GENETIC MANIPULATION

OF DNA AND PROTEIN

EXAMPLES FROM
CURRENT RESEARCH

Edited by **David Figurski**

GENETIC MANIPULATION OF DNA AND PROTEIN – EXAMPLES FROM CURRENT RESEARCH

Edited by **David Figurski**

Genetic Manipulation of DNA and Protein – Examples from Current Research
Edited by David Figurski

Contributors

Deepak Bastia, S. Zzaman, Bidyut K. Mohanty, J. Esclapez, M. Camacho, C. Pire, M.J. Bonete, David H. Figurski, Daniel H. Fine, Brenda A. Perez-Cheeks, Valerie W. Grosso, Karin E. Kram, Jianyuan Hua, Ke Xu, Jamila Hedhli, Jürgen Ludwig, Holger Rabe, Anja Höffle-Maas, Marek Samochocki, Alfred Maelicke, Titus Kaletta, Luis Eduardo S. Netto, Marcos Antonio Oliveira, Toni Petan, Petra Prijatelj Žnidaršič, Jože Pungerčar, Ewa Sajnaga, Ryszard Szyszka, Konrad Kubiński, Jane E. Carland, Amelia R. Edington, Amanda J. Scopelliti, Renae M. Ryan, Robert J. Vandenberg, José Manuel Pérez-Donoso, Claudio C. Vásquez, Kevin Hadi, Oznur Tastan, Alagarsamy Srinivasan, Velpandi Ayyavoo, Ahmed Chraibi, Stéphane Renauld, M. Tang, K.J. Wierenga, K. Lai, Christelle Bonod-Bidaud, Florence Ruggiero, Silvio Alejandro López-Pazos, Jairo Cerón, Juanita Yazmin Damián-Almazo, Gloria Saab-Rincón, Stathis Frillingos, Roman G. Gerlach, Kathrin Blank, Thorsten Wille, Nathan A. Sieracki, Yulia A. Komarova, Shona A. Mookerjee, Elaine A. Sia, Joy Sturtevant, James W. Wilson, Clayton P. Santiago, Jacquelyn Serfecz, Laura N. Quick

Notice

Statements and opinions expressed in the chapters are these of the individual contributors and not necessarily those of the editors or publisher. No responsibility is accepted for the accuracy of information contained in the published chapters. The publisher assumes no responsibility for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained in the book.

Contents

Preface

This diverse collection of research articles is united by the enormous power of modern molecular genetics. The current period is an exciting time both for researchers and the curious who want to know more about genetic approaches to solving problems.

This volume is noteworthy. Every author accomplished two important objectives: (1) making the field and the particular research described accessible to a large audience and (2) explaining fully the genetic tools and approaches that were used in the research. One fact stands out – the importance of a genetic approach to addressing a problem. I encourage you to read several chapters. You will feel the excitement of the scientists, and you will learn about an area of research with which you may not be familiar. Perhaps most importantly, you will understand the genetic approaches; and you will appreciate their importance to the research.

Anyone can benefit from reading these chapters – even those of you who have a solid foundation in modern molecular genetics. This is an eclectic mix of topics (only the surface has been scratched). These chapters are valuable, not only because they reflect the current state of the art and are easy to read, but also because they are concise reviews. The variety will provide you with new knowledge to be sure, but it may also affect your own thoughts about a problem. Thinking about a topic very different from the one you are considering can stimulate fresh and often unconventional ideas.

We all know that the code for all life on the planet is in DNA and RNA. The purpose of genetics is to decipher life's information – to understand why the genome codes for its various functions. Much of the work in this volume is geared to manipulating DNA with that knowledge, not only to provide clues about a function, but also to test an idea or to change a protein to learn how it works or to make it work better.

For a time, the field of molecular genetics was concerned with a few manipulable model organisms. This was necessary to answer basic questions like "How does a gene work?" Now modern molecular genetics has given us the confidence to explore the unknowns in the diversity of life, including complex organisms, like humans. We may need to adapt or develop genetic tools (see the contents section on tools). We have already learned that many of the "paradigms" of the model organisms do not apply to other organisms.

"Manipulate" is a problem word in genetics for some people. This volume has another purpose - to be accessible to those who fear the power of genetics. Those of us who know modern genetics understand that the current precision of genetically modified food, for example, is far safer than the unknowns of genetic crosses, a technology that is strangely acceptable. We have ourselves to blame for the apparent mystery and the public's misperceptions. Too often we discuss our work with our colleagues but fail to explain our work to the public.

By making these chapters freely available to everyone and by the authors clearly describing the question being asked and the approach taken to answer it, this book is partly addressing that concern. People who fear genetics should take comfort in the dissemination of knowledge about this science. Scientists have the same concerns as the public. The more who understand genetics, the more there will be vigilance.

This collection of research articles is testimony to the optimism in the field. Both major and minor problems can be solved. For example, genetics will likely be a part of the solution to hunger, and genetically engineered microorganisms may help solve the problem of global warming. Basic research (see the contents sections on basic research and the development of approaches and tools) is difficult to explain, but it is vitally important for any progress. Genetics will help alleviate suffering by leading to new therapies for disease (see the contents section on disease-related research), and it can generate improved or new molecular activities (see the contents section on applied research).

With a complete understanding of genetics, humankind will reach an important new stage. Humans will be able to change their own genes. Of course, evolution will continue to be an agent of genetic change; but it is slow in humans, and it acts on populations. With the knowledge of genetics, humans will be able to direct change (like the curing of a disease) to an individual; and it can be rapid.

You will be exposed to investigations on bacteria, archaea, fungi, mitochondria, and higher eukaryotes, including humans. You will learn about various genetic approaches, including specific alteration of amino acid residues in proteins, gene fusions, cysteine- and alanine-scanning mutagenesis, recombineering, cloning by "capturing" large segments of DNA, transposable elements, and allelic exchange. The chapters are all very readable, and again I encourage you to sample more than one.

David Figurski
Professor of Microbiology & Immunology at Columbia University,
USA

Molecular Genetics in Basic Research

Site-Directed Mutagenesis and Yeast Reverse 2-Hybrid-Guided Selections to Investigate the Mechanism of Replication Termination

Deepak Bastia, S. Zzaman and Bidyut K. Mohanty
Department of Biochemistry and Molecular Biology,
Medical University of SC, Charleston, SC
USA

1. Introduction

DNA replication in prokaryotes, in budding yeast and in mammalian DNA viruses initiates from fixed origins (*ori*) and the replication forks are extended in either a bidirectional mode or in some cases unidirectionally (Cvetic and Walter, 2005; Sernova and Gelfand, 2008; Wang and Sugden, 2005; Weinreich et al., 2004). In higher eukaryotes there are preferred sequences located in AT-rich islands that serve as origins (Bell and Dutta, 2002). In many prokaryotes, the two replication forks initiated at *ori* on a circular chromosome meet each other at specific sequences called replication termini or *Ter* (Bastia and Mohanty, 1996; Kaplan and Bastia, 2009). The Ter sites bind to sequence-specific DNA binding proteins called replication terminator proteins that allow forks approaching from one direction to be impeded at the terminus, whereas forks coming from the opposite direction pass through the site unimpeded (Bastia and Mohanty, 1996, 2006; Kaplan and Bastia, 2009). Therefore, the mode of fork arrest is polar. The polarity of fork arrest in *Escherichia coli* and *Bacillus subtilis* is caused by the complexes of the terminator proteins called Tus and RTP (Replication Terminator Protein), respectively, with the cognate *Ter* sites to arrest the replicative helicase (such as DnaB in case of *E. coli*) in a polar mode (Kaul et al., 1994; Khatri et al., 1989; Lee et al., 1989; Sahoo et al., 1995). What is the mechanism of polar fork arrest and what might be the physiological functions of *Ter* sites? Using *E. coli* as the main example, with the aid of the techniques of site-directed mutagenesis, yeast reverse 2-hybrid based selection of random mutations (described below), and biochemical characterizations of the mutant forms of the Tus protein, many aspects of the mechanism of replication fork arrest at Tus-*Ter* complexes have been determined. This and a brief description of the current state of the knowledge of replication termination in eukaryotes have also been reviewed below.

Replication termini of *E. coli* **and the plasmid R6K:** Sequence-specific replication termini were first discovered in the drug resistance plasmid R6K (Crosa et al., 1976; Kolter and Helinski, 1978) and in its host *E. coli* (Kuempel et al., 1977). The terminus region of R6K was identified and sequenced (Bastia et al., 1981) and subsequently shown to consist of a pair of *Ter* sites with opposite polarity (Hidaka et al., 1988). An *in vitro* replication system was

developed in which host cell extracts initiated replication of a plasmid DNA template and the moving forks were arrested at the *Ter* sites (Germino and Bastia, 1981). It was also suggested that a terminator protein that might cause fork arrest was likely to be host-encoded. Subsequently, the open reading frame (ORF) encoding the terminator protein was cloned and sequenced and the gene was named TUS (Terminus Utilizing Substance) (Hill et al., 1989). Tus protein was purified from cell extract of *E. coli* and shown to bind to the plasmid *Ter* sequences (Sista et al., 1991; Sista et al., 1989). The *TerC* region of *E. coli* was found to contain several *Ter* sites in two sets of 5 sites each with one cluster having the opposite polarity of fork arrest in comparison with that of the second set (Hill, 1992; Pelletier et al., 1988). Together, these sequences formed a replication trap (Fig.1A). For example, if the clockwise moving fork got arrested at *TerC*, it waited there for the counterclockwise fork to meet it at the site of arrest. The *Ter* consensus sequence is shown in Fig.1B. Site-directed mutagenesis showed the bases that are critical for Tus binding (Duggan et al., 1995; Sista et al., 1991). The complete process of initiation, elongation and termination has been carried out *in vitro* with 22 purified proteins that were necessary and sufficient for fork initiation, propagation and termination (Abhyankar et al., 2003).
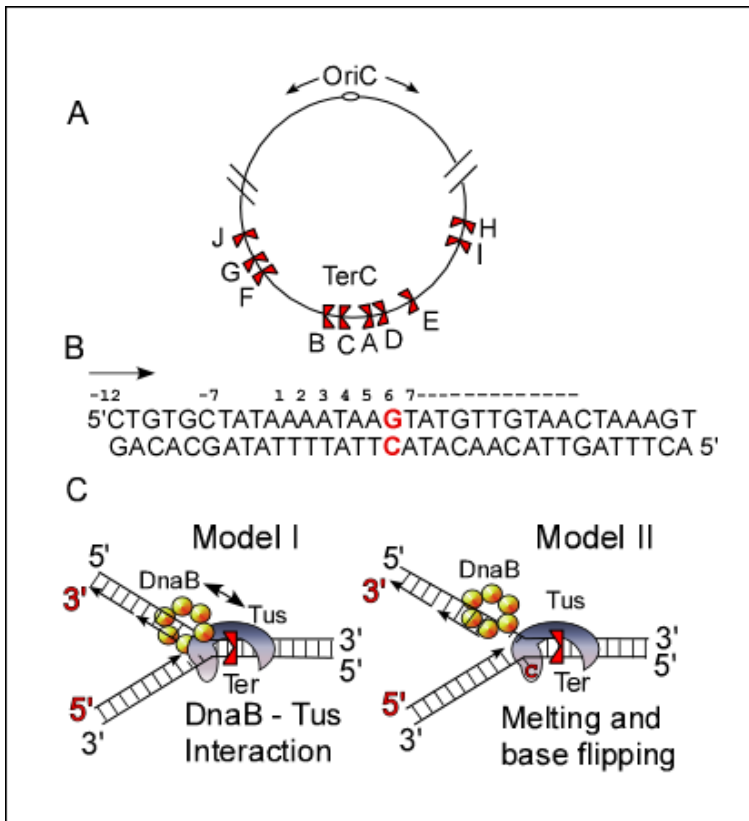
Fig. 1. Replication termini of *E. coli*. A, The bacterial replicon showing the origin and the *TerC* region at its antipode. The flat surfaces of the *Ter* sites indicate the permissive face and

the notched one the nonpermissive face; B, consensus Ter sequence showing the blocking end at the left (arrow) and the nonblocking end at the right; the red C on the bottom strand was reported to flip out upon Tus binding; C, two models of polar fork arrest. Model 1 postulates that both Tus binding to *Ter* and interaction or contact between the nonpermissive face of the Tus-*Ter* complex with DnaB helicase causes polar arrest; model 2 suggests that it is strictly the Tus-*Ter* interaction and the partial melting of the DNA catalyzed by DnaB and the flipping of C6 that causes strong affinity of Tus for *Ter*. The helicase approaching the permissive face fails to induce high-affinity binding of Tus to *Ter*.

Using an *in vitro* helicase assay catalyzed by purified DnaB and Tus proteins, it was shown that Tus binding to *Ter* acts as a polar contra- or anti-helicase and arrests helicase catalyzed DNA unwinding in one orientation of the Tus-*Ter* complex while allowing the helicase to pass through mostly unimpeded in the opposite orientation (Khatri et al., 1989; Lee et al., 1989). It was also shown that the RTP of *B. subtilis* arrested *E. coli* DnaB helicase at the cognate *Ter* sites of the Gram-positive bacterium *in vitro* was able to arrest DnaB of *E. coli* in a polar mode. However, it did not arrest rolling circle replication of a plasmid (Kaul et al., 1994). It is of some interest that not all helicases were arrested at Tus-*Ter* complexes because helicases such as Rep and UvrD were not arrested by either orientations of Tus-*Ter* (Sahoo et al., 1995). The Tus-*Ter* complex of *E. coli* could arrest forks with a very low efficiency *in vivo* in the *B. subtils* host, as contrasted with their ability to arrest forks more efficiently in the natural host. In addition to DnaB, RNA polymerase of bacteriophage T7 and *E. coli* were also arrested in a polar mode, by the Tus-*Ter* complex (Mohanty et al., 1996, 1998). This latter observation had raised the possibility that the Tus-*Ter* complex might just be a steric barrier to unwinding because enzymes apparently as diverse as DnaB helicase and RNA polymerases were arrested by the same complex. This mechanistic issue has been discussed in more detail later.

Crystal structures of Terminator proteins: The first crystal structure of a terminator apoprotein, namely that of RTP, showed that the protein was a symmetrical winged helix dimer (Fig.2B) (Bussiere et al., 1995). The *Ter* sites of *B. subtilis* contain overlapping core and auxiliary sequences with each site binding an RTP dimer (Hastings et al., 2005; Smith and Wake, 1992; Wilce et al., 2001). How can a symmetrical protein arrest forks with polarity? This question was subsequently answered when the crystal structure of two dimeric RTPs bound to a complete bipartite *Ter* site was solved (Wilce et al., 2001). It was shown that the structure of the protein-DNA complex is different at the core complex as contrasted with that of the adjacent auxiliary complex. The crystal structure of Tus bound to *Ter* DNA showed a bi-lobed protein with a positively charged cleft formed by several beta strands that contacted the major groove of the DNA and distorted the latter from the canonical structure (Fig.2A) (Kamada et al., 1996). The transverse view of Tus bound to a space-filling model of DNA shows that the face that arrests replication forks and DnaB has a loop called the L1 loop. The L1 loop appears to play a critical role in fork arrest.

Tus-DnaB interaction: We performed yeast 2-hybrid analysis (described below), confirmed by *in vitro* affinity binding to immobilized Tus, to show that DnaB interacted with Tus (Mulugu et al., 2001). The principles of forward 2-hybrid (Fields and Song, 1989) and reverse 2-hybrid analysis (Mulugu et al., 2001; Sharma et al., 2001) are shown in Fig.3. The open reading frame (ORF) of a protein X is cloned in the correct reading frame to the transcriptional activation domain of Gal4 of yeast (pGAD424-X). A suspected interacting
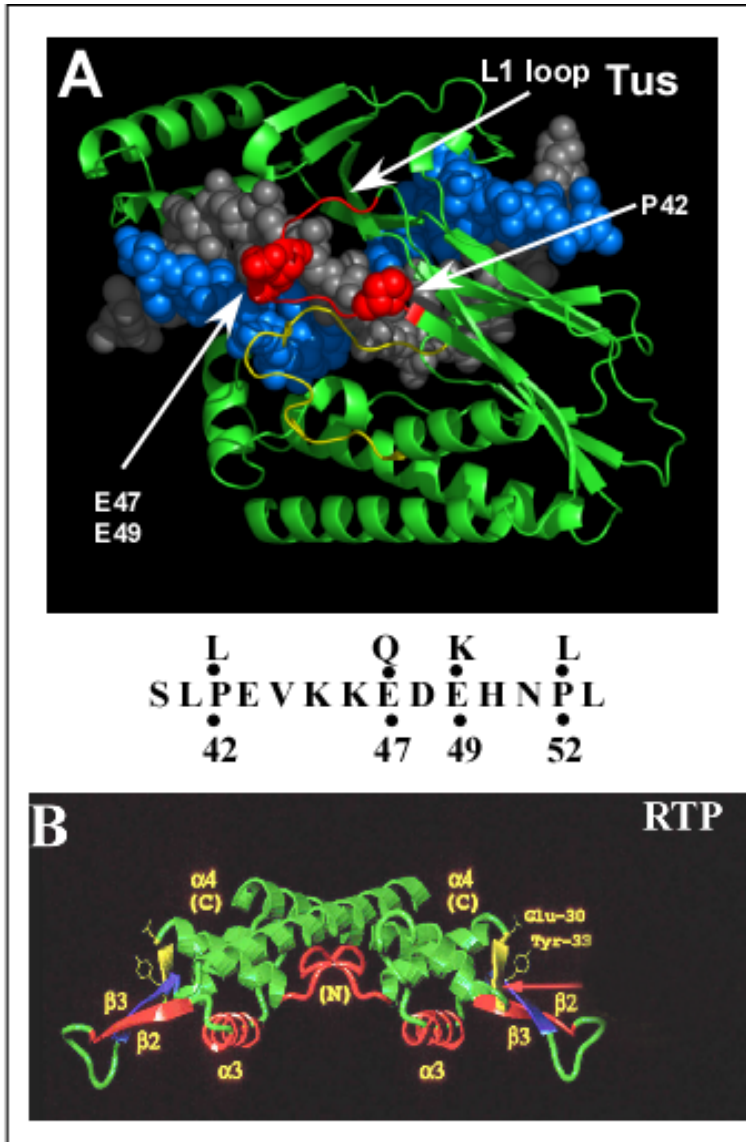
Fig. 2. Crystal structure of Tus-*Ter* complex of *E. coli* and RTP apoprotein of *B. subtilis*. A, crystal structure of Tus-*Ter* complex showing the blocking face with the L1 loop shown in red. Three residues, namely P42, E47 and E49, when mutated (see lower sequence) show impaired helicase arrest. P42L shows slightly reduced DNA binding; E47Q shows stronger DNA binding; and E49K shows no reduction in *Ter* binding but significant reduction in fork and helicase arrest. B, crystal structure of the RTP dimer apoprotein. The Tyr-33 arrow depicts a residue needed for the interaction of Tus with DnaB, as shown by a bifunctional labeled crosslinker that upon cleavage at an S-S bond transfers the label from RTP to DnaB.
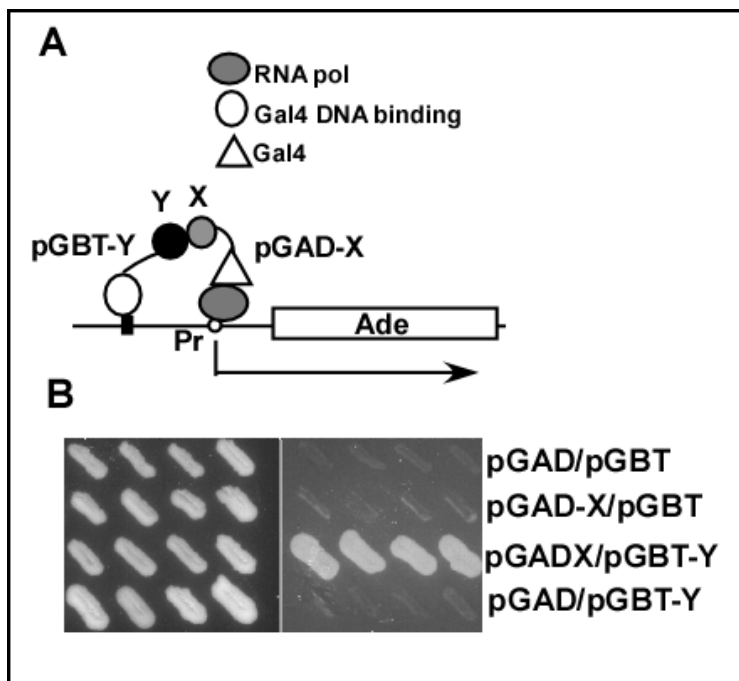
Fig. 3. Schematic representation of forward and reverse 2-hybrid selection. A, The plasmids pGBT-Y and pGAD-X interact through interacting proteins X and Y and turn on the *Ade* reporter gene leading to growth on adenine (ade) dropout minimal medium. Either X or Y is mutagenized by low-fidelity PCR and introduced by transformation in the presence of the other plasmid into the indicator yeast strain. B, X-Y interaction leads to growth on ade-minus plates, and mutants that fail to interact show lack of growth on the selective plates. Trivial mutations, *i.e.*, those containing deletions, nonsense mutations, or frame-shifts are eliminated by Western blotting of cell extracts expressing the presumed X or Y mutant form. Candidates are further characterized by functional and biochemical analyses.

protein Y is similarly fused in-frame to the ORF of the DNA binding domain of Gal4. The yeast strain contains a transcriptional reporter (*Ade*) that is placed next to a promoter and the binding site for the Gal4 DNA binding site. Neither pGAD424-X nor pGBT9-Y can activate the transcription of the reporter gene. However, when both plasmids, each containing a different marker (*e.g.*, *Leu* and *Trp*), are transformed into the reporter yeast strain, X-Y interaction activates the reporter gene. Both plasmids are shuttle vectors that contain an *ori* active in *E. coli* and also an *ori* (*ars*) of yeast. The transcription and translation of the adenine (*Ade*) reporter causes the yeast cells to grow in an adenine dropout minimal medium plate. The reverse 2-hybrid procedure was used to select for missense mutations that break X-Y interaction as follows. Low fidelity PCR amplification of X (or Y) introduces random mutations into the ORF. Then, for example, the mutagenized ORF of X in the pGAD424 vector is used to transform the *Ade* reporter yeast strain containing a resident pGBT9-Y plasmid. Colonies that have mutations that break X-Y interaction are initially selected as clones growing on *Leu-Trp-* medium but failing to grow on *Leu-Trp-Ade-* dropout

plates. The mutations are expected to be a mixture of unwanted ones (*e.g.* missense, nonsense, frame-shifts) and useful ones (missense). The potential mutant clones are grown, cell-free lysates made and subjected to Western blots after polyacrylamide gel electrophoresis and developed with primary antibody raised against X followed by secondary reporter antibody. All clones that fail to produce the protein of the expected length are discarded, and those producing full length X-GAD are saved for further analysis.

Usually, the mutants are confirmed by co-immunoprecipitation of cell lysates with the anti-Y antibody (Ab) retained on agarose beads, stripping of the wild type (WT) X (or mutant X that should be in the wash), separation by gel electrophoresis and visualization with anti-Y Ab. Naturally, the authentic non-interaction mutant forms of X should no longer bind to Y or bind poorly. These "pull down" assays are used to confirm the reverse 2-hybrid results. If the interaction of X and Y is necessary for a biological function (*e.g.*, fork arrest at Tus-*Ter* complex), the X mutants that do not interact with protein Y are then tested by 2-dimensional agarose gel electrophoresis (Brewer and Fangman, 1987, 1988; Mohanty et al., 2006; Mohanty and Bastia, 2004) to determine whether they show the expected biochemical property (in this case, failure to arrest replication forks) (Mulugu et al., 2001). The reverse 2-hybrid approach is a powerful method that can yield mutants that specifically disrupt protein-protein interaction between a pair of known interacting proteins. This procedure can be followed up by isolation of additional mutations isolated by site-directed mutagenesis of residues close to the protein domain (as determined by X-ray crystallography) that contained the mutations recovered from the reverse 2-hybrid approach. A specific example is given below. By mutagenizing Tus by PCR, we were able to collect a pool of random mutants. We performed reverse 2-hybrid analysis of the mutant pool and recovered the mutation P42L (proline at position 42 to leucine) that fails to interact with DnaB. However, a P42L mutation also affected Tus-*Ter* binding to some extent. We mutagenized other residues by site-directed mutagenesis to isolate E47Q (glutamic acid at position 47 to glutamine) and E49K (glutamic acid at position 49 to lysine) (Fig. 2 and 3). Both of the latter mutants were defective in interaction with DnaB and in fork arrest *in vitro*. Whereas the E49K mutant form bound to *Ter* with the same affinity as WT Tus, E47Q had a higher DNA-binding affinity but was defective in fork arrest *in vivo* (Mulugu et al., 2001).

The yeast forward and reverse 2-hybrid analyses followed by biochemical analysis of Tus, showed that it contacted DnaB probably at the L1 loop because the only mutations that impaired helicase arrest and fork arrest without abolishing or significantly reducing Tus-*Ter* interaction were found only at the L1 loop. Another line of evidence for specific replisome-*Ter* interaction is inferred from the observation that that Tus-*Ter* complex works with very low efficiency when placed in *B. subtilis* cells as contrasted with their fork arrest efficiency in *E. coli in vivo* (Andersen et al., 2000).

If there is protein-protein interaction between Tus and DnaB and if this is necessary for fork arrest, how does Tus also promote polar arrest of RNA polymerase, an enzyme apparently different in structure from DnaB? One possible explanation is that Tus might make an equivalent contact with RNA polymerase to inhibit its progression, or else a different mechanism could be operating here. It should, however, be clearly stated that this line of reasoning does not necessarily disprove the first explanation. Based on the data discussed above, we have suggested a model of fork arrest that involves not only stable Tus-*Ter* interaction, but also protein-protein contacts between the DnaB helicase and the L1 loop of Tus (Fig.1C and Fig.2).

Base flipping and DNA melting: An alternative explanation of polar arrest is suggested in model 2 (Fig.1C). X-ray crystallography of Tus bound to linear DNA had shown all Watson-Crick base pairing (Kamada et al., 1996). However, it was reported that a forked DNA that had single stranded regions when co-crystallized with Tus showed a flipped base (C6 in Fig 1C, model 2). It was suggested that both DNA melting and base flipping and the capture of the flipped base by Tus greatly enhanced Tus binding for *Ter* when the helicase approached the blocking end of the Tus-*Ter* complex. The enzyme, when approaching the complex from the non-blocking end, displaced Tus from *Ter*. This interpretation was based on binding studies of Tus to *Ter* on partially single-stranded DNA having a flipped C (Mulcair et al., 2006). Unfortunately, these binding studies were performed between 150 mM-250 mM KCl at which DNA replication and DnaB activity *in vitro* is inhibited by >90% . Curiously, when binding was performed closer to a physiological salt concentration that is permissive of DNA replication, this high binding affinity was greatly reduced to that of the interaction between linear double stranded Ter DNA and Tus (Kaplan and Bastia, 2009). It was therefore necessary to carefully test model 2 to determine its authenticity.

An Independent test of the melting-flipping model shows that it is unnecessary for polar fork arrest: We wished to rigorously test model 2, which postulated that DNA melting and base flipping together could explain polar fork arrest under a physiological salt concentration that permitted DNA replication to occur (Bastia et al., 2008). We reasoned that the model could be tested if one could temporally and spatially separate DNA unwinding by DnaB helicase from its ATP-dependent locomotion on DNA (double- or single-stranded). It is known that when encountering a linear DNA with a 5′ tail and 3′ blunt end, DnaB enters DNA with both strands passing through the central channel of DnaB (Kaplan, 2000). The translocation of DnaB on double-stranded DNA (dsDNA) requires ATP hydrolysis. We constructed the DNA substrate shown in Fig. 4. The DnaB helicase enters the substrate from the left by riding the 5′-single-stranded tail, slides over dsDNA containing a *Ter* site present in both orientations and upon reaching the forked structure with a 3′ overhang, DnaB unwinds this labeled strand (shown in blue). In the blocking orientation of Tus-*Ter* complex, the DnaB helicase slides on the dsDNA until it reached the *Ter* site, at which it is arrested, as shown by its failure to melt off the labeled 3′ tail shown in blue. In the reverse orientation of Tus-*Ter*, the DnaB sliding should displace Tus from *Ter* and continue sliding until it reached the 3′ overhang fork-like structure. At this point it should melt the labeled oligonucleotide, causing its release that can be resolved in a polyacrylamide gel at neutral pH and quantified (Fig.4). Our experiments showed that DnaB sliding, that involved no melting of DNA, not even a transient one, was arrested in a polar mode at a Tus-*Ter* complex. We proceeded to confirm the results further by introducing a pair of site-directed A-T inter-strand cross-links at two residues preceding C6. This covalent interstrand linkage prevented any chance of even transient DNA melting catalyzed by DnaB preceding the C6 residue. We confirmed that in such a substrate, DnaB sliding was arrested in a polar mode by the Tus-*Ter* complex only when present in the blocking orientation. These experiments led us to conclude that under physiological conditions a melting-flipping mechanism is not necessary (and probably does not occur) to cause polar fork arrest (Bastia et al., 2008).

Resolution of daughter DNA molecules at Ter sites: Following fork arrest at *Ter* sites, the daughter DNA molecules are resolved by a special type II topoisomerase, namely Topo IV (Espeli et al., 2003). It has been reported that this topoisomerase is stimulated by the actin-like MreB protein that acts near the resolution site *dif* that resolves dimers generated by recombination (Madabhushi and Marians, 2009).
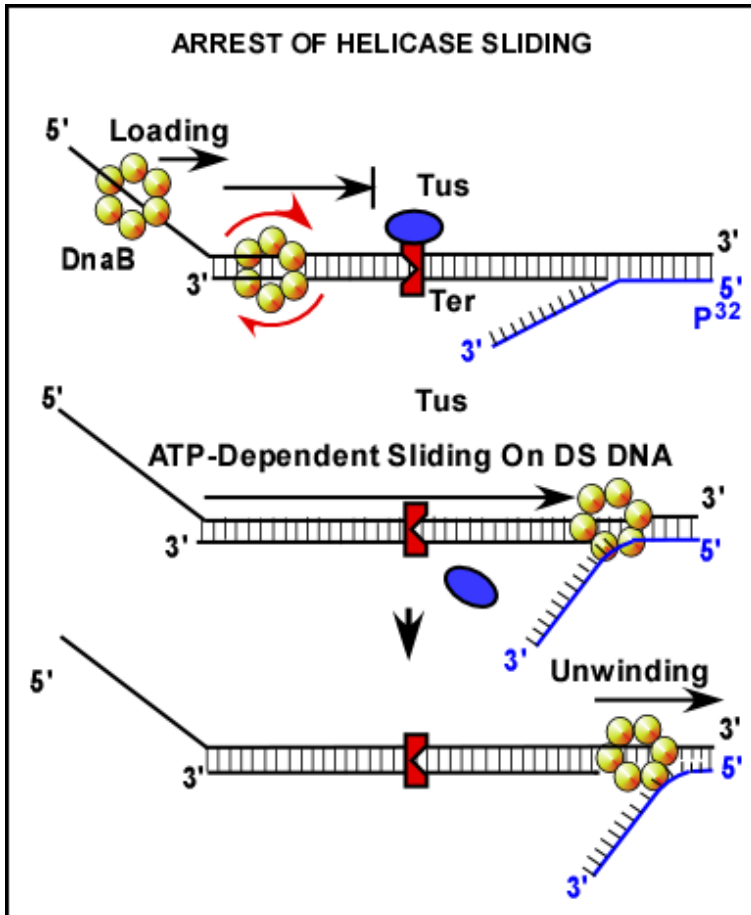
Fig. 4. A substrate designed to separate temporally and spatially DnaB translocation from DNA unwinding. A 5' tailed DNA with otherwise a blunt end on the complementary strand enters the substrate and then slides over the dsDNA until it meets the fork like structure (in blue) and unwinds the labeled strand. If a Tus-*Ter* complex is present in a blocking orientation, the sliding DnaB is arrested, thereby preventing the unwinding of the blue strand; a *Ter* site in the permissive orientation when bound to Tus displaces Tus and slides down the substrate and unwinds the blue strand. The results showed that DnaB sliding, without any DNA melting was arrested in a polar mode by the Tus-*Ter* complex, thereby showing that DNA unwinding (and presumably base flipping) is not necessary for polar helicase/ fork arrest.

Replication termini in eukaryotes: Many, perhaps all, eukaryotes have sequence-specific replication termini located in their ribosomal DNA (rDNA) array. For example, *Saccharomyces cerevisiae* contains a pair of Ter sites in one of the nontranscribed spacers of each rDNA unit between the sequences encoding the 35S RNA and the 5S RNA (Brewer and Fangman, 1988; Brewer et al., 1992; Ward et al., 2000). The second spacer contains a

replication *ori* (*ars*; see Fig.5). The Ter sites bind to the replication terminator protein called Fob1 (fork blockage) (Kobayashi, 2003; Kobayashi and Horiuchi, 1996; Mohanty and Bastia, 2004). The Fob1 protein bound to Ter sites prevents replication forks moving from right to left from colliding with the strong transcription of 35S RNA. It has been shown that transcription-replication collision causes not only fork stalling but also stalled RNA polymerase and an incomplete RNA transcript that can hybridize with DNA to form an R loop. R loops, especially the single stranded DNA therein, is susceptible to physical and enzymatic damage *in vivo* which causes genome instability (Helmrich et al., 2011).
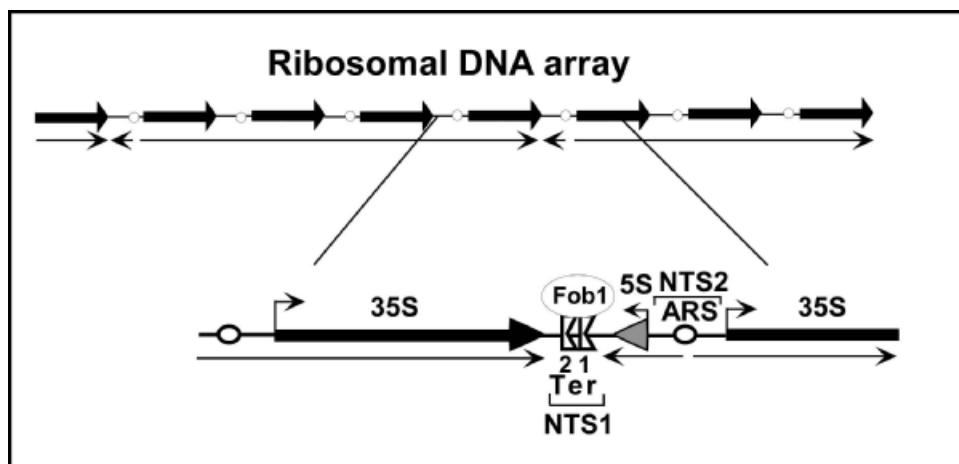
Fig. 5. rDNA repeat region in chromosome XII of *S. cerevisiae* showing the location of the two Ter sites in the nontranscribed spacer 1 (NTS1). The replication is initiated bidirectionally from the *ars* present in nontranscribed spacer 2 (NTS2). The Ter sites prevent replication forks moving to the left from the *ars* from running into RNA polymerase transcribing the 35S rRNA precursor.

The Fob1 protein is multifunctional and loads histone deacetylase to silence intra-chromatid recombination in the tandem array of ~200 rDNA repeats that might otherwise lead to unscheduled loss or gain of rDNA repeats (Bairwa et al., 2010; Huang et al., 2006; Huang and Moazed, 2003). Fob1 protein is also a transcriptional activator and controls exit from mitosis (Bastia and Mohanty, 2006; Stegmeier et al., 2004).

One of the facile techniques to study Fob1 function is to perform segment-directed mutagenesis, which is shown schematically (Fig.6). A segment of an ORF flanked by regions of homology (also from the ORF) is amplified by PCR under conditions of low fidelity synthesis in which one of the dNTPs is present at a suboptimal concentration. This leads to misincorporation of the base into DNA causing random mutations. A plasmid containing a gap corresponding to the segment being mutagenized and the PCR products are used to transform yeast. The mutagenized DNA segment gets incorporated into the plasmid by gap repair caused by the homologous recombination machinery of yeast with high efficiency, thus generating a pool of potential mutants contained in the plasmid. The plasmid contains a marker expressed in yeast (*e.g.*, *Leu*) and an *ars*. Using this protocol, we extensively mutagenized Fob1 and were able to identify many of its functional domains, such as its

DNA binding domain and a domain for its interaction with the silencing linker protein called Net1. Net1 recruits the histone deacetylase Sir2 onto Fob1 by direct protein-protein interaction between Net1 and Sir2 on one hand and between Net1 and Fob1 on the other, and loads Sir2 near the Ter sites. This process, as noted above, causes silencing of rDNA and prevents unwanted recombination (Bairwa et al., 2010; Mohanty and Bastia, 2004). At this time, the detailed mechanism of replication termination in eukaryotes has not been elucidated. However, it is known that two intra-S checkpoint proteins called Tof1 and its interacting partner called Csm3 are necessary for stable fork arrest at Ter because the Tof1-Csm3 complex protects the Fob1 protein from getting displaced from the Ter site by the action of the helicase Rrm3 (Mohanty et al., 2006, 2009). The catenated daughter molecules at Ter sites in *S. cerevisiae* are separated from each other by Topo II (Baxter and Diffley, 2008; Fachinetti et al., 2010).
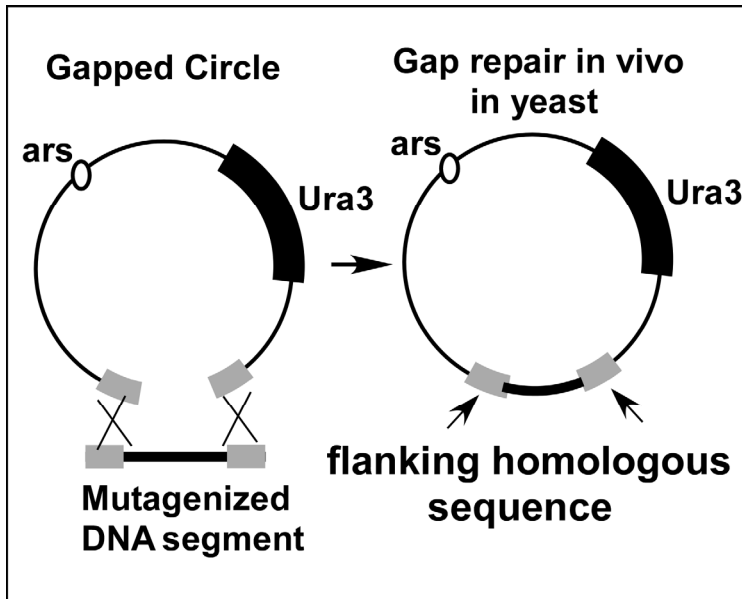
Fig. 6. Schematic diagram showing segment-directed mutagenesis and recovery of mutants by gap repair. The gapped plasmid is prepared by restriction site cutting inside the ORF. The DNA segment is mutagenized by low-fidelity PCR that includes primers with homologous flanking sequence. Transformation of a mixture of mutagenized DNA mixed with the gapped plasmid results in a pool of plasmids, some of which should have random base changes within the mutagenized DNA segment

We have recently reported that the Reb1 terminator protein binding to 2 Ter sites of fission yeast act in a cooperative fashion. The dimeric Reb1 protein, for example, brings into contact a Ter site located on chromosome 2 with two Ter sites located on chromosome 1. Interestingly there was no interaction observed between sites on chromosome 1 and 2 with the Ter sites located in the two rDNA clusters present on chromosome 3. It seems that the Ter-Ter interactions are not random. We further reported that the interactions called "chromosome kissing" modulated the activities of the Ter sites (Singh et al., 2010).

Physiological function of the replication termini: In prokaryotes, the replication termini perform at least 2 functions: (i) these serve as a replication trap and confine the meeting of the two approaching forks to the *TerC* region (Fig.1) where the dimer resolution (*dif*) sites are located. This activity probably facilitates chromosome segregation (Wake, 1997); and (ii) the terminus, in plasmid chromosomes prevents accidental switch to a rolling circle mode of replication that would generate unwanted linearly catenated chromosome (Dasgupta et al., 1991). In eukaryotes, the termini probably serve as barriers to transcription-replication collision that might generate destabilizing R loops. The termini are also known to be involved in cellular differentiation of fission yeast (Dalgaard and Klar, 2000, 2001). As noted above, Fob1 protein has diverse other functions (Bastia and Mohanty, 2006; Kaplan and Bastia, 2009).

In summary, replication termination at site-specific termini is an important part of DNA replication that invites further investigation, especially in eukaryotes, because of its role in various DNA transactions including maintenance of genome stability.

2. References

Abhyankar, M.M., Zzaman, S., and Bastia, D. (2003). Reconstitution of R6K DNA replication in vitro using 22 purified proteins. J Biol Chem *278*, 45476-45484.

Andersen, P.A., Griffiths, A.A., Duggin, I.G., and Wake, R.G. (2000). Functional specificity of the replication fork-arrest complexes of Bacillus subtilis and Escherichia coli: significant specificity for Tus-Ter functioning in E. coli. Mol Microbiol *36*, 1327-1335.

Bairwa, N.K., Zzaman, S., Mohanty, B.K., and Bastia, D. (2010). Replication fork arrest and rDNA silencing are two independent and separable functions of the replication terminator protein Fob1 of Saccharomyces cerevisiae. J Biol Chem *285*, 12612-12619.

Bastia, D., Germino, J., Crosa, J.H., and Ram, J. (1981). The nucleotide sequence surrounding the replication terminus of R6K. Proc Natl Acad Sci U S A *78*, 2095-2099.

Bastia, D., and Mohanty, B.K. (1996). Mechanisms for completing DNA replication. DNA Replication in Eukaryotic Cells (M DePamphilis, Ed) Cold Spring Harbor Laboratory Press, NY, 177-215.

Bastia, D., and Mohanty, B.K. (2006). Termination of DNA Replication. DNA replication and human disease (ed ML DePamphilis), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 155-174.

Bastia, D., Zzaman, S., Krings, G., Saxena, M., Peng, X., and Greenberg, M.M. (2008). Replication termination mechanism as revealed by Tus-mediated polar arrest of a sliding helicase. Proc Natl Acad Sci U S A *105*, 12831-12836.

Baxter, J., and Diffley, J.F. (2008). Topoisomerase II inactivation prevents the completion of DNA replication in budding yeast. Mol Cell *30*, 790-802.

Bell, S.P., and Dutta, A. (2002). DNA replication in eukaryotic cells. Annu Rev Biochem *71*, 333-374.

Brewer, B.J., and Fangman, W.L. (1987). The localization of replication origins on ARS plasmids in S. cerevisiae. Cell *51*, 463-471.

Brewer, B.J., and Fangman, W.L. (1988). A replication fork barrier at the 3' end of yeast ribosomal RNA genes. Cell *55*, 637-643.

Brewer, B.J., Lockshon, D., and Fangman, W.L. (1992). The arrest of replication forks in the rDNA of yeast occurs independently of transcription. Cell *71*, 267-276.

Bussiere, D.E., Bastia, D., and White, S.W. (1995). Crystal structure of the replication terminator protein from B. subtilis at 2.6 A. Cell *80*, 651-660.

Crosa, J.H., Luttropp, L.K., and Falkow, S. (1976). Mode of replication of the conjugative R-plasmid RSF1040 in Escherichia coli. J Bacteriol *126*, 454-466.

Cvetic, C., and Walter, J.C. (2005). Eukaryotic origins of DNA replication: could you please be more specific? Semin Cell Dev Biol *16*, 343-353.

Dalgaard, J.Z., and Klar, A.J. (2000). swi1 and swi3 perform imprinting, pausing, and termination of DNA replication in S. pombe. Cell *102*, 745-751.

Dalgaard, J.Z., and Klar, A.J. (2001). A DNA replication-arrest site RTS1 regulates imprinting by determining the direction of replication at mat1 in S. pombe. Genes Dev *15*, 2060-2068.

Dasgupta, S., Bernander, R., and Nordstrom, K. (1991). In vivo effect of the tus mutation on cell division in an Escherichia coli strain where chromosome replication is under the control of plasmid R1. Res Microbiol *142*, 177-180.

Duggan, L.J., Hill, T.M., Wu, S., Garrison, K., Zhang, X., and Gottlieb, P.A. (1995). Using modified nucleotides to map the DNA determinants of the Tus-TerB complex, the protein-DNA interaction associated with termination of replication in Escherichia coli. J Biol Chem *270*, 28049-28054.

Espeli, O., Levine, C., Hassing, H., and Marians, K.J. (2003). Temporal regulation of topoisomerase IV activity in E. coli. Mol Cell *11*, 189-201.

Fachinetti, D., Bermejo, R., Cocito, A., Minardi, S., Katou, Y., Kanoh, Y., Shirahige, K., Azvolinsky, A., Zakian, V.A., and Foiani, M. (2010). Replication termination at eukaryotic chromosomes is mediated by Top2 and occurs at genomic loci containing pausing elements. Mol Cell *39*, 595-605.

Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. Nature *340*, 245-246.

Germino, J., and Bastia, D. (1981). Termination of DNA replication in vitro at a sequence-specific replication terminus. Cell *23*, 681-687.

Hastings, A.F., Otting, G., Folmer, R.H., Duggin, I.G., Wake, R.G., Wilce, M.C., and Wilce, J.A. (2005). Interaction of the replication terminator protein of Bacillus subtilis with DNA probed by NMR spectroscopy. Biochem Biophys Res Commun *335*, 361-366.

Helmrich, A., Ballarino, M., and Tora, L. (2011). Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. Mol Cell *44*, 966-977.

Hidaka, M., Akiyama, M., and Horiuchi, T. (1988). A consensus sequence of three DNA replication terminus sites on the E. coli chromosome is highly homologous to the terR sites of the R6K plasmid. Cell *55*, 467-475.

Hill, T.M. (1992). Arrest of bacterial DNA replication. Annu Rev Microbiol *46*, 603-633.

Hill, T.M., Tecklenburg, M.L., Pelletier, A.J., and Kuempel, P.L. (1989). tus, the trans-acting gene required for termination of DNA replication in Escherichia coli, encodes a DNA-binding protein. Proc Natl Acad Sci U S A *86*, 1593-1597.

Huang, J., Brito, I.L., Villen, J., Gygi, S.P., Amon, A., and Moazed, D. (2006). Inhibition of homologous recombination by a cohesin-associated clamp complex recruited to the rDNA recombination enhancer. Genes Dev *20*, 2887-2901.

Huang, J., and Moazed, D. (2003). Association of the RENT complex with nontranscribed and coding regions of rDNA and a regional requirement for the replication fork block protein Fob1 in rDNA silencing. Genes Dev *17*, 2162-2176.

Kamada, K., Horiuchi, T., Ohsumi, K., Shimamoto, N., and Morikawa, K. (1996). Structure of a replication-terminator protein complexed with DNA. Nature *383*, 598-603.

Kaplan, D.L. (2000). The 3'-tail of a forked-duplex sterically determines whether one or two DNA strands pass through the central channel of a replication-fork helicase. J Mol Biol *301*, 285-299.

Kaplan, D.L., and Bastia, D. (2009). Mechanisms of polar arrest of a replication fork. Mol Microbiol *72*, 279-285.

Kaul, S., Mohanty, B.K., Sahoo, T., Patel, I., Khan, S.A., and Bastia, D. (1994). The replication terminator protein of the gram-positive bacterium Bacillus subtilis functions as a polar contrahelicase in gram-negative Escherichia coli. Proc Natl Acad Sci U S A *91*, 11143-11147.

Khatri, G.S., MacAllister, T., Sista, P.R., and Bastia, D. (1989). The replication terminator protein of E. coli is a DNA sequence-specific contra-helicase. Cell *59*, 667-674.

Kobayashi, T. (2003). The replication fork barrier site forms a unique structure with Fob1p and inhibits the replication fork. Mol Cell Biol *23*, 9178-9188.

Kobayashi, T., and Horiuchi, T. (1996). A yeast gene product, Fob1 protein, required for both replication fork blocking and recombinational hotspot activities. Genes Cells *1*, 465-474.

Kolter, R., and Helinski, D.R. (1978). Activity of the replication terminus of plasmid R6K in hybrid replicons in Escherichia coli. J Mol Biol *124*, 425-441.

Kuempel, P.L., Duerr, S.A., and Seeley, N.R. (1977). Terminus region of the chromosome in Escherichia coli inhibits replication forks. Proc Natl Acad Sci U S A *74*, 3927-3931.

Lee, E.H., Kornberg, A., Hidaka, M., Kobayashi, T., and Horiuchi, T. (1989). Escherichia coli replication termination protein impedes the action of helicases. Proc Natl Acad Sci U S A *86*, 9104-9108.

Madabhushi, R., and Marians, K.J. (2009). Actin homolog MreB affects chromosome segregation by regulating topoisomerase IV in Escherichia coli. Mol Cell *33*, 171-180.

Mohanty, B.K., Bairwa, N.K., and Bastia, D. (2006). The Tof1p-Csm3p protein complex counteracts the Rrm3p helicase to control replication termination of Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *103*, 897-902.

Mohanty, B.K., Bairwa, N.K., and Bastia, D. (2009). Contrasting Roles of Checkpoint Proteins as Recombination Modulators At Fob1-Ter Complexes With or Without Fork Arrest. Eukaryot Cell *8*, 487-495.

Mohanty, B.K., and Bastia, D. (2004). Binding of the replication terminator protein Fob1p to the Ter sites of yeast causes polar fork arrest. J Biol Chem *279*, 1932-1941.

Mohanty, B.K., Sahoo, T., and Bastia, D. (1996). The relationship between sequence-specific termination of DNA replication and transcription. EMBO J *15*, 2530-2539.

Mohanty, B.K., Sahoo, T., and Bastia, D. (1998). Mechanistic studies on the impact of transcription on sequence-specific termination of DNA replication and vice versa. J Biol Chem *273*, 3051-3059.

Mulcair, M.D., Schaffer, P. M., Oakly, A.J.Cross, H.F., Neylon, C., Hill, T.M. and Dixon, N. (2006) Cell, 125: 1309-1313.

Mulugu, S., Potnis, A., Shamsuzzaman, Taylor, J., Alexander, K., and Bastia, D. (2001). Mechanism of termination of DNA replication of Escherichia coli involves helicase-contrahelicase interaction. Proc Natl Acad Sci U S A *98*, 9569-9574.

Pelletier, A.J., Hill, T.M., and Kuempel, P.L. (1988). Location of sites that inhibit progression of replication forks in the terminus region of Escherichia coli. J Bacteriol *170*, 4293-4298.

Sahoo, T., Mohanty, B.K., Lobert, M., Manna, A.C., and Bastia, D. (1995). The contrahelicase activities of the replication terminator proteins of Escherichia coli and Bacillus subtilis are helicase-specific and impede both helicase translocation and authentic DNA unwinding. J Biol Chem *270*, 29138-29144.

Sernova, N.V., and Gelfand, M.S. (2008). Identification of replication origins in prokaryotic genomes. Brief Bioinform *9*, 376-391.

Sharma, R., Kachroo, A., and Bastia, D. (2001). Mechanistic aspects of DnaA-RepA interaction as revealed by yeast forward and reverse two-hybrid analysis. EMBO J *20*, 4577-4587.

Singh, S.K., Sabatinos, S., Forsburg, S., and Bastia, D. (2010). Regulation of replication termination by Reb1 protein-mediated action at a distance. Cell *142*, 868-878.

Sista, P.R., Hutchinson, C.A., 3rd, and Bastia, D. (1991). DNA-protein interaction at the replication termini of plasmid R6K. Genes Dev *5*, 74-82.

Sista, P.R., Mukherjee, S., Patel, P., Khatri, G.S., and Bastia, D. (1989). A host-encoded DNA-binding protein promotes termination of plasmid replication at a sequence-specific replication terminus. Proc Natl Acad Sci U S A *86*, 3026-3030.

Smith, M.T., and Wake, R.G. (1992). Definition and polarity of action of DNA replication terminators in Bacillus subtilis. J Mol Biol *227*, 648-657.

Stegmeier, F., Huang, J., Rahal, R., Zmolik, J., Moazed, D., and Amon, A. (2004). The replication fork block protein Fob1 functions as a negative regulator of the FEAR network. Curr Biol *14*, 467-480.

Wake, R.G. (1997). Replication fork arrest and termination of chromosome replication in Bacillus subtilis. FEMS Microbiol Lett *153*, 247-254.

Wang, J., and Sugden, B. (2005). Origins of bidirectional replication of Epstein-Barr virus: models for understanding mammalian origins of DNA synthesis. J Cell Biochem *94*, 247-256.

Ward, T.R., Hoang, M.L., Prusty, R., Lau, C.K., Keil, R.L., Fangman, W.L., and Brewer, B.J. (2000). Ribosomal DNA replication fork barrier and HOT1 recombination hot spot: shared sequences but independent activities. Mol Cell Biol *20*, 4948-4957.

Weinreich, M., Palacios DeBeer, M.A., and Fox, C.A. (2004). The activities of eukaryotic replication origins in chromatin. Biochim Biophys Acta *1677*, 142-157.

Wilce, J.A., Vivian, J.P., Hastings, A.F., Otting, G., Folmer, R.H., Duggin, I.G., Wake, R.G., and Wilce, M.C. (2001). Structure of the RTP-DNA complex and the mechanism of polar replication fork arrest. Nat Struct Biol *8*, 206-210.

Biochemical Analysis
of Halophilic Dehydrogenases Altered
by Site-Directed Mutagenesis

J. Esclapez, M. Camacho, C. Pire and M.J. Bonete
Departamento de Agroquímica y Bioquímica, División de Bioquímica y Biología Molecular,
Facultad de Ciencias, Universidad de Alicante, Alicante,
Spain

1. Introduction

Extremely halophilic Archaea are found in highly saline environments such as natural salt lakes, saltern pools, the Dead Sea and so on. These microorganisms require between 2.5 and 5.2 M NaCl for optimal growth. They can balance the external concentration by accumulating intracellular KCl to concentrations that can reach and exceed saturation. The biochemical machinery of these microorganisms has, therefore, been adapted in the course of evolution to be able to function at salt concentrations at which most biochemical systems will cease to function. The biochemical and biophysical properties of several halophilic enzymes have been studied in great detail; and, as a general rule, it was found that the halophilic enzymes are stabilized by multimolar concentration of salts. In most cases the salt also stimulates the catalytic activity. This stabilization of halophilic proteins in solvents containing high salt concentrations has been discussed in terms of apparent peculiarities in their composition. Since the first amino acid composition determinations, it has become clear that halophilic enzymes present a higher proportion of acidic over basic residues, an increase in small hydrophobic residues, a decrease in aliphatic residues and lower lysine content than their non-halophilic homologues (Lanyi, 1974; Eisenberg, et al., 1992; Madern et al., 2000). Since then, structural analyses have revealed two significant differences in the characteristics of the surface of the halophilic enzymes that may contribute to their stability in high salt. The first of these is that the excess of acidic residues are predominantly located on the enzyme surface leading to the formation of a hydration shell that protects the enzyme from aggregation in its highly saline environment. The second is that the surface also displays a significant reduction in exposed hydrophobic character, which arises not from a loss of surface exposed hydrophobic residues but from a reduction in surface-exposed lysine residues. Nevertheless, although the number of halophilic protein sequences has increased during the last years, the number of high resolution structures that permit the details of the protein solvent interactions to be seen is limited. The role of the reduction in the surface lysines has been largely ignored (Britton et al., 1998, 2006). Furthermore, in several studies, the authors have concluded that it is the precise structural organization of surface acidic residues that is important in halophilic adaptation. Not only is there an increase in acidic residue content, but these residues form clusters that bind networks of hydrated ions (Richard et al., 2000).

Halophilic archaea are considered a rather homogeneous group of heterotrophic microorganisms predominantly using amino acids as their source of carbon and energy. However, it has been shown that some halophilic archaea are able to use not only amino acids but different metabolites as well, as, for example, *Haloferax mediterranei,* which grows in a minimal medium containing glucose as the only source of carbon using a modified Entner-Doudoroff pathway (Rodriguez-Valera et al., 1983), or *Haloferax volcanii,* which is also able to grow in minimal medium with acetate as the sole carbon source (Kauri et al., 1990). Isocitrate lyase and malate synthase activities were detected in this organism when it was grown on a medium with acetate as the main carbon source (Serrano et al., 1998).

To understand the molecular basis of salt tolerance responsible for halophilic adaptation of proteins, to analyze the coenzyme specificity, and to study the mode of zinc-binding, we have chosen as model enzymes two halophilic dehydrogenase proteins involved in carbon catabolism. They are the glucose dehydrogenase (GlcDH) and isocitrate dehydrogenase (ICDH) from the extremely halophilic Archaea *Haloferax mediterranei* and *Haloferax volcanii*, respectively.

1.1 *Haloferax mediterranei* glucose dehydrogenase

GlcDH is the first enzyme of a non-phosphorylated Entner-Doudoroff pathway. It catalyses the reaction:

$$\text{Glucose} + \text{NAD(P)}^+ \rightarrow \text{Glucono-1,5-lactone} + \text{NAD(P)H} + \text{H}^+$$
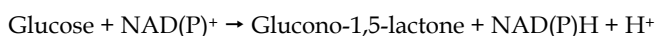
GlcDH from *Hfx. mediterranei* has been characterized and purified using gel filtration and affinity chromatography in the presence of buffers containing a high concentration of salt or glycerol to stabilize its structure. The protein is a dimeric enzyme with a molecular weight of 39 kDa per subunit, and shows a dual cofactor specificity, although it displays a marked preference for NADP^+ to NAD^+. Biochemical studies have established that the presence of a divalent ion such as Mg^{2+} or Mn^{2+} at concentrations of 25 mM enhances enzymatic activity (Bonete et al., 1996). Inactivation by metal chelators and reactivation by certain divalent ions indicated that glucose dehydrogenase from *Hfx. mediterranei* contains tightly bound metal ions that are essential for activity. Studies on the metal content of the enzyme by ICP revealed the presence of zinc ions whose removal by addition of EDTA leads to complete loss of enzyme activity (Pire et al., 2000). Sequence analysis showed that this enzyme belongs to the zinc-dependent medium-chain alcohol dehydrogenase superfamily (MDR), which includes sorbitol dehydrogenases, xylitol dehydrogenases and alcohol dehydrogenases (Pire et al., 2001). The structure of *Hfx. mediterranei* GlcDH has been solved at the highest resolution to date for any water-soluble halophilic enzyme. The structures of the apoenzyme and a D38C mutant in complex with NADP^+ and zinc reveal that the subunit, like that of the other MDR family members, is organized into two domains separated by a deep cleft, with the active site lying at its base. Domain 1 contains the residues involved in substrate binding, catalysis, and coordination of the active-site zinc. Domain 2 consists of a dinucleotide-binding Rossmann fold (Rossmann et al., 1974) that is responsible for binding NADP^+. Its molecular surface is predominantly covered by acidic residues, which are only partially neutralized by bound potassium counterions that also appear to play a role in substrate binding. The surface shows the expected reduction in hydrophobic character associated with the loss of lysines, which is consistent with the

genome-wide reduction of this residue in extreme halophiles. The structure also reveals a highly ordered, multilayered solvation shell that can be seen to be organized into one dominant network covering much of the exposed surface accessible area to an extent not seen in almost any other protein structure solved (Ferrer et al., 2001; Britton et al., 2006). Recently, high-resolution structures of a series of binary and ternary complexes of halophilic GlcDH have allowed an extension of the understanding of the catalytic mechanism in the MDR family. In contrast to the textbook MDR mechanism in which the zinc ion is proposed to remain stationary and attached to a common set of protein ligands, analysis of these structures reveals that in each complex, there are dramatic differences in the nature of the zinc ligation. These changes arise as a direct consequence of linked movements of the zinc ion, a zinc-bound bound water molecule, and the substrate during progression through the reaction. These results provide evidence for the molecular basis of proton traffic during catalysis, a structural explanation for pentacoordinate zinc ion intermediates, and a unifying view for the observed patterns of metal ligation in the MDR family (Esclapez et al., 2005; Baker et al., 2009).

1.2 *Haloferax volcanii* isocitrate dehydrogenase

The citric acid cycle enzyme, ICDH (EC 1.1.1.41 and EC 1.1.1.42), catalyses the oxidative decarboxylation of isocitrate (Kay & Weitzman, 1987):

$$\text{Isocitrate} + \text{NAD(P)}^+ \rightarrow \text{2-oxoglutarate} + \text{NAD(P)H} + \text{H}^+ + \text{CO}_2$$
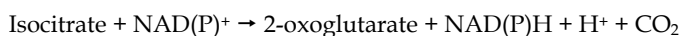
The wild-type enzyme from *Haloferax volcanii* was purified using three steps. The enzyme has been characterized, and it is a dimer with subunit M_r of 62000 Da. Its activity is strictly NADP dependent, and markedly dependent on the concentration of NaCl or KCl, being maximal in 0.5 M NaCl or KCl. The thermostability of the archaeal isocitrate dehydrogenase was investigated incubating the enzyme in buffer containing either 0.5 M or 3 M KCl. Clearly, the thermal stability of the enzyme is substantially reduced at the lower KC1 concentration, with concomitant differences in the activation energies for the thermal inactivation process, 360 kJ mol^{-1} and 610 kJ mol^{-1} at 0.5 M and 3 M KCl, respectively; therefore, the high *in vivo* KC1 concentrations appear to be more important for the stability of the enzyme than for its catalytic ability (Camacho et al., 1995). The gene encoding this protein was sequenced and the derived amino acids were determined. The yields of *Escherichia coli*-expressed enzyme were greater than those obtained by purification of the enzyme from the native organism, but the product was insoluble inclusion bodies. The recombinant ICDH behaves similarly to the native enzyme with respect to the dependence of activity on salt concentration. Kinetic analysis has also shown the purified recombinant and native enzymes to be similar, as are the thermal stabilities (Camacho et al., 2002). *Hfx. volcanii* ICDH dissociation/deactivation has been measured to probe the respective effect of anions and cations on stability. Surprisingly, enzyme stability has been found to be mainly sensitive to cations and very little (or not) to anions. Divalent cations have induced a strong shift of the active/inactive transition towards low salt concentration. A high resistance of ICDH from *Hfx. volcanii* to chemical denaturation has also been found. This study strongly suggests that *Hfx. volcanii* ICDH might be seen as a type of halophilic protein never described before: an oligomeric halophilic protein devoid of intersubunit anion-binding sites (Madern et al., 2004).

2. Materials and methods

2.1 Strains, culture conditions and vectors

Escherichia coli NovaBlue (Novagen) was used as host for plasmids pGEM-11Zf(+) and pET3a. *E. coli* BMH71-18 *mutS* (Promega) and *E. coli* XL1-Blue (Stratagene) were employed in site-directed mutagenesis experiments. *E. coli* BL21(DE3) (Novagen) was used as the expression host. *E. coli* strains were grown in Luria-Bertani medium at 37 °C with shaking at 180 rpm. Plasmids were selected for in solid and liquid media by the addition of 100 μg ampicillin/ml.

Vector pGEM-11Zf(+) (Promega) was used for cloning genes and carrying out some site-directed mutagenesis experiments. The expression vector pET3a was purchased from Novagen.

2.2 Site-directed mutagenesis

Site-directed mutations were introduced into genes cloned in pGEM-11Zf(+) or directly into pET3a expression vector. The synthetic oligonucleotide primers (Applied Biosystems and Bonsai Technology) were designed to contain the desired mutation. Mutant construction was carried out by two different methods. In the first, the gene encoding the halophilic dehydrogenases were cloned into pGEM-11Zf(+) and site-directed mutagenesis was performed using the GeneEditorTM *in vitro* Site-Directed Mutagenesis System (Promega). This method works by the simultaneous annealing of two oligonucleotide primers to one strand of a denaturated plasmid. One primer introduces the desired mutation in the gene; and the other primer mutates the beta-lactamase gene, increasing the resistance to alternate antibiotics as penicillins and cephalosporines. The last change is important to select plasmids derived from the mutant strand. This positive selection results in consistently high mutagenesis efficiencies. The protocols supplied with the kit consist in the annealing of the two oligonucleotide primers to an alkaline-denatured dsDNA template. Following hybridization, the oligonucleotides are extended with DNA polymerase to create a double-stranded structure. The nicks are then sealed with DNA ligase and the duplex structure is used to transform an *E. coli* host. The construction of the mutants was carried out following the Promega protocol but with one modification: the length of the DNA denaturation stage was increased from 5 min at room temperature to 20 min at 37 °C due to the increase of the GC content in the halophilic genomes. In the second, the mutagenesis procedure used followed the method of the Stratagene Quick Change kit, using *Pfu Turbo* DNA polymerase from Stratagene. Extension of the oligonucleotide primers generated a mutated plasmid containing staggered nicks. Following temperature cycling, the product was treated with *Dpn* I (Fermentas). The *Dpn* I endonuclease is specific for methylated and hemimethylated DNA and was used to digest the parental DNA template and to select for mutation containing synthesized DNA. The nicked vector DNA containing the desired mutations was transformed into XL1-Blue competent cells (CNB Fermentation Service). In both methods, putative mutants were screened by dideoxynucleotide sequencing with ABI3100 DNA sequencer (Applied Biosystems).

2.3 Protein preparation

Expression *E. coli* BL21(DE3) cells were transformed with the mutated plasmid. Expression, renaturation and purification of recombinant mutants were as previously described for wild

type halophilic enzymes (Pire et al., 2001; Camacho et al., 2002). The purity of the proteins was checked by SDS–polyacrylamide gel electrophoresis (SDS-PAGE). No protein contamination was detectable after Coomassie-blue staining of the gel. Protein concentration was determined by the method of Bradford (Bradford, 1976).

2.4 Glucose dehydrogenase analysis

2.4.1 Kinetic assays and data processing

Initial velocity studies were performed in 20 mM Tris–HCl buffer pH 8.8, containing 2 M NaCl and 25 mM $MgCl_2$. The reaction was monitored by measuring the appearance of NAD(P)H at 340 nm with a Jasco V-530 spectrophotometer. One unit of enzyme activity was defined as the amount of enzyme required to produce 1 μmol NAD(P)H/min under the assay conditions (40 ^{o}C).

The kinetic constants were obtained from at least triplicate measurements of the initial rates at varying concentrations of D-glucose and NAD(P)$^{+}$. Kinetic data were fitted to the sequential ordered BiBi equation with the program SigmaPlot 9.0.

2.4.2 Effect of EDTA concentration

The samples at different NaCl concentration were incubated with increasing EDTA concentration for 5 min at room temperature. After the incubation, the residual activities of the enzymes were measured in the activity buffer defined previously (Bonete et al., 1996).

2.4.3 Effect of temperature on enzymatic stability and activity

The samples at different NaCl concentration were incubated at various temperatures: 55, 60, 65, 70 and 80 ^{o}C. Aliquots were withdrawn at given times for measurement of residual activity. Furthermore, enzymatic activity was assayed between 25 and 75 ^{o}C at the same conditions described previously.

2.4.4 Effect of salt concentration on enzymatic activity and stability

The enzymatic activity was measured, as previously described, in buffer with KCl or NaCl in the concentration range of 0-4 M. The results are expressed as the percentage of the activity relative to the highest activity obtained.

Salt concentration stability studies were carried out at room temperature and at 40 ^{o}C. Purified preparations of enzyme in 2 M KCl were quickly diluted with 50 mM potassium phosphate buffer pH 7.3 to obtain 0.25 and 0.5 M KCl concentrations. Samples were removed at known time intervals, cooled on ice, and the residual enzymatic activity was then measured. The results are expressed as the percentage of the activity relative to that existing before incubation.

2.4.5 Differential scanning calorimetry (DSC)

DSC experiments were performed using a VP-DSC microcalorimeter (MicroCal). Temperatures from 40 ^{o}C to 90 ^{o}C were scanned at a rate of 60 ^{o}C/h using 50 mM potassium phosphate buffer pH 7.3 containing 1 mM EDTA and 0.5 M or 2.0 M KCl, which also served

for baseline measurements. Prior to scanning, all samples of protein and buffer were degassed under vacuum using a ThermoVac unit (MicroCal). The protein concentrations were in the range of 50–80 μM (approximately 4–6 mg/ml). The data were analyzed using ORIGIN software v 7.0.

2.5 Isocitrate dehydrogenase analysis

2.5.1 Sequence alignment

Initial alignment with *Hfx. volcanii* NADP-dependent ICDH (Q8X277) was obtained with ClustalW (Thompson et al., 1994), taking account of information of *Bacillus subtilis* (P39126) (Singh et al., 2001) and *E. coli* (P08200) (Hurley et al., 1991) NADP-dependent ICDH, and *Thermus thermophilus* NAD-dependent IMDH (P00351) (Imada et al., 1991) and their sequences. The crystalline structures of all of them were solved previously by high-resolution X-ray analysis. Residues critical to substrate binding were identified from high-resolution crystallographic structures of *E. coli* NADP–ICDH with bound isocitrate (Hurley et al., 1991). Critical residues for coenzyme specificity were identified from high-resolution X-ray crystallographic structures of *E. coli* ICDH complexed with $NADP^+$ (Hurley et al., 1991) and *T. thermophilus* IMDH complexed with NAD^+ (Hurley & Dean, 1994).

Oligonucleotide primers containing the necessary mismatches were used for construction of the mutations: R291S, K343D, Y344I, V350A and Y390P.

2.5.2 Kinetic assays and data processing

The activities of native and mutant ICDHs were determined spectrophotometrically at A_{340} and 30 °C in 20 mM Tris-HCl buffer pH 8.0, 1 mM EDTA, 10 mM $MgCl_2$ (Tris/EDTA/Mg^{2+}) containing 2 M NaCl, 1 mM D,L-isocitrate (Camacho et al., 1995, 2002), with $NADP^+$ or NAD^+ as the coenzyme. One unit of enzyme activity is the reduction of 1 μmol of NADP per min. Initial velocities were determined by monitoring the production of NADPH or NADH at 340 nm in a 1-cm light path, based on a molar extinction coefficient of 6200 M^{-1} cm^{-1}. Kinetic parameters K_m and V_{max} were calculated for the $NADP^+$ and NAD^+ and isocitrate, depending on the cases, and the turnover number (K_{cat}) and catalytic efficiency (K_{cat}/K_m) were determined for each of the mutants, by fitting the data to the Eadie–Hofstee equation with the SigmaPlot program (Version 1.02, Jandel Scientific, Erkath, Germany) (Rodriguez-Arnedo et al., 2005).

2.5.3 Modeling ICDH

Native ICDH and the mutant ICDH with all five amino acids substituted (SDIAP mutant) were modeled with the Swiss-Model program on ExPASy Molecular Biology Server (http://swissmodel.expasy. org/) based on sequence homology. The program uses Blast and ExNRL-3D (derived from PDB) database for the search of a potential protein mold. These proteins, previously resolved by X-ray analysis, with more than 20 amino acids in length and more than 25% sequence identity were chosen. The construction of the structural model was done with the Promodll program and the minimization of energy with Gromos96. The program calculates all levels of identity between the sample problem and the sequence pattern, and it calculates the relative standard deviation to the average of the

corresponding structures models and control. *Hfx. volcanii* ICDH shares 56.6% identity with *E. coli* ICDH (Camacho et al., 2002). The final image was refined with Swiss-Pdb Viewer (Rodriguez-Arnedo et al., 2005).

3. Results

3.1 Analysis of acidic surface of *Hfx. mediterranei* GlcDH

3.1.1 Choice of the halophilic GlcDH mutations

Generally, halophilic enzymes present a characteristic amino acid composition, showing an increase in the content of acidic residues and a decrease in the content of basic residues, particularly lysines. The latter decrease appears to be responsible for a reduction in the proportion of solvent-exposed hydrophobic surface. This role was investigated by site-directed mutagenesis of GlcDH from *Hfx. mediterranei*, in which three surface aspartic residues of the 27 per subunit were changed to lysine residues. At the start of the project, an initial GlcDH structure had been solved at medium resolution. Based on direct observation of this structure, the three aspartic residues chosen were D172, D216 and D344, which at least have the carboxyl oxygens exposed to the solvent (Fig. 1).
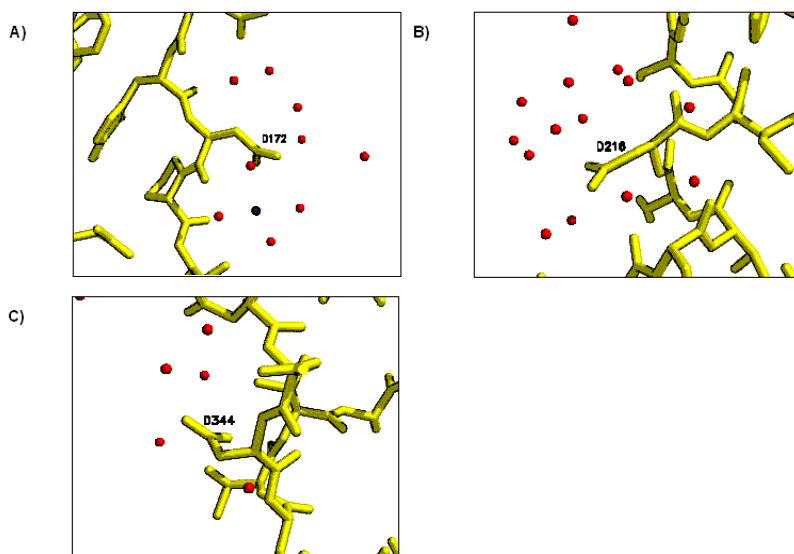
Fig. 1. Diagram showing details of the region surrounding residues D172 (A), D216 (B) and D344 (C) in the high resolution structure of D38C GlcDH with NADP+ and zinc (Britton et al., 2006). The water molecules and the potassium ions are shown in red and black, respectively.

The three selected residues are considered as surface acidic residues, and they are located in different regions of the protein surface. Later, the 1.6 Å resolution GlcDH structure revealed that the side-chain carboxyl of D172 is involved in interactions with a cluster of surface water molecules near a bound potassium counter-ion. In contrast, the side-chain carboxyl of D216 forms interactions with surface waters in a region in which no counter-ions can be

seen. The side-chain carboxyl of D344 lies on the surface, where it interacts with the solvent but also makes hydrogen bonds to the nearby side-chains of T346 and T347. Moreover, multiple alignments (data not shown) with other GlcDH sequences belonging to the MDR superfamily have shown that the acidic residue D216 from *Hfx. mediterranei* GlcDH is conserved in all other halophilic microorganisms. However, residue D344 is only conserved in the *Hfx. volcanii* GlcDH; and residue D172 is not present in any halophilic GlcDH. At the locations corresponding to D172, D216 and D344 in wild type *Hfx. mediterranei* GlcDH, there are non-acidic residues in 100% of the non-halophilic GlcDH sequences analyzed. Therefore, the presence of these acidic residues in the GlcDH from *Hfx. mediterranei* could be an adaptive response to the halophilic environment (Esclapez et al., 2007).

3.1.2 Site-directed mutagenesis and expression of the mutant proteins

Four mutant enzymes were obtained, the triple mutant and the three corresponding single mutant. The triple mutant GlcDH was created with the GeneEditorTM *in vitro* Site-Directed Mutagenesis System (Promega) by introducing the mutations one by one. The single mutant D172K was achieved as the first step in the constructions of the triple mutant GlcDH. The mutants D216K and D344K were constructed by PCR using *Pfu Turbo* DNA polymerase and following digestion with the endonuclease *Dpn*I.

The four mutant genes were cloned into the pET3a expression vector, and the resulting constructs were transformed into *E. coli* BL21(DE3). The expression assays were performed as described previously (Pire et al., 2001). The four mutant proteins were obtained as inclusion bodies, which were solubilized using 20 mM Tris–HCl buffer pH 8.0, 8 M Urea, 50 mM DTT and 2 mM EDTA, like wild type GlcDH. The refolding of each mutant protein was achieved by rapid dilution in 20 mM Tris–HCl buffer pH 7.4, 1 mM EDTA and KCl or NaCl in the concentration range of 1–3 M. The wild-type and triple mutant GlcDHs behave identically in the refolding process under the conditions assayed. The profiles for the triple mutant protein are like the wild type GlcDH, independently of concentration and type of salt. The three single mutants also presented the same profiles. In the presence of NaCl, the recovery of activity was always higher than with KCl; and the highest enzymatic activity was obtained at 3 M NaCl. Furthermore, at low salt concentrations the recovery of activities were lower than at high salt concentrations. No activity was recovered at 1 M KCl or NaCl. Thus, the mutations introduced on the protein surface did not appear to affect refolding in either the triple mutant or the single mutant proteins.

The purification of the GlcDH mutants were carried out as described previously. However, after 3–4 days, protein precipitation was observed in the fractions of triple mutant GlcDH whose protein concentration was greater than 1 mg/ml. This problem was solved by decreasing the protein concentration or by reducing the salt concentration through dialysis against the buffer containing 1 M NaCl or KCl. This fact indicates that the halophilic properties of the triple mutant protein have been altered, since the wild-type and single mutant proteins were stable for months under these conditions.

3.1.3 Properties of the mutant enzymes

The kinetic parameters of the mutant proteins were determined and compared to those that had previously been obtained with wild-type GlcDH. Their K_m values for $NADP^+$ and glucose are essentially similar and no significant differences in the values for V_{max} were

detected. These results indicated that the kinetic parameters were not affected by the mutations. It is unlikely, therefore, that the mutations in position 172, 216 and 344 influenced the active site or the integrity of the enzyme. Similar results were obtained when residues on the surface were mutated on malate dehydrogenase from *Haloarcula marismortui* (Madern et al., 1995) and dihydrolipoamide dehydrogenase from *Hfx. volcanii* (Jolley et al., 1997).

The dependence of enzymatic activity on the concentration of NaCl is shown in Fig. 2. The triple mutant GlcDH shows its maximum activity in a buffer with 0.50–0.75 M NaCl while the wild-type protein has its maximum activity with 1.5 M NaCl. Furthermore, at low salt concentrations the activity of the triple mutant enzyme is higher than the activity of the wild-type GlcDH. At higher salt concentrations, it is lower than the wild-type protein. With the purpose of determining if the observed behavior in the triple mutant protein is due to the presence of just one mutation or of the three modifications, these experiments were also performed with each single mutant protein. The mutants D172K GlcDH and D216K GlcDH show the same profiles as the triple mutant enzyme. In striking contrast, the behavior of the D344K mutant protein is very similar to the profile obtained with the wild-type GlcDH. These results suggest that the D344K modification does not disturb the halophilic characteristics of GlcDH. Therefore, the behavior of the triple mutant GlcDH in the salt concentrations assayed could be due to the introduction of the mutation D172K and D216K. The profiles obtained using buffers with KCl are very similar.

At optimal salt concentration, the activities of the wild-type and mutant GlcDH proteins are very close. The kinetic parameters are very similar too. Therefore, it appears that the different mutations introduced in GlcDH only influence the dependence of enzymatic activity on the salt concentration. However, in similar studies with the dihydrolipoamide dehydrogenase from *Hfx. volcanii*, mutants with only one mutation (E243Q, E423S or E423A) resulted in enzymes less active than the wild-type enzyme and with different kinetic parameters. Based on these results, Jolley and co-workers (Jolley et al., 1997) also supported the view that it is the precise structural organization of acidic residues that is important in halophilic adaptation and not only the increase in acidic residue content (Madern et al., 1995; Irimia et al., 2003).
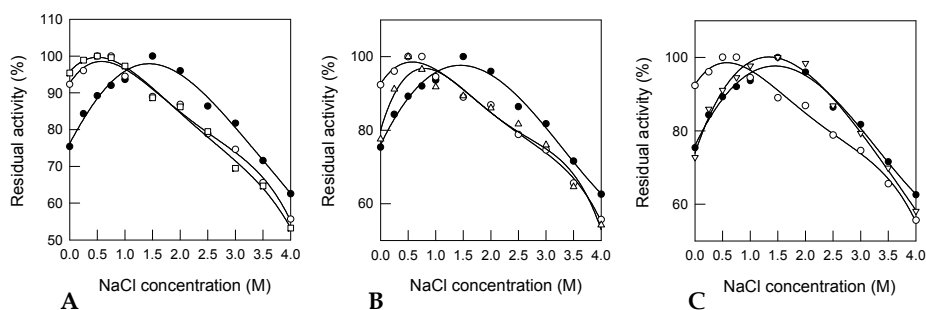
Fig. 2. Effect of NaCl on the activity of wild-type GlcDH (•), triple mutant GlcDH (○) and single mutants: (A) D172K GlcDH (□), (B) D216K GlcDH (△) and (C) D344 K GlcDH (▽). The activity buffer was 20 mM Tris–HCl pH 8.8 with varying concentrations of NaCl.

The effects of different salt concentrations on the residual activity of wild-type halophilic GlcDH and the four mutant proteins were measured after incubation at 25 °C and 40 °C. In the presence of 2 M KCl, neither wild-type enzyme nor mutant proteins were inactivated at the temperatures assayed. In particular, at salt concentrations above 1 M, the proteins were stable for weeks. As salt concentration increases, the proteins were more stable independent of the temperature. However, at low salt concentrations, small differences were observed in the stability of the proteins. The triple mutant and each single mutant protein appeared to be slightly more stable than the wild-type protein at 0.25 and 0.50 M KCl. The behavior of the proteins at 25 °C was similar, although a decrease in the temperature implies an increase of the period over which the enzymes are stable. The half-life time ($t_{1/2}$) for each protein was calculated (Table 1) showing that the mutant protein half-life times, either as a single alteration or altogether, are longer than wild type, both at 25 °C and 40 °C. However, there are no significant differences between the triple mutant and the single mutant proteins. All showed similar half-life times under the conditions assayed.

Biocalorimetry experiments were carried out under two different KCl concentrations using a DSC. In the presence of 2 M KCl, wild-type and single mutant GlcDH denaturing temperatures range from 74.6 °C to 75.9 °C. However, the triple mutant enzyme shows a lower denaturing temperature, between 73.6 °C and 73.7 °C. In other words, the triple mutant enzyme is denatured at slightly lower temperatures than are the wild-type and single mutant GlcDHs in the presence of high salt. At 0.50 M KCl (low salt), the results obtained do not reveal significant data; but the protein denaturing temperatures are lower than those obtained in the presence of high salt, independent of protein type (Fig. 3). This decrease was expected because the halophilic proteins are destabilized in low salt. Consequently the denaturing temperatures of the wild-type and mutant enzymes ranged from 59.8 °C to 60.7 °C. There were no significant differences between the temperatures.

