RNA are also shown as grey sticks. (**b**) Close-up of the interactions involving the amino acids, containing the template nucleotides A3 and A4 that are changed its conformation in the mutated structure (**c**) Interactions involving the template nucleotides A3 and A4 in the wild type structure

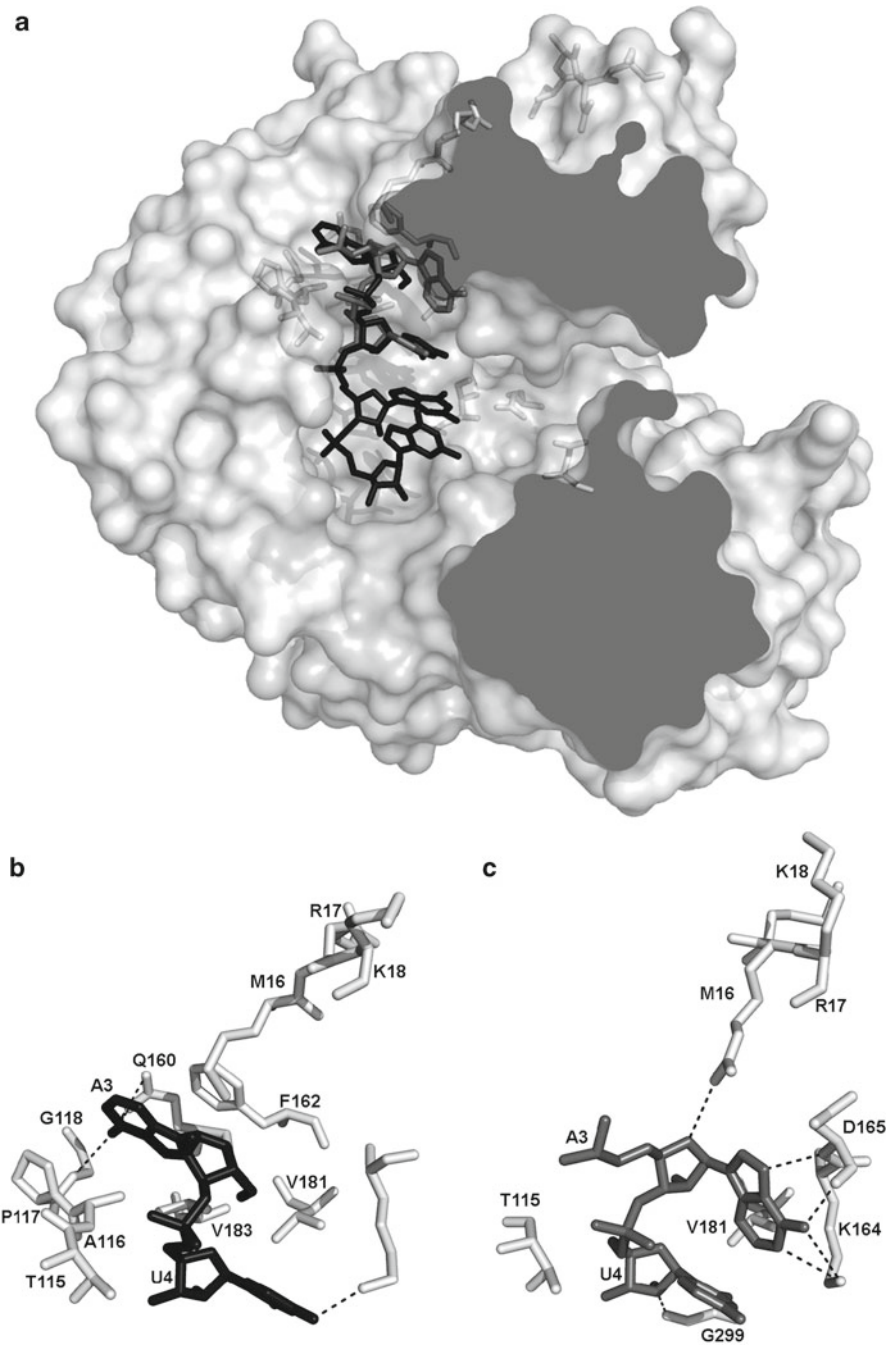


Fig. 17.5 Structure and interactions in the template channel of the mutated polymerase 3D (P44S, P169S, M296I) in complex with a template-primer RNA. (**a**) The polymerase is shown with its molecular surface in gray and the template strand of the RNA depicted in black sticks. The position of the bases A3 and A4 of the RNA template, determined in the wild type 3D-RNA structure, are

References

1. Agudo R, Ferrer-Orta C, Arias A, de la Higuera I, Perales C, Pérez-Luque R, Verdaguer N, Domingo E (2010) A multi-step process of viral adaptation to a mutagenic nucleoside analogue by modulation of transition types leads to extinction-escape. PLoS Pathog. 2010 Aug 26; 6(8) pii: e1001072
2. Arias A, Arnold JJ, Sierra M, Smidansky ED, Domingo E, Cameron CE (2008) Determinants of RNA-dependent RNA polymerase (in)fidelity revealed by kinetic analysis of the polymerase encoded by a foot-and-mouth disease virus mutant with reduced sensitivity to ribavirin. J Virol 82:12346–12355
3. Arnold JJ, Cameron CE (2000) Poliovirus RNA-dependent RNA polymerase (3D(pol)). Assembly of stable, elongation-competent complexes by using a symmetrical primer-template substrate (sym/sub). J Biol Chem 275:5329–5336
4. Crotty S, Maag D, Arnold JJ, Zhong W, Lau JY, Hong Z, Andino R, Cameron CE (2000) The broad-spectrum antiviral ribonucleoside ribavirin is an RNA virus mutagen. Nat Med 6:1375–1379
5. Crotty S, Cameron CE, Andino R (2001) RNA virus error catastrophe: direct molecular test by using ribavirin. Proc Natl Acad Sci USA 98:6895–6900
6. De Francesco R, Migliaccio G (2005) Challenges and successes in developing new therapies for hepatitis C. Nature 436:953–960
7. Domingo E Ed (2005) Virus entry into error catastrophe as a new antiviral strategy. Virus Res 107:115–228
8. Domingo E, Parrish C, Holland JJE (2008) Origin and evolution of viruses, 2nd edn. Elsevier, Oxford
9. Felt JJ, Hoofnagle JH (2005) Mechanism of action of interferon and ribavirin in treatment of hepatitis C. Nature 436:967–972
10. Ferrer-Orta C, Arias A, Perez-Luque R, Escarmis C, Domingo E, Verdaguer N (2004) Structure of foot-and-mouth disease virus RNA-dependent RNA polymerase and its complex with a template-primer RNA. J Biol Chem 279:47212–47221
11. Ferrer-Orta C, Arias A, Agudo R, Perez-Luque R, Escarmis C, Domingo E, Verdaguer N (2006) The structure of a protein primer-polymerase complex in the initiation of genome replication. EMBO J 25:880–888
12. Ferrer-Orta C, Arias A, Escarmis C, Verdaguer N (2006) A comparison of viral RNA-dependent RNA polymerases. Curr Opin Struct Biol 16:27–34
13. Ferrer-Orta C, Arias A, Perez-Luque R, Escarmis C, Domingo E, Verdaguer N (2007) Sequential structures provide insights into the fidelity of RNA replication. Proc Natl Acad Sci USA 104:9463–9468
14. Ferrer-Orta C, Agudo R, Domingo E, Verdaguer N (2009) Structural insights into replication initiation and elongation processes by the FMDV RNA-dependent RNA polymerase. Curr Opin Struct Biol 19:752–758
15. Ferrer-Orta C, Sierra M, Agudo R, de la Higuera I, Arias A et al (2010) Structure of foot-and-mouth disease virus mutant polymerases with reduced sensitivity to ribavirin. J Virol 84:6188–6199
16. Ng KK, Arnold JJ, Cameron CE (2008) Structure-function relationships among RNA-dependent RNA polymerases. Curr Top Microbiol Immunol 320:137–156
17. Paul AV, Yin J, Mugavero J, Rieder E, Liu Y, Wimmer E (2003) A "slide-back" mechanism for the initiation of protein-primed RNA synthesis by the RNA polymerase of poliovirus. J Biol Chem 278:43951–43960
18. Pfeiffer JK, Kirkegaard K (2003) A single mutation in poliovirus RNA-dependent RNA polymerase confers resistance to mutagenic nucleotide analogs via increased fidelity. Proc Natl Acad Sci USA 100:7289–7294