	"Wild type"	Triple mutant	D172K	D216K	D344K
$t_{1/2}$ **40 °C (h)**					
0.25 M KCl	14 ± 2	18 ± 3	23 ± 5	26 ± 8	25 ± 4
0.50 M KCl	86 ± 4	114 ± 9	95 ± 9	117 ± 8	114 ± 7
>1.0 M KCl	>170	>170	>170	>170	>170
$t_{1/2}$ **25 °C (h)**					
0.25 M KCl	142 ± 22	246 ± 23	293 ± 30	219 ± 32	232 ± 30
0.50 M KCl	506 ± 50	613 ± 74	630 ± 113	660 ± 113	537 ± 75
>1.0 M KCl	>170	>170	>170	>170	>170

Table 1. Half-life times of "wild type" and mutant GlcDHs in the presence of different KCl concentrations at 40 °C (A) and 25 °C (B).
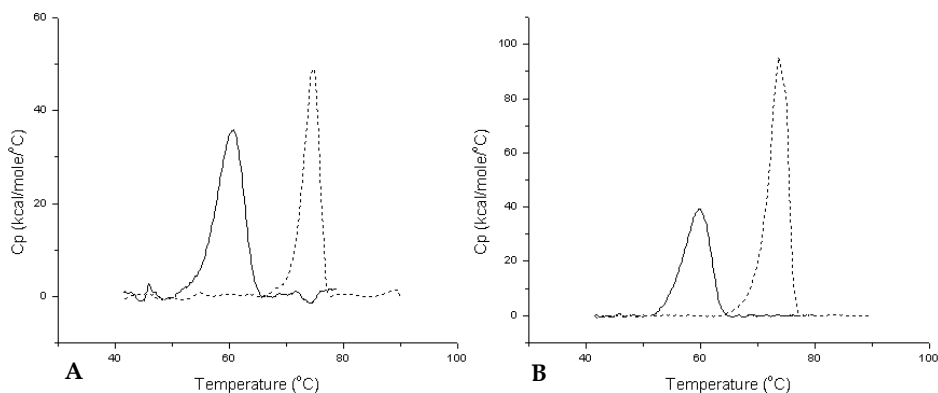
Fig. 3. Calorimetric traces of the thermal transition for wild-type GlcDH (A) and triple mutant GlcDH (B). Thermal transitions were determined in 50 mM potassium phosphate buffer pH 7.3 with 0.5 M (continuous line) or 2 M KCl (dotted line).

The data that we have presented indicate that the halophilic properties of the mutant proteins have been modified. Their enzymatic activity and kinetic parameters have been not affected by the mutations. The triple mutant and the single mutants, D172K GlcDH and D216K GlcDH, have reached their maximum activities at lower salt concentrations than wild-type GlcDH and the D344K mutant. It appears that the D344K substitution has no effect on the salt activity profile. Strikingly, in all the cases the mutant proteins were slightly more stable at low salt concentrations than was the wild-type GlcDH , although they require high salt concentration for maximum stability, like a malate dehydrogenase mutant from *Har. marismortui* (Madern et al., 1995). The biocalorimetry analyses have revealed another difference. The single mutant and the wild-type GlcDHs showed similar denaturing temperatures in the presence of 2 M KCl, while the triple mutant enzyme presented a lower denaturing temperature. Thus, more than one of our substitutions are apparently needed to significantly modify the protein's denaturing temperature at high salt concentration. Probably, these data are the result of an alteration of the hydration shell, which is required for halophilic proteins to be stable at high salt concentrations. Analysis of the high resolution GlcDH structure has shown that the size and order of the hydration shell in the halophilic enzyme is significantly greater than in non-halophilic proteins. Analyses also show that the differences in the characteristics of the molecular surface arise not only from an increase in negative surface charge, but also from the reduction in the percentage of hydrophobic surface area due to lysine side chains. Lysine residues of halophilic enzymes tend to be more buried than those of non-halophilic proteins (Britton et al., 2006).

3.2 Analysis of the zinc-binding site of GlcDH from *Hfx. mediterranei*

3.2.1 Choice of the GlcDH mutations

Whilst sequence analysis clearly identifies *Hfx. mediterranei* GlcDH as belonging to the zinc-dependent medium chain dehydrogenase/reductase family, the zinc-binding properties of the enzymes of this family are known to vary. The family includes numerous

zinc-containing dehydrogenases, which bind one or two zinc atoms per subunit. One of the zinc atoms is essential for catalytic activity, while the other has a structural role and is not present in all the family members. Previous biochemical studies established that the *Hfx. mediterranei* GlcDH appears to have a single zinc atom per subunit. The role of this zinc atom is to participate in the catalytic function of the enzyme (Pire et al., 2000; Pire et al., 2001).

In the crystal structure of horse liver alcohol dehydrogenase (HLADH), three protein ligands, C46, H67 and C174 coordinate the catalytic zinc (Eklund et al., 1981). Residues analogous to C46 and H67 are conserved in the vast majority of members of the MDR family, while in some enzymes the analogous residue for C174 is glutamate as in *Thermoplasma acidophilum* GlcDH (Fig. 4). On the basis of sequence alignment, the residues involved in binding the catalytic zinc in *Hfx. mediterranei* GlcDH are predicted to be D38, H63 and E150. This sequence pattern of residues that bind the catalytic zinc has not previously been observed for any enzyme in the MDR family. The change of the C38 to D38 in the halophilic enzyme could be an adaptive response to the halophilic environment.

In order to investigate the mode of zinc binding to the halophilic GlcDH, two mutant enzymes were constructed by site-directed mutagenesis. We replaced the D38 present in the active center of the protein with C or A.

3.2.2 Site-directed mutagenesis and expression of the mutant proteins

Site-directed mutagenesis was carried out to replace the D38 residue by cysteine and alanine in the recombinant GlcDH using GeneEditorTM *in vitro* Site-Directed Mutagenesis System. The mutant genes were cloned into the pET3a expression vector. The resulting constructs were introduced by transformation into *E. coli* BL21(DE3). After expression, both mutant proteins were obtained as inclusion bodies, as was wild-type GlcDH.
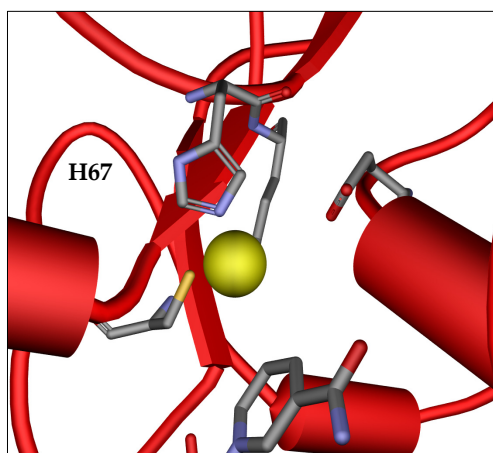
Fig. 4. The catalytic zinc-binding site in *Thermoplasma acidophilum* GlcDH.

The mutant enzymes were refolded and purified as described previously (Pire et al., 2001). In both mutants, the activity was lower than that of the wild-type protein, with the D38A mutant being inactive. This result suggests that D38 is an important residue and that the mutation to A38 leaves the enzyme seriously compromised. With respect to the D38C mutant, the maximum activity observed was approximately 30% of the activity of the wild-type enzyme.

3.2.3 Characterization of D38C GlcDH

The kinetic parameter values for mutant D38C GlcDH were determined and compared with those obtained for wild-type GlcDH (Table 2). K_{mNADP^+} differences are not significant; however, the mutation led to a significant increase of the K_m for glucose. Moreover, as the K_{cat} and $K_{cat}/K_{mglucose}$ parameters show, the catalytic efficiency of the mutant protein is less than the catalytic efficiency of wild-type GlcDH. These results indicate that the replacement of D38 to C38 in the GlcDH probably affects not only the catalytic zinc-binding site, but also the active site of the protein. The C38 GlcDH decreases the enzyme's affinity for glucose and its V_{max} relative to the wild-type enzyme. Consequently, the catalytic efficiency of the mutant enzyme is reduced.

	K_{mNADP^+} (mM)	$K_{mglucose}$ (mM)	V_{max} (U/mg)	K_{cat} (min^{-1})	K_{cat}/K_m (mM^{-1}min^{-1})
"Wild type"	0.035 ± 0.004	2.8 ± 0.3	397 ± 15	31 ± 1	11.10 ± 1.6
D38C	0.044 ± 0.010	12.4 ± 2.3	83 ± 9	7 ± 1	0.52 ± 0.14

Table 2. Kinetic parameters of recombinant wild-type GlcDH and the D38C mutant.

The zinc ion in the wild-type enzyme can be removed by EDTA treatment to yield an inactive enzyme (Pire et al., 2000). In order to compare the strength of zinc binding in "wild type" and in the D38C mutant, a similar treatment was carried out. Fig. 5 shows that zinc is more weakly bound in the D38C mutant than in the wild-type enzyme. The EDTA concentration needed to inactivate the enzyme is lower than that needed for the wild-type enzyme, and this inactivation was independent of salt concentration. For the wild-type enzyme, the capacity of EDTA to sequester the zinc is lower in the D38C mutant; and it is salt concentration-dependent. In the three NaCl concentration tested, the enzyme lost approximately 80% of its activity in the presence of 0.25 mM EDTA, and it was completely inactive at concentrations higher than 2 mM. However, in the case of the wild-type GlcDH, the EDTA necessary to sequester zinc atom at 3 M NaCl is higher than at 1 M, so the behavior of this protein is dependent on the salt concentration. At concentrations above 4 mM of the chelating agent, the enzyme is completely inactive, regardless of the NaCl concentration. Therefore, the substitution of D38 by C38 in the protein has weakened the binding of zinc ion. The D residue at position 38 in the halophilic glucose dehydrogenase instead of C, which is commonly found at the analagous postion in other members of the medium chain dehydrogenase family, could represent a halophilic adaptation.
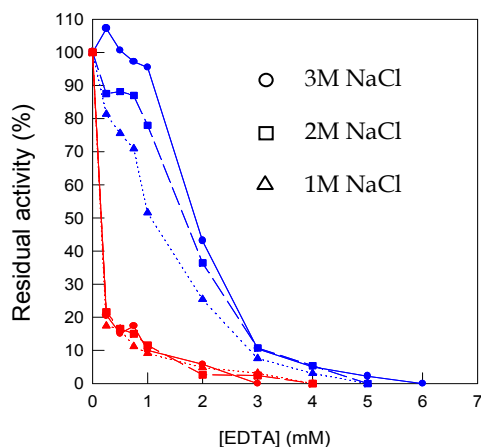
Fig. 5. Deactivation of the wild-type (blue) and D38C mutant (red) GlcDH under various EDTA concentrations at different buffer salt concentrations.

The replacement of D38 by C38 makes the binding of catalytic zinc ion of the halophilic GlcDH very similar to that presented by the thermophilic GlcDHs and other MDR family proteins. In order to clarify if C38 instead of D38 modifies the thermal characteristics of the enzyme at different salt concentrations, the effect of the temperature on enzymatic stability and activity were determined.

Generally at low salt concentration, halophilic proteins are less stable. High temperatures can contribute to their destabilization under these conditions. At high salt concentrations, halophilic proteins are stable; but stability can be perturbed by several factors, such as high temperatures. The thermal inactivation results illustrate that both the wild-type and the D38C mutant proteins show higher thermostability when the concentration of NaCl is raised. However, the D38C GlcDH appears to be slightly more thermostable than "wild type" GlcDH at the NaCl concentration assayed. The half-lives calculated for each protein under the different conditions are shown at Table 3. In general, at temperatures of 60-70 °C the D38C mutant shows a half-life higher than that of wild-type GlcDH. No reliable comparisons can be made at 80 °C, as at that temperature total inactivation of the enzyme is achieved in a few seconds. Below 60 °C the differences between the half-lives are not significant.

| | $t_{1/2}$ 1 M NaCl (h) | | $t_{1/2}$ 2 M NaCl (h) | | $t_{1/2}$ 3 M NaCl (h) | |
	D38C	"Wild type"	D38C	"Wild type"	D38C	"Wild type"
55 °C	33.9	37.2	(a)	(a)	(a)	(a)
60 °C	7.4	4.6	123.7	51.33	210.6	96.27
65 °C	0.3	0.3	9.6	8.3	(a)	(a)
70 °C	(b)	(b)	0.3	0.2	17.2	8.25
80 °C	(b)	(b)	0.04	0.02	0.2	0.2

(a) The enzyme is stable under these conditions.
(b) The enzyme is stable for only a few minutes under these conditions.

Table 3. Half-life time at different temperatures and salt concentrations of wild-type GlcDH and D38C GlcDH.

The replacement of D38 by C38 appears to have a stabilizing effect on the ability of the protein to withstand high temperatures, producing an enzyme that is marginally more stable at high temperature. However, it is clear that the enzymatic activity of the mutant is lower.

3.3 Alteration of coenzyme specificity in *Hfx. mediterranei* GlcDH and *Hfx. volcanii* ICDH

3.3.1 *Hfx. mediterranei* GlcDH

3.3.1.1 Mutations for the reversal of coenzyme specificity

The ability of dehydrogenases to discriminate between NAD$^+$ and NADP$^+$ lies in the amino acid sequence of the nucleotide-binding βαβ motif. This βαβ motif is centered around a highly conserved Gly–X–Gly–X–X–Gly sequence (where X is any amino acid) connecting the first β strand to the α helix. The presence of an aspartic residue at the C-terminal end of the second β strand is conserved in NAD$^+$-specific enzymes. In many NADP$^+$-specific enzymes, this residue is replaced by a smaller and neutral residue and complemented by a nearby positively charged residue that forms a positively charged binding pocket for adenosine 2′-phosphate. The three-dimensional structure of the cofactor binding-site of *Hfx. mediterranei* GlcDH (Fig. 6) indicates the spatial location of the residues mutated here and the interaction of R207 and R208 with the 2′-phosphate group of NADP$^+$ (Britton et al., 2006; Pire et al., 2009).
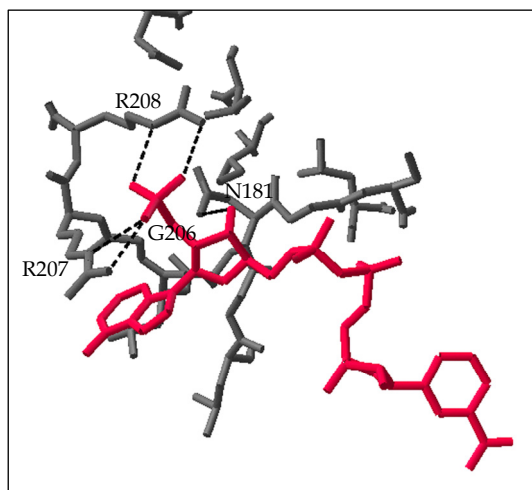
Fig. 6. View of NADP$^+$ bound to "wild type" GlcDH. Interaction through hydrogen bonds is represented with dotted lines. NADP$^+$ is shown in red; a portion of GlcDH, in gray. Crystal structure of GlcDH is from *Hfx. mediterranei* (PDB code 2B5V, Britton et al., 2006).

3.3.1.2 Protein properties

All the reversal coenzyme specificity mutants were expressed as inclusion bodies, and refolding was carried out by rapid dilution in the same way as for the wild-type enzyme

(Pire et al., 2001). To assess that the enzymes reached their maximum activity in terms of proper refolding, enzyme activity was measured as a function of time after rapid dilution. The wild-type and mutated enzymes behaved similarly during refolding, although the refolding kinetics of the mutants were slower. Maximum activity was reached after approximately 24 h with the mutated enzymes, whereas the wild-type enzyme achieved maximum activity 2 h after the rapid dilution of solubilized inclusion bodies (Pire et al, 2009).

Once the protein was folded, the purification procedures were identical for the wild-type and mutant enzymes (Pire et al., 2001).

3.3.1.3 Kinetics of "wild type" and coenzyme specificity reversal mutant enzymes

The kinetic constants of the wild-type and mutant forms of GlcDH were determined with both coenzymes, NAD$^+$ and NADP$^+$. The kinetic constants for the enzymes are compared in Table 4 A and B.

The K_m value of the wild-type enzyme was 11-fold lower for NADP$^+$ than for NAD$^+$, indicating that the enzyme has a strong preference for NADP$^+$. The single substitution G206D increased the K_m 74-fold for NADP$^+$ and decreased K_{cat} 2-fold, resulting in a 150-fold decrease in the K_{cat}/K_m when using NADP$^+$. This was to be expected as the negative charge of D206 would be likely to repel the adenosine 2'-phosphate of NADP$^+$. This single substitution had a positive effect on catalysis with NAD$^+$. In NAD$^+$-dependent enzymes, an aspartic residue in this position confers specificity towards NAD$^+$ by the bidentate hydrogen bonding with the 2' and 3' hydroxyl groups of the adenosine of NAD$^+$. The K_m in the presence of NAD$^+$ was similar to that of the "wild type", but K_{cat} showed a 2-fold increase. The G206D mutant preferred NAD$^+$ to NADP$^+$, showing a K_{cat} value with NAD$^+$ similar to that of the wild-type enzyme with NADP$^+$; however, the K_{cat}/K_m ratio was still better in the wild-type enzyme with NADP$^+$.

The single mutant R207I showed an increase of 48 times in K_m value with NADP$^+$ when compared with the "wild type"; this again was as expected, considering the role of D207 in the stabilization of the negative charge of the adenosine 2'-phosphate group of NADP$^+$. This increase was accompanied by a decrease in K_{cat}, which clearly makes the R207I mutant less efficient in catalysis with NADP$^+$. For NAD$^+$ the K_m value also increased, but at a ratio of 3 times, much lower than the K_m increase with NADP$^+$. The R207I substitution also makes the enzyme less efficient with NAD$^+$, with a decrease of 4 times in K_{cat}/K_m; this substitution also increases the K_m for glucose. A similar effect was even more pronounced in the single substitution R208N, in which saturation with glucose cannot be achieved, and attempts to calculate K_m and K_{cat} with both coenzymes led to very high standard deviation values.

The activity of the G206D/R207I double mutant with NADP$^+$ was very low (almost undetectable), and as such the kinetic parameters could not be calculated. However, when the coenzyme NAD$^+$ was incubated with this double mutant, it reached the highest K_{cat} value, between 1.5 and 2 times higher than the K_{cat} of the wild-type enzyme with NADP$^+$, and between 3 and 4 times higher than the K_{cat} of the "wild type" with NAD$^+$. These values indicate that the local rearrangement of the active centre due to the mutations makes catalysis more efficient. The dissociation constant for NAD$^+$ in the double mutant decreased 1.7-fold in comparison with K_{iNAD+} in the "wild type", but the K_{mNAD+} value of NAD$^+$ registered a 2-fold increase. The G206D/R207I/R208N triple substitution produced an

A)	NADP$^+$ Wild type		G206D	R207I	
K$_{iNAD(P)}^+$ (mM)	0.09 ± 0.02		0.69 ± 0.10	1.3 ± 0.5	
K$_{mNAD(P)}^+$ (mM)	0.035 ± 0.009		2.6 ± 0.4	1.7 ± 0.3	
K$_{mglucose}$ (mM)	2.8 ± 0.3		52 ± 8	24 ± 8	
[a]K$_{cat}$ (min^{-1}) (x10^{-3})	31.1 ± 1.2		15.7 ± 1.5	11.8 ± 1.5	
K$_{cat}$/K$_{mNAD(P)}^+$ (mM^{-1}min^{-1}) (x10^{-3})	900 ± 30		6.0 ± 1.5	7 ± 2	
K$_{cat}$/K$_{mglucose}$ (mM^{-1}min^{-1}) (x10^{-3})	11.1 ± 1.6		0.30 ± 0.07	0.5 ± 0.2	
B)	NAD$^+$ Wild type	G206D	R207I	G206D/ R207I	G206D/R207I/ R208N
K$_{iNAD(P)}^+$ (mM)	2.0 ± 0.2	1.0 ± 0.2	2.1 ± 0.9	1.17 ± 0.13	1.7 ± 0.4
K$_{mNAD(P)}^+$ (mM)	0.40 ± 0.12	0.49 ± 0.16	1.3 ± 0.4	2.7 ± 0.6	1.3 ± 0.3
K$_{mglucose}$ (mM)	12.9 ± 1.4	15 ± 4	32 ± 13	80 ± 14	57 ± 11
[a]K$_{cat}$ (min^{-1}) (x10^{-3})	15.8 ± 0.9	34 ± 5	4.3 ± 0.8	54 ± 7	2.8 ± 0.3
K$_{cat}$/K$_{mNAD(P)}^+$ (mM^{-1}min^{-1}) (x10^{-3})	44 ± 17	70 ± 30	3.3 ± 0,6	20 ± 7	2.1 ± 0.7
K$_{cat}$/K$_{mglucose}$ (mM^{-1}min^{-1}) (x10^{-3})	1.2 ± 0.2	2.3 ± 0.9	0.13 ± 0.08	0.7 ± 0.2	0.049 ± 0.015

Substrate concentrations range used in kinetic parameters determination: Wild-type: [glucose]= 2-20 mM, [NADP$^+$]= 0.04-0.2 mM, [NAD$^+$]= 0.286-1mM; G206D: [NAD$^+$]= 0.286 -1mM, [glucose]= 2.5-20 mM, [NADP$^+$]= 0.286-4 mM, [glucose]= 2.5-40 mM; R207I: [NAD$^+$]= 0.8-2 mM, [glucose]= 25-100 mM, [NADP$^+$]= 0.4-2 mM, [glucose]= 20-100 mM; G206D/R207I: [NAD$^+$]= 0.4-4 mM, [glucose]= 10-100 mM; G206D/R207I/R208N:[NAD$^+$]= 0.8-4 mM, [glucose]= 16.67-100 mM.
[a]K$_{cat}$ values are referred to the GlcDH dimer kinetic parameters and are expressed ± standard deviation.

Table 4. Kinetic constants of wild-type and mutant GlcDH.

inactive enzyme with NADP$^+$, confirming that these two arginines are necessary for NADP$^+$ stabilization. Regarding the kinetic parameters with NAD$^+$, as in the double mutant G206D/R207I, the K$_m$ values of both substrates were higher than in the "wild type"; but in the triple mutant the K$_{cat}$ was also lower, and it was the worst catalyst.

In contrast with our results, in alcohol dehydrogenase from gastric tissues of *Rana perezi*, the complete reversal of coenzyme specificity from NADP(H) to NAD(H) was reached with the concerted mutation of three residues G223D/T224I/H225N (Rosell et al., 2003) . The single mutation G223D had no effect on catalysis with NAD$^+$ and the double mutant G223D/T224I was a better catalyst than the single mutant with NAD$^+$, but worse than the triple mutant. The great increase in the K$_{cat}$ observed with the GlcDH double mutant was not observed in alcohol dehydrogenase. It appears that one or two substitutions in alcohol dehydrogenase were not sufficient to transform coenzyme specificity, whereas multiple substitutions could be effective (Rosell et al., 2003). The same, but reverse, effect was observed with NAD$^+$-dependent xylitol dehydrogenase, in that the reverse of coenzyme specificity from NAD$^+$ to NADP$^+$ was achieved with the triple mutant D207A/I208R/F209S (Watanabe et al., 2005). In yeast alcohol dehydrogenase, a NAD$^+$ specific enzyme, the D201G substitution produces an

enzyme with low activity with NADP$^+$, but the G203R substitution neither affects affinity for NAD$^+$ or NADH nor enables reactivity with NADP$^+$, and the D201G/G203R enzyme has kinetic characteristics similar to the single D201G enzyme. R203 should be able to interact with the 2′-phosphate, but it seems that a greater change is required in the amino acid sequence to transform the specificity (Fan & Plapp, 1999).

Although there are some examples of a coenzyme specificity change from NAD$^+$ to NADP$^+$ with only one mutation, it seems that the specificity change from NADP$^+$ to NAD$^+$ is more difficult to reach with single substitutions. This study shows that G206, R207 and R208 are determinant for coenzyme specificity in *Hfx. mediterranei* GlcDH. The substitution G206D hampered the binding of NADP$^+$ and increased by a factor of two the activity with NAD$^+$, resulting in an enzyme that preferred NAD$^+$ over NADP$^+$. Double mutation G206D/R207I was enough to make an unproductive enzyme with NADP$^+$, although the more important findings were that in double mutant G206D/R207I, the specific activity of the enzyme with NAD$^+$ was almost twice than in the wild type with NADP$^+$, though the K_{cat}/K_m ratio was low due to the increase of K_m. In this sense, as some authors point out, we have to be cautious to interpret the K_{cat}/K_m ratio as catalytic efficiency, since at certain substrate concentrations the wild-type GlcDH catalyzes the oxidation of glucose, using NADP$^+$ as a coenzyme at a lower rate than double mutant in the presence of NAD$^+$ (Pire et al., 2009)

3.3.2 *Hfx. volcanii* ICDH

One of the most interesting features of proteins is the fact that they keep in their amino acid sequences a substantial record of their evolutionary histories. Surprisingly, homologous proteins in organisms that diverged billions of years are still similar enough to recognize a correspondence in the organization of conserved and variable regions. They can even be used as markers of the evolutionary process itself. Such comparisons have been performed using protein sequence alignments obtained with different algorithms. The alignments may reveal amino acids with a common origin and/or having similar positions in the corresponding three-dimensional structures of each protein.

Molecular evolution is based on the use of alignments to reconstruct gene trees representing, as closely as possible, the historic process of sequence divergence. This reconstruction requires the development of statistical models able to reproduce the process of mutation, drift and selection.

If one represents the structural alignment of a family of proteins in the form of linear sequence of amino acids, one can see that the spatial correspondence of identical amino acids is reflected in the form of sequence identity. The presence of insertions and deletions of specific parts of structure is revealed in the form of holes or gaps (Gómez-Moreno & Sancho, 2003).

The aim of alignment algorithms for amino acid sequences is the relative structural correspondence of residues. Comparative modeling is the extrapolation of the structure to a new amino acid sequence (model) from a known three-dimensional structure of at least one member (mould) of the same family of proteins. The obtained models contain sufficient information to permit experimental design with an acceptable degree of reliability or to allow structural comparison (Gómez-Moreno & Sancho, 2003).

One of the purposes of such an analysis is to select residues for mutation that may cause changes in some of the biochemical characteristics of the protein, such as the variation in coenzyme specificity.

3.3.2.1 Sequence alignment

The ICDHs belong to an ancient and divergent family of decarboxylating dehydrogenases that includes NAD-isopropylmalate dehydrogenase (IMDH) (Dean & Golding, 1997). "This family of dehydrogenases shares a common protein fold, topologically distinct from other dehydrogenases of known structure that lacks the αβαβ binding motif characteristic of the nucleotide-binding Rossman fold (Rossman et al., 1974; Chen et al, 1995). In ICDHs the adenosine moiety of the coenzyme binds in a pocket constructed from two loops and an α-helix (Hurley et al., 1991), although the latter is substituted by a β-turn in IMDH (Imada et al., 1991; Chen et al., 1995).

Dehydrogenases discriminate between nicotinamide coenzymes through interactions established between the protein and the 2′-phosphate of NADP+ and the 2′- and 3′-hydroxyls of NAD+ (Chen et al., 1995). In the NAD-binding site, the introduction of positively charged residues changes the preference of an NAD-dependent enzyme to neutralize the negatively charged 2′-phosphate of NADP+, as it has been demonstrated with engineered dihydrolipomide and malate dehydrogenases (Bocanegra et al., 1993; Nishiyama et al., 1993).

Specificity in *E. coli* ICDH is conferred by interactions among R395, K344, Y345, Y391 and R292′ with the 2′-phosphate of bound NADP+ (Hurley et al., 1991; Dean & Golding, 1997). These residues are conserved in oligomeric NADP-dependent ICDHs and some monomeric NADP–ICDHs from prokaryotes. They are replaced with a variety of residues in the NAD-dependent ICDHs (Lloyd & Weitzman, 1988; Steen et al., 1997; Yasutake et al., 2003). Except for the substitution of Y391 with glutamine in *Aeropyrum pernix*, these residues are conserved in the archaeal NADP-dependent ICDHs and are in accordance with the cofactor specificity (Steen et al., 2002). However, K344 and Y345, which interact with NADP+ in ICDH, are substituted by D278 and I279 in IMDH. This enzyme preferentially uses NAD+ as a coenzyme; and D278 hydrogen bonds to the 2′-hydroxyl group of NAD+, while repelling the 2′-phosphate of NADP+ (Chen et al., 1996; Hurley et al., 1996). Thus, this residue is a major determinant of coenzyme specificity toward NAD+ and is strictly conserved in all known IMDHs (Chen et al., 1996). The specificity in IMDH is conferred by the strictly conserved D278 (D344 in ICDH numbering), which forms a double hydrogen bond with the 2′- and 3′-hydroxyls of the adenosine ribose of NAD+ (Dean & Golding, 1997). Not only are these movements incompatible with the strong 2′-phosphate interactions seen in NADP-ICDH, but the negative charge on D344 may also repel NADP+ (Hurley et al., 1996; Rodriguez-Arnedo et al., 2005).

Specificity is governed by (1) residues that interact directly with the unique 2′-hydroxyl and phosphate groups of NAD+ and NADP+, respectively; (2) more distant residues that modulate the effects of the first group; and (3) remote residues (Hurley et al., 1996). The first group of residues includes L344, Y345, Y391 and R395 (*E. coli* ICDH numbering), and it is easy to imagine that changes at these residues destroy the 2′-phosphate binding site (Hurley et al., 1996). The second group of residues includes V351. The adenine ring of NAD+ approaches the C_α of residue 351 as a result of a conformational change in the adenosine

ribose of NAD$^+$, which also brings it close to D344, suggesting an important role for the ring shift in specificity. The V351A mutant was designed to avoid obstructing this ring shift. A second key role for V351A is to accommodate the correct packing of the introduced I345 (Hurley et al., 1996). Sequence alignment of *Hfx. volcanii*, *E. coli* and *B. subtilis* NADP-dependent ICDHs with *T. thermophilus* NAD-dependent IMDH (Fig. 7) showed that amino acid residues involved in NADP$^+$ binding are conserved in the NADP-dependent ICDHs and are located in a characteristic binding pocket constructed from two loops and an α-helix. In IMDH from *T. thermophilus*, this pocket comprises two loops and a β-turn (Dean & Golding, 1997; Rodriguez-Arnedo et al., 2005).

```
                                                                 Loop1
H. volcanii [Q8X277]      KVVVKDRIADNMLQQLLTRTADYDVIATMNLNGDYMSDAAGAQIGGLGIAPGANFGEGL--CLA
E. coli [P08200]          EIVIKDVIADAFLQQILLRPAEYDVIACMNLNGDYISDALAAQVGGIGIAPGANIGDEC--ALF
B. subtilis [P39126]      KIIIKDSIADIFLQQILTRPNEFDVVATMNLNGDYISDALAAQVGGIGIAPGANINYETGHAIF
T. thermophilus [P00351]  DVALEHQYVDAMAMHLVRSPARFDVVVTGNIFGDILSDLASVLPGSLGLLPSASLGRGT--PVF
                                            291

                          Loop 2                                                α-helix
H. volcanii [Q8X277]      EPVHGSAPKYAGKDKVNPTAMILSGRLMFEYMGW-KDAGKLIRDAVEKTISDGDVTYDLERQIE
E. coli [P08200]          EATHGTAPKYAGQDKVNPGSIILSAEMMLRHMGW-TEAADLIVKGMEGAINAKTVTYDFERLMD
B. subtilis [P39126]      EATHGTAPKYAGLDKVNPSSVILSGVLLLEHLGW-NEAADLVIKSMEKTIASKVVTYDFARLMD
T. thermophilus [P00351]  EPVHGSAPDIAGKGIANPTAAILSAAMMLEHAFGLVELARKVEDAVAKALLE-TPPFLETPPPD
                                   343   350                                    390
```

Fig. 7. Sequence alignments of *Hfx. volcanii*, *E. coli* and *B. subtilis* NADP-dependent isocitrate dehydrogenases (ICDHs) and *T. thermophilus* NAD-dependent isopropylmalate dehydrogenase. Amino acids critical to coenzyme binding and catalysis are enclosed in black boxes, and numbers correspond to relative position in the ICDH sequence from *Hfx. volcanii*.

3.3.2.2 Site-directed mutagenesis

In the halophilic enzyme, the R291S, K343D, Y344I, V350A and Y390P (halophilic ICDH numbering) mutations were selected based on homology. The substitutions were made by site-directed mutagenesis. The changes carried out are positively charged residues, such as Arg and Lys; uncharged amino acids, such as Ser; or negatively charged, such as Asp. Lys is a residue that appears to be conserved in many species, which could mean that its positive charge is crucial for proper catalysis by the enzyme. The first mutant made and characterized was R291S. Arg forms a hydrogen bond with the 2'-phosphate of a NADP$^+$, as can be seen in the *E. coli* ICDH structure, and is replaced by Ser in *T. thermophilus* IMDH and by a wide variety of amino acids in other NAD-dependent enzymes (Chen et al., 1995). We found that, after the R291S substitution in *Hfx. volcanii* ICDH, the specificity for NADP$^+$ decreased; but no activity for NAD$^+$ was detected. The double mutant was obtained with a second mutation at residue Y390 in the halophilic enzyme (Rodriguez-Arnedo et al., 2005). This amino acid was replaced with Pro, as in IMDH from *T. thermophilus* (Chen et al., 1995), in order to remove a hydrogen bond to the 2'-phosphate and alter the local secondary structure from an α-helix to a β-turn. When examined, however, the secondary structure was unchanged by this substitution. Furthermore, preliminary studies (Chen et al., 1995) have shown that it is unnecessary to replace an α-helix by a β-turn to eliminate interactions with the 2'-phosphate of NADP$^+$. Mutants three (SDP mutant) and four (SDIP mutant) involved replacing K343 and Y344 with Asp and Ile, respectively, to obtain triple and quadruple mutants. Both D343 (D278 in IMDH numbering) and I344 (I279 in IMDH

numbering) are found in all known prokaryotic NAD-dependent decarboxylating dehydrogenases (Chen et al., 1996). In *Pyrococcus furiosus* NAD-dependent ICDH, the introduction of the double mutation of D328L/I329Y did not greatly decrease the preference for NAD$^+$; but it improved the preference for NADP$^+$ by reducing the catalytic efficiency for NAD$^+$ (Steen et al., 2002). In comparison, the loss of hydrogen bonding and the introduction of a negatively charged Asp greatly reduced binding with NADP$^+$, but the changes had no effect on binding with NAD$^+$ in *Hfx. volcanii* ICDH. The final mutation was V350A to obtain a quintuple mutant (SDIAP mutant) (Rodriguez-Arnedo et al., 2005).

After all five amino acids were changed in *Hfx. volcanii* ICDH, coenzyme binding switched from NADP$^+$ to NAD$^+$. The importance of V350A could explain why the quadruple mutation (SDIP mutant) destroyed NADP$^+$ binding without promoting NAD$^+$ binding. The corresponding residue is Val or Ile in all known NADP-dependent ICDHs and Ala in most, but not all, NAD-dependent dihydrodiol dehydrogenases (Hurley et al., 1996). Coenzyme specificity was changed to NAD$^+$ only after the fifth mutation, V350A. When V350A was the only change (single mutant), the coenzyme specificity was not switched to NAD$^+$, although specificity for NADP$^+$ decreased, as occurred when R291S was the only change (Rodriguez-Arnedo et al., 2005).

Molecular models of both the native enzyme and the mutants showed no significant changes in secondary structure when they were compared with the model of *E. coli* ICDH, and with each other (data not shown). The high sequence identity (56.6%) of halophilic organism isocitrate dehydrogenase with that of *E. coli* allows a reasonable model which would present a deviation of 1 Å.

The introduced mutations in ICDH from *Hfx. volcanii* do not produce apparent changes in the structure of the enzyme. Therefore, the change of coenzyme specificity would have occurred by the elimination of the hydrogen bonds formed by the amino acids involved in the binding of NADP$^+$. The position of the NADP$^+$ in the mutant is the same as it is in the native enzyme. The loop formed to accommodate the coenzyme is the same as the one seen in the *E. coli* enzyme. This loop has also been observed in the three-dimensional structure of ICDH from *B. subtilis* (Singh et al., 2001).

3.3.2.3 Kinetic characterization of mutants

Five amino acid substitutions introduced into wild-type *Hfx. volcanii* ICDH caused a complete shift in preference from NADP$^+$ to NAD$^+$. All mutations were based on sequence homology with NAD-dependent enzymes. Wild-type ICDH showed total dependence on NADP$^+$ and had a Michaelis constant (K_m) of 101 ± 30 μM and a K_{cat} of 0.176 ± 0.020 min^{-1} (Table 5). The affinity for the coenzyme was not changed until all five mutations were made, but specificity for NADP$^+$ decreased after the first mutation. Our SDIAP mutant displayed a preference for NAD$^+$ over NADP$^+$, had a K_m for NAD$^+$ of 144 ± 60 μM, and a K_{cat} of 0.0422 ± 0.0017 min^{-1} (Table 5). The K_m for NADP$^+$ increased approximately threefold in the single mutant and remained constant in the three subsequent mutants. The K_{cat}/K_m for the NADP$^+$ coenzyme was reduced 24-fold in the quadruple mutant. In contrast to the results obtained with other NAD(P)-dependent dehydrogenases (Nishiyama et al., 1993; Chen et al., 1997; Steen et al., 2002), our experiments showed that none of the mutants of *Hfx. volcanii* ICDH displayed dual cofactor specificity. Table 5 indicates that the K_{cat} for NAD$^+$ is lower than the K_{cat} for NADP$^+$ for the wild-type and some mutants, suggesting that most of the changes in

specificity arise from discrimination in binding, rather than direct changes in catalysis. Catalytic efficiency of the quintuple mutant (SDIAP) with NAD^+ was 17% of that of the wild-type enzyme with $NADP^+$, suggesting that a hydrogen bond between the adenosine ribose of NAD^+ and D343, as seen in the X-ray structure of the IMDH binary complex (Dean & Golding, 1997), may have been successfully established in the halophilic enzyme purified from *E. coli*. Thus, we have obtained an ICDH mutant with clearly changed coenzyme specificity (Rodriguez-Arnedo et al., 2005).

Isocitrate specificity changed with the first mutation, but in this case, specificity for isocitrate increased 3- to 10-fold with increasing mutations, suggesting that these mutations favored substrate binding. The maximum value for isocitrate binding occurred when the mutant showed specificity for NAD^+ only. Thus, the mutations markedly influenced not only the K_m for $NAD(P)^+$, but also the K_m for isocitrate. The effect of the mutation on the efficiency for $NADP^+$ or NAD^+ was evaluated by incorporating the K_m for the substrate. This parameter is called the overall catalytic efficiency and is defined as: $(K_{cat}/(K_m^{\ IC} \times K_m^{\ NAD(P)})$ (Table 5) (Nishiyama et al., 1993). We speculate that some specific interaction between the substrate and $NADP^+$, which differs from the native substrate–coenzyme complex, is responsible for the decrease in activity. The ratio K_{cat}/K_m is a measure of both enzyme efficiency and the degree to which an enzyme stabilizes the transition state (Dean & Golding, 1997).

Residues at	$NADP^+$			NAD^+		
291 343 344 350 390	K_m (µM)	K_{cat} (min^{-1})	K_{cat}/K_m x 10^3 (min^{-1}/µM)	K_m (µM)	K_{cat} (min^{-1})	K_{cat}/K_m x 10^3 (min^{-1}/µM)
R K Y V Y	101±30	0.176±0.020	1.7±0.7	____	_____	_____
S - - - -	350±30	0.244±0.020	0.70±0.12	____	_____	_____
S - - - P	327±50	0.060±0.004	0.18±0.04	____	_____	_____
S D - - P	203±30	0.145±0.008	0.70±0.14	____	_____	_____
S D I - P	282±20	0.020±0.003	0.071±0.016	____	_____	_____
S D I A P	____	_____	_____	144±60	0.0422± 0.0017	0.29±0.13

Residues at	Isocitrate			
291 343 344 350 390	K_m (µM)	K_{cat} (min^{-1})	K_{cat}/K_m x 10^3 (min^{-1}/µM)	$K_{cat}/(K_m^{\ IC} \times K_m^{\ NAD(P)})$
R K Y V Y	108±30	0.139±0.010	1.3±0.5	1.3 x 10^{-5}
S - - - -	18±3	0.123±0.002	6.8±1.2	2.0 x 10^{-5}
S - - - P	36±4	0.032±0.003	0.88±0.18	2.7 x 10^{-6}
S D - - P	36±5	0.086±0.003	2.4±0.4	1.2 x 10^{-5}
S D I - P	40±7	0.0149±0.0003	0.37±0.07	1.3 x 10^{-6}
S D I A P	11±1*	0.0327±0.0013	3.0±0.4*	2.0 x 10^{-5}*

Initial rates were measured spectrophotometrically at 30 °C in 20 mM Tris-HCl buffer pH 8.0, 1 mM EDTA and 10 mM MgCl$_2$ (Tris/EDTA/Mg^{2+}) containing 2 M NaCl. Substrate and coenzyme concentrations were varied between 0.019 and 0.10 mM for D,L-isocitrate and between 0.044 and 0.4 mM for $NADP^+$ or NAD^+. Single-letter amino acid codes denote amino acid residues. Catalytic efficiency was measured as K_{cat}/K_m. Asterisks indicate values that were obtained with NAD^+ as the coenzyme. Abbreviations: K_m = Michaelis constant; K_{cat} = turnover number; K_{cat}/K_m = catalytic efficiency; and $K_{cat}/(K_m^{IC} \times K_m^{NAD(P)})$ = overall catalytic efficiency.

Table 5. Kinetic parameters of purified wild-type and mutant *Hfx. volcanii Ds-threo*-isocitrate dehydrogenases (ICDHs) toward $NADP^+$ and NAD^+.

One might think that a local conformational change induced by specific binding of NADP⁺ or NAD⁺ is responsible for the variation in the behavior of the *Hfx. volcanii* enzyme versus isocitrate. This effect is probably due to less repulsion of the charges in the active site.

The comparison of the sequence with that of the *T. thermophilus* IMDH (Imada et al., 1991) reveals a consistent framework in the evolution of the substrate specificity of the decarboxylating dehydrogenases belonging to the class of isocitrate dehydrogenases. All residues (R117, R127, R151, Y158 and K228) interact with the α- and β-carboxylates and α-hydroxyls of isocitrate, which are common with isopropylmalate and malate; all residues are conserved in all known ICDHs and IMDHs. Two aspartic acid residues, D282 and D306, coordinate Mg^{2+}. They are also conserved in all available sequences. Residues that interact with the γ-carboxylate of isocitrate, S111 and N113 (*Hfx. volcanii* ICDH numbering), are only found in ICDH sequences and are not found in IMDH sequences. Common features of ICDH and IMDH are due to conserved residues that interact with the α- and β-groups and the metal ion. Difference in the substrate specificity is determined by non-conserved residues that interact with the γ-group (Hurley et al., 1991).

4. Conclusion

Site–directed mutagenesis has allowed us to (1) extend the understanding of the molecular basis of salt tolerance for halophilic adaptation, (2) analyze the role of sequence differences between thermophilic and halophilic dehydrogenases involving a ligand to the zinc ion, and (3) identify the residues implicated in coenzyme specificity.

The replacement of aspartic residues by lysine residues on the GlcDH surface have led to a modification of the halophilic properties of the mutant enzymes, D172K and D216K being the most significant mutations (Esclapez et al., 2007).

The mutation of D38, a residue that lies close to the catalytic zinc ion, to C38 or A38 led to a significant reduction in and abolition of activity, respectively. These results suggest that this residue is important in catalysis, either in forming a key aspect of the zinc-binding site or in some other process related with substrate recognition. The replacement of D38 by C38 results in the production of a less efficient enzyme with lower enzymatic activity and catalytic efficiency. Furthermore, this mutant shows slightly more thermostability. Although the D38C GlcDH is less active, it has been crystallized in the presence of several combinations of products and substrates. This fact has allowed us to describe many aspects of the mechanism of the zinc-dependent MDR superfamily (Esclapez et al., 2005; Baker et al., 2009).

Structural analysis of the GlcDH from *Hfx. mediterranei* revealed that the adenosine 2'-phosphate of NADP⁺ is stabilized by the side chains of R207 and R208. The first attempt to change coenzyme specificity involved making the G206D mutant. Further substitutions of uncharged residues for these residues were made to analyze their importance in NADP⁺ binding and to improve specificity for NAD⁺. The single mutants G206D and R207I were less efficient with NADP⁺ than the wild type, and the double and triple mutants, G206D/R207I and G206D/R207I/R208N, showed no activity with NADP⁺ (Pire et al., 2009).

The results obtained in our study with the halophilic ICDH and the complete switch of coenzyme specificities in IMDH from *T. thermophilus* (Imada et al., 1991) and ICDHs from *E. coli* show that coenzyme specificity in the β-decarboxylating dehydrogenases are principally determined by interactions between the nucleotides and surface amino acid residues lining the binding pockets (Rodriguez-Arnedo et al., 2005).

5. Acknowledgment

6. References

Bonete, M. J., Pire, C., Llorca, F. I. & Camacho, M. L. (1996) Glucose dehydrogenase from the halophilic Archaeon *Haloferax mediterranei*: enzyme purification, characterisation and N-terminal sequence. *FEBS Lett.*, 383: 227-229.

Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.*, 72: 248-254.

Baker, P. J., Britton, K. L., Fisher, M., Esclapez, J., Pire, C., Bonete, M. J., Ferrer, J. & Rice, D. W. (2009) Active site dynamics in the zinc-dependent medium chain alcohol dehydrogenase superfamily. *Proc. Natl. Acad. Sci. USA,* 106: 779-784.

Bocanegra, J. A., Scrutton, N. S. & Perham, R. N. (1993) Creation of an NADP-dependent pyruvate dehydrogenase multienzyme complex by protein engineering. *Biochemistry*, 32: 2737–2740.

Britton, K. L., Stillman, T. J., Yip, K. S. P., Forterre, P., Engel, P. C. & Rice, D. W. (1998) Insights into the molecular basis of salt tolerance from the study of glutamate dehydrogenase from *Halobacterium salinarum*. *J. Biol. Chem.*, 273: 9023-9030.

Britton, K. L., Baker, P. J., Fisher, M., Ruzheinikov, S., Gilmour, D. J., Bonete, M. J., Ferrer, J., Pire, C., Esclapez, J. & Rice, D. W. (2006) Analysis of protein solvent interactions in glucose dehydrogenase from the extreme halophile *Haloferax mediterranei*. *Proc. Natl. Acad. Sci. USA,* 103: 4846-4851.

Camacho, M. L., Brown, R. A., Bonete, M.-J., Danson, M. J. & Hough, D. W. (1995) Isocitrate dehydrogenases from *Haloferax volcanii* and *Sulfolobus solfataricus:* enzyme purification, characterisation and N-terminal sequence. *FEMS Microbiol. Lett.*, 134: 85-90.

Camacho, M. L., Rodríguez-Arnedo, A. & Bonete, M. J. (2002) NADP-dependent isocitrate dehydrogenase from the halophilic archaeon *Haloferax volcanii*: cloning, sequence determination and overexpression in *Escherichia coli*. *FEMS Microbiol. Lett.*, 209: 155-160.

Chen, R., Greer, A. & Dean, A. M. (1995) A highly active decarboxylating dehydrogenase with rationally inverted coenzyme specificity. *Proc. Natl. Acad. Sci. USA,* 92: 11666–11670.

Chen, R., Greer, A. & Dean, A. M. (1996) Redesigning secondary structure to invert coenzyme specificity in isopropylmalate dehydrogenase. *Proc. Natl. Acad. Sci. USA*, 93: 12171–12176.

Chen, R., Greer, A. & Dean, A. M. (1997) Structural constraints in protein engineering: the coenzyme specificity of *Escherichia coli* isocitrate dehydrogenase. *Eur. J. Biochem.*, 250: 578–582.

Dean, A. M. & Golding, G. B. (1997) Protein engineering reveals ancient adaptive replacements in isocitrate dehydrogenase. *Proc. Natl. Acad. Sci. USA,* 94: 3104–3109.

Eisenberg, H., Mevarech, M. & Zaccai, G. (1992) Biochemical, structural and molecular genetic aspects of halophilism. *Adv. Protein Chem.*, 43: 1-62.

Eklund, H., Samma, J. P., Wallen, L., Brändén, C. L., Akeson, A. & Jones, T. A. (1981) Structure of a triclinic ternary complex of horse liver alcohol dehydrogenase at 2,9 Å resolution. *J. Mol. Biol.*, 146: 561-587.

Esclapez, J., Britton, K. L., Baker, P. J., Fisher, M., Pire, C., Ferrer, J., Bonete, M. J. & Rice, D. W. (2005) Crystallization and preliminary X-ray analysis of binary and ternary complexes of *Haloferax mediterranei* glucose dehydrogenase. *Acta Cryst. F*, 61: 743-746.

Esclapez, J., Pire, C., Bautista, V., Martínez-Espinosa, R. M., Ferrer, J. & Bonete, M. J. (2007) Analysis of acidic surface of *Haloferax mediterranei* glucose dehydrogenase by site-directed mutagenesis. *FEBS Lett.*, 581: 837-842.

Fan, F. & Plapp, B. V. (1999) Probing the affinity and specificity of yeast alcohol dehydrogenase I for coenzymes. *Arch. Biochem. Biophys.*, 367: 240-249.

Ferrer, J., Fisher, M., Burke, J., Sedelnikova, S. E., Baker, P. J., Gilmour, D. J., Bonete, M. J., Pire, C., Esclapez, J. & Rice, D. W. (2001) Crystallization and preliminary X-ray analysis of glucose dehydrogenase from *Haloferax mediterranei*. *Acta Cryst. D*, 57: 1887-1889.

Gómez-Moreno, C. & Sancho, J. (Eds) (2003). *Estructura de Proteínas*, Ariel Ciencia, ISBN 9788434480612, Barcelona, Spain.

Hurley, J. H., Dean, L. A. M., Koshland, Jr., D. E. & Stroud, R. M. (1991) Catalytic Mechanism of NADP$^+$-Dependent Isocitrate Dehydrogenase: Implications from the Structures of Magnesium-Isocitrate and NADP$^+$ Complexes. *Biochemistry*, 30: 8671-8678.

Hurley, J. H. & Dean, A. M. (1994) Structure of 3-isopropylmalate dehydrogenase in complex with NAD$^+$: ligand induced loop closing and mechanism for cofactor specificity. *Structure,* 2: 1007–1016.

Hurley, J. H., Chen, R. & Dean, A. M. (1996) Determinants of Cofactor Specificity in Isocitrate Dehydrogenase: Structure of an Engineered NADP$^+$ → NAD$^+$ Specificity-Reversal Mutant. *Biochemistry*, 35: 5670–5678.

Imada, K., Sato, M., Tanaka, N., Katsube, Y., Matsuura, Y. & Oshima, T. (1991) Three-dimensional structure of a highly thermostable enzyme, 3-isopropylmalate dehydrogenase of *Thermus thermophilus* at 2.2 Å resolution *J. Mol. Biol.*, 222: 725–738.

Irimia, A., Ebe, C., Madern, D., Richard, S. B., Cosenza, L. W., Zaccai, G. & Vellieux, F. M. D. (2003) The oligomeric states of *Haloarcula marismortui* malate dehydrogenase are modulated by solvent components as shown by crystallographic and biochemical studies. *J. Mol. Biol.*, 326: 859-873.

Jolley, K. A., Rusell, R. J. M., Hough, D. H. & Danson, M. J. (1997) Site-directed mutagenesis and halophilicity of dihydrolipoamide dehydrogenase from the halophilic archaeon, *Haloferax volcanii*. *Eur. J. Biochem.*, 248: 362-368.

Kauri, T., Wallace, R. & Kushner, D. J. (1990) Nutrition of the halophilic archaebacterium, *Haloferax volcanii*. *Syst. Appl. Microbiol.*, 13: 14–18.

Kay, J. & Weitzman, P. D. J. (Eds.) (1987) Krebs' Citric Acid Cycle: Half a Century and Still Turning. *Biochem. Soc. Symp.*, 54: l-198.

Lanyi, J. K. (1974) Salt-dependent properties of proteins from extremely halophilic bacteria. *Bacteriol. Rev.*, 38: 272-290.

Lloyd, A. J. & Weitzman, P. D. J. (1988) Purification and characterization of NAD-linked isocitrate dehydrogenase from *Methylophilus methylotrophus. Biochem. Soc. Trans.*, 16: 871–872.

Madern, D., Pfister, C. & Zaccai, G. (1995) Mutation at a single acidic amino acid enhances the halophilic behavior of malate dehydrogenase from *Haloarcula marismortui* in physiological salts. *Eur. J. Biochem.*, 230: 1088-1095.

Madern, D., Ebel, C. & Zaccai, G. (2000) Halophilic adaptation of enzymes. *Extremophiles*, 4: 91-98.

Madern, D., Camacho, M., Rodríguez-Arnedo, A., Bonete, M.-J. & Zaccai, G. (2004) Salt-dependent studies of NADP-dependent isocitrate dehydrogenase from the halophilic archaeon *Haloferax volcanii*. *Extremophiles*, 8: 377-384.

Nishiyama, M., Birktoft, J. J. & Beppu, T. (1993) Alteration of Coenzyme Specificity of Malate Dehydrogenase from *Thermus flavus* by Site-directed Mutagenesis. *J. Biol. Chem.*, 268: 4656–4660.

Pire, C., Camacho, M. L., Ferrer, J., Hough, D. W. & Bonete, M. J. (2000) NAD(P)+- glucose dehydrogenase from *Haloferax mediterranei*: kinetic mechanism and metal content. *J. Mol. Catal. B: Enzym.*, 10: 409-417.

Pire, C., Esclapez, J., Ferrer, J. & Bonete, M. J. (2001) Heterologous overexpression of glucose dehydrogenase from the halophilic archaeon *Haloferax mediterranei*, an enzyme of the medium chain dehydrogenase/reductase family. *FEMS Lett.*, 200: 221-227.

Pire, C., Esclapez, J., Díaz, S., Pérez-Pomares, F., Ferrer, J. & Bonete, M. J. (2009) Alteration of coenzyme specificity in halophilic NAD(P)+ glucose dehydrogenase by site-directed mutagenesis. *J. Mol. Catal. B: Enzym.*, 59: 261-265

Richard, S. B., Madern, D., Garcin, E. & Zaccai, G. (2000) Halophilic adaptation: novel solvent-protein interactions observed in the 2.9 and 2.6 Å resolution structures of the wild type and a mutant of malate dehydrogenase from *Haloarcula marismortui. Biochemistry,* 39: 992-1000.

Rodríguez-Arnedo, A., Camacho, M. & Bonete, M. J. (2005) Complete reversal of coenzyme specificity of isocitrate dehydrogenase from *Haloferax volcanii. Protein J.,* 24: 259-266.

Rodríguez Valera, F., Juez, G. & Kushner, D. J. (1983) *Halobacterium mediterranei* spec. nov., a new carbohydrate utilizing extreme halophile. *System. Appl. Microbiol.*, 4: 369-381.

Rosell, A., Valencia, E., Ochoa, W. F., Fita, I, Parés, J. & Farrés, X. (2003) Complete reversal of coenzyme specificity by concerted mutation of three consecutive residues in alcohol dehydrogenase. *J. Biol. Chem.,* 278: 40573-40580.

Rossmann, M. G., Moras, D. & Olsen, K. W. (1974) Chemical and biological evolution of nucleotide-binding protein. *Nature,* 250: 194-199.

Serrano, J. A., Camacho, M. & Bonete, M. J. (1998). Operation of glyoxylate cycle in halophilic archaea: presence of malate synthase and isocitrate lyase in *Haloferax volcanii*. *FEBS Lett.*, 434: 13–16.

Singh, S. K., Matsuno, K., LaPorte, D. C. & Banaszak, L. J. (2001) Crystal Structure of *Bacillus subtilis* Isocitrate Dehydrogenase at 1.55 Å. Insights Into The Nature of Substrate Specificity Exhibited by *Escherichia coli* Isocitrate Dehydrogenase Kinase/Phosphatase. *J. Biol. Chem.,* 276: 26154–26163.

Steen, I. H., Lien, T. & Birkeland, N.-K. (1997) Biochemical and phylogenetic characterization of isocitrate dehydrogenase from a hyperthermophilic archaeon, *Archaeoglobus fulgidus*. *Arch. Microbiol.*, 168: 412–420.

Steen, I. H., Lien, T., Madsen, M. S. & Birkeland, N.-K. (2002) Identification of cofactor discrimination sites in NAD-isocitrate dehydrogenase from *Pyrococcus furiosus. Arch. Microbiol.*, 178: 297–300.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.,* 22: 4673-4680.

Watanabe, S., Kodaki, T. & Makino, S. (2005) Complete reversal of coenzyme specificity of xylitol dehydrogenase and increase of thermostability by the introduction of structural zinc. *J. Biol. Chem.,* 280: 10340-10349.

Yasutake, Y., Watanabe, S., Yao, M., Takada, Y., Fukunaga, N. & Tanaka, I. (2003) Crystal Structure of the Monomeric Isocitrate Dehydrogenase in the Presence of NADP+. Insight into the cofactor recognition, catalysis, and evolution. *J. Biol. Chem.*, 278: 36897–36904.

Targeted Mutagenesis in the Study of the Tight Adherence (*tad*) Locus of *Aggregatibacter actinomycetemcomitans*

David H. Figurski[1], Daniel H. Fine[2], Brenda A. Perez-Cheeks[1],
Valerie W. Grosso[1], Karin E. Kram[1], Jianyuan Hua[1],
Ke Xu[1] and Jamila Hedhli[1]

*[1]Department of Microbiology & Immunology, College of Physicians & Surgeons,
Columbia University, New York, NY, USA*
*[2]Department of Oral Biology, The University of Medicine & Dentistry of New Jersey,
Newark, NJ, USA*

1. Introduction

Aggregatibacter actinomycetemcomitans is a Gram-negative, capnophilic (CO_2 loving), coccobacillus found only in humans and Old World primates (for reviews, see Henderson et al., 2003, 2010). This bacterium is primarily known as the etiologic agent of Localized Aggressive Periodontitis (LAP), which is predominantly an infection of adolescents (Slots & Ting, 1999; Zambon, 1985). *A. actinomycetemcomitans* also causes extraoral infections, including infective endocarditis, septicemia, and abscesses (Fine et al., 2006; Fives-Taylor et al., 1999; Rahamat-Langendoen et al., 2011; van Winkelhoff & Slots 1999). *A. actinomycetemcomitans* is a member of the HACEK group of Gram-negative bacteria, all of which can cause infective endocarditis (Paturel et al., 2004). Most HACEK bacteria-caused cases of infective endocarditis result from *A. actinomycetemcomitans*. Poor oral health is a risk factor for developing severe extraoral infections by *A. actinomycetemcomitans* (Paturel et al., 2004; van Winkelhoff & Slots 1999). One study showed that 16% of patients with infective endocarditis from *A. actinomycetemcomitans* had dental procedures immediately prior to the onset of disease. A staggering 42% of *A. actinomycetemcomitans*-caused infective endocarditis patients had generalized dental disease – an indication of poor overall oral health.

LAP is a severe form of periodontitis (Genco et al., 1986; Slots & Ting, 1999). For reasons so far unknown, the disease is localized to the premolar and incisor teeth. Infection leads to inflammation and rapid destruction of the periodontal ligament and the alveolar bone and culminates in loss of teeth. The prevalence of LAP has been estimated to be 0.5-1% (Henderson et al., 2002; Löe & Brown, 1991; Rylev & Kilian, 2008). However, prevalence varies considerably with different ethnic groups. For example, in the African-American population, the prevalence is 10-15 times higher than the average. There is evidence that race and socioeconomic status play key roles in determining prevalence.

The molecular mechanisms behind the pathogenesis of *A. actinomycetemcomitans* are not understood. The organism elaborates a number of factors that have been implicated in

virulence (Fine et al., 2006; Fives-Taylor et al., 1999). Several adhesins have been identified. The best studied adhesins are (1) Aae, an autotransporter that binds to buccal epithelial cells (Fine et al., 2005, 2010; Rose et al., 2003); (2) ApiA, another autotransporter that binds to buccal epithelial cells, but with lower affinity than does Aae (Komatsuzawa et al., 2002; Yue et al., 2007); and (3) EmaA, which binds collagen (Jiang et al., 2012; Mintz, 2004). A leukotoxin affects leukocytes by inducing one of several pathways (Tsai et al., 1979; Kachlany, 2010). The pathway that is activated depends on the cell type. The toxic effects include apoptosis, degranulation, and a novel lysosome-mediated cell-death pathway (DiFranco et al., 2012; Fong et al., 2006; Kelk et al., 2003, 2011; Lally et al., 1999). Another factor is the cytolethal distending toxin, which causes death by cell-cycle arrest (De Rycke & Oswald, 2001; Fine et al., 2006; Pickett & Whitehouse, 1999).

1.1 Non-specific adherence and virulence

A particularly striking property of a fresh clinical isolate of *A. actinomycetemcomitans* is its ability to form an extremely tenacious biofilm on inert surfaces, such as glass, plastic, and hydroxyapatite (Fig. 1) (Fine et al., 1999a; Kachlany et al., 2000, 2001a). Freshly isolated clinical strains of *A. actinomycetemcomitans* form small, rough colonies on agar plates (Fig. 2A). In broth, the clinical isolates tightly adhere to the surfaces of culture vessels; and aggregates may be visible at the bottom of a tube. The cells express long, bundled, protein fibrils, termed "Flp fibers," which are required to form the tenacious biofilm. (A Flp fiber is composed of several Flp pili.) Propagation of adherent wild-type strains generally leads to spontaneous variants (see Section 4 below) that produce smooth colonies (Fig. 2B), have cells that do not autoaggregate, and cannot produce tenacious biofilms because the cells do not express Flp pili (Fig. 1, right) (Fine et al., 1999b; Kachlany et al., 2001a).
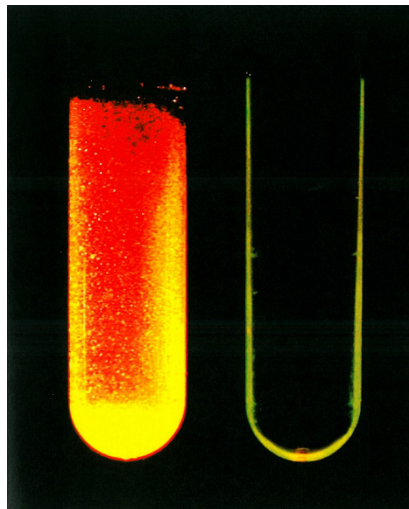
Fig. 1. Tenacious adherence of *A. actinomycetemcomitans*.
Cultures of a clinical isolate (left) and a spontaneous smooth-colony variant (right) were grown in broth. After the culture was mixed and the broth was poured out, the remaining adherent cells were stained with ethidium bromide and illuminated with UV light.

Studies in our LAP-infection model demonstrated that a wild-type strain is able to colonize and persist in the mouths of rats (Fine et al., 2001). In contrast, an isogenic smooth-colony-forming variant failed to persist. Despite this unequivocal evidence for the importance of tenacious adherence in the colonization of the oral cavity by *A. actinomycetemcomitans*, the genetic and molecular bases underlying this remarkable property were unknown.
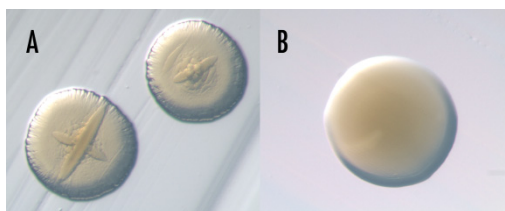
Fig. 2. Colony morphologies of *A. actinomycetemcomitans*.
(A) "Rough" colony morphology of a clinical isolate. (B) "Smooth" colony morphology of a spontaneous variant.

1.2 The *tad* locus

The study of *A. actinomycetemcomitans* had been hampered by a paucity of molecular tools for genetic manipulation, especially tools for clinical isolates. We adapted the transposon Tn*903* to provide inducible random mutagenesis of the chromosome of *A. actinomycetemcomitans* (Thomson et al., 1999). A clinical isolate of *A. actinomycetemcomitans* from a 13-year-old African-American female was mutagenized using our synthetic Tn*903* transposon (IS*903ϕkan*). The random-mutant bank was screened for mutant strains that formed smooth colonies. Identifying and sequencing the transposon insertion sites of smooth mutants led to the identification of a 14-gene segment, designated the *tad* (*t*ight *ad*herence) locus (Fig. 3A) (Kachlany et al., 2000; Planet et al., 2003; Tomich et al, 2007). The *tad* locus encodes a novel macromolecular transport system that is used in the biogenesis of Flp pili (Fig. 4). One hypothetical gene of the *tad* locus, *flp-2*, is not needed for Flp pili
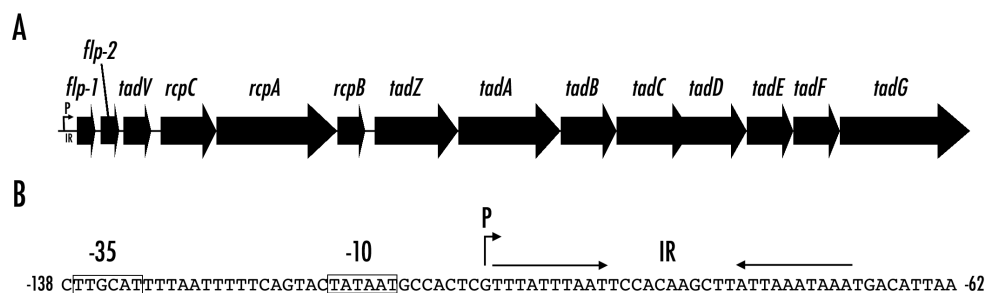
Fig. 3. The *tad* locus of *A. actinomycetemcomitans*.
(A) The size of the locus is approximately 12 kb. Filled arrows designate the genes. Their relative sizes are approximately correct. "P" and the bent arrow indicate the transcriptional start site (Haase et al., 2003); "IR," inverted repeat. (B) The promoter region magnified.

biosynthesis in *A. actinomycetemcomitans* (Perez et al., 2006). We determined whether the gene products are cytoplasmic, in the inner membrane, or in the outer membrane (Clock et al., 2008). Our studies have also revealed the functions or features of the products of several *tad* locus genes (Tomich et al., 2007). *flp-1* encodes the Flp1 prepilin (Inoue et al., 1998; Kachlany et al., 2001b); *tadV* encodes a protease that processes the Flp1 prepilin to the mature Flp1 pilin that is needed for assembly into Flp pili (Tomich et al., 2006); the product of *rcpA* (Haase et al., 1999) has the properties of an outer membrane pore (secretin) (Clock et al., 2008); the product of *rcpB* (Haase et al., 1999) is an outer-membrane protein that may gate the pore (Clock et al., 2008; Perez et al., 2006); *tadZ* encodes a protein that may localize the Tad secretion machine to a pole (Perez-Cheeks et al., 2012; Xu et al., 2012; see Section 3 below); the *tadA* product is an ATPase (Bhattacharjee et al., 2001); and the *tadE* and *tadF* products are "pseudopilins," whose functions are not known; but the pseudopilins are processed by TadV in the same way that the prepilin is (Tomich et al., 2006).
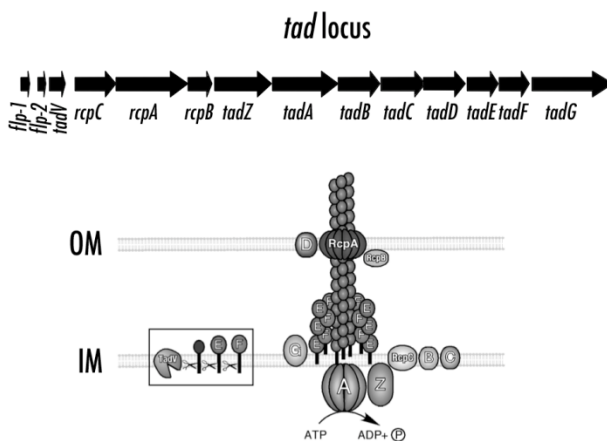
Fig. 4. Hypothetical structure of the Tad secretion apparatus.
Shown is a cartoon of the Flp pilus and its secretion machine formed (at least in part) by a complex of the *tad* gene products. The lack of detail reflects the paucity of data for the gene products. "IM" means "inner membrane"; "OM," "outer membrane."

1.3 The *tad* locus in *A. actinomycetemcomitans* is a virulence factor

Using our rat model of LAP, we demonstrated that a functional *tad* locus is needed for the ability of *A. actinomycetemcomitans* to colonize and persist in the oral cavity and to cause tissue and bone destruction (Schreiner et al., 2003). Unlike the wild-type strain, isogenic *tad* locus mutant strains, defective in either *flp-1* (encoding the pilin) or *tadA* (encoding an ATPase), did not persist in the mouths of the rats; nor did they cause bone loss. In addition to our evidence in *A. actinomycetemcomitans*, studies in other bacterial species have demonstrated the involvement of *tad* homologs in virulence: *Haemophilus ducreyi* (Spinola et al., 2003), the etiologic agent of chancroid, a sexually transmitted disease, and *Pasteurella multocida* (Fuller et al., 2000), which causes fowl cholera. The *tad* genes of *Pseudomonas aeruginosa* are needed for biofilm formation and for adherence to epithelial cells (de Bentzmann et al., 2006).

1.4 *tad* loci are widespread in prokaryotes

Searches of completed and ongoing microbial sequencing projects have revealed that closely related *tad* gene clusters are present in the genomes of a wide variety of Gram-negative and Gram-positive bacteria (Kachlany et al., 2001a; Tomich et al, 2007; P. Planet & D. Figurski unpublished results, P. Planet, unpublished results), some of which represent significant threats to human health. Among the pathogens are *Yersinia pestis*, the agent of bubonic plague; *Mycobacterium tuberculosis*, the agent of tuberculosis; *Bordetella pertussis*, the agent of whooping cough; *Burkhoderia cepacia*, a frequent colonizer of the lungs of patients with cystic fibrosis; and *P. aeruginosa*, an opportunistic human pathogen. About 40% of over 3400 bacterial genomes sequenced to date have *tad* loci. In addition, potentially all Archaea may harbor homologs of *tad* genes (P. Planet, personal communication). Given our evidence that the *tad* locus is foreign to the chromosome of *A. actinomycetemcomitans* and that *tad* loci are widely distributed in prokaryotes, we have also referred to the *tad* locus as the "Widespread Colonization Island" (WCI) (Planet et al., 2003).

2. The Flp pilus

The Flp pili of *A. actinomycetemcomitans* (Fig. 5) are proteinaceous fibers that are attached to the exterior of the bacterial cell and are necessary for tenacious adherence (Kachlany et al., 2000, 2001b). Flp pili are polymers of the mature Flp1 pilin protein, and they are assembled and secreted by a complex of proteins encoded by the *tad* locus. Flp pili are abundant, extremely adhesive, and bundled. The *flp-1* and *flp-2* genes of *A. actinomycetemcomitans* and their predicted products and the *flp* genes from other organisms and their predicted products form a distinct, monophyletic group with homology to other pilin genes, particularly to those of subclass b (Kachlany et al., 2001b). We (M. Tomich and D. Figurski, unpublished results) and others (Inoue et al., 2000) have shown that the Flp1 pilin is a glycoprotein, but the structure and function of the modification is unknown.
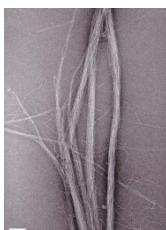
Fig. 5. Flp pili.
Transmission electron micrograph of purified Flp pili from *A. actinomycetemcomitans.*

After their translocation to the inner membrane, the prepilins are cleaved (processed) to maturity by cognate prepilin peptidases (Giltner et al., 2012). We have shown that the 75-amino acid Flp1 prepilin of *A. actinomycetemcomitans* is cleaved by TadV protein at the sequence G^XXXXEY (Tomich et al., 2006). The mature Flp1 pilin is only 49 amino acids in length, which is much smaller than other known type IV pilins (Giltner et al., 2012; Kachlany et al., 2001b). Because of its small size, we believe Flp1 pilin is an attractive subject for genetic and structural analysis. In addition to learning about the molecular details of Flp1, we also wish to understand the basis of the three most obvious phenotypes of Flp pili:

tenacious adherence to surfaces, binding to *A. actinomycetemcomitans* cells, and binding to each other (bundling).

2.1 Alanine-scanning mutagenesis of the coding region for the mature Flp1 pilin

To begin to study the properties of Flp1, we constructed and characterized a series of Flp1 pilin mutants, each with an alanine substitution for a specific non-alanine residue of the mature Flp1 pilin. The codon for each non-alanine residue was changed to a codon for alanine. (The mutant genes were constructed with the fewest possible nucleotide changes.) In this way, translation of the mutant gene would give a mutant Flp1 prepilin, which, after being processed, gave rise to a mutant mature Flp1 pilin. (Alanine was chosen because it is the smallest amino acid that is relatively neutral and can maintain an α-helix in a polypeptide.) We changed the non-alanine residues in the mature Flp1 pilin by overlap extension PCR (polymerase chain reaction) (Ho et al., 1989). In this method, the *flp-1* gene was divided into two segments, each of which could be amplified by PCR with a pair of primers. For each segment, one primer annealed just beyond an end of the gene; the other primer was directed to the internal region of the gene where the mutation was to be introduced. The internal primers for the two segments carried the appropriately changed nucleotides. As a result, the two amplified segments overlapped slightly. When the two segments were added together, denatured, and reannealed, a single strand of one segment could anneal in the overlap region with the opposite strand of the other segment. One half of the hybrid molecules would have free 5′ ends facing the single-stranded portion. These are dead-ends. The other half would have free 3′ ends, which could prime DNA synthesis to give full-length duplex molecules with the mutation in both strands. By using the end-specific primers, the full-length, mutated *flp-1* gene could be amplified and cloned. The mutation was then confirmed in the clone by nucleotide sequencing.

2.2 Characterization of mutant Flp1 pilins

The mutant genes were inserted into a plasmid vector downstream of the IPTG (isopropyl β-D-thiogalactopyranoside)-inducible *tac* promoter. (The *tac* promoter is a strong promoter that is a hybrid of the *trp* and *lac* promoters. Like the *lac* promoter, the *tac* promoter is inhibited by the *Escherichia coli* LacI repressor protein, whose gene was already added to the plasmid cloning vector. LacI is inactivated by IPTG.) We wanted to know (1) if the mutant pilin had wild-type abundance, (2) if the mutant *flp-1* gene could complement a *flp-1* chromosomal mutant gene, (3) if the mutant pilin allowed Flp pili to be made, and (4) if any Flp pili assembled with the mutant pilin promoted adherence. Mutant Flp1 pilin abundance was indicated by immunological detection of pilin in protein extracts that were separated into bands by electrophoresis in a sodium dodecyl sulfate polyacrylamide gel (Western blot). Genetic complementation was detected by the conversion of the smooth-colony morphology of a *flp-1⁻* mutant strain to a rough-colony morphology reminiscent of the wild-type strain. The presence of Flp pili was shown by electron microscopy of *flp-1⁻* mutant cells carrying the mutant pilin gene. To quantify adherence, a slightly modified crystal violet assay of O'Toole and Kolter (1998) was used. Basically, wells of microtiter dishes were inoculated with the strains to be tested. Nonadherent cells were removed by washes. Adherent cells were then stained with crystal violet. After more washes to remove free crystal violet and any remaining nonadherent cells, the crystal violet in the adherent cells

was then eluted with DMSO (dimethyl sulfoxide). The eluted crystal violet gave a color to the solution that was quantified with a spectrophotometer. Increasing crystal violet in the eluate is indicative of increasing adherence.

Each of our Flp1 mutants had one non-alanine residue changed to alanine. (See Table 1 for the single-letter and three-letter codes for the amino acids. Table 2 shows the residue change in the Flp1 pilin for each mutant and the phenotype.) In the mutants, every non-alanine residue of the mature Flp1 pilin was changed to alanine. The Flp1 prepilin has 75 residues; but, after cleavage, the mature Flp1 pilin has 49 residues. Nine residues are already alanine. The other 40 residues of the mature pilin were changed to alanine. The small size of mature Flp1 pilin made it reasonable to create this series of mutant pilins.

Amino acid	Single-letter code	Three-letter code
alanine	A	ala
arginine	R	arg
asparagine	N	asn
aspartic acid	D	asp
cysteine	C	cys
glutamic acid	E	glu
glutamine	Q	gln
glycine	G	gly
histidine	H	his
isoleucine	I	ile
leucine	L	leu
lysine	K	lys
methionine	M	met
phenylalanine	F	phe
proline	P	pro
serine	S	ser
threonine	T	thr
tryptophan	W	try
tyrosine	Y	tyr
valine	V	val

Table 1. Single-letter and three-letter codes for the amino acids.

Mutant Residue	Colony Morph.[a]	Adher.[b]	Auto-aggreg.[c]	Protein exp.[d]	Piliation[e]	Class[f]
V27A	S	-	-	+/-	+	IV
T28A	S	-	-	+	-	III
I30A	S	-	-	+	+	IV
E31A	S	-	-	+	-	III
Y32A	S	-	-	+/-	-	III
G33A	R	+	+	ND	ND	I

Mutant Residue	Colony Morph.[a]	Adher.[b]	Auto-aggreg.[c]	Protein exp.[d]	Piliation[e]	Class[f]
L34A	S	-	-	+/-	+	IV
I35A	S	-	-	-	-	II
I37A	S	-	-	+	-	III
V39A	R	+	+	ND	ND	I
V41A	R	+	+	ND	ND	I
L42A	R	+	+	ND	ND	I
I43A	S	-	-	+	+	IV
V44A	R	+	+	ND	ND	I
V46A	S	-	-	+	+	IV
F47A	S	-	-	+	-	III
Y48A	R	+	+	ND	ND	I
S49A	R	+	+	ND	ND	I
N50A	S	-	-	+	-	III
N51A	R	+	+	ND	ND	I
G52A	S	-	-	+	+	IV
F53A	R	+	+	ND	ND	I
I54A	S	-	-	-	-	II
N56A	R	+	+	ND	ND	I
L57A	S	-	-	+	+	IV
Q58A	R	+	+	ND	ND	I
S59A	R	+	+	ND	ND	I
K60A	R	+	+	ND	ND	I
F61A	S	-	-	+	-	III
N62A	R	+	+	ND	ND	I
S63A	R	+	+	ND	ND	I
L64A	S	-	-	+	-	III
S66A	R	+	+	ND	ND	I
T67A	R	+	+	ND	ND	I
V68A	R	+	+	ND	ND	I
S70A	R	+	+	ND	ND	I
N72A	R	+	+	ND	ND	I
V73A	S	-	-	+/-	-	III
T74A	R	+	+	ND	ND	I
K75A	S	-	-	+/-	+	IV

Table 2. Flp1 mutants and phenotypes.
[a] R, rough colony morphology; S, smooth colony morphology; [b] +, ≥60% wild-type adherence; -, <60% wild-type adherence; [c] +, autoaggregation; -, no autoaggregation; [d] +, wild-type Flp1 level; +/-, intermediate Flp1 level; -, no Flp1 observed; [e] +, pili observed; -, no pili observed; ND = not determined; [f] I-IV, phenotypic class

The mutant pilins were assayed for the phenotypes described above. We divided the mutant pilins into four phenotypic classes (Tables 2 and 3, Fig. 6). Class I Flp1 pilin mutants (21 in number) were indistinguishable from the wild-type pilin in our assays. In other *in vitro* assays or in animal experiments, some of these Class I mutants might show phenotypes that differ from those of the wild-type pilin. Six of the seven residues in the hydrophobic-region Class I mutants were originally glycine (1), valine (4), and leucine (1), all of which are similar to the alanine replacement. The seventh mutant in the hydrophobic region had alanine in place of tyrosine, a larger hydrophobic residue. There were two Class II mutants defined as showing no or very little pilin. Class III (9) and Class IV (8) mutants all showed abundant pilin, approximately equal to the abundance of the wild type pilin. However, electron micrographs showed that Class III pilins do not form pili. In contrast, Class IV mutant pilins could be assembled into pili; but the pili were nonadherent. Consequently, Class IV mutant cells did not autoaggregate, nor could a Class IV mutant strain form the tenacious Tad biofilm. One of the Class IV mutants (K75A) is particularly interesting because it produces curved and non-bundled pili.
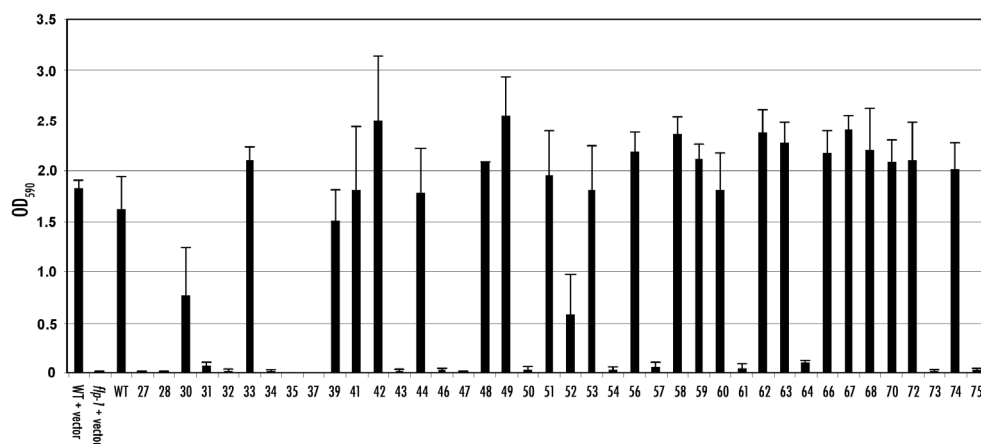
Fig. 6. Effect of mutant Flp1 pilins on adherence by *A. actinomycetemcomitans*.
Shown are the results of crystal violet assays for adherence. The first three results on the left are the controls: (1) positive controls, which are the wild-type *A. actinomycetemcomitans* strain (WT) with and without the empty plasmid used for cloning (vector) and (2) a negative control, which is a Flp1 pilin mutant strain (*flp-1*) with the empty vector. The mutant strain does not show tenacious adherence because it cannot make Flp pili. The numbers on the X-axis indicate the residues that were changed to alanine. The OD_{590} values on the Y-axis indicate relative amounts of crystal violet eluted from adherent cells (see text). The error bars show the calculated ranges of values from three experiments.

Because there is no 3D structure yet for Flp1 pilin, we used structure-prediction software with the sequence. Like other type IV pilins, the mature Flp1 pilin was predicted to be divided into a largely hydrophobic N-terminal domain and a distinct C-terminal domain. In Flp1, the C-terminal domain was predicted to contain an 11-residue amphipathic α-helix and a 12-residue, mostly-polar, C-terminal tail. It is thought that pilin subunits interact for polymerization in the hydrophobic region (Giltner et al., 2012). We therefore expected the

"assembly" mutants (Class III) to be caused by alanine substitutions of residues predominantly in the hydrophobic N-terminal region. Conversely, the C-terminal region of a pilin is thought to interact with the environment (Giltner et al., 2012), so we expected Class IV mutants to occur mostly from alterations of C-terminal residues. However, we were somewhat surprised by the number of Class III mutants with a substitution in a C-terminal residue, and by the number of Class IV mutants with a change in an N-terminal residue.

Mutant Class	Adherence	Protein Expression	Piliation	Pilus Morphology
I	+	ND	ND	ND
II	-	-	ND	ND
III	-	+	-	NA
IV	-	+	+	Normal and altered

NA, "not applicable"; ND, "not determined."

Table 3. Summary of Flp1 mutant classes.

A previous study indicated that seven serine and asparagine residues in the C-terminal region are modified (Inoue et al., 2000). We found that only one substitution of those seven residues, a Class III mutant (N50A), gave a mutant phenotype. We do not know if the defect is due to the loss of modification at this residue or to the change in that amino acid.

Alanine-scanning mutagenesis has been an important step in beginning to understand the Flp1 pilin. We now have a collection of mutant pilins that can be studied for information on pilin stability, pili assembly, pili bundling, and pili adherence. These mutants will guide experiments in which substitutions can be made with amino acids that are of different sizes, have similar properties, or have very different properties. Similar experiments can also be done at the alanine residues.

A 3D structure of the Flp1 pilin is needed. There are a few structures of type IV pilins, most of which were determined from crystals formed after removing the N-terminal hydrophobic domain (Giltner et al., 2012). The Flp1 structure will clearly be different from the few structures of other type IV pilins. For example, Flp1 is 2 to 3-fold smaller than other type IV pilins. Also the mature Flp1 pilin has no cysteine residues, which are thought to be needed to form a disulfide bond to make the D region structure that seems to be conserved in type IV pilins. The phenotypes of the Flp1 mutants have also underscored the differences of Flp1 pilin and the "typical" type IV pilins. A structure would help us to understand (1) what is truly "typical" in type IV pilins, (2) the importance of certain Flp1 residues, and (3) the molecular basis of the phenotypes of the mutants.

3. TadZ

The *tadZ* gene has no known homolog, and it is unique to *tad* loci. The presence of a *tadZ* homolog is taken to indicate that a series of genes is, is part of, or once was a *tad* locus. In the *tad* locus of *A. actinomycetemcomitans*, *tadZ* is essential for the biogenesis of Flp pili and, therefore, for the tenacious biofilm. Our recent fluorescence-microscopy study of a TadZ-

EGFP (enhanced green fluorescent protein) fusion showed that TadZ protein localizes to the old cell pole (*i.e.*, opposite the pole formed by cell division) (Fig. 7) (Perez-Cheeks et al., 2012). It localized in the absence of any other protein encoded by the *tad* locus. (The TadZ-EGFP fusion was formed by eliminating the translational stop codon of the *tadZ* gene and attaching it in-frame to the coding sequence for EGFP minus its ribosome-binding-site and translational start codon.) In contrast, a TadA-EGFP fusion also localized to a pole, but only when TadZ protein was present. We proposed that TadZ is responsible for localizing the entire Tad secretory apparatus to a pole (Perez-Cheeks et al., 2012).

We did a large phylogenetic analysis and showed that the *tadZ* genes belong in the *parA*/*minD* superfamily of genes (Perez-Cheeks et al., 2012). The prototypical bacteriophage P1 ParA protein (and each of the various ParA-like proteins) is needed for the proper segregation of DNA at cell division. The *E. coli* MinD protein, which is thought to be the prototype for the other MinD-like proteins encoded by the family, has been studied extensively. It is needed to localize the septum properly at cell division. Other gene families in the *parA*/*minD* superfamily are the *nifH* family (named for nitrogen fixation), the *fleN* family (named for flagellar synthesis), and the *bcsQ* family (named for cellulose biosynthesis).
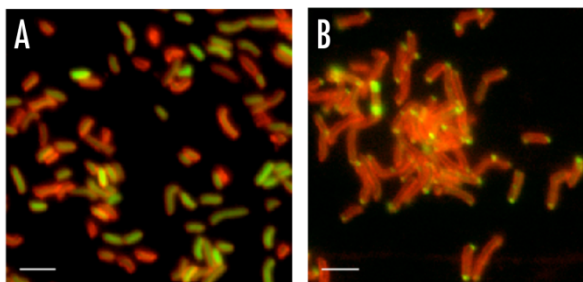
Fig. 7. Polar foci of the TadZ-EGFP fusion protein.
Shown are micrographs of *tadZ* mutant cells of *A. actinomycetemcomitans* slightly elongated by treatment with a sub-inhibitory concentration of the antibiotic piperacillin, stained with the membrane-specific fluorescent dye TMA-DPH [1-(4-trimethylammoniumphenyl)-6-phenyl-1,3,5-hexatriene p-toluenesulfonate] (fluoresces red), and containing a plasmid expressing (A) green fluorescent protein as a control or (B) a TadZ-EGFP fusion. The cells were illuminated with UV light. Polar foci are seen only in the cells expressing the TadZ-EGFP fusion protein (Perez-Cheeks et al., 2012).

3.1 The atypical Walker-like A box of TadZ proteins

Each of the protein products of the *parA*/*minD* superfamily of genes has a Walker-like A box [KGGXX(S/T)] (Fig. 8), which forms a structure involved in binding and hydrolyzing ATP. The Walker-like A box is a variation of the Walker A box [GXXGXGK(S/T)], which also allows a protein to bind and hydrolyze ATP. However, the ATPases of Walker A box proteins are considerably stronger than the ATPases of the Walker-like A box proteins. Each product of the *tadZ* gene family also has a Walker-like A box, but the TadZ Walker-like A box is unique. The second lysine (K6 in the numbering system of Fig. 8) is missing. We call

the unique Walker-like A motif in TadZ proteins the "atypical Walker-like A box" (Perez-Cheeks et al., 2012).

Fig. 8. Walker-like A boxes of proteins from the *parA*/*minD* superfamily.
Ec, *Escherichia coli*; P1, bacteriophage P1; *Aa*, *Aggregatibacter actinomycetemcomitans*;
Cc, *Caulobacter crescentus*

3.2 Phenotypes of mutants altered in the atypical Walker-like A box of TadZ

We wanted to know the effect of changing the residue in position 6 of the atypical Walker-like A box from alanine in the TadZ protein of *A. actinomycetemcomitans* (AaTadZ) to the lysine residue of the canonical Walker-like A box. We used overlap extension PCR (Section 2.1) to change the codon in *tadZ* for residue 155 of the AaTadZ protein. We cloned the mutant gene into a plasmid vector and expressed it from the *tac* promoter, as described in Section 2.2. The mutant and wild-type proteins were found to be equally abundant in cells. After we introduced the plasmid into a *tadZ* mutant strain, we assayed three phenotypes of *tad* locus function: (1) change in colony morphology of a *tadZ*⁻ mutant from smooth (the mutant phenotype) to rough (the wild-type phenotype), (2) autoaggregation of cells, and (3) formation of the tenacious biofilm. Whereas wild-type AaTadZ protein was completely functional in all three phenotypes, the A155K mutant protein was completely defective in all
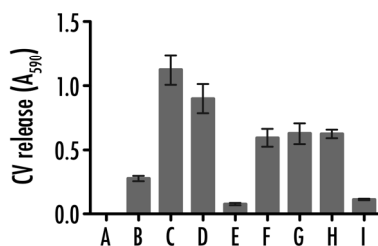
Fig. 9. Adherence of *A. actinomycetemcomitans* strains expressing the wild-type TadZ protein and mutants of TadZ.
Adherence was shown by the crystal violet (CV) assay described in the text. The Y-axis is the same as in Fig. 6. A-D are controls: (A) media alone; (B) the *tadZ*⁻ mutant strain with an empty vector plasmid; (C) a wild-type strain with vector; and (D) the *tadZ*⁻ mutant strain with a plasmid carrying *tadZ*⁺. E-I are *tadZ*⁻ strains expressing mutant TadZ proteins: (E) TadZ A155K, (F) TadZ K150R, (G) TadZ K150A, (H) TadZ S156T, and (I) TadZ S156A. The error bars show the calculated ranges of values.

three phenotypes (see Fig. 9 for the adherence result). Therefore, the presence of alanine at position 6 of the Walker-like A box is essential for AaTadZ function.

We noticed that a common feature of the atypical Walker-like A boxes from other TadZ proteins was the absence of lysine at Walker-like box position 6, not the presence of alanine. The *tadZ* genes from various *tad* loci had codons for other amino acids at position 6.

We confirmed this property for AaTadZ. When we mutated the *tadZ* gene to substitute a glycine, valine, asaparagine, or serine residue for alanine in the protein, all the mutant proteins were functional.

Even though it seemed that the absence of lysine was the primary requirement for the position 6 residue of the Walker-like A box of TadZ proteins, other residues of the Walker-like A motif are conserved. This observation indicated that the other residues in the Walker-like A boxes of TadZ proteins are important for function. To test this, we made mutants of AaTadZ in which the lysine residue at Walker-like A box position 1 (AaTadZ K150) was changed to arginine or alanine. Likewise, we made mutants at position 7 (AaTadZ S156) with threonine or alanine in place of the conserved serine residue. The mutant AaTadZ proteins did not allow wild-type biofilm formation (Fig. 9). Strains with the K150R, K150A, and the S156T mutants showed some biofilm formation; but it was reduced relative to wild type. The S156A mutant was completely unable to adhere. We concluded that the other residues of the atypical Walker-like A box of TadZ proteins are important to a function leading to tenacious adherence and biofilm formation.

The MinD proteins from *E. coli* (Hu & Lutkenhaus, 1999; Raskin & de Boer, 1999) and *Neisseria gonorrhoea* (Ramirez-Arcos et al., 2002), the ParA protein from plasmid pB171 of *E. coli* (Ebersbach & Gerdes, 2002), and the ParA-like Soj protein of *Bacillus subtilis* (Marston & Errington, 1999; Quisel et al., 1999) form polar foci and oscillate from pole to pole in the cell. The mobility of MinD-like and ParA-like proteins depends on their ATPase activity (Lutkenhaus & Sundaramoorthy, 2003). Changing a residue in the Walker-like A box of ParA from pB171 (Ebersbach & Gerdes, 2001), of the MinD protein from *N. gonorrhoea* (Ramirez-Arcos et al., 2002), or of Soj (Quisel et al., 1999) causes the protein to lose its mobility. Our TadZ-EGFP experiments have led us to suggest that TadZ positions the Tad secretion apparatus to a pole. Our hypothesis requires that TadZ does not oscillate. Maybe AaTadZ is a natural variant selected to allow it to remain stationary at a pole. The rest of the Walker-like A box may be needed to bind ATP, which was found in crystals of TadZ from *Eubacterium rectale* (Xu et al., 2012). The binding of ATP is not for the polar localization nor for dimer formation (see Sections 3.3 and 3.4 below). We note that ATP binding is necessary for the interaction of MinD with MinC and MinE (Hayashi et al., 2001; Hu et al., 2002), of Soj with Spo0J (Marston & Errington, 1999; Quisel et al., 1999), and of bacteriophage P1 ParA with ParB (Bouet & Funnell, 1999). Perhaps ATP is needed for TadZ to interact with one or more proteins of the Tad secretion apparatus.

3.3 Mutants altered in the atypical Walker-like A box of TadZ localize properly

We used PCR to amplify the *tadZ* genes that encoded the atypical Walker-like A box mutant proteins. The amplified mutant genes were used to make mutant TadZ-EGFP fusions, as described for the wild-type gene in Section 3. All of the mutant fusion proteins formed normal-appearing fluorescent foci at the old poles of cells. The absence of a defect in focus

formation or location was particularly striking for the A155K and S156A mutants. A mutant TadZ protein with either allele was completely defective in the phenotypes we assayed (Fig. 9 and Section 3.2). In fact, the mutant fusions consistently showed a higher number of cells with polar foci than was seen with the wild-type TadZ-EGFP fusion. We do not understand the significance of the increase, but we concluded that the residues of the atypical Walker-like A box are not needed for the polar localization of TadZ.

3.4 Mutants altered in the atypical Walker-like A box of TadZ form dimers

We used a bacterial reporter strain to indicate TadZ dimer formation *in vivo* (Hu et al., 2000). The basis of the reporter is the following. Bacteriophage λ cI protein represses the λ p_R promoter. In the reporter strain, the *E. coli lacZ* gene, which encodes β-galactosidase, was fused to p_R. Because there is a convenient colorimetric assay for β-galactosidase, expression of *lacZ* from the synthetic *p_R-lacZ* operon can be measured. The level of β-galactosidase is a function of p_R activity and is, therefore, an indication of cI activity. cI protein is a dimer (Pabo et al., 1979). Each monomeric polypeptide has an N-terminal domain for DNA binding and a C-terminal domain for dimerization. If the C-terminal domain is removed, the N-terminal DNA-binding domain (λ cIDB) is unable to function because it cannot form dimers. Fusing a polypeptide that can dimerize to λ cIDB can make a functional repressor and reduce expression of *p_R-lacZ*.

We created a chimeric gene that encoded a fusion of the coding regions for λ cIDB and TadZ. The product of the fusion repressed *p_R-lacZ* and indicated that TadZ can form dimers (Table 4) (Perez-Cheeks et al., 2012). Knowing that TadZ can dimerize, we then asked if the atypical Walker-like A box mutants are proficient or defective in the dimerization activity (Perez-Cheeks et al., 2012). To do this, it was necessary to make chimeras of the coding region for λ cIDB and the mutant *tadZ* genes. Each fusion protein was then tested in the reporter strain for the ability to repress *p_R-lacZ*. Each mutant fusion protein made a functional repressor that was as good as the fusion with wild-type TadZ (Table 4). We concluded that individual residues of the atypical Walker-like A box of TadZ are not needed for dimer formation.

Protein	% repression of *p_R-lacZ*
Empty vector	0.0 ± 0.0
λcIDB	9.7 ± 4.0
λcI	66.0 ± 1.4
λcIDB-TadZ	60.1 ± 0.7
λcIDB-TadZ K150R	49.0 ± 1.9
λcIDB-TadZ K150A	60.3 ± 1.5
λcIDB-TadZ A155K	49.7 ± 2.1
λcIDB-TadZ S156T	53.7 ± 0.8
λcIDB-TadZ S156A	57.6 ± 4.2

Table 4. *In vivo* dimerization assay results for wild-type TadZ and mutant TadZ proteins.

4. Regulation of the *tad* locus

Logic and evidence indicate that the expression of the *tad* locus genes in *A. actinomycetemcomitans* is controlled. The Tad⁻ variants that spontaneously arise from Tad⁺ parents form larger colonies on solid medium, grow to a higher density in liquid medium, and have a shorter generation time than do their parents. We conclude that it is energetically expensive for a cell to make Flp pili. Therefore, it would be advantageous to the cell for *tad* locus expression to be regulated. One way to accomplish this is to control *tad* gene transcription.

There is evidence for transcriptional regulation of *tad* loci. *P. aeruginosa* has a *tad* locus, and it is regulated by the PprA-PprB two-component system (Bernard et al., 2009). The *pprA* and *pprB* genes map within a locus that has five contiguous, but divergent, transcriptional units. Four encode the *tad* genes and *pprA*; the fifth encodes *pprB* only. After being activated by the histidine kinase (PprA), the response regulator (PprB) binds to the promoters and activates transcription of the *tad* genes. In another example, the expression of the *tad* genes (the *flp* operon) in the human pathogen *Haemophilus ducreyi* is affected by the CpxR-CpxA two-component system (Labandeira-Rey et al., 2010). Overexpression or constitutive activation of the response regulator CpxR causes repression of *flp* operon transcription and a reduction in the level of Flp1 protein. The authors suggested that activated CpxR directly represses transcription of the *flp* operon because CpxR bound to a target in the *flp* promoter region.

For *A. actinomycetemcomitans*, Kaplan et al. (2003) have shown the presence of probable nonadherent cells within a Tad⁺ colony. They found sequestered, loosely packed, non-aggregating, and probably nonadherent cells in an adherent colony of Tad⁺ cells. The authors suggested that transiently nonadherent cells are produced by the colony as part of a developmental pathway to expand the biofilm. Temporary nonadherence would allow cells to escape and seed new adherent colonies that are needed for the biofilm to spread. One possibility is that the nonadherent and non-autoaggregating cells lack Flp pili due to the inhibition of *tad* gene transcription. Indeed, there is a provocative 31-bp (base-pair) inverted repeat (IR) adjacent to the *tad* promoter (Fig. 3B) (G. Hovel-Miner, P. Planet, and D. Figurski, unpublished results). (The 31-bp IR has a spacer of 11 bp flanked by 10-bp arms that are inverted complementary repeats of each other.) The IR is conserved in all six serotypes of *A. actinomycetemcomitans*, hinting that it is important to this organism. However, the function of the IR is currently unknown. Because the *tad* promoter is very strong (Kram et al., 2008), if the IR is a binding site for a protein that regulates transcription, it seems likely that the protein that binds IR would be a repressor that reduces *tad* transcription.

Our studies have indicated that *tad* locus transcription is regulated by a termination cascade to maintain the correct stoichiometry of the *tad* gene products (Kram et al., 2008). We isolated three transcriptional terminators (T). T1 and T3 are factor-independent terminators, whereas T2 is a Rho-dependent terminator. RNA sizes and results from a *lacZ* transcriptional reporter indicated that T1 accounts for ~99% termination and is located after the pilin gene, *flp-1*, and before *flp-2*. T2 is located between *tadV* and *rcpC* and seems to terminate ~36% of the remaining transcripts in our assay. T3 terminates transcription after *tadG*, *i.e.*, at the end of the locus opposite the promoter. Rough RNA quantitation indicates that for every full-length *tad* transcript, there are ~1.5 *flp-1*/*flp-2*/*tadV* transcripts and ~160 *flp-1* transcripts. This means that, assuming all *tad* mRNAs are translated with equal

efficiencies, for every Tad secretion apparatus, there are about 50% more TadV protease molecules to process the abundance of Flp1 prepilin protein.

4.1 Tad⁺ and Tad⁻ bacteria make different biofilms

Biofilm formation depends on adherence. Biofilms with different characteristics may indicate different modes of adherence. Tad+ *A. actinomycetemcomitans* synthesize a pili-based, resilient biofilm. Tad- variants have been reported to form weak biofilms (Inoue et al., 2003). We noticed that cells of Tad- variants displayed adherence when the culture was handled gently (Fig. 10). Most or all of the biofilm of a Tad- strain is lost under the conditions we use to allow the biofilm of a Tad+ strain to remain intact. Three-dimensional light microscopy showed that the biofilm of a Tad- strain is very different from the biofilm of its Tad+ parent. Biofilms of Tad+ strains showed distinct, tightly packed microcolonies of cells. In contrast, the biofilms of Tad- strains showed loosely packed cells in an extracellular matrix that stained readily with DAPI (4′,6-diamidino-2-phenylindole, a fluorescent DNA stain). The biofilm of the Tad- strain showed cells in structures that were interpreted to be columns and mushroom shapes. We suggested that the Tad- biofilm resembles a "typical" biofilm.
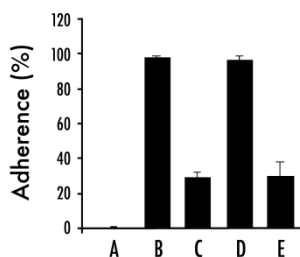
Fig. 10. Adherence of isogenic Tad+ and Tad- strains of *A. actinomycetemcomitans*.
Shown are results from the crystal violet assay for adherence (see text) for two isogenic pairs of *A. actinomycetemcomitans* strains (B, C and D, E). A is a no-cells control. B and D are the Tad+ parents: DF2200N (serotype a) and CU1000N (serotype f). The two spontaneous Tad- (smooth-colony) variants are DF2261N (C) and CU1060N (E). The adherence of DF2200N is taken as 100%. The error bars show the Standard Deviation calculated from the results of three experiments. Some adherence of the variants can be detected in conditions that favor the tenacious adherence of Tad+ strains.

4.2 Choosing proteins that may be needed for adherence to inert surfaces

We sought to find a non-*tad*-locus protein needed for adherence in *A. actinomycetemcomitans*. We identified what we thought were prime candidates for adherence-essential proteins in *A. actinomycetemcomitans* (Table 5). We considered two-component systems because the two known examples of *tad* regulation by a non-*tad* determinant involve two-component systems. Only four two-component systems are predicted to be encoded by the genome of *A. actinomycetemcomitans*: ArcAB, CpxRA, NarPQ, and QseBC. We also selected two other proteins that we thought might be required for adherence: OxyR and PgaC. OxyR was chosen because *Aggregatibacter aphrophilus*, a bacterium closely related to *A.*

actinomycetemcomitans, has a nearly identical *tad* locus (P. Planet, C. Sheth, and D. Figurski, unpublished results; Di Bonaventura et al., 2009). *A. aphrophilus* makes more Flp pili in higher oxygen than in lower (S. Kachlany, C. Sheth, and D. Figurski, unpublished results). Therefore, we selected OxyR, a transcriptional activator that responds to oxygen (Bauer et al., 1999), because the *tad* genes in this bacterium appear to respond to the presence of oxygen. In addition, there is evidence that OxyR induces the adhesin *apiA* in *A. actinomycetemcomitans* (Ramsey & Whiteley, 2009). PGA (poly-β-1,6-*N*-acetyl-D-glucosamine) is an extracellular polymer that is synthesized by *A. actinomycetemcomitans* and is associated with the biofilm of Tad+ cells (Izano et al., 2008). We considered PgaC, an enzyme in the pathway for PGA synthesis, to be a candidate for a protein that affects a biofilm in *A. actinomycetemcomitans*.

Gene	Function
arcB	Two-component sensor kinase; response to anaerobic and aerobic stress
cpxR	Two-component response regulator; response to outer membrane stress
narP	Two-component response regulator; response to nitrate and nitrite stress
qseB	Two-component response regulator; response to quorum sensing
oxyR	Transcriptional regulator; response to redox stress
pgaC	Integral membrane glycosyltransferase; required for PGA biosynthesis

Table 5. Selected putative regulators of adherence by *A. actinomycetemcomitans*.

4.3 Mutagenesis of possible adherence-required genes by allelic exchange

We constructed a series of six mutant Tad+ strains and six mutant Tad- strains. Each strain had a mutation in one of our six candidates genes (Section 4.2). We asked if any of the mutants was defective in the strain's biofilm. The mutant strains were made by allelic exchange, *i.e.*, the mutant gene was substituted for the wild-type gene in the chromosome (Fig. 11). In allelic exchange, the mutant gene is marked (usually with an antibiotic resistance gene) and introduced into a wild-type strain. Homologous recombination allows the mutant gene and its marker to replace the wild-type gene in the chromosome. If the gene is not essential for viability, strains in which the exchange has taken place can be selected because they are able to grow in the presence of the marker antibiotic. Strains that have not done the exchange will not grow in the presence of the antibiotic because the introduced DNA cannot replicate.

We used plasmid pMB78 (Bhattacharjee et al., 2007), a "suicide" plasmid, *i.e.*, one that can replicate or be maintained in one host and not in another. Plasmid pMB78 can be maintained in *E. coli*, but not in *A. actinomycetemcomitans*. The wild-type gene was amplified from the chromosome of *A. actinomycetemcomitans* and cloned into pMB78 in *E. coli.* To construct the mutant allele, the internal part of the wild-type gene was deleted and replaced with the gene for kanamycin resistance. Plasmid pMB78 has the uptake signal sequence (USS) needed for transformation of *A. actinomycetemcomitans.* After transformation and recombination, the recombinants (*i.e.*, the mutant strains) were selected by their growth on medium containing kanamycin. The presence of the mutant alleles was confirmed by PCR and gel electrophoresis (Fig. 12A).
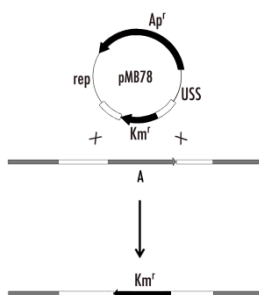
Fig. 11. Schematic for allelic exchange.
pMB78 is an ampicillin resistance (Ap[r])-encoding suicide plasmid that cannot be maintained in *A. actinomcetemcomitans* (see text) (Bhattacharjee et al., 2007). The wild-type gene is cloned and mutated by insertion of DNA encoding (in this example) kanamycin resistance (Km[r]). After recombination ("X") in the homologous regions (white boxes), the mutated gene replaces the wild-type gene in the chromosome. The mutant strains (recombinants) can grow in the presence of kanamycin. "rep" is "replication region"; "USS," "uptake signal sequence," which is needed for transformation of *A. actinomycetemcomitans* (Thomson et al., 1999).
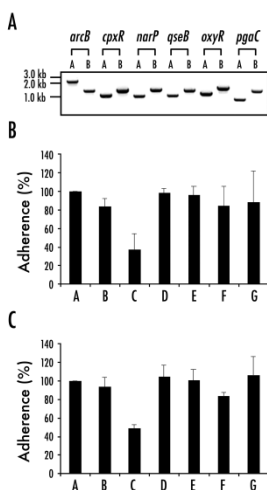
Fig. 12. Directed mutations and their effects on the adherence of Tad[+] and Tad[-] strains of *A. actinomycetemcomitans*.
Panel A shows agarose gel electrophoresis of PCR products of the wild-type gene and the mutant gene after recombination for each of the six possible regulators of adherence. Panels B and C show the results of the crystal violet assay for adherence (see text) of the isogenic Tad[+] and Tad[-] strains, respectively. In Panels B and C, adherence of the parent strain (A) is normalized to 100%. The mutant strains are defective in *arcB* (B), *cpxR* (C), *narP* (D), *qseB* (E), *oxyR* (F), and *pgaC* (G).

Adherence of the Tad+ (Fig. 12B) and Tad- (Fig. 12C) mutants was quantified by the crystal violet assay (Section 2.2). One mutation, Δ*cpxR*::Km^r ("::" indicates "insertion of"), consistently caused a significant decrease in adherence (relative to the wild-type strain) for both the Tad+ and Tad- strains (Fig. 12 B and C, respectively). This result showed that the Δ*cpxR*::Km^r mutation affected adherence in both types of strains. An important confirmation was to ask if adding the intact *cpxR*+ gene *in trans* restored adherence to the mutant strains (genetic complementation). If it did, then we could conclude that the defect in the mutant strain was the result of not having a functional CpxR protein. The *cpxR*+ gene was added *in trans* as a cloned gene on a plasmid that replicates in *A. actinomycetemcomitans.* We were surprised to learn that *cpxR*+ *in trans* restored adherence to the Tad- strain (Fig. 13A), but not to the Tad+ strain (Fig. 13B). For the Tad+ strain, genetic complementation occurred only when the complete *cpxRA* operon was added *in trans*, indicating the need for CpxA in Tad+ adherence (Fig. 13B). More needs to be done to make this a solid conclusion. However, we can conclude that both types of adherence in *A. actinomycetemcomitans* respond to a function encoded by the *cpxRA* operon. We wish to point out that adherence in *A. actinomycetemcomitans* is positively regulated by the *cpxRA* operon, and we do not know whether it functions directly or indirectly to regulate adherence. In contrast, the activation of CpxR in *H. ducreyi* leads the molecule to bind to the *tad* (*flp* operon) promoter as a negative regulator (Labandeira-Rey et al., 2010).

5. A strategy for making precise genomic deletions: the new Vector Excision (VEX) method

In Section 4.3, allelic exchange was described as a method of exchanging a wild-type segment of the chromosome with a mutated segment. In the examples given, a single gene was mutated. Allelic exchange can also be used for more than one gene. However, because the technique usually relies on standard molecular cloning methods, allelic exchange often becomes more difficult with larger segments. (See the chapter by Gerlach et al. for a discussion of recombineering – a method of cloning that overcomes the limitation caused by the locations of restriction endonuclease cleavage sites.) Another problem can be caused by the insertion of the marker, which can affect the expression of downstream genes. A method based on site-specific recombination has been developed to remove the marker (Datsenko and Wanner, 2000).

We have developed a straightforward method for making chromosomal deletions that can be any size [one bp to several kb (kilobase pairs)] and do not affect the expression of downstream genes. The strategy is based on the Vector Excision (VEX) method that we developed (Ayres et al., 1993). With VEX, the deleted portion can also be "captured" on a self-transmissible, broad-host-range plasmid. After transfer to another bacterium, the expression and/or functions of the captured genes can be assessed in a different host (see the chapter by Wilson et al.).

The new VEX strategy is illustrated here as a deletion (~11 kb) of all the *tad* genes of *A. actinomycetemcomitans* (Fig. 14). A "double cointegrate" is generated by two cycles of homologous recombination. Essentially the double cointegrate is formed by two successive allelic exchanges (Section 4.3). The plasmid suicide vector we used was based on pBBR1MCS (Kovach et al., 1994), which replicates in *E. coli*, but not in *A.*

actinomycetemcomitans. The plasmid can be mobilized into both species by conjugation, and the USS required for transformation of *A. actinomycetemcomitans* was added along with the gene for ampicillin resistance by recombineering. Thus, the vector can be introduced into *A. actinomycetemcomitans* either by transformation (as was done here) or by conjugative mobilization (crucial for strains that transform poorly or not at all).
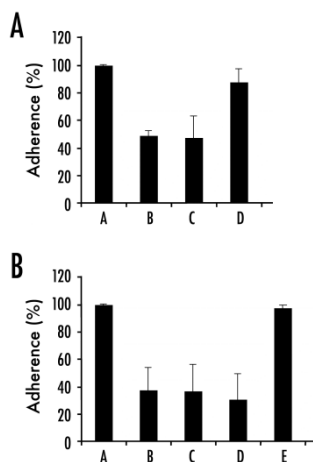
Fig. 13. Genetic complementation of the adherence defect from the *cpxR-* mutation in isogenic Tad- and Tad+ strains of *A. actinomycetemcomitans*.
Shown are crystal violet assays for adherence (see text). Panel A is the Tad- strain; panel B is the Tad+ strain. Column A in panel A is the smooth-colony variant with the wild-type *cpxR* gene. Its adherence is normalized to 100%. Column B is the *cpxR-* mutant strain; column C, the *cpxR-* mutant strain with the empty plasmid vector; and column D, the *cpxR-* mutant strain with the *cpxR+* plasmid. Column A in panel B is the Tad+ parent. B-D are analogous to those in Panel A, but with the Tad+ stain. Column E is the *cpxR-* mutant strain with a *cpxRA+* plasmid. Note that complementation of the *cpxR-* mutation in the Tad+ strain requires *cpxRA+ in trans*, whereas *cpxR+ in trans* is sufficient for complementation of the *cpxR-* mutation in the Tad- strain.

The purpose of the two homologous recombination events is to integrate two directly repeated *loxP* sites. The homology cassette determines where the *loxP* sequence is inserted into the chromosome. In our experiment, one *loxP* sequence was inserted in the *flp-1* gene; and the other, in *tadG*. The "left" (in Fig. 14) *loxP* sequence was marked by the *aacC1* gene for gentamicin resistance. The "right" *loxP* sequence was marked by *aadA*, the gene for spectinomycin resistance. Each *loxP* cassette was formed by cloning three fragments into the multiple cloning site (MCS) of the vector: (1) homology region I (HRI), (2) *loxP* and *aacC1* or *aadA*, and (3) homology region II (HRII). The fragments were generated by PCR. The 34-bp *loxP* sequence was added to the appropriate primer for fragment (2). Fragments were
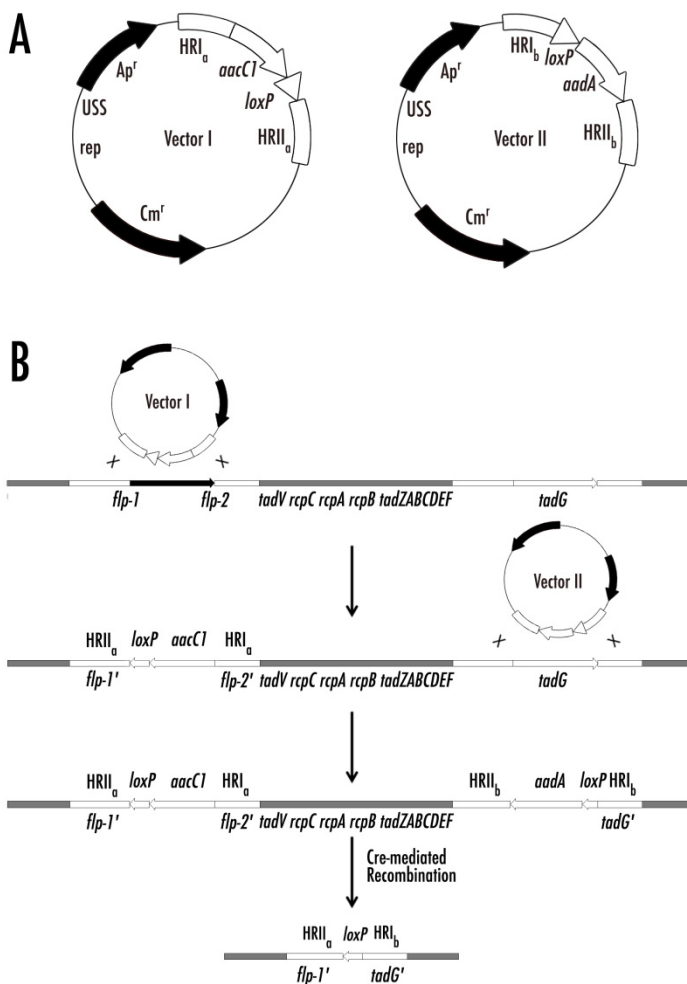
Fig. 14. Schematic for the "new VEX" method of making precise genomic deletions. Panel A shows the two plasmids that will be used in the construction of the double cointegrate. Vectors I and II are suicide plasmids that can replicate in *E. coli*, but not in *A. actinomycetemcomitans*. They are made from the same plasmid. The symbols and abbreviations are the following: the *loxP* cassettes (with cloned fragments) are in white; "Ap^r," ampicillin resistance; "Cm^r," chloramphenicol resistance; "rep," replicon; "USS," uptake signal sequence for transformation of *A. actinomycetemcomitans*; "HR," homology region; "*aacC1*," the gene for gentamicin resistance; "*aadA*," the gene for spectinomycin resistance; and "*loxP*," a 34-bp sequence that is the target for Cre-mediated site-specific recombination. Panel B depicts the two homologous recombination events and resolution by the Cre recombinase to delete the genes of the *tad* locus in *A. actinomycetemcomitans*. See the text for details.

cloned with the following restriction endonucleases: *Xba*I and *Bam*HI for fragment (1), *Bam*HI and *Sal*I for fragment (2), and *Sal*I and *Kpn*I for fragment (3). The site-specific recombinase Cre binds to a *loxP* sequence (Sternberg and Hoess, 1983), forms a synapse with another Cre-bound *loxP* sequence, and catalyzes recombination between *loxP* sequences. Cre-mediated recombination is very efficient (>95%). Directly repeated *loxP* sites cause the intervening DNA to cyclize and be deleted by the action of Cre. Inverted repeats of *loxP* sites lead to inversion of the intervening DNA. Cre was supplied to the double cointegrate by conjugation with a *cre*-encoding plasmid.

After the Cre-mediated deletion (resolution) (Fig. 15), a single *loxP* sequence is left in place of the deleted DNA (Fig. 14). In the *flp-1* to *tadG* deletion, ~11 kb of DNA was removed, whereas one 34-bp *loxP* sequence was inserted. Our results show that a double *loxP*-containing homology cassette with antibiotic resistance genes blocks transcription of a downstream gene. In contrast, the single *loxP* sequence present after resolution allows transcription and expression of the downstream gene (T. McConville and D. Figurski, unpublished results).
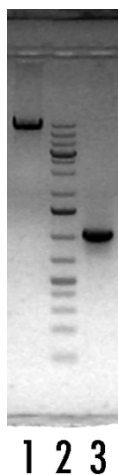
Fig. 15. Deletion of the *tad* locus of *A. actinomycetemcomitans*.
Shown is an agarose gel of PCR products that show the wild-type *tad* locus (1) and a genomic deletion of ~11 kb that removes all the genes of the *tad* locus (3). The deletion was made using the new VEX strategy shown in Fig. 14. Lane 2 shows DNA markers of different molecular weights.

6. Conclusion

The 14-gene *tad* locus of *A. actinomycetemcomitans* is needed for the synthesis and secretion of Flp pili. The ability of *A. actinomycetemcomitans* to cause periodontal disease depends on Flp pili-mediated tenacious adherence in the oral cavity. (Flp pili likely mediate colonization of

teeth.) We think we may understand the function of Flp pili in the etiology of oral disease caused by *A. actinomycetemcomitans.* However, *tad* loci occur in nearly 40% of prokaryotes. We have suggested that *tad* loci help prokaryotes colonize a niche (Kachlany et al., 2001a). For *A. actinomycetemcomitans*, a major niche is known to be the oral cavity. Tenacious adherence may be a property of *A. actinomycetemcomitans* because it must be able to colonize in the presence of extensive normal flora. Therefore, the strong phenotype of tenacious adherence may be a special property of the *tad* locus of *A. actinomycetemcomitans*. The strong phenotype allowed us to discover the *tad* genes. The phenotypes may be important, but more subtle, in other organisms.

Several proteins are unique products of *tad* loci. Such proteins may be good targets for therapeutic drugs. For example, *tadZ* is present in every *tad* locus. Our genetic studies in *A. actinomycetemcomitans* have shown that the *tadZ* gene is required for the function of its *tad* locus. We believe it is likely to be important in all *tad* loci. Therefore, a drug specific for TadZ protein might inactivate all *tad* loci. We have shown that the *tad* locus of *A. actinomycetemcomitans* is required for colonization of the mouth and periodontal disease. If we are correct that *tad* loci in other prokaryotes are needed to colonize their specific niches, inhibiting TadZ to inactivate *tad* loci may be useful in preventing and/or minimizing some diseases. Such a drug that targets a non-essential colonization factor should also be largely refractory to the selective pressure for the emergence of resistant strains.

Genetic studies of *tad* loci and the development of new genetic tools are helping us to determine and understand (1) the functions of the individual *tad* genes, (2) whether and how the various proteins encoded by a *tad* locus act to colonize a specific niche, and (3) the importance of the *tad* genes for *A. actinomycetemcomitans* and for other prokaryotes.

7. Acknowledgments

8. References

Ayres, E.K., Thomson, V.J., Merino, G., Balderes, D. & Figurski, D.H. (1993). Precise deletions in large bacterial genomes by vector-mediated excision (VEX). The *trfA* gene of promiscuous plasmid RK2 is essential for replication in several gram-negative hosts, *J. Mol. Biol*. 230, 174-185.

Bauer, C.E., Elsen, S. & Bird, T.H. (1999). Mechanisms for redox control of gene expression, *Annu. Rev. Microbiol*. 53, 495-523.

Bernard, C.S., Bordi, C., Termine, E., Filloux, A. & de Bentzmann, S. (2009). Organization and PprB-dependent control of the *Pseudomonas aeruginosa tad* Locus, involved in Flp pilus biology, *J. Bacteriol*. 191, 1961-1973.

Bhattacharjee, M.K., Kachlany, S.C., Fine, D.H. & Figurski, D.H. (2001). Nonspecific adherence and fibril biogenesis by *Actinobacillus actinomycetemcomitans*: TadA protein is an ATPase, *J. Bacteriol*. 183, 5927-5936.

Bhattacharjee, M.K., Fine, D.H. & Figurski, D.H. (2007). *tfoX* (*sxy*)-dependent transformation of *Aggregatibacter* (*Actinobacillus*) *actinomycetemcomitans*, *Gene* 399, 53-64.

Bouet, J.Y. & Funnell, B.E. (1999). P1 ParA interacts with the P1 partition complex at parS and an ATP-ADP switch controls ParA activities, *EMBO J*. 18, 1415-1424.

Clock, S.A., Planet, P.J., Perez, B.A. & Figurski, D.H. (2008). Outer membrane components of the Tad (tight adherence) secreton of *Aggregatibacter actinomycetemcomitans*, *J. Bacteriol*. 190, 980-990.

Datsenko, K.A. & Wanner, B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products, *Proc. Natl. Acad. Sci. USA* 97(12):6640-5.

de Bentzmann, S., Aurouze, M., Ball, G. & Filloux, A. (2006). FppA, a novel *Pseudomonas aeruginosa* prepilin peptidase involved in assembly of type IVb pili, *J. Bacteriol*. 188, 4851-4860.

De Rycke, J. & Oswald, E. (2001). Cytolethal distending toxin (CDT): a bacterial weapon to control host cell proliferation? *FEMS Microbiol. Lett*. 203, 141–148.

Di Bonaventura, M.P., DeSalle, R., Pop, M., Nagarajan, N., Figurski, D.H., Fine, D.H., Kaplan, J.B. & Planet, P.J. (2009). Complete genome sequence of *Aggregatibacter* (*Haemophilus*) *aphrophilus* NJ8700, *J. Bacteriol*. 191, 4693-4694.

DiFranco, K.M., Gupta, A., Galusha, L.E., Perez, J., Nguyen, T.V., Fineza, C.D. & Kachlany, S.C. (2012). Leukotoxin (Leukothera®) targets active leukocyte function antigen-1 (LFA-1) protein and triggers a lysosomal mediated cell death pathway, *J. Biol. Chem*. 287, 17618-17627.

Ebersbach, G. & Gerdes, K. (2001). The double *par* locus of virulence factor pB171: DNA segregation is correlated with oscillation of ParA, *Proc. Natl. Acad. Sci. USA* 98, 15078–15083

Fine, D.H., Furgang, D., Kaplan, J., Charlesworth, J. & Figurski. D.H. (1999a). Tenacious adhesion of *Actinobacillus actinomycetemcomitans* strain CU1000 to salivary-coated hydroxyapatite. *Arch. Oral. Biol*. 44, 1063-1076.

Fine, D.H., Furgang, D., Schreiner, H.C., Goncharoff, P., Charlesworth, J., Ghazwan, G., Fitzgerald-Bocarsly, P. & Figurski, D.H. (1999b). Phenotypic variation in *Actinobacillus actinomycetemcomitans* during laboratory growth: implications for virulence, *Microbiology* 145, 1335-1347.

Fine, D.H., Goncharoff, P., Schreiner, H., Chang, K.M., Furgang, D. & Figurski, D. (2001). Colonization and persistence of rough and smooth colony variants of *Actinobacillus actinomycetemcomitans* in the mouths of rats, *Arch. Oral Biol*. 46, 1065-1078.

Fine, D.H., Velliyagounder, K., Furgang, D. & Kaplan, J.B. (2005) The *Actinobacillus actinomycetemcomitans* autotransporter adhesin Aae exhibits specificity for buccal epithelial cells from humans and old world primates, *Infect. Immun*. 73, 1947-1953.

Fine, D.H., Kaplan, J.B., Kachlany, S.C. & Schreiner, H.C. (2006). How we got attached to *Actinobacillus actinomycetemcomitans*: a model for infectious diseases, *Periodont. 2000* 42, 114–157.

Fine, D.H., Kaplan, J.B., Furgang, D., Karched, M., Velliyagounder, K. & Yue, G. (2010). Mapping the epithelial-cell-binding domain of the *Aggregatibacter actinomycetemcomitans* autotransporter adhesin Aae, *Microbiology* 156, 3412-3420.

Fives-Taylor, P.M., Meyer, D.H., Mintz, K.P. & Brissette, C. (1999). Virulence factors of *Actinobacillus actinomycetemcomitans*, *Periodontol. 2000* 20, 136-167.

Fong, K.P., Pacheco, C.M., Otis, L.L., Baranwal, S., Kieba, I.R., Harrison, G., Hersh, E.V., Boesze-Battaglia, K. & Lally, E.T. (2006). *Actinobacillus actinomycetemcomitans* leukotoxin requires lipid microdomains for target cell cytotoxicity, *Cell Microbiol.* 8, 1753-1767.

Fuller, T.E., Kennedy, M.J. & Lowery, D.E. (2000). Identification of *Pasteurella multocida* virulence genes in a septicemic mouse model using signature-tagged mutagenesis, *Microb. Pathog.* 29, 25-38.

Genco, R.J., Christersson, L.A. & Zambon, J.J. (1986). Juvenile periodontitis, *Int. Dent. J.* 36, 168-176.

Giltner, C.L., Nguyen, Y. & Burrows, L.L. (2012). Type IV Pilin Proteins: Versatile Molecular Modules, *Microbiol. Mol. Biol. Rev.* 76, 740-772.

Haase, E.M., Zmuda, J.L. & Scannapieco, F.A. (1999). Identification and molecular analysis of rough-colony-specific outer membrane proteins of *Actinobacillus actinomycetemcomitans*, *Infect. Immun.* 67, 2901-2908

Haase, E.M., Stream, J.O. & Scannapieco, F.A. (2003). Transcriptional analysis of the 5' terminus of the *flp* fimbrial gene cluster from *Actinobacillus actinomycetemcomitans*, *Microbiology* 149(Pt 1), 205-215.

Hayashi, I., Oyama, T. & Morikawa, K. (2001). Structural and functional studies of MinD ATPase: implications for the molecular recognition of the bacterial cell division apparatus, *EMBO J.* 20, 1819-1828.

Henderson, B., Wilson, M., Sharp, L. & Ward, J.M. (2002). *Actinobacillus actinomycetemcomitans*, *J. Med. Microbiol.* 51, 1013-1020.

Henderson, B., Nair, S.P., Ward, J.M. & Wilson, M. (2003). Molecular pathogenicity of the oral opportunistic pathogen *Actinobacillus actinomycetemcomitans*, *Annu. Rev. Microbiol.* 57, 29-55.

Henderson B., Ward J.M. & Ready D. (2010). *Aggregatibacter* (*Actinobacillus*) *actinomycetemcomitans*: a triple A* periodontopathogen? *Periodontol. 2000* 54, 78-105.

Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. & Pease, L.R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction, *Gene* 77, 51-59.

Hu, J.C., Kornacker, M.G. & Hochschild, A. (2000). *Escherichia coli* one- and two-hybrid systems for the analysis and identification of protein-protein interactions, *Methods* 20, 80-94.

Hu, Z. & Lutkenhaus, J. (1999). Topological regulation of cell division in *Escherichia coli* involves rapid pole to pole oscillation of the division inhibitor MinC under the control of MinD and MinE, *Mol. Microbiol.* 34, 82–90.

Hu, Z., Gogol, E.P. & Lutkenhaus, J. (2002). Dynamic assembly of MinD on phospholipid vesicles regulated by ATP and MinE, *Proc. Natl. Acad. Sci. USA* 99, 6761-6766.

Inoue, T., Tanimoto, I., Ohta, H., Kato, K., Murayama, Y. & Fukui, K. (1998). Molecular characterization of low-molecular-weight component protein, Flp, in *Actinobacillus actinomycetemcomitans* fimbriae, *Microbiol. Immunol*. 42, 253–258.

Inoue, T., Ohta, H., Tanimoto, I., Shingaki, R. & Fukui, K. (2000). Heterogeneous post-translational modification of *Actinobacillus actinomycetemcomitans* fimbrillin, *Microbiol. Immunol*. 44, 715-718.

Izano, E.A., Sadovskaya, I., Wang, H., Vinogradov, E., Ragunath, C., Ramasubbu, N., Jabbouri, S., Perry, M.B. & Kaplan, J.B. (2008). Poly-N-acetylglucosamine mediates biofilm formation and detergent resistance in *Aggregatibacter actinomycetemcomitans*, *Microb. Pathog*. 44, 52-60.

Jiang, X., Ruiz, T. & Mintz, K.P. (2012). Characterization of the secretion pathway of the collagen adhesin EmaA of *Aggregatibacter actinomycetemcomitans*, *Mol Oral Microbiol*. 27, 382-396.

Kachlany, S.C., Planet, P.J., Bhattacharjee, M.K., Kollia, E., DeSalle, R., Fine, D.H. & Figurski, D.H. (2000). Nonspecific adherence by *Actinobacillus actinomycetemcomitans* requires genes widespread in bacteria and archaea. *J. Bacteriol*. 182, 6169-6176.

Kachlany, S.C., Planet, P.J., DeSalle, R., Fine, D.H., & Figurski, D.H. (2001a). Genes for tight adherence of *Actinobacillus actinomycetemcomitans*: from plaque to plague to pond scum. *Trends Microbiol*. 9, 429-437.

Kachlany, S.C., Planet, P.J., DeSalle, R., Fine, D.H., Figurski, D.H. & Kaplan, J.B. (2001b). *flp-1*, first representative of a new pilin gene subfamily, is required for nonspecific adherence of *Actinobacillus actinomycetemcomitans*, *Mol. Microbiol*. 40, 542-554.

Kachlany, S.C. (2010). *Aggregatibacter actinomycetemcomitans* leukotoxin: from threat to therapy, *J. Dent. Res*. 89, 561-570.

Kaplan, J.B., Meyenhofer, M.F. & Fine, D.H. (2003). Biofilm growth and detachment of *Actinobacillus actinomycetemcomitans*, *J. Bacteriol*. 185, 1399-1404.

Kelk, P., Johansson, A., Claesson, R., Hänström, L. & Kalfas, S. (2003). Caspase 1 involvement in human monocyte lysis induced by *Actinobacillus actinomycetemcomitans* leukotoxin, *Infect. Immun*. 71, 4448-4455.

Kelk, P., Abd, H., Claesson, R., Sandström, G., Sjöstedt, A. & Johansson, A. (2011). Cellular and molecular response of human macrophages exposed to *Aggregatibacter actinomycetemcomitans* leukotoxin. *Cell Death Dis*. 2, e126.

Kovach, M.E., Phillips, R.W., Elzer, P.H., Roop II, R.M. & Peterson, K.M. (1994). pBBR1MCS: a broad-host-range cloning vector, *Biotechniques* 16, 800-802.

Kram, K.E., Hovel-Miner, G.A., Tomich, M. & Figurski, D.H. (2008). Transcriptional regulation of the *tad* locus in *Aggregatibacter actinomycetemcomitans*: a termination cascade, *J. Bacteriol*. 190, 3859-3868.

Komatsuzawa, H., Asakawa, R., Kawai, T., Ochiai, K., Fujiwara, T., Taubman, M.A., Ohara, M., Kurihara, H. & Sugai, M. (2002). Identification of six major outer membrane proteins from *Actinobacillus actinomycetemcomitans*, *Gene* 288, 195–201.

Labandeira-Rey, M., Brautigam, C.A. & Hansen, E.J. (2010). Characterization of the CpxRA regulon in *Haemophilus ducreyi*, *Infect. Immun*. 78, 4779-4791.

Lally, E.T., Hill, R.B., Kieba, I.R. & Korostoff, J. (1999). The interaction between RTX toxins and target cells, *Trends Microbiol*. 7, 356-361.

Löe, H. & Brown, L.J. (1991). Early onset periodontitis in the United States of America, *J. Periodontol.* 62, 608–616.

Lutkenhaus, J. & Sundaramoorthy, M. (2003). MinD and role of the deviant Walker A motif, dimerization and membrane binding in oscillation, *Mol. Microbiol*. 48, 295–303.

Marston, A.L. & Errington, J. (1999). Dynamic movement of the ParA-like Soj protein of *B. subtilis* and its dual role in nucleoid organization and developmental regulation, *Mol. Cell* 4, 673–682.

Mintz, K.P. (2004). Identification of an extracellular matrix protein adhesin, EmaA, which mediates the adhesion of *Actinobacillus actinomycetemcomitans* to collagen. *Microbiology* 150, 2677-2688.

O'Toole, G. A. & Kolter. R. (1998). Initiation of biofilm formation in *Pseudomonas fluorescens* WCS365 proceeds via multiple, convergent signalling pathways: a genetic analysis, *Mol. Microbiol.* 28, 449-461.

Pabo, C.O., Sauer, R.T., Sturtevant, J.M. & Ptashne, M. (1979). The lambda repressor contains two domains, *Proc. Natl. Acad. Sci. USA* 76, 1608-1612.

Paturel, L., Casalta, J.P., Habib, G., Nezri, M. & Raoult, D. (2004). *Actinobacillus actinomycetemcomitans* endocarditis, *Clin. Microbiol. Infect*. 10, 98–118.

Perez, B.A., Planet, P.J., Kachlany, S.C., Tomich, M., Fine, D.H. & Figurski, D.H. (2006). Genetic analysis of the requirement for *flp-2*, *tadV*, and *rcpB* in *Actinobacillus actinomycetemcomitans* biofilm formation. *J. Bacteriol*. 188, 6361-6375.

Perez-Cheeks, B.A., Planet, P.J., Sarkar, I.N., Clock, S.A., Xu, Q. & Figurski, D.H. (2012). The product of *tadZ*, a new member of the *parA/minD* superfamily, localizes to a pole in *Aggregatibacter actinomycetemcomitans*, *Mol. Microbiol*. 83, 694-711.

Pickett, C.L. & Whitehouse, C.A. (1999). The cytolethal distending toxin family, *Trends Microbiol.* 7, 292-297.

Planet, P.J., Kachlany, S.C., Fine, D.H., DeSalle, R., and Figurski, D.H. (2003) The Widespread Colonization Island (WCI) of *Actinobacillus actinomycetemcomitans*, *Nat. Genet.* 34, 193-198.

Quisel, J.D., Lin, D.C. & Grossman, A.D. (1999). Control of development by altered localization of a transcription factor in *B. subtilis*, *Mol. Cell* 4, 665–672.

Rahamat-Langendoen J.C., van Vonderen, M.G.A., Engström, L.J., Manson, W.L., van Winkelhoff, A.J. & Mooi-Kokenberg, E.A.N.M. (2011). Brain abscess associated with *Aggregatibacter actinomycetemcomitans*: case report and review of literature, *J. Clin. Periodontol*. 38, 702–706.

Ramirez-Arcos, S., Szeto, J., Dillon, J-A.R. & Margolin, W. (2002). Conservation of dynamic localization among MinD and MinE orthologues: oscillation of *Neisseria gonorrhoeae* proteins in *Escherichia coli*, *Mol. Microbiol*. 46, 493–504.

Ramsey, M.M. & Whiteley, M. (2009). Polymicrobial interactions stimulate resistance to host innate immunity through metabolite perception, *Proc. Natl. Acad. Sci. USA* 106, 1578-1583.

Raskin, D.M., & de Boer, P.A. (1999). Rapid pole-to-pole oscillation of a protein required for directing division to the middle of *Escherichia coli*, *Proc. Natl. Acad. Sci. USA* 96, 4971–4976.

Rose, J.E., Meyer, D.H. & Fives-Taylor, P.M. (2003). Aae, an autotransporter involved in adhesion of *Actinobacillus actinomycetemcomitans* to epithelial cells. *Infect. Immun.* 71, 2384–2393.

Rylev, M. &, Kilian, M. (2008). Prevalence and distribution of principal periodontal pathogens worldwide, *J. Clin. Periodontol.* 35, 346-361.

Schreiner, H.C., Sinatra, K., Kaplan, J.B., Furgang, D., Kachlany, S.C., Planet, P.J., Perez, B.A., Figurski, D.H. & Fine, D.H. (2003). Tight adherence genes of *Actinobacillus actinomycetemcomitans* are required for virulence in a rat model, *Proc. Natl. Acad. Sci. USA* 12, 7295-7300.

Slots, J. & Ting, M. (1999). *Actinobacillus actinomycetemcomitans* and *Porphyromonas gingivalis* in human periodontal disease: occurrence and treatment, *Periodontol. 2000* 20, 82-121.

Spinola, S. M., Fortney, K.R., Katz, B.P., Latimer, J.L., Mock, J.R., Vakevainen, M. & Hansen, E. J. (2003). *Haemophilus ducreyi* requires an intact *flp* gene cluster for virulence in humans, *Infect. Immun.* 71, 7178-7182.

Sternberg, N. & Hoess, R. (1983). The molecular genetics of bacteriophage P1, *Annu. Rev. Genet.* 17, 123-154.

Thomson, V.J., Bhattacharjee, M.K., Fine, D.H., Derbyshire, K.M. & Figurski, D.H. (1999). Direct selection of IS*903* transposon insertions by use of a broad host range vector: Isolation of catalase-deficient mutants of *Actinobacillus actinomycetemcomitans*, *J. Bacteriol.* 181, 7298-7307.

Tomich, M., Fine, D.H. & Figurski, D.H. (2006). The TadV protein of *Actinobacillus actinomycetemcomitans* is a novel aspartic acid prepilin peptidase required for maturation of the Flp1 pilin and TadE and TadF pseudopilins, *J. Bacteriol.* 188, 6899-6914.

Tomich, M., Planet, P.J. & Figurski, D.H. (2007). The *tad* locus: Postcards from the Widespread Colonization Island, *Nat. Rev. Microbiol.* 5, 363-375.

Tsai, C.C., McArthur, W.P., Baehni, P.C., Hammond, B.F. & Taichman, N.S. (1979). Extraction and partial characterization of a leukotoxin from a plaque-derived Gram-negative microorganism, *Infect. Immun.* 25, 427-439.

Van Winkelhoff, A. J. & Slots, J. (1999). *Actinobacillus actinomycetemcomitans* and *Porphyromonas gingivalis* in nonoral infections, *Periodontol. 2000* 20, 122-135.

Xu, Q., Christen, B., Chiu, H., Jaroszewski, L., Klock, H.E., Knuth, M.W., Miller, M.D., Elsliger, M., Deacon, A.M., Godzik, A., Lesley, S.A., Figurski, D.H., Shapiro, L. & Wilson, I.A. (2012). Structure of pilus assembly protein TadZ from *Eubacterium rectale*: implications for polar localization, *Mol. Microbiol.* 83, 712-727.

Yue, G., Kaplan, J.B., Furgang, D., Mansfield, K.G. & Fine, D.H. (2007). A second *Aggregatibacter actinomycetemcomitans* autotransporter adhesin exhibits specificity for buccal epithelial cells in humans and Old World primates, *Infect. Immun.* 75, 4440-4448.

Zambon, J.J. (1985). *Actinobacillus actinomycetemcomitans* in human periodontal disease, *J. Clin. Periodont.* 12, 1–20.

4

Directed Mutagenesis of Nicotinic Receptors to Investigate Receptor Function

Jürgen Ludwig, Holger Rabe, Anja Höffle-Maas,
Marek Samochocki, Alfred Maelicke and Titus Kaletta
Galantos Pharma GmbH
Germany

1. Introduction

Nicotinic acetylcholine receptors (nAChR) are the archetypes of drug receptors. In 1905, John Newport Langley introduced the concept of a receptive substance on the surface of skeletal muscle that mediated the action of a drug, such as nicotine and curare (a neurotoxin made from a plant) (Langley et al., 1905). He also proposed that these receptive substances were different in different species and tissues, and they undergo conformational changes in response to the respective drug. Today nAChRs are considered prototypes of receptors that function as integral signal transducers. They have the response element and the ion channel domain within the same molecular entity as the ligand-binding domain that is activated by acetylcholine. In contrast, muscarinic acetylcholine receptors (mAChR) are prototypic G protein-coupled receptors. They also sense molecules outside the cell, but they require the G proteins to induce cellular responses by coupling to intracellular signalling pathways. More important than their historical role, nicotinic receptors continue to be at the forefront of science, as they are drug targets for muscle and nerve diseases, such as Alzheimer's and Parkinson's diseases, Schizophrenia and Myasthenia gravis. Therefore, this receptor family serves as an excellent example for demonstrating the suitability of site-directed mutagenesis for investigating receptor function and exploring drug action.

The neurotransmitter acetylcholine binds to the extracellular domain of the nAChR and, consequently, opens the receptor-integral membrane channel for Na^+, K^+ and Ca^{2+} ions. The channel can close in two ways. (1) The acetylcholine dissociates from its extracellular binding site, a process that is enhanced by rapid cleavage of the neurotransmitter by the acetylcholine esterase in the synaptic cleft. Thus released, the neurotransmitter has only a short time to act on nAChRs before the signal is terminated again. (2) The channel closes spontaneously despite the presence of transmitter, a process called desensitization. Desensitization is a protective mechanism against exposure to acetylcholine and its agonists that is too long or too strong. Desensitization thus avoids excessive influx of ions into the cell, which can result in impairment of cellular function and cell death.

Ligand binding to nAChR usually occurs in the submillisecond range. The receptor-integral channel is opened only for a few milliseconds and is then closed. Nicotinic receptors are therefore extremely fast and efficient signal transducers. They play key roles in such life-important properties as muscle contraction and brain action.

1.1 Structure

A nAChR is composed of five subunits that form a central pore. The extracellular domain contains the binding site for the neurotransmitter acetylcholine (ACh). The four transmembrane helices from each subunit make up the integral ion channel. Depicted are some of the features described in this section. The C-loop is an important part of the acetylcholine binding site. The binding site of the modulating substance galantamine (GAL) is situated at another part of the C-loop. A phosphorylation site (P) is important for modulating the activity of the receptor. A glycosylation site (G) seems to play a role in cobratoxin resistance. Ivermectin and PNU-120596 are other substances that modulate the activity of the receptor, but they have different locations (IVE and PNU). The cytoplasmatic loop is important for receptor targeting.
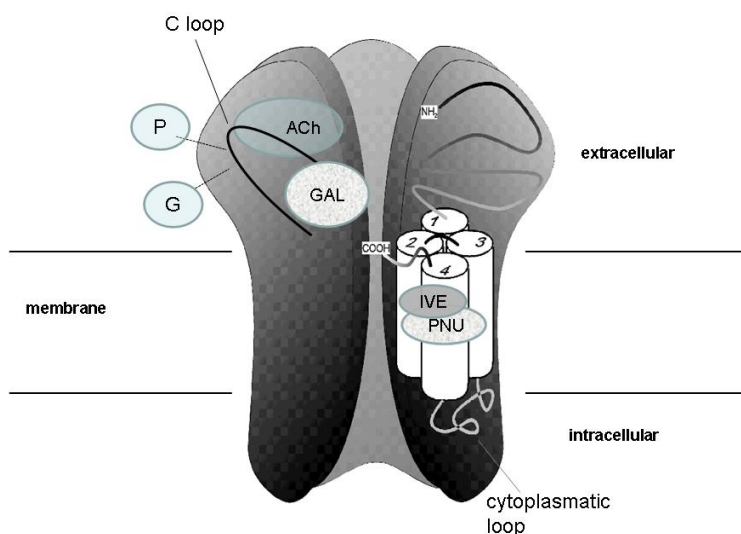
Fig. 1. Schematic representation of a nicotinic Acetylcholine Receptor (nAChR)

The nAChR is assembled from five subunits that are organized around a central pore. Seventeen homologous subunits (α1 – α10, β1 – β4, γ, δ and ε) are known in vertebrates (Zouridakis et al., 2009). This pool of subunits accounts for the vast number of nACh receptor subtypes that exhibit extensive functional diversity with respect to their pharmacological profile, spatiotemporal expression patterns and kinetic properties.

All subunits share the same architectural blueprint, *i.e.*, they consist of a large aminoterminal extracellular domain containing the name-giving cysteine loop, followed by a transmembrane domain and a small intracellular domain. The transmembrane domain consists of four transmembrane regions (TM1 to TM4). TM1, TM2, and TM3 are linked by two short loops. A long and highly variable intracellular loop occurs between TM3 and TM4. Except for those of the long intracellular loop, the amino acid residues are

substantially conserved. This pool of subunits accounts for the vast number of nACh receptor subtypes that exhibit extensive functional diversity with respect to their pharmacological profile, spatiotemporal expression patterns and kinetic properties.

1.2 Use of directed mutagenesis in nicotinic acetylcholine receptor research

Site-directed mutagenesis is a powerful tool to investigate the role of individual amino acids within a protein, to understand the function of a given protein or to understand pharmacological interactions between the protein and compounds. However, modifying individual amino acids or larger parts of a protein with unknown function bears a number of risks. (1) The impact of a particular mutation might appear to be subtle, but it could also lead to a dysfunctional receptor or to the failure of receptor assembly. (2) A mutation might exert an effect on a distant site of the molecule due to conformational changes. This might be interpreted wrongly. (3) The mutated protein might adopt new properties that are unrelated to the natural protein. Hence, it is advisable always to design a battery of receptor mutants or chimeras. It is also important to include similar mutants, either by mutating an amino acid to several different amino acids or by creating similar chimeras. If the results of these mutants are consistent, the risk of erroneously assigning a wrong function can be reduced. Combining mutagenesis studies with other approaches, such as molecular modelling, will further substantiate a hypothesis. A number of such cases will be described in Section 2. Here, the two major mutant types will be described.

1.2.1 Single amino acid changes

Single amino acid changes can be used to investigate the role of amino acids in the binding of the natural ligands and drugs. Furthermore, this type of mutagenesis can be applied to cases in which computer algorithms have predicted motifs, such as glycosylation or phosphorylation sites. Other possible applications are assembly, the targeting of the receptor, or the identification of signals for expression.

A common way to designate a mutant is to use a letter-number-letter scheme. The first letter indicates the wild-type amino acid, using the one-letter code (see Table 1 in the chapter by Figurski et al. for the amino acid codes); the number refers to the position of the amino acid in the protein; and the final letter designates the amino acid that now occupies the position of the original amino acid (*e.g.*, T197A refers to a mutant in which threonine at position 197 was changed to alanine).

1.2.2 Chimeric receptors

Chimeric receptors combine parts of different receptors. This type of mutagenesis is useful in cases in which, instead of a single amino acid, a whole region is in the centre of interest. Examples are functional domains, like the ligand-binding domain or the channel domain, and segments critical for protein signalling, sorting or targeting.

A common way to designate a receptor chimera is to use the names of the receptor types and the name of the joining amino acid (*e.g.*, alpha7-V201-5HT$_3$ refers to a chimeric receptor with the N-terminal part of the α7 nAChR joined with the C-terminal part of the 5HT$_3$ receptor at the amino acid valine 201.
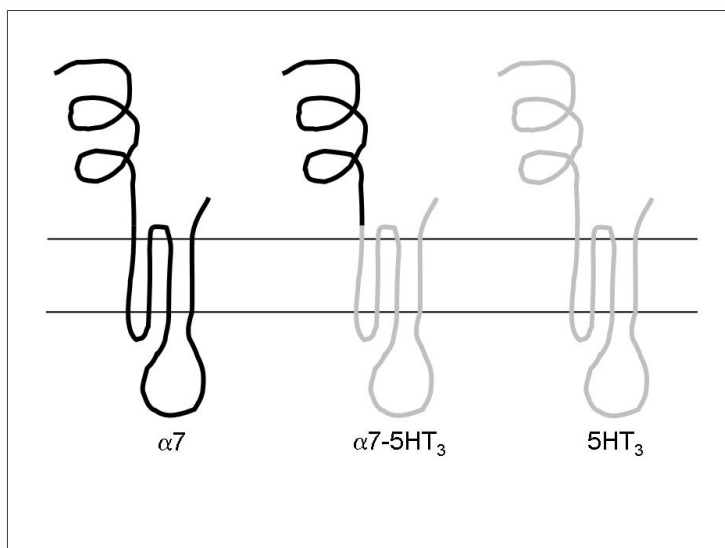
Fig. 2. Schematic representation of a single subunit each from the α7 receptor, the a7-5HT$_3$ chimera, and the 5HT$_3$ receptor. The black line denotes the α7 protein; the grey line denotes the 5HT$_3$ protein.

1.3 In vitro systems to test functional properties of mutated receptors

Electrophysiology is a critical tool for receptor research. This section gives a basic introduction to the reader not familiar with this technique. Two methods are commonly used: (1) the two-electrode voltage clamp method using *Xenopus* oocytes and (2) the whole-cell voltage clamp method amenable for cell lines.

1. Oocytes from the frog *Xenopus laevis* produce almost all ion channel receptor types in high amounts upon injection of their mRNAs. Since only transient protein expression is possible in this system, oocytes can be used only for a short period of a few days. As the name suggests, two sharp microelectrodes filled with a high molarity potassium chloride solution are pricked into 0.6 to 1 mm oocytes to initiate the recording of channel activity. Fortunately an oocyte can be used for several hours. This is enough time to allow the study of a series of agonists and antagonists.
2. The whole-cell patch-clamp technique is the most common electrophysiological method used for many mammalian cell lines such as HEK-293, CHO, GH4C1 or PC-12 cells. These cells can be easily transfected with nAChR expression vectors. It is possible to generate stable cell lines that have integrated the nAChR sequence into the genomic DNA. The whole-cell voltage clamp method employs thin pipettes, which enable good electrical access to the interior of a cell combined with full external electrical insulation. This technique allows measuring small electric currents generated by ion flow through a single receptor molecule or through a couple of receptors. The fast gating and desensitizing channels require fast and direct drug application methods, such as U-tubes and Y-tubes. These systems can only be used with the small mammalian cells (10

– 20 µm). Heterologously expressed receptors in mammalian cell lines might differ in their biophysical and pharmacological characteristics as compared to those analyzed in *Xenopus laevis* oocytes. Therefore, results from the two systems are not always in agreement.

2. Lessons learned from directed mutagenesis

2.1 Identification of intramolecular signals for nicotinic receptor targeting

The physiological role of nAChRs depends on their localisation in specific regions of the cell. For example, presynaptic receptors regulate the transmitter release from synaptic vesicles into the synaptic cleft. Postsynaptic receptors modulate the postsynaptic potential that stimulates action potential formation in neurons or at the neuromuscular junction (Wonnacott et al., 1997; Albuquerque et al., 2009). The observation of differential expression of nAChRs at pre- and postsynaptic sites triggered the search for specific cellular localization signals. This section illustrates mutagenesis approaches to investigate receptor targeting.

Chicken ciliary neurons proofed to be an excellent tool to study receptor targeting (Williams et al., 1998; Temburni et al., 2000). Ciliary ganglion neurons express two nAChR subtypes: (1) α7 nAChRs, for which localisation is restricted to the perisynaptic dendritic membrane and (2) heteromeric nAChRs consisting of α3, α5 and β4 subunits, which are expressed primarily in postsynaptic membranes (Jacob et al., 1986; Conroy and Berg, 1995). The α3 and α7 nAChR subunits are highly homologous to each other, with one exception: the long cytoplasmatic loop shows great diversity in sequence and length (Lindstrom et al. 1996). It also has been shown that this loop is required for the cellular sorting and trafficking machinery. Therefore, it might be the candidate domain for subcellular targeting. Chimeric α7 nAChR subunits were constructed, in which the cytoplasmatic loop was replaced by the homologous region of the α3 nAChR subunit. Furthermore, a myc-epitope tag was added at the C-terminus to allow detection of the receptor without affecting function. Then the chimaeric α7 nAChR subunit with the α3 nAChR cytoplasmatic loop was ectopically expressed in chicken ciliary neurons. Indeed this chimeric receptor was targeted to the postsynaptic membrane, as shown by antibody staining of the myc-epitope tag. This result demonstrates that the cytoplasmatic loop of the α3 nAChR governs the subcellular targeting of the receptor. Other endogenous nAChR subunits do not play this role because the nAChR subunit does not co-assemble with the α3 nAChR subunit (Conroy and Berg, 1995). In addition, α7 nAChR chimeric subunits containing the cytoplasmatic loop of the α5 or β4 nAChR subunit were designed and expressed. These chimeras were targeted to the perisynaptic site (Temburni et al., 2000). This result means that not only the α7 nAChR subunit contains signals for perisynaptic localisation, but that the α5 and β4 nAChR subunits also do.

Interestingly, in the case of the α3α5β4 heteromer, in which perisynaptic and postsynaptic signals are both present, it is the cytoplasmatic loop of the α3 nAChR subunit that determines the targeting to the postsynaptic site of ciliary ganglion neurons. Taken together the results show that the cytoplasmatic loop contains the cellular localisation signal.

The next step was to determine the exact signal peptide within the cytoplasmatic loop. Before initiating the studies, it was tried to transfer the described approach to hippocampal

neurons (Xu et al., 2006). Again the cytoplasmatic loops between the α4 and α7 nAChR subunits were swapped and ectopically expressed in various combinations in hippocampal neurons. Unfortunately, no surface expression could be detected in cells expressing these chimeric nAChR subunits, either alone or in various combinations. It was possible that the design of the chimera was too aggressive. Possibly a critical peptide sequence needed to enable proper receptor assembly and expression in hippocampal cells was accidentally removed.

Therefore, another mutagenesis strategy was chosen. Swapping internal protein domains may affect receptor assembly. Instead model proteins were used. They were left intact, but they were tagged with putative signal peptides from the nAChR. Two non-neuronal transmembrane proteins were chosen. CD4 and the Interleukin 2 receptor when heterologously expressed in neurons are evenly distributed (Gu et al., 2003). It was tested whether the intracellular loops of the α4 and α7 nAChR subunits are able to target these proteins to specific sites in neurons. When the cytoplasmatic loop of the α7 nAChR was fused to the intracellular-oriented C-terminus of CD4, the chimeric protein was only detected in the dendrites. In contrast, the homologous α4 nAChR cytoplasmatic loop leads to axonal expression.

In order to narrow down the precise localisation signals in the cytoplasmatic loop of the nAChR subunits, the following strategy was chosen. Various overlapping fragments covering the loop from its N-terminal region to its C-terminal region were fused to the C-terminus of the Interleukin-2 receptor. The chimeric receptors were expressed in hippocampal neurons. A specific 25-amino acid-fragment (residue positions 30-54) of the α4 nAChR cytoplasmatic loop targeted the chimera to axons. A 48-residue fragment (positions 33-80) of the α7 nAChR cytoplasmatic loop targeted the chimaera to dendrites (Xu J. et al. 2006).

In a last step, site-directed mutagenesis of specific amino acid residues identified in the targeting sequences a leucine motif (DEXXXLLI) in the α4 nAChR cytoplasmatic loop and a tyrosine motif (YXXx) in the α7 nAChR loop.

In conclusion, an iterative approach of chimera design has pin-pointed the precise targeting sequence of a receptor. It is important to note that chimera design may destroy receptor expression or function. Therefore, it is advisable to generate a set of chimeras and to recognize that the change of an expression system may require an adaption of chimera design.

2.2 Confirming computer-based predictions for posttranslational modifications

Posttranslational modifications are important mechanisms for regulating protein expression and protein activity in eukaryotes. Three posttranslational modifications are known for the nicotinic acetylcholine receptor family: glycosylation, phosphorylation and palmitoylation (Albuquerque et al., 2009). Modern computational algorithms effectively help to identify sequence motifs for putative posttranslational modifications. However, experimental approaches are needed to confirm that the site is actually used for posttranslational modifications and to understand its physiological role. By illustrating the role of phosphorylation, this section exemplifies how to combine computational tools with mutagenesis strategies.

Phosphorylation of α7 nAChR negatively regulates its activity. For example, tyrosine kinase inhibition by genistein, a kinase inhibitor, decreases α7 nAChR phosphorylation and, as a consequence, strongly increases acetylcholine-evoked currents (Charpantier et al., 2005). Therefore, it is interesting to know where and which phosporylation sites are present in the protein sequence. Computer analysis predicts two putative phosphorylation sites (tyrosines at residues 386 and 442) in the long cytoplasmatic loop between TM3 and TM4 of the human α7 nAChR. A site-directed mutagenesis that replaced the tyrosines with alanines (Y386A and Y442A) was carried out, and a receptor double mutant with both mutated phosphorylation sites was tested in *Xenopus* oocytes. Indeed the activity of the receptor double mutant was increased to an extent comparable to inhibition of the wild-type receptor by genistein (a kinase inhibitor). This result confirmed that at least one of the two sites is a physiological phosphorylation site. As the receptor double mutant was insensitive to genistein, it can also be assumed that there are no additional physiologically relevant phosphorylation sites in the protein sequence, which might have been overlooked by the computer algorithm. We note that the value of the receptor mutant not only lies in confirming the phosphorylation site, but also in establishing a physiological role for this posttranslational modification. It is primarily required for the regulation of receptor activity, rather than for receptor expression.

2.3 Receptor chimeras demonstrate the modular domain structure of nAChRs

As mentioned in the introduction, nAChRs are integral signal transducers in which the signalling domain and the response element are within one protein. A key question for understanding the molecular design of a receptor family is whether this is achieved through a modular architecture with functionally independent and separable elements.

In 1993 it was impressively demonstrated that domains are interchangeable between different ligand-gated ion channels. This was exemplified with chimeras made from the α7 nAChR and the 5HT$_3$ receptors. The 5HT$_3$ receptor is modulated by the ligand serotonin (5HT), permeable for Na$^+$ and K$^+$ ions, but it is blocked by Ca^{2+} ions. The authors constructed recombinant chimeric receptors with the N-terminal part of the α7 type nAChR, containing the ligand-binding site, and the C-terminal part of the 5HT$_3$ receptor, containing the ion channel domain. They constructed five different chimeras with different junction points for the two receptor parts, using conserved residues W173, Y194, V201, L208 and P217 (the numbers refer to the residue of the α7 receptor). Four of the five chimeric subunit constructs produced properly assembled membrane receptors with an intact extracellular ligand-binding domain, as confirmed by radioactive α-bungarotoxin-binding assays. (α–bungarotoxin is a competitive inhibitor of acetylcholine and binds with high affinity to the acetylcholine-binding site.) Two of the five chimaeras were functional, as shown by the two-electrode voltage clamp technique in *Xenopus* oocytes. One chimera (V201) was able to form large acetylcholine-evoked currents; another (Y194) was able to form small currents.

The V201 chimera was obviously the most interesting chimera and, therefore, was investigated thoroughly by electrophysiological methods and compared to the respective wild-type receptors.

First the ligand-binding properties of the chimaera were pharmacologically studied with a set of modulators and ligands. The agonists were acetylcholine and nicotine, and the

antagonists were α-bungarotoxin and curare, both of which have little or no effect on the $5HT_3$ receptor. In contrast 5HT, the natural ligand of the $5HT_3$ receptor, has no effect on the α7 nAChR receptor. In the *Xenopus* oocyte system, the chimera responded to these ligands in a manner similar to the response of the α7 nAChR wild-type receptor, *i.e.*, acetylcholine increased the current; curare inhibited the currents; and 5HT had no effect.

Second the ion channel properties of the chimera were investigated. α7 type nAChRs and $5HT_3$ receptors differ in their sensitivity to external calcium. While α7 type nAChRs currents increase with higher calcium concentrations, $5HT_3$ receptor currents decrease. The α7 nAChR channel domain is highly conductive for calcium ions, whereas the $5HT_3$ receptor channel domain is blocked by calcium. In addition, external calcium ions have a potentiating effect of the acetylcholine action on α7 nAChRs. In *Xenopus* oocyte studies, the α7 nACh-V201-$5HT_3$ receptor chimera behaved in a manner similar to the response of the $5HT_3$ receptor towards the external calcium concentration. This means that the ion channel properties of the $5HT_3$ receptor are independent of its ligand-binding properties. It appears that, by swapping the ligand-binding domains, a $5HT_3$ receptor can be engineered to respond to acetylcholine and other nAChR modulators like a natural nAChR.

Interestingly, the onset and desensitization kinetics of the current of the chimera is in between the rapid kinetics of wild-type α7 nAChR and the slow kinetics of the $5HT_3$ receptor. This suggests that not all properties of a receptor can be exchanged by simply swapping the ligand-bind domains. The specific kinetic properties of a receptor apparently require interplay between different domains.

In conclusion, computer-based predictions and traditional biochemistry can certainly suggest the domain structure of a protein. However, in order to distinguish whether domains are functionally independent or just building blocks of a larger functional unit, chimeras are invaluable tools.

2.4 Concatemeric nACh receptors revealed the subunit order of nAChRs

The principal architecture of heteromeric nAChRs allows for, at least theoretically, numerous combinatorial arrangements of the various α, β, γ, δ and ε subunit types around the central pore. These arrangements could result in many different receptors with different properties. Which of the possible arrangements are realized in nature, and what distinct properties might they have? For example, it is conceivable that different arrangements of the subunits have different acetylcholine-binding properties because the binding site is located at the interface of two subunits.

In neuronal tissue, many different nAChR combinations are usually present. It can be difficult to study individual nAChR types. In addition the actual arrangement of the subtypes cannot easily be assessed. Directed mutagenesis can be used to force subtypes to form specific nAChR arrangements. Thus, it can be a powerful tool for investigating the roles of different nAChR arrangements. This section focuses on the α4β2 type nAChR, which is the most abundant form of heteropentameric nACh receptors in the mammalian brain.

A pentameric α4β2 nAChR could consist of three α4 and two β2 subunits, or vice versa. Earlier research using *Xenopus laevis* oocytes has shown that the use of equal amounts of α4

and β2 subunit mRNAs generates an $(\alpha4)_2(\beta2)_3$ nAChR (*i. e.*, the α4 and β2 subunits are in a 2:3 stoichiometry, respectively) (Anand et al., 1991; Cooper et al., 1991). In contrast, under conditions in which the α4 subunit mRNA was in excess, an $(\alpha4)_3(\beta2)_2$ nAChR was obtained (3:2 stoichiometry) (Zwart and Vijverberg, 1998). The two different α4β2 nAChRs differ in their pharmacological properties. The $(\alpha4)_2(\beta2)_3$ nAChR is highly sensitive to acetylcholine, whereas the $(\alpha4)_3(\beta2)_2$ nAChR is less sensitive. Since both receptor forms are present in the brain (Marks et al., 1999; Gotti et al., 2008), it has been suggested that the ratio of the two nAChRs is part of a regulatory mechanism for neuronal cell response to nicotine (Tritto et al., 2002; Kim et al., 2003). In addition to their differential sensitivity to acetylcholine and nicotine, the two receptor forms differ in other properties, such as desensitization kinetics and Ca^{2+} permeability (Nelson et al., 2003; Zwart et al., 2008; Moroni et al., 2006; Tapia et al., 2007; Moroni et al., 2008).