19. Pfeiffer JK, Kirkegaard K (2005) Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. PLoS Pathog 1:e11
20. Sierra M, Airaksinen A, González-López C, Agudo R, Arias A, Domingo E (2007) Foot-and-mouth disease virus mutant with decreased sensitivity to ribavirin: implications for error catastrophe. J Virol 81:2012–2024
21. Steitz TA, Steitz JA (1993) A general two-metal-ion mechanism for catalytic RNA. Proc Natl Acad Sci USA 90:6498–6502

Chapter 18
Ribosomes: Ribozymes that Survived Evolution Pressures but Is Paralyzed by Tiny Antibiotics

Ada Yonath

Abstract An impressive number of crystal structures of ribosomes, the universal cellular machines that translate the genetic code into proteins, emerged during the last decade. The determination of ribosome high resolution structure, which was widely considered formidable, led to novel insights into the ribosomal function, namely, fidelity, catalytic mechanism, and polymerize activities. They also led to suggestions concerning its origin and shed light on the action, selectivity and synergism of ribosomal antibiotics; illuminated mechanisms acquiring bacterial resistance and provided structural information for drug improvement and design. These studies required the pioneering and implementation of advanced technologies, which directly influenced the remarkable increase of the number of structures deposited in the Protein Data Bank.

18.1 Introduction

The translation process requires a complex apparatus composed of many components, among which the ribosome is the key player, as it is actively involved in the translation process. Ribosomes are universal ribozymes performing two main tasks: decoding the genetic information and polymerizing amino acids, while providing the framework for the proper positioning of all other participants, including mRNA, its substrates (tRNAs) and initiation, elongation, release and recycling factors that ensure that protein synthesis occurs progressively and with high specificity. They operate is each living cell continuously since the constant programmed cell death, which implies constant proteins degradation, requires simultaneous production of proteins. Hundreds of thousands of ribosomes are present in typical mammalian

A. Yonath (✉)
Department of Structural Biology, Weizmann Institute of Science, 76100 Rehovot, Israel
e-mail: ada.yonath@weizmann.ac.il

M.A. Carrondo and P. Spadon (eds.), *Macromolecular Crystallography*,
NATO Science for Peace and Security Series A: Chemistry and Biology,
DOI 10.1007/978-94-007-2530-0_18, © Springer Science+Business Media B.V. 2012

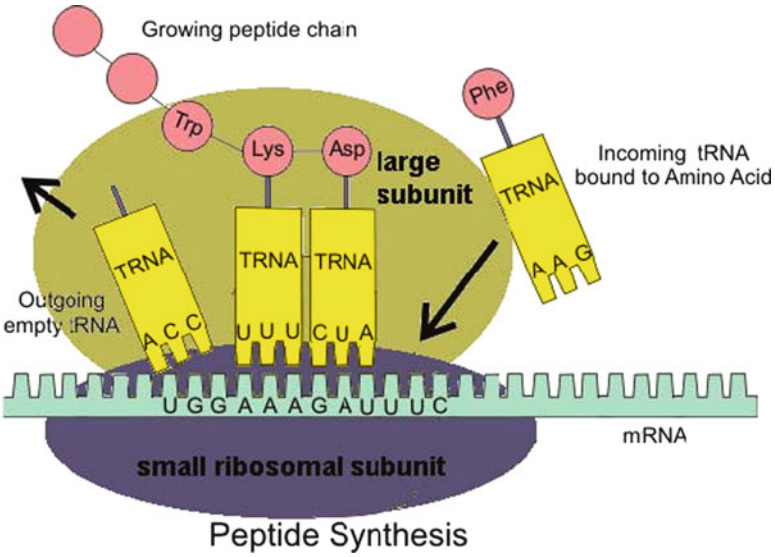


Fig. 18.1 Schematic view of the translation process

cells. Fast replicating cells, e.g. liver cells, may contain a few millions ribosomes. Even bacterial cells may contain to 100,000 ribosomes during their log period. mRNA chains, produced by the transcription of the segments of the DNA that should be translated, carry the genetic information to the ribosomes, and tRNA molecules bring the cognate amino acids to the ribosome (Fig. 18.1). The tRNA molecules from all living cells are built of double helical L-shape molecules containing an anticodon loop that matches its three-nucleotide codes on the mRNA on one of their edges, ~70 Å away, their 3′ends are single strands with the universal sequence CCA to which the cognate amino acid is bound by an ester bond. For increasing efficiency, a large number of ribosomes act simultaneously as polymerases synthesizing proteins by one-at-a-time addition of amino acids to a growing peptide chain, while translocating along the mRNA template, producing proteins on a continuous basis in an incredible speed (5–15 new peptide bonds per second, in eukaryotes and prokaryotes, respectively).