For each stoichiometric nAChR, there exist two possible orders of the subunits. The $(\alpha4)_2(\beta2)_3$ nAChR can be α4–α4–β2–β2–β2 (1.1) or α4–β2–α4–β2–β2 (1.2). The $(\alpha4)_3(\beta2)_2$ nAChR can be α4–α4–α4–β2–β2 (2.1) or α4–α4–β2–α4–β2 (2.2). Which of the possible arrangements are realized and what functional significance is conferred by specific positions of each subunit within the receptor complex? Insight into this issue has been gained by using a specific type of directed mutagenesis, *i.e.*, concatemers. A concatemer is a long synthetic gene that contains smaller genes (*e.g.*, the subunit genes) linked in series. In this case, the resulting protein consists of a defined sequence of subunits in which the carboxyl-terminus of the preceding subunit is covalently linked with the amino-terminus of the following subunit (Zhou et al., 2003; Nelson et al, 2003). This technique allows one to enforce a predefined subunit order for the receptor. Studies employing this technique with tandem and triple concatemers of the α4 and β2 subunits showed that only two arrangements were functional. For the $(\alpha4)_2(\beta2)_3$ nAChR, it was the α4–β2–α4–β2–β2 (1.2) arrangement; and for the $(\alpha4)_3(\beta2)_2$ nAChR, it was the α4–α4–β2–α4–β2 (2.2) arrangement (Carbone et al., 2009). Thus, in both receptor forms a triplet of the same subunit was avoided. In summary, the mutagenesis studies described improved our understanding of how the subunit types are arranged (Zhou et al., 2003; Nelson et al, 2003).

Recently, the use of a pentameric nAChR concatemer revealed that there exists also at the α4/α4 interface a functional acetylcholine-binding site (Mazzaferro et al., 2011). Originally, the acetylcholine-binding sites were thought to be located at the α4/β2 subunit interfaces. As these interfaces are present in both receptor isoforms, it is unlikely that they account for differences in acetylcholine sensitivities. In the new study the authors clearly identified the α4/α4 interface in α4–α4–β2–α4–β2 receptors as an additional acetylcholine-binding site. They used a combined approach of a pentameric nAChR concatemer, chimeric α/β subunits with mutagenesis of loop C and structural modelling to determine that this α4/α4 interface accounts for isoform-specific characteristics, *i.e.*, for the low acetylcholine sensitivity. In conclusion, directed mutagenesis permitted a defining of the order of the nAChR subunits and, thus, allowed a determination of the way agonist-binding sites are formed.

2.5 Cobratoxin

The α-neurotoxins from snake venoms are potent antagonists of nicotinic acetylcholine receptors. In mouse some of them are over ten times more toxic than nicotine (*e.g.*, the LD_{50}

of α-cobratoxin from *Naja naja* in mouse is 0.4 mg/kg versus 7.1 mg/kg for nicotine). Despite the high toxicity of the snake toxins, some animals are resistant to α-neurotoxins. This is the case for animals that feed on cobras, such as the mongoose, and, of course, the snake itself. A substantial effort using directed mutagenesis has been made to characterize the interaction between α-neurotoxins and the nicotinic acetylcholine receptors to identify the mechanism of this resistance.

2.5.1 Understanding the cobratoxin – α7 nicotinic acetylcholine receptor interaction

The binding site of the α7 nAChR is composed of six loops, A to F. In order to identify which of these loops interact(s) with α-cobratoxin (α-Cbtx), extensive site-directed mutagenesis was carried out to generate 40 receptor mutants (Fruchart-Gaillard et al., 2002). The possible role of a given amino acid in a α7 nAChR loop thought to interact with α-Cbtx was determined by comparing a mutant receptor with a changed amino acid to the wild-type receptor using a competition binding assay with radioactive iodide-labelled α-bungarotoxin. Only mutations in loops C, D and F reduced the affinity to α-Cbtx. Hence, these loops may be critical for α7 nAChR - α-Cbtx interaction. Mutations of loop C at residues F186 and Y187 showed the greatest effect. They reduced the affinity by 100- to 200-fold. It is important to notice that not every mutation at these two positions reduced affinity. For example, F186R reduced affinity by a factor of 100; in contrast, F186A reduced affinity by a factor of only 4; and F186T, not at all. Hence, when determining a possible role of an amino acid by site-directed mutagenesis, amino acid properties may matter (*e.g.*, size or charge). It is often important to test several amino acid exchanges for a full understanding.

Similarly, a site-directed mutagenesis approach was taken to generate 36 toxin mutants. The objective was to identify the interaction sites of α-Cbtx with the receptor (Antil-Delbeke et al., 2000). This study found loop II and the C-terminal tail of α-Cbtx to interact with α7 nAChR.

To determine which amino acid(s) of the receptor interacts with which amino acid(s) of the toxin, α7 nAChR receptor mutants were tested with α-Cbtx mutants in the competition assay. The studies revealed that the amino acid R33 in loop II of α-Cbtx interacts with a number of amino acids in loop C of the α7 nAChR, such as Y187, W148, P193, and Y194 (Fruchart-Gaillard et al., 2002). Another example is the amino acid K35 of α-Cbtx that interacts with the amino acids F186 and D163 of the α7 nAChR.

This information was then used in a computational 3D model to orient α-Cbtx in the binding pore of α7 nAChR and to help understand the mechanism of the antagonistic action of α-Cbtx. How can the large α-Cbtx molecule exert its antagonistic action on a binding site that is configured to fit small ligands, such as acetylcholine or nicotine? The docking study revealed that only the tip of loop II of the toxin plugs into the cavity between two receptor subunits. About 75% of the remaining surface of the toxin stays outside the toxin-receptor complex. In this way, α-Cbtx behaves like a small ligand and effectively antagonizes α7 nAChR.

2.5.2 Establishing resistance against snake toxin

A snake is usually resistant to its own venom. Hence, it is interesting to find out whether it is resistant from a specific difference in the target molecule of its toxin. In this case, the

target molecule is the nicotinic acetylcholine receptor. There are considerable differences in the protein sequences of the nAChR ligand-binding domains of snakes (*Naja spes*) and mammals. The differences may suggest the different sensitivities. It is possible to make the snake α1 nAChR sensitive to α-Bungarotoxin, a venom of the elapid family. It has been shown by site-directed mutagenesis that introducing the mutation N189F, which substitutes the asparagine in the snake protein sequence with the phenylalanine in the mouse sequence, abolishes resistance (Takacs et al., 2001). This suggests not only that the snake α1 nAChR contains a ligand-binding domain for snake toxins, but also that a single amino acid can cause sensitivity or resistance. In order to test whether the asparagine indeed confers the resistance, a F189N mutation was introduced into the nAChR of the mouse. Two-electrode voltage clamp analysis in *Xenopus* oocytes showed that this mouse receptor mutant was resistant to α-Bungarotoxin. This means that a single amino acid substitution can determine sensitivity or resistance to a snake venom.

Interestingly, N189 is an N-glycosylation site, and it has been postulated that the bulky glycosyl residue may prevent the toxin from entering the binding site (Barchan et al., 1992). This hypothesis might explain the resistance of the mongoose to snake venoms and the sensitivity of mammals. In fact, the mongoose nAChR has an N-glycosylation site at N187, whereas other mammals, such as mouse, cat and humans, lack an N-glycosylation site in the ligand-binding domain.

Another resistance mechanism may be deduced from the receptor-toxin interaction, as described in the previous section. The F189 residue in the α1 nAChR is analogous to F186 in the α7 nAChR. F186 was identified to be critical for the receptor-binding interaction. Therefore, it is conceivable that two mechanisms in the snake receptor confer resistance to α-neurotoxins: (1) disruption of a critical protein-protein interaction and (2) steric hindrance via glycosylation.

2.6 Mapping of the binding sites of allosteric potentiating ligands

Modulators of ligand-gated ion channels (LGICs) have become therapeutically important because, in contrast to traditional agonists or antagonists, these substances change receptor activity only in the presence of the natural ligand. This allows a more physiological control of a LGIC. This concept has been broadly applied for the treatment of epilepsy. For example, benzodiazepines, such as diazepam ("Valium"), present a major class of allosteric modulators of the $GABA_A$-receptors.

Cholinergic neurotransmission is a prominent therapeutic target for the treatment of diseases like Alzheimer's disease. The drug galantamine exerts its therapeutic action by allosteric modulation of the nAChRs (Bertrand and Gopalakrishnan, 2007; Maelicke et al., 2001). Understanding the mechanism of allosteric modulation is therefore important for developing novel drugs for treating Alzheimer´s disease (Faghih et al., 2007).

Today a range of allosteric modulators is known. Synonymously used terms are "allosteric potentiating ligands" (APL) and "positive allosteric modulators" (PAM). They fall into different classes. Galantamine is a representative of the type I class of PAMs (PAM I), which enhance nAChR activity by increasing the current without affecting receptor desensitization. In contrast, members of the type II class of PAMs (PAM II) increase the current of nAChRs, but also reduce their desensitization (Hurst et al., 2005). Hence, a

number of studies were directed to the identification of possible different binding sites, which would help to explain the mechanistic differences of the two classes.

2.6.1 PNU-120596

The group around Neil Millar ran a combined approach to locate the binding site of the type II modulator PNU-120596 a well-studied developmental compound. In a first round of experiments, they compared the action of PNU-120596 on the α7 nAChR and the 5HT$_3$ receptor with a set of α7 nACh/5HT$_3$ receptor chimeras (Young et al., 2008). In these chimeras, principally the extracellular part of the α7 nAChR was combined with at least the first three transmembrane domains from the 5HT$_3$ receptor. PNU-120596 could, of course, potentiate α7 nAChR. However it could potentiate neither the original 5HT$_3$ receptor nor any of the receptor chimeras. This indicated that the PNU-120596 binding site is located in the transmembrane part of the nACh receptor.

In a second round of mutant receptor designs, the amino acid sequence of the transmembrane part of the α7 nAChR was compared with the one from the 5HT$_3$ receptor. A number of differences were identified. Amino acids that are not conserved in the α7-nAChR (which is potentiated by PNU-120596) and the 5HT$_3$ receptor (which is not potentiated by PNU-120596) formed the basis for a set of α7 nAChR mutants. Amino acids in the nAChR were mutated to the corresponding amino acids of the 5HT$_3$ receptor. As expected, some of the mutants were simply not functional, whereas others showed responses to PNU-120596 that were similar to that of wild-type nAChR. Five mutants having otherwise normal function were less responsive than wild type to the potentiating activity of PNU-120596. Two of them (A225D and M253L) were nearly resistant to PNU-120596.

In order to understand why these five amino acids of wild-type nAChR conferred sensitivity to PNU-120596 modulation, the authors investigated several computer models of the nAChR. It turned out that the five amino acids were part of an intra-subunit cavity. Docking simulations revealed that the most favourable docking position of PNU-120596 was at a location very near to the locations of the five amino acids.

Complementary studies using a set of different α7/5HT3 receptor chimeras showed that the α7 nAChR ion channel domain is essential for the action of PNU-120596 (Bertrand et al., 2008). This has led to the model that PNU-120596 exerts its function by stabilizing the cavity in an agonist-like fashion (Barron et al., 2009).

In summary, a combined approach of site-directed mutagenesis experiments, molecular modelling and docking studies was needed to identify the binding site of PNU-120596.

2.6.2 Galantamine

The following studies revealed a totally different location for the binding site of galantamine. They offer an explanation of why the mechanisms of action of the two allosteric modulating ligands are substantially different.

The identification of the galantamine binding site on the extracellular domain close to the acetylcholine binding site has taken more than a decade. Several different approaches were required to finally locate it precisely (Schröder et al., 1994; Ludwig et al., 2010).

A first step was the combination of results from earlier studies about an antibody called FK1 (Schröder et al., 1994; Brejc et al., 2001; Luttmann et al., 2009). An important feature of antibody FK1 is its ability to block the potentiating effect of galantamine on the nAChR without affecting the response to acetylcholine or agonists. The strategy described in this paper helped to identify two stretches (27 and 28 amino acids in length) that contain part of the binding site of galantamine. However, the question remained as to which amino acids participate in the interaction (Schröder et al., 1994). A protein from a freshwater snail assisted in narrowing the range of possible amino acids (Brejc et al., 2001; Smit et al., 2001). The freshwater snail *Lymnaea stagnalis* produces a protein that is homologous to the ligand-binding domain of nAChRs. It is called Acetyl Choline Binding Protein, and it assembles as a homopentamer amenable to X-ray crystallography. A 2.7 Å resolution structure was used as a template to model the ligand-binding domains of α7 and α4β2 nAChRs (Luttmann et al., 2009). In these models only a small proportion of the amino acid stretches of the receptor lie in a position at the outer surface that would be accessible to the FK1 antibody (Figure 3). As both amino acid stretches contribute to the FK1 epitope, it seems highly probable that the epitope is located at the junction of the two stretches.
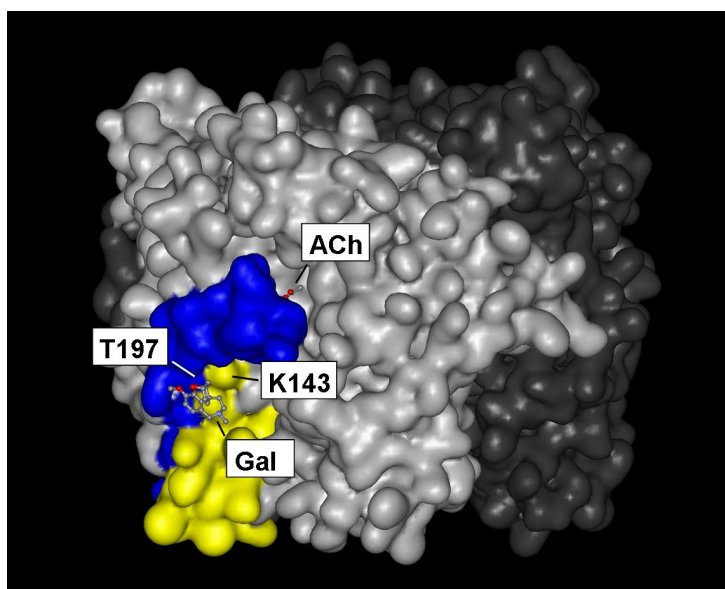
Fig. 3. Surface model of the α7 nAChR ligand-binding domain. Only the ligand-binding domain (LBD) is shown. The cell membrane and channel domain would be beneath the LBD. Two of the five subunits are shown in light grey, with the other ones depicted in dark grey. One molecule of acetylcholine (ACh) is bound. It is almost buried inside the binding site. The amino acids in yellow and blue belong to amino acid stretches identified in Schröder et al., 1994 as contributing to the epitope for the galantamine blocking antibody FK1. Mutation of the amino acids T197 and K143 showed no (T197) or reduced (K143) galantamine effect when stimulated with acetylcholine. This is in line with the assumption that the galantamine binding site is at the junction of these two amino acid stretches. The amino acids T197 and K143 are possible binding sites for galantamine (Gal), as predicted by docking studies.

These insights led to a hypothetical binding site, which was proven with a set of eight different α7 nAChR mutants (Ludwig et al., 2010). All of the mutants were functional, albeit two had much lower affinities for acetylcholine and agonists than wild type. Four of the mutants, all of which showed a normal response to agonists, had no or a reduced response to galantamine. These four mutants were altered in amino acids at the borders of the two stretches mentioned above and were adjacent in the α7 nAChR model (Ludwig et al., 2010). Docking studies pinpointed two amino acids, threonine 197 and lysine 143, as the galantamine binding site (Luttmann et al., 2009). The two amino acids that were replaced in the other two mutants were shown to be oriented in a way in which they would be unlikely to interact with galantamine directly.

A mechanistic model to explain how the binding of acetylcholine opens the ion channel proposes that the C-loop (Fig. 1) acts as a lever that moves upon binding of acetylcholine. Since the galantamine binding site is located at the lower part of the C-loop, it is assumed that binding of galantamine enhances the action of this lever.

2.6.3 Ivermectin

Ivermectin is a member of the PAM I class of modulators. It has a potentiating action on the nAChR activation without affecting the desensitization. Collins and Millar performed an approach similar to the one described for PNU-120596 to identify the amino acids that play a critical role in the interaction with ivermectin (Collins and Millar, 2010).

The authors conducted experiments with α7 nAChRs, $5HT_3$ receptors and the already described α7-$5HT_3$ chimeric receptors in *Xenopus* oocytes. The experiments led to the conclusion that the transmembrane domain plays a critical role in the allosteric modulation by ivermectin. The results showed that, while ivermectin potentiated the effect of acetylcholine on the α7 nAChR, it had no effect on the $5HT_3$ receptor. In contrast, the α7-$5HT_3$ chimeric receptor, which contains the ligand-binding domain of the α7 nAChR and the ion channel domain of the $5HT_3$ receptor, was surprisingly inhibited by ivermectin. The reason for this unexepected response of the chimeric receptor to ivermectin is not known, but it might be that the $5HT_3$ receptor extracellular domain blocks the access of ivermectin to a transmembrane domain binding site that is accessible in the chimeric receptor.

In subsequent experiments, the authors changed selected amino acids in the α7 nAChR. The mutations A225D, Q272V, T456Y, and C459Y almost completely prevented allosteric modulation by ivermectin. It is interesting to note that some of these mutants react similarly to PNU-120596 and to ivermectin. The A225D and C459Y mutants showed a reduced response to both compounds. In contrast, mutations at Q272V and T456V reduced the allosteric potentiation of ivermectin, but not of PNU-120596. This suggests that the amino acids responsible for the allosteric potentiating action of ivermectin partially overlap the ones responsible for the action of PNU-120596.

2.6.4 Conclusion

Section 2.6 describes the use of directed mutagenesis for the identification and investigation of the binding sites of the three nAChR modulators PNU-120596, galantamine and ivermectin. Galantamine and ivermectin are classified as PAM I

modulators, while PNU-120596 is a PAM II class allosteric modulator, based on electrophysiological properties. However, the binding site of galantamine is on the extracellular ligand-binding domain, while ivermectin and PNU-120596 bind to the channel domain of the receptor. Apparently similar electrophysiological properties do not reflect similar binding sites.

3. Concluding remarks

This chapter has highlighted milestones of the directed mutagenesis research performed on nAChRs. It includes research about receptor targeting, the modular domain architecture, the order of the heteromeric subunit assembly, the mechanism of toxin resistance and the receptor interaction with modulatory ligands. Future research will continue to investigate the nicotinic receptor family and its role in diseases, such as Alzheimer's or Parkinson's. The interaction between α7 nAChR and β-amyloid is just one of the many urgent questions that need to be resolved (Tong et al., 2011). Directed mutagenesis will remain one of the most powerful tools on the journey towards the full understanding of this molecular machine.

4. References

Anand R., Conroy, W.G., Schoepfer, R., Whiting, P., & Lindstrom, J. (1991). Neuronal nicotinic acetylcholine receptors expressed *Xenopus* oocytes have a pentameric quaternary structure. *Journal of Biological Chemistry*, Vol.266, No.17 (June 1991), pp. 11192-11198, ISSN 0021-9258.

Antil-Delbeke, S., Gaillard, C., Tamiya, T., Corringer, P., Corringer, P., Servent, D., & Ménez, A. (2000). Molecular determinants by which a long chain toxin from snake venom interacts with the neuronal alpha7-nicotinic acetylcholine receptor. *Journal of Biological Chemistry* Vol.275, No.38 (September 22), pp.29594–29601 ISSN 1083-351X

Albuquerque, E., Pereira, E., Alkondon, M., & Rogers, S. (2009). Mammalian nicotinic acetylcholine receptors: from structure to function. *Physiological Reviews*, Vol.89, No.1, (January 2009), pp. 73-120. ISSN 1522-1210

Barchan, D., Kachalsky, S., Neumann, D., Vogel, Z., Ovadia, M., Kochava, E., & Fuchs, S. (1992). How the mongoose can fight the snake: The binding site of the mongoose acetylcholine receptor. *PNAS* Vol. 89, No.16 (August 15), pp.7717-7721, ISSN 0027-8424

Barron, S.C., McLaughlin, J.T., See, J.A., Richards, V.L., & Rosenberg, R.L., (2009). An allosteric modulator of alpha7 nicotinic receptors, N-(5-Chloro-2,4-dimethoxyphenyl)-N'-(5-methyl-3-isoxazolyl)-urea (PNU-120596), causes conformational changes in the extracellular ligand binding domain similar to those caused by acetylcholine. *Molecular Pharmacology*, Vol.76, No.2 (August 2009), pp. 253-263, ISSN 0026-895X

Bertrand, D., Bertrand, S., Cassar, S., Gubbins, E., Li, J., & Gopalakrishnan, M. (2008). Positive allosteric modulation of the alpha7 nicotinic acetylcholine receptor: ligand interactions with distinct binding sites and evidence for a prominent role of the M2-M3 segment. *Molecular Pharmacology*, Vol.74, Nr.5, (November 2008), pp.1407-1416, ISSN 0026-895X

Brejc, K., van Dijk, W., Klaassen, R., Schuurmans, M., van Der Oost, J., Smit, A. and Sixma, T. (2001). Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature*, Vol.411, No.6835, (May 2001), pp.269-276, ISSN 0028-0836

Carbone, A.L., Moroni, L.M., Groot-Kormelink, P.J. & Bermudez, I. (2009). Pentameric concatenated $(\alpha 4)_2(\beta 2)_3$ and $(\alpha 4)_3(\beta 2)_2$ nicotinic acetylcholine receptors: subunit arrangement determines functional expression. *British Journal of Pharmacology*, Vol. 156, No.6 (March 2009), pp. 970–981, ISSN 0007-1188.

Charpantier, E., Wiesner, A., Huh, K., Ogier, R., Hoda J., Allaman, G., Raggenbass, M., Feuerbach, D., Bertrand, D., & Fuhrer, C. (2005). $\alpha 7$ Neuronal nicotinic acetylcholine receptors are negatively regulated by tyrosine phosphorylation and src-family kinases. *The Journal of Neuroscience*, Vol.25, No.43, (October 26), pp.9836 – 9849, ISSN: 1529-2401

Collins, T., and Millar, N.S. (2010). Nicotinic acetylcholine receptor transmembrane mutations convert ivermectin from a positive to a negative allosteric modulator. Molecular Pharmacology, Vol.78, No.2, (August 2010), pp. 198-204, ISSN 0026-895X

Conroy, W.G., & Berg D.K. (1995). Neurons can maintain multiple classes of nicotinic acetylcholine receptors distinguished by different subunit compositions. *The Journal of Biological Chemistry,* Vol.270, No. 9, (March 1995), pp. 4424- 4431, ISSN 0021-9258

Cooper, E., Couturier, S. & Ballivet, M. (1991). Pentameric structure and subunit stoichiometry of a neuronal nicotinic acetylcholine receptor. *Nature*, Vol.350, No.6315 (March 1991), pp. 235-238, ISSN 0028-0836.

Drisdel, R., Manzana, M. & Green, W (2004). The role of palmitoylation in functional expression of nicotinic $\alpha 7$ receptors. *The Journal of Neuroscience*, Vol.24 No.46, (November 17), pp. 10502–10510, ISSN: 1529-2401

Eiselé, J., Bertrand, S., Galzi, J., Devillers-Thiéry, A., Changeux J., & Bertrand D. (1993) Chimaeric nicotinic-serotonergic receptor combines distinct ligand binding and channel specificities. *Nature*, Vol.366, No.6454, (December 1993), pp.479-483, ISSN: 0028-0836

Fruchart-Gaillard, C., Gilquin, B., Antil-Delbeke, S., Le Novère, N., Tamiya, T., Corringer, P., Changeux, J., Me´nez, A., & Serven, D. (2002). Experimentally based model of a complex between a snake toxin and the $\alpha 7$ nicotinic receptor. *PNAS*, Vol.99, No.5, (March 5), pp.3216–3221, ISSN 0027-8424

Faghih, R., Gfesser, G., & Gopalakrishnan, M. (2007). Advances in the discovery of novel positive allosteric modulators of the alpha7 nicotinic acetylcholine receptor. *Recent Patents on CNS Drug Discovery*, Vol.2, No.2 (June 2007), pp. 99-106, ISSN 1574-8898

Gahring, L., & Rogers, S. (2006). Neuronal nicotinic acetylcholine receptor expression and function on nonneuronal cells. *AAPS Journal*, Vol.7, No.4, (January 2006), pp. E885-894, ISSN ISSN 1550-7416

Geerts, H., Guillaumat, P., Grantham, C., Bode, W., Anciaux, K., & Sachak, S. (2005), Brain levels and acetylcholinesterase inhibition with galantamine and donepezil in rats, mice, and rabbits. *Brain Research*, Vol.1033, No.2, (February 2005), pp. 186-193, ISSN 0006-8993

Gotti, C., Zoli, M., & Clementi, F. (2006). Brain nicotinic acetylcholine receptors: native subtypes and their relevance. *Trends In Pharmacological Science*, Vol.27, No.9 (September 2006), pp. 482–491, ISSN 0165-6147.

Gotti C., Clementi F., Fornari A., Gaimarri A., Guiducci S., Manfredi I., Moretti M., Pedrazzi P., Pucci L., & Zoli M. (2009). Structural and functional diversity of native brain neuronal nicotinic receptors. *Biochemical Pharmacology,* Vol.78, No.7, (October 2009), pp. 703-711. ISSN 0006-2952

Hucho, F., Tsetlin, V. & Machold, J. (1996). The emerging three-dimensional structure of a receptor. The nicotinic acetylcholine receptor. European Journal of Biochemistry, Vol.239, No.3, (August 1996), pp.539-557, ISSN 1432-1033

Hurst, R., Hajós, M., Raggenbass, M., Wall, T., Higdon, N., Lawson, J., Rutherford-Root, K., Berkenpas, M., Hoffmann, W., Piotrowski, D., Groppi, V., Allaman, G., Ogier, R., Bertrand, S., Bertrand, D. & Arneric, S. (2005). A novel positive allosteric modulator of the alpha7 neuronal nicotinic acetylcholine receptor: in vitro and in vivo characterization. *Journal of Neuroscience*, Vol.25, No.17, pp. 4396-4405, ISSN 1529-2401

Jacob, M.H., Lindstrom, J.M., & Berg, D.K. (1986). Surface and intracellular distribution of a putative neuronal nicotinic acetylcholine receptor. *The journal of cell biology,* Vol. 103, No. 1, (July 1986), pp. 205-214, ISSN 0021-9525

Kim H., Flanagin B.A., Qin C., Macdonald R.L., & Stitzel J.A. (2003). The mouse Chrna4 A529T polymorphism alters the ratio of high to low affinity α4β2 nAChRs. *Neuropharmacology* Vol.45, No.3 (September 2003), pp. 345–354, ISSN 0028-3908.

Langley, J. (1905). On the reaction of cells and of nerve-endings to certain poisons, chiefly as regards the reaction of striated muscle to nicotine and to curari.. *The Journal of Physiology*, Vol.33, No.4-5 (December 1905), pp. 374-413, ISSN 1469- 7793

Lindstrom, J., Anand, R., Gerzanich, V., Peng, X., Wang, F. and Wells, G. (1996). Structure and function of neuronal nicotinic acetylcholine receptors. *Progress in Brain Research,* Vol.109, No.1 (1996), pp. 125-137, ISSN 0079-6123

Ludwig J., Höffle-Maas A., Samochocki M., Luttmann E., Albuquerque E., Fels G., & Maelicke A. (2010). Localization by site-directed mutagenesis of a galantamine binding site on α7 nicotinic acetylcholine receptor extracellular domain. *Journal of Receptors and Signal Transduction*, Vol.30, No.6, (December 2006), pp. 469-483, ISSN 1532-4281

Luttmann, E., Ludwig, J., Höffle-Maas, A., Samochocki, M., Maelicke, A., & Fels G. (2009). Structural model for the binding sites of allosterically potentiating ligands on nicotinic acetylcholine receptors. *ChemMedChem*, Vol.4, No.11, (November 2009), pp. 1874-1882, ISSN 1860-7187

Macklin, K., Maus, A., Pereira, E., Albuquerque, E., & Conti-Fine, B. (1998). Human vascular endothelial cells express functional nicotinic acetylcholine receptors. *Journal of Pharmacology and Experimental Therapeutics,* Vol.287, No.1, (October 1998), pp. 435-439. ISSN 1521-0103

Maelicke, A., & Albuquerque, E. (2000). Allosteric modulation of nicotinic acetylcholine receptors as a treatment strategy for Alzheimer's disease. *European Journal of Pharmacology*, Vol.393, No.1-3, pp.165-170, ISSN 0014-2999

Maelicke, A., Samochocki, M., Jostock, R., Fehrenbacher, A., Ludwig, J., Albuquerque E., & Zerlin, M. (2001) Allosteric sensitization of nicotinic receptors by galantamine, a new treatment strategy for Alzheimer's disease. *Biological Psychiatry*, Vol.49, No.3, (February 2001), pp. 279-288, ISSN 0006-3223

Marks, M.J., Whiteaker, P., Calcaterra, J., Stitzel, J.A., Bullock, A.E., Grady, S.R. et al. (1999). Two pharmacologically distinct components of nicotinic receptor-mediated rubidium efflux in mouse brain require the β2 subunit. *The Journal of Pharmacology and Experimental Therapeutics* Vol.289, No.2 (May 1999), pp. 1090–1103, ISSN 0022-3565.

Maus, A., Pereira, E., Karachunski, P., Horton, R., Navaneetham, D., Macklin, K., Cortes, W., Albuquerque, E., & Conti-Fine, B. (1998). Human and rodent bronchial epithelial cells express functional nicotinic acetylcholine receptors. *Molecular Pharmacology*, Vol.54, No.5, (November 1998), pp. 779-788. ISSN 0026-895X

Mazzaferro, S., Benallegue, N., Carbone, A., Gasparri, F., Vijayan, R., Biggin, P.C., Moroni, M., & Bermudez, I. (2011). An additional ACh binding site at the α4/α4 interface of the (α4β2)$_2$α4 nicotinic receptor influences agonist sensitivity. *Journal of Biological Chemistry*, Vol.286, No.35 (September 2011), pp. 31043-31054, ISSN 0021-9258.

Miwa, J., Stevens, T., King, S., Caldarone, B., Ibanez-Tallon, I., Xiao, C., Fitzsimonds, R., Pavlides, C., Lester, H., Picciotto, M., & Heintz, N. (2006) The prototoxin lynx1 acts on nicotinic acetylcholine receptors to balance neuronal activity and survival in vivo. *Neuron*, Vol.51, No.5, (September 2006), pp. 587-600. ISSN 0896-6273

Moroni, M, Zwart, R., Sher, E., Cassels, B.K., & Bermudez, I. (2006). α4β2 nicotinic receptors with high and low acetylcholine sensitivity: pharmacology, stoichiometry, and sensitivity to long-term exposure to nicotine. *Molecular Pharmacology* Vol.70 No.2 (August 2006), pp. 755–768, ISSN 0026-895X

Nelson, M.E., Kuryatov, A., Choi, C., Zhou, Y., & Lindstrom, J. (2003). Alternate stoichiometries of α4β2 nicotinic acetylcholine receptors. *Molecular Pharmacology*, Vol.63, No.2 (February 2003), pp. 332-341, ISSN 0026-895X

Sambrook, J., & Russell, D. (2000). *Molecular Cloning: A Laboratory Manual, 3 Vol.* 3rd edition, Cold Spring Harbor Laboratory, ISBN 0879695773, New York

Schröder, B., Reinhardt-Maelicke, S., Schrattenholz, A., McLane, K., Kretschmer, A., Conti-Tronconi, B., & Maelicke, A. (1994). Monoclonal antibodies FK1 and WF6 define two neighboring ligand binding sites on Torpedo acetylcholine receptor alpha-polypeptide. *Journal of Biological Chemistry*, Vol.269, No.14 (April 1994), pp. 10407-10416, ISSN 0021-9258

Smit, A., Syed, N., Schaap, D., van Minnen, J., Klumperman, J., Kits, K., Lodder, H., van der Schors, R., van Elk, R., Sorgedrager, B., Brejc, K., Sixma, T., & Geraerts, W. (2001) A glia-derived acetylcholine-binding protein that modulates synaptic transmission. Nature, Vol.411, No.6835, (May 2001), pp. 261-268, ISSN 0028-0836

Storch, A., Schrattenholz, A., Cooper, J., Abdel Ghani, E., Gutbrod, O., Weber K., Reinhardt, S., Lobron, C., Hermsen, B., Soskiç, V., Pereira, E., Albuquerque, E., Methfessel, C., & Maelicke, A. (1995) Physostigmine, galanthamine and codeine act as 'noncompetitive nicotinic receptor agonists' on clonal rat pheochromocytoma cells.

European Journal of Pharmacology, Vol.290, No.3, (August 1995), pp. 207-219, ISSN 0014-2999

Takacs, Z., Wilhelmsen, K., & Sorota, S. (2001). Snake a-neurotoxin binding site on the Egyptian cobra (*Naja haje*) nicotinic acetylcholine receptor is conserved. *Mol. Biol. Evol.* Vol.18, No.9, pp.1800–1809, ISSN: 0737-4038

Tapia, L., Kuryatov, A., & Lindstrom, J. (2007). Ca^{2+} permeability of the $(\alpha4)_3(\beta2)_2$ stoichiometry exceeds that of $(\alpha4)_2(\beta2)_3$ human acetylcholine receptors. *Molecular Pharmacology*, Vol.71, No.3, (March 2007), pp. 769-776, ISSN 0026-895X

Temburni, M., Blitzblau, R., & Jacob M. (2000). Receptor targeting and heterogeneity at interneuronal nicotinic cholinergic synapses in vivo. *Journal of Physiology,* Vol.525.1 (March 2000), pp. 21-29, ISSN 0022-3751

Thomsen, T., Kaden, B., Fischer, J., Bickel, U., Barz, H., Gusztony, G., Cervos-Navarro, J., & Kewitz, H. (1991). Inhibition of acetylcholinesterase activity in human brain tissue and erythrocytes by galanthamine, physostigmine and tacrine. *European Journal of Clinical Chemistry and Clinical Biochemistry*, Vol.29, No.8, (August 1991), pp. 487-492, ISSN 1437-4331

Tritto, T., Stitzel, J.A., Marks, M.J., Romm, E., & Collins, A.C. (2002). Variability in response to nicotine in the LSXSS RI strains: potential role of polymorphism in alpha4 and alpha6 nicotinic receptor genes. *Pharmacogenetics* Vol.12, No.3 (April 2002), pp. 197–208, ISSN 0960-314X.

Tong, M., Arora, K., White, M. M., & Nichols, R. A. (2011). Role of key aromatic residues in the ligand-binding domain of alpha7 nicotinic receptors in the agonist action of beta-amyloid. *Journal of Biological Chemistry*, Vol.286, No.31, (September 2011), pp. 34373-34381, ISSN 0021-9258

Unwin, N. (2005). Refined structure of the nicotinic acetylcholine receptor at 4A resolution. Journal of Molecular Biology, Vol.346, No.4, (March 2005), pp.967-989, ISSN 0022-2836

Williams, B., Temburni, M., Levey, M., Bertrand, S., Bertrand, D., & Jacob, M. (1998). The long internal loop of the α3 subunit targets nAChR to subdomains within individual synapses on neurones in vivo. *Nature Neuroscience,* Vol. 1 (November 1998), pp. 557-562 ISSN, 1097-6256

Wonnacott, S. (1997). Presynaptic nicotinic ACH receptors. *Trends in Neurosciences,* Vol. 20, No.2 (February 1997), pp. 92-98, ISSN 0166-2236

Xu, J., Zhu, Y., & Heinemann, S. (2006). Identification of sequence motifs that target neuronal nicotinic receptors to dendrites and axons. *The journal of neuroscience,* Vol.26, No.38, (September 2006), pp. 9780- 9793, ISSN 0270-6474

Young, G., Zwart, R., Walker, A., Sher, E., & Millar, N. (2008). Potentiation of alpha7 nicotinic acetylcholine receptors via an allosteric transmembrane site. *Proceedings of the National Acadademy of Sciences USA,* Vol.105, No.38, (September 2008) pp. 14686-1491, ISSN 1091-6490

Zhou, Y., Nelson, M.E., Kuryatov, A., Choi, C., Cooper, J., & Lindstrom, J. (2003). Human α4 β2 acetylcholine receptors formed from linked subunits. *The Journal of Neuroscience*, Vol.23, Nr.27 (October 2003), pp. 9004 –9015, ISSN 0270-6474.

Zouridakis, M., Zisimopoulou, P., Poulas, K., & Tzartos, S. (2009). Recent advances in understanding the structure of nicotinic acetylcholine receptors. *IUBMB Life*, Vol.61, No.4, (April 2009), pp. 407-423, ISSN 1521-6551

Site-Directed Mutagenesis as a Tool to Characterize Specificity in Thiol-Based Redox Interactions Between Proteins and Substrates

Luis Eduardo S. Netto[1] and Marcos Antonio Oliveira[2]
[1]Instituto de Biociências – Universidade de Sao Paulo
[2]Universidade Estadual Paulista – Campus do Litoral Paulista
Brazil

1. Introduction

Redox pathways are involved in several processes in biology, such as signal transduction, regulation of gene expression, oxidative stress and energy metabolism. Proteins are the central mediators of electron transfer processes. Many of these proteins rely on non-proteinaceous redox cofactors (such as NAD^+; FAD; heme; or Cu, Fe or other transition metals) for their redox activity. In contrast, other proteins use cysteine residues for this property (Netto et al., 2007). The amino acid cysteine has low reactivity for redox transitions (Winterbourn and Metodiewa, 1999; Wood et al., 2003; Marino and Gladishev, 2011). However, protein folding can generate environments in which cysteine residues are reactive. Examples are reduction (or isomerization or formation) of disulfide bonds, reduction of methionine thioesther-sulfoxide, degradation of peptide bonds, peroxide reduction, and others (Lindahl et al., 2011).

Glutathione (GSH) is by far the major non-proteinaceous thiol in cells that plays a central role in several redox processes, such as xenobiotic excretion and antioxidant defense. GSH is composed of three amino acids: glutamate, cysteine and glycine. GSH synthesis is performed in two steps, which occur mainly in liver cells. In the first step, the γ-glutamylcysteine synthetase enzyme catalyses the rate-limiting step, with the formation of an unusual peptide bond between the gamma-carboxyl group of the side chain of glutamate and the primary amino group of a cysteine in an ATP-dependent reaction. Then GSH synthetase catalyzes the formation of a peptide bond between the carboxyl group of cysteine (from the dipeptide γ-glutamylcysteine) with the amino group of a glycine. This tripeptide is considered to be the major redox buffer in mammalian cells, and it is a substrate of two relevant groups of enzymes: glutathione transferases and glutathione peroxidases. It is thought that most healthy cells have higher GSH/GSSG (glutathione disulfide) ratios than sick ones (Berndt et al., 2007; Jacob et al., 2003; Jones, 2006).

Besides glutathione, there is a high number of Cys-based (see Table 1 in the chapter by Figurski et al. for amino acid abbreviations) redox proteins. These Cys-based proteins are very versatile. The oxidation states of their sulfur atoms can vary from +6 to -2 (Jacob et al., 2003). One of the most widespread functions of Cys-based proteins is the catalysis of thiol-

disulfide exchange reactions, by which these enzymes control the oxidation state (dithiol or disulfide) in their targets/substrates (Netto et al., 2007). Thioredoxins and glutaredoxins (also known as thioltransferases) are disulfide reductases, whereas protein disulfide isomerases are also involved in the oxidation of dithiols and/or the shuffling of disulfides. Furthermore, Cys-based proteins can also control the levels of other eletrophiles, such as peroxides (in the cases of peroxiredoxins and GSH peroxidases), xenobiotics (GSH transferases) and sulfoxides (methionine sulfoxide reductases). Therefore, this large repertoire of proteins, together with GSH, is part of a complex network that, in a dynamic fashion, controls intracellular redox balance.

The classical view is that the reactivity of a cysteine sulfhydryl group is related to its pK_a, since its deprotonated form (thiolate = RS$^-$) is more nucleophilic and, therefore, reacts faster than the equivalent protonated form (R-SH). According to this view, the lower the pK_a of a thiol, the higher will be the availability of the more nucleophilic species, the thiolate. The sulfhydryl groups of most cysteines (either linked to a polypeptide backbone or the free amino acid) possess low reactivity, which has been related to the fact that their pK_a values are around 8.5 (Benesch and Benesch, 1955). In contrast, most redox proteins possess a reactive cysteine that is stabilized in the thiolate form by a basic residue - in most cases by a lysine, histidine or arginine residue (Copley et al., 2004).

However, a decrease in the pK_a value of several orders of magnitude would give rise to an increase in thiolate concentration, with a maximum increase of one order of magnitude (Ferrer-Sueta et al., 2011). However, as an example of Cys-based redox proteins, peroxiredoxins reacts one to ten million times faster with peroxides than the corresponding reaction with the free amino acid cysteine (Winterbourn and Hampton, 2008). Therefore, factors other than thiolate availability should be taken into account. Indeed the stabilization of the transition state by active-site residues was recently proposed to be the catalytic power of peroxiredoxins. Site-directed mutagenesis was employed to test these hypotheses (Hall et al., 2010; Nagy et al., 2011).

It is clear that Cys-based proteins present reactive Cys residues to specific reactions, most of them being the nucleophilic substitution (S_N2) type. Indeed peroxiredoxins are effective in reducing peroxides; but they are poor in reducing other eletrophiles, such as chroloamines (Peskin et al., 2007). In contrast, glutaredoxins are powerful GSH-dependent disulfide reductants. In spite of the fact that their reactive Cys residues have low pK_a values (<4.0), these oxido-reductases are unable to reduce O-O bonds (Discola et al., 2009).

In line with the observation that Cys-based circuits display high specificity, a new concept of oxidative stress was proposed by Jones (2006). Since several antioxidant interventions failed to have therapeutic effects, it was thought that oxidative stress leads to alterations of discrete pathways, rather than to an overall redox imbalance. Therefore, perhaps an antioxidant intervention would be more effective if it were directed to specific pathways, *i.e.*, the oxidative stress would be better defined as a disruption of a specific pathway (Jones, 2006). For instance, some signal transduction pathways are activated by oxidized, but not by reduced, thioredoxin (Trx) (Berndt et al., 2007), *e.g.*, only reduced Trx1 binds Ask-1, thereby inhibiting the kinase activity of Ask-1, whereas oxidation of Trx1 leads to dissociation of the complex and activation of Ask-1, which can trigger apoptosis (Saitoh et al., 1998). Another example is the activation of NF-κB. Binding of subunit p50 to its target sequence in DNA requires the reduction of a single cysteinyl residue in the nucleus by Trx1 (Matthews et al.,

1992; Hayashi et al., 1993). Also circadian cycles depend on Cys-based redox signaling (O'Neill et al., 2011). The specificity of these pathways involves protein–protein interactions. The identification of the amino acids involved is relevant for the comprehension of patho-physiological phenomena.

2. Approach

Our group has followed an approach for studying Cys-based redox systems that involves multiple methodologies. Initially we decided to study yeast thiol-based systems, such as the thioredoxin and glutaredoxin ones. Yeast is a convenient system because it is very amenable to genetic manipulation. We obtained from *EUROSCARF* (http://web.unifrankfurt.de/ fb15/mikro/euroscarf/complete.html) a collection of about four thousand strains, each one with a single deletion of a specific gene. In the case of the cytosolic thioredoxin system from *Saccharomyces cerevisiae*, we have elucidated the three-dimensional structures of all proteins by NMR (*i.e.*, ScTrx1 and ScTrx2), in collaboration with the group of Dr. Almeida (Pinheiro et al., 2008; Amorim et al., 2007), as well as by crystallography, *i.e.*, ScTrxR1, also known as Trr1 (Oliveira et al., 2010). We have also elucidated structures of the yeast glutaredoxins (Discola et al., 2009). With this information, together with available public structural, biochemical, and enzymatic data, we were able to generate a hypothesis about the mechanistic aspects of these redox pathways that could be tested by site-directed mutagenesis.

Recently we have done this approach with the bacterial thiol-based systems, as a consequence of our participation in the genome sequencing project for the phytopathogen *Xylella fastidiosa* (Simpson et al., 2000). Again site-directed mutagenesis was employed to test a hypothesis generated by experimental work. The hypothesis proposed the involvement of amino acid residues in catalysis. The interpretation of the data was not always straightforward, probably because amino acids can interact with other residues in a protein to give unpredictable effects. In the following sections, we will describe what we have learned in different thiol-based systems, using site-directed mutagenesis to test hypotheses.

3. Characterization of Cys-based proteins

3. 1. Cys-based proteins from *Saccharomyces cerevisiae*

3.1.1 Molecular aspects of specific redox protein-protein interactions in the cytosolic thioredoxin system

Thioredoxin appears to be an ancient protein, since it is widespread among all living organisms. These small proteins (12–13 kDa) possess disulfide reductase activity endowed by two vicinal cysteines present in a CXXC residue motif - typically CGPC. The cysteines are used to reduce target proteins that are recognized by other domains of thioredoxin polypeptide. The reduction of target proteins results in a disulfide bridge between the two cysteines from the thioredoxin CXXC motif, which is then reduced by thioredoxin reductase using reducing equivalents from NADPH. Some of the target proteins of thioredoxin include ribonucleotide reductase (important for DNA synthesis), methionine sulfoxide reductase, peroxiredoxins and transcription factors such as p53 and NF-*k*B (reviewed by Powis and Montfort, 2001). Therefore, the thioredoxin system is composed of NADPH, thioredoxin reductase and thioredoxin.

Proteins endowed with thioredoxin reductase activity are also widespread and comprise enzymes with different redox centers. Thioredoxin reductase enzymes belong to the nucleotide pyridine disulfide oxidoreductase family, which includes glutathione reductase, alkyl hydroperoxide reductase F (AhpF), and lipoamide dehydrogenase (Williams et al., 2000). Constituents of this family are homodimeric flavoproteins that also contain one or two dithiol-disulfide motifs - CXXXXC and/or CXXC. Thioredoxin reductase catalyzes the disulfide reduction of oxidized thioredoxin, using NADPH via the FAD molecule and the redox-active cysteine residues (Waksman et al., 1994). Initially thioredoxin reductases were divided into two sub-groups (low and high molecular weight TrxR) based on the absence or presence of a dimerization domain (Williams et al., 2000). However, there are some thioredoxin reductase enzymes with distinct extra domains that do not fit well into these two classes. Therefore, based on structural and biochemical considerations, we proposed that thioredoxin reductases should be divided into five sub-classes. In spite all these differences, all thioredoxin reductases share a common core, containing two domains (a NADPH-binding domain and a FAD-binding domain) and two redox centers: a FAD molecule and a dithiol-disulfide group (Oliveira et al., 2010).

Cytosolic thioredoxin system from yeast *Saccharomyces cerevisiae* is composed of one low molecular weight thioredoxin reductase (yTrxR1) and two thioredoxin enzymes (yTrx1 and yTrx2). Interestingly, most thioredoxin systems are composed of only one thioredoxin. yTrx1 and yTrx2 share 78% amino-acid identity and were initially considered as fully redundant enzymes. We are investigating whether these two oxido-reductases have specific roles. The expression of yTrx2 is highly inducible by peroxides in a process mediated by Yap1, whereas expression of yTrx1 is more constitutive (Kuge and Jones, 1994; Lee et al., 1999). The relevance of the cytosolic thioredoxin system from yeast can be attested to by the fact that deletion of ScTrxR1 gene renders yeast inviable (Giaever et al., 2002). Some of their targets include at least three peroxiredoxins (Tsa1, Tsa2 and Ahp1), methionine sulfoxide reductase, ribonucleotide reductase, a Cys-based peroxidase involved in the oxidative stress response (Gpx3-Orp1) (Fourquet et al., 2008) and the system involved in sulfate assimilation (PAPS). yTrx1 and yTrx2 present specificity towards their targets, *i.e.*, they cannot reduce all disulfide bonds. Some protein-protein interactions should occur in order for a protein disulfide reduction to take place. Once yTrx1 (or yTrx2) reduces a target disulfide bond, it gets oxidized. While studying the reduction of yTrx1 (or yTrx2) by yTrxR1, we observed that this flavoprotein exhibited remarkable specificity, *i.e.*, yTrxR1 only reduces yeast thioredoxins (yTrx1 and yTrx2), but not mammalian or bacterial thioredoxins (Oliveira et al., 2010). yTrxR1 can also reduce yeast mitochondrial thioredoxin (yTrx3). Probably this species-specificity phenomenon involves recognition of certain amino acid residues by yTrxR1 through protein-protein interactions.

The identification of protein–protein interactions is a major challenge in cell biology. The interactions for the various pathways are specific, directing signals to specific targets. Thiol-based systems are emerging as relevant pathways in signal transduction. Because there are several thiol-disulfide oxido-reductases in each genome, it is reasonable to think that each one of them interacts with different partners. Although all thioredoxin reductases catalyze the same overall reaction (*i.e.*, reduction of thioredoxin at the expense of NADPH), apparently the species-specificity phenomenon is restricted to the low molecular weight enzymes. This is probably because high molecular weight thioredoxin reductases have an external selenocysteine residue (a cysteine analog with a selenium instead of the sulfur atom
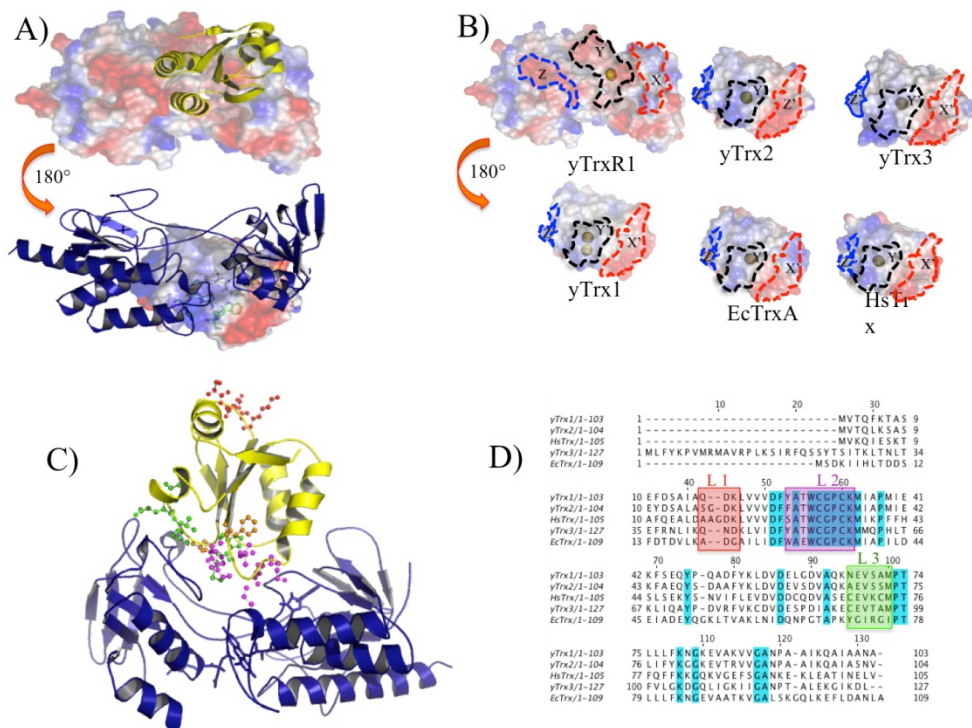
Fig. 1. **Structural features involved in the species–specificity phenomenon**. (A) Theoretical representation of the yTrx-yTrxR complex. The electrostatic surface of yTrxR1 is depicted at the top (Red = negatively charged atoms; blue = positively charged atoms; white = no charge) and yTrx1 is represented by the cartoon (yellow). At the bottom is shown the yTrxr1 electrostatic surface and yTrxr1 (cartoon-blue). (B) Comparison of electrostatic surfaces among five distinct thioredoxin enzymes, three of them from yeast. yTrx = thioredoxin from *S. cerevisiae*; EcTrxA = thioredoxin A from *E. coli*; and HsTrx = thioredoxin from *H. sapiens*. (C) Loop 2 (grape) and loop 3 (green) are in close proximity to thioredoxin reductase. (D) Amino-acid alignment among five thioredoxin enzymes. Three loops are candidates for physical interaction with thioredoxin reductase. Since the loop 3 (L3) amino-acid sequences display higher variability than the loop 2 (L2) sequences, loop 3 is implicated in the species–specificity phenomenon.

in the side chain) that can reduce target substrates with low physical interaction. Therefore, species specificity probably requires extensive protein-protein interactions. As mentioned above, thioredoxin reductase 1 from *S. cerevisiae* (yTrxR1 = yTrr1) can reduce cytosolic and mitochondrial thioredoxins from yeast; but it cannot reduce thioredoxin from *Escherichia coli* or from *Homo sapiens*.

Since thioredoxins present high sequence similarity, we are interested in identifying factors involved in species–specific interactions. As mentioned above, we obtained the structures of all proteins belonging to the cytosolic thioredoxin system from the yeast *S. cerevisiae* (yTrx1 = PDB 2I9H; yTrx2 = PDB 2HSY; yTrxR1 = PDB 3ITJ). This allowed us to test models for protein–protein interactions. The analysis indicated that complementary electrostatic surfaces between yTrxR1 and yTrxs are partially responsible for the species-specificity phenomenon (Fig. 1A). Furthermore, residues that belong to loop 3 appear to be directly related to protein-protein interactions (Fig. 1B). Indeed site-directed mutagenesis was a valuable tool for testing hypotheses raised by crystal structure analysis and by biochemical assays (Oliveira et al., 2010).

3.1.2 Aspects involved in the high oxido-reductase activity of yeast Glutaredoxin 2 in comparison with yeast Glutaredoxin 1

Like thioredoxin, glutaredoxin enzymes are thiol-disulfide oxido-reductases, whose genes are widespread among both eukaryotic and prokaryotic genomes. These small, heat stable enzymes are ubiquitously distributed and endowed with disulfide reductase activity (Discola et al., 2009). In the case of the yeast *S. cerevisiae*, eight isoforms have been identified so far. Three of them are dithiolic glutaredoxins, with two vicinal cysteines in a CXXC motif (mostly CPYC). The other five are monothiolic enzymes. They are characterized less well and will not be considered here.

yGrx1 and yGrx2 are the two major dithiolic glutaredoxins from *S. cerevisiae*. They display high amino acid sequence similarity to each other (85%). These enzymes can reduce disulfide bonds through two distinct mechanisms. In the most studied one, a mixed disulfide between a target protein and GSH is reduced by the monothiolic mechanism. In this case, only the N-terminal Cys of the CXXC motif takes part (Figure 2, reactions f and g). The most used assay to measure glutaredoxin activity, the HED (β-hydroxyethyl disulfide) assay, operates through this monothiolic pathway. Alternatively, a glutaredoxin with two Cys residues can reduce disulfides through the dithiolic pathway (Figure 2, reactions a–e).

The monothiolic pathway has received increased attention because it appears to control the levels of glutathionylated enzymes in cells. Glutathiolation is an emerging post-translational modification. This modication protects reactive Cys residues from irreversible oxidation to the sulfinic (RSO_2H) or sulfonic (RSO_3H) states. In analogy to phosphatases, glutaredoxins catalyze removal of the glutathionyl moiety and thereby regulate signaling processes (Gallogly and Mieyal, 2007). Like phosphorylation and other post-translational modifications, glutathionylation is reversible. To evaluate if a specific glutathiolation event is regulatory, some criteria were proposed, such as (1) a change in the activity of the target protein; (2) occurrence in response to a stimulus; or (3) occurrence at a physiological GSH/GSSG ratio (normally high in the cytosol), with both the modification and its

disappearance being fast (Gallogly and Mieyal, 2007). Some proteins that fulfill these criteria are actin (Wang et al., 2001; Wang et al., 2003), Ras (Adachi et al., 2004) and Protein Tyrosine Phosphatase (Kanda et al., 2006).
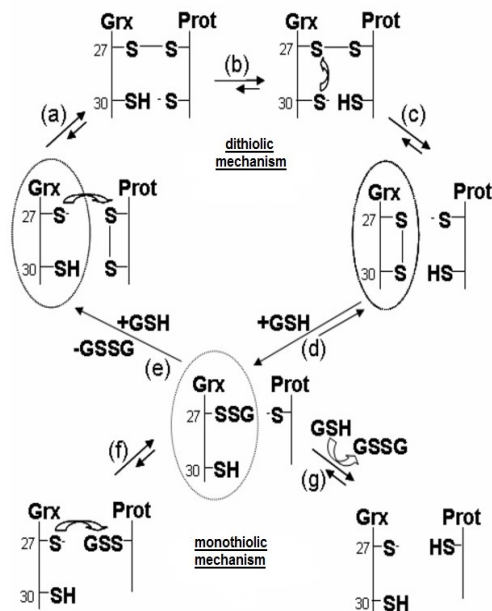
Fig. 2. **Mechanisms of disulfide reduction by glutaredoxin**. Dithiolic: Reactive Cys (in thiolate form) from glutaredoxin (Grx) performs a nucleophilic attack (S_N2 type) on a disulfide of a target proteins, leading to the formation of a mixed disulfide (reaction a); a thiolate is formed in the second Cys (reaction b) and this thiolate performs a nucleophilic attack (S_N2 type) on the mixed disulfide, generating a intramolecular disulfide on glutaredoxin (reaction c). Reduction of glutaredoxin takes places by two consecutive reactions with GSH (reactions d and f). Monothiolic: Reactive Cys (in thiolate form) from glutaredoxin (Grx) performs a nucleophilic attack (S_N2 type) on a mixed disulfide between GSH and a target protein, leading to the formation of glutathiolated glutaredoxin (reaction f); reduction takes place by reaction with a second GSH molecule (reaction g). Reaction g is considered the rate-limiting step in the monothiolic pathway (Srinivasan et al., 1997).

In collaboration with Dr. Demasi, our group has shown that the proteasome is also post-translationally modified by glutathiolation in response to oxidative stress (Demasi et al., 2001; Demasi et al., 2003) and also that glutaredoxin can reduce the mixed disulfide bond between the proteasome and GSH (Silva et al., 2008). Site-directed mutagenesis of Cys residues in the 20S proteasome is underway in order to clarify mechanistic details of this process. Therefore, it is relevant to comprehend features that control the deglutathionylase activity of glutaredoxins (Figure 2, reactions f and g) to better appreciate the function of this post-translational modification in cell biology.

In this regard, it was relevant to observe that the two main dithiolic glutaredoxins from yeast display markedly distinct monothiolic (HED assay) specific activities (Discola et al., 2009). Although these two enzymes share a high degree of similarity in their amino acid sequences, yGrx2 is two orders of magnitude more active than yGrx1 (Discola et al, 2009). These data are consistent with results from studies with knockout strains (*i.e.*, strains with null alleles) that indicate that yGrx2 accounts for most of the oxido-reductase activity observed in yeast extracts (Luikenhuis et al., 1998). In an attempt to gain insights on this phenomenon, our group obtained two crystallographic structures of yGrx2 (intramolecular disulfide = PDB 3D4M; mixed disulfide with GSH = PDB 3D5J, both of which are related to the short form yGrx2) and compared them with the crystal structures of yGrx1 (reduced = PDB 2JAD; mixed disulfide with GSH = PDB 2JAC) available in the literature (Håkansson and Winther, 2007). The overall structures are highly similar (Fig. 3A). However, differences in the active sites were hypothesized to be involved in the distinct catalytical efficiencies between yGrx1 and yGRx2 (Fig. 3B). In order to obtain the structures of these complexes, it was necessary to mutate the C-terminal Cys (Cys30) to Ser in order to slow reaction d (reverse) (see d in Fig. 2). The analysis of the structures of yGrx1C30S and yGrx2C30S (short isoform) in complex with GSH revealed that the distances between Ser30 (Cys 30 in both yGrx1 and yGrx2, short isoform, wild-type proteins) and the reactive Cys (Cys47) are markedly distinct (3.47 Å in yGrx1C30S and 5.14 Å in yGrx2C30S).
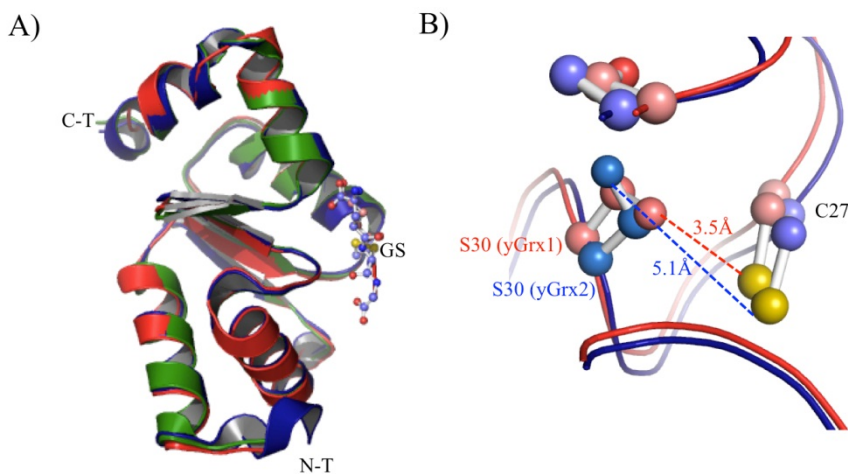
Fig. 3. **Crystal structures of dithiolic glutaredoxins from yeast**. (A) Cartoon representation of overall glutaredoxin structures. (Red= yGrx1C30S, PDB code 2JAC; Green = yGrx2 disulfide, PDB code 3D4M; Blue = yGrx2C30S, PDB code = 3D5J); (B) Active sites of the complex with glutathione. The distances between the resolving Cys and the reactive cysteine (sulfur atoms colored in yellow) are shown by dashed lines. Colors are defined in A.

In principle, any factor that would slow reaction d (reverse) (see d in Fig.2) should favor reaction g and, consequently, the monothiolic activity, which is the mechanism by which the HED assay operates. According to this hypothesis, anything that increases the distance between the two sulfur atoms of the CXXC motif should favor the monothiolic activity and, consequently, the rates in the HED assay. Accordingly, Ser23 is in close proximity to Ser30

in yGRx2 (short isoform). This interaction between two serine residues is probably stabilizing a configuration, in which the distances between the two sulfur atoms would be high in the wild-type yGrx2, thereby accounting for its high catalytical efficiency (Discola et al., 2009). In contrast, yGrx1 has an Ala residue at position 23; and this side chain cannot make a salt bridge. In this case, the distances for residues 27 and 30 (cysteines in the wild-type protein) are short. A short distance between the sulfur atoms favors the dithiolic mechanism over the monothiolic mechanism (Fig. 2). This hypothesis was tested by site-directed mutagenesis. Indeed the relevance of a serine at position 23 for the monothiolic activity was demonstrated (Discola et al., 2009).

Probably biochemical and structural features other than a serine/alanine residue at position 23 are involved with the higher catalytical efficiency of yGrx2 over yGrx1. Indeed Ser89 in yGrx2 (short isoform) and Asp89 in yGrx1 were recently implicated in the different catalytical properties of these two oxido-reductases (Li et al., 2010). Ser89 is involved in the binding of GSH in glutaredoxin (Discola et al., 2009). The authors also employed site-directed mutagenesis to show that their hypothesis was correct (Li et al., 2010).

Since glutathionylation is emerging as a key concept in redox signaling, it is reasonable that the combined approach of biochemical and structural assays together with site-directed mutagenesis will be followed to establish the involvement of other factors in this post-translational modification.

3.1.3 Site-directed mutagenesis to characterize residues that allow reduction of 1-Cys peroxiredoxin by ascorbate

Peroxiredoxins are ubiquitous, Cys-based peroxidases, whose importance is underlined by their high abundance and their involvement in multiple cellular processes probably related to their capacity to decompose hydroperoxides (Rhee and Woo, 2011; Wood et al., 2003). Indeed several groups have shown independently that peroxiredoxins compete with heme-peroxidases and Se-GSH peroxidases for hydroperoxides (Horta et al., 2010; Ogosucu et al., 2007; Parsonage et al., 2005; Toledo Jr et al., 2011). As a consequence of their high abundance and reactivity, peroxiredoxins are major sinks for peroxides (Winterbourn and Hampton, 2008).

A peroxiredoxin can be classified as a 1-Cys or 2-Cys Prx, depending on the number of Cys residues that participate in the catalytic cycle (Rhee and Woo, 2011; Wood et al., 2003). For most 2-Cys Prxs, the biological reductant is thioredoxin. For 1-Cys Prxs, the situation is far more complex. In many cases the identity of the reductant is not known. Our group has shown that ascorbate can support the peroxidase activity of 1-Cys Prx. This represented a change of the thiol-specific antioxidant paradigm (Monteiro et al., 2007).

Since 1-Cys and 2-Cys Prxs share amino acid sequence similarity, we asked ourselves which amino acids are responsible for the ability of 1-Cys Prx enzymes to accept ascorbate as the electron donor. Through a multiple approach involving amino acid sequence alignment, mass spectrometry and enzymatic assays, we postulated that two features are required: (1) the absence of a Cys involved in disulfide formation (resolving Cys) and (2) the presence of a His residue fully conserved in 1-Cys Prxs and absent in the 2-Cys Prx counterparts. By site-directed mutagenesis, we were able to engineer a 2-Cys Prx to be reducible by ascorbate by taking into account the two factors described above (Monteiro et al., 2007). Further

studies are underway in order to comprehend the physiological significance of ascorbating acting as a reducing agent for 1-Cys Prx.

3.2 Cys-based proteins from bacteria

3.2.1 Residues of Peroxiredoxin Q involved in redox-dependent secondary structure change

Our group is also interested in the analyses of Cys-based proteins from bacteria, as a consequence of our participation in the genome-sequencing project of the bacterium *Xylella fastidiosa*. *X. fastidiosa* is a gram-negative bacterium that is the etiologic agent of several plant diseases, such as Citrus Variegated Chlorosis, which imposes great losses in orange production in Brazil (Lambais, 2000). *X. fastidiosa* also causes Pierce disease in grapevines, phony peach disease, and leaf scorch diseases in almond and oleander (Hendson et al., 2001).

Animal and plant hosts generate oxidative insults against pathogens, such as *X. fastidiosa*, in an attempt to avoid infection. The oxidants include hydrogen peroxide, organic hydroperoxides, and peroxynitrite (Koszelak-Rosenblum et al., 2008; Tenhaken et al., 1999; Wrzaczek et al., 2009). To counteract this host response, bacteria present a large repertoire of antioxidants, including Cys-based peroxidases (Horta et al., 2010). Therefore, in principle, any intervention that results in the decrease of antioxidants from pathogens can have a therapeutic property. Indeed the mechanism of action of several antibiotics is based on the generation of oxidants (Kohanski et al., 2010).

After completion of the genome-sequencing project of *X. fastidiosa* (Simpson et al., 2000), we decided to characterize Cys-based peroxidases from this plant pathogen. Analysis of the *X. fastidiosa* genome revealed the presence of five genes that encode proteins potentially involved in hydroperoxide decomposition: one catalase, one glutathione peroxidase (GPx), one organic hydroperoxide resistance protein (Ohr) and two peroxiredoxins (AhpC and PrxQ), both of which probably display the 2-Cys Prx mechanism (Horta et al., 2010). All of them, except GPx protein, were identified in the whole-cell extract and extracellular fraction of the citrus-isolated strain 9a5c (Smolka et al., 2003). We decided to characterize peroxiredoxins from *X. fastidiosa*.

As noted above (Section 3.1.3), peroxiredoxins can be classified into two groups, 2-Cys Prxs and 1-Cys Prxs, depending on the mechanism of catalysis. Besides this mechanistic classification, others were proposed that are based on amino acid sequence similarity. Later structural features were incorporated into the classification proposals. They provided insights on the evolution of proteins within the Trx superfamily, which includes the Prxs (Copley et al., 2004; Nelson et al., 2011). Adopting the classification described in Copley et al. (2004) for Prx classes, class 1 is the most ancestral, but the least characterized of all 4 classes. The other 3 classes of Prxs were derived from those of class 1. We therefore decided to investigate a class 1 Prx, peroxiredoxin Q (XfPrxQ) from *X. fastidiosa* (Horta et al., 2010).

Historically all classes of Prx have been considered only moderately reactive. The reason was that their catalytic efficiencies (k_{cat}/Km) toward hydroperoxides, as determined by steady-state kinetics, were in the 10^4-10^5 M^{-1} s^{-1} range. In contrast, selenocysteine-containing GPx (10^8 M^{-1} s^{-1}) and heme-containing catalases (10^6 M^{-1} s^{-1}) presented considerably higher

values (Wood et al., 2003). More recently, with the development of new assays, Prx enzymes were considered as reactive as selenium- and heme-containing proteins (Ogusucu et al., 2007; Parsonage et al., 2005; Trujillo et al., 2007). At that time, only class 3 Prx enzymes and class 4 Prx enzymes (composed mostly of typical 2-Cys Prx enzymes, but also some 1-Cys Prx proteins) were analyzed by these assays. Consequently, the catalytic efficiencies for enzymes of the other Prx classes remained to be determined. Through a competitive-kinetics approach (Toledo et al., 2011), we demonstrated that the second-order rate constants of the peroxidase reactions of XfPrxQ with hydrogen peroxide and peroxynitrite lay in the order of 10^7 and 10^6 M^{-1} s^{-1}, respectively. These reactions are as fast as the most efficient peroxidases. Furthermore, the catalytic cycle of XfPrxQ was elucidated by multiple approaches, such as X-ray crystallography, circular dichroism, biochemical assays, mass spectrometry, and site-directed mutagenesis. Using data obtained by site-directed mutagenesis, we were able to propose a model for the redox-dependent structural changes in PrxQ proteins (Fig. 4) that was consistent with all of our data and data from the literature (Horta et al., 2010). Site-
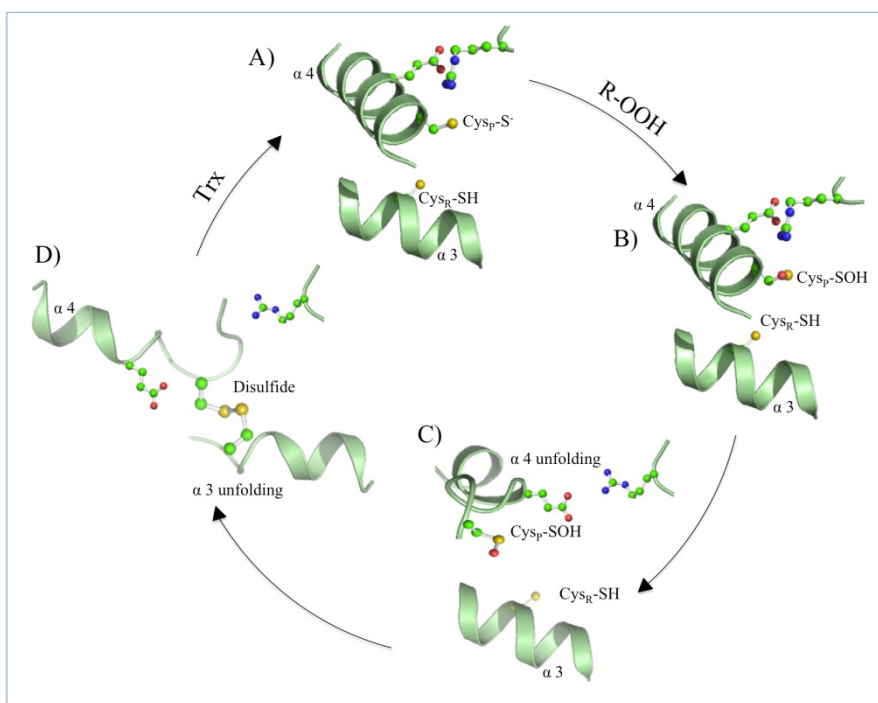
Fig. 4. **Model for the redox-dependent conformational changes in XfPrxQ**. Proposed sequence of structure snapshots along the catalytic cycle of the PrxQ subfamily proteins. The protein is represented in cartoon (light green) and residues are represented as sticks. Atoms are colored as follow: C=green, O=red, N=blue and S=orange. Peroxidatic and resolving cysteines are indicated as Cys_P and Cys_R, respectively. (A) Reduced species based on the crystal structure of XfPrxQ C47S. (B) and (C) Hypothetical conformational intermediates based on circular dichroism data (Horta et al., 2010). (D) Oxidized species based on the crystal structure of Bcp from *X. campestris* (PDB code = 3GKK).

directed mutagenesis revealed that Cys47 is the center responsible for the changes in secondary structure measured by circular dichroism (Horta et al., 2010).

3.2.2 Site-directed mutagenesis as a valuable tool to characterize a new antioxidant protein: Organic Hydroperoxide Resistance Protein

During the annotation of the *X. fastidiosa* genome (Simpson et al., 2000), the function of one gene caught our attention: *ohr*. What was reported at that time was that the deletion of *ohr* gene in *X. campestris pv. Phaseoli* rendered those cells sensitive to the oxidative insult by organic hydroperoxides, but not by hydrogen peroxide (Mongkolsuk et al., 1998). Furthermore, the transcription of the *ohr* gene was specifically induced by organic hydroperoxides, such as *tert*-butyl-hydroperoxide (Mongkolsuk et al., 1998). Therefore, this gene was named *ohr* (organic hydroperoxide resistance gene). However, the biological activity for the corresponding protein was not known.

Alignment of the deduced amino acid sequences of putative Ohr proteins from several bacteria revealed the presence of two fully conserved Cys residues. Therefore, we hypothesized that the *ohr* gene probably encodes a Cys-based, thiol-dependent peroxidase. In order to test this hypothesis, recombinant Ohr was obtained by cloning and expressing the *ohr* gene from *X. fastidiosa*. Indeed it displayed a thiol-dependent peroxidase (Cussiol et al., 2003). Remarkably, the peroxidase activity of Ohr was specifically supported by dithiols, such as DTT, but not by monothiols, such as 2-mercapthoethanol. In contrast, both mono- and dithiols support the enzymatic activity of peroxiredoxins and glutathione peroxidase. Furthermore, Ohr shows high activity towards organic hydroperoxides. Site-directed mutagenesis of the two conserved Cys residues unequivocally revealed that the most N-terminal one (Cys62) is the redox center (Cussiol et al., 2003). Another major achievement was the elucidation of the X-ray structure of Ohr (Oliveira et al., 2006). It showed a unique structure. The "Ohr fold" was quite different from the thioredoxin fold that is present in peroxiredoxin and the GSH peroxidases. At the same time, the structure and biochemical activity of the Ohr from *Pseudomonas aeruginosa* was elucidated (Lesniak et al., 2002). Essentially it has the same features described above.

In contrast to peroxiredoxins and GSH peroxidases, in which the reactive Cys residue is solvent-exposed, the reactive Cys residue in Ohr is buried in the polypeptide chain. The microenvironment where the reactive Cys is located is surrounded by several hydrophobic residues that probably confer to Ohr higher affinity for organic hydroperoxides. Proteins with folds similar to Ohr are only present in bacteria. This fact might indicate that this peroxidase is a target for drug development. Another unique property of Ohr is that only lipoamide, and neither GSH nor thioredoxin can support its Cys-based peroxidase activity (Cussiol et al., 2010). Due to these distinct properties, we are searching for Ohr inhibitors. Therefore, we are pursuing the characterization of enzyme-substrate interactions. We found in the active site of Ohr an electronic density of the polyethylene glycol molecule. Polyethylene glycol is a polymeric compound with elongated shape that was used in the crystallization trials. Since peroxides derived from fatty acids present an elongated structure and fit very well into this electronic density, we proposed that this kind of substrate may be the physiological target of the Ohr enzyme. Amino acid residues possibly involved in enzyme–substrate interactions were identified (Oliveira et al., 2006). Currently site-directed mutagenesis of these residues is underway in order to test the hypothesis that they are

involved in the binding of organic hydroperoxides. This information will be relevant in the search for Ohr inhibitors. We have already found some chemicals that can inhibit Ohr and also inhibit the growth of of *X. fastidiosa*.

4. Concluding remarks

In the characterization of Cys–based proteins, the involvement of amino acids in catalysis was analyzed by several enzymatic and biochemical assays, as well as by X-ray crystallography. Using an approach that combined these methodologies with site-directed mutagenesis allowed several hypotheses to be raised. Research is currently testing them.

5. References

Adachi T, Pimentel DR, Heibeck T, Hou X, Lee YJ, Jiang B, Ido Y, Cohen RA (2004) S-Glutathiolation of Ras mediates redox-sensitive signaling by angiotensin II in vascular smooth muscle cells. *J Biol Chem.* 279, 29857–29862.

Amorim GC, Pinheiro AS, Netto LE, Valente AP, Almeida FC (2007) NMR solution structure of the reduced form of thioredoxin 2 from *Saccharomyces cerevisiae*. *J Biomol NMR.* 38, 99–104.

Benesch RE, Benesch R (1955) The acid strength of the -SH group in cysteine and related compounds. *J Amer Chem Soc.* 77, 5877–5881.

Berndt C, Lillig CH, Holmgren A (2007) Thiol-based mechanisms of the thioredoxin and glutaredoxin systems: implications for diseases in the cardiovascular system. *Am J Physiol Heart Circ Physiol.* 292, H1227–1236.

Copley SD, Novak WR, Babbitt PC (2004) Divergence of function in the thioredoxin fold suprafamily: evidence for evolution of peroxiredoxins from a thioredoxin-like ancestor. *Biochemistry,* 43, 13981–13995.

Cussiol JR, Alegria TG, Szweda LI, Netto LE (2010) Ohr (organic hydroperoxide resistance protein) possesses a previously undescribed activity, lipoyl-dependent peroxidase. *J Biol Chem.* 285, 21943–21950.

Cussiol JR, Alves SV, Oliveira MA, Netto LE (2003) Organic hydroperoxide resistance gene encodes a thiol-dependent peroxidase. *J Biol Chem.* 278, 11570–11578.

Demasi M, Shringarpure R, Davies KJ (2001) Glutathiolation of the proteasome is enhanced by proteolytic inhibitors. *Arch Biochem Biophys.* 389, 254–263.

Demasi M, Silva GM, Netto LE (2003) 20 S proteasome from *Saccharomyces cerevisiae* is responsive to redox modifications and is S-glutathionylated. *J Biol Chem.* 278, 679–685.

Discola KF, Oliveira MA, Rosa Cussiol JR, Monteiro G, Bárcena JA, Porras P, Padilla CA, Guimarães BG, Netto LE (2009) Structural aspects of the distinct biochemical properties of glutaredoxin 1 and glutaredoxin 2 from *Saccharomyces cerevisiae*. *J Mol Biol.* 385, 889–901.

Ferrer-Sueta G, Manta B, Botti H, Radi R, Trujillo M, Denicola A (2011) Factors affecting protein thiol reactivity and specificity in peroxide reduction. *Chem Res Toxicol.* 24, 434–450.

Gallogly MM, Mieyal JJ (2007) Mechanisms of reversible protein glutathionylation in redox signaling and oxidative stress. *Curr Opin Pharmacol.* 7, 381–391.

Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, Dow S, Lucau-Danila A, Anderson K, André B, Arkin AP, Astromoff A, El-Bakkoury M, Bangham R, Benito

R, Brachat S, Campanaro S, Curtiss M, Davis K, Deutschbauer A, Entian KD, Flaherty P, Foury F, Garfinkel DJ, Gerstein M, Gotte D, Güldener U, Hegemann JH, Hempel S, Herman Z, Jaramillo DF, Kelly DE, Kelly SL, Kötter P, LaBonte D, Lamb DC, Lan N, Liang H, Liao H, Liu L, Luo C, Lussier M, Mao R, Menard P, Ooi SL, Revuelta JL, Roberts CJ, Rose M, Ross-Macdonald P, Scherens B, Schimmack G, Shafer B, Shoemaker DD, Sookhai-Mahadeo S, Storms RK, Strathern JN, Valle G, Voet M, Volckaert G, Wang CY, Ward TR, Wilhelmy J, Winzeler EA, Yang Y, Yen G, Youngman E, Yu K, Bussey H, Boeke JD, Snyder M, Philippsen P, Davis RW, Johnston M (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature,* 418, 387–391.

Håkansson KO, Winther JR (2007) Structure of glutaredoxin Grx1p C30S mutant from yeast. *Acta Crystallog Sect D: Biol Crystallogr.* 63, 288–294.

Hall A, Parsonage D, Poole LB, Karplus PA (2010) Structural evidence that peroxiredoxin catalytic power is based on transition-state stabilization. *J Mol Biol.* 402, 194–209.

Hayashi T, Ueno Y, Okamoto T (1993) Oxidoreductive regulation of nuclear factor kappa B. Involvement of a cellular reducing catalyst thioredoxin. *J Biol Chem.* 268, 11380–11388.

Hendson M, Purcell AH, Chen D, Smart C, Guilhabert M, Kirkpatrick B (2001) Genetic diversity of Pierce's disease strains and other pathotypes of *Xylella fastidiosa. Appl Environ Microbiol.* 67, 895–903.

Horta BB, de Oliveira MA, Discola KF, Cussiol JR, Netto LE (2010) Structural and biochemical characterization of peroxiredoxin Qbeta from *Xylella fastidiosa*: catalytic mechanism and high reactivity. *J Biol Chem.* 285, 16051–16065

Jacob C, Giles GI, Giles NM, Sies H (2003) Sulfur and selenium: the role of oxidation state in protein structure and function. *Angew Chem Int Ed Engl.* 42, 4742–4758.

Jones DP (2006) Redefining oxidative stress. *Antioxid Redox Signal.* 8, 1865–1879.

Kanda M, Ihara Y, Murata H, Urata Y, Kono T, Yodoi J, Seto S, Yano K, Kondo T (2006) Glutaredoxin modulates platelet-derived growth factor-dependent cell signaling by regulating the redox status of low molecular weight protein-tyrosine phosphatase. *J Biol Chem.* 281, 28518–28528.

Kohanski MA, DePristo MA, Collins JJ.(2010) Sublethal antibiotic treatment leads to multidrug resistance via radical-induced mutagenesis. *Mol Cell*. 12;311-320.

Koszelak-Rosenblum M, Krol AC, Simmons DM, Goulah CC, Wroblewski L, Malkowski MG (2008) *J Biol Chem*. 283, 24962–24971.

Kuge S, Jones N (1994) YAP1 dependent activation of TRX2 is essential for the response of *Saccharomyces cerevisiae* to oxidative stress by hydroperoxides. *EMBO J.* 13, 655–664.

Lambais MR, Goldman MH, Camargo LE, Goldman GH (2000) A genomic approach to the understanding of *Xylella fastidiosa* pathogenicity. *Curr Opin Microbiol*. 3, 459–462.

Lee J, Godon C, Lagniel G, Spector D, Garin J, Labarre J, Toledano MB (1999) Yap1 and Skn7 control two specialized oxidative stress response regulons in yeast. *J Biol Chem.* 274, 16040–16046.

Lesniak J, Barton WA, Nikolov DB (2002) Structural and functional characterization of the *Pseudomonas* hydroperoxide resistance protein Ohr. *EMBO J.* 24, 6649–6659.

Li WF, Yu J, Ma XX, Teng YB, Luo M, Tang YJ, Zhou CZ (2010) Structural basis for the different activities of yeast Grx1 and Grx2. *Biochim Biophys Acta*. 1804, 1542–1547.

Lindahl M, Mata-Cabana A, Kieselbach T (2011) The disulfide proteome and other reactive cysteine proteomes: analysis and functional significance. *Antioxid Redox Signal*. 12, 2581–2642.

Luikenhuis S, Perrone G, Dawes IW, Grant CM (1998) The yeast *Saccharomyces cerevisiae* contains two glutaredoxin genes that are required for protection against reactive oxygen species. *Mol Biol Cell*, 9, 1081–1091.

Marino SM, Gladyshev VN (2011) Redox biology: computational approaches to the investigation of functional cysteine residues. *Antioxid Redox Signal*. 15, 135–146.

Matthews JR, Wakasugi N, Virelizier JL, Yodoi J, Hay RT (1992) Thioredoxin regulates the DNA binding activity of NF-kappa B by reduction of a disulphide bond involving cysteine 62. *Nucleic Acids Res.* 20, 3821–3830.

Monteiro G, Horta BB, Pimenta DC, Augusto O, Netto LE (2007) Reduction of 1-Cys peroxiredoxins by ascorbate changes the thiol-specific antioxidant paradigm, revealing another function of vitamin C. *Proc Natl Acad Sci USA*. 104, 4886–4891.

Nagy P, Karton A, Betz A, Peskin AV, Pace P, O'Reilly RJ, Hampton MB, Radom L, Winterbourn CC (2011) Model for the exceptional reactivity of peroxiredoxins 2 and 3 with hydrogen peroxide: a kinetic and computational study. *J Biol Chem.* 286, 18048–18055.

Nelson KJ, Knutson ST, Soito L, Klomsiri C, Poole LB, Fetrow JS (2011) Analysis of the peroxiredoxin family: using active-site structure and sequence information for global classification and residue analysis. *Proteins*, 79, 947–964.

Netto LE, Oliveira MA, Monteiro G, Demasi AP, Cussiol JR, Discola KF, Demasi M, Silva GM, Alves SV, Faria VG, Horta BB (2007) Reactive cysteine in proteins: protein folding, antioxidant defense, redox signaling and more. *Comp Biochem Physiol C Toxicol Pharmacol*. 146, 180–193.

Ogusucu R, Rettori D, Munhoz DC, Netto LE, Augusto O (2007) Reactions of yeast thioredoxin peroxidases I and II with hydrogen peroxide and peroxynitrite: rate constants by competitive kinetics. *Free Radic Biol Med.* 42, 326–334.

Oliveira MA, Discola KF, Alves SV, Medrano FJ, Guimarães BG, Netto LE (2010) Insights into the specificity of thioredoxin reductase-thioredoxin interactions. A structural and functional investigation of the yeast thioredoxin system. *Biochemistry*, 49, 3317–3326.

Oliveira MA, Guimarães BG, Cussiol JR, Medrano FJ, Gozzo FC, Netto LE (2006) Structural insights into enzyme-substrate interaction and characterization of enzymatic intermediates of organic hydroperoxide resistance protein from *Xylella fastidiosa*. *J Mol Biol.* 359, 433–445.

O'Neill JS, Reddy AB (2011) Circadian clocks in human red blood cells. *Nature*, 469, 498–503.

Parsonage D, Youngblood DS, Sarma GN, Wood ZA, Karplus PA, Poole LB (2005) Analysis of the link between enzymatic activity and oligomeric state in AhpC, a bacterial peroxiredoxin. *Biochemistry*, 44, 10583–10592.

Peskin AV, Low FM, Paton LN, Maghzal GJ, Hampton MB, Winterbourn CC (2007) The high reactivity of peroxiredoxin 2 with H_2O_2 is not reflected in its reaction with other oxidants and thiol reagents. *J Biol Chem.* 282, 11885–11892.

Pinheiro AS, Amorim GC, Netto LE, Almeida FC, Valente AP (2008) NMR solution structure of the reduced form of thioredoxin 1 from *Sacharomyces cerevisiae*. *Proteins*, 70, 584–587.

Powis G, Montfort WR (2001) Properties and biological activities of thioredoxins. *Annu Rev Biophys Biomol Struct*. 30, 421–455.

Rhee SG, Woo HA (2011) Multiple functions of peroxiredoxins: peroxidases, sensors and regulators of the intracellular messenger H_2O_2, and protein chaperones. *Antioxid Redox Signal*. 15, 781–794.

Saitoh M, Nishitoh H, Fujii M, Takeda K, Tobiume K, Sawada Y, Kawabata M, Miyazono K, Ichijo H (1998) Mammalian thioredoxin is a direct inhibitor of apoptosis signal-regulating kinase (ASK) 1. *EMBO J*. 17, 2596–2606.

Silva GM, Netto LE, Discola KF, Piassa-Filho GM, Pimenta DC, Bárcena JA, Demasi M (2008) Role of glutaredoxin 2 and cytosolic thioredoxins in cysteinyl-based redox modification of the 20S proteasome. *FEBS J*. 275, 2942–2955.

Simpson AJ et al (2000) The genome sequence of the plant pathogen *Xylella fastidiosa*. The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis. *Nature*, 406, 151-159.

Smolka MB, Martins-de-Souza D, Martins D, Winck FV, Santoro CE, Castellari, RR, Ferrari F, Brum IJ, Galembeck E, Della Coletta Filho H, Machado MA, Marangoni S, Novello JC (2003) Proteome analysis of the plant pathogen *Xylella fastidiosa* reveals major cellular and extracellular proteins and a peculiar codon bias distribution. *Proteomics*, 3, 224–237.

Srinivasan U, Mieyal PA, Mieyal, JJ (1997) pH profiles indicative of rate-limiting nucleophilic displacement in thioltransferase catalysis. *Biochemistry*, 36, 3199–3206.

Tenhaken R, Levine A, Brisson LF, Dixon RA, Lamb C (1995) Function of the oxidative burst in hypersensitive disease resistance. *Proc Natl Acad Sci USA*. 92, 4158–4163.

Toledo JC Jr, Audi R, Ogusucu R, Monteiro G, Netto LE, Augusto O (2011) Horseradish peroxidase compound I as a tool to investigate reactive protein-cysteine residues: from quantification to kinetics. *Free Radic Biol Med*. 50, 1032–1038.

Trujillo M, Ferrer-Sueta G, Thomson L, Flohé L, Radi R (2007) Kinetics of peroxiredoxins and their role in the decomposition of peroxynitrite. *Subcell Biochem*. 44, 83–113.

Waksman G, Krishna TSR, Williams CH Jr, Kuriyan J (1994) Crystal structure of *Escherichia coli* thioredoxin reductase refined at 2 Å resolution. Implications for a large conformational change during catalysis. *J Mol Biol*. 236, 800–816.

Wang J, Boja ES, Tan W, Tekle E, Fales HM, English S, Mieyal JJ, Chock PB (2001) Reversible glutathionylation regulates actin polymerization in A431 cells. *J Biol Chem*. 276, 47763–47766.

Wang J, Tekle E, Oubrahim H, Mieyal JJ, Stadtman ER, Chock PB (2003) Stable and controllable RNA interference: investigating the physiological function of glutathionylated actin. *Proc Natl Acad Sci USA*. 100, 5103–5106.

Williams CH Jr, Arscott LD, Muller S, Lennon BW, Ludwig ML, Wang PF, Veine DM, Becker K, Schirmer RH (2000) Thioredoxin reductase: Two modes of catalysis have evolved. *Eur J Biochem*. 267, 6110–6117.

Winterbourn CC, Metodiewa D (1999) Reactivity of biologically important thiol compounds with superoxide and hydrogen peroxide. *Free Radic Biol Med*. 27, 322–328.

Winterbourn CC, Hampton MB (2008) Thiol chemistry and specificity in redox signaling. *Free Radic Biol Med*. 45, 549–561.

Wood ZA, Schröder E, Robin Harris J, Poole LB (2003) Structure, mechanism and regulation of peroxiredoxins. *Trends Biochem Sci*. 1, 32–40.

Wrzaczek M, Brosché M, Kollist H, Kangasjärvi J (2009) Arabidopsis GRI is involved in the regulation of cell death induced by extracellular ROS. *Proc Natl Acad Sci USA*. 106, 5412–5417.

Protein Engineering in Structure-Function Studies of Viper's Venom Secreted Phospholipases A2

Toni Petan[1], Petra Prijatelj Žnidaršič[2] and Jože Pungerčar[1]
[1]Department of Molecular and Biomedical Sciences, Jožef Stefan Institute, Ljubljana,
[2]Department of Chemistry and Biochemistry, Faculty of Chemistry and Chemical
Technology, University of Ljubljana, Ljubljana,
Slovenia

1. Introduction

Secreted phospholipases A_2 (sPLA2s) constitute a large family of interfacial enzymes that hydrolyze the *sn*-2 ester bond of membrane glycerophospholipids, releasing free fatty acids and lysophospholipids (Murakami et al., 2011). They are abundant in snake venoms, frequently being their major toxic components, and display a variety of pharmacological effects, such as neurotoxicity, myotoxicity, anticoagulant activity, cardiotoxicity and haemolytic activity. The molecular mechanisms underlying these effects are still poorly understood; but they are most probably based on the existence of specific, high-affinity binding sites for toxic sPLA2s on the surface of target cells in specific target tissues (Kini, 2003). Snakes have developed, through evolution, an arsenal of structurally similar molecules that target a specific tissue or function, in order to capture and digest their prey. Interestingly, very often a particular sPLA2 molecule may display, despite its simple, globular and compact structure, several different toxic activities. Thus, it significantly expands the possible prey-damaging mechanisms of the venom, which also depend on the type of prey, site of injection of the toxin and the tissue involved. The remarkable variety of pharmacological effects exerted by sPLA2 toxins is a consequence of several factors that have intrigued scientists working in the field, but have also greatly complicated the study of their actions. The complications include (1) the apparently indiscriminate enzymatic activities of sPLA2 toxins on different cellular and non-cellular phospholipid membranes and aggregates; (2) the diverse effects of the products of the hydrolysis of sPLA2 toxins, especially on membrane structural integrity and dynamics, thus affecting the major structural and functional features of the cell; and (3) the ever increasing diversity of sPLA2 intra- and extracellular binding proteins discovered in mammalian tissues, which are involved in very different biological processes. Additionally, many of the actions of sPLA2 toxins involve complex, multi-step molecular mechanisms, in which a specific combination of enzymatic activity and/or protein binding is probably essential for a particular step. Although sPLA2s are structurally highly conserved proteins, it is clear that subtle evolutionary changes of residues on the surface of the molecule have empowered these enzymes with this wide range of toxic activities. Their ability to recognize specific molecular

targets has been gradually optimized and thus interferes with a range of physiological processes (Kini & Chan, 1999). Snakes have even developed, through their evolution, catalytically inactive sPLA2-homologues specialized in membrane damage that occurs independently of enzymatic activity (Lomonte et al., 2009). Interestingly, a range of structurally very similar sPLA2 enzymes, as well as an enzymatically inactive sPLA2 homologue, are also present in mammals. The mammalian sPLA2 family consists of 10 or 11 enzymes (Lambeau & Gelb, 2008) that display different cell- and tissue-specific expression patterns. The proteins act with a broad range of enzymatic activities on a variety of cellular and non-cellular phospholipid membranes (Murakami et al., 2011). They bind with high affinity to various soluble and membrane protein targets, many of which were discovered using toxic sPLA2s (Pungerčar & Križaj, 2007; Valentin & Lambeau, 2000). Furthermore, apart from their direct effects on membrane structure and function, the products of their catalysis are precursors of hundreds of bioactive lipid signalling molecules, such as the eicosanoids. The mammalian sPLA2 enzymes display a similarly broad range of roles, mostly incompletely understood and often contradictory, in various physiological and pathophysiological processes, such as lipid digestion and homeostasis, innate immunity, inflammation, fertility, blood coagulation, asthma, atherosclerosis, autoimmune diseases and cancer (Lambeau & Gelb, 2008; Murakami et al., 2011). In this they are analogous to their venom counterparts, owing their functions to a combination of enzymatic activity, direct and indirect effects of the products of their hydrolysis, and specific interactions with molecular partners inside or outside the cell. The research on the action of exogenous snake venom sPLA2 enzymes, which target particular physiological processes in their mammalian prey, has been providing important clues for deciphering the biological roles of the mammalian endogenous sPLA2 enzymes as well (Rouault et al., 2006; Valentin & Lambeau, 2000).

The most potent sPLA2 toxins display presynaptic (ß-)neurotoxicity by attacking the presynaptic site of neuromuscular junctions. The venom of the nose-horned viper, *Vipera ammodytes ammodytes*, contains three presynaptically neurotoxic sPLA2s, ammodytoxins (Atxs) A, B and C, two non-toxic ammodytins (Atns), $AtnI_1$ and $AtnI_2$, and a myotoxic and catalytically inactive Ser49 sPLA2 homologue, ammodytin L (AtnL). They are all group IIA sPLA2s. The presynaptically acting (ß-neurotoxic) Atxs interfere specifically with the release of acetylcholine from motoneurons and cause irreversible blockade of neuromuscular transmission. The exact mechanism of their action is not yet fully understood, but it must include specific binding to receptor(s) on the presynaptic membrane and enzymatic activity (Montecucco et al., 2008; Pungerčar & Križaj, 2007). The binding to highly specific, and yet unknown, primary molecular targets of ß-neurotoxins (ß-ntxs) on the presynaptic membrane is most probably followed by entry of the toxin into the nerve cell (Logonder et al., 2009; Pražnikar et al., 2008; Rigoni et al., 2008). It has been proposed that different sPLA2-toxins exploit different internalization routes (Pungerčar & Križaj, 2007). In the motoneuron, they may impair the cycling of synaptic vesicles by phospholipid hydrolysis and by binding to specific intracellular protein targets, like calmodulin (Kovačič et al., 2009, 2010; Šribar et al., 2001) and 14-3-3 proteins (Šribar et al., 2003b) in the cytosol, and R25 (Šribar et al., 2003a) in mitochondria. Although the role of enzymatic activity in ß-neurotoxicity of sPLA2s is still somewhat controversial, accumulated results speak largely in favour of its being indispensable for full expression of the ß-neurotoxic effect (Montecucco et al., 2008; Pungerčar & Križaj, 2007; Rouault et al., 2006).

In spite of numerous attempts to identify the surface residues of sPLA2s crucial for a particular pharmacological effect (i.e., the "ß-neurotoxic site" or the "anticoagulant site") – based initially on structural analysis, chemical modification and, later on, site-directed mutagenesis, the molecular basis of their toxicity has yet to be resolved (Kini, 2003; Križaj, 2011; Pungerčar & Križaj, 2007; Rouault et al., 2006). We have addressed this issue in studies based on protein engineering of the nosed-horned viper sPLA2s. These have resulted in more than fifty mutants and chimeric sPLA2 proteins that have been produced and characterized in terms of their biochemical and biological activities (most of them are shown in Table 1). The site-directed mutagenesis studies have provided answers to, or at least significantly improved, our knowledge concerning many important questions regarding the toxic and enzymatic activities of Atxs and other sPLA2s, such as:

- Which regions of the molecule are important for the ß-neurotoxicity of Atxs and homologous toxic sPLA2s?
- Which surface residues of Atxs are crucial for interaction with their high-affinity binding proteins? Is there a correlation between the affinity for binding to a particular protein target and the neurotoxic potency of Atxs, suggesting a role for the receptor protein in the process?
- Which residues are responsible for interfacial membrane binding of Atxs? How do the interfacial membrane and kinetic properties of Atxs influence their toxic action? How do the enzymatic properties of Atxs compare with those of the best-studied mammalian sPLA2s?
- What is the importance of the loss in evolution of enzymatic activity in the case of the AtnL and other myotoxic sPLA2-homologues?

The results obtained have contributed significantly to a better understanding of the molecular mechanisms of action of snake venom sPLA2s and provided clues to the action of the homologous groups of mammalian sPLA2s.

2. Search for the "neurotoxic site" and the role of enzymatic activity

The significant structural similarities of toxic and non-toxic sPLA2s, which differ in their pharmacological actions, have enticed a large number of researchers hoping to find the "holy grail" of sPLA2-toxin research – the toxic site. The site is presumed to comprise only a small number of crucial amino acid residues (Kini, 2003). To explain the wide range of pharmacological effects induced by snake venom sPLA2s, Kini & Evans (1989) proposed a model comprising specific "target sites" present on the surface of particular cell types. The target sites are proposed to be recognized by complementary "pharmacological sites" on the toxin molecule, these being structurally distinct from, and independent of, the "catalytic site." Thus, high-affinity binding (at least in the nM range) of the toxin to specific target sites ensures that, upon entering the circulation, each toxin binds primarily to its proper target tissue. It is highly likely that the primary target, or acceptor, sites are proteins. This is because of the much lower affinity (mM–μM range) of sPLA2s for binding to the abundant zwitterionic phospholipid surfaces (i.e., cell membranes) (Bezzine et al., 2002; Petan et al., 2005; Singer et al., 2002) and, following enzyme adsorption to the membrane surface, indiscriminate binding to and hydrolysis of phospholipid molecules at the catalytic site. Therefore, separate pharmacological sites on an sPLA2 molecule recognizing different target binding sites should be the main structural determinants that differentiate their respective

pharmacological actions, such as presynaptic or central neurotoxicity. However, according to the results of our mutagenesis structure-function studies of Atxs reviewed below, it is unlikely that there is a structurally distinct, single "presynaptic neurotoxic site" located in a specific part of the molecule, in contrast to the strict physical localization of the enzyme active site. Rather, different parts of the toxin molecule are likely to be involved in different stages of the complex multi-step process of neurotoxicity, all contributing to the final outcome. In this view, structurally different ß-ntxs may have different surface regions that bind to different (extra- and intracellular) targets, which are nevertheless involved in the same process, most probably the recycling of synaptic vesicles. However, they all share the nonspecific $sPLA_2$ activity, i.e., the ability to bind and hydrolyze different phospholipid molecules embedded in membranes of various compositions – an essential step in the complete, irreversible blockade of neuromuscular transmission. Therefore, at least in the case of ß-neurotoxic $sPLA_2$s, the use of the term "presynaptic neurotoxic site" appears unsuitable for describing the multiple regions distributed on the surface of the $sPLA_2$ molecule (Prijatelj et al., 2008).

Given the multi-step, and as yet incompletely known, molecular events leading to presynaptic neurotoxicity of $sPLA_2$s, a simple correlation between their *in vitro* enzymatic activity and their lethal potency would not be expected (Rosenberg, 1997). Indeed, there are numerous examples of $sPLA_2$ ß-ntxs with significantly different enzymatic properties, which are, however, not reflected in differences in toxicity; in fact even the most potent $sPLA_2$ ß-ntxs are weak enzymes (Petan et al., 2005; Pražnikar et al., 2008; Prijatelj et al., 2006b, 2008; Rosenberg, 1997). Nevertheless, $sPLA_2$ enzymatic activity is necessary for full expression of the ß-neurotoxic effect (Montecucco et al., 2008; Pungerčar & Križaj, 2007). Its role in the process is most probably obscured by the numerous factors affecting both $sPLA_2$ activity and neurotoxicity, especially the, as yet unknown, (sub)cellular location, accessibility, composition and physical properties of the target membrane. The enzymatic action of $sPLA_2$s could lead to structural and functional destruction of cell membranes and organelles, like mitochondria or synaptic vesicles (Pražnikar et al., 2008, 2009; Pungerčar & Križaj, 2007; Rigoni et al., 2008), since the products of phospholipid hydrolysis are disruptive to many physiological processes by impairing the function of peripheral and integral membrane proteins and promoting membrane dysfunction by altering membrane asymmetry, curvature and fusogenicity (Montecucco et al., 2008; Paoli et al., 2009; Rigoni et al., 2005). The apparent lack of correlation between *in vitro* enzymatic activity and lethal potency of Atxs or other neurotoxic $sPLA_2$s (Montecucco et al., 2008; Petan et al., 2005; Pungerčar & Križaj, 2007) can be explained by the strict localization of the $sPLA_2$ activity to particular target membrane(s) due to binding to highly specific extra- and intracellular protein acceptors (Paoli et al., 2009; Petan et al., 2005; Pungerčar & Križaj, 2007). Our studies investigating the interfacial binding and kinetic properties of toxic $sPLA_2$s and their mutants have provided important clues to understanding the role of enzymatic activity in the process of presynaptic neurotoxicity of $sPLA_2$s. As described in detail below, despite their potent neurotoxic activity, Atxs are quite effective in hydrolysing pure phosphatidylcholine (PC) vesicles as well as PC-rich plasma membranes of mammalian cells, similarly to the most active mammalian group V and X $sPLA_2$ enzymes (Petan et al., 2005, 2007; Pražnikar et al., 2008). We have also shown that, when tightly bound to the membrane surface, the Ca^{2+} requirements of Atxs are in the micromolar range (Petan et al., 2005), opening up the possibility that such neurotoxins are also catalytically active in the subcellular

compartments where Ca^{2+} concentrations are low (Kovačič et al., 2009; Petan et al., 2005). Moreover, Atxs are rapidly internalized in motoneuronal cells and are, surprisingly, translocated to the cytosol, where they specifically bind calmodulin (CaM) and 14-3-3 proteins, strongly opposing the dogma of the exclusively extracellular action of not only sPLA2-neurotoxins, but also of sPLA2s in general (Pražnikar et al., 2008). In agreement with these findings, we have recently shown that high-affinity binding to the cytosolic Ca^{2+}-sensor molecule CaM leads to structural stabilization (increased resistance to the reducing environment of the cytosol) and a significant augmentation of the enzymatic activity of Atxs and, intriguingly, also of the mammalian group V and X sPLA2s (Kovačič et al., 2009, 2010). These findings strongly support the possibility of augmentation of Atx enzymatic activity by CaM in the cytosol during the process of ß-neurotoxicity. They also point to a new mechanism of modulating the enzymatic activity of mammalian group V and X sPLA2s or some other non-toxic endogenous sPLA2 (Kovačič et al., 2010).

3. Structural determinants of presynaptic neurotoxicity of sPLA2s

The subtlety of the structure-function relationship of sPLA2 neurotoxins is obvious on examination of the primary structures and toxicities of Atxs. The three sPLA2 toxins each consist of 122 amino acid residues and differ at only five positions (Križaj, 2011). AtxC may be considered as a natural double mutant (F124I/K128E) and AtxB as a triple mutant (Y115H/R118M/N119Y) of AtxA. Nevertheless, their lethal potencies in mice differ considerably. AtxA is the most lethal; and its protein isoforms, AtxC and AtxB, are 17- and 28-fold less potent, respectively (Thouin et al., 1982). The crystal structures of recombinant AtxA (PDB code 3G8G) and natural AtxC (PDB code 3G8H) demonstrate the absence of significant structural differences between the two toxins (Saul et al., 2010). There is only a minor conformational difference at positions 127 and 128 in the C-terminal region, caused by the charge-reversal substitution of Lys128 for Glu, which does not significantly influence the toxicity (Saul et al., 2010). An illustrative example of the subtle structure-function relationships of sPLA2s is the conversion, by a single mutation (F22Y), of the gene for bovine pancreatic group IB sPLA2 to a gene encoding a molecule able to compete with crotoxin, a ß-neurotoxic sPLA2 from the South American rattlesnake, *Crotalus durissus terrificus*, for binding to its 45-kDa neuronal-binding protein. This led the authors to suggest the conversion of the non-toxic pancreatic sPLA2 to a neurotoxic molecule (Tzeng et al., 1995).

By substituting several basic residues in the C-terminal region (AtxA[NNTETE] mutant: AtxA-K108N/K111N/K127T/K128E/E129T/K132E) and in the ß-structure region (AtxA[SSL] mutant: AtxA-K74S/H76S/R77L) with acidic and non-ionic residues, we have shown, contrary to previous beliefs, that the basic character of Atxs, and probably of other ß-neurotoxic sPLA2s, is not obligatory for presynaptic toxicity (Ivanovski et al., 2004; Prijatelj et al., 2000; Table 1). According to our earlier structure-function analyses, the more than one order of magnitude lower toxicity of AtxC than that of AtxA is a consequence of the substitution of the aromatic Phe124 by Ile (Pungerčar et al., 1999). Furthermore, in accordance with the three substitutions responsible for the difference in toxicities of AtxA and AtxB, several other C-terminal residues of AtxA, namely the Tyr115/Ile116/Arg118/Asn119 (YIRN) cluster, were shown to be important for the neurotoxicity of Atxs (Ivanovski et al., 2000). Thus, the lethal potency of the AtxA-Y115K/I116K/R118M/N119L (AtxA[KKML]) mutant was 290-fold lower than that of AtxA (Ivanovski et al., 2000 and Table 1).

$sPLA_2$	LD_{50} (µg/kg)	IC_{50} (nM)			References
		CaM	R25	R180	
AtxA	21	6 ± 2	10 ± 3	16 ± 3	a, b
AtxB	580	23 ± 4	n. d.	n. d.	a, c
AtxC	360	21 ± 3	50	155	a, c, d, e
12-AtxA	280	72 ± 15	5 ± 2	$> 10^4$	f
I-AtxA	500	250 ± 60	3.4 ± 0.3	$> 10^4$	f
P-AtxA	420	380 ± 85	3.5 ± 0.5	$> 10^4$	f
$AtnI_2$/$AtxA^{K108N}$	$> 10^4$	1300 ± 200	20 ± 6	490 ± 100	g
$AtnI_2^{N24F}$/$AtxA^{K108N}$	> 5000	1700 ± 300	24 ± 6	850 ± 200	g
$AtnI_2$	$> 10^4$	$> 10^4$	$> 10^4$	610 ± 100	g
$AtxA^{NNTETE}$	660	27 ± 5	16	100	c, d
$AtxA^{K108N/K111N}$	67	17 ± 4	38	68	c, d, h
$AtxA^{K127T}$	35	20 ± 3	22	300	c, d
$AtxA^{K128E}$	45	14 ± 3	n. d.	n. d.	c, h
$AtxA^{F24A}$	90	7.0 ± 0.9	15 ± 1	17 ± 2	b
$AtxA^{F24N}$	2800	5.6 ± 0.7	14 ± 2	26 ± 1	b
$AtxA^{F24S}$	380	7.4 ± 0.2	14 ± 1	16 ± 2	b
$AtxA^{F24W}$	175	13.6 ± 0.3	37 ± 5	26 ± 3	b
$AtxA^{F24Y}$	330	9.2 ± 0.9	14 ± 3	19 ± 4	b
$AtxA^{V31W}$	135	n. d.	n. d.	n. d.	i
$AtxA^{R72E}$	84	71 ± 3	14 ± 1	78 ± 7	j
$AtxA^{R72I}$	32	17 ± 2	18.0 ± 0.4	28 ± 3	j
$AtxA^{R72K}$	50	46 ± 2	11.7 ± 0.4	83 ± 9	j
$AtxA^{R72S}$	55	24 ± 2	16 ± 1	35 ± 1	j
$AtxA^{SSL}$	276	18 ± 1	18 ± 1	107 ± 8	j
$AtxA^{K86A}$	24	8.1 ± 0.3	15 ± 1	20 ± 3	j
$AtxA^{K86E}$	32	7 ± 1	18 ± 1	42 ± 4	j
$AtxA^{K86G}$	34	7.5 ± 0.2	12.1 ± 0.1	65 ± 4	j
$AtxA^{K86R}$	31	8.8 ± 0.5	16.3 ± 0.5	36 ± 1	j
$AtxA^{KKML}$	~ 6000	50 ± 9	86	257	c, k
$AtxA^{KK}$	~ 5000	21 ± 3	380	118	c, k
$AtxA$/$DPLA_2^{YIRN}$	45	27 ± 5	180 ± 25	100 ± 16	l
$AtxA^{KEW}$/$DPLA_2^{YIRN}$	910	43 ± 12	110 ± 15	22 ± 4	l
$AtxA^{KE}$/$DPLA_2^{YIRN}$	790	14 ± 4	87 ± 9	24 ± 4	l
$AtxA^{W}$/$DPLA_2^{YIRN}$	107	28 ± 7	78 ± 12	19 ± 5	l
$AtxA$/$DPLA_2$	2600	110 ± 10	45%#	280 ± 19	l
$DPLA_2^{YIRN}$	~ 17000	43 ± 14	200 ± 30	120 ± 21	c, l
$DPLA_2$	3100	300 ± 36	75%#	300 ± 45	c, l
AtnL	> 10000	n. d.	n. d.	n. d.	m
$AtnL^{YVGD}$	> 7000	n. d.	n. d.	n. d.	m
$AtnL^{YWGD}$	2200	n. d.	n. d.	n. d.	m

Table 1. Lethal potency and protein-binding affinity of Atxs, Atns, $DPLA_2$ and their mutants. IC_{50} values (the concentration of competitor $sPLA_2$ required to reduce the binding of 10 nM ^{125}I-AtxC by 50%) were determined from competition binding experiments for binding to calmodulin (CaM), the mitochondrial receptor R25 and the neuronal M-type $sPLA_2$ receptor, R180. #The recombinant toxin did not completely inhibit the binding of ^{125}I-AtxC. [a]Thouin et al., 1982; [b]Petan et al., 2002; [c]Prijatelj et al., 2003; [d]Prijatelj et al., 2000; [e]Čopič et al., 1999; [f]Prijatelj et al., 2006b; [g]Prijatelj et al., 2002; [h]Pungerčar et al., 1999; [i]Petan et al., 2005; [j]Ivanovski et al., 2004; [k]Ivanovski et al., 2000; [l]Prijatelj et al., 2008; [m]Petan et al., 2007.

The KKML cluster is present in the weakly neurotoxic DPLA$_2$ from the venom of Russell's viper, *Daboia* (*Vipera*) *russelii russelii*, which shares a high level of amino acid identity (82%) with AtxA. However, the latter is almost 150-fold more toxic in mice (Prijatelj et al., 2003). To our great surprise, the introduction of the YIRN cluster into DPLA$_2$ did not increase its toxicity; on the contrary, the DPLA$_2$YIRN mutant was more than five times less toxic than DPLA$_2$ (Prijatelj et al., 2003). Additionally, our study on the importance of the N-terminal residue Phe24, in which it was replaced by other aromatic (tyr or trp), polar uncharged (ser or asn) or hydrophobic (ala) residues, suggested that Phe24 is also involved in the neurotoxicity of Atxs, apparently at a stage not involving enzymatic activity or interactions with the high-affinity binding proteins R25, R180 and CaM (Petan et al., 2002). The aromatic Phe24 was chosen for this study on the basis of several interesting characteristics. It is located in a region immediately preceding the Ca^{2+}-binding loop, but it is spatially close to the important Phe124. It is important for membrane binding as part of the interfacial binding surface (IBS, see below) of the enzyme, and it is replaced by ser in the weakly neurotoxic DPLA$_2$. These facts prompted us to propose that a particular combination of both C-terminal and N-terminal residues must be involved in ß-neurotoxicity. In order to identify the N-terminal residues that supplement the role of the YIRN cluster in the high neurotoxic potency of AtxA, we selectively mutated some of the remaining residues that differentiate DPLA$_2$ from AtxA. First, we introduced the N-terminal half of AtxA into DPLA$_2$ by preparing the chimeric AtxA/DPLA$_2$ protein. Its lethal potency was relatively low, in the range of those of DPLA$_2$ and the AtxA^{KKML} mutant (Table 1 and Ivanovski et al., 2000), confirming that it is primarily the presence of the KKML cluster in the C-terminus of the chimera that has a strong negative influence on toxicity. Secondly, by substituting the KKML cluster in DPLA$_2$ with the YIRN cluster of AtxA, we produced a chimeric mutant (AtxA/DPLA$_2$YIRN) that is 58-fold higher in lethal potency than is AtxA/DPLA$_2$, reaching a level of toxicity similar to that of the highly neurotoxic AtxA. Thus, only in combination with the N-terminal part of AtxA is the presence of the YIRN cluster sufficient for the high neurotoxic potency of AtxA and AtxA/DPLA$_2$YIRN. This allowed us to exclude the importance of the additional eleven C-terminal residues present in AtxA and absent in DPLA$_2$ – Thr70, His76, Glu78, Gly85, Arg100, Asn114, Ser130, Glu131 – and was in accordance with the findings of our early mutagenesis study on the three C-terminal lysines at positions 108, 111 and 128 (Pungerčar et al., 1999) (Figure 1). These results clearly confirmed our hypothesis that a particular combination of C-terminal residues, especially those in the region 115-119, i.e., the YIRN cluster, and certain N-terminal residues are necessary for the potent ß-neurotoxicity of Atxs.

Our next objective was to determine the contribution to the high neurotoxic potency of AtxA of the remaining nine N-terminal residues that differentiate it from DPLA$_2$ (Met7, Gly11, Asn17, Pro18, Leu19, Thr20, Phe24, Val31 and Ser67). As described above, the importance of the aromatic Phe24 had already been established (Petan et al., 2002). In order to assess the importance of the remaining N-terminal residues, we first substituted Met7, Gly11 and Val31 in the highly toxic chimera AtxA/DPLA$_2$YIRN by the corresponding residues present in DPLA$_2$ – lys, glu and trp, respectively. The mutant protein AtxA^{KEW}/DPLA$_2$YIRN displayed a 20-fold lower lethal potency than the AtxA/DPLA$_2$YIRN chimera, suggesting involvement of the group of residues at positions 7, 11 and 31 in AtxA neurotoxicity. The lethality of the partial mutants, AtxA^{KE}/DPLA$_2$YIRN and AtxA^{W}/DPLA$_2$YIRN (Table 1), revealed that the contribution of the pair of residues in the N-terminal helix, Met7 and Gly11, to the

```
                     1        10        20        30        40        50        60

AtxA                 SLLEFGMMILG-ETGKNPLTSYSFYGCYCGVGGKGTPKDATDRCCFVHDCCYGNLP--D-C-----S
AtxB                 ..........-..............................................--.-.-----.
AtxC                 ..........-..............................................--.-.-----.
AtxA^KKML            ..........-..............................................--.-.-----.
AtxA/DPLA2^YIRN      ..........-..............................................--.-.-----.
AtxA^KEW/DPLA2^YIRN  ......K...E-.....................W........................--.-.-----.
AtxA/DPLA2           ..........-..............................................--.-.-----.
DPLA2^YIRN           ......K...E-....LAIP...S......W...........................--.-.-----N
DPLA2                ......K...E-....LAIP...S......W...........................--.-.-----N
AtnL                 .VI...K..QE-..D.............H..L.N..K.............S...AK.S--.-.-----.
AtnL^YVGD            .VI...K..QE-..D...............V....K.............AK.S--.-.-----.
AtnL^YWGD            .VI...K..QE-..D...............W....K.............AK.S--.-.-----.
AtnI2                N.YQ..N..FK-M.K.SA.L...N......W....K.Q..............RVN--G-.-----D
hGIIA               N.VN.HR..KL-T...EAAL..G....H.....R.S.........VT.....KR.EKRG-.-----G
hGV                 G..DLKS..EK-V....A..N.G.......W..R.....G..W..WA..H...R.EEKG-.-----N
hGX                 GI..LAGTVGC-V-.PRTPIA.MK...F..L..H.Q.R..I.W..HG.....TRAEEAG-.-----.
```

```
                     70        80        90        100       110       120       130

AtxA                 PKTDRYKYHRENGAIVCGK-GTSCENRICECDRAAAICFRKNLKTYNYIYRNYPD-FLCKKESEKC
AtxB                 .................-.............................H..MY...-..........
AtxC                 .................-.............................-I...E.....
AtxA^KKML            .................-.............................KK.ML...-..........
AtxA/DPLA2^YIRN      ..S.....K.V......E.-............K.......Q..N..S.........-....G.-L..
AtxA^KEW/DPLA2^YIRN  ..S.....K.V......E.-............K.......Q..N..S.........-....G.-L..
AtxA/DPLA2           ..S.....K.V......E.-............K.......Q..N..SKK.ML...-....G.-L..
DPLA2^YIRN           ..S.....K.V......E.-............K.......Q..N..S.........-....G.-L..
DPLA2                ..S.....K.V......E.-............K.......Q..N..SKK.ML...-....G.-L..
AtnL                 ...N..E...........S-S.P.KKQ.............E......KK.KV.LR-.K..GV....
AtnL^YVGD            ...N..E...........S-S.P.KKQ.............E......KK.KV.LR-.K..GV....
AtnL^YWGD            ...N..E...........S-S.P.KKQ.............E......KK.KV.LR-.K..GV....
AtnI2               ..LSI.S.SF...D....G-DDP.LRAV.....V.....GE..N..DKK.K...S-SH.TET-.Q.
hGIIA               T.FLS..FSNSGSR.T.A.-QD..RSQL....K...T..AR.KT...KK.QY.SN-KH.RGSTPR.
hGV                 IR.QS...RFAW.VVT.EP-.PF.HVNL.A...KLVY.LKR..RS..PQ.QYF.N-I..S
hGX                 ...E..SWQCV.QSVL..PAENK.QELL.K..QEI.N.LAQT--E..LK.LF..Q-...EPD.P..D
```

Fig. 1. Amino acid alignment of snake venom group IIA sPLA$_2$s, including the sPLA$_2$-homologue ammodytin L (AtnL), some of their mutants and the human group IIA, V and X sPLA$_2$s. The residues comprising the putative IBS of Atxs are presented in bold type, while those most important for the neurotoxicity of ammodytoxins (Atxs) are underlined. The weakly neurotoxic sPLA$_2$ from Russell's viper, *Daboia r. russelii*, DPLA$_2$, differs from AtxA in only 22 residues (82% identity). AtnL, the enzymatically inactive but myotoxic Ser49

structural homologue of Atxs, displays 74% amino acid identity with AtxA. The neutral ammodytin I$_2$ (AtnI$_2$) is a non-toxic homologue of Atxs from the same venom with 58% amino acid identity with AtxA. Atxs also display a relatively high degree of identity (48%) with the human groups IIA (hGIIA), V (hGV) sPLA$_2$s, and X (hGX) sPLA$_2$ (41%). The common sPLA$_2$ numbering of residues was used (Renetseder et al., 1985). Gaps, represented by dashes, were used to align the homologous sPLA$_2$s according to conserved residues. Identical amino acid residues are shown by dots. Amino acid single-letter symbols are shown in Table 1 in the chapter by Figurski et al.

neurotoxic potency of AtxA is substantially higher than that of Val31 – in accordance with the negligible effect of the V31W mutation on the lethality of AtxA, despite its outstanding effect on enzymatic activity (Petan et al., 2005). Interestingly, the bulky Trp at position 31 of the AtxA^{V31W} and AtxA^W/DPLA$_2$^{YIRN} mutants had no significant impact on neurotoxicity, despite being spatially very close to Phe24 and also to the YIRN cluster. The collective contribution of the N-terminal residues Met7, Gly11 and Phe24 to the neurotoxicity of AtxA is very significant and similar to that of the YIRN cluster in the C-terminus. For example, the substitution of Phe24 with ser in AtxA caused an approximately 19-fold decrease in neurotoxicity (Petan et al., 2002), which is similar to the reduction seen after adding the N-terminal region of AtxA^{KEW}, in which the residues Met7, Gly11 and Val31 were substituted with lys, glu and trp, respectively, to DPLA$_2$^{YIRN}, resulting in the AtxA^{KEW}/DPLA$_2$^{YIRN} chimeric mutant. Therefore, it is highly likely that the remaining N-terminal residues differentiating AtxA and DPLA$_2$, i.e., the Asn17/Pro18/Leu19/Thr20 cluster and Ser67, are not greatly involved in neurotoxicity.

Our structure-function studies of ß-neurotoxic sPLA$_2$s (Ivanovski et al., 2000, 2004; Petan et al., 2002, 2005; Prijatelj et al., 2000, 2002, 2003, 2006b, 2008; Pungerčar et al., 1999) clearly show that different parts of the toxin molecule have separate roles in the distinct steps of the complex mechanism of presynaptic neurotoxicity (Pungerčar & Križaj, 2007). Most significantly, by selectively mutating parts of the DPLA$_2$ molecule, we were able to map the residues that separate the weakly ß-neurotoxic sPLA$_2$ from the 150-fold more potent AtxA (Prijatelj et al., 2008). In summary, the most important structural features responsible for the high neurotoxic potency of Atxs (Figure 2, A) are: the "upper" part of the molecule concentrated around the C-terminal region 115-119 (the YIRN cluster) and including the spatially close aromatic Phe124 and Phe24, and the N-terminal helix region with the Met7/Gly11 pair in the "lower right" part of the molecule.

4. Enzymatic activity of Atxs: factors influencing interfacial binding and hydrolysis of phospholipid membranes

Secreted PLA$_2$s are prototypical interfacial enzymes (Berg et al., 2001). In order to gain access to their phospholipid substrate, sPLA$_2$s first have to bind at the lipid/water membrane interface by their interfacial binding surface (IBS), a group of residues located on a relatively flat exposed region surrounding the entrance to the active site pocket (Lin et al., 1998; Ramirez & Jain, 1991). Only then, after the enzyme is bound to the surface, can binding of a single phospholipid molecule and its hydrolysis occur in the active site. The catalytic turnover cycle of sPLA$_2$s includes a highly conserved His48–Asp99 dyad and an activated water molecule that acts as a nucleophile during hydrolysis of the *sn*-2 ester bond of the phospholipid (Scott et al., 1990). The resulting tetrahedral intermediate is stabilized by the
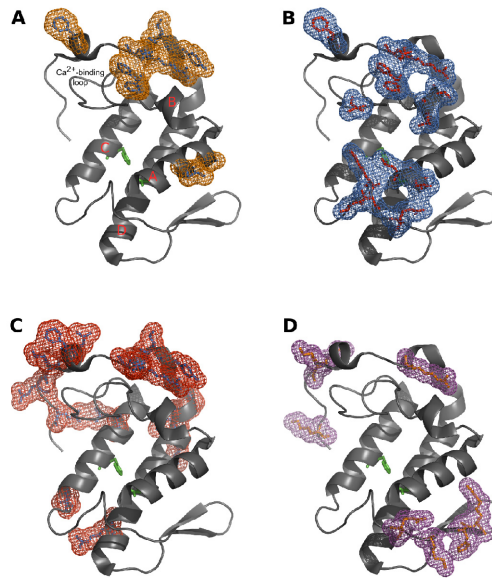
Fig. 2. Surface residues of ammodytoxin A (AtxA) responsible for A) presynaptic neurotoxicity, B) interfacial membrane binding, C) binding to calmodulin and D) binding to factor Xa. The molecule is oriented with the interfacial binding surface (IBS) and the N-terminal residues facing the viewer, while the C-terminal region is located in the upper-left corner of the molecule; and the ß-structure, in the lower-right corner. A) The conserved sPLA$_2$ helical structures are labelled with alphabet letters (red) (see section 4.1 for details; Saul et al., 2010). Residues important for neurotoxicity are presented in orange mesh surface and extend from the C-terminal region 115–119 (the YIRN cluster), including the spatially close aromatic Phe124 and Phe24, to the N-terminal helix region with the Met7/Gly11 pair in the "lower right" part of the molecule. B) The IBS residues of AtxA are presented in blue mesh surface (Leu2, Leu3, Leu19, Thr20, Phe24, Val31, Ser67, Lys69, Thr70, Arg72, Arg118, Asn119 and Phe124) and surround the active site pocket with His48 and Asp99 (presented in green). C) The Atx–CaM interaction surface comprises most of the C-terminal residues of AtxA in the region 108–131 and several basic residues in α-helices C and D (red mesh surface representation; more details in Kovačič et al., 2010). D) The anticoagulant activity of AtxA is a consequence of its binding to factor Xa, which involves mostly basic residues (magenta mesh surface): Arg118, Lys127, Lys128 and Lys132 in the C-terminal end and Arg72, Lys74, His74 and Arg77 in the ß-structure region (Prijatelj et al., 2006a).

Ca^{2+} cofactor, which is coordinatively bound by three main-chain carbonyl oxygen atoms of residues in the conserved Ca^{2+}-binding loop of the enzyme (Tyr28, Gly30 and Gly32) by two carboxylate oxygen atoms of Asp49, and by two oxygen atoms from the phospholipid substrate (Scott et al., 1990). Calcium ion is required for the initial binding of a phospholipid molecule to the active site of sPLA$_2$s and for the catalytic step, but it is not necessary for adsorption of sPLA$_2$s to the membrane (Yu et al., 1993). Given that the binding of sPLA$_2$s to a membrane surface is structurally and kinetically independent of the subsequent binding

and catalytic steps at the active site (Berg et al., 2001), the term substrate specificity in the case of sPLA$_2$s is a combination of two independent "specificities": (1) the affinity of the enzyme for binding to a membrane surface, governed by the interaction of the IBS and 20–40 phospholipids on the surface, and (2) the relative velocity of hydrolysis of different phospholipid species by the membrane-bound enzyme that obeys Michaelis-Menten kinetics (Berg et al., 2001; Lambeau & Gelb, 2008). The latter is determined by many factors influencing the interactions of the substrate molecule in the active site cleft, the rate of the catalytic reaction and the rate of release of the reaction products. In general, the active sites of sPLA$_2$s display low specificity for different phospholipid head-groups and acyl chains (Singer et al., 2002). As a consequence, the physiological functions of some sPLA$_2$s, e.g., the human group IIA (hGIIA), V (hGV) and X (hGX) enzymes, are significantly influenced or even determined by their different interfacial binding specificities and not by the specificity of their catalytic sites (Beers et al., 2003; Bezzine et al., 2000, 2002; Pan et al., 2002; Singer et al., 2002). Therefore, factors influencing interfacial binding of mammalian and toxic sPLA$_2$s, such as the composition and physical properties of the membrane, the nature of the IBS of the enzyme, and the concentration of phospholipids that are accessible to the sPLA$_2$ (Mounier et al., 2004), are crucial determinants of sPLA$_2$ biological activity.

The IBS of most sPLA$_2$s comprises a ring of conserved hydrophobic residues, whereas, around them and on the edges of the IBS, some polar, basic and acidic residues are present (Petan et al., 2005; Snitko et al., 1997). Given that most sPLA$_2$s have high activities on anionic membrane surfaces, it was long thought that electrostatic forces between cationic residues of the enzyme and anionic membrane phospholipids are crucial for the membrane binding and activity of sPLA$_2$s (Scott et al., 1994). However, a number of studies have shown that, in fact, hydrophobic, aromatic and hydrogen-bonding interactions account for most of the binding energy, even to negatively charged membrane surfaces (Gelb et al., 1999; Ghomashchi et al., 1998; Lin et al., 1998). Nevertheless, it has been suggested that electrostatic interactions are responsible for proper orientation and initial association of sPLA$_2$s with both anionic and zwitterionic membranes. They may significantly modulate the dynamics of establishing strong hydrophobic interactions, when aliphatic and aromatic residues on the IBS penetrate the membrane surface, needed to form a stable sPLA$_2$-membrane complex (Beers et al., 2003; Petan et al., 2005; Prijatelj et al., 2008; Stahelin and Cho, 2001).

4.1 Basic structural, interfacial kinetic and binding properties of Atxs

Atxs are highly basic proteins (AtxA has a pI of 10.2, net charge +6), with structural features typical of group IIA sPLA$_2$ enzymes (Figure 2, A): an N-terminal α-helix A (residues 1–14), a short α-helix B (residues 16–22), a Ca^{2+}-binding loop (residues 25–35), a long α-helix C (residues 39–57), a loop preceding an antiparallel two-stranded ß-sheet (ß-structure; residues 75–78 and 81–84), a long α-helix D (antiparallel to helix C; residues 89–109) and a C-terminal extension (mostly disordered, with two short helical turns; residues 110–133). The active site cleft is buried at the end of a hydrophobic channel (formed by the residues Phe5, Gly6, Tyr22, Ser23, Cys29, Gly30, Ala102, Ala103, Phe106) leading to the highly conserved His48–Asp99 dyad located between the antiparallel α-helices C and D. The entrance to the catalytic site of all sPLA$_2$s is surrounded by the IBS group of residues, forming a relatively flat surface on one side of the molecule that is responsible for membrane binding – the first and prerequisite step in sPLA$_2$ catalytic action. The IBS of Atxs (Figure 2, B) is formed by Leu2, Leu3, Leu19, Thr20, Phe24, Val31, Ser67, Lys69, Thr70,

Arg72, Arg118, Asn119, and Phe124 (Petan et al., 2005). The slight structural flexibility observed for the exposed side chains of residues in the IBS (e.g., Phe24 in Figure 2, B) is in keeping with the role of these residues in supporting optimal interactions of the molecule with the dynamic structure of phospholipid aggregates (Saul et al., 2010).

According to the enzymatic activities and membrane-binding affinities of Atxs determined on different phospholipid vesicles, it is clear that these snake venom sPLA$_2$s are very effective in binding to and hydrolyzing different phospholipid membranes (Petan et al., 2005). This property exists despite the fact that they have evolved to be specific and potent neurotoxic molecules and despite the fact that they share striking structural and functional similarities with the mammalian (non-toxic) sPLA$_2$s. The enzymatic activities of Atxs on vesicles composed of anionic phosphatidylglycerol (PG), which is, in general, the best sPLA$_2$ substrate, are comparable to those displayed by the most potent mammalian sPLA$_2$, the pancreatic group IB sPLA$_2$, and 5-fold higher than those displayed by the mammalian group IIA sPLA$_2$s on these vesicles (Singer et al., 2002). Atxs also have particularly high activities on phosphatidylserine (PS) vesicles, the main anionic phospholipid in eukaryotic membranes, well above the activities of the group IB and IIA sPLA$_2$s that are the most active mammalian sPLA$_2$s on these vesicles. Most importantly, the activities of Atxs were high also on pure zwitterionic phosphatidylcholine (PC) vesicles, much higher than that of the highly homologous and cationic hGIIA enzyme, which cannot bind to the PC-rich plasma membranes of mammalian cells (Beers et al., 2003; Bezzine et al., 2000, 2002; Birts et al., 2009; Singer et al., 2002). The hGIIA sPLA$_2$ is well known for its preference for anionic phospholipid substrates and its negligible activity on PC-rich membrane surfaces (Beers et al., 2003; Bezzine et al., 2000, 2002). These properties strongly influence the physiological role of the hGIIA enzyme, for example, by enabling its high antibacterial concentrations in human tears without affecting the corneal epithelial cells (Birts et al., 2009). Although the activities of Atxs were lower than those displayed by group V and X sPLA$_2$s, which are by far the most potent among the mammalian sPLA$_2$s in hydrolyzing PC vesicles and releasing fatty acids from cell membranes (Bezzine et al., 2000, 2002; Singer et al., 2002), they were able to hydrolyze plasma membranes of different intact mammalian cells at a rate that correlated well with their specific activities on PC vesicles (Petan et al., 2005; Pražnikar et al., 2008). The high activity of Atxs on PC-rich vesicles is a consequence of the ability of these venom sPLA$_2$s to bind well to such membrane surfaces, with affinities comparable to those of mammalian group V and X sPLA$_2$s. Furthermore, unlike the neutral hGX and similarly to the highly cationic hGIIA enzyme, the presence of anionic phospholipids in the membrane surface greatly enhances the membrane-binding affinity of Atxs and consequently the rate of phospholipid hydrolysis (Bezzine et al., 2000, 2002; Petan et al., 2005, 2007). This property may have an important influence on both localization of the toxin to its target membrane and its enzymatic effectiveness *in vivo*. It is now clear that, when conditions of high affinity binding apply (i.e., binding of hGIIA or Atxs to anionic PG vesicles), sPLA$_2$s reach their half-maximal enzymatic activities at low micromolar concentrations of Ca^{2+} (Petan et al., 2005; Singer et al., 2002). The proposed cytosolic action of Atxs (Pražnikar et al., 2008), and most probably other structurally similar sPLA$_2$s, is strongly supported by several factors. They may enable the enzymatic activity of Atxs in a reducing environment containing insufficient (nanomolar) concentrations of calcium. In support of this idea are (1) the high degree of stability of AtxA under conditions resembling those in the cytosol of eukaryotic cells (Kovačič et al., 2009; Petrovič et al., 2004), (2) the additional and very significant structural

stabilization and augmentation of sPLA$_2$ enzymatic activity by CaM (Kovačič et al., 2009, 2010), (3) the transient cytosolic microdomains of high local calcium concentrations (~100 μM) (Meldolesi et al., 2002) or calcium entry through the damaged plasmalemma due to sPLA$_2$ action prior to internalization (Montecucco et al., 2008), and (4) the presence of anionic phospholipids (PS) on the cytosolic face of the plasma membrane and internal cellular organelles (Okeley & Gelb, 2004). In conclusion, Atxs are effective enzymes that bind strongly to and hydrolyze rapidly both anionic and zwitterionic phospholipid aggregates, including mammalian plasma membranes, presenting a broad combination of properties characteristic of different mammalian sPLA$_2$s. The potential for the high enzymatic activity of Atxs appears to be at odds with their specific neurotoxic action. However, it is in line with the possible limitations imposed on their intracellular activity on a particular target membrane by the harsh conditions in the cytosol during the process of neuromuscular transmission blockade.

4.2 Role of different IBS residues in supporting interfacial binding and activity of Atxs

The presence of tryptophan on the IBS of different sPLA$_2$s is a well-known determinant of their ability to bind with high affinity and hydrolyze PC-rich membranes, crucially influencing their biological roles. Its role has been highlighted in the cases of the hGV and hGX enzymes (Bezzine et al., 2002; Han et al., 1999), the acidic sPLA$_2$ from *Naja naja atra* snake venom (Sumandea et al., 1999), a range of mutants of hGIIA (Beers et al., 2003) and the pancreatic group IB sPLA$_2$s (Lee et al., 1996). Despite the fact that Atxs do not contain a Trp residue, they display a relatively high activity on PC vesicles. Furthermore, the substitution of Val31 by Trp led to a dramatic 27-fold increase in the activity of AtxA (Petan et al., 2005), reaching a level of activity on PC-rich vesicles higher than those of hGV and hGX and in the range of the best-acting snake venom sPLA$_2$s, e.g., the cobra venom sPLA$_2$ (Sumandea et al., 1999). However, despite its very high affinity for PC-rich surfaces and its order of magnitude higher potency in releasing fatty acids from plasma membranes of intact HEK293 cells, C2C12 myocytes and motoneuronal NSC34 cells (Pražnikar et al., 2008), the AtxA-V31W mutant did not display a major change in neurotoxic potency *in vivo*. This again highlights the dependence of the toxicity of sPLA$_2$ ß-ntxs on a combination of toxin-acceptor interactions leading to localization of the toxin to the target membrane at which the enzymatic activity of AtxA is sufficient for the observed effects. In contrast, the same substitution, V31W, on the IBS of the enzymatically active quaternary mutant AtnL[YVGD] of the myotoxic Atx homologue, AtnL (see below), caused a 100-fold increase in its activity on PC vesicles, as well as a significant increase in its toxicity *in vivo* and *in vitro* (Petan et al., 2007). Thus, in the case of the enzymatically active mutants of AtnL, the correlation between enzymatic activity on PC membranes and toxicity *in vivo* suggests that the mechanism of their toxicity differs from that used by Atxs and that it depends largely on their interfacial binding affinity and enzymatic activity on PC-rich target membranes, both of which are significantly affected by Trp31 (Petan et al., 2007).

The role of aromatic residues in the interfacial binding of sPLA$_2$s depends to a high degree on the nature of the residue itself, its position on the IBS and the orientation of its side-chain (Stahelin & Cho, 2001; Sumandea et al., 1999). This is evident from the fact that the substitution of Phe24 with Trp did not cause a substantial increase in enzymatic activity or interfacial binding affinity of AtxA (Petan et al., 2002, 2005) or hGIIA (Beers et al., 2003).

Thus, both Phe and Trp can have similar roles in interfacial binding, despite the differences in their interactions with the membrane: the highly amphiphilic Trp favours partitioning in the interfacial phospholipid head group region of the bilayer (Yau et al., 1998), while the aromatic Phe penetrates deeper into the hydrophobic core of the phospholipid acyl chains (Stahelin & Cho, 2001; Sumandea et al., 1999). In the case of AtxA and hGIIA, Trp31 is obviously in a much better position to influence interfacial binding than is Trp24. Additionally, AtxA and AtxA^{F24W} display higher activities and binding affinities than do the F24S, F24Y and F24N mutants on PC-rich vesicles containing anionic phospholipids. Although the structures of Trp and Phe differ significantly, both are obviously better suited to take advantage of the presence of anionic phospholipid in the interface than the polar Ser, Asn and even the aromatic Tyr. The importance of non-polar interactions in interfacial binding to negatively charged surfaces is also clear in the case of AtxB, AtxA^{KKML} and AtxA^{V31W}. These molecules already display very high activities on PC vesicles containing 10% PS (10% PS/PC vesicles), and they reach the level of maximal activity already on 30% PS/PC vesicles, indicating that the enzymes are fully bound to these vesicles (Petan et al., 2005). In general, we have observed the greatest increases in activity upon introduction of 10% anionic phospholipids (PS or PG) into PC vesicles for mutants that have numbers of hydrophobic and aromatic residues on their IBS similar to or higher than those in AtxA. Besides providing the basis for electrostatic interactions, the presence of anionic phospholipids in PC vesicles may facilitate non-polar interactions as a result of membrane perturbation (Buckland & Wilton, 2000).

Despite the remarkable impact of Trp at position 31 on interfacial binding of Atxs and other sPLA$_2$s, such as hGIIA (Beers et al., 2003), our subsequent site-directed mutagenesis studies revealed that the positive influence of Trp on membrane binding of sPLA$_2$s may be significantly diminished, depending on the delicate balance of contributions of each IBS residue to interfacial binding. While Atxs and hGIIA sPLA$_2$ do not contain a Trp on their IBS, DPLA$_2$, a weakly neurotoxic sPLA$_2$ from Russell's viper that differs from AtxA in only 22 residues, has a Trp residue at position 31 and yet displays interfacial binding and kinetic properties similar to those of AtxA (Petan et al., 2005; Prijatelj et al., 2003, 2008). Besides V31W, DPLA$_2$ contains several other substitutions of equivalent AtxA IBS residues: L19I, T20P, F24S, S67N, T70S, R118M and N119L (Petan et al., 2005; Prijatelj et al., 2008). By analysing the properties of a range of AtxA/DPLA$_2$ mutants and chimeras (Prijatelj et al., 2008), we were able to pinpoint the IBS residues that are crucial for the intriguing differences in interfacial binding and activity of AtxA and DPLA$_2$. Not surprisingly, the introduction of Trp31 to the AtxA/DPLA$_2$^{YIRN} chimera (see Figure 1 for sequence data) caused a very significant 50-fold increase in the rate of hydrolysis of PC vesicles. However, the impact of the same Trp residue, when introduced along with two additional substitutions, producing the AtxA^{KEW}/DPLA$_2$^{YIRN} mutant, was almost completely abolished, resulting in a low level of activity on PC vesicles, only 2 to 4-fold higher than those of AtxA/DPLA$_2$^{YIRN}, AtxA and DPLA$_2$. These results clearly confirmed our earlier suggestion (Petan et al., 2005) that the well-known positive impact of Trp31 on the interfacial binding and enzymatic activity of mammalian and venom sPLA$_2$s is, to a large extent, counterbalanced by the presence of Lys7 and Glu11 on the edges of the IBS in the case of DPLA$_2$, instead of the hydrophobic Met7 and nonpolar Gly11 in AtxA. This is most probably a consequence of altered orientation of DPLA$_2$ on the membrane, due to strong electrostatic interactions between Lys7 and Glu11 and the zwitterionic membrane surface,

which prevents the productive interaction of Trp31, and thus DPLA$_2$, with the interface. Similarly, two basic residues at nearly equivalent positions in the hGIIA enzyme, Arg7 and Lys10, were shown to have a negative influence on interfacial binding (Bezzine et al., 2002; Snitko et al., 1997). In agreement with this, we have shown that the substitution of Arg at position 72 of the IBS of Atxs and DPLA$_2$ has a positive impact on interfacial binding and activity only upon introduction of a hydrophobic (Ile), but not a polar (Ser), acidic (Glu) or other basic (Lys) residue (Ivanovski et al., 2004). An additional, albeit smaller, negative effect on interfacial binding and enzymatic activity of DPLA$_2$ is provided by the presence of the polar residue Ser24 instead of the aromatic and hydrophobic Phe in AtxA (Petan et al., 2002, 2005). However, there are several residues on the IBS of DPLA$_2$, but not Atxs, that significantly improve its interfacial binding and activity. The YIRN cluster in the C-terminal part of AtxA, shown to be important for its neurotoxic effect (Ivanovski et al., 2000), contains the hydrophilic IBS residues Arg118 and Asn119, while the hydrophobic Tyr115 and Ile116 are not part of the presumed IBS. The introduction of the YIRN cluster into DPLA$_2$, instead of its KKML cluster, or into the various AtxA/DPLA$_2$ variants, resulted in a substantial decrease of the initial rates of hydrolysis of charge-neutral PC vesicles (Table 2). Therefore, the presence of a hydrophobic or aromatic IBS residue at positions 118 and 119 has a significant positive influence on interfacial binding to anionic and, particularly, zwitterionic phospholipid surfaces. Accordingly, this positive impact is evident in the case of Met118 and Leu119 in DPLA$_2$ and the AtxA[KKML] mutant, as well as of Met118 and Tyr119 in AtxB, in contrast to the hydrophilic Arg118 and Asn119 in AtxA and mutants containing the YIRN cluster (Ivanovski et al., 2000; Petan et al., 2005; Prijatelj et al., 2003, 2008). Considering the similar activities on PC-rich vesicles displayed by AtxB and the AtxA[KKML] mutant, it is obvious that the roles of Tyr and Leu at position 119 are in fact very similar. This indicates that the removal of the polar cluster Arg118/Asn119 from AtxA has a greater positive effect on interfacial binding than the aromatic or hydrophobic nature of the substituting residue at position 119, which is in accordance with the negligible impact of the substitution Y119W on interfacial binding of hGIIA sPLA$_2$ to 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) vesicles (Beers et al., 2003). Thus, the role of aromatic residues, including Trp, in interfacial binding of sPLA$_2$s, depends strongly, as well as on their nature, position on the IBS, and side-chain orientation, on the counterbalancing effects of polar/electrostatic interactions provided by hydrophilic/charged residues on or near the IBS. In conclusion, the most significant differences in interfacial binding and enzymatic activity of AtxA and DPLA$_2$ are therefore a consequence of the natural substitutions that have led to significant changes in the hydrophobic/aromatic character of residues on or near the IBS – M7K, G11E, F24S, V31W, R118M and N119L.

5. Interaction of Atxs with high-affinity binding proteins

Several high-affinity binding proteins for Atxs have been identified in porcine cerebral cortex. Two of these are membrane proteins of 25 kDa (R25) and 180 kDa (R180). R25, which is located in the mitochondrial membrane (Šribar et al., 2003a), binds only Atxs (Vučemilo et al., 1998), whereas R180, identified as the plasma membrane M-type sPLA$_2$ receptor (sPLA$_2$R), binds both toxic and non-toxic sPLA$_2$s of groups IB and IIA (Čopič et al., 1999; Vardjan et al., 2001). Surprisingly, several cytosolic, high-affinity binding proteins for Atxs were identified as well – calmodulin (CaM) (Šribar et al., 2001), the γ and

sPLA$_2$	Specific activity (μmol/(min × mg))				
	POPG	POPS	POPC	10% POPS/POPC	30% POPS/POPC
AtxA	1042 ± 160	1251 ± 188	3.8 ± 0.5	56 ± 6	450 ± 21
AtxB	1149 ± 34	1189 ± 153	14 ± 2	240 ± 22	1133 ± 150
AtxC	1116 ± 91	477 ± 101	1.9 ± 0.2	14.0 ± 1.3	166 ± 21
AtnI$_2$	1070 ± 150	57 ± 5	12.3 ± 1.7	49 ± 2	159 ± 16
AtxA^{KKML}	1148 ± 126	1252 ± 40	19 ± 1	400 ± 22	1322 ± 42
AtxA^{V31W}	2102 ± 88	1964 ± 154	102 ± 7	525 ± 28	1957 ± 36
AtxA^{F24W}	914 ± 43	906 ± 56	4.4 ± 0.5	50 ± 1	209 ± 6
AtxA^{F24Y}	822 ± 41	1304 ± 102	2.1 ± 0.1	9.5 ± 0.5	163 ± 11
AtxA^{F24N}	780 ± 29	535 ± 43	1.09 ± 0.03	5.1 ± 0.5	68 ± 3
AtxA^{F24A}	813 ± 107	1291 ± 65	1.2 ± 0.1	4.4 ± 0.3	92 ± 5
AtxA^{F24S}	400 ± 73	430 ± 42	0.71 ± 0.06	4.4 ± 0.1	24 ± 4
AtxA^{R72E}	438 ± 26	n. d.	0.33 ± 0.07	n. d.	n. d.
AtxA^{R72I}	1655 ± 108	n. d.	6.9 ± 0.8	n. d.	n. d.
AtxA^{R72K}	686 ± 102	n. d.	1.2 ± 0.1	n. d.	n. d.
AtxA^{R72S}	800 ± 82	n. d.	1.6 ± 0.3	n. d.	n. d.
DPLA$_2$	1600 ± 130	1195 ± 47	1.1 ± 0.1	88 ± 9	442 ± 31
DPLA$_2$^{YIRN}	1100 ± 120	996 ± 129	0.07 ± 0.02	4.2 ± 0.4	141 ± 20
AtxA/DPLA$_2$	1300 ± 140	n. d.	3.14 ± 0.05	n. d.	n. d.
AtxA^{KEW}/DPLA$_2$^{YIRN}	1200 ± 170	n. d.	2.2 ± 0.5	n. d.	n. d.
AtxA^{KE}/DPLA$_2$^{YIRN}	830 ± 80	n. d.	0.14 ± 0.01	n. d.	n. d.
AtxA^{W}/DPLA$_2$^{YIRN}	1900 ± 210	n. d.	26 ± 3	n. d.	n. d.
AtxA/DPLA$_2$^{YIRN}	1000 ± 120	n. d.	0.53 ± 0.06	n. d.	n. d.
AtnL	No activity	No activity	No activity	n. d.	n. d.
AtnL^{YVGD}	225 ± 35	91 ± 7	~0.005	0.49 ± 0.04	43 ± 2
AtnL^{YWGD}	180 ± 25	89 ± 10	0.22 ± 0.03	8.3 ± 0.3	96 ± 5
12-AtxA	5.2 ± 0.7	n. d.	0.76 ± 0.06	n. d.	n. d.
I-AtxA	4.0 ± 0.2	n. d.	0.96 ± 0.04	n. d.	n. d.
P-AtxA	1.29 ± 0.04	n. d.	0.13 ± 0.03	n. d.	n. d.
hGIIA	220 ± 90	40 ±18	Lag, 0.7 ± 0.2	n. d.	n. d.
hGX	14 ± 0.8	4 ± 2	30 ± 0.2	n. d.	n. d.

Table 2. Specific enzymatic activity of sPLA$_2$s on phospholipid vesicles. The specific enzymatic activities were determined on extruded phospholipid vesicles composed of 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC), 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoserine (POPS), 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoglycerol (POPG), and on vesicles with mixed compositions as stated (in molar percentage) in the Table. The initial rates of phospholipid hydrolysis were measured using a sensitive fluorescence, fatty acid displacement assay; the enzymatic activities were calculated after calibrating the responses with known amounts of oleic acid (Petan et al., 2005). The enzymatic activity value for each sPLA$_2$ is the mean ± S.D. of at least five independent measurements. Data taken from Ivanovski et al. (2004), Petan et al. (2005, 2007), Prijatelj et al. (2006b, 2008), and Singer et al. (2002).

ε isoforms of 14-3-3 proteins (Šribar et al., 2003b), and protein disulphide isomerase (Šribar et al., 2005). This suggests the possibility of not only intracellular, but also cytosolic, action of Atxs and other sPLA$_2$s. They have since been shown to be surprisingly stable in the reducing environment of the eukaryotic cytosol. They are enzymatically active at low micromolar concentrations of Ca^{2+} ions, and they are also structurally stabilized and activated by binding to CaM (Kovačič et al., 2009, 2010; Petan et al., 2005; Petrovič et al.,

2005). Atxs were also found to bind factor Xa (FXa) and display anticoagulant activity (Prijatelj et al., 2006a). Given that the topic of neuronal high-affinity binding proteins for Atxs and their putative roles in ß-neurotoxicity has been reviewed recently (Pungerčar & Križaj, 2007), here we only summarize the results obtained with a range of Atx mutants for binding to CaM, R25, FXa and the M-type sPLA$_2$ receptor.

5.1 The C-terminal region of Atxs has a novel calmodulin binding motif

Our protein engineering and competition binding studies using a range of Atxs, DPLA$_2$ and AtnI$_2$ mutants and chimeras, as well as synthetic peptides corresponding to different regions of Atxs, ultimately resulted in the identification of a novel binding motif for the ubiquitous, highly conserved Ca^{2+}-sensor CaM (Kovačič et al., 2010; Prijatelj et al., 2003). The C-terminal region of AtxA, with a distinctive hydrophobic patch within the region 107–125 surrounded by several basic residues, was identified as the most important element for binding to the N-terminal methionine-rich pocket of CaM in its Ca^{2+}-induced dumbbell conformation. This work was based on the intriguing 150-fold lower lethal potency and the correspondingly 50-fold lower CaM-binding affinity of DPLA$_2$ compared to AtxA (Prijatelj et al., 2003). Most notably, the DPLA$_2$[YIRN] mutant, containing the C-terminal YIRN cluster of AtxA (see above and Figure 1), displayed a 7-fold increase in binding affinity for CaM, only 7-fold lower than that of AtxA. This result suggested a major role of this region for the interaction with CaM. On the other hand, introduction of the whole N-terminal half of AtxA to DPLA$_2$ and DPLA$_2$[YIRN] caused only a slight, 2 to 3-fold increase in binding affinities for CaM, suggesting that the N-terminal part of AtxA has only a supporting role in the binding and is probably located on the periphery of the CaM binding site (Prijatelj et al., 2008). Very recently, we confirmed unambiguously the crucial importance of the C-terminal region of Atxs in CaM binding by building several three-dimensional structural models of sPLA$_2$–CaM complexes, based on AtxA–CaM interaction site mapping, our previous mutagenesis data and protein-docking algorithms (Kovačič et al., 2010). According to the model, the closed conformation of CaM forms a "clamp" around AtxA, with most of the C-terminal residues of AtxA in the region 108-131 being in direct contact with CaM (Figure 2, C). There are only a few additional, but important, contacts between the proteins, formed mainly by basic residues in α-helices C (Arg43 and Asn54) and D (Arg94, Lys108, Asn109 and Lys111) of AtxA. Importantly, the model was validated by superimposing the structures of several mammalian sPLA$_2$s on that of AtxA in the complex and correctly predicting the favourable interactions of group V and X mammalian sPLA$_2$s and CaM. The model also showed the basis for the unfavourable interactions, which prevented complex formation between the structurally very similar mammalian group IB and IIA sPLA$_2$s and CaM. Formation of the complexes was confirmed by competition binding experiments and the augmentation of the enzymatic activity of those sPLA$_2$s interacting with CaM. Most importantly, we have clearly demonstrated that CaM stabilizes and protects Atxs from denaturation in the reducing environment of the cell cytosol and acts as a non-essential activator of Atxs and of the mammalian group V and X sPLA$_2$s. The fact that CaM-binding results in stabilization and augmentation of enzymatic activity of sPLA$_2$s, ranging from the neurotoxic Atxs to the mammalian group V and X sPLA$_2$s, suggests a novel role for CaM as a regulator of the intracellular actions of some toxic and non-toxic sPLA$_2$s, which may be active even in the reducing environment of the cytosol (Kovačič et al., 2009, 2010). The significance of CaM–sPLA$_2$ interactions in the pharmacological and (patho)physiological roles of sPLA$_2$s remain to be confirmed.

5.2 The N- and C-terminal regions of Atxs are involved in binding to the neuronal M-type sPLA$_2$ receptor (R180)

The M-type sPLA$_2$ receptor, which was cloned and revealed as a member of the C-type lectin superfamily, has been well-characterized and is currently the best-known sPLA$_2$ binding target. It was shown recently that the M-type receptor is involved in regulating cell senescence (Augert et al., 2009) and is the main antigen target in autoimmune human membranous nephropathy (Beck et al., 2009). However, the pathophysiological implications of the interaction between the M-type sPLA$_2$ receptors and both toxic and non-toxic sPLA$_2$s in general are intriguing and still unknown (Lambeau & Gelb, 2008; Pungerčar & Križaj, 2007; Rouault et al., 2006). Although it has been suggested that the neuronal M-type sPLA$_2$ receptor located on the plasma membrane could be responsible for specific targeting and internalization of sPLA$_2$ neurotoxins in presynaptic nerve terminals, the results of our mutagenesis studies have shown that it may not be involved in the neurotoxicity of Atxs (Prijatelj et al., 2006b). In order to prevent the interaction between AtxA and the receptor, we prepared three mutants of AtxA, containing either the 12-amino acid-long peptide ARIRARGSIEGR, named 12-AtxA, or one each of two variants of a shorter 5-amino acid peptide (ASIGQ and ASPGQ; named I-AtxA and P-AtxA, respectively) fused to the N-terminus of AtxA. These peptides are similar to the N-terminal propeptides present in proenzymes of mammalian group IB (e.g., DSGISPR in human pancreatic proPLA$_2$) and group X sPLA$_2$s (e.g., EASRILRVHRR in the human proenzyme), which do not bind to the M-type sPLA$_2$ receptor (Hanasaki & Arita, 2002; Lambeau et al., 1995). The presence of the fusion peptides in AtxA indeed completely abolished the interaction of the toxin with the M-type receptor. However, the mutants displayed only one order of magnitude lower lethality in mice and were able to induce neurotoxic effects on a mouse phrenic nerve-hemidiaphragm preparation (see Table 1). Although the N-terminal fusion peptides acted effectively as "propeptides" for AtxA and lowered specific enzymatic activity to ~1% of the wild-type enzyme, their similar neurotoxic profiles on the neuromuscular junction indicate that minimal enzymatic activity suffices for presynaptic toxicity of sPLA$_2$s. Additionally, antibodies targeting the sPLA$_2$-binding C-type lectin-like domain 5 of the M-type sPLA$_2$ receptor were unable to abolish the neurotoxic action of AtxA on the neuromuscular preparation. Thus, the interaction of AtxA with the neuronal M-type sPLA$_2$ receptor, R180, is apparently not essential for its presynaptic neurotoxicity. Since we performed the binding experiments with the N-terminal fusion AtxA mutants, using porcine sPLA$_2$ receptor, the question still remains as to whether the fusion peptides would also abolish the binding of AtxA to the mouse R180 receptor.

Nevertheless, our site-directed mutagenesis studies have provided important information about the parts of the AtxA molecule involved in binding to R180. Interestingly, the C-terminal YIRN cluster, which is essential for neurotoxicity and binding to CaM and R25, appears to be involved in binding to R180 as well. The first clues were provided by the lower affinity of the AtxA^{KKML} mutant and DPLA$_2$ for binding to R180 (Ivanovski et al., 2000). (Both contain KKML instead of the YIRN cluster.) The higher binding affinity of mutants containing the YIRN cluster (DPLA$_2$^{YIRN} or AtxA/DPLA$_2$^{YIRN}) confirmed that it is an important part of the AtxA/R180 interaction surface. However, other parts of the molecule also contribute considerably to the binding affinity of AtxA. Interestingly, the N-terminal part of AtxA in the chimera AtxA/DPLA$_2$ did not increase the binding affinity of DPLA$_2$ to R180. However, the introduction of three N-terminal DPLA$_2$ residues (K, E

and/or W) to AtxA/DPLA$_2^{YIRN}$ caused an increase in the binding affinity for R180 up to the level of AtxA, confirming the expected involvement of the N-terminal part of the molecule in binding to R180 as well. Therefore, the interaction surface of AtxA with its neuronal M-type receptor, R180, extends from the C-terminal top of the molecule, across the area close to the Ca^{2+}-binding loop and the IBS, reaching the N-terminal helix on the edges of the IBS.

5.3 Atxs bind to the mitochondrial receptor R25 through their C-terminal region

The C-terminal region of Atxs appears to be essential also for binding to the, as yet unidentified, mitochondrial membrane receptor R25. The affinities for binding of the chimeric mutants to R25 ranged from that of the non-toxic AtnI$_2$, which does not bind to R25, to those of AtxA, AtnI$_2$/AtxA^{K108N} and AtnI$_2^{N24F}$/AtxA^{K108N}, which were similar to that of AtxA (Prijatelj et al., 2002). Additionally, substitutions of Phe24 in the N-terminal part of AtxA had a negligible effect on its affinity for R25, but nevertheless indicated that this residue is in the vicinity of the toxin-receptor binding surface, most probably located in the C-terminal part of AtxA (Petan et al., 2002). Furthermore, while DPLA$_2$ and the chimera AtxA/DPLA$_2$ could not completely inhibit the binding of AtxC to R25 in competition binding experiments, the substitution of their KKML cluster with the YIRN cluster led to complete inhibition with a relatively high binding affinity. This implies that residues in the region 115–119 are very important for effective binding to R25, which is apparently present in multiple isoforms. However, the complete inhibition of binding of radiolabelled AtxC to R25 by the AtxA^{KKML} mutant, in contrast to DPLA$_2$ and AtxA/DPLA$_2$, indicated that the interaction surface extends beyond the C-terminal region. Indeed there was a slight increase in binding affinity upon introduction of the N-terminal Lys7, Glu11 and/or Trp31 of DPLA$_2$ to the AtxA/DPLA$_2^{YIRN}$ chimera (K, E and/or W; Table 1). While our site-directed mutagenesis and competition binding studies have so far revealed a central role for the C-terminal residues of Atxs in high affinity binding to R25, the interaction surface probably extends beyond the C-terminal region.

5.4 Atxs exert anticoagulant effects by binding to factor Xa

In the analysis of non-neurotoxic effects of Atxs, our study of an array of site-directed mutants (Prijatelj et al., 2006a) showed that basic residues in the C-terminal and ß-structure regions are important for AtxA binding to human blood coagulation factor Xa (Figure 2, D). Hence, AtxA has anticoagulant activity. In addition, the 10-fold lower affinity of AtxC for factor Xa, in comparison with AtxA, is due to the (natural) charge-reversal substitution of Lys128 by Glu, leading to a small local conformational change in the C-terminal end of AtxC, which perturbs the interaction with factor Xa (Saul et al., 2010).

6. Restoration of enzymatic activity reduces the Ca^{2+}-independent membrane damage induced by the myotoxic ammodytin L

Ammodytin L (AtnL) is a myotoxic and enzymatically inactive structural homologue of Atxs (Petan et al., 2007; Pungerčar et al., 1990). The main structural feature of snake venom sPLA$_2$-like myotoxins is the substitution of the highly-conserved aspartic acid residue at position 49, effectively preventing Ca^{2+} binding and thus abolishing catalytic

activity (Petan et al., 2007; Ward et al., 2002). The most common substitution is Lys, but AtnL is one of the two known Ser49 sPLA$_2$ homologues. In addition to this replacement, several other substitutions of highly-conserved residues in the enzymatically inactive sPLA$_2$s are found concentrated in the region of the Ca^{2+}-binding loop, where the carbonyl oxygen atoms of Tyr28, Gly30 and Gly32 provide three additional coordinative bonds for Ca^{2+}-binding. The myotoxic activity of the Lys49 sPLA$_2$s, as well as of the Ser49-containing AtnL, is best characterized by their ability to induce myonecrosis *in vivo* and show a potent Ca^{2+}-independent membrane-damaging activity *in vitro* (Lomonte et al., 2009). Although the involvement of a skeletal muscle protein acceptor in the process of sPLA$_2$-induced myotoxicity cannot be ruled out, a considerable body of evidence has been accumulated indicating that the myotoxic effects are a consequence of a Ca^{2+}-independent protein-lipid interaction that causes direct damage of the plasma membrane of target muscle cells and may allow entry of extracellular Ca^{2+} and irreversible cell damage (Cintra-Francischinelli et al., 2010; Lomonte et al., 2009; Rufini et al., 1992). The calcium-independent, and thus enzymatic activity-independent, mechanism of membrane damage is also supported by experiments in which inhibition of certain enzymatically active Asp49 myotoxic sPLA$_2$s did not eliminate their myotoxic activity (Soares et al., 2001).

We recently described the first example of restoration of activity in AtnL or in any enzymatically inactive Ser49/Lys49 sPLA$_2$ homologue (Petan et al., 2007). We prepared two enzymatically active quaternary mutants of AtnL (H28Y/L31V,W/N33G/S49D), differing at position 31. Although Asn33 is not directly involved in Ca^{2+} coordination, we suspected that it might have a negative impact on the optimal local structure of the Ca^{2+}-binding loop. We therefore replaced it with a glycine residue, which is very often found at this position in catalytically active sPLA$_2$s, as well as in Atxs. Val31 was present in the AtnL^{YVGD} mutant in order to recreate the Ca^{2+}-binding loop of Atxs, while the AtnL^{YWGD} mutant had Trp31 in order to enhance its membrane binding affinity, and thus enzymatic activity, as previously shown in the case of the AtxA^{V31W} mutant described above. The successful restoration of enzymatic activity by such a small number of substitutions indicates that, apart from the residues involved in Ca^{2+} coordination, the remainder of the substrate-binding and catalytic network of AtnL is very well conserved. Apparently, in evolution, Lys49 and Ser49 sPLA$_2$ myotoxins have lost their Ca^{2+}-binding ability and enzymatic activity through subtle changes in the Ca^{2+}-binding network alone, without affecting the rest of the catalytic machinery. Our results strongly suggest that these changes were selected for their Ca^{2+}-independent membrane-damaging ability and increase the specificity of their myotoxic activity. Although the restoration of enzymatic activity in AtnL increased its cytotoxic potency and lethality in general, it had a negative effect on its Ca^{2+}-independent mechanism of membrane damage and on its ability to specifically target differentiated muscle cells *in vitro*. Given that AtnL shares a high level of identity with AtxA (74%), it was not surprising that the enzymatically active mutants of AtnL displayed a combination of the properties of both toxins. In other words, the restoration of enzymatic activity of AtnL reduced its ability to act as a potent and specific myotoxic molecule. The latter supports the idea of an evolutionary specialization of Ser49/Lys49 sPLA$_2$ homologues to perform the role of abundant and weakly lethal, but specific, muscle-targeting toxins in the arsenal of pharmacologically active molecules found in snake venom.