The ribosomes are giant assemblies composed of many different proteins (r-proteins) and long ribosomal RNA (rRNA) chains. The ratio of rRNA to r-proteins (~2:1) is maintained throughout evolution, except in mitochondrial ribosome (mitoribosome) in which almost half of the bacterial rRNA is replaced by r-proteins. In all organisms ribosomes are built of two subunits, which associate to form functionally active ribosomes. In prokaryotes, the small subunit, denoted as 30S, contains an RNA chain (16S) of ~1,500 nucleotides and ~20 different proteins. The large subunit (50S in prokaryotes) has two RNA chains (23S and 5S RNA) of about 3,000 nucleotides in total, and different <31 proteins. The available three dimensional structures of the bacterial ribosome and their subunits show that in each of the two subunits the ribosomal proteins are entangled within the complex

rRNA conformation, thus maintaining a striking dynamic architecture that is ingeniously designed for their functions: precise decoding; substrate mediated peptide-bond formation and efficient polymerase activity. The structural bases for ribosomal functions, as obtained by high resolution crystallographic studies are summarized in several recommended recent reviews [34, 46, 58, 73]. Further insights obtained from the combination of the crystallographic results with those emerging from single-molecule techniques (cryogenic electron microscopic and fluorescence resonance energy transfer) are outlined in [23]. Selected topics of ribosome function are discussed below. As so far high resolution structures are available only for prokaryotic ribosomes, the discussion is confined to these ribosomes and to insights evolved from their structures.

18.2 Snapshots Along the Birth of the Nascent Chains

While the elongation of the nascent chain proceeds, the two subunits perform cooperatively. The tRNA molecules are the non-ribosomal entities combine the two subunits, as each of their three binding sites, A-(aminoacyl), P-(peptidyl), and (exit), (Fig. 18.2) resides on both subunits. Their anticodon loops interact with the mRNA on the small subunit, and their acceptor stems with the aminoacylated or peptidylated 3'ends are located on the large subunit.

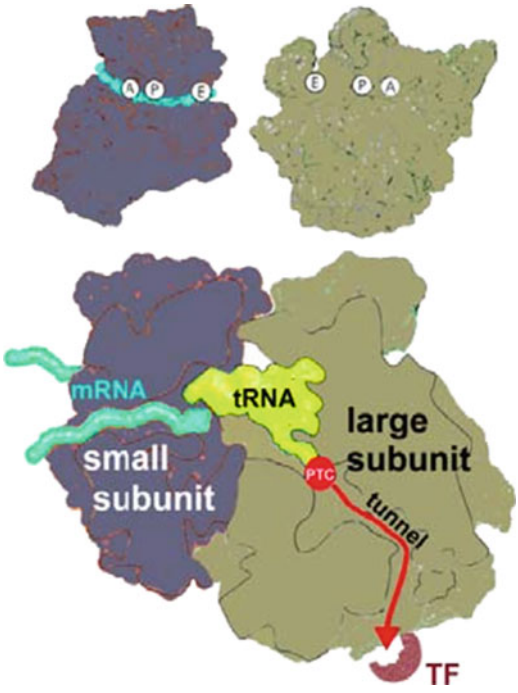


Fig. 18.2 (*Left*) The two subunits. The approximate positions of the mRNA and three tRNA sites are marked. (*Right*) A slice through the center of the translating ribosome showing the P-site tRNA (*blue*), the nascent chain path and direction are shown as a *red arrow*, and the first chaperone encountering the emerging nascent chain (*TF* trigger factor) is represented by a *half circle*

The small subunit provides the path along which the mRNA progresses, the decoding center and the mechanism controlling translation fidelity [46, 51]. Translation initiation is the rate-limiting step of the entire process. It starts by the correct selection and placement of the mRNA reading frame, with the help of the initiation factors, and then proceeds through a tightly regulated steps. Within the initiation complex the initiation codon is decoded in the P-site of the ribosomal subunit, and involves GTP-binding [57].

The large subunit contains the site for the main ribosomal catalytic function, namely polymerization of the amino acids and provides the dynamic protein exit tunnel [13]. The structure of its larger subunit revealed that the ribosome is a ribozyme with RNA at the core of its enzymatic activity. Simultaneously with the advancement of the mRNA along the path in the small subunit, peptide bonds are being formed in the large subunit. This inherently dynamic process requires small and large-scale motions of the ribosomal substrates (e.g., the intersubunit rotational movements during tRNA-mRNA translocation), coupled to conformational rearrangements of its components, facilitating the translocation of the tRNA 3′end from A- to P-site, the detachment of the P-site tRNA from the growing polypeptide chain, the passage of the deacylated tRNA molecule to the E-site and its subsequent release. The nascent proteins progress along a dynamic tunnel and emerge from the large subunit (Fig. 18.2) into a shelter formed by ribosome-bound trigger-factor, acting as a chaperone preventing aggregation and misfolding [8, 53].

The current consensus view is consistent with ribosomal positional catalysis assisted by its P-site tRNA substrate (e.g. [9, 11]) and not by acid/base mechanism [43]. All known structures indicate that the ribosomes provide the suitable stereochemistry for peptide bond formation, the guided path for the A- to P- site translocation and the appropriate geometrical means for substrate mediated catalysis. In all of the so far determined structures the ribosomal catalytic site, called the peptidyl transferase center (PTC), is situated within a highly conserved symmetrical region (Fig. 18.3) that connects all ribosomal functional centers involved in amino-acid polymerization, namely the tRNA entrance/exit dynamic stalks, the PTC, the nascent protein exit tunnel, and the bridge connecting the PTC cavity with the vicinity of the decoding center in the small subunit. Hence, it can serve as the central feature for signaling between all the functional regions involved in protein biosynthesis, that are located remotely from each other (up to 200 Å away), but must “talk” to each other during elongation [1, 2, 9]. As the symmetry relates the backbone fold and nucleotides orientations, but not nucleotide sequence, it emphasizes the superiority of functional requirement over sequence conservation.