7. Future perspectives

Recently, we have been able to introduce a free cysteine residue into the molecule of AtxA by substituting Asn79 by Cys at the turn of the antiparallel ß-sheet. This mutation allowed for specific labelling of the neurotoxin molecule by either fluorescent or nanogold probes (Pražnikar et al., 2008, 2009). The conjugates were less toxic than the wild-type toxin, but retained its basic properties. They have proved to be a valuable tool in the study of presynaptic neurotoxicity. As an illustration, we have provided evidence that AtxA, a neurotoxic snake venom sPLA$_2$, is indeed internalized into mammalian (mouse) motor nerve terminals (Logonder et al., 2009).

We believe that similar approaches can be used in further protein engineering of Atxs and also other ß-ntxs, especially in monitoring their trafficking and cellular action (enzymatic activity, binding to protein target molecules). This should finally lead to a complete understanding of the molecular mechanism of presynaptic toxicity of sPLA$_2$ neurotoxins, which may also open the way for better medical treatment of snake envenomation. Moreover, similar research approaches can also be exploited in the studies of homologous mammalian sPLA$_2$s whose biological roles are still to be clarified.

8. Acknowledgment

9. References

Augert, A., Payré, C., De Launoit, Y., Gil, J., Lambeau, G. & Bernard, D. (2009). The M-type receptor PLA$_2$R regulates senescence through the p53 pathway. *EMBO Reports*, Vol. 10, No. 3, pp. 271–277

Beck, L.H., Bonegio, R.G.B., Lambeau, G., Beck, D.M., Powell, D.W., Cummins, T.D., Klein, J.B. & Salant, D.J. (2009). M-type phospholipase A$_2$ receptor as target antigen in idiopathic membranous nephropathy. *New England Journal of Medicine*, Vol. 361, No. 1, pp. 11–21

Beers, S.A., Buckland, A.G., Giles, N., Gelb, M.H. & Wilton, D.C. (2003). Effect of tryptophan insertions on the properties of the human group IIA phospholipase A$_2$: Mutagenesis produces an enzyme with characteristics similar to those of the human group V phospholipase A$_2$. *Biochemistry*, Vol. 42, No. 24, pp. 7326–7338

Berg, O.G., Gelb, M.H., Tsai, M.D. & Jain, M.K. (2001). Interfacial enzymology: The secreted phospholipase A$_2$-paradigm. *Chemical Reviews*, Vol. 101, No. 9, pp. 2613–2654

Bezzine, S., Koduri, R.S., Valentin, E., Murakami, M., Kudo, I., Ghomashchi, F., Sadilek, M., Lambeau, G. & Gelb, M.H. (2000). Exogenously added human group X secreted phospholipase A$_2$ but not the group IB, IIA, and V enzymes efficiently release arachidonic acid from adherent mammalian cells. *Journal of Biological Chemistry*, Vol. 275, No. 5, pp. 3179–3191

Bezzine, S., Bollinger, J.G., Singer, A.G., Veatch, S.L., Keller, S.L. & Gelb, M.H. (2002). On the binding preference of human groups IIA and X phospholipases A$_2$ for membranes with anionic phospholipids. *Journal of Biological Chemistry*, Vol. 277, No. 50, pp. 48523–48534

Birts, C.N., Barton, C.H. & Wilton, D.C. (2010). Catalytic and non-catalytic functions of human IIA phospholipase A_2. *Trends in Biochemical Sciences*, Vol. 35, No. 1, pp. 28–35

Buckland, A.G. & Wilton, D.C. (2000). Anionic phospholipids, interfacial binding and the regulation of cell functions. *Biochimica et Biophysica Acta*, Vol. 1483, No. 2, pp. 199–216

Cintra-Francischinelli, M., Pizzo, P., Angulo, Y., Gutiérrez, J.M., Montecucco, C. & Lomonte, B. (2010). The C-terminal region of a Lys49 myotoxin mediates Ca^{2+} influx in C2C12 myotubes. *Toxicon*, Vol. 55, No. 2-3, pp. 590–596

Čopič, A., Vučemilo, N., Gubenšek, F. & Križaj, I. (1999). Identification and purification of a novel receptor for secretory phospholipase A_2 in porcine cerebral cortex. *Journal of Biological Chemistry*, Vol. 274, No. 37, pp. 26315–26320

Gelb, M.H., Cho, W. & Wilton, D.C. (1999). Interfacial binding of secreted phospholipases A_2: More than electrostatics and a major role for tryptophan. *Current Opinion in Structural Biology*, Vol. 9, No. 4, pp. 428–432

Ghomashchi, F., Lin, Y., Hixon, M.S., Yu, B.Z., Annand, R., Jain, M.K. & Gelb, M.H. (1998). Interfacial recognition by bee venom phospholipase A_2: Insights into nonelectrostatic molecular determinants by charge reversal mutagenesis. *Biochemistry*, Vol. 37, No. 19, pp. 6697–6710

Han, S.K., Kim, K.P., Koduri, R., Bittova, L., Munoz, N.M., Leff, A.R., Wilton, D.C., Gelb, M.H. & Cho, W. (1999). Roles of Trp31 in high membrane binding and proinflammatory activity of human group V phospholipase A_2. *Journal of Biological Chemistry*, Vol. 274, No. 17, pp. 11881–11888

Hanasaki, K. & Arita, H. (2002). Phospholipase A_2 receptor: A regulator of biological functions of secretory phospholipase A_2. *Prostaglandins & Other Lipid Mediators*, Vol. 68-69, pp. 71–82

Ivanovski, G., Čopič, A., Križaj, I., Gubenšek, F. & Pungerčar, J. (2000). The amino acid region 115-119 of ammodytoxins plays an important role in neurotoxicity. *Biochemical and Biophysical Research Communications*, Vol. 276, No. 3, pp. 1229–1234

Ivanovski, G., Petan, T., Križaj, I., Gelb, M.H., Gubenšek, F. & Pungerčar, J. (2004). Basic amino acid residues in the ß-structure region contribute, but not critically, to presynaptic neurotoxicity of ammodytoxin A. *Biochimica et Biophysica Acta*, Vol. 1702, No. 2, pp. 217–225

Kini, R.M. & Evans, H.J. (1989). A model to explain the pharmacological effects of snake venom phospholipases A_2. *Toxicon*, Vol. 27, No. 6, pp. 613–635

Kini, R.M. & Chan, Y.M. (1999). Accelerated evolution and molecular surface of venom phospholipase A_2 enzymes. *Journal of Molecular Evolution*, Vol. 48, No. 2, pp. 125–132

Kini, R.M. (2003). Excitement ahead: Structure, function and mechanism of snake venom phospholipase A_2 enzymes. *Toxicon*, Vol. 42, No. 8, pp. 827–840

Kovačič, L., Novinec, M., Petan, T., Baici, A. & Križaj, I. (2009). Calmodulin is a nonessential activator of secretory phospholipase A_2. *Biochemistry*, Vol. 48, No. 47, pp. 11319–11328

Kovačič, L., Novinec, M., Petan, T. & Križaj, I. (2010). Structural basis of the significant calmodulin-induced increase in the enzymatic activity of secreted phospholipases A_2. *Protein Engineering, Design & Selection*, Vol. 23, No. 6, pp. 479–487

Križaj, I. (2011). Ammodytoxin: A window into understanding presynaptic toxicity of secreted phospholipases A_2 and more. *Toxicon*, Vol. 58, No. 3, pp. 219–229

Lambeau, G., Ancian, P., Nicolas, J.P., Beiboer, S.H., Moinier, D., Verheij, H. & Lazdunski, M. (1995). Structural elements of secretory phospholipases A_2 involved in the

binding to M-type receptors. *Journal of Biological Chemistry*, Vol. 270, No. 10, pp. 5534–5540

Lambeau, G. & Gelb, M.H. (2008). Biochemistry and physiology of mammalian secreted phospholipases A$_2$. *Annual Review of Biochemistry*, Vol. 77, pp. 495–520

Lee, B.I., Yoon, E.T. & Cho, W. (1996). Roles of surface hydrophobic residues in the interfacial catalysis of bovine pancreatic phospholipase A$_2$. *Biochemistry*, Vol. 35, No. 13, pp. 4231–4240

Lin, Y., Nielsen, R., Murray, D., Hubbell, W.L., Mailer, C., Robinson, B.H. & Gelb, M.H. (1998). Docking phospholipase A$_2$ on membranes using electrostatic potential-modulated spin relaxation magnetic resonance. *Science*, Vol. 279, No. 5358, pp. 1925–1929

Logonder, U., Jenko-Pražnikar, Z., Scott-Davey, T., Pungerčar, J., Križaj, I. & Harris, J.B. (2009). Ultrastructural evidence for the uptake of a neurotoxic snake venom phospholipase A$_2$ into mammalian motor nerve terminals. *Experimental Neurology*, Vol. 219, No. 2, pp. 591–594

Lomonte, B., Angulo, Y., Sasa, M. & Gutiérrez, J.M. (2009). The phospholipase A$_2$ homologues of snake venoms: Biological activities and their possible adaptive roles. *Protein and Peptide Letters*, Vol. 16, No. 8, pp. 860–876

Meldolesi, J. (2002). Rapidly exchanging Ca^{2+} stores: Ubiquitous partners of surface channels in neurons. *News in Physiological Sciences*, Vol. 17, pp. 144–149

Montecucco, C., Gutiérrez, J.M. & Lomonte, B. (2008). Cellular pathology induced by snake venom phospholipase A$_2$ myotoxins and neurotoxins: Common aspects of their mechanisms of action. *Cellular and Molecular Life Sciences*, Vol. 65, No. 18, pp. 2897–2912

Mounier, C.M., Ghomashchi, F., Lindsay, M.R., James, S., Singer, A.G., Parton, R.G. & Gelb, M.H. (2004). Arachidonic acid release from mammalian cells transfected with human groups IIA and X secreted phospholipase A$_2$ occurs predominantly during the secretory process and with the involvement of cytosolic phospholipase A$_2$-α. *Journal of Biological Chemistry*, Vol. 279, No. 24, pp. 25024–25038

Murakami, M., Taketomi, Y., Miki, Y., Sato, H., Hirabayashi, T. & Yamamoto, K. (2011). Recent progress in phospholipase A$_2$ research: From cells to animals to humans. *Progress in Lipid Research*, Vol. 50, No. 2, pp. 152–192

Okeley, N.M. & Gelb, M.H. (2004). A designed probe for acidic phospholipids reveals the unique enriched anionic character of the cytosolic face of the mammalian plasma membrane. *Journal of Biological Chemistry*, Vol. 279, No. 21, pp. 21833–21840

Pan, Y.H., Yu, B.-Z., Singer, A.G., Ghomashchi, F., Lambeau, G., Gelb, M.H., Jain, M.K. & Bahnson, B.J. (2002). Crystal structure of human group X secreted phospholipase A$_2$. Electrostatically neutral interfacial surface targets zwitterionic membranes. *Journal of Biological Chemistry*, Vol. 277, No. 32, pp. 29086–29093

Paoli, M., Rigoni, M., Koster, G., Rossetto, O., Montecucco, C. & Postle, A.D. (2009). Mass spectrometry analysis of the phospholipase A$_2$ activity of snake pre-synaptic neurotoxins in cultured neurons. *Journal of Neurochemistry*, Vol. 111, No. 3, pp. 737–744

Petan, T., Križaj, I., Gubenšek, F. & Pungerčar, J. (2002). Phenylalanine-24 in the N-terminal region of ammodytoxins is important for both enzymic activity and presynaptic toxicity. *Biochemical Journal*, Vol. 363, Pt. 2, pp. 353–358

Petan, T., Križaj, I., Gelb, M.H. & Pungerčar, J. (2005). Ammodytoxins, potent presynaptic neurotoxins, are also highly efficient phospholipase A$_2$ enzymes. *Biochemistry*, Vol. 44, No. 37, pp. 12535–12545

Petan, T., Križaj, I. & Pungerčar, J. (2007). Restoration of enzymatic activity in a Ser-49 phospholipase A_2 homologue decreases its Ca^{2+}-independent membrane-damaging activity and increases its toxicity. *Biochemistry*, Vol. 46, No. 44, pp. 12795–12809

Petrovič, U., Šribar, J., Pariš, A., Rupnik, M., Kržan, M., Vardjan, N., Gubenšek, F., Zorec, R. & Križaj, I. (2004). Ammodytoxin, a neurotoxic secreted phospholipase A_2, can act in the cytosol of the nerve cell. *Biochemical and Biophysical Research Communications*, Vol. 324, No. 3, pp. 981–985

Petrovič, U., Šribar, J., Matis, M., Anderluh, G., Peter-Kataliníc, J., Križaj, I. & Gubenšek, F. (2005). Ammodytoxin, a secretory phospholipase A_2, inhibits G2 cell-cycle arrest in the yeast *Saccharomyces cerevisiae*. *Biochemical Journal*, Vol. 391, Pt. 2, pp. 383–388

Pražnikar, Z.J., Kovačič, L., Rowan, E.G., Romih, R., Rusmini, P., Poletti, A., Križaj, I. & Pungerčar, J. (2008). A presynaptically toxic secreted phospholipase A_2 is internalized into motoneuron-like cells where it is rapidly translocated into the cytosol. *Biochimica et Biophysica Acta*, Vol. 1783, No. 6, pp. 1129–1139

Pražnikar, Z.J., Petan, T. & Pungerčar, J. (2009). A neurotoxic secretory phospholipase A_2 induces apoptosis in motoneuron-like cells. *Annals of the New York Academy of Sciences*, Vol. 1152, pp. 215–224

Prijatelj, P., Čopič, A., Križaj, I., Gubenšek, F. & Pungerčar, J. (2000). Charge reversal of ammodytoxin A, a phospholipase A_2-toxin, does not abolish its neurotoxicity. *Biochemical Journal*, Vol. 352, Pt. 2, pp. 251–255

Prijatelj, P., Križaj, I., Kralj, B., Gubenšek, F. & Pungerčar, J. (2002). The C-terminal region of ammodytoxins is important but not sufficient for neurotoxicity. *European Journal of Biochemistry*, Vol. 269, No. 23, pp. 5759–5764

Prijatelj, P., Šribar, J., Ivanovski, G., Križaj, I., Gubenšek, F. & Pungerčar, J. (2003). Identification of a novel binding site for calmodulin in ammodytoxin A, a neurotoxic group IIA phospholipase A_2. *European Journal of Biochemistry*, Vol. 270, No. 14, pp. 3018–3025

Prijatelj, P., Charnay, M., Ivanovski, G., Jenko, Z., Pungerčar, J., Križaj, I. & Faure, G. (2006a). The C-terminal and ß-wing regions of ammodytoxin A, a neurotoxic phospholipase A_2 from *Vipera ammodytes ammodytes*, are critical for binding to factor Xa and for anticoagulant effect. *Biochimie*, Vol. 88, No. 1, pp. 69–76

Prijatelj, P., Vardjan, N., Rowan, E.G., Križaj, I. & Pungerčar, J. (2006b). Binding to the high-affinity M-type receptor for secreted phospholipases A_2 is not obligatory for the presynaptic neurotoxicity of ammodytoxin A. *Biochimie*, Vol. 88, No. 10, pp. 1425–1433

Prijatelj, P., Jenko Pražnikar, Z., Petan, T., Križaj, I. & Pungerčar, J. (2008). Mapping the structural determinants of presynaptic neurotoxicity of snake venom phospholipases A_2. *Toxicon*, Vol. 51, No. 8, pp. 1520–1529

Pungerčar, J., Liang, N.S., Štrukelj, B. & Gubenšek, F. (1990). Nucleotide sequence of a cDNA encoding ammodytin L. *Nucleic Acids Research*, Vol. 18, No. 15, p. 4601

Pungerčar, J., Križaj, I., Liang, N.S. & Gubenšek, F. (1999). An aromatic, but not a basic, residue is involved in the toxicity of group-II phospholipase A_2 neurotoxins. *Biochemical Journal*, Vol. 341, Pt. 1, pp. 139–145

Pungerčar, J. & Križaj, I. (2007). Understanding the molecular mechanism underlying the presynaptic toxicity of secreted phospholipases A_2. *Toxicon*, Vol. 50, No. 7, pp. 871–892

Ramirez, F. & Jain, M.K. (1991). Phospholipase A_2 at the bilayer interface. *Proteins: Structure, Function, and Bioinformatics*, Vol. 9, No. 4, pp. 229–239

Renetseder, R., Brunie, S., Dijkstra, B.W., Drenth, J. & Sigler, P.B. (1985). A comparison of the crystal structures of phospholipase A_2 from bovine pancreas and *Crotalus atrox* venom. *Journal of Biological Chemistry*, Vol. 260, No. 21, pp. 11627–11634

Rigoni, M., Caccin, P., Gschmeissner, S., Koster, G., Postle, A.D., Rossetto, O., Schiavo, G. & Montecucco, C. (2005). Equivalent effects of snake PLA_2 neurotoxins and lysophospholipid-fatty acid mixtures. *Science*, Vol. 310, No. 5754, pp. 1678–1680

Rigoni, M., Paoli, M., Milanesi, E., Caccin, P., Rasola, A., Bernardi, P. & Montecucco, C. (2008). Snake phospholipase A_2 neurotoxins enter neurons, bind specifically to mitochondria, and open their transition pores. *Journal of Biological Chemistry*, Vol. 283, No. 49, pp. 34013–34020

Rosenberg, P. (1997) Pitfalls to avoid in the study of correlations between enzymatic activity and pharmacological properties of phospholipase A_2 enzymes, In: *Venom Phospholipase A_2 Enzymes: Structure, Function and Mechanism* (Kini, R. M., Ed.), pp. 155-183, John Wiley & Sons, Chichester, England.

Rouault, M., Rash, L.D., Escoubas, P., Boilard, E., Bollinger, J., Lomonte, B., Maurin, T., Guillaume, C., Canaan, S., Deregnaucourt, C., Schrével, J., Doglio, A., Gutiérrez, J.M., Lazdunski, M., Gelb, M.H. & Lambeau, G. (2006). Neurotoxicity and other pharmacological activities of the snake venom phospholipase A_2 OS2: The N-terminal region is more important than enzymatic activity. *Biochemistry*, Vol. 45, No. 18, pp. 5800–5816

Rufini, S., Cesaroni, P., Desideri, A., Farias, R., Gubenšek, F., Gutiérrez, J.M., Luly, P., Massoud, R., Morero, R. & Pedersen, J.Z. (1992). Calcium ion independent membrane leakage induced by phospholipase-like myotoxins. *Biochemistry*, Vol. 31, No. 49, pp. 12424–12430

Saul, F.A., Prijatelj-Žnidaršič, P., Vulliez-le Normand, B., Villette, B., Raynal, B., Pungerčar, J., Križaj, I. & Faure, G. (2010). Comparative structural studies of two natural isoforms of ammodytoxin, phospholipases A_2 from *Vipera ammodytes ammodytes* which differ in neurotoxicity and anticoagulant activity. *Journal of Structural Biology*, Vol. 169, No. 3, pp. 360–369

Scott, D.L., White, S.P., Otwinowski, Z., Yuan, W., Gelb, M.H. & Sigler, P.B. (1990). Interfacial catalysis: The mechanism of phospholipase A_2. *Science*, Vol. 250, No. 4987, pp. 1541–1546

Scott, D.L., Mandel, A.M., Sigler, P.B. & Honig, B. (1994). The electrostatic basis for the interfacial binding of secretory phospholipases A_2. *Biophysical Journal*, Vol. 67, No. 2, pp. 493–504

Singer, A.G., Ghomashchi, F., Le Calvez, C., Bollinger, J., Bezzine, S., Rouault, M., Sadilek, M., Nguyen, E., Lazdunski, M., Lambeau, G. & Gelb, M.H. (2002). Interfacial kinetic and binding properties of the complete set of human and mouse groups I, II, V, X, and XII secreted phospholipases A_2. *Journal of Biological Chemistry*, Vol. 277, No. 50, pp. 48535–48549

Snitko, Y., Koduri, R.S., Han, S.K., Othman, R., Baker, S.F., Molini, B.J., Wilton, D.C., Gelb, M.H. & Cho, W. (1997). Mapping the interfacial binding surface of human secretory group IIa phospholipase A_2. *Biochemistry*, Vol. 36, No. 47, pp. 14325–14333

Soares, A.M., Andrião-Escarso, S.H., Bortoleto, R.K., Rodrigues-Simioni, L., Arni, R.K., Ward, R.J., Gutiérrez, J.M. & Giglio, J.R. (2001). Dissociation of enzymatic and pharmacological properties of piratoxins-I and -III, two myotoxic phospholipases A_2 from *Bothrops pirajai* snake venom. *Archives of Biochemistry and Biophysics*, Vol. 387, No. 2, pp. 188–196

Šribar, J., Čopič, A., Pariš, A., Sherman, N.E., Gubenšek, F., Fox, J.W. & Križaj, I. (2001). A high affinity acceptor for phospholipase A_2 with neurotoxic activity is a calmodulin. *Journal of Biological Chemistry*, Vol. 276, No. 16, pp. 12493–12496

Šribar, J., Čopič, A., Poljšak-Prijatelj, M., Kuret, J., Logonder, U., Gubenšek, F. & Križaj, I. (2003a). R25 is an intracellular membrane receptor for a snake venom secretory phospholipase A_2. *FEBS Letters*, Vol. 553, No. 3, pp. 309–314

Šribar, J., Sherman, N.E., Prijatelj, P., Faure, G., Gubenšek, F., Fox, J.W., Aitken, A., Pungerčar, J. & Križaj, I. (2003b). The neurotoxic phospholipase A_2 associates, through a non-phosphorylated binding motif, with 14-3-3 protein γ and ε isoforms. *Biochemical and Biophysical Research Communications*, Vol. 302, No. 4, pp. 691–696

Šribar, J., Anderluh, G., Fox, J.W. & Križaj, I. (2005). Protein disulphide isomerase binds ammodytoxin strongly: Possible implications for toxin trafficking. *Biochemical and Biophysical Research Communications*, Vol. 329, No. 2, pp. 733–737

Stahelin, R.V. & Cho, W. (2001). Differential roles of ionic, aliphatic, and aromatic residues in membrane-protein interactions: A surface plasmon resonance study on phospholipases A_2. *Biochemistry*, Vol. 40, No. 15, pp. 4672–4678

Sumandea, M., Das, S., Sumandea, C. & Cho, W. (1999). Roles of aromatic residues in high interfacial activity of *Naja naja atra* phospholipase A_2. *Biochemistry*, Vol. 38, No. 49, pp. 16290–16297

Thouin, L.G., Ritonja, A., Gubenšek, F. & Russell, F.E. (1982). Neuromuscular and lethal effects of phospholipase A from *Vipera ammodytes* venom. *Toxicon*, Vol. 20, No. 6, pp. 1051–1058

Tzeng, M.C., Yen, C.H., Hseu, M.J., Dupureur, C.M. & Tsai, M.D. (1995). Conversion of bovine pancreatic phospholipase A_2 at a single site into a competitor of neurotoxic phospholipases A_2 by site-directed mutagenesis. *Journal of Biological Chemistry*, Vol. 270, No. 5, pp. 2120–2123

Valentin, E. & Lambeau, G. (2000). What can venom phospholipases A_2 tell us about the functional diversity of mammalian secreted phospholipases A_2? *Biochimie*, Vol. 82, No. 9–10, pp. 815–831

Vardjan, N., Sherman, N.E., Pungerčar, J., Fox, J.W., Gubenšek, F. & Križaj, I. (2001). High-molecular-mass receptors for ammodytoxin in pig are tissue-specific isoforms of M-type phospholipase A_2 receptor. *Biochemical and Biophysical Research Communications*, Vol. 289, No. 1, pp. 143–149

Vučemilo, N., Čopič, A., Gubenšek, F., & Križaj, I. (1998). Identification of a new high-affinity binding protein for neurotoxic phospholipases A_2. *Biochemical and Biophysical Research Communications*, Vol. 251, No. 1, pp. 209–212

Ward, R.J., Chioato, L., De Oliveira, A.H.C., Ruller, R. & Sá, J.M. (2002). Active-site mutagenesis of a Lys49-phospholipase A_2: Biological and membrane-disrupting activities in the absence of catalysis. *Biochemical Journal*, Vol. 362, Pt. 1, pp. 89–96

Yau, W.M., Wimley, W.C., Gawrisch, K. & White, S.H. (1998). The preference of tryptophan for membrane interfaces. *Biochemistry*, Vol. 37, No. 42, pp. 14713–14718

Yu, B.Z., Berg, O.G. & Jain, M.K. (1993). The divalent cation is obligatory for the binding of ligands to the catalytic site of secreted phospholipase A_2. *Biochemistry*, Vol. 32, No. 25, pp. 6485–6492

Site-Directed Mutagenesis in the Research of Protein Kinases - The Case of Protein Kinase CK2

Ewa Sajnaga, Ryszard Szyszka and Konrad Kubiński
Department of Molecular Biology, Institute of Biotechnology,
The John Paul II Catholic University of Lublin,
Poland

1. Introduction

Protein kinases constitute one of the largest and best-explored superfamilies of mammalian genes. The human genome encodes approximately 518 protein kinase genes, and the majority of their proteins have been characterized to some extent. Many of protein kinases have been implicated in signal transduction pathways that regulate growth and survival of cells, indicating their potential role in cancer (Manning et al., 2002a). Indeed, most cancers are associated with disregulation of protein kinases or by loss or damage of cellular protein kinase inhibitors (Brognard & Hunter, 2011; Johnson, 2009; Pearson & Fabbro, 2004).

A typical protein kinase (based on 478 known enzymes) is characterized by the presence of a highly homologous kinase catalytic domain of about 300 amino acid residues. This domain serves three distinct roles: (a) binding and orientation of the phosphate donor (ATP or, rarely, a different nucleoside triphosphate) in a complex with a divalent cation (usually Mg^{2+} or Mn^{2+}); (b) binding and orientation of the protein substrate; and (c) transfer of the γ-phosphate from the NTP to the hydroxyl moiety of the acceptor residue of the protein substrate.

In response to a variety of regulatory signals, protein kinases phosphorylate specific serine, threonine, or tyrosine residues within target proteins to modify their biological activities. Phosphorylation of several proteins in the cell creates recognition and/or regulatory sites that influence many properties of the target proteins (*e.g.*, catalytic activity, localization, sensitivity to proteolytic degradation, protein-protein interaction, etc). In eukaryotes, Ser/Thr and Tyr protein kinases (Note: The single- and three-letter codes for the amino acids are given in Table 1 in the chapter by Figurski et al.) play a key role in molecular networks controlling the activity of various signaling proteins (Brognard & Hunter, 2011; Cohen, 2002). Ser/Thr and Tyr-protein kinases form the largest protein family in the human genome (Pandit et al., 2004). They constitute about 2-3% of the proteomes of other model organisms, such as *Saccharomyces cerevisiae*, *Caenorhabditis elegans* and *Drosophila melanogaster* (Manning et al., 2002 a, b; Goldberg et al., 2006; Plowman et al., 1999). Based on the conserved features of catalytic domains of eukaryotic protein kinases, Hanks and coworkers have placed the kinases into various classes, groups, and subfamilies (Hanks et al., 1988).

The first 3D structure of a protein kinase was determined for PKA by X-ray crystallography. It revealed the basic bi-lobed scaffold formed by N- and C-terminal lobes that has been observed in all the protein kinase structures solved to date. The N-terminal lobe of the kinase fold comprises of an anti-parallel β-sheet made of five β-strands (β1 – β5) and a single αC-helix. The C-terminal lobe is larger and is mainly composed of α-helices. The nucleotide- and substrate-binding pockets are located in the cleft between the two lobes. The phosphate groups of ATP are positioned for phosphotransfer by their interactions with conserved residues in the N- and C-terminal lobes. These include a glycine-rich loop characterized by the GXGXXG motif (where X represents any amino acid) between the β1 and β2 strands, a Lys residue localized by a salt bridge formed by a Lys-Glu pair (K72 and E91), and Mg^{2+} ions. The conserved Asn (N171) and Asp (D184) further coordinate the metal ions. The catalytic loop situated in the C-terminal lobe contains aspartate (D166), referred to as the catalytic base that facilitates extraction of a proton from the hydroxyl side-chains of the phospho-sites of the substrates. The activation segment (20–30 residues in length) caps the C-terminal lobe. This segment forms a part of the substrate-binding pocket and shows high structural variation in the active and inactive kinase structures.

The grouping of the protein kinases based on catalytic subunit sequence similarity results in clustering of kinases that share functional features, such as preferred sites of phosphorylation, the mode of regulation and cellular localization. The similarity in the amino acid sequence of the catalytic domains of protein kinases has proven to be a good indicator of other features held in common by the different members of the family.

The diversity of essential functions mediated by kinases is shown by the conservation of approximately 50 distinct kinase families that have been identified in yeast, invertebrates, and mammals. Protein kinases can be clustered into groups, families, and sub-families. These classifications are based on sequence similarity and biochemical function. Among the 518 human protein kinases, 478 belong to a single superfamily whose catalytic domains are related in sequence.

Protein kinases are divided into 10 main groups, which organize diversity and compare genes between distant organisms (Miranda-Saavedra & Barton, 2007). The groups are named as follows: AGC, CAMK, CK1, CMGC, STE, RGC, Other, TK, TK, and Atypical. This classification was first used for characterization of the human kinome (all the kinases encoded in the genome) (Manning et al., 2002) and is based on an earlier classification by Hanks and Hunter (1995).

1.1 The CMGC kinases

This group of Ser/Thr protein kinases was named after the initials of some members (CDK, MAPK, GSK3, and CLK). It includes key kinases involved in growth, stress-response, and the cell cycle, and kinases involved in splicing and metabolic control. The four well characterized subfamilies of this group include the following: cyclin-dependent kinases (CDK) (Liolli, 2010); mitogen-activated protein kinases (MAPK) (Biondi & Nebreda, 2003; Zhang & Dong, 2007); glycogen synthase kinases (GSK) (Biondi & Nebreda, 2003); and cell kinases 2 (CK2), better known as protein kinase CK2 or casein kinase 2 (St-Denis & Litchfield, 2009). The CMGC group also contains other members. These are the following: SR protein kinases [phosphorylating serine- and arginine-rich proteins engaged in regulation of splicing and nuclear transport (Ghosh & Adams, 2011)] and DYRK protein

kinases, *i.e.* dual-specificity tyrosine-phosphorylated protein kinases, presumably involved in brain development (Becker et al., 1998).

1.2 Protein kinase CK2

Protein kinase CK2, formerly called casein kinase II, is a ubiquitous second messenger-independent protein kinase found in all eukaryotic organisms examined (Jensen et al., 2007; Niefind et al., 2009; Litchfield, 2003, Kubiński et al., 2007). This enzyme, which has been studied for over 50 years, is able to phosphorylate more than 300 substrates, on serine, threonine and tyrosine (Meggio & Pinna, 2003; Vilk et al., 2008). As the list of targets for CK2 continues to grow, it is becoming evident that CK2 has the potential to participate in the regulation of various cellular processes. Most of the CK2 substrates reported so far correspond to proteins that participate in cell signaling (Ahmed et al., 2002; Meggio & Pinna, 2003).
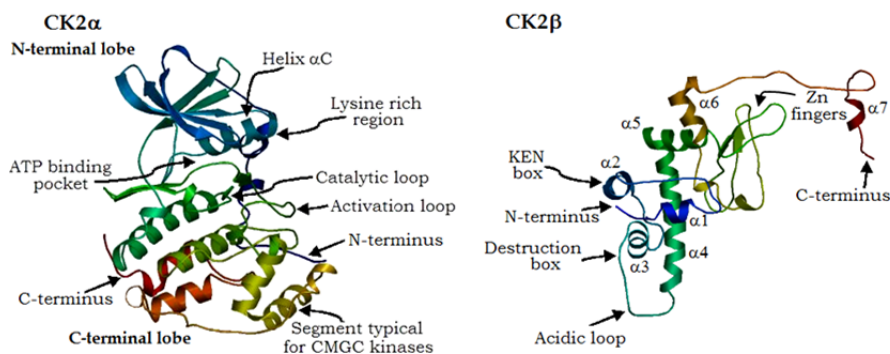
Fig. 1. Model structures of the CK2α and CK2β subunits from *Mytilus galloprovincialis* (Mediterranean mussel) (Koyanou-Koutsokou et al., 2011b)
The structural features of the CK2a and CK2β subunits were elaborated using the SWISS-MODEL Workspace for protein structure homology modeling (Arnold et al., 2006; Kopp and Schwede 2004) and 1ds5D (Batistuta et al., 2000) or 3EED (Raaf et al., 2008) as templates, respectively.

Protein kinase CK2 is distributed ubiquitously in eukaryotic organisms, where it appears as a tetrameric complex composed of two catalytic subunits (α/α') associated with a dimer of regulatory β subunits (Figs. 1 & 2). The CK2 tetramer exhibits constitutive activity that can be easily detected in most cellular or tissue extracts in the absence of any stimulatory compounds. In many organisms, distinct isoenzymic forms of the catalytic subunit of CK2 have been identified (Glover, 1998; Kolaiti et al., 2011; Kouyanou-Koutsoukou et al., 2011a, b; Maridor et al., 1991; Litchfield et al., 1990; Shi et al., 2001). In humans, only a single regulatory CK2β subunit has been identified; but multiple forms of CK2β have been identified in other organisms, such as *Saccharomyces cerevisiae* (Glover, 1998). Complementary evidence indicates that dimers of CK2β are at the core of the tetrameric CK2 complexes (Graham & Litchfield, 2000; Pinna & Meggio, 1997). Tetrameric CK2 complexes may contain identical (*i.e.*, $\alpha_2\beta_2$ or $\alpha'_2\beta_2$) or non-identical (*i.e.*, $\alpha\alpha'\beta_2$) catalytic subunits (Gietz et al., 1995). Holoenzyme composition may influence CK2 properties, namely nucleotide and protein substrate specificity and sensitivity to effectors (Janeczko et

al., 2011). Protein kinase CK2 holoenzyme and its catalytic subunit alone can use both ATP and GTP as phosphate donors (Issinger, 1993).

The catalytic subunits of CK2α and CK2α′ are the products of separate genes located in different chromosomes. The 330 N-terminal amino acids exhibit over 90% sequence identity. However, the C-terminal sequences are unrelated (Olsten & Litchfield, 2004). The unique C-terminal domains of the catalytic subunits are highly conserved among species (*e.g.*, the amino acid sequences of the C-termini of the catalytic subunits of human and chicken CK2α and CK2α′ exhibit 98% and 97% identity, respectively), indicating a possible functional importance for this domain (Litchfield, 2003).

Although there is no known difference between the catalytic activities of CK2α and CK2α′, there is evidence that they exhibit functional specialization (Duncan & Lichfield, 2008; Faust & Montenarch, 2000). The CK2α subunit is phosphorylated at C-terminal sites (Thr344, Thr360, Ser362 and Ser360) by p34[cdc2] during cell cycle progression, while CK2α′ is not phosphorylated (St-Denis et al., 2009). Further evidence to support the idea that CK2α and CK2α′ have independent functions in the cell is provided by the different specificities of cellular binding proteins, such as CKIP-1, Hsp90, Pin-1, and PP2A (Olsten et al., 2005).

Despite the many isoforms of catalytic subunits, only one regulatory subunit has been identified for CK2β in mammals (Allende and Allende, 1995). In contrast to the activity of regulatory subunits of other kinases, such as PKA (cAMP-dependent protein kinase) and CDK (cyclin-dependent protein kinase), CK2β does not switch on or off the intrinsic activity of the catalytic subunits (Bolanos-Garcia et al., 2006).

The CK2β regulatory subunit is remarkably conserved among species, but it does not have homology with the regulatory subunits of other protein kinases (Bibby & Litchfield, 2005). The amino acid sequence of the CK2β regulatory subunit is almost identical in *Homo sapiens*, *Drosophila melanogaster*, *Ceratitis capitata* (Mediterranean fruit fly), *Danio rerio* (zebrafish), *Ciona intestinalis* (sea squirt), and *Mytilus galloprovincialis* (Mediterranean mussel) (Kouyanou-Koutsoukou et al., 2011a, b; Kolaiti et al., 2011). It is completely identical in birds and mammals (Maridor et al., 1991; Wirkner et al., 1994). In contrast, the fruit fly *D. melanogaster* has four regulatory subunit genes. They are used for one CK2α (DmCK2α) and three CK2βs (DmCK2β, DmCK2β' and DmCK2βtes) (Jauch et al., 2002). *Zea mays* has three isoforms of the catalytic α-subunit (CK2a-1, CK2a-2 and CK2a-3) and three regulatory β-subunits (CK2b-1, CK2b-2 and CK2b-3) (Riera et al., 2001). *S. cerevisiae* CK2 holoenzyme contains two regulatory β-subunits (β and β′). They cannot substitute for each other, and both of them are needed to form a fully active enzymatic unit (Kubinski et al., 2007).

Results presented by several groups and obtained by the use of a variety of approaches, including X-ray crystallography, have determined that a dimer of the CK2β subunits forms the core of the CK2 tetramer (Chantalat et al., 1999; Sarno et al., 2000; Canton et al., 2001).

The CK2β regulatory subunit is a compact, globular homodimer that shows high amino acid sequence conservation across species. The N-terminal domain (amino acids 1-104) is globular and contains four α-helices (marked as α1-α4 in Fig. 1). Helices α1 (residues 9–14), α2 (residues 27–31) and α3 (residues 46–54) wrap around α4 (residues 66–89) (Bolanos-Garcia et al., 2006). This part of the protein contains autophosphorylation sites, consisting of serines 2, 3, and possibly 4 (Boldyreff et al., 1993). Studies conducted by Zhang and coworkers (2002) indicate that phosphorylation of these sites enhances CK2β stability. The

first 20 N-terminal amino acids of the CK2β regulatory subunit are also involved in the interaction with Nopp140, a protein that binds a nuclear localization sequence and shuttles between the nucleus and the cytoplasm (Li et al., 1997). This part of the protein also contains two motifs that have been previously characterized as motifs that regulate cyclin degradation. The CK2β regulatory subunit has a sequence resembling the nine-amino-acid motif called the destruction box, which plays a key role in the specific degradation of cyclin B at the end of mitosis (King et al., 1996). This motif, located in helix α3, contains three highly conserved residues that conform to the general destruction box consensus (RXXLXXXXN/D) (Bolanos-Garcia et al., 2006). Interestingly, this motif is located on a surface-exposed α3 helix, where it would be available for recognition by the cellular degradation machinery. A signal known as the KEN box, which was found previously in mitotic cyclins and which has been shown to play a role in mediating cell cycle-dependent protein degradation, is also present in CK2β. This degradation motif is characterized by the minimal consensus sequence KEN, but it is often followed shortly by either an N or D residue and is often preceded by another N or D residue. A similar sequence (D_{32}KFNLTGLN$_{40}$) forms helix α2 of the CK2β protein (Bibby & Litchfield, 2005).

The N-terminal part of the CK2β also contains an "acidic loop" between helices α3 and α4. This acidic, surface-exposed region of the protein, encoded by residues 55-64, has been identified as the site on CK2 that binds polyamines, which are known to stimulate CK2 activity *in vitro* (Meggio et al., 1994; Leroy et al., 1997).

The analysis of the CK2β regulatory subunit structure by X-ray crystallography revealed the importance of the zinc finger in CK2β regulatory subunit dimerization (Chantal et al., 1999). The zinc-finger region is characterized by four conserved cysteine residues (residues 109, 114, 137 and 140), which mediate the interaction that allows the CK2β dimer to form the core of the CK2 holoenzyme (Chantal et al., 1999; Canton et al., 2001).

The C-terminal part of the CK2β regulatory subunit (residues 178–205) contains a large loop (residues 178–193) and helix α7 (residues 194–200). Although helix α7 is located away from helices α1-α6, the C-terminal amino acids (190-205) contribute to the formation of the CK2β regulatory subunit dimer (Niefind et al., 2001). This part of the regulatory subunit contains two phosphorylation sites: Thr213, which is phosphorylated by the checkpoint kinase Chk1 (Kristensen et al., 2004) and Ser209, which is phosphorylated *in vitro* and in mammalian cells by p34[cdc2] in a cell-cycle-dependent manner (Litchfield et al., 1995).

The traditional view of the CK2β regulatory subunit is that it functions as a component of tetrameric CK2 complexes and that it is the regulator of the catalytic CK2α and CK2α' subunits, enhancing their stability, specificity and activity. As an example, the CK2β regulatory subunit stimulates CK2 holoenzyme activity towards certain protein substrates, such as topoisomerase II (Leroy et al., 1999), and inhibits others, like calmodulin (Marin et al., 1999).

It was shown that CK2β does not exist exclusively within stable CK2 complexes. This observation raises the prospect that CK2β has functions that are independent of its role as the regulatory subunit of CK2. For example, overexpression of CK2β in the fission yeast *Schizosaccharomyces pombe* revealed severe growth defects and a multiseptated phenotype, whereas CK2α overexpression had no effect (Roussou & Draetta, 1994).

CK2β seems to interact directly with more than 40 different proteins, including other protein kinases such as A-Raf, Chk1, Chk2, PKC-ζ, Mos and p90[rsk] (Bibby & Lichfield, 2005; Bolanos-Garcia et al., 2006; Olsen & Guerra, 2008). It was shown that association of the human protein kinases Chk1, Mos, and A-Raf is mediated by the C-terminal region of the CK2β subunit and that these associations involve some residues that interact with the catalytic CK2α subunit (Chen et al., 1997; Lieberman & Ruderman, 2004; Olsen & Guerra, 2008). The interaction between Chk1 and CK2β leads to an increase in the Cdc25C phosphorylation activity of Chk1. Screening of several cell lines has shown that the association between CK2β and Chk1 is also formed *in vivo* (Guerra at al., 2003).

Overexpression of CK2 has been linked to several pathological conditions, ranging from cardiovascular pathologies and cancer progression to neurodegenerative disorders (*e.g.,* Alzheimer's disease, Parkinson's disease, brain ischemia) and infectious diseases (Guerra & Issinger, 2008; Ahmad et al., 2008; Trembley et al., 2009). Various specific, potent small molecule inhibitors of protein kinase CK2 have been developed in recent years, including condensed polyphenolic compounds, tetrabromobenzimidazole/triazole derivatives, and indoloquinazolines (Gianoncelli et al., 2009; Pagano et al., 2008; Raaf et al., 2008). Inhibition of CK2 kinase activity by these compounds display a remarkable pro-apoptotic efficacy on a number of tumor-derived cell lines, indicating a possibility of developing novel antineoplastic drugs (Batistuta, 2009; Duncan et al., 2010; Prudent et al., 2010; Unger et al., 2004).

2. Mutagenesis in studies on protein kinase CK2

Within the last 2 decades, a number of studies have produced mutants of both CK2α and CK2β that provide a valuable, yet incomplete, basis to rationalize the biochemical features of the enzyme, i.e., its constitutive activity, dual-cosubstrate specificity, acidophilic substrate specificity and tetrameric structure (Fig. 2).

2.1 Mutagenesis of the CK2α catalytic subunit

2.1.1 Mutations of CK2α in the regions responsible for constitutive activity

A majority of protein kinases need to be activated. Phosphorylation within the kinase activation loop is the most popular mode of activation. In contrast to other known protein kinases, CK2 has constitutive activity and does not demand activation. In this case, activation is achieved by the interaction between the N-terminal tail and the activation loop in the kinase domain. The role of the N-terminal segment in stable opening of the activation loop was confirmed in mutagenesis studies (Sarno et al., 2001). In particular, the Δ2-12 CK2α mutant, in comparison with the wild-type kinase, displayed an almost complete loss of activity, which was reflected by increased Km values for ATP and the peptide substrate (from 10 to 206 μM and from 26 to 140 μM, respectively). Further experiments revealed that holoenzyme reconstitution restored the activity of the mutant to the wild-type level. This demonstrates an alternative CK2β subunit-dependent mechanism to provide constitutive activity in the case of CK2 holoenzyme (Sarno et al., 2002).

Recently, molecular dynamics (MD) simulation has been carried out in order to explore the role of the CK2α N-terminal segment in the conformational behavior of the kinase (Cristiani
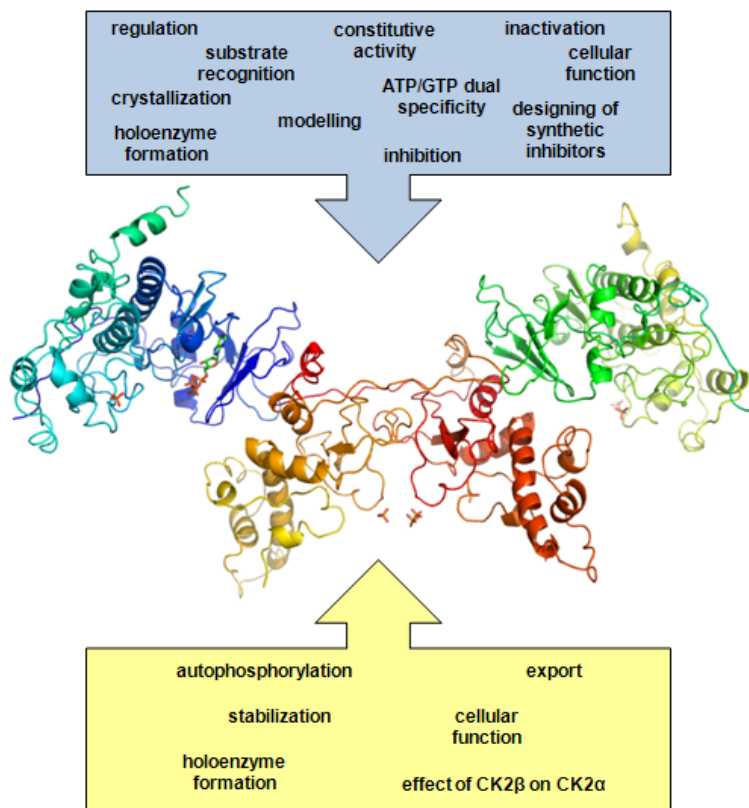
Fig. 2. Multiple applications of mutagenesis in studies on CK2.
The blue box presents various aspects of research using mutagenesis on CK2α; and the yellow box, on CK2β. The model of the human CK2 holoenzyme was developed using the PyMOL software based on the structure of the human CK2 holoenzyme (PDB code 1JWH) from the Protein Data Bank. The catalytic α subunits are presented in blue and green; the regulatory β subunits are in red and yellow.

et al., 2011). Comparison of the αC-helix RMSD (root mean square deviation) values obtained for the Δ2-12 CK2α mutant (*i.e.*, deleted for residues 2 through 12) and the wild-type kinase models show an increase in this parameter for the mutant form of the enzyme. This effect is due to instability of the CK2α conformation in the case of absence of an N-terminal segment and its interaction with the αC-helix. These results are consistent with the data presented by Sarno and collaborators, and they indicate that the complete N-terminal segment is essential for proper conformation and constitutive activity of protein kinase CK2α (Cristiani et al., 2011).

The experiment presented above is an example of the validation of *in vitro* mutagenesis studies with the use of computing analysis, but the opposite direction of studies is also possible. Two CK2α mutants, the triple mutant Y206F/R10A/Y261F and the single mutant Y125F, were constructed *in silico*. MD simulations were then carried out to study the relation

between CK2 conformation and activity (Cristiani et al., 2011). The amino acids substituted in the first virtual mutant are engaged in the most important bonds between the N-terminal segment and other regions of CK2α to maintain kinase activity. The CK2α Y125F mutant is also very useful in studying the influence of Tyr125 on the conformational change of Phe121. According to Niefind and Issinger (2010), Phe121 can assume two different conformations: in and out, which regulate the activity of CK2α. Preliminary MD simulations on the two protein mutant models are very promising. The authors are currently working on the construction of both CK2 mutants. Biochemical characterization of the mutants will be carried out (Cristiani et al., 2011).

2.1.2 Mutation of CK2α in the basic regions

Protein kinase CK2 is characterized by its special aptitude to interact with negatively charged ligands. This ability correlates with the presence of several basic residues in CK2α that are not conserved in a majority of other protein kinases. These residues are located mainly in the "Lys-rich segment" and in the "p+1 loop." The Lys-rich segment ($K_{74}KKKIKR_{80}$) at the beginning of the αC-helix is a distinctive feature of CK2α (Tuazon & Traugh, 1991; Guerra et al., 1999). Results from mutational studies support the notion that this cluster is involved in substrate recognition, inhibition by heparin, down-regulation by the CK2β subunit and interaction with heat shock protein 90, and nuclear targeting (Guerra et al., 1999; Pinna & Meggio, 1997) (Table 1). CK2α mutants from *Caenorhabditis elegans* and *Xenopus laevis* (K74E/K75E and K75E/K76E, respectively) had lysines replaced by glutamic acid residues, which greatly affected the charge of this region in both mutant enzymes. The changes produced neither a significant increase in the K_m of the CK2α subunit for the casein and model peptide substrates nor changes in the affinity of the mutated CK2α subunit for the CK2β subunit during assembling a fully competent CK2 holoenzyme**.** The same mutations, however, had a significant effect on the affinity of CK2α for heparin and for other polyanionic inhibitors (Hu & Rubin, 1990; Gatica at al., 1994). Complete suppression of heparin inhibition was observed with the quadruple mutated K74-77A CK2α used by Vaglio and collaborators (1996). These authors showed (1) that all the four basic residues at positions 74, 75, 76, and 77 are implicated in heparin binding and (2) that the mutation of all of them was necessary to minimize heparin inhibition. Further mutagenesis studies showed that the additional basic residues cooperated with high heparin binding (apart from the 74-77 quartet). These were mainly Arg191, Arg195 and Lys198 located in the p+1 loop. However, the triple mutant for the three non-Lys-rich segment residues was less effective in heparin inhibition than was the mutant resulting from quadruple mutation of the 74-77 cluster (Vaglio et al., 1996). The triple mutant in which Lys79, Arg80 and Arg83 were changed into alanines did not alter the IC_{50} (concentration needed to give 50% inhibition) value for heparin. However, the mutant did show a reduction in the phosphorylation efficiency of the peptide substrate (and derivatives in which individual aspartyl residues were replaced by alanines). Because of these properties, it was specified that the basic residues in positions 77-83 are mainly involved in substrate recognition, rather than in heparin inhibition (Sarno et al., 1995; Vaglio et al., 1996). These authors concluded that the highly conserved 74-80 basic stretch is composed of two functionally distinct entities: (1) an N-terminal moiety mostly involved in heparin inhibition as well as in down-regulation by the β subunit and (2) the C-terminal part implicated in recognition of the crucial specificity determinant at positions n+3, but irrelevant to heparin.

Extended mutagenesis analysis combined with biochemical characterization provided clear evidence that residues responsible for both substrate recognition and down-regulation of CK2α catalytic activity are located mainly in the Lys-rich loop and p+1 loop spanning sequences 74-83 and 191-198, respectively. This corroborates the concept that the CK2β subunit down-regulates the CK2β by acting as a pseudosubstrate (Meggio et al., 1994; Sarno et al., 1996, 1997a, 1999).

Sarno and collaborators (1997b) analyzed the relative contribution of basic residues, presumably implicated in CK2-substrate interaction, in the recognition of peptide substrates varying in the number and position of acidic determinants. Sixteen derivatives of the optimal peptide substrate RRRA-DDSDDDDD, wild-type CK2 and twelve CK2α mutants defective in substrate recognition were used in the experiments. In the CK2α mutants, different basic residues implicated in substrate recognition were replaced by alanine (*e.g.*, K49A, K74-77A, or K79A/R80A/K83A). The results obtained support the idea that the acidic residues at positions n+1 and n+3 are essential, while additional acidic residues are required for efficient phosphorylation of CK2 substrates. Kinetic analysis with CK2α mutants revealed that Lys48 was implicated in the recognition of the determinant at position n+2. Lys77 interacts with the determinants at n+3 and n+4, while Lys198 recognized the determinant at n+1 (Sarno et al., 1997b). Molecular modeling based on crystallographic data supported these observations. It showed that several of these basic residues are clustered around the active site, where they make contact with individual acidic residues of the peptide substrate, polyanionic inhibitors, regulatory elements present in the β subunit, N-terminal segment of the CK2α, and possibly other proteins interacting with CK2 (Sarno et al., 1999).

2.1.3 Mutations of CK2α in catalytic subdomains

Subdomains II and VII of CK2α involved in nucleotide binding and phosphotransfer are in close proximity to each other in the three-dimensional structure. CK2α differs from more than 95% of other known protein kinases in having Val66 instead of the corresponding alanine within conserved region II and Trp176 instead of the corresponding phenylalanine within region VII (Allende & Allende, 1995). To investigate whether these variant amino acid residues might be responsible for effective GTP utilization, Jakobi and Traugh (1995) mutated both of these residues back to the consensus amino acids. Their results indicated that both single mutants of CK2α and the double mutant CK2α could still use GTP as a phosphate donor. The single and double mutations only altered the relative affinities for ATP and GTP. This finding indicated that at least one other amino acid residue must be responsible for the effective utilization of GTP by CK2. The same authors studied the above-mentioned mutants with respect to the catalytic activity of the reconstructed holoenzyme. The relatively lower affinity for GTP of the holenzyme reconstructed from the mutated CK2α was caused by changes in both the K_m and V_{max} values for GTP and ATP, while for the catalytic subunits, it was a result of changes in the K_m values only. These studies showed that the unique property of the effective utilization of GTP by CK2 was correlated with stimulation of the activity by the regulatory subunits and with the ability to undergo a conformational change upon formation of the holoenzyme.

Srinivasan and collaborators (1999) showed that the dual specificity of CK2 probably originated from the loop situated around the stretch $H_{115}VNNTD_{120}$ in CKα. In their work, they combined site-directed mutagenesis of CK2α with comparative 3D-structure modeling.

Due to significant amino acid sequence similarity (69,5%), kinase CDK2 was chosen to be a good comparative model for CK2α. Based on modeling, a ΔN118 CK2α mutant was constructed. The kinase assay showed decreased affinity of this protein to GTP, in comparison to the wild-type CK2α. The K_m values were 146 and 37 µM, respectively. The results obtained clearly indicate that the adenine/guanine binding region (His115–Asp120) is responsible for the dual specificity of kinase towards phosphate donors (Srinivasan et al., 1999).

The latter study was extended by Jakob and collaborators (2000), who created several mutants of *Xenopus laevis* CK2α with substitutions at positions 118 and 129. They tested them for cosubstrate specificity after their combination with CK2β. The region containing Asn118, known to participate in the recognition of the guanine base, is a part of the sequence $N_{117}NTD_{120}$. This sequence closely resembles the conserved sequence NKXD that is present in G proteins and other GTPases. The study demonstrated that both the CK2α ΔN118 and CK2α N118E mutants produced a 5 to 6-fold increase in the K_m for GTP with little effect on the affinity for ATP.

The mutagenesis by Yde and collaborators (2005) resulted in the first stable and fully active mutant of the human catalytic subunit of protein kinase CK2 that is devoid of dual cosubstrate specificity. The resulting mutant hsCK2α1-335 (human CK2 deleted for the last 56 amino acids) V66A/M163L was designed on the basis of several structures of the enzyme from *Zea mays* in a complex with various ATP-competitive ligands. As structural research revealed the existence of a purine base-binding plane harboring the purine base of ATP and GTP. This plane is flanked in human CK2α by two side-chains of Val66 and Met163, and it adopts a significantly different orientation than it does in other kinase homologues. By exchanging these two flanking amino acids, the cosubstrate specificity is shifted towards strongly favoring ATP. These findings demonstrated that CK2α possesses a sophisticated structural adaptation that favors dual-cosubstrate specificity, a property that may have biological significance.

The mutagenesis studies also provided much insight into the significance of the sequence of the catalytic domain with respect to the CK2α/CK2β interaction. It was reported that CK2α V66A and V66A/W176F were able to interact with CK2β, but this interaction failed to stimulate catalytic activity on the peptide substrate. These results were in contrast to the result with the wild-type α subunit, which was stimulated 4-fold. Nevertheless, the stimulatory response to the cationic modulatory compounds, spermine and polylysine, was the same for holoenzymes reconstituted with the wild-type subunit and all three above-mentioned mutants of the α subunit. The results showed that there must be at least two different interactions between the catalytic α and regulatory β subunit: one that is responsible for stimulation by the β subunit itself and another for mediating the stimulation by polycationic compounds (Jakobi & Traugh, 1992). However, experiments using calmodulin as a substrate for phosphorylation revealed that the insensitivity of the CK2α mutant V66A to CK2β was only apparent. Down-regulation of calmodulin phosphorylation by the CK2β subunit is even enhanced by the V66A mutant. This observation indicated a possible indirect role for Val66 in conferring to the α-subunit a conformation less sensitive to down-regulation (Sarno et al. 1997a).

It is known that the hydrophobic and polar residues of domain II and VII are responsible for the selectivity of a number of specific, potent CK2 ATP-competitive inhibitors, like TBBz

(tetrabromobenzimidazole) and TBBt (tetrabromobenzotriazole) (Sarno et al., 2005a). The importance of the same key residues in the hydrophobic portion of the binding site was corroborated by mutational analysis of residues of the human CK2α. Their side chains contribute to the reduction in the internal size of the hydrophobic pocket adjacent to the ATP/GTP-binding site in CK2 (Battistutta et al., 2001; Sarno et al, 2005). Three of these residues (Val66 or Ile66, Ile174, and Met163) are specific to CK2. They are generally replaced by smaller ones in other protein kinases. Both single and double mutants with substitutions for Val66 and Ile174 gave rise to catalytically active CK2α with altered susceptibility to various inhibitors. However, replacement of Met163 by glycine produced a catalytically inactive mutant (Sarno et al., 2005b). Similar data were obtained with yeast CK2α. Mutants with alterations to V67 and I213 (analogous to V66 and I174 of human CK2α) displayed considerably higher K_i values toward inhibitors TBBz and TBBt and only a slight change in the affinity for ATP (Sajnaga et al, 2008). The structural basis for decreased emodin binding to human CK2α resulting from a single point mutation (V66A) has been examined by molecular dynamics (MD) simulations and energy analysis (Zhang & Zhong, 2010). It was found that the V66A mutation resulted in a packing defect due to a change in hydrophobicity. It led to abnormal behavior of the glycine-rich loop, α-helix, and C-loop. The critical role of Ile66 in cosubstrate binding and selection, besides forcing the nucleotide ligands to adopt different positions in the binding pocket, was also demonstrated in a mutational study (Jakobi et al., 1994; Jakobi & Traugh, 1992, 1995).

Chaillot and collaborators (2000) studied the role of Gly177 in conserved region VII of the catalytic domain, which is close to the active site. It was revealed that the CK2α G177K mutant exhibited improved catalytic efficiency for acid peptidic substrates, probably by establishing interactions with the acidic residues.

The acidic residue Asp or Glu of the catalytic loop (corresponding to Glu170 in PKA and conserved in most Ser/Thr protein kinases) is responsible for the binding of basic residues that specify the protein/peptide substrates. In CK2, the residue is replaced by a histidine (His160). Such a substitution could explain the acidophilic properties of CK2, in contrast to the basophilic properties of PKA and other Ser/Thr kinases. The actual role of the His160 in the determination of the site specificity of CK2 was assessed by Dobrowolska and collaborators (1994). Interestingly, subsequent mutational studies in which His160 was replaced with alanine or aspartic acid ruled out any significant role of this residue in substrate recognition (Sarno et al., 1997b).

A CK2α inactive mutant (D156A) was produced based on structural homology to kinase PKA. The mutant protein was able to compete efficiently with the wild-type CK2α for the regulatory β subunits. Although it does not exhibit kinase activity, the D156A mutant can bind CK2β to form an inactive holoenzyme. Moreover, the mutant abolishes the inhibitory effect of CK2β on CK2α-mediated phosphorylation of calmodulin. These results suggest that CK2α D156A may be a useful dominant-negative mutant for elucidation of the cellular functions of the CK2 regulatory subunit (Cosmelli et al., 1997).

2.1.4 Mutations of CK2α in the glycine-rich loop

The glycine-rich sequence (G-loop) is one of the most critical structures of protein kinases, since it contributes in many ways to enzyme activity. This multifunctional structural

element participates in nucleotide binding, substrate recognition, catalysis, and regulation of activity (Bossemeyer et al., 1994). In their extensive mutational studies combined with biochemical characterization, Sarno and collaborators (1999) confirmed that some basic residues in the glycine-rich loop of the CK2α, particularly Lys49, are implicated in substrate recognition and inhibition by polyanions. Another residue located within this region, Gly48, is involved in binding the ATP phosphate moiety. Replacement of Gly48 by alanine in CK2α affected its catalytic efficiency and specificity. It is thought that alanine causes this phenotype by creating an electrostatic barrier between ATP and the peptide substrate (Chaillot et al., 2000).

2.1.5 Mutations of CK2α in the C-terminal region

The C-terminal region of vertebrate CK2α is composed of 54 amino acids. Knowledge of this segment is rather poor, except for phosphorylation by kinase p34[Cdc2] and interaction with isomerase Pin1 (Bosc et al., 1995; Messenger et al., 2002). It is known from the publications on crystallization of CK2 that the catalytic subunits are particularly sensitive to degradation, which makes the crystallization process of the entire subunit difficult (Niefind et al., 2000, 2001). Truncation at the C-terminus reduced the intrinsic degradability of CK2α and allowed its crystallization and the determination of its 3D structure. Starting from sequence alignments of C-termini from different CK2αs, Grasselli and collaborators (2004) constructed a mutant carrying the substitution of two distal prolines with alanines (P382A/P384A). Most intriguing was the resistance of the mutant to proteolytic degradation, which makes this protein an excellent candidate for crystallization of the entire CK2α subunit.

Bischoff and collaborators (2011) have recently determined for the first time the structure of the full-length human CK2α`[C336S] subunit. A point mutation of CK2α` was necessary to prevent covalent dimerization from intermolecular disulfide bridges formed by Cys336. However, these results shed light on the differences between the two catalytic subunits, α and α` (*e.g.*, significantly lower affinity of CKα` towards CK2β relative to that of CK2α).

2.1.6 Mutagenesis of CK2α in other regions

Determination of the structure of the CK2 holoenzyme and individual subunits provided knowledge about the nature and location of the interface between catalytic and regulatory subunits (Niefind et al., 2001). Using structure-guided alanine-scanning mutagenesis combined with isothermal titration calorimetry (ITC), energetic "hot spots" were identified on the surface of CK2α that determine the α/β subunit interaction (Raaf et al., 2011). Three single and one double CK2α subunit mutants were produced, in which individual hydrophobic amino acids located within the CK2α interface were replaced by alanine. The ITC analysis of CK2α mutants revealed that substitution of Leu41 and Phe54 were most disruptive to binding of CK2β. Moreover, the L41A and F54A mutants retained their kinase activity, compared to the wild-type CK2. Based on the results mentioned above, it can be claimed that these residues are suspected of being interaction "hot spots" (Raaf et al., 2011).

The amino-acid sequence and the structure of yeast protein kinase CK2α differ from those of CK2α' and other eukaryotic CK2α subunits. CK2α is unique in containing a 38-amino-acid loop consisting of two α-helical structures situated close to structures engaged in ATP/GTP

and substrate binding (Niefind et al., 2001). Modeling of the tertiary structure of the CK2α showed that, after removing both α-helical motifs, the CK2α subunit assumes a structure that is more similar to that of CK2α' than it is to the structure of intact CK2α. The deletion of the 38 amino acids from CK2α drastically decreases its catalytic efficiency. Its characteristics are similar to yeast CK2α' with respect to sensitivity to salt, heparin and spermine (Sajnaga et al., 2008) (Fig. 3).
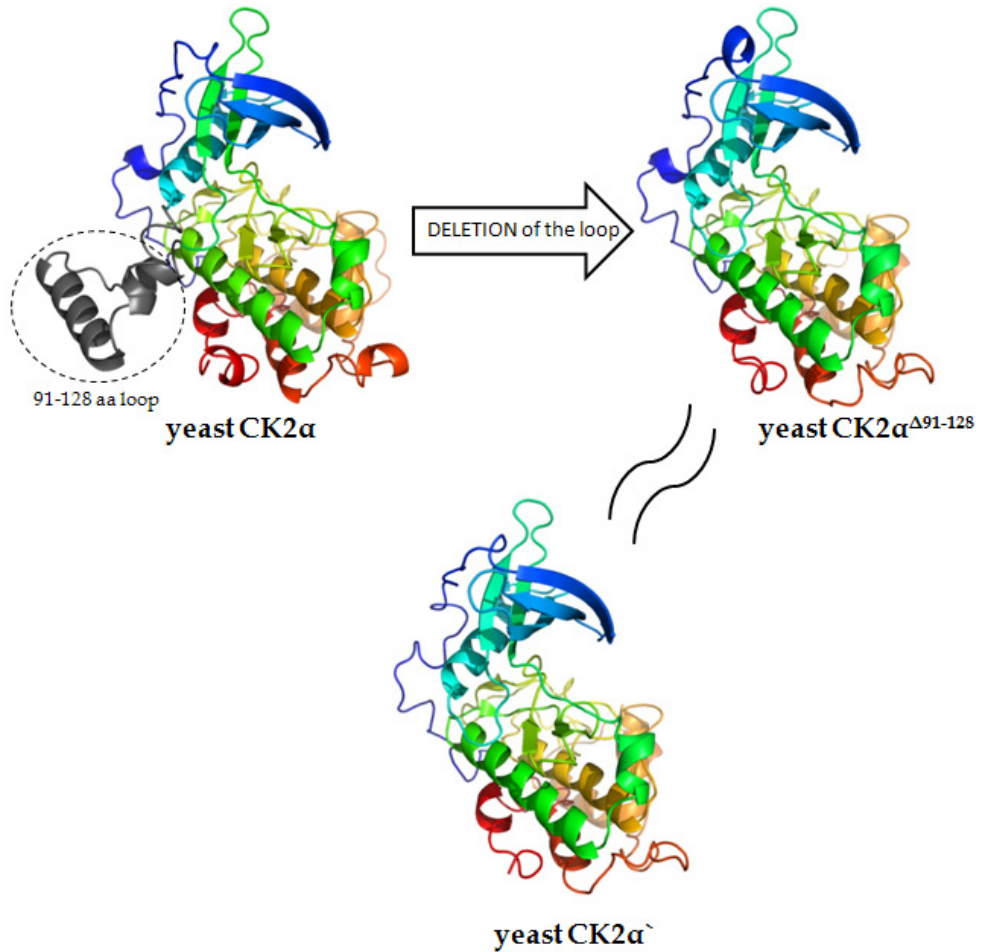
Fig. 3. Conformational consequences of mutagenesis of the yeast CK2α catalytic subunit.

CK2 residues	Location	Mutant	Reference/source
Substrate recognition and inhibition by polyanions			
K49	Subd. I; Gly loop	K49A	Sarno et al., 1999
K74			Sarno et al., 1997a, 1998, 1999, Vaglio et al. 1996
K75	Subd. II/III, Lys rich loop	K74-77A, K77A	
K76			Gatica et al., 1994[1]
K77			
K79		K79A,	
R80	Subd. III, Helix C	R80A/K83A	Sarno et al., 1998, 1999
K83			
K122	Subd.V, Linker region	K122A	Sarno et al., 1997a, 1999
H160	Subd. VIb, Catalytic loop	H160D	Dobrowolska et al., 1994
R191		R191, 195, K190A	Sarno et al., 1997a, 1998, 1999; Vaglio et al., 1996
R195	Subd. VIII, p+1 loop		
K190		K198A	
K198			
Catalytic efficiency and specificity			
G48[2]	Subd. I, Gly loop	G48D	Chaillot et al., 2000
V66	Subd. II	V66A M163A	Yde et al., 2005
M163	Subd. VIb	CK2α $^{1-335}$	

V66	Subd. II	V66A/I174A	Sarno et al., 2005
I174	Subd. VII		
V66/W176	Subd. II, Subd. VII	V66A/W176F	Jakobi and Traugh, 1992, 1995
N118	Subd. V, ATP/GTP binding region	N118A, CK2α$^{\Delta N118}$	Srinivasan et al., 1999; Jakob et al., 2000
D156	Subd. VIb	D156A	Cosmelli et al., 1997
M163	Subd. VIb	M163G	Sarno et al., 2005
G177	Subd. VII	G117K	Chaillot et al., 2000
N189	Subd. VII, Activating segment	N189R	Srinivasan et al., 1999
Regulation by β subunit			
L41	Subd. I	L41A	Raaf et al., 2011
L54	Subd. I, ATP/GTP binding region	L54A	Raaf et al., 2011
V66	Subd. II	V66A,	Sarno et al., 1997b
W176	Subd. VII	V66A/W176F	Jakobi &Traugh, 1992
Constitutive activity			
M6 –V30	N-terminus	Δ2-12, Δ2-18, Δ2-24, Δ2-30	Sarno et al., 2001, 2002; Cristiani et al., 2011
Y125	Subd. V, Hinge region	Y125F[4]	Cristiani et al., 2011
E180	Subd. VII, Activation segment	E180A	Sarno et al., 2002

E182	Subd. VII, Activation segment	Y182F	Sarno et al., 2002
Stability			
M336-Q393	C-terminus	Δ336-393	Ermakova et al., 2003
P382		P382A	
	C-terminus		Grasselini et al., 2004
P384		P384A	
C336[3]	C-terminus	C336S	Bischoff et al., 2011

[a]The residue numbers correspond with those of human CK2α, unless otherwise indicated. The Roman numerals indicate the eleven conserved subdomains present in the catalytic domain of all protein kinases (Hanks & Hunter, 1995). Abbreviations: [1]CK2α from *Xenopus laevis;* [2]CK2α from *Yarrovia lipolytica;* [3]Human CK2α'; [4]*in silico* mutation.

Table 1. Summary of CK2α mutants[a]

The deletion of the loop of amino acids 91-128 from yeast CK2α led to behavioral and structural similarity to CK2α` (Sajnaga et al., 2008). The 3D models of proteins were created using the SWISS-MODEL software based on protein structure templates (PDB code 1ds5D) available in the Protein Data Bank and visualized with the PyMOL software.

Chimeras of different kinases can be easily engineered using recombinant DNA technology and used in studies on the structure and function of kinase. To study the effect of CK2β on the activity of CK1α, Jedlicki and collaborators (2008) generated CK2α/CK1α chimeras that were able to bind tightly to the CK2β regulatory subunit, but maintain the peptide substrate specificity of CK1. This is related to the capacity of the CK2β to regulate the activity of CK2α, as well as other protein kinases, such as A-Raf, C-Mos, and Chk1. It has been shown that a chimera combining a large part of the CK1α kinase with the N-terminal region of CK2α that is responsible for binding CK2β can be stimulated by this subunit. It is possible that such chimeras could be used to test the presence of "the docking site" on the CK2β subunit, which would bring substrate molecules near the catalytic subunits.

2.2 Mutagenesis of the regulatory subunit CK2β

From the primary sequence of the β subunit, it is obvious that the charged amino acids are not equally distributed. The acidic residues are clustered in the N-terminal half, whereas the basic residues are clustered in the C-terminal part of the molecule. Mutational studies have shown that, in contrast to cyclins, which invariably act as indispensable activators of CK2-related CDKs, the CK2β subunit fulfills antagonist functions. The features of CK2β can be explored by generating large synthetic fragments, some of which reproduce the C-terminal moiety and thus stimulate its catalytic activity. Fragments reproducing segments of the N-terminal sequence are inhibitory, which becomes especially evident when calmodulin is the substrate (Marin et al, 1992, 1995; Meggio et al 1994; Sarno et al, 1997a).

2.2.1 Mutations of CK2β that affect autophosphorylation

The CK2β subunit is known to be autophosphorylated by the catalytic subunit. Autophosphorylation occurs on serine residues at positions 2 and 3 in the amino-terminal region of the molecule. Both these serines fit CK2 consensus specificity requirements (Marin et al, 1992). This finding was corroborated by the fact that the mutant S2,3G (*i.e.*, S2G/S3G) is completely incapable of autophosphorylation (Hinrichs et al, 1993). Deletion of the first four amino acids (CK2β Δ1-4), which eliminated autophosphorylation of CK2β, had no significant effect on the reconstruction of CK2 holoenzymes nor on their catalytic activity, thermostability, and responsiveness to polylysine. Unlike the wild-type CK2β, however, CK2β Δ1-4 failed to confer to the reconstituted holoenzyme the typical responsiveness to NaCl stimulation. These results indicated that autophosphorylation sites are not required on CK2 for conferring a stable structure and full catalytic activity. In contrast an autophosphorylation site is implicated in the NaCl-dependent fine-tuning of CK2 activity (Meggio et al., 1993). Interestingly, the acidic stretch heavily influences autophosphorylation of the β subunit, even though Ser2 is more than 50 amino acids away in the primary sequence (Boldyreff et al., 1994).

2.2.2 Mutations of CK2β that affect binding with CK2α

In order to shed light on the mechanisms by which the CK2β subunits affect the catalytic properties of CK2 and to elucidate the molecular interactions between the catalytic and regulatory subunits of CK2, Boldyreff and collaborators (1992, 1993) generated a number of mutants of the CK2β subunit, which were tested for their ability to functionally replace the wild-type CK2β. These authors showed that deletion of the last 44 residues of the C-end (CK2β Δ171-215) eliminated the capacity to form tetramers with CK2α and to stimulate activity. However, deletion of the last 34 amino acids (CK2βΔ181-215) yielded an active CK2β that had lower affinity for CK2α. Shorter deletions (*e.g.*, CK2β Δ194-215) did not affect the interaction between the catalytic and regulatory subunits of CK2. Boldyreff and collaborators demonstrated that deletion mutants in which the last 45 or more amino acids are missing were not able to assemble with the α subunit. These data identified the C-terminal segment of CK2β as essential for association with the CK2α subunit, with special reference to its 171-180 stretch, which is indispensable both to form tetrameric CK2 and to stimulate activity of the CK2α catalytic subunit (Boldyreff et al., 1994). Tight interaction between the CK2α and CK2β subunits, accomplished by the C-terminal part of the CK2β subunit, was also described (Kusk et al., 1995; Marin et al., 1997).

Mutagenesis along with crosslinking and peptide studies have shown that the acidic amino acid stretch of CK2β from residues 55-64 interacts with a corresponding basic stretch of the CK2α subunit. However, these weak electrostatic interactions seem to determine the activity of, but not the formation of, the CK2 holoenzyme (Krehan et al., 1996, Sarno et al, 1997b).

Kusk and collaborators (1995) used mutagenesis of CK2 subunits with a yeast two-hybrid system to explore domains involved in intersubunit contact. [In the yeast two-hybrid system, a peptide or protein is fused to part A of a transcriptional activator. Another peptide or protein is fused to part B. Transcriptional activation of an easily assayed reporter gene occurs only when part A and part B come together. Parts A and B

themselves cannot interact to form the transcriptional activator, nor can either part individually (part A, the part A fusions, part B, and the part B fusions) cause the reporter to be expressed. However, if the fusions interact, part A and part B can come together, and the reporter is activated. This is an indication that the peptides or proteins in the fusions can interact.] A series of plasmid constructs was prepared. They encoded N-terminal or C-terminal truncations of the CK2α and CK2β subunits to indicate which regions of the subunits were engaged in CK2 holoenzyme formation in yeast cells. The data revealed that the regulatory CK2β subunit has a modular structure. An N-terminal domain (residues 20-145) is responsible for homodimerization (CK2β/CK2β). A C-terminal domain (residues 152-200) is necessary for heterodimerization (CK2α/CK2β). Amino acid residues 1 to 20 in the N-terminus and 351 to 391 in the C-terminus of CK2α are dispensable for interaction with the regulatory subunit.

2.2.3 Mutations of CK2β that affect the activity of CK2α

The modulation of CK2α subunit activity by CK2β has a stimulatory effect on most substrates. However, when calmodulin is used as the substrate, the CK2β subunit almost completely inhibits the activity of the catalytic subunit (Guerra et al., 1999). This inhibition can be overcome by addition of polylysine (Meggio et al, 1992). Mutagenesis studies on the CK2β subunit revealed an acidic stretch (amino acids 55-64) that is responsible for the inhibitory effect and for the stimulation by polylysine (Meggio et al., 1994). Interestingly, mutants of CK2β bearing substitutions at positions 55, 57, and 59-64 to alanine produced up to 4-fold more active holoenzyme after assembling with the catalytic α subunit than did the wild type. At the same time, these mutants were refractory to the stimulatory effect of polylysine. This finding revealed that the acidic N-terminal cluster of CK2β, especially Asp55 and Glu57, is involved in intrinsic down-regulation of CK2 basal activity and has been implicated in responsiveness to various effectors (Boldyreff et al., 1993, 1994).

Other data provided by Hinrichs and collaborators (1995) demonstrated that Pro58 located in the center of the acidic segment also constitutes an important structural feature affecting the function of down-regulation of CK2β towards the catalytic subunits. The effect of a mutation of proline to alanine resulted in an effect that was similar to mutation of the acidic residues alone. It produced hyperactive CK2β subunits that stimulated the CK2α activity to a greater extent than did the wild-type CK2β subunit.

2.2.4 Mutations of CK2β that affect export of the holoenzyme

It is known that protein kinase CK2 is present in not only the cytoplasm, nuclei, and several other cell organelles, but also on the external side of the cellular membrane (Kubler et al, 1983). Rodrigez and collaborators (2008) have studied the role of CK2β in the export of the holoenzyme to the extracellular membrane through deletion and point mutations. The region of CK2β between amino acids 20 and 33 was found to be necessary, but not sufficient, to allow the catalytic subunits to function as an ectokinase. An important function of this region is fulfilled by Phe21 and Phe22, which anchor the loop of the 20-33 sequence. Another key element of this region is constituted by the acidic residues in positions 26-28. They are exposed to the medium, free to interact with other proteins (Bolanos-Garcia et al, 2006).

2.2.5 Mutation of CK2β that affects its stability

Overexpression of CK2 catalytic subunits leads to increased cell proliferation and transformation, while overexpression of the regulatory CK2 subunit is associated with decreased proliferation in yeast and mammalian cells (Li et al., 1999; Lebrin et al., 2001; Vilk et al., 2001). Moreover, CK2β is physiologically expressed at a higher level than CK2α, and the excess of the regulatory subunit is rapidly ubiquitinated and degraded in a proteasome-dependent manner (Luscher & Litchfield, 1994; Zhang et al., 2002). To protect CKβ from the degradation machinery and to stabilize it, six surface-exposed lysine residues were mutated to arginine (French et al., 2007). The 6KR mutant functioned as normal CK2β, but it was not sensitive to proteasome inhibition. The physiological role of mutagenesis-mediated CK2β stabilization was also examined with the use of cell proliferation assays. A significant decrease in proliferation was observed in cells expressing the 6KR mutant when compared to wild-type CK2β. The authors suggest that the stabilized form of the CK2 regulatory subunits can be utilized to inhibit cell proliferation in cancer cells (French et al., 2007).

2.3 Mutagenesis of CK2 substrates

Protein kinase CK2 is a multi-substrate enzyme with a large number of cellular partners. In 2003, Meggio and Pinna updated the list of 307 CK2 substrates with 308 sites phosphorylated by CK2 (Meggio & Pinna, 2003). This number is now out-of-date, as novel CK2 protein substrates are discovered every year. A *bona fide* CK2 substrate may possess one or several phosphoacceptor sites affected by CK2, but an analysis of the initial amino acid sequences of possible CK2 partners may show a dozen or so putative CK2 sites. Site-directed mutagenesis is a useful tool to create CK2 substrate mutants. Such proteins are produced (1) to indicate precisely the phosphorylatable amino acid, (2) to study the physiological significance of CK2-mediated phosphorylation of a given protein substrate, or (3) to confirm the physiological relevance of CK2-mediated phosphorylation. Presented below are several examples of the mutagenesis of CK2 substrates.

Mdm2 is a cellular oncoprotein that down-regulates the growth suppressor protein p53 (Barak et al., 1992). Computer analysis of the amino acid sequence of Mdm2 revealed 19 putative CK2 phosphorylation sites. Three Mdm2 mutants with deletions at codons 1-114, 93-285, and 271-491 were produced to exclude sites that are not affected by CK2. The phoshorylation assays revealed that only the central part of Mdm2 is phosphorylated. Based on further detailed analysis of the remaining CK2 consensus sites, Ser269 was chosen to be the most promising. Using overlap extension PCR (see section 2.7 in the chapter by Sturtevant), the Mdm2 point mutant S269A was constructed and the relevant CK2 phosphorylation site was finally discovered (Götz et al., 1999).

In some protein substrates, putative CK2 phosphorylation sites are located close to one another, and thus several point mutants had to be produced to score them. The consensus sequence analysis of the N-terminal domain of the human transcription factor Tcf-4 indicated multiple sites that fit the motif for CK2 phosphorylation. No CK2-mediated phosphorylation was detected on the Tcf-4 fragments comprising amino acids 1-30 and 1-49. Thus, the best candidates for CK2-affected amino acids were the serine residues located in the Tcf-4 peptide T_{54}NQDSSSDSEAERRP$_{68}$. Three Tcf-4 mutants, one triple point mutant (S58A/S59A/S60A) and two single point mutants (S58E and S60E) were made to help indicate the phosphorylatable amino acid. *In vitro* phosphorylation assays revealed that all three adjacent serines are modified by CK2 with different efficiencies (Miravet et al., 2002).

Sic1 is a yeast protein that specifically inhibits Clb/Cdk activity in the G1 phase, so that DNA replication is suppressed (Verma et al., 2001; Nash et al., 2001). Moreover, Sic1 undergoes multistep phosphorylation. Therefore, Sic1 phosphorylation occurs at several positions. One looks like the CK2 consensus site. CK2-mediated phosphorylation of Sic1 within the Q_{199}ESEDEED sequence was confirmed both *in vitro* and *in vivo* in *Saccharomyces cerevisiae* cells (Coccetti et al., 2004, 2006). Mutations of the CK2 consensus site on Sic1 (S201A and S201E) alter the coordination between cell growth and division. They also change the level and time-course of S-Cdk kinase activity. These mutation data strongly support the physiological relevance of Sic1 phosphorylation for inhibitory activity (Coccetti et al., 2004).

The regulatory effect of CK2 activity on the Wnt signaling pathway is widely known (Pinna, 2002; Litchfield, 2003). Kinase phosphorylates and interacts with β-catenin and thus enhances the stability and transcriptional activity of β-catenin (Song et al, 2003; Seldin et al, 2005). The AKT/PKB kinase is also a well-known CK2 substrate and interacting partner. CK2-mediated phosphorylation at Ser129 causes AKT hyperactivation (Di Maira et al, 2005; Guerra, 2006). CK2 may link the two pathways..