The linkage between the elaborate architecture of the symmetrical region and the position of the A-site tRNA, as observed crystallographically [9] indicates that the translocation of the tRNA 3′end is performed by a combination of two independent, albeit synchronized motions: a sideways shift, performed as a part of the overall mRNA/tRNA translocation, and a rotatory motion of the A-tRNA 3’end along a path confined and navigated by the PTC walls, of which all nucleotides have been classified as essential by a comprehensive genetic selection analysis

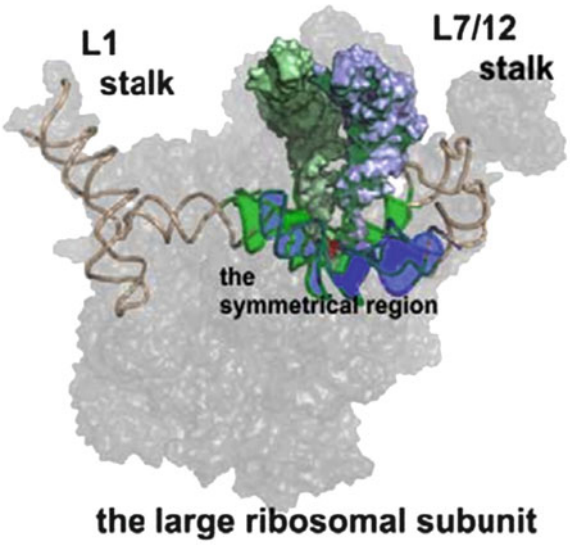


Fig. 18.3 The symmetrical region within the large ribosomal subunit (colored in *blue* and *green*). Its extensions are shown in *gold*. The A-site tRNA is shown in *metal blue* and the P-site tRNA in *light green*

[49]. This motion enables the formation of all interactions that are prerequisite for substrate positioning [33, 56, 63], for mediating acceleration [64], and for the formation of the transition state (TS) of this reaction [24]. This stunning architecture allows for the PTC remarkable ability to rearrange itself upon substrate binding, explaining the pace difference between the formation of single peptide bond by minimal substrates and possessive amino acid polymerization and verifying the finding that the peptidyl transfer reaction is modulated by conformational changes at the active site [9, 11, 12, 54, 67].

18.3 The Proto Ribosome Concept

The high level of conservation suggests that the modern ribosome evolved from a simpler entity that can be described as a pro-ribosome, by gene fusion or gene duplication. In particular, the preservation of the three-dimensional structure of the two halves of the ribosomal frame regardless of the sequence demonstrates the rigorous requirements of accurate substrate positioning in stereochemistry supporting peptide bond formation. This, as well as the universality of the symmetrical region led to the assumption that the ancient ribosome contained a pocket confined by two self folded RNA chains, which associated to form a pocket like dimer (Fig. 18.4).

As RNA chains can act as gene-like molecules coding for their own reproduction, it is conceivable that the surviving pockets became the templates for the ancient ribosomes. In later stage these primitive RNA genes underwent initial optimization

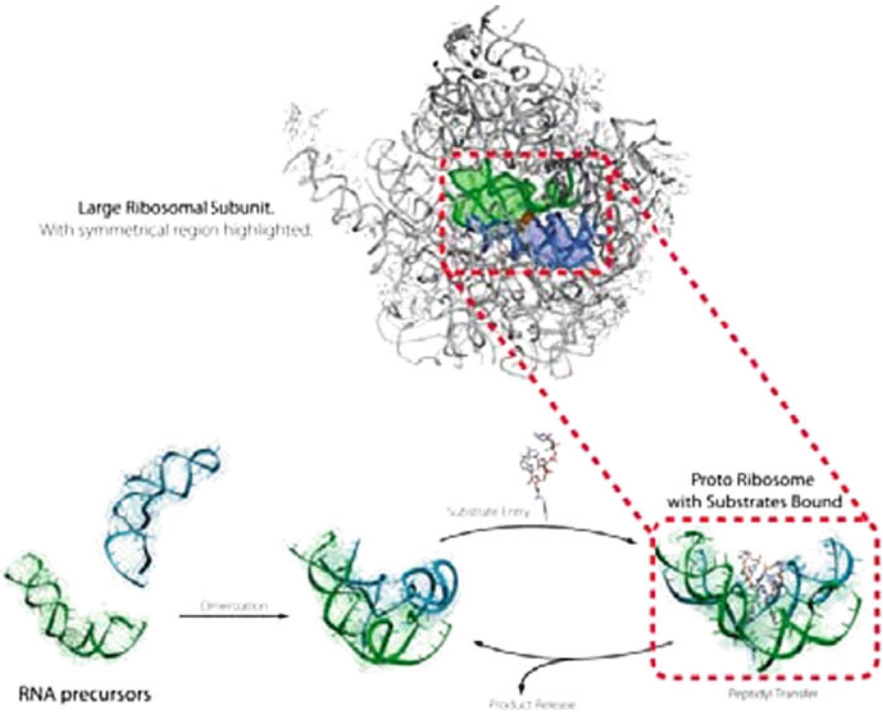


Fig. 18.4 Shows how the proto ribosome could have evolved. The symmetrical region is highlighted within the contemporary ribosome. A view showing its pocket-like nature is shown on *bottom right*

to produce a more defined, relatively stable pocket, and when a clear distinction was made between the amino acid and the growing peptidyl sites, each of the two halves was further optimized for its task so that their sequences evolved differently. In parallel, the substrates of the ancient ribosomes, which were initially activated amino acids (presumably by binding to single or oligo nucleotides), evolved to allow accurate binding. Later, for increasing specificity, these short RNA segments were extended to larger structures by their fusion with RNA stable features, to form the ancient tRNA. Later, RNA chains capable of storing, selecting and transferring instructions for producing useful proteins became available. Subsequently, the decoding process was combined with peptide bond formation. Then single molecules evolved, capable of not only carrying the amino acids while bound to them, but also translating the genomic instructions, by adding a feature similar to the modern anticodon arm to the ancient tRNA structure [10, 21]. Importantly, the notion that the ribosome evolved around an ancient core is also supported by computational and biochemical studies [15, 31].

In short: analysis of substrate binding modes to unbound ribosomal subunits and to functionally active ribosomes illuminated the significance of the PTC mobility and supported the hypothesis that the ancient ribosome could have

evolved from an RNA molecular machine that was functionally active in the RNA world era, which produced single peptides bonds and non-coded chains. Genetic control of the reaction seems to evolve after polypeptides capable of enzymatic function were created, and a stable RNA primitive carrier fold was converted into tRNA molecules.

18.4 Structures for Improving Antibiotics

The intensive research on ribosomes has practical aspects; one of them has clinical relevance since many antibiotics target the ribosome. The increasing incidence of antibiotic resistance and toxicity creates serious problems in modern medicine; combating resistance to antibiotics has been a major concern in recent years. Useful antibiotics that target ribosomes inhibit cell growth by selectively paralyzing the ribosome's activity in pathogens (always eubacteria) and not the eukaryotes. They act by diverse mechanism, all based on a common strategy: coinciding with functionally critical centers of the ribosome Examples are causing miscoding, minimizing essential mobility, interfering with substrate binding at the decoding center and at the PTC, or blocking the protein exit tunnel (Suggested reviews and recent work are: [3–5, 16–20, 32, 37–39, 45, 53, 55, 59, 61, 68, 69]).

By its nature, X-ray crystallography should be the choice method for investigating ribosome-antibiotics interactions. However, since X-ray crystallography requires diffracting crystals, and since so far no ribosomes from pathogenic bacteria could be crystallized, currently the crystallographic studies are confined to the currently available crystals of suitable pathogen models. Currently available are high-resolution structures of complexes of antibiotics with ribosomal particles from the eubacteria *E. coli, Thermus thermophilus* and *Deinococcus radiodurans*, all suitable to serve as a pathogen model. Also available are complexes obtained from antibiotics bound to ribosomes from the Dead Sea archaeon *Haloarcula marismortui* that resembles eukaryotes in respect to antibiotics binding site, hence requiring enormously high antibiotics concentrations for obtaining these complexes. Comparisons between the two types of complexes proved indispensable for increasing our understanding on antibiotics action (Fig. 18.5).

A major issue concerning the clinical usefulness of ribosomal antibiotics is their selectivity, namely their capabilities in the discrimination between the ribosomes of the eubacterial pathogens and those of eukaryotes. Although prokaryotic and eukaryotic ribosomes differ in size (~2.4 and 4 Mega Dalton, respectively), their functional regions, which are the targets for the antibiotics, are highly conserved. Therefore the imperative distinction between eubacterial pathogens and mammals, the key for antibiotics usefulness, is achieved generally, albeit not exclusively, by subtle structural difference within the antibiotics binding pockets of the prokaryotic and eukaryotic ribosomes. In fact, even among the pathogens, there are examples for species selectivity that determines the susceptibility and the fitness cost of the ketolides (e.g. [44]). Selectivity (and resistance) can also be obtained by exploiting induced fit mechanisms based on network of remote interactions by utilizing

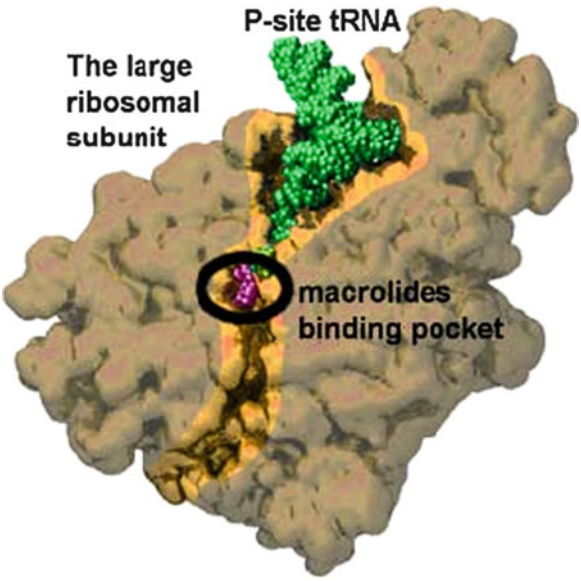


Fig. 18.5 Shows a section through the large ribosomal subunit at the level of the protein exit tunnel, together with P-site tRNA. The location of the macrolide binding pocket is circled

nucleotides that are less conserved, as they do not directly involved in the ribosome functions [19, 20]. Another intriguing issue relates to the contributions of two ribosomal proteins, namely L4 and L22. These proteins line a small part of the exit tunnel at its constriction, and do not interact directly with most of the members of the macrolides family, yet their mutations acquire resistance to them [14, 22, 42, 72], presumably by perturbing the rRNA structure at the tunnel walls [26, 35].

Current attempts to overcome antibiotics resistance and increase their selectivity are being made (e.g. [18, 66]).These include developments of synergetic antibiotics, such as the recent potent antibiotic drug, synercid [29, 69] and reviving "forgotten" antibiotics families, such as the lankacidins [5]. Other strategies are based on insertions of additional moieties that should bind to the ribosome and compensate for the lost interactions in the resistant strains. In parallel, comprehending the factors allowing for selectivity should provide powerful tools to understand many of the mechanisms exploited for acquiring resistance. Thus, the lessons learned from ribosome crystallography for combating resistance of antibiotics targeting the ribosome paved new paths for antibiotics improvement.

18.5 Historical Comments

Owing to the huge size and the complexity of the ribosome, it was widely assumed that ribosomes cannot be crystallized. Twenty years passed from the first indications for potential high resolution by examining the initial microcrystals that diffracted to

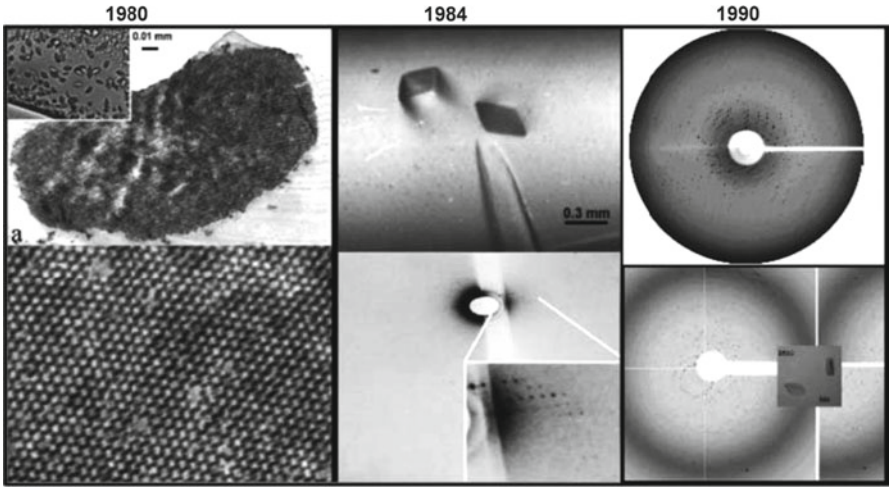


Fig. 18.6 Shows the progress in crystallization. On the *left*, the microcrystals are shown in an insert (*top left*), and a positively stains section of these micro crystals as seen by EM

relatively high resolution, namely 3.5 Å [71] to the first 3D structures. The shift from poorly diffracting microcrystals to high-resolution structures was achieved gradually, based on the assumptions that the higher the conformational homogeneity the better the crystals, and that the preferred conformation is that of functionally active ribosomes. Assuming that the ribosomes of bacteria that grow under robust conditions are less sensitive to external conditions, we focused on such sources, and, indeed, the first three dimensional microcrystals were obtained (Fig. 18.6) from the large ribosomal subunits from *Bacillus stearothermophilus* [71], a source considered to be an extremophile at the beginning of the eighties. Extensive systematic explorations for suitable bacterial sources indicated that the key for obtaining crystals suitable for crystallographic studies is to use ribosomes from relatively robust bacteria, such as *H. marismortui, T. thermophilus* and *D. radiodurans* [25]. A parallel strategy is to crystallize complexes of ribosomes with substrates, inhibitors and/or factors that can trap them at preferred orientations. Among such complexes are the initial crystals of the whole ribosome from *T. thermophilus* with mRNA and tRNA molecules [27]. Efforts aimed at crystals improvement included a thorough examination of the influence of the relative concentrations of mono- and di-valent ions [62] and constant refinements of bacterial growth pathways [6]. Remarkably, flexible functional regions could be traced in maps obtained from crystals grown under conditions mimicking their physiological environment [28], whereas in crystals obtained under far from physiological environment these regions are highly disordered [7].