To elucidate the roles of CK2 in the Wnt and AKT/PKB signaling pathways, the AKT phosphorylation-deficient mutant (S129A) was overexpressed in an embryonic cell line. The β-catenin-dependent transcriptional activity was analyzed. The data obtained indicate that blockage of AKT phosphorylation by CK2 impairs β-catenin activity and decreases its stability. Therefore, CK2-mediated AKT phosphorylation at Ser129 is a necessary step in the up-regulation of the β-catenin transcriptional activity in human embryonic kidney cells (Ponce et al., 2011).

Besides phosphorylation of numerous cellular proteins, CK2 directly interacts with many of them forming protein-protein complexes (Litchfield, 2003). Both catalytic and regulatory CK2 subunits can interact with different proteins, independently of the holoenzyme (Bibby et al., 2005). Wee1 kinase, involved in cell cycle progression, is one such CK2 protein partner. The Wee1 kinase is a key inhibitor of cyclin-dependent kinase (CDK1) and mitotic entry in eukaryotes. Several deletion mutants of the Wee1 catalytic domain were produced to investigate the interaction with CK2 subunits. Immunoprecipitation experiments revealed that Wee1 binds CK2β via two domains of Wee1 (comprising amino acids 59-71 and 232-332) and two regions of CK2β (comprising residues 1-5 and 155-170). Although the interaction does not affect Wee1 activity, it up-regulates CDK1 by reversing the Wee1-mediated inhibitory effect on CDK1. These findings reinforce the notion that CK2β can serve other protein kinases. It may be a universal regulatory subunit that can act independently of the CK2 holoenzyme (Olsen et al., 2010).

3. Conclusion

Even 58 years after its first description (Burnett & Kennedy, 1954), the story of protein kinase CK2 has not been fully clarified. This enzyme catalyzes phosphorylation of over 300 substrates. They are characterized by having multiple acidic residues surrounding the phospho-acceptor amino acid. Consequently, CK2 plays a key role in several physiological and pathological processes (Guerra & Issinger, 2008). After all those years of research, we are still asking the question: how is it possible that one kinase can be involved in so many

different biochemical processes in the cell? Using different biochemical and genetic methods, we have solved several problems connected with the structure and mechanism of the catalytic action of this enigmatic protein kinase. The application of mutagenesis methods in many cases has helped us and will continue to help us get answers to many problems connected with CK2 activity. Among them are the following:

- The interaction between subunits
- Catalytic specificity and efficiency
- Substrate recognition
- Regulation by the β-subunit
- Stability of the subunits
- Interactions with modulators and substrates
- The effect of phosphorylation on catalytic activity
- Constitutive CK2 activity.

A protein kinase, such as CK2, is difficult to explore with respect to its physiological functions. CK2 has been shown to be involved in numerous aspects of cell proliferation and survival, including cell cycle progression and apoptosis control (Ahmad et al., 2008; Ahmed et al., 2002; Batistuta, 2009; Gyenis & Litchfield, 2008; Meggio & Pinna, 2003; Litchfield, 2003). Alterations in the levels or activity of CK2 have been implicated in a variety of human diseases, including cancers (Guerra & Issinger, 2008). All these observations raise important questions regarding the mechanisms that control CK2 activity and specificity. These questions have a special value, since defects in regulation of these processes could contribute to tumorigenesis.

In this context, the application of mutagenesis methods, together with other techniques (*e.g.*, molecular modeling), may be very useful in designing highly effective and specific inhibitors that are promising for CK2-based target therapy.

4. Acknowledgement

5. References

Ahmad K.A., Wang G., Unger G., Slaton J., Ahmed K. (2008) Protein kinase CK2 – A key suppressor of apoptosis. *Advances in Enzyme Regulation,* Vol. 48, No. 1, (April 2008), pp. 179-187, ISSN 0065-2571.

Ahmed, K., Gerber, D.A., Cochet, C. (2002) Joining the cell survival squad: an emerging role for protein kinase CK2. Trends in Cellular Biology, Vol. 12, No. 5, (May 2002), pp. 226-230, ISSN 0962-8924.

Adler, V., Pincus, M.R., Minamoto, T., Fuchs, S. Y., Bluth, M.J., Brandt-Rauf, P.W., Friedman, F.K., Robinson, R.C., Chen, J.M., Wang, X.W., Harris, C.C. & Ronai, Z. (1997). Conformation-dependent phosphorylation of p53. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 94, No. 5, (March 1997), pp. 1686-1691, ISSN 0027-8424.

Allende, J. E. & Allende, C.C. (1995). Protein kinases. 4. Protein kinase CK2: an enzyme with multiple substrates and a puzzling regulation. *The FASEB Journal*, Vol. 9, No. 5, (March 1995), pp. 313-323, ISSN 0892-6638.

Arnold K., Bordoli L., Kopp J., Schwede T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* Vol.22, No. 2, (January 2006), pp. 195-201, ISSN 1367-4803

Barak, Y. & Oren, M. (1992). Enhanced binding of a 95 kDa protein to p53 in cells undergoing p53-mediated growth arrest. *The EMBO Journal*, Vol. 11, No. 6, (June 1992), pp. 2115-2121, ISSN 0261-4189.

Battistutta, R. (2009) Protein kinase CK2 in health and disease: Structural bases of protein kinase CK2 inhibition. *Cellular and Molecular Life Sciences,* Vol. 66, No. 11-12, (June 2009), pp. 1868-1889, ISSN 1420-682X.

Battistutta, R., Sarno, S., De Moliner, E., Marin, O., Issinger, O.-G., Zanotti, G., Pinna, L.A. (2000) The crystal structure of the complex of *Zea mays* alpha subunit with a fragment of human beta subunit provides the clue to the architecture of protein kinase CK2 holoenzyme. *European Journal of Biochemistry* 267, No. 16, (August 2000), pp. 5184-5190, ISSN 0014-2956.

Battistutta, R., De Moliner, E., Sarno, S., Zanotti, G. & Pinna, L.A. (2001). Structural features underlying selective inhibition of protein kinase CK2 by ATP site-directed tetrabromo-2-benzotriazole. *Protein Science*, Vol. 10, No. 11, (November 2001), pp. 2200-2206, ISSN 0961-8368

Bibby, A.C. & Litchfield, D.W. (2005). The multiple personalities of the regulatory subunit of protein kinase CK2: CK2 dependent and CK2 independent roles reveal a secret identity for CK2beta. *International Journal of Biological Sciences*, Vol. 1, No. 2, (April 2005), pp. 67-79, ISSN 1449-2288

Bischoff, N., Olsen, B., Raaf, J., Bretner, M., Issinger, O.-G. & Niefind, K. (2011). Structure of the human protein kinase CK2 catalytic subunit CK2alpha' and interaction thermodynamics with the regulatory subunit CK2beta. *Journal of Molecular Biology*, Vol. 407, No. 1, (March 2011), pp. 1-12, ISSN 1089-8638.

Becker, W., Weber, Y., Wetzel, K., Eirmbter, K., Tejedor, F.J., Joost, H.-G. (1998) Sequence characteristics, subcellular localization, and substrate specificity of DYRK-related kinases, a novel family of dual specificity protein kinases. *The Journal of Biological Chemistry*, Vol. 273, No. 40, (October 1998), pp. 25893-25902, ISSN 0021-9258.

Biondi R.M. and Nebreda A.R. (2003) Signalling specificity of Ser/Thr protein **kinases** through docking-site-mediated interactions. *Biochemical Journal,* Vol. 372, Pt. 1, (May 2003), pp. 1-13, ISSN 0264-6021.

Bolanos-Garcia, V.M., Fernandez-Recio, J., Allende, J.E. & Blundell, T.L. (2006). Identifying interaction motifs in CK2beta--a ubiquitous kinase regulatory subunit. *Trends in Biochemical Sciences*, Vol. 31, No. 12, (December 2006), pp. 654-661, ISSN 0968-0004.

Boldyreff, B., Meggio, F., Pinna, L.A. & Issinger, O.-G. (1992). Casein kinase-2 structure-function relationship: creation of a set of mutants of the beta subunit that variably surrogate the wildtype beta subunit function. *Biochemical and Biophysical Research Communications*, Vol. 188, No. 1, (October 1992), pp. 228-234, ISSN 0006-291X

Boldyreff, B., Meggio, F., Pinna, L.A. & Issinger, O.-G. (1993). Reconstitution of normal and hyperactivated forms of casein kinase-2 by variably mutated beta-subunits. *Biochemistry*, Vol. 32, No. 47, (November 1993), pp. 12672-12677, ISSN 0006-2960

Boldyreff, B., Meggio, F., Pinna, L.A. & Issinger, O.-G. (1994). Protein kinase CK2 structure-function relationship: effects of the beta subunit on reconstitution and activity. *Cellular & Molecular Biology Research*, Vol. 40, No. 5-6, (January 1994), pp. 391-399, ISSN 0968-8773

Boldyreff, B., James, P., Staudenmann, W., Issinger, O.-G. (1993) Ser2 is the autophosphorylation site in the beta subunit from bicistronically expressed human casein kinase-2 and from native rat liver casein kinase-2 beta. *European Journal of Biochemistry,* Vol. 218, No. 2, (December 1), pp. 515-521, ISSN 0014-2956.

Boldyreff, B., Mietens, U., Issinger O.-G. (1996) Structure of protein kinase CK2: dimerization of the human beta-subunit. *FEBS Letters*, Vol. 379, No.2, (January 1996), pp. 153-156, ISSN 0014-5793.

Brognard, J. & Hunter, T. (2011) Protein kinase signaling networks in cancer. *Current Opinion in Genetics and Development,* Vol. 21, No.1 , (February 2011), pp. 4-11, ISSN 0959-437X.

Bosc, D. G., Slominski, E., Sichler, C. & Litchfield, D. W. (1995). Phosphorylation of casein kinase II by p34cdc2. Identification of phosphorylation sites using phosphorylation site mutants in vitro. *The Journal of Biological Chemistry*, Vol. 270, No. 43, (October 1995), pp. 25872-25878, ISSN 0021-9258

Bossemeyer, D. (1994). The glycine-rich sequence of protein kinases: a multifunctional element. *Trends in Biochemical Sciences*, Vol. 19, No. 5, (May 1994), pp. 201-205, ISSN 0968-0004

Burnett, G. & Kennedy E.P. (1954) The enzymatic phosphorylation of proteins. *The Journal of Biological Chemistry*, Vol. 211, No. 2, (December 1954), pp. 969-980, ISSN 0021-9258.

Canton, D.A., Zhang, C., Litchfield, D.W. (2001) Assembly of protein kinase CK2: investigation of complex formation between catalytic and regulatory subunits using a zinc-finger-deficient mutant of CK2beta. *Biochemical Journal*, Vol. 358, Pt. 1, (August 2001), pp. 87-94, ISSN 0264-6021.

Chantalat, L., Leroy, D., Filhol, O., Nueda, A., Benitez, M.J., Chambaz, E.M., Cochet, C., Dideberg, O. (1999) Crystal structure of the human protein kinase CK2 regulatory subunit reveals its zinc finger-mediated dimerization. *The EMBO Journal,* Vol. 18, No. 11, (June 1999), pp. 2930-2940, ISSN 0261-4189.

Chaillot, D., Declerck, N., Niefind, K., Schomburg, D., Chardot, T. & Meunier, J.C. (2000). Mutation of recombinant catalytic subunit alpha of the protein kinase CK2 that affects catalytic efficiency and specificity. *Protein Engineering*, Vol. 13, No. 4, (April 2000), pp. 291-298, ISSN 0269-2139.

Chen, M., Li D., Krebs, E.G., Cooper, J.A. (1997) The casein kinase II beta subunit binds to Mos and inhibits Mos activity. *Molecular and Cellular Biology*, Vol. 17, No. 4, (April 1997), pp. 1904-1912, ISSN 0270-7306.

Coccetti, P., Rossi, R. L., Sternieri, F., Porro, D., Russo, G. L., di Fonzo, A., Magni, F., Vanoni, M. & Alberghina, L. (2004). Mutations of the CK2 phosphorylation site of Sic1 affect cell size and S-Cdk kinase activity in Saccharomyces cerevisiae. *Molecular Microbiology*, Vol. 51, No. 2, (January 2004), pp. 447-460, ISSN 0950-382X.

Coccetti, P., Zinzalla, V., Tedeschi, G., Russo, G. L., Fantinato, S., Marin, O., Pinna, L. A., Vanoni, M. & Alberghina, L. (2006). Sic1 is phosphorylated by CK2 on Ser201 in budding yeast cells. *Biochemical and Biophysical Research Communications*, Vol. 346, No. 3, (August 2006), pp. 786-793, ISSN 0006-291X.

Cohen, P. (2002) Protein kinases - the major drug targets of the twenty-first century? *Nature Reviews Drug Discovery*, Vol. 1, No. 4, (April 2002), pp. 309-315, ISSN 1474-1776.

Cosmelli, D., Antonelli, M., Allende, C. C. & Allende, J. E. (1997). An inactive mutant of the alpha subunit of protein kinase CK2 that traps the regulatory CK2beta subunit. *FEBS letters*, Vol. 410, No. 2-3, (June 1997), pp. 391-396, ISSN 0014-5793.

Cristiani, A., Costa, G., Cozza, G., Meggio, F., Scapozza, L. & Moro, S. (2011). The role of the N-terminal domain in the regulation of the "constitutively active" conformation of protein kinase CK2alpha: insight from a molecular dynamics investigation. *ChemMedChem*, Vol. 6, No. 7, (July 2011), pp. 1207-1216, ISSN 1860-7187.

Di Maira, G., Salvi, M., Arrigoni, G., Marin, O., Sarno, S., Brustolon, F., Pinna, L. A. & Ruzzene, M. (2005). Protein kinase CK2 phosphorylates and upregulates Akt/PKB. *Cell death and differentiation*, Vol. 12, No. 6, (June 2005), pp. 668-677, ISSN 1350-9047.

Dobrowolska, G., Meggio, F., Marin, O., Lozeman, F. J., Li, D., Pinna, L. A. & Krebs, E. G. (1994). Substrate recognition by casein kinase-II: the role of histidine-160. *FEBS letters*, Vol. 355, No. 3, (December 1994), pp. 237-241, ISSN 0014-5793.

Duncan J.S., Litchfield D.W. (2008) Too much of a good thing: The role of protein kinase CK2 in tumorigenesis and prospects for therapeutic inhibition of CK2. *Biochemica et Biophysica Acta,* Vol. 1784, No. 1, (January 2008), pp. 33-47, ISSN 0006-3002.

Duncan, J.S., Turowec, J.P., Vilk, G., Li, S.S.C., Gloor, G.B., Litchfield, D.W. (2010) Regulation of cell proliferation and survival: Convergence of protein kinases and caspases. *Biochemica et Biophysica Acta,* Vol. 1804, No. 3, (March 2010) pp. 505-510, ISSN 0006-3002.

Ermakova, I., Boldyreff, B., Issinger, O. G. & Niefind, K. (2003). Crystal structure of a C-terminal deletion mutant of human protein kinase CK2 catalytic subunit. *Journal of Molecular Biology*, Vol. 330, No. 5, (July 2003), pp. 925-934, ISSN 0022-2836.

Faust, M., Montenarh, M. (2000) Subcellular localization of protein kinase CK2. A key to its function? Cell and Tissue Research, Vol. 301, No. 3, (September 2000), pp. 329–340, ISSN 0302-766X.

French, A. C., Luscher, B. & Litchfield, D. W. (2007). Development of a stabilized form of the regulatory CK2beta subunit that inhibits cell proliferation. *The Journal of Biological Chemistry*, Vol. 282, No. 40, (October 2007), pp. 29667-29677, ISSN 0021-9258.

Gatica, M., Jedlicki, A., Allende, C. C. & Allende, J. E. (1994). Activity of the E75E76 mutant of the alpha subunit of casein kinase II from Xenopus laevis. *FEBS Letters*, Vol. 339, No. 1-2, (February 1994), pp. 93-96, ISSN 0014-5793.

Ghosh & Adams (2011) Phosphorylation mechanism and structure of serine-arginine protein kinases. *The FEBS Journal*, Vol. 278, No. 4, (February 2011), pp. 587-597, ISSN 1742-4658

Gianoncelli, A., Cozza, G., Orzeszko, A., Meggio, F., Kazimierczuk, Z., Pinna, L.A. (2009) Tetraiodobenzimidazoles are potnt inhibitors of protein kinase CK2. *Bioorganic & Medicinal Chemistry*, Vol. 17, No 20, (October 2009), pp. 7281-7289, ISSN 0968-0896.

Gietz, R.D., Graham, K.C. and Litchfield, D.W. (1995) Interactions between the subunits of casein kinase II. *The Jornal of Biological Chemistry*, Vol. 270, No. 22, (June 1995), pp. 13017–13021. ISSN 0021-9258.

Glover C.V. (1998) On the physiological role of casein kinase II in *Saccharomyces cerevisiae*. *Progress in Nucleic Acid Research & Molecular Biology*, Vol. 59, (May 1998), pp. 95–133, ISSN 0079-6603.

Gotz, C., Kartarius, S., Scholtes, P., Nastainczyk, W. & Montenarh, M. (1999). Identification of a CK2 phosphorylation site in mdm2. *European Journal of biochemistry / FEBS*, Vol. 266, No. 2, (December 1999), pp. 493-501, ISSN 0014-2956

Gotz, C., Scholtes, P., Prowald, A., Schuster, N., Nastainczyk, W. & Montenarh, M. (1999). Protein kinase CK2 interacts with a multi-protein binding domain of p53. *Molecular and Cellular Biochemistry*, Vol. 191, No. 1-2, (January 1999), pp. 111-120, ISSN 0300-8177

Graham, K.C. and Litchfield, D.W. (2000) The regulatory beta subunit of protein kinase CK2 mediates formation of tetrameric CK2 complexes. J. Biol. Chem. 275: 5003–5010.

Grasselli, E., Tomati, V., Bernasconi, M. V., Nicolini, C. & Vergani, L. (2004). C-terminal region of protein kinase CK2 alpha: How the structure can affect function and stability of the catalytic subunit. *Journal of Cellular Bbiochemistry*, Vol. 92, No. 2, (May 2004), pp. 270-284, ISSN 0730-2312

Guerra, B. (2006). Protein kinase CK2 subunits are positive regulators of AKT kinase. *International Journal of Oncology*, Vol. 28, No. 3, (March 2006), pp. 685-693, ISSN 1019-6439

Guerra B. & Issinger O.-G. (2008) Protein kinase CK2 in human diseases. *Current Medicinal Chemistry*, Vol. 15, No. 19, pp. 1870-1886, ISSN 0929-8673.

Guerra B., Issinger O.-G., Wang J.Y. (2003) Modulation of human checkpoint kinase Chk1 by the regulatory beta-subunit of protein kinase CK2. *Oncogene*, Vol. 22, No. 32, (August 7), pp. 4933-4942, ISSN 0950-9232.

Guerra, B., Boldyreff, B., Sarno, S., Cesaro, L., Issinger, O. G. & Pinna, L. A. (1999). CK2: a protein kinase in need of control. *Pharmacology & Therapeutics*, Vol. 82, No. 2-3, (May-June 1999), pp. 303-313, ISSN 0163-7258.

Gyenis, L. & Litchfield, D.W. (2008) The emerging CK2 interactome: insights into the regulation and functions of CK2. *Molecular and Cellular Biochemistry*, Vol. 316, No. 1-2, (September 2008), pp. 5-14, ISSN 0300-8177

Hanks, S. K. & Hunter, T. (1995). Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *The FASEB Journal*, Vol. 9, No. 8, (May 1995), pp. 576-596, ISSN 0892-6638.

Hanks, S.K., Quinn, A.M., Hunter, T. (1988) The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science*, Vol. 241, No. 4861 , (July 1), pp. 42-52, ISSN 0036-8075.

Hinrichs, M. V., Jedlicki, A., Tellez, R., Pongor, S., Gatica, M., Allende, C. C. & Allende, J. E. (1993). Activity of recombinant alpha and beta subunits of casein kinase II from Xenopus laevis. *Biochemistry*, Vol. 32, No. 28, (July 1993), pp. 7310-7316, ISSN 0006-2960

Hinrichs, M. V., Gatica, M., Allende, C. C. & Allende, J. E. (1995). Site-directed mutants of the beta subunit of protein kinase CK2 demonstrate the important role of Pro-58. *FEBS Letters*, Vol. 368, No. 2, (July 1995), pp. 211-214, ISSN 0014-5793

Hu, E. & Rubin, C. S. (1990). Expression of wild-type and mutated forms of the catalytic (alpha) subunit of Caenorhabditis elegans casein kinase II in Escherichia coli. *The Jornal of Biological Chemistry,*, Vol. 265, No. 33, (November 1990), pp. 20609-20615, ISSN 0021-9258.

Issinger O.-G. (1993) Casein kinases: pleiotropic mediators of cellular regulation. *Pharmacology & Therapeutics*, Vol. 59, No. 1, (January 1993), pp. 1-30, ISSN 0163-7258.

Jacob, G., Neckelman, G., Jimenez, M., Allende, C. C. & Allende, J. E. (2000). Involvement of asparagine 118 in the nucleotide specificity of the catalytic subunit of protein kinase CK2. *FEBS Letters*, Vol. 466, No. 2-3, (January 2000), pp. 363-366, ISSN 0014-5793.

Jakobi, R. & Traugh, J. A. (1992). Characterization of the phosphotransferase domain of casein kinase II by site-directed mutagenesis and expression in Escherichia coli. *The Journal of Biological Chemistry*, Vol. 267, No. 33, (November 1992), pp. 23894-23902, ISSN 0021-9258.

Jakobi, R., Lin, W. J. & Traugh, J. A. (1994). Modes of regulation of casein kinase II. *Cellular & Molecular Biology Research*, Vol. 40, No. 5-6, pp. 421-429, ISSN 0968-8773.

Jakobi, R. & Traugh, J. A. (1995). Site-directed mutagenesis and structure/function studies of casein kinase II correlate stimulation of activity by the beta subunit with changes in conformation and ATP/GTP utilization. *European Journal of Biochemistry*, Vol. 230, No. 3, (June 1995), pp. 1111-1117, ISSN 0014-2956

Janeczko, M., Masłyk, M., Szyszka, R., Baier, A. (2011) Interactions between subunits of protein kinase CK2 and their protein substrates influences its sensitivity to specific inhibitors. *Molecular & Cellular Biochemistry*, Vol. 356, No. 1-2, (October 2011), pp. 121-126, ISSN 0300-8177.

Jauch, E., Melzig, J., Brkulj, M., Raabe, T. (2002) *In vivo* functional analysis of *Drosophila* protein kinase casein kinase 2 (CK2) beta-subunit. *Gene*, Vol. 298, No.1, (September 18), pp. 29-39, ISSN 0378-1119.

Jedlicki, A., Allende, C. C. & Allende, J. E. (2008). CK2alpha/CK1alpha chimeras are sensitive to regulation by the CK2beta subunit. *Molecular & Cellular Biochemistry*, Vol. 316, No. 1-2, (September 2008), pp. 25-35, ISSN 0300-8177

Jensen, B.C., Kifer, C.T., Brekken, D.L., Randall A.C., Wang, Q., Drees, B.L. & Parsons M. (2007) Characterization of protein kinase CK2 from *Trypanosoma brucei*. *Molecular & Biochemical Parasitology*, Vol. 151, No. 1, (January 2007), pp. 28-40, ISSN 0166-6851.

Johnson S.N. (2009) Protein kinase inhibitors: contributions from structure to clinical compounds. *Quarterly Reviews of Biophysics*, Vol. 42, No. 1, (February 2009), pp. 1-40, ISSN 0033-5835.

King, R.W., Glotzer, M., Kirschner, M.W. (1996) Mutagenic analysis of the destruction signal of mitotic cyclins and structural characterization of ubiquitinated intermediates. *Molecular Biology of the Cell*, Vol 7, No. 9, (September 1996), pp. 1343-1357, ISSN 1939-4586.

Knighton, D.R., Zheng, J.H., Ten Eyk, L.F., Ashford, V.A., Xuong, N.H., Taylor, S.S., Sowadski, J.M. (1991) Crystal structure of the catalytic subunit of cyclic adenosine

monophosphate-dependent protein kinase. *Science*, Vol. 253, No. 5018, (July 1991), pp. 407-414, ISSN 0036-8075.

Kolaiti R.-M., Baier A., Szyszka R., Kouyanou-Koutsoukou S. (2011) Isolation of a CK2α subunit and the holoenzyme from the mussel *Mytilus galloprovincialis* and construction of the CK2α and CK2β cDNAs. *Marine Biotechnology (New York)*, Vol. 13, No. 3, (June 2011), pp. 505-516, ISSN 1436-2228.

Kopp, J., Schwede, T. (2004) Automated protein structure homology modeling: a progress report. *Pharmacogenomics*, Vol. 5, No. 4, (June 2004), pp. 405-416, ISSN 1462-2416.

Kouyanou-Koutsoukou S., Baier A., Kolaitis R.-M., Maniatopoulou E., Thanopoulou K., Szyszka R. (2011a) Cloning and purification of protein kinase CK2 recombinant alpha and beta subunits from the Mediterranean fly *Ceratitis capitata*. *Molecular and Cellular Biochemistry*, Vol. 356, No. 1-2, (October 2011), pp. 261-267, ISSN 0300-8177.

Kouyanou-Koutsoukou S., Kalpaxis D.L., Pytharopoulou S., Kolaitis R.-M., Baier A. and Szyszka R. (2011b) Translational control of gene expression in the mussel *Mytilus galloprovincials*: The impact of cellular stress on protein synthesis, the ribosomal stalk and the protein kinase CK2 activity. In: *Mussels: Anatomy, Habitat and Environmental Impact.* (Ed. Lauren E. McGevin), Nova Publisher, pp. 97-128, ISBN 9781617617638 1617617636.

Krehan, A., Lorenz, P., Plana-Coll, M. & Pyerin, W. (1996). Interaction sites between catalytic and regulatory subunits in human protein kinase CK2 holoenzymes as indicated by chemical cross-linking and immunological investigations. *Biochemistry*, Vol. 35, No. 15, (April 1996), pp. 4966-4975, ISSN 0006-2960

Kristensen, L.P., Larsen, M.R., Højrup, P., Issinger, O.-G., Guerra, B. (2004) Phosphorylation of the regulatory β-subunit of protein kinase CK2 by checkpoint kinase Chk1: identification of the *in vitro* CK2β phosphorylation site. *FEBS Letters*, Vol.569, No. 1-3, (July 2004), pp. 217-223, ISSN 0014-5793.

Krupa, A., Abhinandan, K.R., Srinivasan, N. (2004) KinG: a database of protein kinases in genomes. *Nucleic Acids Research*, Vol. 32 (Database Issue), (January 2004), pp. D153-D155, ISSN 0305-1048.

Kubiński, K., Domańska, K., Sajnaga, E., Mazur, E., Zieliński, R. & Szyszka, R. (2007) Yeast holoenzyme of protein kinase CK2 requires both β and β' regulatory subunits for its activity. *Molecular & Cellular Biochemistry*, Vol. 295, No. 1-2, (January 2007), pp. 229-235, ISSN 0300-8177.

Kubler, D., Pyerin, W., Burow, E. & Kinzel, V. (1983). Substrate-effected release of surface-located protein kinase from intact cells. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 80, No. 13, (July 1983), pp. 4021-4025, ISSN 0027-8424.

Kusk, M., Bendixen, C., Duno, M., Westergaard, O. & Thomsen, B. (1995). Genetic dissection of intersubunit contacts within human protein kinase CK2. *Journal of Molecular Biology*, Vol. 253, No. 5, (November 1995), pp. 703-711, ISSN 0022-2836.

Lebrin, F., Chambaz, E. M. & Bianchini, L. (2001). A role for protein kinase CK2 in cell proliferation: evidence using a kinase-inactive mutant of CK2 catalytic subunit alpha. *Oncogene*, Vol. 20, No. 16, (April 2001), pp. 2010-2022, ISSN 0950-9232.

Leroy, D., Alghisi, G.C., Roberts, E., Filhol-Cochet, O., Gasser, S.M. (1999) Mutations in the C-terminal domain of topoisomerase II affect meiotic function and interaction with

the casein kinase 2 beta subunit. *Molecular & Cellular Biochemistry*, Vol. 191, No. 1-2, (January 1999), pp. 85-95, ISSN 0300-8177.

Leroy, D., Heriche, J.K., Filhol, O., Chambaz, E.M. Cochet, C. (1997) Binding of polyamines to an autonomous domain of the regulatory subunit of protein kinase CK2 induces a conformational change in the holoenzyme. A proposed role for the kinase stimulation. *The Journal of Biological Chemistry,* Vol. 272, No. 33, (August 1997), pp. 20820-20827, ISSN 0021-9258.

Li, D., Dobrowolska, G., Aicher, L. D., Chen, M., Wright, J. H., Drueckes, P., Dunphy, E. L., Munar, E. S. & Krebs, E. G. (1999). Expression of the casein kinase 2 subunits in Chinese hamster ovary and 3T3 L1 cells provides information on the role of the enzyme in cell proliferation and the cell cycle. *The Journal of Biological Chemistry*, Vol. 274, No. 46, (November 1999), pp. 32988-32996, ISSN 0021-9258

Li, D., Meier, U.T., Dobrowolska, G. & Krebs E.G. (1997) Specific interaction between casein kinase 2 and the nucleolar protein Nopp140. *The Journal of Biological Chemistry*, Vol. 272, No. 6, (February 1997), pp. 3773-3779, ISSN 0021-9258

Lieberman, S.L. & Ruderman, J.V. (2004) CK2 beta, which inhibits Mos function, binds to a discrete domain in the N-terminus of Mos. *Developmental Biology*, Vol. 268, No. 2, (April 2004), pp. 271-279, ISSN 0012-1606.

Liolli G., 2010. Structural dissection of cyclin dependent kinases regulation and protein recognition properties. *Cell Cycle*, Vol. 9, No. 8, (April 2010), pp. 1551-1561, ISSN 1538-4101.

Litchfield, D. W. (2003). Protein kinase CK2: structure, regulation and role in cellular decisions of life and death. *The Biochemical Journal*, Vol. 369, Pt 1, (January 2003), pp. 1-15, ISSN 0264-6021

Litchfield, D.W., Bosc, D.G., Slonimski, E. (1995) The protein kinase from mitotic human cells that phosphorylates Ser-209 on the casein kinase II beta-subunit is p34cdc2. *Biochemica et Biophysica Acta*, Vol. 1269, No. 1, (October 1995), pp. 69-78, ISSN 0006-3002.

Litchfield, D.W., Lozeman, F.J., Piening, C., Sommercorn, J., Takio, K., Walsh, K.A. & Krebs E.G. (1990) Subunit structure of casein kinase II from bovine testis: demonstration that the α and α´ subunits are distinct polypeptides. *The Journal of Biological Chemistry*, Vol. 265, No. 13, (May 1990), pp. 7638-7644, ISSN 0021-9258.

Luscher, B. & Litchfield, D. W. (1994). Biosynthesis of casein kinase II in lymphoid cell lines. *European Journal of Biochemistry*, Vol. 220, No. 2, (March 1994), pp. 521-526, ISSN 0014-2956.

Manning G., Plowman G.D., Hunter T. & Sudarsanam S. (2002a) Evolution of protein kinase signaling from yeast to man. *Trends in Biochemical Sciences*, Vol. 27, No. 10, (October 2002), pp. 514-520, ISSN 0968-0004

Manning, G., Whyte, D.B., Martinez, R., Hunter, T. & Sudarsanam S. (2002b) The protein kinase complement of the human genome. *Science*, Vol. 298, No. 5600, (December 2002), pp. 1912-1934, ISSN 0036-8075.

Maridor G., Park W., Krek W. & Nigg E.A. (1991) Casein kinase II. cDNA sequences, developmental expression and tissue distribution of mRNAs for a, a´ and b subunits of the chicken enzyme. *The Journal of Biological Chemistry*, Vol. 266, No. 4, (February 1991), pp. 2362-2368, ISSN 0021-9258.

Marin, O., Meggio, F. & Pinna, L.A. (1999) Structural features underlying the unusual mode of calmodulin phosphorylation by protein kinase CK2: A study with synthetic calmodulin fragments. *Biochemical and Biophysical Research Communications*, Vol. 256, No. 2, (March 1999), pp. 442–446, ISSN 0006-291X.

Miranda-Saavedra, D. & Barton, G.J. (2007) Classification and functional annotation of eukaryotic protein kinases. *Proteins*, Vol. 68, No. 4, (September 2007), pp. 893-914, ISSN 0887-3585.

Marin, O., Meggio, F., Draetta, G. & Pinna, L. A. (1992). The consensus sequences for cdc2 kinase and for casein kinase-2 are mutually incompatible. A study with peptides derived from the beta-subunit of casein kinase-2. *FEBS Letters*, Vol. 301, No. 1, (April 1992), pp. 111-114, ISSN 0014-5793.

Marin, O., Meggio, F., Boldyreff, B., Issinger, O. G. & Pinna, L. A. (1995). Dissection of the dual function of the beta-subunit of protein kinase CK2 ('casein kinase-2'): a synthetic peptide reproducing the carboxyl-terminal domain mimicks the positive but not the negative effects of the whole protein. *FEBS Letters*, Vol. 363, No. 1-2, (April 1995), pp. 111-114, ISSN 0014-5793.

Marin, O., Meggio, F., Sarno, S. & Pinna, L. A. (1997). Physical dissection of the structural elements responsible for regulatory properties and intersubunit interactions of protein kinase CK2 beta-subunit. *Biochemistry*, Vol. 36, No. 23, (June 1997), pp. 7192-7198, ISSN 0006-2960.

Meggio, F., Boldyreff, B., Marin, O., Marchiori, F., Perich, J. W., Issinger, O. G. & Pinna, L. A. (1992). The effect of polylysine on casein-kinase-2 activity is influenced by both the structure of the protein/peptide substrates and the subunit composition of the enzyme. *European Journal of Biochemistry,* Vol. 205, No. 3, (May 1992), pp. 939-945, ISSN 0014-2956.

Meggio, F., Boldyreff, B., Issinger, O. G. & Pinna, L. A. (1993). The autophosphorylation and p34cdc2 phosphorylation sites of casein kinase-2 beta-subunit are not essential for reconstituting the fully-active heterotetrameric holoenzyme. *Biochimica et Biophysica Acta*, Vol. 1164, No. 2, (July 1993), pp. 223-225, ISSN 0006-3002.

Meggio, F., Boldyreff, B., Issinger, O. G. & Pinna, L. A. (1994). Casein kinase 2 down-regulation and activation by polybasic peptides are mediated by acidic residues in the 55-64 region of the beta-subunit. A study with calmodulin as phosphorylatable substrate. *Biochemistry*, Vol. 33, No. 14, (April 1994), pp. 4336-4342, ISSN 0006-2960.

Meggio, F. & Pinna, L. A. (2003). One-thousand-and-one substrates of protein kinase CK2? *The FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, Vol. 17, No. 3, (March 2003), pp. 349-368, ISSN 1530-6860

Messenger, M. M., Saulnier, R. B., Gilchrist, A. D., Diamond, P., Gorbsky, G. J. & Litchfield, D. W. (2002). Interactions between protein kinase CK2 and Pin1. Evidence for phosphorylation-dependent interactions. *The Journal of Biological Chemistry*, Vol. 277, No. 25, (June 2002), pp. 23054-23064, ISSN 0021-9258

Miller, S. J., Lou, D. Y., Seldin, D. C., Lane, W. S. & Neel, B. G. (2002). Direct identification of PTEN phosphorylation sites. *FEBS letters*, Vol. 528, No. 1-3, (September 2002), pp. 145-153, ISSN 0014-5793

Miravet, S., Piedra, J., Miro, F., Itarte, E., Garcia de Herreros, A. & Dunach, M. (2002). The transcriptional factor Tcf-4 contains different binding sites for beta-catenin and

plakoglobin. *The Journal of Biological Chemistry*, Vol. 277, No. 3, (January 2002), pp. 1884-1891, ISSN 0021-9258

Nash, P., Tang, X., Orlicky, S., Chen, Q., Gertler, F. B., Mendenhall, M. D., Sicheri, F., Pawson, T. & Tyers, M. (2001). Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. *Nature*, Vol. 414, No. 6863, (November 2001), pp. 514-521, ISSN 0028-0836.

Niefind, K., Guerra, B., Ermakowa, I. & Issinger, O.-G. (2000). Crystallization and preliminary characterization of crystals of human protein kinase CK2. *Acta Crystallographica. Section D, Biological Crystallography*, Vol. 56, No. Pt 12, (December 2000), pp. 1680-1684, ISSN 0907-4449.

Niefind, K., Guerra, B., Ermakowa, I. & Issinger, O.-G. (2001). Crystal structure of human protein kinase CK2: insights into basic properties of the CK2 holoenzyme. *The EMBO journal*, Vol. 20, No. 19, (October 2001), pp. 5320-5331, ISSN 0261-4189.

Niefind, K. & Issinger, O. G. (2010). Conformational plasticity of the catalytic subunit of protein kinase CK2 and its consequences for regulation and drug design. *Biochimica et Biophysica Acta*, Vol. 1804, No. 3, (March 2010), pp. 484-492, ISSN 0006-3002.

Niefind K., Raaf J. & Issinger O.-G. (2009) Protein kinase CK2 in health and disease: Protein kinase CK2: from structures to insights. *Cellular and Molecular Life Sciences*, Vol. 66, No. 11-12, (June 2009), pp. 1800-1816, ISSN 1420-682X.

Olsen, B. B., Guerra, B., Niefind, K. & Issinger, O. G. (2010). Structural basis of the constitutive activity of protein kinase CK2. *Methods in Enzymology*, Vol. 484, No. pp. 515-529, ISSN 1557-7988.

Olsen, B. B. & Guerra, B. (2008) Ability of CK2β to selectively regulate cellular protein kinases. *Molecular & Cellular Biochemistry*, Vol. 316, No. 1-2, (September 2008), pp. 115-126, ISSN 0300-8177.

Olsten, M.E., Litchfield D.W., (2004) Order or chaos? An evaluation of the regulation of protein kinase CK2. *Biochemistry & Cell Biology*, Vol. 82, No. 6, (December 2004), pp. 681–693, ISSN 0829-8211.

Olsten, M.E., Weber J.E., Litchfield D.W. (2005) CK2 interacting proteins: emerging paradigms for CK2 regulation? *Molecular & Cellular Biochemistry*, Vol. 274, No. 1-2, (June 2005), pp. 115-124, ISSN 0300-8177.

Pagano, M.O., Bain, J., Kazimierczuk, Z., Sarno, S., Ruzzene, M., Di Maria, G., Elliott, M., Orzeszko, A., Cozza, G., Meggio, F. & Pinna, L.A. (2008) The selectivity of inhibitors of protein kinase CK2: an update. *Biochemical Journal*, Vol. 415, No. 3, (November 2008), pp. 353-365, ISSN 0264-6021.

Pandit, S.B., Balaji, S., Srinivasan,N. (2004) Structural and functional characterization of gene products encoded in the human genome by homology detection. *IUBMB Life*, Vol. 56, No. 6, (June 2004), pp. 317-331, ISSN 1521-6543.

Pearson, M.A. & Fabbro, D. (2004) Targetting protein kinases in cancer therapy: a success? *Expert Review of Anticancer Therapy*, Vol. 4, No.6., (December 2004), pp. 1113-1124, ISSN 1473-7140.

Pinna, L. A. (2002). Protein kinase CK2: a challenge to canons. *Journal of Cell Science*, Vol. 115, Pt. 20, (October 2002), pp. 3873-3878, ISSN 0021-9533

Pinna, L. A. & Meggio, F. (1997). Protein kinase CK2 ("casein kinase-2") and its implication in cell division and proliferation. *Progress in Cell Cycle Research*, Vol. 3, No. pp. 77-97, ISSN 1087-2957

Plowman, G.D., Sudarsanam, S., Bingham, J., Whyte, D., Hunter, T., (1999) The protein kinases of *Caenorhabditis elegans*, a model for signal transduction in multicellular organisms. *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 96, No. 24, (November 1999), pp. 13603-13610, ISSN 0027-8424.

Ponce, D., Maturana, J. L., Cabello, P., Yefi, R., Niechi, I., Silva, E., Armisen, R., Galindo, M., Antonelli, M. & Tapia, J. C. (2011). Phosphorylation of AKT/PKB by CK2 is necessary for the AKT-dependent up-regulation of beta-catenin transcriptional activity. *Journal of Cellular Physiology*, Vol.226, No. 7, (July 2010), pp. 1953-1959, ISSN 1097-4652

Prudent, R., Sautel, C.F., Cochet, C. (2010) Structure-based discovery of small molecules targeting different surfaces of protein-kinase CK2. *Biochimica et Biophysica Acta*, Vol. 1804, No. 3, (March 2010), pp. 493-498, ISSN 0006-3002.

Raaf, J., Bischoff, N., Klopffleisch, K., Brunstein, E., Olsen, B. B., Vilk, G., Litchfield, D. W., Issinger, O.-G. & Niefind, K. (2011) Interaction between CK2alpha and CK2beta, the subunits of protein kinase CK2: thermodynamic contributions of key residues on the CK2alpha surface. *Biochemistry*, Vol. 50, No. 4, (February 2011), pp. 512-522, ISSN 1520-4995.

Raaf, J., Brunstein, E., Issinger, O.-G., Niefind, K. (2008) The interaction of CK2alpha and CK2beta, the subunits of protein kinase CK2, requires CK2beta in a preformed conformation and is enthalpically driven. *Protein Science*, Vol. 17, No. 12, (December 2008), pp. 2180-2186, ISSN 0961-8368.

Riera, M., Peracchia, G., de Nadal, E., Ariño, J., Pagès, M. (2001) Maize protein kinase **CK2**: regulation and functionality of three beta regulatory subunits. *The Plant Journal*, Vol. 25, No. 4, (February 2001), pp. 365-374, ISSN 0960-7412.

Rodriguez, F. A., Contreras, C., Bolanos-Garcia, V. & Allende, J. E. (2008). Protein kinase CK2 as an ectokinase: the role of the regulatory CK2beta subunit. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 105, No. 15, (April 2008), pp. 5693-5698, ISSN 1091-6490

Roussou, I, Draetta, G. (1994) The *Schizosaccharomyces pombe* casein kinase II alpha and beta subunits: evolutionary conservation and positive role of the beta subunit. *Molecular and Cellular Biology*, Vol. 14, No. 1, (January 1994), pp. 576-586, ISSN 0270-7306.

Sajnaga, E., Kubiński, K. & Szyszka, R. (2008). Catalytic activity of mutants of yeast protein kinase CK2alpha. *Acta Biochimica Polonica*, Vol. 55, No. 4, (November 2008), pp. 767-776, ISSN 0001-527X.

Salamon, J. A., Acuna, R. & Dawe, A. L. (2010). Phosphorylation of phosducin-like protein BDM-1 by protein kinase 2 (CK2) is required for virulence and G beta subunit stability in the fungal plant pathogen Cryphonectria parasitica. *Molecular Microbiology*, Vol. 76, No. 4, (May 2010), pp. 848-860, ISSN 1365-2958.

Sarno, S., Boldyreff, B., Marin, O., Guerra, B., Meggio, F., Issinger, O.-G. & Pinna, L. A. (1995). Mapping the residues of protein kinase CK2 implicated in substrate recognition: mutagenesis of conserved basic residues in the alpha-subunit.

Biochemical and Biophysical Research Communications, Vol. 206, No. 1, (January 1995), pp. 171-179, ISSN 0006-291X

Sarno, S., Ghisellini, P., Cesaro, L., Battistutta, R. & Pinna, L.A. (2001). Generation of mutants of CK2alpha which are dependent on the beta-subunit for catalytic activity. *Molecular & Cellular Biochemistry*, Vol. 227, No. 1-2, (November 2001), pp. 13-19, ISSN 0300-8177.

Sarno, S., Ghisellini, P. & Pinna, L.A. (2002). Unique activation mechanism of protein kinase CK2. The N-terminal segment is essential for constitutive activity of the catalytic subunit but not of the holoenzyme. *The Journal of Biological Chemistry*, Vol. 277, No. 25, (June 2002), pp. 22509-22514, ISSN 0021-9258.

Sarno, S., Marin, O., Boschetti, M., Pagano, M.A., Meggio, F., Pinna, L.A. (2000) Cooperative modulation of protein kinase CK2 by separate domains of its regulatory beta-subunit. *Biochemistry*, Vol. 39, No.40, (October 10), pp. 12324-12329, ISSN 0006-2960.

Sarno, S., Ruzzene, M., Frascella, P., Pagano, M. A., Meggio, F., Zambon, A., Mazzorana, M., Di Maira, G., Lucchini, V. & Pinna, L. A. (2005). Development and exploitation of CK2 inhibitors. *Molecular & Cellular Biochemistry*, Vol. 274, No. 1-2, (June 2005), pp. 69-76, ISSN 0300-8177.

Sarno, S., Salvi, M., Battistutta, R., Zanotti, G. & Pinna, L.A. (2005). Features and potentials of ATP-site directed CK2 inhibitors. *Biochimica et Biophysica Acta*, Vol. 1754, No. 1-2, (December 2005), pp. 263-270, ISSN 0006-3002.

Sarno, S., Vaglio, P., Cesaro, L., Marin, O. & Pinna, L.A. (1999). A multifunctional network of basic residues confers unique properties to protein kinase CK2. *Molecular & Cellular Biochemistry*, Vol. 191, No. 1-2, (January 1999), pp. 13-19, ISSN 0300-8177.

Sarno, S., Vaglio, P., Marin, O., Meggio, F., Issinger, O.-G. & Pinna, L.A. (1997a). Basic residues in the 74-83 and 191-198 segments of protein kinase CK2 catalytic subunit are implicated in negative but not in positive regulation by the beta-subunit. *European Journal of Biochemistry*, Vol. 248, No. 2, (September 1997), pp. 290-295, ISSN 0014-2956.

Sarno, S., Vaglio, P., Marin, O., Issinger, O. G., Ruffato, K. & Pinna, L.A. (1997b). Mutational analysis of residues implicated in the interaction between protein kinase CK2 and peptide substrates. *Biochemistry*, Vol. 36, No. 39, (September 1997), pp. 11717-11724, ISSN 0006-2960.

Sarno, S., Vaglio, P., Meggio, F., Issinger, O. G. & Pinna, L. A. (1996). Protein kinase CK2 mutants defective in substrate recognition. Purification and kinetic analysis. *The Journal of Biological Chemistry*, Vol. 271, No. 18, (May 1996), pp. 10595-10601, ISSN 0021-9258.

Seldin, D. C., Landesman-Bollag, E., Farago, M., Currier, N., Lou, D. & Dominguez, I. (2005). CK2 as a positive regulator of Wnt signalling and tumourigenesis. *Molecular & Cellular Biochemistry*, Vol. 274, No. 1-2, (June 2005), pp. 63-67, ISSN 0300-8177.

Shi, X., Potvin, B., Huang, T., Hilgard, P., Spray, D.C., Suadicani, S.O., Wolkoff, A.W., Stanley, P. and Stockert, R.J. (2001) A novel casein kinase 2 alpha-subunit regulates membrane protein traffic in the human hepatoma cell line HuH-7. *The Journal of Biological Chemistry*, Vol. 276, No. 3, (January 2001), pp. 2075-2082, ISSN 0021-9258 276: 2075–2082.

Song, D. H., Dominguez, I., Mizuno, J., Kaut, M., Mohr, S. C. & Seldin, D. C. (2003). CK2 phosphorylation of the armadillo repeat region of beta-catenin potentiates Wnt signaling. *The Journal of Biological Chemistry*, Vol. 278, No. 26, (June 2003), pp. 24018-24025, ISSN 0021-9258.

Srinivasan, N., Antonelli, M., Jacob, G., Korn, I., Romero, F., Jedlicki, A., Dhanaraj, V., Sayed, M. F., Blundell, T. L., Allende, C. C. & Allende, J. E. (1999). Structural interpretation of site-directed mutagenesis and specificity of the catalytic subunit of protein kinase CK2 using comparative modelling. *Protein Engineering*, Vol. 12, No. 2, (February 1999), pp. 119-127, ISSN 0269-2139.

St-Denis, N.A., Derksen, D.R. & Litchfield, D.W. (2009) Evidence for regulation of mitotic progression through temporal phosphorylation and dephosphorylation of CK2alpha. *Molecular & Cellular Biology*, Vol. 29, No. 8, (April), pp. 2068-2081, ISSN 0270-7306.

Trembley J.H., Wang G., Unger G., Slaton J. & Ahmed K. (2009) CK2: a key player in cancer biology. *Cellular and Molecular Life Sciences*, Vol. 66, No 11-12, (June 2009), pp. 1858-1867, ISSN 1420-682X.

Tuazon, P. T. & Traugh, J. A. (1991). Casein kinase I and II--multipotential serine protein kinases: structure, function, and regulation. *Advances in Second Messenger and Phosphoprotein Research*, Vol. 23, No. (January 1991), pp. 123-164, ISSN 1040-7952.

Vaglio, P., Sarno, S., Marin, O., Meggio, F., Issinger, O. G. & Pinna, L. A. (1996). Mapping the residues of protein kinase CK2 alpha subunit responsible for responsiveness to polyanionic inhibitors. *FEBS letters*, Vol. 380, No. 1-2, (February 1996), pp. 25-28, ISSN 0014-5793.

Verma, R., McDonald, H., Yates, J. R., 3rd & Deshaies, R. J. (2001). Selective degradation of ubiquitinated Sic1 by purified 26S proteasome yields active S phase cyclin-Cdk. *Molecular Cell*, Vol. 8, No. 2, (August 2001), pp. 439-448, ISSN 1097-2765.

Vilk, G., Derksen, D. R. & Litchfield, D. W. (2001). Inducible expression of the regulatory protein kinase CK2beta subunit: incorporation into complexes with catalytic CK2 subunits and re-examination of the effects of CK2beta on cell proliferation. *Journal of Cellular Biochemistry*, Vol. 84, No. 1, pp. 84-99, ISSN 0730-2312.

Vilk G., Weber J.E., Turowiec J.P., Duncan J.S., Wu C., Derksen D.R., Zień P., Sarno S., Donella-Deana A., Lajoie G., Pinna L.A., Li S.S. and Litchfield D.W. (2008) Protein kinase CK2 catalyzes tyrosine phosphorylation in mammalian cells. *Cellular Signalling*, Vol. 20, No. 11, (November 2008), pp. 1942-1951, ISSN 0898-6568.

Wirkner, U., Voss, H., Lichter, P., Pyerin, W. (1994) Human protein kinase CK2 genes. *Cellular & Molecular Biology Research*, Vol. 40, No. 5-6, pp 489-99., ISSN 0968-8773.

Yde, C. W., Ermakova, I., Issinger, O. G. & Niefind, K. (2005). Inclining the purine base binding plane in protein kinase CK2 by exchanging the flanking side-chains generates a preference for ATP as a cosubstrate. *Journal of Molecular Biology*, Vol. 347, No. 2, (March 2005), pp. 399-414, ISSN 0022-2836

Zhang, C., Vilk, G., Canton, D. A. & Litchfield, D. W. (2002). Phosphorylation regulates the stability of the regulatory CK2beta subunit. *Oncogene*, Vol. 21, No. 23, (May 2002), pp. 3754-3764, ISSN 0950-9232

Zhang, N. & Zhong, R. (2010). Structural basis for decreased affinity of Emodin binding to Val66-mutated human CK2 alpha as determined by molecular dynamics. *Journal of Molecular Modeling*, Vol. 16, No. 4, (April 2010), pp. 771-780, ISSN 0948-5023

Zhang, Y. & Dong, C. (2007) Regulatory mechanisms of mitogen-activated kinase signaling. *Cellular & Molecular Life Sciences*, Vol. 64, No. 21, (November 2007), pp. 2771-2789, ISSN 1420-682X.

Directed Mutagenesis in Structure Activity Studies of Neurotransmitter Transporters

Jane E. Carland, Amelia R. Edington, Amanda J. Scopelliti,
Renae M. Ryan and Robert J. Vandenberg
Department of Pharmacology, The University of Sydney
Australia

1. Introduction

The delicate balance between excitation and inhibition within the central nervous system is critical to the maintenance of normal brain function. Players key to this balance are neurotransmitter transporters. Neurotransmitter transporters are drawn from two families of solute carriers (SLC), SLC1 and SLC6. The transporters for glutamate and small neutral amino acids belong to the SLC1 family, while transport of monoamines (5-hydroxytryptamine, dopamine, noradrenaline) and amino acid neurotransmitters (γ-aminobutyric acid, glycine) are mediated by members of the SLC6 family. These integral membrane proteins regulate the concentration of neurotransmitters, such as glutamate and glycine, within the synapse. They utilise pre-existing electrochemical gradients to drive the transport of neurotransmitters across neuronal and glial membranes, terminating neurotransmission and replenishing intracellular levels of neurotransmitter for future release. Neurotransmitter transporters are targeted by a number of substances, both therapeutic (antidepressants, anticonvulsant, antipsychotics, analgesics, anxiolytics) and addictive (cocaine, methampetamine). Their dysfunction is associated with multiple disorders, including epilepsy, ischaemic stroke, neuropathic pain and schizophrenia (Dohi et al., 2009; Sur & Kinney, 2004). Thus, structure activity studies of transporters are essential to provide new insights into their function and direct the design of novel, transporter-specific therapeutics.

Since the cloning of the GABA transporter, GAT1, in 1990 (Guastella et al., 1990), directed mutagenesis studies have underpinned our understanding of the secondary structure and function of neurotransmitter transporters. This work has subsequently been supported, and significantly advanced, by the high resolution crystal structures of prokaryotic homologues of the SLC1 and SLC6 families. The crystal structure of the SLC1 homologue from *Pyrococcus horikoshii* (Glt$_{Ph}$) was the first to be solved at 3.5 Å resolution (Fig 1A) (Yernool et al., 2004). This was followed in 2005 by the crystal structure of a homologue of the SLC6 family from *Aquifex aeolicus (*LeuT$_{Aa}$) at 1.65 Å resolution (Yamashita et al., 2005) (Fig 3A). These crystal structures have provided important insights into the interactions of transporters with substrates, ions, lipids and inhibitors, allowing the postulation of numerous functional mechanisms. However, the details provided by these high-resolution structures are insufficient to fully understand transport mechanisms. The knowledge obtained from

crystal structures and 3D models can be used to direct mutagenesis work (utilising chimeric transporters and site-directed mutagenesis), electrophysiology and uptake studies to determine the molecular basis for transporter function. The techniques described in this chapter can be applied to other membrane proteins, such as G-protein coupled receptors and ligand-gated ion channels.

1.1 The SLC1 family of transporters

The SLC1 family of neurotransmitter transporters includes five human excitatory amino acid transporters (EAAT1 to EAAT5) (Slotboom et al., 1999) and two neutral amino acid transporters (ASCT1 and ASCT2) (Arriza et al., 1993; Kanai & Hediger, 2004). The EAATs exhibit 40-44% sequence identity with ASCTs and approximately 36% sequence identity with the related Na^+-coupled aspartate transporter, Glt_{Ph} (Fig. 2). The high resolution crystal structure of Glt_{Ph} reveals that SLC1 transporters exist as bowl-shaped trimers (Fig. 1A) (Yernool et al., 2004). Each protomer is comprised of eight transmembrane domains (TM1-8) and two re-entrant hairpin loops (HP1 and HP2). TM1, TM2, TM4, and TM5 mediate intersubunit contacts and support the "transport" domain, which is composed of TM3, TM6, TM7, TM8, HP1 and HP2 (Fig. 1B). This transport domain mediates substrate and ion translocation, and each protomer has an independent translocation pathway.

The EAATs are critical components of excitatory synapses, where they mediate the high affinity uptake of the dominant excitatory neurotransmitter, glutamate, as well as L- and D-aspartate. Both EAAT1 and EAAT2 are expressed in glia. Of these two subtypes, EAAT2 is the more widely distributed and is the major regulator of glutamate concentrations in the central nervous system. EAAT3 is expressed on neuronal membranes throughout the brain, while EAAT4 is selectively expressed on cerebellar Purkinje cells. EAAT5 is expressed on retinal neurons. Glutamate uptake is coupled to the co-transport of three Na^+ ions and one H^+ and the counter-transport of one K^+ ion, rendering them electrogenic (Zerangue & Kavanaugh, 1996). In addition, glutamate transport is associated with a thermodynamically uncoupled Cl^- conductance (Fig. 1C) (Fairman et al., 1995, Wadiche et al., 1995).

1.2 The SLC6 family of transporters

Members of the SLC6 transporter family are responsible for the transport of monoamine (dopamine, serotonin/5-hydroxytryptamine, noradrenaline) and amino acid (GABA and glycine) neurotransmitters across cell membranes. Two glycine transporters (GLYT1 and GLYT2) have been cloned, and five GLYT1 splice variants (GLYT1a to GLYT1e) and three GlyT2 splice variants (GLYT2a to GLYT2c) have been identified. The crystal structure of the prokaryotic transporter $LeuT_{Aa}$ serves as a useful template for unravelling the functional implications of transporter structures (FIG. 3). Members of the SLC6 family are traditionally thought to exist as monomers (Horiuchi et al., 2001; Lopez-Corcuera et al., 1993), although more recent work suggests that they may form dimers *in vivo* (Bartholomaus et al., 2008). Each subunit is formed by twelve transmembrane domains (TM1-12), with amino- and carboxy-termini located on the intracellular side of the membrane. Each subunit exhibits a two-fold axis of symmetry, with TM1 to TM5 corresponding to TM6 to TM10 with an inverted topology repeat. TM1 to TM10 form the core of the transporter, with TM1 and TM6 exhibiting the highest degree of sequence homology. They run antiparallel and exist with a central unwound section. The area surrounding this central unwound section is critical for substrate and ion binding.
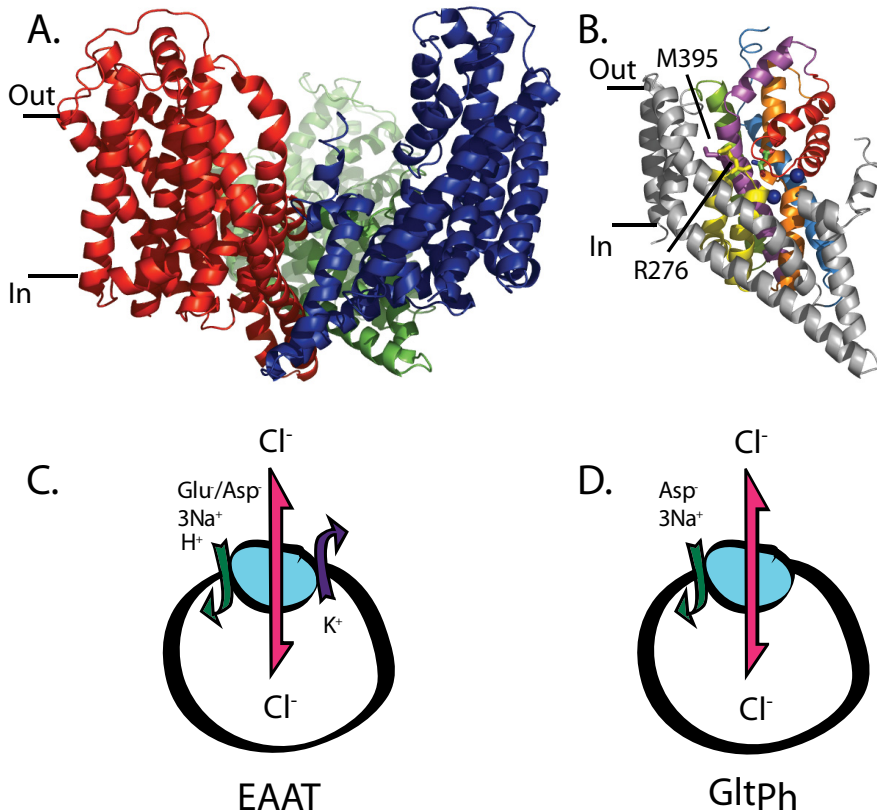
Fig. 1. The crystal structure of the prokaryotic transporter Glt_{Ph} and the stoichiometry of transport by the excitatory amino acid transporters (EAAT1- EAAT5) and Glt_{Ph}. A. Glt_{Ph} is a bowl shaped trimer viewed in the plane of the membrane. Individual protomers are coloured red, green and blue. B. A single protomer of the Glt_{Ph} trimer (PDB 2NWX). The C-terminal domain is shown in colour; HP1 (yellow), TM7 (orange), HP2 (red) and TM8 (magenta). Bound aspartate is shown in stick representation and two Na^+ ions are shown as blue spheres. R276 (HP1) and M395 (TM8), which are discussed in section 3.5, are also shown in stick representation. Structures were viewed and rendered in PyMOL (http://www.pymol.org) (Schrodinger, 2010). C. Glutamate or aspartate transport via the EAATs is coupled to three Na^+ ions and one H^+, followed by the counter-transport of one K^+ ion. Binding of Na^+ and substrate to the EAATs activates a thermodynamically uncoupled Cl^- conductance (*pink arrow*). D. Aspartate transport via Glt_{Ph} is coupled to the co-transport of three Na^+ ions, but is not coupled to the movement of either H^+ or K^+ ions. Na^+ and aspartate binding to Glt_{Ph} also activates an uncoupled Cl^- conductance (*pink arrow*).
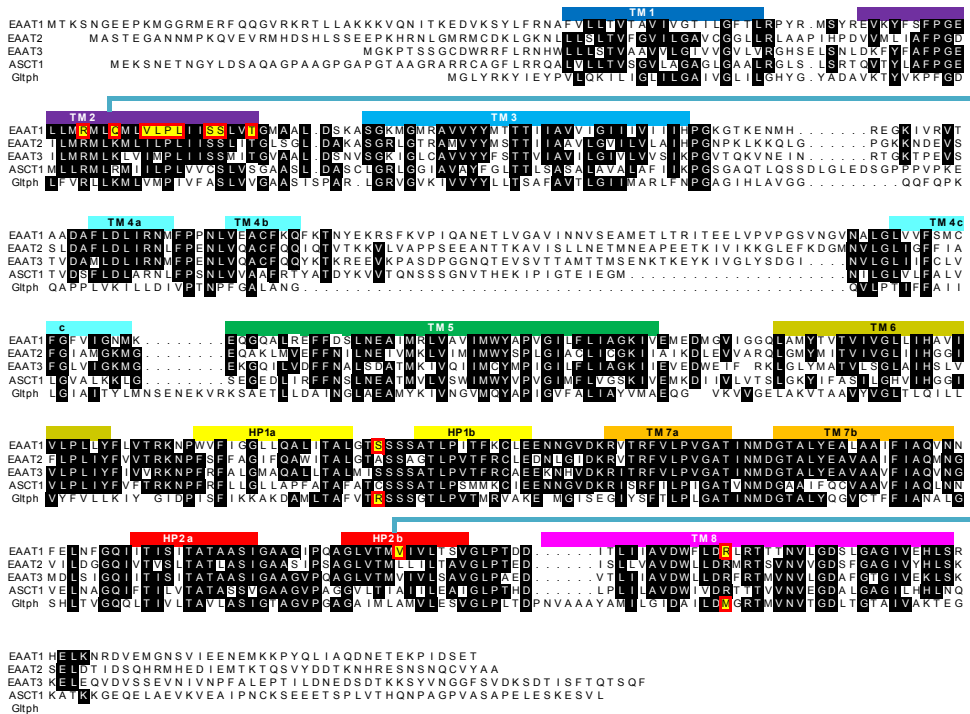
Fig. 2. Amino acid sequence alignment of SLC1 family members. Sequences for EAAT1-3, ASCT1 and Glt$_{Ph}$ are shown. Transmembrane domains are indicated using the colour scheme as for the structure of Glt$_{Ph}$ in Fig. 1. Homologous regions are highlighted in black. Residues highlighted in red boxes with yellow background are discussed in the text. The blue line (connecting Q93 to V452 in EAAT1) indicates that cysteine mutants of these two residues can be cross-linked (see section 3.4 for details).
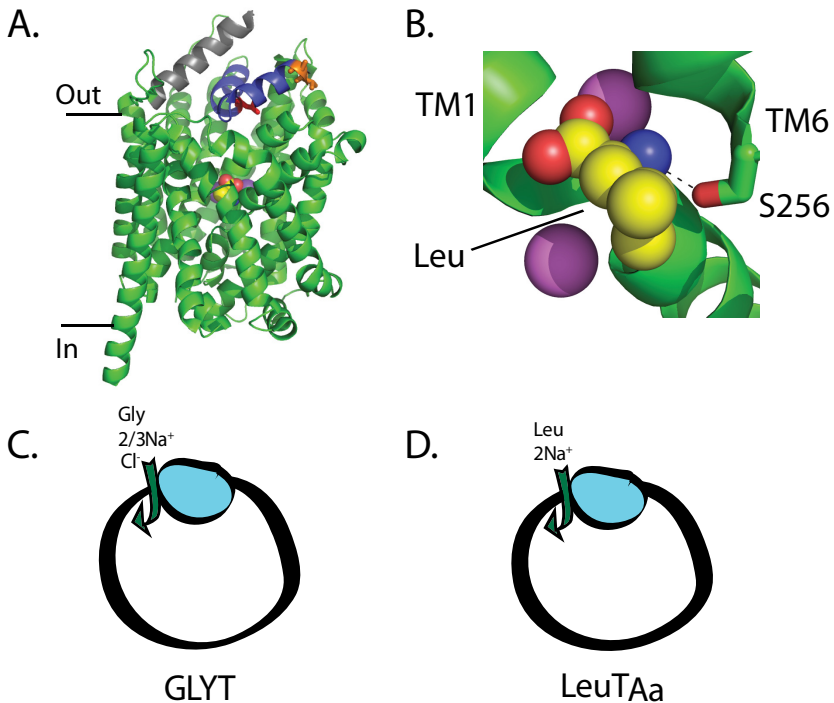
Fig. 3. The crystal structure of the prokaryotic transporter LeuT$_{Aa}$ and the stoichiometry of transport by the GLYTs and LeuT$_{Aa}$. A. The structure of LeuT$_{Aa}$ viewed in the plane of the membrane (PDB 2A65). Bound leucine is shown in A space-filling representation, and the 2 Na$^+$ ions are shown as purple spheres. Extracellular loop 2 (EL2, grey) and extracellular loop 4 (EL4, blue) are highlighted (see section 2.2 for details). EL4 residues R531 and K532 (orange stick representation) and I545 (red stick representation) are also highlighted (see section 3.2 for discussion). B. Bound leucine interacts with transmembrane domains TM1 and TM6. The hydrogen bond between the nitrogen atom of leucine and the side chain of serine 265 is represented by a dashed line (see section 3.2 for details). Na$^+$ ions are shown as purple spheres. C. Glycine transport via the GLYTs is coupled to two (GLYT1) or three (GLYT2) Na$^+$ ions and one Cl$^-$ ion. D. Alanine transport via LeuT$_{Aa}$ is coupled to the co-transport of two Na$^+$ ions.

Mammalian members of the SLC6 family share 20-25% sequence identity with the prokaryotic transporter LeuT$_{Aa}$ (FIG. 4). GLYT1 and GLYT2 are structurally similar and exhibit 48% sequence homology (FIG. 4), but they display significant functional differences. GLYT1 is predominantly expressed in glia at excitatory glutamatergic synapses, where they are responsible for regulating glycine, which acts as a co-agonist at NMDA receptors. Of the five GLYT1 isomers identified, GLYT1b and GLYT1c are nervous system-specific. In contrast, GLYT2 is typically localized with glycine receptors at inhibitory glycinergic synapses (spinal cord). The translocation of glycine by both transporters is coupled to the co-transport of Na$^+$ and Cl$^-$ (FIG. 3), but the stoichiometry of ion-flux coupling by GLYT1 and GLYT2 differs. The transport of one glycine molecule is coupled to the co-transport of

two Na+ ions and one Cl- ion for GLYT1 transporters, while the movement of three Na+ ions and one Cl- ion is coupled to glycine transport for GLYT2 (Fig. 3C).
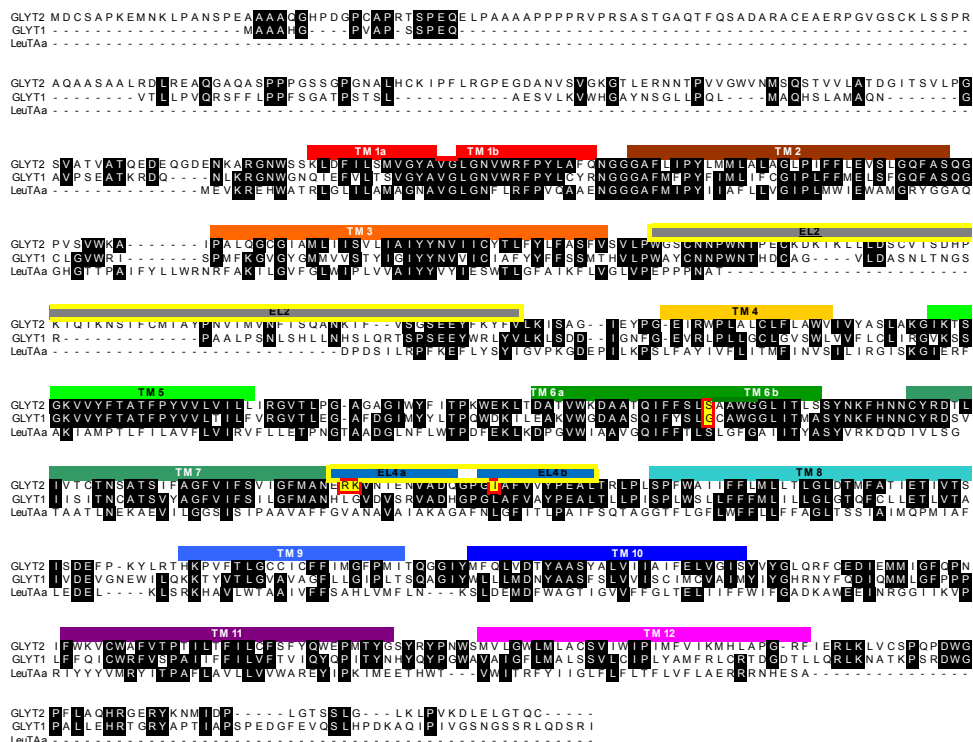
Fig. 4. Amino acid sequence alignment of GLYT1, GLYT2 and LeuT$_{Aa}$. Transmembrane domains are indicated using the colour scheme used for the structure of LeuT$_{Aa}$. Homologous regions are highlighted in black. Chimeric transporters were generated between GLYT1 and GLYT2 in which extracellular loops 2 and 4 (highlighted with yellow edges) were switched between the two transporters (see section 2.2 for details). Residues highlighted in red boxes with yellow background are discussed in the text.

2. The use of chimeras in studies of neurotransmitter transporters

Chimeras provide an excellent tool for the study of neurotransmitter transporters. Switching specific regions/structures between transporters can provide insights into transporter function and substrate selectivity and inform the design of directed mutagenesis studies. Further, chimeras between mammalian and bacterial transporters have the potential to assist with the crystallisation of transporters, thereby facilitating the determination of the structure of mammalian transporters. We employ a fusion PCR technique to create chimeric junctions at specific amino acid locations, allowing for the precise design of chimeras. We have produced numerous chimeras using this method, including chimeras of GLYT1 and GLYT2.

2.1 PCR fusion methodology

Conventional chimera construction methodology relies on restriction enzyme cloning, in which unique restriction enzyme sites are introduced into both the acceptor and the donor proteins. While this technique allows for the production of chimeras with specific/known junction points, the availability of unique restriction enzyme sites can impose limitations on the design of potential chimeras. In contrast, the PCR fusion technique (Shevchuk et al., 2004) creates chimeric junctions at any amino acid, without the need for restriction enzyme sites. In this method (Fig. 5), each segment of the final chimera is amplified in individual PCR reactions. The primers are designed to engineer complementary overlapping sequences onto the junction-forming ends of each product. The PCR primers possess typical properties (18-24 nucleotides in length and a melting temperature of ~64°C). The overlapping sequences correspond to the desired chimeric junction sites between subunits. The strands of the PCR products (duplex DNAs) are routinely separated and allowed to reanneal by cycles of heating and cooling. A partially duplex chimera can form as a result of annealing by the complementary regions of different fragments, one half of the chimeras will have free 3'-ends that can be elongated by DNA polymerase. The other half will have free 5'-ends that cannot be elongated. If more than two PCR products are involved, eventually a full-length chimera forms. The resulting duplex chimera is then amplified by PCR with oligonucleotide primers containing restriction enzyme sites at the 5' and 3' ends of the DNA of the chimera to facilitate subcloning into a suitable vector. For chimeras that require more than three fragments, it is often best to produce an initial chimera of three fragments and then incorporate additional fragments as required. It is important to obtain complete DNA sequences of any clones generated in this manner to confirm the junction sites and also to ensure that there have been no spurious sequence changes.

2.2 Identifying determinants of drug selectivity

Drugs that have selective effects *in vivo* are much sought after. Such compounds potentially have minimal side effects, making them attractive options as therapeutics. Studies suggest that the GLYTs may provide a novel therapeutic target for the development of drugs to treat neurological disorders and pain. In particular, GLYT1 is considered a potential target for the development of agents to treat schizophrenia (Sur et al., 2004). GLYT2 is a key target for studies to develop molecules to treat chronic pain (Aragon & Lopez-Corcuera, 2003). *N*-Arachidonylglycine (NAGly) is an endogenous derivative of arachidonic acid. NAGly has been shown to induce analgesia in rat models of neuropathic and inflammatory pain (Succar et al., 2007; Vuong et al., 2008). One of the mechanisms of action of NAGly is the inhibition of GLYT2 (IC_{30} = 3 µM), whilst it has no effect on GLYT1 (Wiles et al., 2006). Understanding the molecular basis of NAGly selectivity may aid in the development of novel analgesic compounds.

Extracellular loops EL2 and EL4 have been implicated in mediating inhibition of the GLYTs. Residues that contribute to the binding site of Zn^{2+}, a non-competitive inhibitor of GLYT1, have been identified in EL2 and EL4 (Ju et al., 2004). Zn^{2+} has been proposed to inhibit GLYT1 by binding to EL2 and EL4, restricting the movement of these loops and thus preventing glycine transport. LeuT$_{Aa}$, the bacterial homologue of the GLYTs, has been crystallized in the presence of clomipramine, a non-competitive inhibitor (Singh et al., 2007), and tryptophan, a competitive inhibitor (Singh et al., 2008). Clomipramine was shown to
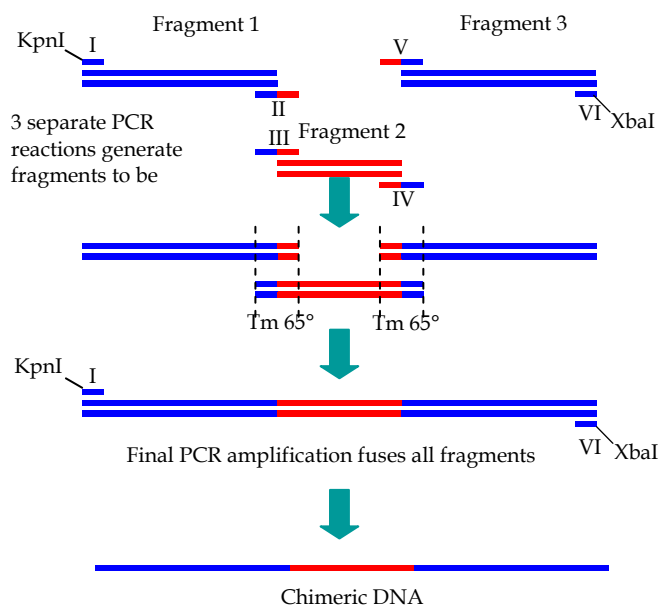
Fig. 5. A schematic summary of the construction of chimera GLYT1(EL2) using the PCR fusion methodology. The DNA sequence of the donor cDNA (GLYT1) is in blue and the acceptor cDNA (GLYT2) is in red. cDNAs of GLYT1 and GLYT2 in the presence of their corresponding primers (I and II, III and IV, V and VI) undergo PCR to generate three fragments with the appropriate homologous ends. The fragments are 'fused' together in another reaction, in which the single strands from the overlapping regions serve as internal primers (see text). The final PCR amplification reaction fuses ('zips') all the fragments in the presence of primers I and VI, which, in this case, include restriction site for the enzymes *Kpn*I and *Xba*I. The final product is the chimera GLYT1(EL2) with two unique restriction sites engineered on either end to allow for insertion of the chimeric gene into the vector pOTV. The restriction sites may be altered for subcloning into different vectors.

stabilize LeuT$_{Aa}$ in an occluded state by interacting with a number of transmembrane domains and displacing the tip of EL4. In contrast, the presence of the competitive inhibitor tryptophan appeared to trap the transporter in an open-to-out conformation. Four separate tryptophan molecules were identified in this crystal structure, with one of them found to be interacting with EL2 and EL4. These observations suggest that these loops play key roles in the inhibition of GLYTs.

In order to ascertain the molecular basis for the inhibitory activity of NAGly on GLYT2, chimeras were generated between GLYT1 and GLYT2, in which EL2 and/or EL4 were switched (Fig. 6). A chimera is named according to its parental transporter, with the inserted loop in parentheses. For example, GLYT2(EL2) is predominantly GLYT2 with the EL2 of GLYT1. One of the important controls required when using a chimeric protein to understand structure-function relationships is the ability to maintain the functional properties of the chimera. For chimeras that are predominantly GLYT2, the EC$_{50}$ for glycine

was very similar to that of the parental GLYT2 transporter, which indicates that these chimeras transport glycine similar to GLYT2. With the GLYT1-based chimeras, GLYT1(EL2) and GLYT1(EL2, EL4), the EC_{50} for glycine was increased by 6-8 fold compared to GLYT1. Thus, for the GLYT1-based chimeras, the transport of glycine differs slightly from the parent transporter; and, therefore, we are not able to be as confident that other functional changes are solely due to the region of interest. Nevertheless, the experiments using the GLYT2-based chimeras yielded valuable information concerning the domains responsible for NAGly sensitivity. GLYT2 is inhibited by NAGly, whereas GLYT2(EL2) and GLYT2(EL4) have reduced sensitivity (Edington et al., 2009). These observations suggest that EL2 and EL4 contribute to forming an inhibitory site on GLYT2.
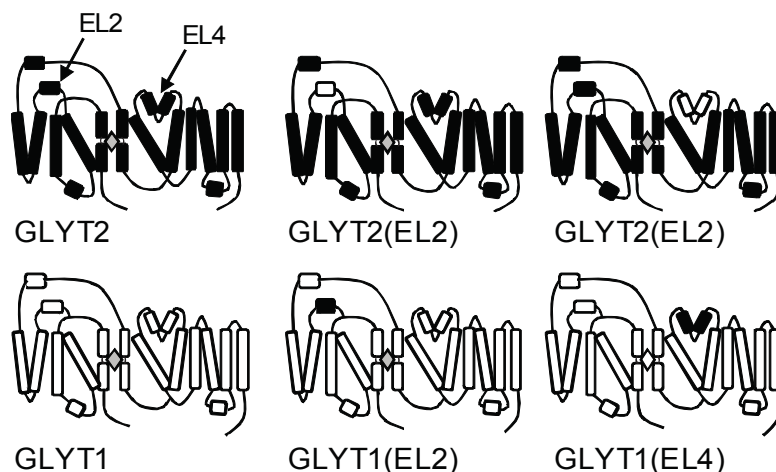
Fig. 6. A schematic diagram of the topology of the wild-type transporters GLYT2 and GLYT1 and their EL2 and EL4 chimeras. *Black* indicates the regions of the transporter from GLYT2 and *white* indicates the regions from GLYT1. Shown are GLYT2, GLYT2(EL2), GLYT2(EL4), GLYT1, GLYT1(EL2), and GLYT1(EL4).

3. Site-directed mutagenesis

Identifying domains that are responsible for conferring functional differences between transporters is only the first step in understanding the molecular processes that dictate neurotransmitter transporter function. Chimeric and directed mutagenesis studies are mutually beneficial and, when used in conjunction, can increase the efficiency and efficacy of experimentation. Information obtained in the study of chimeric transporters helps to focus site-directed mutagenesis studies to specific domains. The introduction of point mutations within these regions can then provide very useful information about the function of specific residues and the location of substrate and ion binding sites. Conventional mutagenesis studies utilise amino acid sequence alignments between multiple members of a transporter family to direct and design mutagenesis studies. However, this work often relies on the mutation knocking out a particular function and thereby assigning that function to the residue of interest. This approach can generate misleading conclusions because loss of function can be the result of a number of changes, many of which do not necessarily reflect a

direct disruption of the interaction being investigated. To avoid this issue, we employ knowledge obtained from recent advances in high resolution transporter crystal structures and homology models to improve our accuracy in predicting residues of interest and successfully creating functional mutants. Crystal structures provide a powerful 3-dimensional tool that enables us to visualise each residue and their individual contacts with the substrate/ions/other residues, thus improving the selection process. In addition, homology models provide a computational prediction of the global effect of a mutation. Thus, they can be used to predict which mutations will be accommodated by individual transporters. We have undertaken numerous mutagenesis studies that have, among other things, identified the glycine substrate binding site in GLYT1 and GLYT2, helped to characterise selective drug-binding sites on the GLYTs and clarified the specificity of ion and substrate interactions with the EAATs and Glt$_{Ph}$.

3.1 Directed mutagenesis methodology

Site-directed mutagenesis is a powerful tool that makes select changes to the genetic code of a protein to alter the amino acid sequence. The resultant mutant protein exhibits subtle structural changes. The ability to selectively manipulate amino acids in a controlled manner provides a powerful tool that is exploited by researchers undertaking structure-function studies. It should be noted that the importance of loss of function mutants must be interpreted with caution. Disruption of transport functions following a mutation can result in loss of function. However, loss of function could also result from misfolding of a protein or altered expression levels. Therefore, mutants that lose function are only useful if the cause of loss of function can be accurately ascertained.