While developing crystallographic procedures, we obtained a starting model by electron microscopy, using three-dimensional image reconstruction from two dimensional sheets (Fig. 18.7). These studies revealed that nascent proteins progress zcated protection of nascent chains by the ribosome [36, 48]. However, the common notion

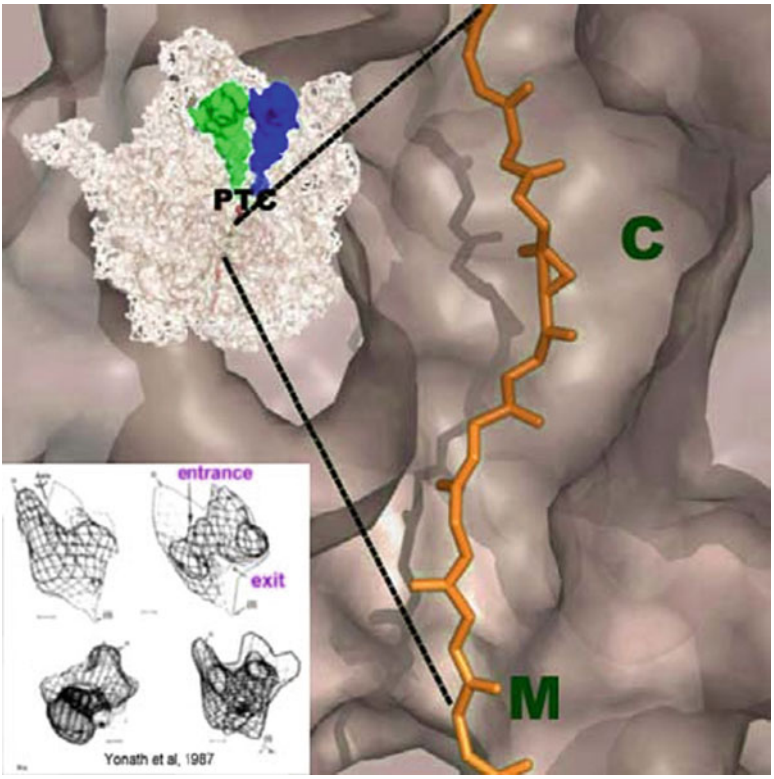


Fig. 18.7 A section through the large subunit (*top left*) and the protein exit tunnel, in which poly-alanine is docked. *C* denotes a crevices where initial folding can take place and *M* is the location of the macrolides binding pocket. The initial three dimensional reconstructions (from 2D sheets) are shown in the *bottom-left* inset

that nascent proteins progress on the ribosome surface until its maturation, raised doubts in the existence of the tunnel [41, 47], even after its visualization.

Alongside the improvement of the quality of the micro- or poorly diffracting crystals, our studies required the development of innovative methodologies. Among these is the pioneering of bio-crystallography at cryogenic temperatures, which was introduced because of the extreme radiation sensitivity of the ribosomal crystals [30] and became almost instantaneously the routine method all over the world, thus enabling structure determination from crystals considered not useful previously. Also, we introduced an unconventional use of multi-heavy atom clusters [60] (Fig. 18.8).

One of them, the heteropolytungstate (NH4)6P2W18O62 was found to play a dual role in the determination of the structure of structure of the small ribosomal subunit from *T. thermophilus*. Thus, in addition to significant phasing power and anomalous signal, post crystallization treatment with minute amount of one of this cluster increased dramatically the resolution of the X-ray diffraction from the initial low resolution (7–9 Å) to ~3Å [25, 50] presumably by minimizing the internal

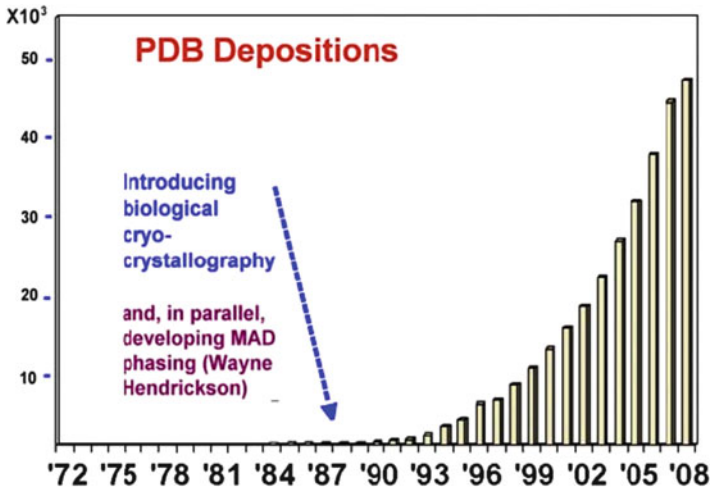


Fig. 18.8 PDB depositions by the year

flexibility involved naturally in mRNA binding to the ribosome and its progression through the ribosome.

18.6 Conclusions

By providing molecular snapshots of various intermediates in ribosome-mediated translation in atomic detail, the high resolution structures have revolutionized our understanding of the mechanism of protein synthesis. Despite this impressive progress, countless new questions arose. Many of which concern structural dynamics and intricate localized rearrangements, with answers that may emerge by combination of approaches like X-ray crystallography, cryo EM, FRET and biochemistry. An striking advance in this direction is the recent ability to follow translation by single ribosomes, one codon at a time) using mRNA hairpins tethered by the ends to optical tweezers [65].

Acknowledgments Support was provided by the US National Inst. of Health (GM34360), and the Kimmelman Center for Macromolecular Assemblies. AY holds the Martin and Helen Kimmel Professorial Chair.

References

1. Agmon I, Bashan A, Zarivach R, Yonath A (2005) Symmetry at the active site of the ribosome: structural and functional implications. Biol Chem 386:833–844

2. Agmon I, Bashan A, Yonath A (2006) On ribosome conservation and evolution. Isr J Ecol Evol 52:359–379
3. Auerbach T, Bashan A, Yonath A (2004) Ribosomal antibiotics: structural basis for resistance, synergism and selectivity. Trends Biotechnol 22:570–576
4. Auerbach T, Mermershtain I, Bashan A, Davidovich C, Rozenberg H, Sherman DH, Yonath A (2009) Structural basis for the antibacterial activity of the 12-membered-ring mono-sugar macrolide methymycin. Biotechnology 84:24–35
5. Auerbach T, Mermershtain I, Davidovich C, Bashan A, Belousoff M et al (2010) The structure of ribosome-lankacidin complex reveals ribosomal sites for synergistic antibiotics. Proc Natl Acad Sci USA 107:1983–1988
6. Auerbach-Nevo T, Zarivach R, Peretz M, Yonath A (2005) Reproducible growth of well diffracting ribosomal crystals. Acta Crystallogr D Biol Crystallogr 61:713–719
7. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. Science 289:905–920
8. Baram D, Pyetan E, Sittner A, Auerbach-Nevo T, Bashan A, Yonath A (2005) Structure of trigger factor binding domain in biologically homologous complex with eubacterial ribosome reveals its chaperone action. Proc Natl Acad Sci USA 102:12017–12022
9. Bashan A, Agmon I, Zarivach R, Schluenzen F, Harms J, Berisio R, Bartels H, Franceschi F et al (2003) Structural basis of the ribosomal machinery for peptide bond formation, translocation, and nascent chain progression. Mol Cell 11:91–102
10. Belousoff MJ, Davidovich C, Zimmerman E, Caspi Y, Wekselman I et al (2010) Ancient machinery embedded in the contemporary ribosome. Biochem Soc Trans 38:422–427
11. Beringer M, Rodnina MV (2007) The ribosomal peptidyl transferase. Mol Cell 26:311–321
12. Beringer M, Bruell C, Xiong L, Pfister P, Bieling P, Katunin VI, Mankin AS, Bottger EC, Rodnina MV (2005) Essential mechanisms in the catalysis of peptide bond formation on the ribosome. J Biol Chem 280:36065–36072
13. Berisio R, Schluenzen F, Harms J, Bashan A, Auerbach T, Baram D, Yonath A (2003) Structural insight into the role of the ribosomal tunnel in cellular regulation. Nat Struct Biol 10:366–370
14. Berisio R, Corti N, Pfister P, Yonath A, Bottger EC (2006) 23S rRNA 2058A–<G alteration mediates ketolide resistance in combination with deletion in L22. Antimicrob Agents Chemother 50:3816–3823
15. Bokov K, Steinberg SV (2009) A hierarchical model for evolution of 23S ribosomal RNA. Nature 457:977–980
16. Bommakanti AS, Lindahl L, Zengel JM (2008) Mutation from guanine to adenine in 25S rRNA at the position equivalent to E. coli A2058 does not confer erythromycin sensitivity in Sacchromyces cerevisae. RNA 14:460–464
17. Bottger EC (2006) The ribosome as a drug target. Trends Biotechnol 24:145–147
18. Bottger EC (2007) Antimicrobial agents targeting the ribosome: the issue of selectivity and toxicity – lessons to be learned. Cell Mol Life Sci 64:791–795
19. Davidovich C, Bashan A, Auerbach-Nevo T, Yaggie RD, Gontarek RR, Yonath A (2007) Induced-fit tightens pleuromutilins binding to ribosomes and remote interactions enable their selectivity. Proc Natl Acad Sci USA 104:4291–4296
20. Davidovich C, Bashan A, Yonath A (2008) Structural basis for cross-resistance to ribosomal PTC antibiotics. Proc Natl Acad Sci USA 105:20665–20670
21. Davidovich C, Belousoff MJ, Bashan A, Yonath A (2010) Ancient machinery embedded in the contemporary ribosome. Res Microbiol 18:18–25
22. Davydova N, Streltsov V, Wilce M, Liljas A, Garber M (2002) L22 ribosomal protein and effect of its mutation on ribosome resistance to erythromycin. J Mol Biol 322:635–644
23. Frank J, Gonzalez RL Jr (2010) Structure and dynamics of a processive Brownian motor: the translating ribosome. Annu Rev Biochem 79:381–412
24. Gindulyte A, Bashan A, Agmon I, Massa L, Yonath A, Karle J (2006) The transition state for formation of the peptide bond in the ribosome. Proc Natl Acad Sci USA 103:13327–13332

25. Gluehmann M, Zarivach R, Bashan A, Harms J, Schluenzen F et al (2001) Ribosomal crystallography: from poorly diffracting microcrystals to high-resolution structures. Methods 25:292–302
26. Gregory ST, Dahlberg AE (1999) Erythromycin resistance mutations in ribosomal proteins L22 and L4 perturb the higher order structure of 23S ribosomal RNA. J Mol Biol 289:827–834
27. Hansen HA, Volkmann N, Piefke J, Glotz C, Weinstein S, Makowski I, Meyer S, Wittmann HG, Yonath A (1990) Crystals of complexes mimicking protein biosynthesis are suitable for crystallographic studies. Biochim Biophys Acta 1050:1–7
28. Harms J, Schluenzen F, Zarivach R, Bashan A, Gat S, Agmon I, Bartels H, Franceschi F, Yonath A (2001) High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. Cell 107:679–688
29. Harms J, Schluenzen F, Fucini P, Bartels H, Yonath A (2004) Alterations at the peptidyl transferase center of the ribosome induced by the synergistic action of the streptogramins dalfopristin and quinupristin. BMC Biol 2:1–10
30. Hope H, Frolow F, von Bohlen K, Makowski I, Kratky C (1989) Cryocrystallography of ribosomal particles. Acta Crystallogr B 45:190–199
31. Hsiao C, Mohan S, Kalahar BK, Williams LD (2009) Mol Biol Evol 23:23
32. Ippolito JA, Kanyo ZF, Wang D, Franceschi FJ, Moore PB, Steitz TA, Duffy EM (2008) Crystal structure of the oxazolidinone antibiotic linezolid bound to the 50S ribosomal subunit. J Med Chem 51:3353–3356
33. Korostelev A, Trakhanov S, Laurberg M, Noller HF (2006) Crystal structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements. Cell 126:1065–1077
34. Korostelev A, Ermolenko DN, Noller HF (2008) Structural dynamics of the ribosome. Curr Opin Chem Biol 12:674–683
35. Lawrence M, Lindahl L, Zengel JM (2008) Effects on translation pausing of alterations in protein and RNA components of the ribosome exit tunnel. J Bacteriol 190:5862–5869
36. Malkin LI, Rich A (1967) Partial resistance of nascent polypeptide chains to proteolytic digestion due to ribosomal shielding. J Mol Biol 26:329–346
37. Mankin AS (2001) Ribosomal antibiotics. Mol Biologia 35:509–520
38. Mankin AS (2006) Nascent peptide in the “birth canal” of the ribosome. Trends Biochem Sci 31:11–13
39. Mankin AS (2008) Macrolide myths. Curr Opin Microbiol 11:414–421
40. Milligan RA, Unwin PN (1986) Location of exit channel for nascent protein in 80S ribosome. Nature 319:693–695
41. Moore PB (1988) The ribosome returns. Nature 331:223–227
42. Moore SD, Sauer RT (2008) Revisiting the mechanism of macrolide-antibiotic resistance mediated by ribosomal protein L22. Proc Natl Acad Sci USA 105:18261–18266
43. Nissen P, Hansen J, Ban N, Moore PB, Steitz TA (2000) The structural basis of ribosome activity in peptide bond synthesis. Science 289:920–930
44. Pfister P, Corti N, Hobbie S, Bruell C, Zarivach R, Yonath A, Bottger EC (2005) 23S rRNA base pair 2057-2611 determines ketolide susceptibility and fitness cost of the macrolide resistance mutation 2058A–<G. Proc Natl Acad Sci USA 102:5180–5185
45. Poehlsgaard J, Douthwaite S (2005) The bacterial ribosome as a target for antibiotics. Nat Rev Microbiol 3:870–881
46. Ramakrishnan V (2008) What we have learned from ribosome structures. Biochem Soc Trans 36(Pt4):567–574
47. Ryabova LA, Selivanova OM, Baranov VI, Vasiliev VD, Spirin AS (1988) Does the channel for nascent peptide exist inside the ribosome? Immune electron microscopy study. FEBS Lett 226:255–260
48. Sabatini DD, Blobel G (1970) Controlled proteolysis of nascent polypeptides in rat liver cell fractions: II. Location of the polypeptides in rough microsomes. J Cell Biol 45:146–157
49. Sato NS, Hirabayashi N, Agmon I, Yonath A, Suzuki T (2006) Comprehensive genetic selection revealed essential bases in the peptidyl-transferase center. Proc Natl Acad Sci USA 103:15386–15391

50. Schluenzen F, Tocilj A, Zarivach R, Harms J, Gluehmann M, Janell D, Bashan A et al (2000) Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. Cell 102:615–623
51. Schluenzen F, Zarivach R, Harms J, Bashan A, Tocilj A, Albrecht R, Yonath A, Franceschi F (2001) Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. Nature 413:814–821
52. Schluenzen F, Harms JM, Franceschi F, Hansen HA, Bartels H, Zarivach R, Yonath A (2003) Structural basis for the antibiotic activity of ketolides and azalides. Structure 11:329–338
53. Schluenzen F, Wilson DN, Tian P, Harms JM, McInnes SJ, Hansen HA, Albrecht R, Buerger J, Wilbanks SM, Fucini P (2005) The binding mode of the trigger factor on the ribosome: implications for protein folding and SRP interaction. Structure (Camb) 13:1685–1694
54. Schmeing TM, Huang KS, Strobel SA, Steitz TA (2005) An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA. Nature 438:520–524
55. Schuwirth BS, Day JM, Hau CW, Janssen GR, Dahlberg AE, Cate JH, Vila-Sanjurjo A (2006) Structural analysis of kasugamycin inhibition of translation. Nat Struct Mol Biol 13(10):879–886
56. Selmer M, Dunham CM, Murphy Iv FV, Weixlbaumer A et al (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. Science 313:1935–1942
57. Simonetti A, Marzi S, Myasnikov AG, Fabbretti A, Yusupov M, Gualerzi CO, Klaholz BP (2008) Structure of the 30S translation initiation complex. Nature 455(7211):416–420
58. Steitz TA (2008) A structural understanding of the dynamic ribosome machine. Nat Rev Mol Cell Biol 9:242–253
59. Tenson T, Mankin A (2006) Antibiotics and the ribosome. Mol Microbiol 59:1664–1677
60. Thygesen J, Weinstein S, Franceschi F, Yonath A (1996) The suitability of multi-metal clusters for phasing in crystallography of large macromolecular assemblies. Structure 4:513–518
61. Vazquez-Laslop N, Thum C, Mankin AS (2008) Molecular mechanism of drug-dependent ribosome stalling. Mol Cell 30:190–202
62. von Bohlen K, Makowski I, Hansen HA, Bartels H et al (1991) Characterization and preliminary attempts for derivatization of crystals of large ribosomal subunits from Haloarcula marismortui diffracting to 3Å resolution. J Mol Biol 222:11–15
63. Voorhees RM, Weixlbaumer A, Loakes D, Kelley AC, Ramakrishnan V (2009) Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. Nat Struct Mol Biol 16:528–533
64. Weinger JS, Strobel SA (2006) Translation. Biochemistry 45:5939–5948
65. Wen JD, Lancaster L, Hodges C, Zeri AC, Yoshimura SH, Noller HF, Bustamante C, Tinoco I (2008) Following translation by single ribosomes one codon at a time. Nature 452:598–603
66. Yassin A, Fredrick K, Mankin AS (2005) Deleterious mutations in small subunit ribosomal RNA identify functional sites and potential targets for antibiotics. Proc Natl Acad Sci USA 102:16620–16625
67. Yonath A (2003) Ribosomal tolerance and peptide bond formation. Biol Chem 384:1411–1419
68. Yonath A (2005) Antibiotics targeting ribosomes: resistance, selectivity, synergism and cellular regulation. Annu Rev Biochem 74:649–679
69. Yonath A, Bashan A (2004) Ribosomal crystallography: initiation, peptide bond formation, and amino acid polymerization are hampered by antibiotics. Annu Rev Microbiol 58:233–251
70. Yonath A, Leonard KR, Wittmann HG (1987) A tunnel in the large ribosomal subunit revealed by three-dimensional image reconstruction. Science 236:813–816
71. Yonath A, Muessig J, Tesche B, Lorenz S, Erdmann VA, Wittmann HG (1980) Crystallization of the large ribosomal subunits from *Bacillus stearothermophilus*. Biochem Int 1:315–428
72. Zaman S, Fitzpatrick M, Lindahl L, Zengel J (2007) Novel mutations in ribosomal proteins L4 and L22 that confer erythromycin resistance in Escherichia coli. Mol Microbiol 66:1039–1050
73. Zimmerman E, Yonath A (2009) Biological implications of the ribosome's stunning stereochemistry. Chembiochem 10:63–72