Site-directed mutagenesis involves the use of specifically designed primers (sense and antisense) that include a point mutation of interest in the centre of the sequence. Each primer is 18-24 nucleotides in length with a melting temperature of ~64°C. The primers incorporate the mutation into an intronless gene, using the cDNA as template in a PCR reaction, with the elongation by DNA polymerase. Following amplification, the endonuclease *Dpn*I, which recognizes methylated and hemimethylated DNA, can be added to the amplification reaction to digest the parental cDNA. It is important to obtain DNA sequences to confirm the mutation. The mutant DNA is subcloned into the appropriate vector. In the studies we describe, the Oocyte Transcription Vector (pOTV) is used, as it enables efficient RNA production and facilitates high expression levels of protein in *Xenopus laevis* oocytes. To construct a double point mutation, one set of primers may be used when the two residues are next to each other or in close proximity. However, if the two mutations are further apart, DNA encoding one mutation is initially made and used as the template for the second mutation. Site-directed mutagenesis kits are sold by many companies (e.g., Stratagene, Promega, Clonetech). We routinely use the QUIKCHANGE(tm) Site-Directed Mutagenesis kit from Stratagene.

3.2 Mutagenesis to identify determinants of drug selectivity of glycine transporters

The use of chimeric GLYT transporters allowed the identification of EL2 and EL4 as being regions critical for determining the selective activity of NAGly on GLYT2 (Edington et al., 2009). Having identified the regions of interest, we used site-directed mutagenesis to identify the residues within these regions that form the selective drug-binding site. Our

chimeric study (see Section 2.2) revealed that NAGly interacted with EL4, so the aim of the subsequent mutagenesis study was to identify the key residues in EL4 that play a role in determining the differential sensitivity of GLYT2 compared to GLYT1.

The EL4 of GLYT2 exhibits 60% sequence identity with the GLYT1 EL4. Point mutations were introduced at all positions within the GLYT2 EL4 that differed from GLYT1. In total, eleven mutations were produced. Each resulted in a functional mutant GLYT2 transporter that exhibited glycine transport unchanged from that of the wild-type transporter. Three mutations resulted in changes in sensitivity to NAGly. Mutation of arginine at position 531 to leucine (R531L) and lysine at position 532 to glycine (K532G) resulted in modest reductions in NAGly sensitivity ($IC_{30} = 13 \pm 2$ µM and 9 ± 1 µM, respectively) compared to GLYT2 ($IC_{30} = 3.4 \pm 0.6$ µM). In contrast, the mutation of isoleucine at position 545 to leucine (I545L) markedly reduced NAGly sensitivity ($IC_{30} > 30$µM) (Edington et al., 2009).

This work revealed that three residues within EL4 are critical to the inhibitory activity of NAGly on GLYT2. Modelling studies place I545 in the middle of EL4, while R531 and K532 are located at the edge of EL4. It is surprising that a conservative mutation at position 545 (an isoleucine to a leucine) would have the most dramatic impact. Two likely possibilities follow. (1) NAGly may bind to I545, and the I545L mutation may distort the way that NAGly fits into the binding site. (2) The I545L mutation may alter the conformation of the two arms of EL4, which then impacts on the way that this domain interacts with other elements that may be crucial for NAGly binding. The carboxyl groups of NAGly may interact with the positively charged R531 and K532 residues. Further structural studies are required to fully characterize the specific interaction sites between the NAGly and GLYT2.

3.3 Substrate selectivity of the GLYTs

GLYT1 and GLYT2 can be differentiated by their substrate selectivity and inhibitor sensitivity. GLYT1 transports both glycine and the N-methyl derivative of glycine, sarcosine, while GLYT2 only transports glycine (Supplisson & Bergman, 1997). In addition, GLYT1 is selectively inhibited by *N*[3-(4-fluorophenyl)-3-(4'-phenylphenoxy) propylsarcosine (NFPS) (Aubrey & Vandenberg, 2001). The crystal structure of LeuT$_{Aa}$ provides a good working model for the study of GLYTs. As with the GLYTs, the substrate binding site of LeuT$_{Aa}$ is formed at the junction between TM1–5 and TM6–10. It is composed of amino acid residues from TM1 and TM6 (Yamashita et al., 2005). Both of these transmembrane domains contain an unwound segment, and many of the substrate contact sites are with the main chain atoms of these unwound segments. Sequence alignments between LeuT$_{Aa}$ and the GLYTs revealed that there are a number of identical residues and some key differences in the predicted substrate binding site. In particular, in the crystal structure of LeuT$_{Aa}$, the amino group of leucine is hydrogen-bonded to the hydroxyl group of the side chain of serine at position 256 (Fig 3B). This residue is located in TM6. We focused on the role of the corresponding residues in GLYT1b and GLYT2a to investigate if this residue is a determinant of GLYT substrate selectivity (Vandenberg et al., 2007).

Serine at position 256 in LeuT$_{Aa}$ corresponds to a glycine residue (G305) in GLYT1b and a serine residue (S481) in GlyT2a. The G305 in GLYT1b was mutated to a serine (GLYT1-G305S) and S481 in GLYT2a was mutated to a glycine (GLYT2-S481G) using site directed mutagenesis (Vandenberg et al., 2007). In contrast to wild type GLYT2a, sarcosine is a

substrate of the GLYT2a-S481G mutant. The maximal current (97 ± 2%) and EC_{50} (26.2 ± 1.3 μM) of sarcosine at the mutant receptor are similar to that for wild-type GLYT1b (87 ± 1%, 22 ± 1 μM). The introduction of the corresponding mutation, G305S, into GLYT1b reduced levels of surface expression to approximately 10% of wild type. To overcome this limitation, two additional mutants were generated, GLYT1b-G305A and GLYT2a-S481A. Glycine transport of both mutant receptors is similar to that of wild type. However, incorporation of the S481A substitution into GLYT2a produced a transporter that could transport sarcosine with an efficacy similar to that of GLYT1b (70 ± 3%), but with a reduced affinity (590 ± 50 μM) relative to wild type. Combined, these findings demonstrate that residues at this position are important for sarcosine transport. For GLYT2a, sarcosine can be transported if an alanine or a glycine residue is present at this site, but not if a serine residue is present.

3.4 Mutagenesis to identify transport and channel domains of glutamate transporters

Glutamate transporters have two distinct functions: ion coupled glutamate transport and glutamate-activated chloride channel activity. In 1995, two studies demonstrated that the two functions co-exist in the same protein (Fairman et al., 1995; Wadiche et al., 1995). Glutamate binding is required for activation of the channel, but the direction of Cl^- ion flow through the channel domain is uncoupled from the direction of glutamate transport. This raised the question as to how the protein could support the dual functions. In the following section, we will describe how mutagenesis has been used to understand the structural basis for the dual functions of glutamate transporters. This is an interesting example of the complementary nature of mutagenesis and crystallography approaches to understand the functional properties of this class of transporters.

Prior to the determination of the crystal structure of Glt_{Ph}, mutagenesis was used to identify residues that may play a role in transporter function. Valine 452 (V452) of EAAT1 is located in the HP2 domain, and the V452C mutation does not alter the functional properties of the transporter. After modification of the V452C mutant with the methanethiosulfonate (MTS) reagent, [2-(trimethylammonium)ethyl] methanethiosulfonate (MTSET), the protein is no longer capable of transporting glutamate; but it still retains the glutamate-activated chloride channel (Ryan & Vandenberg, 2002). This suggests that the two functions are mediated by distinct conformational states of the transporter. In a separate study, our group attempted to identify regions of the transporter that form the chloride channel. We focussed our mutagenesis studies on TM2. TM2 contains a number of positively charged residues at the extracellular edge of the helix and a number of uncharged serine and threonine residues in the middle of the helix. We postulated that positive charges at the extracellular edge would attract anions into the channel and the hydrophilic residues within the channel would facilitate anion movement through the channel. To address this hypothesis, we mutated the positively charged residues at the extracellular edge to cysteine residues and probed the reactivity of the cysteine residues to both positively and negatively charged MTS reagents. The negatively charged MTS reagents had faster rates of reactivity than the positively charged MTS reagents, which suggested that the positively charged residues do attract negative charges to the extracellular edge of TM2. Substitutions of the hydrophilic serine and threonine residues in the middle of TM2 to small aliphatic residues significantly altered the anion permeability of the channel without affecting the transport function. This confirmed that the two functional properties of the transporter are mediated by separate

domains and also that the serine and threonine residues are likely to line the pore of the channel.

The studies described above were carried out prior to any knowledge of the three dimensional structure of the transporter, and so we attempted to identify how close the channel domain was from various other sites on the transporter by cross-linking experiments. Cysteine residues were introduced within TM2 and then at various other sites of the transporter, including V452 (see above). A disulfide bond formed spontaneously between V452C (in HP2) and Q93C (in TM2), indicating that these two residues must come into close proximity. From this study a crude structural model for these parts of the transporter was developed (Fig. 7A).
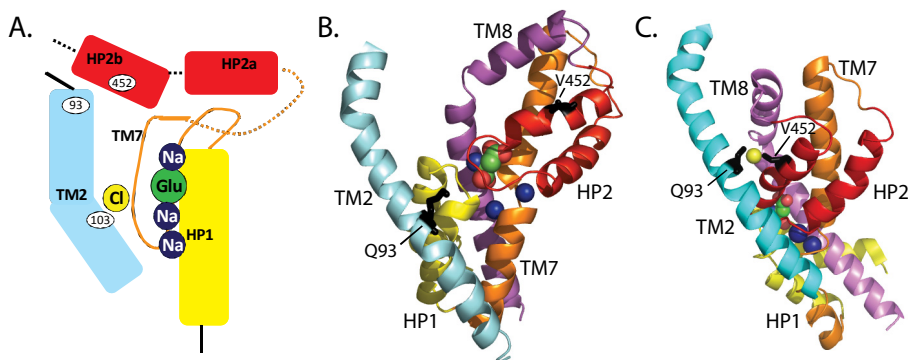
Fig. 7. Structural predictions of the relationship between the glutamate binding and translocation domain and the chloride conducting domain. A. A structural model for glutamate transport and Cl- ion permeation of EAAT1. We have omitted the K+ ion and H+ for simplicity. The *thick line* is in the plane of the paper and the *dashed line* is behind the plane of the paper. V452C can form a disulfide bond with Q93C. We propose that Cl- ions interact with residues along TM2. In this model, we suggest that glutamate and Na+ ions permeate the same pore as Cl- ions, but that there are separate molecular determinants for the two functions. B, C. The structure of the transport domain composed of HP1 (yellow), TM7 (orange), HP2 (red) and TM8 (purple) relative to TM2, which contains molecular determinants for Cl- permeation. Bound aspartate is shown in space-filling representation, and two Na+ ions are represented as blue spheres. The residues equivalent to Q93 and V452 are in stick representation and coloured black. B. Shows the distance between Q93 and V452 in the occluded state (PDB 2NWX), while C is the Hg^{2+} cross-linked structure showing the conformational changes required to bring Q93 and V452 into close proximity (PDB 3KBC, Hg^{2+} shown as a yellow sphere).

The crystal structure of Glt$_{Ph}$ was published in late 2004 (see above for a description of the structure); and, whilst the structure revealed many important details about substrate binding, the nature of the mechanism of activation of the chloride channel was not clear. The equivalent residues to V452 and Q93 in Glt$_{Ph}$ are approximately 20 Å apart (Fig. 7B, equivalent residues in a Glt$_{Ph}$ protomer are represented in black), which suggested that these residues were unlikely to come sufficiently close to form a disulfide bond. The first step in resolving this apparent contradiction came from confirming that hydrophilic

residues in TM2 of Glt$_{Ph}$ also form part of the lining of the chloride channel lumen, as observed for EAAT1 (Ryan & Mindell, 2007). The laboratory of Olga Boudker then repeated the crosslinking experiments in Glt$_{Ph}$, using cysteine mutants and adopting a similar approach to the one our group had done for EAAT1. Whilst spontaneous disulfide bonds did not form between the two residues in Glt$_{Ph}$, it was possible to catalyse the formation of a bond between the two cysteine residues using Hg^{2+}. It was concluded that the two domains can indeed move sufficiently to allow the two residues to come into close proximity. The cross-linked Glt$_{Ph}$ was also crystallized, and its structure was compared with the original structure (Fig. 7C). This study identified the conformational changes required to bring about the formation of the crosslinks and also suggested a mechanism for the transport process and how this process can lead to channel activation (Reyes et al., 2009). Briefly, the three transport domains (consisting of TM3, TM6, TM7 and TM8 and HP1 and HP2 from each protomer) move as three separate units through a rigid trimerization scaffold. TM2 is part of the scaffold, whilst HP2 is part of the transport domain. It would appear that the sliding movement of the transport domain relative to the rigid trimerization scaffold allows chloride ions to pass through the gap between the two functional domains (Vandenberg et al., 2008). Further mutagenesis will be required to verify this proposal. This series of experiments starting with mutagenesis, followed by crystallography, further mutagenesis and then further crystallography, which will also be followed up by further mutagenesis, highlights how the two approaches to understanding structure and function relationships can complement one another and provoke new ideas and concepts in protein function.

3.5 Substrate affinity and K$^+$ ion coupling in EAAT1

For the last section of this chapter, we will focus on an example of how mutagenesis approaches have been used to understand substrate and ion binding properties of the glutamate transporter family. Many of the residues that have been implicated in substrate and ion binding/translocation (Bendahan et al., 2000; Kavanaugh et al., 1997; Vandenberg et al., 1995) and chloride permeation (Ryan et al., 2004) are conserved throughout the glutamate transporter family. In particular, the carboxy-termini of both the EAATs and Glt$_{Ph}$ are highly conserved and contain the substrate and Na$^+$ binding sites. Despite their significant amino acid identity, the EAATs and Glt$_{Ph}$ display several functional differences. The EAATs transport aspartate and glutamate with similar affinity, while Glt$_{Ph}$ is selective for aspartate over glutamate. In addition, Glt$_{Ph}$ transport is not coupled to the co-transport of H$^+$ or the counter-transport of K$^+$ (Boudker et al., 2007; Ryan et al., 2009). Examination of the amino acid sequences of the substrate binding site of the EAATs and Glt$_{Ph}$ does not reveal any residues that can clearly account for the differences observed in substrate selectivity or affinity. However, an arginine residue is in close proximity to the substrate binding site of both the EAATs and Glt$_{Ph}$, but it is located in TM8 in the EAATs and in HP1 of Glt$_{Ph}$ (Fig 1B). The aim of our study was to investigate the functional effect of the location of a positively charged arginine residue in two members of the glutamate transporter family, EAAT1 and Glt$_{Ph}$.

In order to examine the role of this arginine residue, two double mutant transporters were produced. In EAAT1 the arginine residue was moved from TM8 to HP1 (EAAT1S363R/R477M), and the reverse double mutation was introduced into the gene for Glt$_{Ph}$ (Glt$_{Ph}$R276S/M395R). Switching the arginine residue from TM8 to HP1 in EAAT1 had

no effect on substrate selectivity, but it did increase affinity for both glutamate and aspartate and abolished K^+ coupling. The apparent affinity for both L-glutamate and L-aspartate was increased ~130 fold, and it was similar to the affinity of L-aspartate in Glt_{Ph} (Ryan et al., 2009). The counter-transport of one K^+ ion per transport cycle is thought to be important for the relocation of the EAATs to the outward-facing state. The movement of an arginine residue from TM8 to HP1 has potentially slowed the return of the empty transporter to the extracellular facing side, thus contributing to the decrease in observed affinity values. In contrast, the inverse changes in Glt_{Ph} (Glt_{Ph}R276S/M395R) resulted in a functional transporter that has a ~4-fold reduction in the affinity for aspartate compared to wild type. The substitutions did not affect substrate selectivity or introduce K^+ dependence.

The crystal structure of Glt_{Ph} reveals that the backbone carbonyl group of the arginine residue in HP1 forms a direct contact with the substrate. However, our mutagenesis studies suggest that it is the side chain that is influencing transport properties. A possible explanation is that the conformation of the HP1 loop region and also the proximal TM8 is influenced by the arginine side chain and neighbouring residues. Thus, mutating this arginine may influence the conformation of the backbone carbonyl group, which in turn may influence substrate affinity.

The movement of K^+ ions through the transporter is likely to rely upon multiple conformational changes and interactions. Disruption of any of these interactions via a mutation is liable to result in loss of K^+ coupling. However, to introduce K^+ coupling will require multiple mutations. This may explain why the double mutation is sufficient to abolish K^+ coupling in EAAT1, but not introduce it into Glt_{Ph}.

4. Conclusion

Directed mutagenesis has been particularly useful in understanding the structure and function of mammalian membrane proteins. In this chapter we have outlined our approach to structure-activity studies of neurotransmitter transporters in the solute carrier (SLC) families, SLC1 and SLC6. We have used examples from our work on the excitatory amino acid transporters (EAATs), the archaeal aspartate transporter (Glt_{Ph}), glycine transporters (GLYTs) and the prokaryotic leucine transporter ($LeuT_{Aa}$).

5. Acknowledgments

6. References

Aragon C. & Lopez-Corcuera B. (2003). Structure, function and regulation of glycine neurotransporters. *European Journal of Pharmacology* 479(1-3): 249-262.

Arriza J., Kavanaugh M., Fairman W., Wu Y., Murdoch G., North R. & Amara S. (1993). Cloning and expression of a human neutral amino acid transporter with structural similarity to the glutamate transporter gene family. *Journal of Biological Chemistry* 268(21): 15329.

Aubrey K.R. & Vandenberg R.J. (2001). N[3-(4'-fluorophenyl)-3-(4'-phenylphenoxy)propyl]sarcosine (NFPS) is a selective persistent inhibitor of glycine transport. *British Journal of Pharmacology* 134(7): 1429-1436.

Bartholomaus I., Milan-Lobo L., Nicke A., Dutertre S., Hastrup H., Jha A., Gether U., Sitte H.H., Betz H. & Eulenburg V. (2008). Glycine transporter dimers: evidence for occurrence in the plasma membrane. *Journal of Biological Chemistry* 283(16): 10978-10991.

Bendahan A., Armon A., Madani N., Kavanaugh M. & Kanner B. (2000). Arginine 447 plays a pivotal role in substrate interactions in a neuronal glutamate transporter. *Journal of Biological Chemistry* 275(48): 37436.

Boudker O., Ryan R., Yernool D., Shimamoto K. & Gouaux E. (2007). Coupling substrate and ion binding to extracellular gate of a sodium-dependent aspartate transporter. *Nature* 445(7126): 387-393.

Dohi T., Morita K., Kitayama T., Motoyama N. & Morioka N. (2009). Glycine transporter inhibitors as a novel drug discovery strategy for neuropathic pain. *Pharmacology and Therapeutics* 123(1): 54-79.

Edington A.R., McKinzie A.A., Reynolds A.J., Kassiou M., Ryan R.M. & Vandenberg R.J. (2009). Extracellular loops 2 and 4 of GLYT2 are required for N-arachidonylglycine inhibition of glycine transport. *Journal of Biological Chemistry* 284(52): 36424-36430.

Fairman W.A., Vandenberg R.J., Arriza J.L., Kavanaugh M.P. & Amara S.G. (1995). An excitatory amino-acid transporter with properties of a ligand-gated chloride channel. *Nature* 375(6532): 599-603.

Guastella J., Nelson N., Nelson H., Czyzyk L., Keynan S., Miedel M.C., Davidson N., Lester H.A. & Kanner B.I. (1990). Cloning and expression of a rat brain GABA transporter. *Science* 249(4974): 1303-1306.

Horiuchi M., Nicke A., Gomeza J., Aschrafi A., Schmalzing G. & Betz H. (2001). Surface-localized glycine transporters 1 and 2 function as monomeric proteins in Xenopus oocytes. *Proc Natl Acad Sci U S A* 98(4): 1448-1453.

Ju P., Aubrey K.R. & Vandenberg R.J. (2004). Zn2+ inhibits glycine transport by glycine transporter subtype 1b. *Journal of Biological Chemistry* 279(22): 22983-22991.

Kanai Y. & Hediger M. (2004). The glutamate/neutral amino acid transporter family SLC1: molecular, physiological and pharmacological aspects. *Pflügers Archiv European Journal of Physiology* 447(5): 469-479.

Kavanaugh M., Bendahan A., Zerangue N., Zhang Y. & Kanner B. (1997). Mutation of an amino acid residue influencing potassium coupling in the glutamate transporter GLT-1 induces obligate exchange. *Journal of Biological Chemistry* 272(3): 1703.

Lopez-Corcuera B., Alcantara R., Vazquez J. & Aragon C. (1993). Hydrodynamic properties and immunological identification of the sodium- and chloride-coupled glycine transporter. *J Biol Chem* 268(3): 2239-2243.

Reyes N., Ginter C. & Boudker O. (2009). Transport mechanism of a bacterial homologue of glutamate transporters. *Nature* 462(7275): 880-885.

Ryan R.M., Compton E.L. & Mindell J.A. (2009). Functional characterization of a Na+-dependent aspartate transporter from Pyrococcus horikoshii. *Journal of Biological Chemistry* 284(26): 17540-17548.

Ryan R.M. & Mindell J.A. (2007). The uncoupled chloride conductance of a bacterial glutamate transporter homolog. *Nature Structural and Molecular Biology* 14(5): 365-371.

Ryan R.M., Mitrovic A.D. & Vandenberg R.J. (2004). The chloride permeation pathway of a glutamate transporter and its proximity to the glutamate translocation pathway. *Journal of Biological Chemistry* 279(20): 20742-20751.

Ryan R.M. & Vandenberg R.J. (2002). Distinct conformational states mediate the transport and anion channel properties of the glutamate transporter EAAT-1. *Journal of Biological Chemistry* 277(16): 13494-13500.

Schrodinger L. (2010). The PyMOL Molecular Graphics System, Version 1.3r1.

Shevchuk N.A., Bryksin A.V., Nusinovich Y.A., Cabello F.C., Sutherland M. & Ladisch S. (2004). Construction of long DNA molecules using long PCR-based fusion of several fragments simultaneously. *Nucleic Acids Research* 32(2): e19.

Singh S.K., Piscitelli C.L., Yamashita A. & Gouaux E. (2008). A competitive inhibitor traps LeuT in an open-to-out conformation. *Science* 322(5908): 1655-1661.

Singh S.K., Yamashita A. & Gouaux E. (2007). Antidepressant binding site in a bacterial homologue of neurotransmitter transporters. *Nature* 448(7156): 952-956.

Slotboom D.J., Konings W.N. & Lolkema J.S. (1999). Structural features of the glutamate transporter family. *Microbiology and Molecular Biology Reviews* 63(2): 293-307.

Succar R., Mitchell V.A. & Vaughan C.W. (2007). Actions of N-arachidonyl-glycine in a rat inflammatory pain model. *Molecular Pain* 3: 24.

Supplisson S. & Bergman C. (1997). Control of NMDA receptor activation by a glycine transporter co-expressed in Xenopus oocytes. *Journal of Neuroscience* 17(12): 4580-4590.

Sur C. & Kinney G.G. (2004). The therapeutic potential of glycine transporter-1 inhibitors. *Expert Opinion on Investigational Drugs* 13(5): 515-521.

Vandenberg R.J., Arriza J.L., Amara S.G. & Kavanaugh M.P. (1995). Constitutive ion fluxes and substrate binding domains of human glutamate transporters. *Journal of Biological Chemistry* 270(30): 17668-17671.

Vandenberg R.J., Huang S. & Ryan R.M. (2008). Slips, leaks and channels in glutamate transporters. *Channels (Austin)* 2(1): 51-58.

Vandenberg R.J., Shaddick K. & Ju P. (2007). Molecular basis for substrate discrimination by glycine transporters. *Journal of Biological Chemistry* 282(19): 14447-14453.

Vuong L.A., Mitchell V.A. & Vaughan C.W. (2008). Actions of N-arachidonyl-glycine in a rat neuropathic pain model. *Neuropharmacology* 54(1): 189-193.

Wadiche J.I., Amara S.G. & Kavanaugh M.P. (1995). Ion fluxes associated with excitatory amino acid transport. *Neuron* 15(3): 721-728.

Wiles A.L., Pearlman R.J., Rosvall M., Aubrey K.R. & Vandenberg R.J. (2006). N-Arachidonyl-glycine inhibits the glycine transporter, GLYT2a. *Journal of Neurochemistry* 99(3): 781-786.

Yamashita A., Singh S.K., Kawate T., Jin Y. & Gouaux E. (2005). Crystal structure of a bacterial homologue of Na+/Cl--dependent neurotransmitter transporters. *Nature* 437(7056): 215-223.

Yernool D., Boudker O., Jin Y. & Gouaux E. (2004). Structure of a glutamate transporter homologue from Pyrococcus horikoshii. *Nature* 431(7010): 811-818.

Zerangue N. & Kavanaugh M. (1996). Flux coupling in a neuronal glutamate transporter. *Nature* 383(6601): 634-637.

Site-Directed Mutagenesis as a Tool for Unveiling Mechanisms of Bacterial Tellurite Resistance

José Manuel Pérez-Donoso[1] and Claudio C. Vásquez[2]
[1]Universidad de Chile
[2]Universidad de Santiago de Chile
Chile

1. Introduction

Tellurium (Te) is a scarce element in the earth's crust and is not essential for living organisms. It is rarely found in the non-toxic, elemental state (Te[o]); and the soluble oxyanions, tellurite (TeO_3^{2-}) and tellurate (TeO_4^{2-}), are toxic for most forms of life. Tellurite toxicity has been extensively exploited as a selective agent in diverse microbiological culture media.

A few bacterial tellurite resistance mechanisms have been proposed; but the genetic, biochemical and/or physiological bases underlying TeO_3^{2-} resistance are still poorly understood.

One of our strategies to study bacterial resistance to TeO_3^{2-} has been the cloning and characterization of genes from tellurite-resistant bacteria using *Escherichia coli* as a sensitive host. Using this experimental approach, we have previously shown that the genes *cysK*, *iscS* and *cobA* [encoding cysteine synthase, cysteine desulfurase and S-Adenosyl-L-methionine:uroporphirin-III C-methyltransferase (SUMT), respectively] from the thermophilic rod *Geobacillus stearothermophilus* V mediate tellurite resistance when expressed in *E. coli*. All of these genes were subjected to site-directed mutagenesis to demonstrate their participation in tellurite resistance in this mesophilic host (Vásquez et al., 2001; Tantaleán et al., 2003; Araya et al., 2009). More recently we conducted similar mutagenesis experiments with the *Aeromonas caviae* ST *lpdA* gene, encoding dihydrolipoil dehydrogenase, and found that two amino acid residues are involved in the tellurite reductase branch-activity of this enzyme (unpublished data).

Site-directed mutagenesis, also referred to as site-specific or oligonucleotide-directed mutagenesis, is a technique in molecular biology that allows the creation of mutations at a defined DNA sequence. In general, a synthetic primer containing the desired base change is hybridized to a single-stranded DNA containing the gene of interest; the rest of the gene is then copied using a DNA polymerase. The double-stranded DNA molecule thus obtained is ligated to an appropriate vector and introduced into a host cell for mutant selection.

This chapter does not intend to be an extensive review of tellurite resistance. Instead it was written as an example to make young scientists see how simple observations can help to

state the basis of much more complex networks underlying a particular, defined phenomenon.

2. The enigma of tellurite toxicity

The ability of bacteria to counteract the effect of heavy metals has interested microbiologists for many years. Toxic heavy metals are often encountered in nature in many different forms. In air, they exist as metal or oxide dust. In surface and ground water, they are found attached to humic substances; and they also bind to soil and sediments.

Tellurium has applications in the semiconductor industry and electronics (in the production of thermoelements, photoelements and other devices in automation equipment). The increasing demand for new and different semiconductors necessitates research work on the application of various tellurium compounds as semiconductor components.

As a group, microorganisms display resistance to nearly all metal and non-metal ions that are considered toxic to the environment, including Ag^+, As^{3+}, Cd^{2+}, Cr^{3+}, Hg^{2+}, Sb^{3+}, Te^{4+}, Te^{6+} and Zn^{2+}, among others (Silver, 2006). Although the literature on the subject is vast and continuously updated (Silver, 2011), in most cases, however, the knowledge of the biochemical and/or genetic mechanisms underlying the metal resistance phenomena is still very limited. This is particularly true for bacterial tellurite resistance, in which much effort has been expended to understand how bacteria counteract the toxic effects of the tellurium salt.

Tellurium (Te) was considered almost an exotic element and was treated with certain indifference by most serious chemists. However, the impressive number of publications on Te compounds during the last few years shows that Te is now widely used in applied chemical reactions.

The natural Te cycle has not been investigated in depth, and the role of microbes – if any - in this process has not yet been elucidated. Nevertheless, tellurite-resistant bacteria do exist in nature; and they often reduce tellurite to its elemental, less toxic, form (Te^o), which accumulates as black deposits inside the cell (Taylor, 1999; Chasteen et al., 2009).

As a result of the accumulated knowledge, several tellurite-resistance determinants (Te^R) have been localized on plasmids and on the chromosome. Structure and organization vary greatly among bacterial species (Taylor, 1999). It has been argued that tellurite toxicity results from the ability of tellurite to act as a strong oxidizing agent that damages a number of cell components (Taylor, 1999; Pérez et al., 2007). In the last years, however, available evidence shows that tellurite toxicity results from the generation of reactive oxygen species (ROS) (Borsetti et al., 2005; Calderón et al., 2006; Tremaroli et al., 2007; Pérez et al., 2007). ROS, such as hydrogen peroxide (H_2O_2), superoxide anion (O_2^-) and hydroxyl radical (OH^-), are typical byproducts of aerobic metabolism. However, they can also be produced upon exposure of the cell to free radical-generating compounds, like metals and metalloids.

Our group has been interested in studying tellurite resistance (Te^R)/toxicity for many years. First we focused on thermophilic, Gram-negative, rods of the genus *Thermus* and later on *G. stearothermophilus* V, a thermotolerant, spore-forming, Gram-positive bacterium that was isolated in our laboratory from a soil sample. In both cases we demonstrated the existence of

Site-Directed Mutagenesis as a Tool
for Unveiling Mechanisms of Bacterial Tellurite Resistance
187

cellular reductases that convert tellurite into elemental tellurium at the expense of NAD(P)H oxidation *in vitro* (Chiong et al., 1988; Moscoso et al., 1998).

In an attempt to identify genetic determinants for Te[R] in these bacteria, we constructed gene libraries that were used to transform sensitive *E. coli* hosts to tellurite resistance. While the cloning of resistance determinants from *Thermus* has been unsuccesful so far, we did clone tellurite resistance determinants from *G. stearothermophilus* V into *E. coli*. These genes were subjected to site-directed mutagenesis in order to unveil their participation in the resistance phenomenon (Vásquez et al., 2001; Tantaleán et al., 2003; Araya et al., 2009). More recently, we have shown that overproduction of the *Aeromonas caviae* ST dihydrolipoil dehydrogenase results in enhanced tellurite resistance in *E. coli*. This enzyme exhibits NADH-dependent tellurite reductase (TR) activity (Castro et al., 2008, 2009). The change of two defined amino acid residues at the enzyme active site decreased TR activity (unpublished data). What follows is a chronological description of the above-mentioned results.

2.1 CysK

Cysteine synthases (CysKs) are enzymes that catalyze the last step in cysteine biosynthesis. They have been related to tellurite resistance in different microorganisms (Moore & Kaplan, 1992; O'Gara et al., 1997; Alonso et al., 2000; Vásquez et al., 2001; Lithgow et al., 2004). All cysteine synthases described to date require the cofactor pyridoxal 5'-phosphate (PLP) for activity. PLP-dependent enzymes catalyze a broad spectrum of aminoacid transformations involved in the development of an organism, such as transaminations, β-eliminations, β-γ replacements and racemizations.

Searching for tellurite-resistance determinants, we identified and characterized a new thermophilic CysK in *G. stearothermophilus* V (Saavedra et al., 2004). *E. coli* ovexpressing this *cysK* gene shows a tellurite resistance that is over 10-fold more than that observed for the wild-type controls. Despite the fact that it is known that the ping-pong catalytic mechanism of this enzyme is similar to that of other CysKs, the *G. stearothermophilus* V enzyme has not been fully characterized. The catalytic amino acid residues are also not known. In addition, the importance of cysteine synthase in bacterial tellurite resistance has not been totally documented; and our group has proposed that the resistance increase is paralleled by increased levels of reduced thiols, such as glutathione.

G. stearothermophilus V CysK is a homodimer (32 kDa/monomer) that requires one PLP molecule per subunit (Saavedra et al., 2004). It belongs to the β family of PLP-dependent enzymes and shares some similarities with other enzymes involved in deamination reactions, as do tryptophan synthase, threonine deaminase and O-acetyl-serine sulphydrilase (Alexander et al., 1994).

A general mechanism for the CysK-catalyzed reaction has been proposed (Cook and Wedding, 1977; Tai et al., 1998). The enzyme binds PLP by a lysine group forming a Schiff base, known as internal aldimine. This intermediate absorbs in the 400-430 nm region and exhibits two resonant forms. Addition of the O-acetyl-L-serine (OAS) substrate allows the formation of the geminaldiamine intermediate, which produces the external aldimine. Then the quinonoid intermediate is formed; and when the substituent in the β position is released (acetate), the α-aminoacrylate is finally formed.

To study the residues and motifs that define the catalytic properties of this enzyme, we used site-directed mutagenesis to assess the importance of the C-terminus and of some putative catalytic residues for CysK activity and CysK-mediated bacterial tellurite resistance (unpublished data).

As a first approach, a set of CysK C-terminal deletions of 10 (CysK ΔTyr298), 20 (CysK ΔLeu288), 30 (CysK ΔAla278), 40 (CysK ΔGly268) and 60 (CysK ΔAla248) amino acids were constructed, overexpressed, purified and characterized. Binding of the PLP cofactor was evaluated through the absorption spectrum of the purified proteins. An absorbance peak at 412 nm is characteristic of the α-aminoacrylate intermediate.

All the CysK deletion mutants larger than ΔTyr298 were inactive, unable to bind PLP and did not confer tellurite resistance. This result indicated a direct relationship between enzymatic activity and tellurite tolerance. It is also in agreement with a role for thiols, such as cysteine, in tellurite tolerance. In this context, increased levels of intracellular reduced thiols, particularly glutathione, were observed in cells overproducing CysK. This observation suggested that increased concentrations of cell antioxidants could be responsible for protecting *cysK*-overexpressing cells from tellurite-mediated oxidation. Interestingly, the mutant having Tyr298 (Tyrosine as residue 298, Y298) (see Table 1 in the chapter by Figurski et al. for the amino acid codes) as the C-terminal residue [CysK Δ(Tyr298)] was inactive, despite retaining the ability to bind PLP and to form the α-aminoacrylate intermediate. As expected, overproduction of this mutant enzyme did not enhance tellurite resistance in *E. coli*.

As shown in Fig. 1, CysK displays the conserved amino acid sequence motif SVKDRIA near the amino terminus, which is required for PLP binding. Most of the C-terminal truncated mutants were unable to bind PLP, despite the presence of this motif and the finding that the proteins were correctly translated and folded in the cytoplasm. This observation suggests that other residues are involved in stabilizing PLP binding. The residues are probably located near the CysK C-terminus.

As can be deduced from these results, protein deletions allow a global view about the importance of some protein motifs and not a detailed interpretation on the participation of defined amino acids in the enzyme's functioning. In this context, site-directed mutagenesis offers a more versatile alternative to study the role of a defined motif or amino acid residue. These two experimental approaches can complement each other in order to obtain a detailed analysis for understanding the enzymatic mechanism.

To choose the appropriate residues to be subjected to site-directed mutagenesis, the first approach involved sequence conservation studies using BLASTP and ClustalW software and other programs. However, a better idea about which domains and/or residues could be interacting with defined molecules or atoms in a reaction can be obtained by constructing 3D models of the protein. In those cases where there is no crystallographic information regarding the protein, bioinformatic tools can offer a useful alternative in order to predict a model based on sequence homology.

Sequence analysis allowed identification of the conserved [43]SVKDRIA[49] domain (NCBI accession number AAG28533.1, see Fig. 1) in the *G. stearothermophilus* V CysK. In this conserved sequence, K45 is required to form the protonated Schiff base with PLP. K45 also
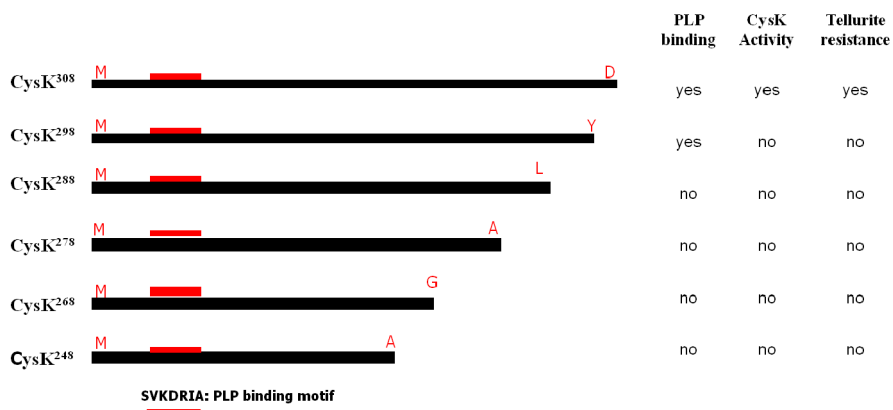
Fig. 1. Schematic model representing the different truncated CysKs and their properties (see the text for details).

participates directly in binding the cofactor and NH_3 transamination. A replacement of K45 by alanine (K45A) was made in CysK to evaluate the importance of the positive charge required for PLP interaction. The mutated gene was overexpressed in *E. coli*, and the enzyme was purified to homogeneity. No enzymatic activity or cofactor binding was observed by the purified enzyme, confirming that this residue is required for PLP attachment, as predicted.

No structural data are available for *G. stearothermophilus* V CysK, so a 3D model based on the available crystallographic structure (2 A° resolution) for the *Salmonella enterica* serovar Typhimurium CysK was constructed. With this model, we expected to make predictions about residues located at the enzyme's C-terminus that could be interacting with PLP, stabilizing it and allowing the transamination reaction.

As mentioned before, the deletion mutant CysK Δ(Tyr298) was inactive, but still able to bind PLP. Based on this observation, a homology model of this deletion mutant was constructed to define the position of specific residues that can participate in PLP stabilization both in wild-type CysK and in the CysK Δ(Tyr298) mutant.

The model suggested that the CysK Δ(Tyr298) deletion mutant exhibits amino acids in positions that can interact with PLP so as to orientate the cofactor for the reaction to proceed (Fig. 2). Tyrosine 298 is most probably interacting directly with PLP, given that it is located at a distance (less than 4 A°) that allows the formation of a hydrogen bond with the cofactor. Experiments with this deletion mutant confirmed that the enzyme was able to bind PLP and was folded correctly. However, studies performed to assess PLP- mediated fluorescence indicated that the cofactor displayed a different orientation than that in wild-type enzyme. Altogether, these results suggested that the Tyr298 residue could be important for CysK activity. Most probably Tyr298 forms a hydrogen bond through the hydroxyl group of the tyrosine, thus stabilizing and favoring a correct orientation of PLP. To test this possibility, homology analyses and new models of Y298A and Y298P CysK mutants were constructed to

predict the role of the Tyr298 residue. The results suggested that the -OH group of Tyr298 is required for PLP binding. It is thought to affect catalysis because of an interaction with the α–aminoacrylate intermediate through a water molecule.
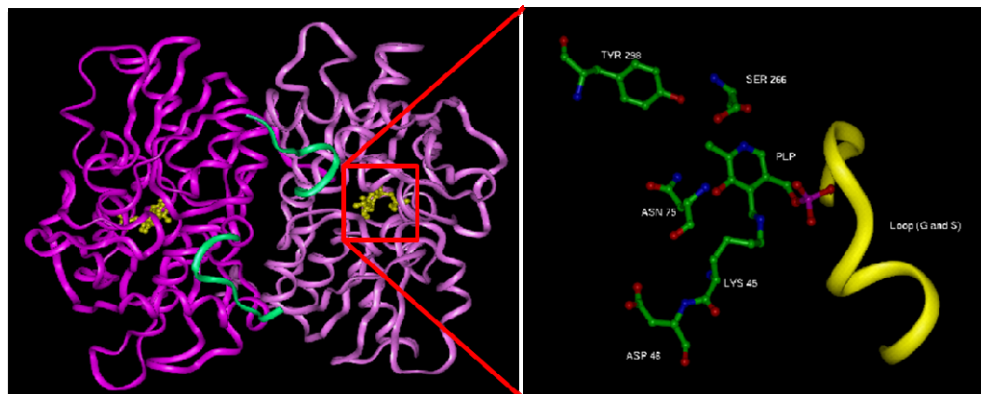
Fig. 2. Model (homology) of the *G. stearothermophilus* V CysK. Left, CysK homodimer showing one PLP molecule (yellow) bound per subunit and the C-terminal 10 amino acids (green). Right, schematic view of those amino acid residues involved in PLP binding (LYS45) and in orienting and/or stabilizing the cofactor (SER266, TYR298, and ASN75).

Two additional Tyr298 site-directed variants were constructed and characterized (Y298A and Y298F). Mutant proteins were purified and assayed for PLP binding, α-aminoacrylate intermediate formation, and enzymatic activity (lyase and cysteine synthase). Both mutant enzymes were inactive and did not bind PLP nor formed the α-aminoacrylate intermediate, suggesting that the -OH group of Tyr298 is required for CysK activity.

Based on these observations, we propose that the Tyr298 residue of *G. stearothermophilus* V CysK is involved in stabilization and proper orientation of PLP by interacting with the phosphate group of the cofactor through hydrogen bonding. The expression of the *G. stearothermophilus* V *cysK* gene was later shown to mediate low-level tellurite resistance in *E. coli* (Vásquez et al., 2001).

2.2 IscS

One of the superoxide anion targets is a family of dehydratases that use exposed [4Fe-4S] clusters to bind and dehydrate their substrates during the biosynthesis of branched-chain amino acids. Oxidation of these proteins results in the dismantling of these [Fe-S] centers with the concomitant loss of enzyme activity. Crude *E. coli* extracts catalyzing the formation of these centers *in vitro* contain at least four enzymatic activities that provide the required sulfur atoms: O-acetyl-serine sulfhydrilase A (CysK), O-acetyl-serine sulfhydrilase B (CysM) and β-cistationase; the fourth protein displayed cysteine desulfurase activity. Cysteine desulfurases remove the sulfur atom from cysteine to construct and repair [Fe-S] clusters in protein substrates that, in turn, catalyze essential redox reactions in critical metabolic pathways. Very soon it became clear that the IscS cysteine desulfurase played an important role in transferring the sulfur from cysteine for [Fe-S] center synthesis *in vivo*.

Various tellurite-resistant (Te^R) *E. coli* clones were isolated upon transformation with a *G. stearothermophilus* V *Hin*dIII library. In particular, one contained a 3.5 kilobase (kb) DNA insert that specified three open reading frames (ORFs). By comparison with sequences deposited in protein data banks, it was found that ORF2 [1200 base pairs (bp)] encoded a ~45 kDa cysteine desulfurase (IscS). The expression of the *G. stearothermophilus* V IscS cysteine desulfurase conferred tellurite resistance in *E. coli*. The enzyme was induced and purified to homogeneity; the purified enzyme displayed cysteine desulfurase activity. We showed that tellurite resistance depends in part on the activity of the IscS enzyme, supporting the hypothesis that essential proteins with iron-sulfur [Fe-S] clusters are among the main targets of the oxidative damage caused by tellurite in *E. coli*. Unlike the case for other microbes, the *G. stearothermophilus* V *iscS* gene does not appear to be within an operon containing other genes involved in *de novo* [Fe-S] cluster formation.

Because *G. stearothermophilus* V is a thermophile, IscS purification included an incubation of soluble cell extracts at 70 °C for 20 min. This step eliminated almost 75% of the total starting protein, without an appreciable loss of IscS protein or cysteine desulfurase activity (Fig. 3, lane 1). Subsequent column chromatography steps resulted in enzyme preparations that were >98% pure. The amino-terminus of the purified IscS (MNLEQIRKDTPLHKKYSYIN), determined by Edman degradation, matched precisely the predicted primary sequence of the product of the *iscS* gene. The native form of the enzyme is a homodimer with an apparent molecular mass of 93-97 kDa, as determined by size-exclusion chromatography. This cysteine desulfurase belongs to the α-family of PLP-dependent enzymes and exhibits an absorbance maximum for PLP centered at about 420 nm, the characteristic UV-visible spectrum of other cysteine desulfurases.

To further confirm that IscS activity was responsible for tellurite resistance in *E. coli*, a series of mutant derivatives was constructed. Plasmids containing truncated versions of the *iscS* gene (90, 150 or 210 bp deletions of the 3' end) did not confer resistance to tellurite in *E. coli*; nor did they exhibit desulfurase activity in crude extracts. The induced mutant proteins formed inclusion bodies, suggesting that the carboxyl terminus of IscS is essential for proper folding, dimerization and/or function.

We decided to make a directed change of Lys213, a residue that likely binds the PLP cofactor. This lysine is conserved in the sequence of cysteine desulfurases from both Gram-positive and Gram-negative bacteria, as well as in the yeast *Saccharomyces cerevisiae*. Lys213 in IscS was replaced by alanine to yield the *iscS_K213A* mutant gene, which was cloned into the expression plasmid pET21b and introduced into *E. coli* JM109(DE3) to induce high-level transcription from the promoter. Cells expressing *iscS K213A* did not exhibit tellurite resistance.

The K213A enzyme was purified to homogeneity, using the same procedure as for the native enzyme, but omitting the initial heat-treatment because the mutant protein did not show the thermostability of the wild-type IscS (Fig. 3). Unlike extracts containing IscS, those of the IscS K213A mutant protein did not exhibit the typical intense yellow colour, consistent with the idea that Lys213 is critical for PLP binding. In fact, the absorbance peak characteristic of PLP-containing enzymes was missing in the UV-visible spectrum of the purified mutant protein. The mutant IscS enzyme showed less than 10% of the specific activity exhibited by the wild-type IscS.
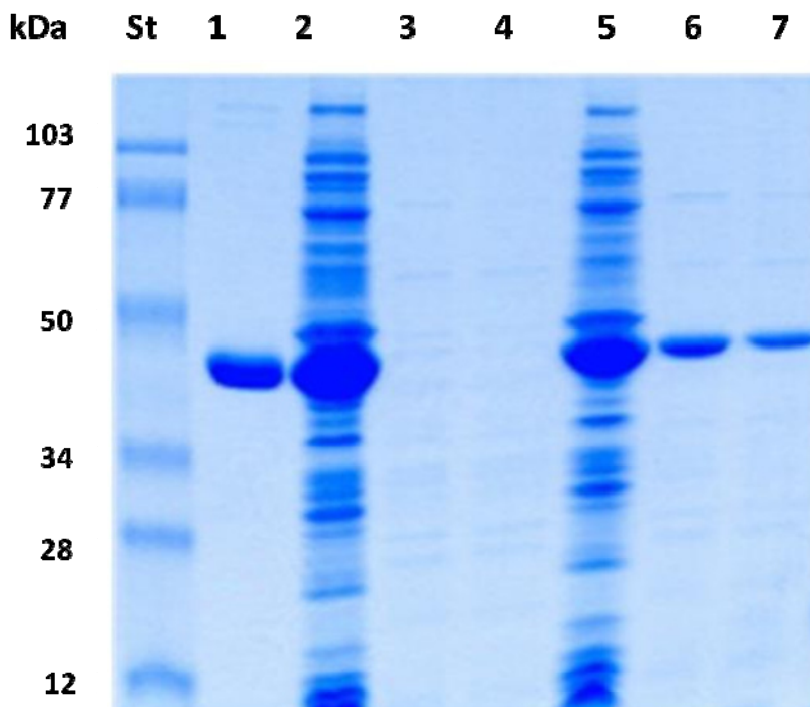
Fig. 3. Effect of temperature on the IscS K213A and K213R mutants. Lane 1, wild-type IscS. Lanes 2, 3 and 4, crude extracts of *E. coli* overproducing the K213A mutant cysteine desulfurase incubated for 10 min at 37 °C, 10 min at 70 °C and 20 min at 70 °C, respectively. Lanes 5, 6, and 7, as in lines 2-4, but using *E. coli* extracts overproducing the K213R mutant cysteine desulfurase.

On the other hand, an IscS K223R mutant was also constructed to confirm the importance of the positive charge in protein stabilization, PLP binding and desulfurase activity. As shown in Figs. 3 and 4, the K213R IscS displayed the same thermostability behavior of the wild-type desulfurase; and the enzyme was also easily purified by heating. In addition, the purified K213R mutant displayed the characteristic intense yellow color of PLP enzymes, an observation that was further confirmed by the presence of the 412 nm absorbance peak in the UV-visible spectra. As expected, this mutant displayed desulfurase activity levels identical to those of the wild-type desulfurase (Fig. 4). These results confirmed that the arginine residue was able to maintain IscS function, indicating that the positive charge at this position is required for proper PLP binding (Fig. 5).

The *G. stearothermophilus* V *iscS* gene not only complemented successfully an *E. coli iscS* mutation, but also conferred tellurite resistance to an *E. coli sodAsodB* double mutant, arguing that superoxide causes specific damage to one or more critical [Fe-S] cluster-containing proteins.
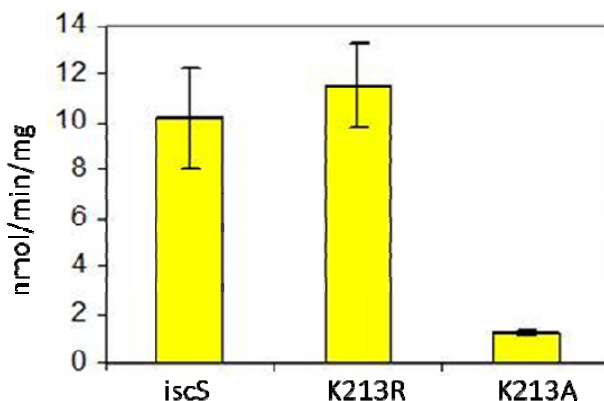
Fig. 4. Desulfurase activity of wild-type IscS and the indicated mutant IscSs.

2.3 CobA

A third gene mediating tellurite resistance in *E. coli* was identified from a *G. stearothermophilus* V *Hin*dIII library. The *cobA* gene was initially identified as one of three main ORFs present in a 3.8 kb *G. stearothermophilus* V DNA insert in the recombinant plasmid p1VH. *E. coli* carrying p1VH exhibited over 10 fold the resistance to potassium tellurite observed in the same strain harboring the pSP72 cloning vector alone (Araya et al., 2004).
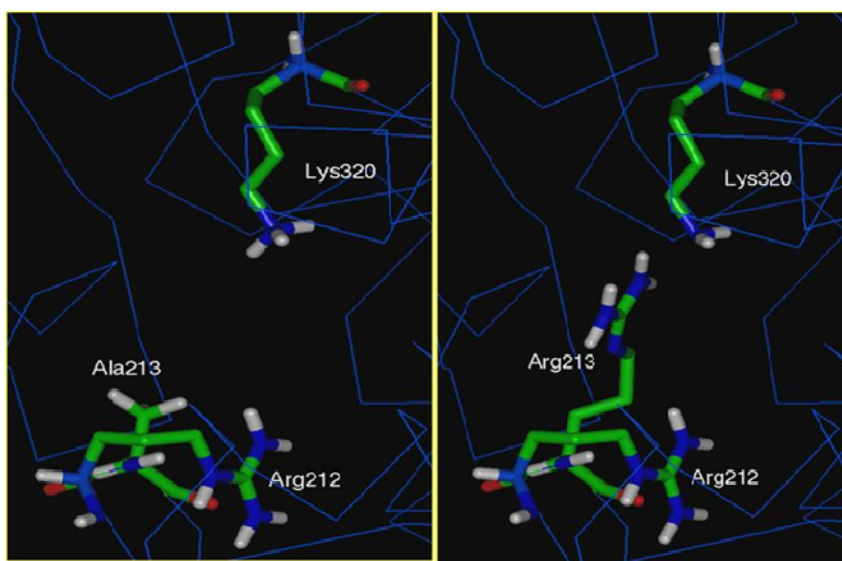
Fig. 5. Molecular models of the active site of the K213A and K213R mutants of the IscS enzyme from *G. stearothermophilus* V.

The *cobA* gene of *G. stearothermophilus* V encodes a 28 kDa protein exhibiting 71% identity with the *Bacillus megaterium cobA* gene, which encodes the enzyme S-adenosyl-L-methionine:uroporphirin-III C-methyltransferase, also referred to as SUMT.

The *cobA* gene was amplified by PCR using appropriate primers, inserted into the pET21b expression vector and introduced into *E. coli* JM109(DE3). Transformants exhibited higher tellurite resistance than that observed for cells carrying the cloning vector alone. Minimal inhibitory concentration (MIC) determinations were carried out in the absence of the inducer (IPTG), since the protein seemed to be toxic to the cell when overexpressed.

G. stearothermophilus V SUMT was induced with IPTG and judged >95% pure after being purified in two chromatographic steps (Cibacron blue and Sephadex column chromatography). After being fractionated by polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulphate (PAGE-SDS) and transferred to a polyvinylidene fluoride (PVDF) membrane, the purified SUMT was sent for EDMAN microsequencing analysis. The first 10 amino acids (MTNGKVYIVG) matched 100% of those predicted by the nucleotide sequence of the *cobA* gene. A M_r of 60 kDa, compatible with a homodimeric quaternary structure, was deduced for the SUMT enzyme by electrophoresis under non-denaturing conditions and by gel chromatography.

The *cobA*-encoded amino acid sequence was compared with that of other methylases involved in corrinoid biosynthesis. Highly conserved regions were present in all analyzed sequences. One of them, whose consensus is GXGXGD, has been described as a SAM-binding motif (Fig. 6), suggesting that the *G. stearothermophilus V cobA* gene product is actually an uroporphirinogen III-like C-methyltransferase. Appropriate primers were designed to introduce a change of A to G at position 12 of the SUMT enzyme by recombinant PCR. The product was cloned into the pET21b(+) expression vector and the A12G mutant protein was purified as above. As the wild-type counterpart, the SUMT A12G protein exhibited a homodimeric structure, as determined by non-denaturing polyacrylamide gel electrophoresis and size exclusion chromatography. Cells expressing the mutant SUMT showed low K_2TeO_3 resistance (MIC 2.5 µg/ml) as compared to the wild-type clone (MIC 18 µg/ml).

Various deletion mutants of the *cobA* gene were constructed by PCR that included 60-, 120- and 180-bp deletions from its 3' end. The truncated DNA fragments were amplified and cloned into the pET21b(+) expression vector. Plasmids carrying *cobA*Δ60, *cobA*Δ120 and *cobA*Δ180 were introduced into *E. coli* JM109(DE3) by transformation. Although expressed in high amounts, purification of the truncated proteins failed, as they formed inclusion bodies. Several attempts to solubilize them were carried out also without results. All clones expressing truncated genes were sensitive to tellurite (MIC 1.25 µg/ml).

Addition of methyl-^{3}H SAM to the purified SUMT enzyme followed by size exclusion chromatography revealed two radioactive peaks, corresponding to enzyme-bound and free SAM. When the wild-type enzyme was replaced by the SUMT A12G mutant in the SAM binding assay, only the free SAM peak was observed.

Another strategy of a cell to cope with the toxic effects of tellurite is to form volatile, less toxic, compounds with it. In this context, headspace analysis from cultures of different bacteria by gas chromatography- fluorine induced chemiluminescence detection (GC-

F2ICD) has proven to be useful for detecting the evolution of sulfur compounds, such as methanethiol (MeSH), dimethyl sulfide (DMS), dimethyl disulfide (DMDS), dimethyl trisulfide (DMTS), and organotellurides, like dimethyl telluride (DMTe) (Chasteen and Bentley, 2003).
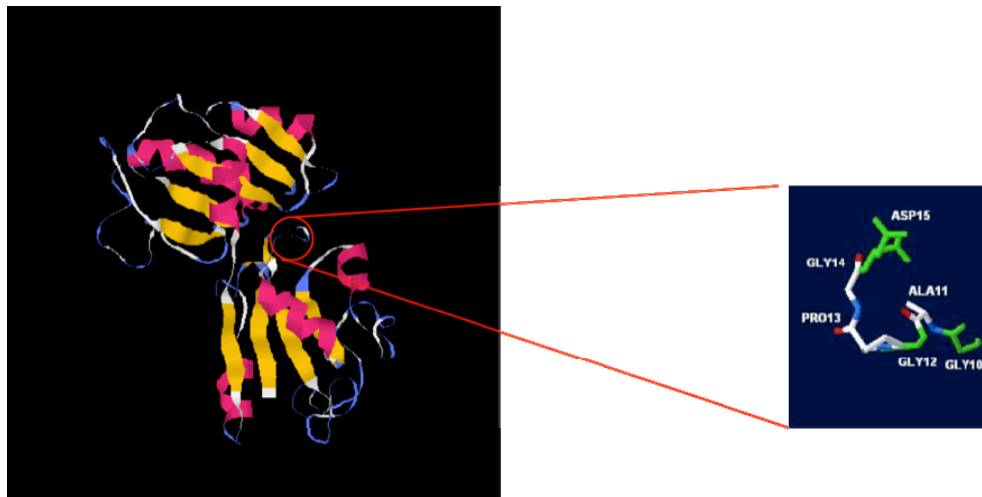
Fig. 6. Model (homology) of the SUMT methyl transferase. The inset shows the SAM-binding motif.

Given that SUMT is a methyltransferase, it was tempting to correlate it *a priori* with the resulting volatile Te derivatives. Since methylcobalamins (methyl-B12) have been involved in biomethylating a number of heavy metals and metalloids, one putative way by which SUMT could participate in tellurite resistance in *E. coli* could be precisely through this kind of methylation. However, two lines of evidence allow this assumption to be discarded. First, *E. coli* does not synthesize methyl-B12 *de novo*; and, second, amending the culture medium with 10 mM $CoCl_2$ did not change the K_2TeO_3 MIC. It is well known that Co salts inhibit any methyl-B12-mediated methylation. Thus, a putative role of SUMT in K_2TeO_3 resistance would be the utilization of Te as a substrate and to catalyze the transfer of methyl groups from SAM to the metalloid. In this context, some work from other authors has indicated that the gene products of the *tpm* and *tehB* genes from *Pseudomonas syringae* and *E. coli*, respectively, are able to biomethylate tellurium (Cournoyer et al., 1998; Liu et al., 2000). Unfortunately, we were unable to detect the genesis of methylated tellurium derivatives in the headspace of cells cultured in the presence of tellurite or tellurate. To date there is no clear experimental evidence that sheds light on the enzymatic mechanism underlying tellurium biomethylation.

On the other hand, since sulfite reductase (reduces sulfite to sulfur) utilizes siroheme as a prosthetic group, SUMT could participate in tellurite tolerance by enhancing the biosynthesis of this cofactor and, hence, that of cysteine. In the same context, it was found that enzymes that reduce thiols (glutathione and thioredoxin reductases) and their metabolites (thioredoxins, glutaredoxins and glutathione) would be involved in tellurite

resistance. Recent results from our laboratory indicate that when grown in the presence the toxicant, the total thiol content is higher in cells expressing the *cobA* gene than in cells carrying the vector alone. Our interpretation is that SUMT could participate in the generation of reducing power (cysteine, for example) that would be used to compensate (or to recover) GSH or another metabolically important thiol that could have been consumed during tellurite reduction.

2.4 LpdA

As mentioned before, one of the most relevant properties of potassium tellurite is its high toxicity for microorganisms. In this context, our approach to understand the basis of the toxic effects has been the search of resistance determinants in tellurite-resistant strains, such as *G. staerothermpophilus* V. Following the same idea, another highly TeO_3^{2-}-resistant bacterial strain was isolated from environmental water. This new strain exhibited a tellurite MIC close to 300 µg/ml. It was identified as the Gram-negative *Aeromonas caviae* ST. In addition to its high resistance to tellurite, this strain exhibited high levels of tellurite reduction, as determined by the darkness of cells exposed to the toxic salt and by tellurite reductase (TR) enzymatic assays performed with cell-free extracts. Interestingly, most of this TR activity was dependent of NADH and tracked to the pyruvate dehydrogenase multienzymatic complex (PDH), specifically to the E3 component encoded by the *lpdA* gene (Castro et al., 2008, 2009).

The *lpdA* gene was cloned; and the recombinant plasmid was used as template to construct three different mutants by site-directed mutagenesis: C45A, H322Y and E354K (Fig. 7). These mutants were chosen based on previous work on the *E. coli* E3 component indicating that C45 is highly conserved and is involved in the formation of a disulfide bond with C50, required for appropriate protein conformation (Kim et al., 2008). H322 and E354 were changed to Y and K, respectively, because it was previously shown that these mutations affect NADH binding, a substrate required for PDH as well as for TR activity (Castro et al., 2008, 2009).

In this case a different and easier approach to construct the mutants was carried out. Using a high-fidelity and highly processive DNA polymerase and two complementary primers, the plasmid was amplified by PCR and then the methylated template was digested with *Dpn*I restriction endonuclease and used to transform *E. coli*.

As expected, changes of these amino acids resulted in negative effects on pyruvate dehydrogenase activity in cells overproducing these proteins as compared to controls. Decreased PDH activity was observed, particularly in the cases of the H322Y and E354K mutants. The effect was not so pronounced in mutants that do not affect NADH binding. Regarding tellurite reductase activity (TR), an important decrease (~70%) was observed in all three mutants, as determined with purified proteins or in crude extracts of cells overproducing the respective mutant (Fig. 8). These results confirm the importance of NADH for PDH and TR activities and also indicated that C45, while relevant for LpdA-mediated tellurite reduction, is not absolutely required for PDH activity. This idea is in agreement with previous observations of our group and others regarding the importance of cysteine in tellurite resistance (Vásquez et al., 2001; Fuentes et al., 2007).
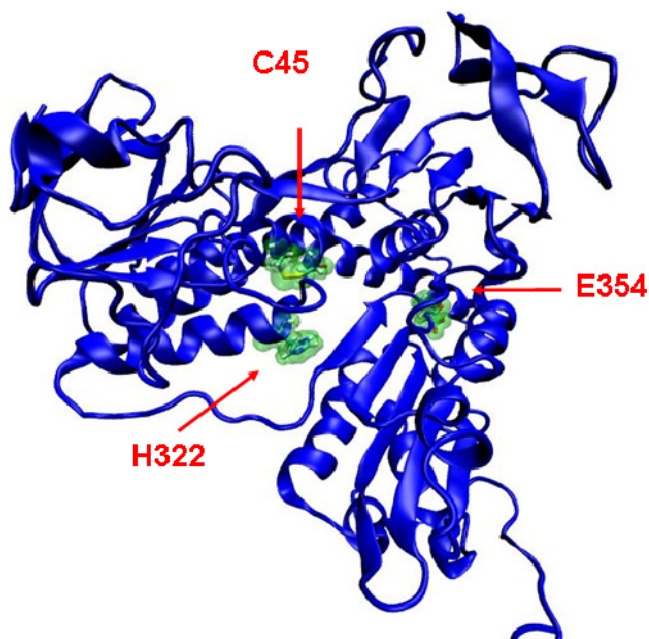
Fig. 7. LpdA model (homology) indicating the spatial position of the amino acids targeted for site-directed mutagenesis.
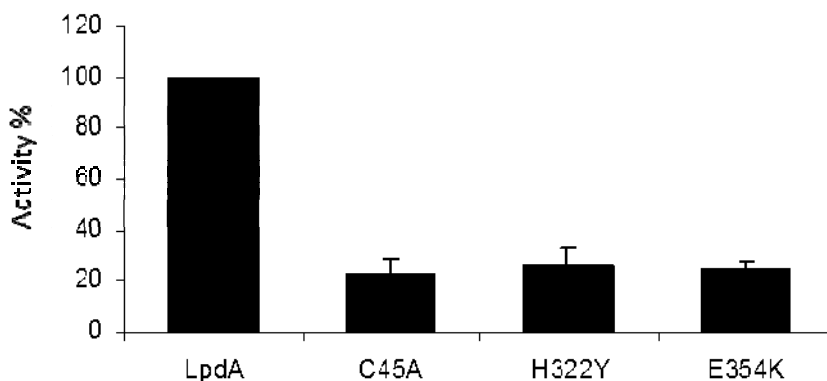
Fig. 8. Tellurite reductase activity of purified LpdA and the indicated LpdA mutants.

3. Conclusion

Several different mechanisms have been proposed to account for the toxicity of tellurite. Tellurium may replace sulfur and/or selenium in critical metabolites or enzymes and abate their essential functions. Alternatively, tellurite is a strong oxidizing agent that may cause general oxidative damage; or it may cause specific damage to critical thiol groups or [Fe-S]

clusters present in essential enzymes. The results of this chapter point out different instances in which diverse metabolic pathways, their substrates or products play a still not well-defined role in bacterial tellurite resistance.

4. Acknowledgments

5. References

Alexander, F.W., Sandmeier, E., Mehta, P.K. & Christen, P. (1994). Evolutionary relationships among pyridoxal-5'-phosphate-dependent enzymes. Regio-specific alpha, beta and gamma families. *Eur. J. Biochem.*, Vol. 219, No. 3, (February 1994), pp. 953-960, ISSN 0014-2956

Alonso, G., Gomes, C., González, C. & Rodríguez-Lemoine, V. (2000). On the mechanism of resistance to channel-forming colicins (PacB) and tellurite, encoded by plasmid Mip233 (IncHI3). *FEMS Microbiol. Lett.*, Vol. 192, No. 2, (November 2000), pp. 257-261, ISSN 0378-1097

Araya, M.A., Tantaleán, J.C., Fuentes, D.E., Pérez, J.M., Calderón, I.L., Saavedra, C.P., Chasteen, T.G. & Vásquez, C.C. (2009). Cloning, purification and characterization of *Geobacillus stearothermophilus* V uroporphirinogen-III C-methyltransferase: evaluation of its role in resistance to potassium tellurite in *Escherichia coli*. *Res. Microbiol.*, Vol. 160, No. 2, (March 2009), pp. 125-133, ISSN 0923-2508

Borsetti, F., Tremaroli,V., Michelacci, F., Borghese, R., Winterstein, C., Daldal, F. &, Zannoni, D. (2005). Tellurite effects on *Rhodobacter capsulatus* cell viability and superoxide dismutase activity under oxidative stress conditions. *Res Microbiol.*, Vol. 156, No. 7, (August 2005), pp. 807-813, ISSN 0923-2508

Calderón, I.L., Arenas, F.A., Pérez, J.M., Fuentes, D.E., Araya, M.A., Saavedra, C.P., Tantaleán, J.C., Pichuantes, S.E., Youderian, P.A. & Vásquez, C.C. (2006). Catalases are NAD(P)H-dependent tellurite reductases. *PLoS ONE*, Vol. 1, No. 1, (December 2006), pp. e70, ISSN 1932-6203

Castro, M.E., Molina, R., Díaz, W., Pichuantes, S.E. & Vásquez, C.C. (2008). The dihydrolipoamide dehydrogenase of *Aeromonas caviae* ST exhibits NADH dependent tellurite reductase activity. *Biochem. Biophys. Res. Commun.*, Vol. 375, No. 1, (October 2008), pp. 91-94, ISSN 0006-291X

Castro, M.E., Molina, R.C., Díaz, W.A., Pradenas, G.A. & Vásquez, C.C. (2009). Expression of *Aeromonas caviae* ST pyruvate dehydrogenase complex components mediate tellurite resistance in *Escherichia coli*. *Biochem. Biophys. Res. Commun.*, Vol. 380, No. 1, (February 2009), pp. 148-152, ISSN 0006-291X

Chasteen, T.G. & Bentley, R. (2003). Biomethylation of selenium and tellurium: microorganisms and plants. *Chem. Rev.*, Vol. 103, No. 1, (January 2003), pp. 1-25, ISSN 0009-2665

Chasteen, T.G., Fuentes, D.E., Tantaleán, J.C. & Vásquez, C.C. (2009). Tellurite: history, oxidative stress and molecular mechanisms of resistance. *FEMS Microbiol. Rev.*, Vol. 33, No. 4, (July 2009), pp. 820-832, ISSN 1574-6976

Chiong, M., Barra, R., González, E. & Vásquez C. (1988). Resistance of *Thermus* spp. to potassium tellurite. *Appl. Environ. Microbiol.*, Vol. 54, No. 2, (February 1988), pp. 610-612, ISSN 0099-2240

Cook, P.F. & Wedding R.T (1977). Overall mechanism and rate equation for O-acetylserine sulfhydrylase. *J. Biol. Chem.*, Vol. 252, No. 10, (May 1977), pp. 3549-3553, ISSN 0021-9258

Cournoyer, B., Watanabe, S. & Vivian, A. (1998). A tellurite-resistance genetic determinant from phytopathogenic pseudomonads encodes a thiopurine methyltransferase: evidence of a widely-conserved family of methyltransferases. *Biochim Biophys Acta.*, Vol. 1397, No. 2, (April 1998), pp. 161-168, ISSN 0304-4165

Fuentes, D.E., Fuentes, E.L., Castro, M.E., Pérez, J.M., Araya, M.A., Chasteen, T.G., Pichuantes, S.E. & Vásquez, C.C. (2007). Cysteine metabolism-related genes and bacterial resistance to potassium tellurite. *J. Bacteriol.*, Vol. 189, No. 24, (October 2007), pp. 8953-8960, ISSN 0021-9193

Kim, Y., Ingram, L. & Shanmugam, T. (2008). Dihydrolipoamide dehydrogenase mutation alters the NADH sensitivity of pyruvate dehydrogenase complex of *Escherichia coli* K-12. *J. Bacteriol.*, Vol. 190, No. 11, (March 2008), pp. 3851-3858, ISSN 0021-9193

Lithgow, J.K., Hayhurst, E.J., Cohen, G., Aharonowitz, Y. & Foster, S.J. (2004). Role of a cysteine synthase in *Staphylococcus aureus*. *J. Bacteriol.* Vol. 186, No. 6, (March 2004), pp. 1579-1590, ISSN 0021-9193

Liu, M., Turner, R.J., Winstone, T.L., Saetre, A., Dyllick-Brenzinger, M., Jickling, G., Tari, L.W., Weiner, J.H. & Taylor, D.E. (2000). *Escherichia coli* TehB requires S-adenosylmethionine as a cofactor to mediate tellurite resistance. *J. Bacteriol.*, Vol. 182, No. 22, (November 2000), pp. 6509-6513, ISSN 0021-9193

Moore, M. & Kaplan, S. (1992). Identification of intrinsic high-level resistance to rare-earth oxides and oxyanions in members of the class Proteobacteria: characterization of tellurite, selenite, and rhodium sesquioxide reduction in *Rhodobacter sphaeroides*. *J. Bacteriol.*, Vol. 174, No. 5, (March 1992), pp. 1505-1514, ISSN 0021-9193

Moscoso, H., Saavedra, C., Loyola, C., Pichuantes, S. & Vásquez C. (1998). Biochemical characterization of tellurite-reducing activities of *Bacillus stearothermophilus* V. *Res. Microbiol.*, Vol. 149, No. 6, (June 1998), pp. 389-397, ISSN 0923-2508

O'Gara, J., Gomelsky, M. & Kaplan, S. (1997). Identification and molecular genetic analysis of multiple loci contributing to high-level tellurite resistance in *Rhodobacter sphaeroides* 2.4.1. *Appl. Environ. Microbiol.*, Vol. 63, No. 12, (December 1997), pp. 4713-4720, ISSN ISSN 0099-2240

Pérez, J.M., Calderón, I.L., Arenas, F.A., Fuentes, D.E., Pradenas, G.A., Fuentes, E.L., Sandoval, J.M., Castro, M.E., Elías, A.O. & Vásquez, C.C. (2007). Bacterial toxicity of potassium tellurite: unveiling an ancient enigma. *PLoS ONE*. Vol. 2, No. 2, (February 2007), pp. e211, ISSN 1932-6203

Saavedra, C.P., Encinas, M.V., Araya, M.A., Pérez, J.M., Tantaleán, J.C., Fuentes, D.E., Calderón, I.L., Pichuantes, S.E. & Vásquez, C.C. (2004). Biochemical characterization of a thermostable cysteine synthase from *Geobacillus stearothermophilus* V. *Biochimie*, Vol. 86, No. 7, (July 2004), pp. 481-485, ISSN 0300-9084

Silver S. (1996). Bacterial resistances to toxic metal ions--a review. *Gene*, Vol. 179, No. 1, (November 1996), pp. 9-19, ISSN 0378-1119

Silver, S. (2011). BioMetals: a historical and personal perspective. *Biometals*, Vol. 24, No. 3, (June 2011), pp. 379-390, OnlineISSN 1572-8773

Tai, C.H., Yoon, M.Y., Kim, S.K., Rege, V.D., Nalabolu, S.R., Kredich, N.M., Schnackerz, K.D. & Cook, P.F. (1998). Cysteine 42 is important for maintaining an integral active site for O-acetylserine sulfhydrylase resulting in the stabilization of the alpha-aminoacrylate intermediate. *Biochemistry*, Vol. 37, No. 30, (July 1998), pp. 10597-10604, ISSN 0006-2960

Tantaleán, J.C., Araya, M.A., Saavedra, C.P., Fuentes, D.E., Pérez, J.M., Calderón, I.L., Youderian, P. & Vásquez, C.C. (2003). The *Geobacillus stearothermophilus* V *iscS* gene, encoding cysteine desulfurase, confers resistance to potassium tellurite in *Escherichia coli* K-12. *J. Bacteriol.*, Vol. 185, No. 19, (October 2003), pp. 5831-5837, ISSN 0021-9193

Taylor, D.E. (1999). Bacterial tellurite resistance. *Trends Microbiol.*, Vol. 7, No. 3, (March 1999), pp. 111-115, ISSN 0966-842X

Tremaroli, V., Fedi, S. & Zannoni, D. (2007). Evidence for a tellurite-dependent generation of reactive oxygen species and absence of a tellurite-mediated adaptive response to oxidative stress in cells of *Pseudomonas pseudoalcaligenes* KF707. *Arch Microbiol.*, Vol. 187, No. 2, (February 2007), pp. 127-135, ISSN 1432-072X

Vásquez, C., Saavedra, C., Loyola, C., Araya, M. & Pichuantes, S. (2001). The product of the *cysK* gene of *Bacillus stearothermophilus* V mediates potassium tellurite resistance in *Escherichia coli*. Curr. Microbiol., Vol. 43, No. 6, (December 2001), pp. 418-421, ISSN 1432-0991

Molecular Genetics in Disease-Related Research

A Mutagenesis Approach for the Study of the Structure-Function Relationship of Human Immunodeficiency Virus Type 1 (HIV-1) Vpr

Kevin Hadi[1], Oznur Tastan[2],
Alagarsamy Srinivasan[3] and Velpandi Ayyavoo[1,*]
[1]University of Pittsburgh, Pittsburgh, PA,
[2]Microsoft Research, Cambridge, MA,
[3]NanoBio Diagnostics, West Chester, PA,
USA

1. Introduction

Before the era of molecular biology, the methods available for an understanding of gene function were limited. Such studies typically relied on the ability to identify and isolate naturally occurring variants exhibiting a defect in function. Hence, the progress was slow in this regard. The discoveries in the field of microbiology, combined with advances in technology in the later part of 20th century, dramatically changed this scenario. The current molecular biological techniques enable site-directed mutagenesis approaches for generating a gene with a specific amino acid substitution, mutation, or a deletion or for truncating a gene anywhere in a matter of days.

The present review highlights the studies conducted on human immunodeficiency virus type 1 (HIV-1) Vpr, an auxiliary protein associated with virus particles. Vpr contains 96 amino acids, and it is a multifunctional protein. To analyze the contribution of specific residues to protein function and to the cytopathic effects in HIV-1-infected individuals, investigators from several laboratories, including ours, took advantage of a novel approach. Specifically, this approach involved exchange of residues through mutagenesis. The choice of residue was based on the information available regarding the naturally occurring polymorphisms at the level of individual amino acids in Vpr. The results from these studies support a link between polymorphisms in these genes and the disease status of infected individuals, who are known as progressors or non-progressors. In addition the studies have shed light on the structure-function relationship of Vpr.

The Joint United Nations Program on HIV/AIDS (UNAIDS) (2010) reports that the worldwide prevalence of those living with human immunodeficiency virus (HIV) is between 31-35 million as of the end of 2009. Roughly 2.6 million new cases of HIV infection occurred in the same year. That number has likely remained approximately constant. Although the most common route of HIV infection is via sexual contact, the use of

---

* Corresponding Author

contaminated drug paraphernalia, mother-child transmission via pregnancy or breastfeeding, and tainted blood transfusions comprise other means of infection.

The symptomatic outcome of infection is AIDS, usually occurring ~10 years after initial infection. CD4+ T-cell counts drop below 200, and subsequent severe immune dysfunction results. This eventually leads to fatal coinfection by opportunistic pathogens. The advent of highly active retroviral therapy (HAART) in the 1990s led to a drastically improved prognosis for AIDS patients (Peters and Conway, 2011). This triple-drug cocktail controls viremia and allows immune function to recover to nearly uninfected levels, with the caveat of near-perfect patient adherence to a difficult combination of drug regimens. The extremely rigid treatment schedules have resulted in low compliance, leading to the emergence of viruses that exhibit resistance to drugs.

2. Genetic organization of HIV

The causative agents of AIDS have been identified as human immunodeficiency virus types 1 (HIV-1) and 2 (HIV-2). Both HIV-1 and HIV-2 are members of the lentivirus family of retroviruses. HIV-1 is the predominant virus responsible for AIDS throughout the world. The schematic representation of the genome organization of HIV-1 is shown in **Figure 1.** The genome of HIV-1 codes for two regulatory proteins (Tat and Rev) and four auxiliary proteins (Vif, Vpr Vpu and Nef), in addition to structural proteins Gag, Gag-Pol and Env. The genome organization of HIV-2 is similar to HIV-1. The unique genes are *vpu* and *vpx* for HIV-1 and HIV-2, respectively. With respect to viral morphogenetic events, the lentiviruses are similar to alpharetroviruses, gammaretroviruses and deltaretroviruses. During virus infection, Gag and Gag-Pol proteins are synthesized in the cytoplasm and transported to the cell membrane, where virus assembly occurs. In the case of HIV-1, the non-structural proteins Vpr, Vif, and Nef are also packaged into the virus particles.
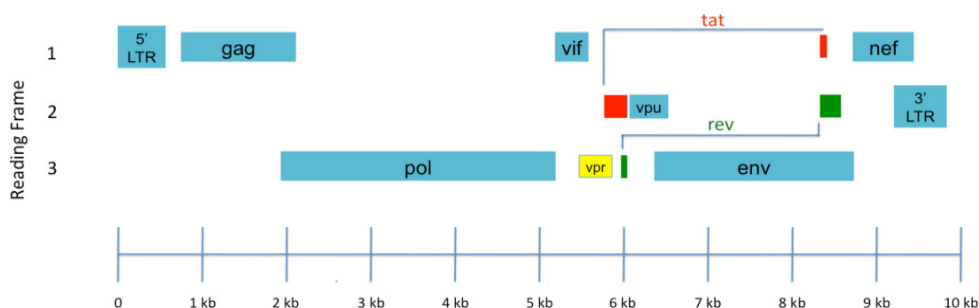
Fig. 1. Organization of the HIV-1 Genome.

3. Heterogeneity in the human immunodeficiency virus

Within HIV-1-positive patients, the high error rate of reverse transcription can produce many variants, or quasispecies. After seroconversion to HIV-positive status, the viral loads measure at 10000-50000 viral RNA copies per ml of patient sera during the asymptomatic

stage; in the later stages of disease, viral RNA genomes can increase to several million copies per ml (Poropatich and Sullivan, 2011; Tungaturthi et al., 2004). Thus, a high replication rate produces tremendous variation in the viral population within a single patient. The pressures of the host immune response drive the selection of variants from early stages of infection, until the host immune response cannot cope with the viral diversity (Fischer et al., 2010; Wolinsky et al., 1996).

Studies of polymorphisms in a number of the genes of HIV-1 confirm the propensity of the virus to escape host immunity (Fischer et al., 2010; Fischer et al., 2007; Gaschen et al., 2002; Korber et al., 2001). As expected, the antigenic Env protein exhibits the highest mutability. Comparisons of current strains of HIV show a staggering 20% variability within a subtype and up to 35% variability between subtypes (Gaschen et al., 2002). To highlight the problem of this variability in vaccine development, the evolutionary dynamics of influenza virus provides a revealing picture. The influenza genome varies by 1-2% per year, enabling influenza virus escape from polyclonal vaccine responses and necessitating annual vaccine changes (Fischer et al., 2007; Gaschen et al., 2002). Along with intra- and inter-subtype variation in HIV, *env* shows on average a 10% change in genetic diversity over the course of an infection in a single patient (Korber et al., 2001). In addition to the *env* gene, Korber et al. (2001) estimated the percent variation in sequences culled from an HIV-1 database compared to the HXB2 strain consensus in *tat* and *gag*. They found 9% and 5% variation, respectively. From another study, *gag* and *pol* remained relatively conserved; but the rest of the genes exhibited high variability comparable to that of *env* (Gaschen et al., 2002) **(Figure 1)**. The vast genetic diversity seen in these studies thus far exemplifies the major obstacle to vaccine development.

The variation in the viral genome throughout the population potentially correlates with the variation among HIV-infected individuals. In the progressor group, in which AIDS develops 10 years after initial infection, CD4+ T-cell counts (the primary marker of AIDS progression) generally fall below 200 cells per µl and causing loss of cell-mediated immunity (Levy, 2009). Along with the normal progressors, several other categories of disease progression exist. Long term non-progressors (LTNPs) and elite controllers (EC) do not receive HAART and do not show the aforementioned clinical signs of AIDS for up to 20 years after infection. CD4+ T-cell counts are maintained to levels above 350 cells per µl and viral loads to under 2000 copies per ml. Viral loads in these groups can range as low as 50-2000 copies per ml (Poropatich and Sullivan, 2011). In rapid progressors (RP), CD4+ T-cell counts generally decline 3-8 years postinfection. Although individual host genetics likely play a role in differential disease progression (Goulder et al., 1997), the presence of multiple quasispecies in patients potentially explains the differences in disease outcome.

4. Polymorphisms in specific HIV-1 genes

Several genes of HIV have a functional role in disease symptoms (Caly et al., 2008; Casartelli et al., 2003; Tolstrup et al., 2006). As Nef downregulates MHC class I antigen presentation involved in generating a cytotoxic T-lymphocyte (CTL) response, polymorphisms in this gene correlate with the LTNP group (Caly et al., 2008; Casartelli et al., 2003; Tolstrup et al., 2006). Similarly the role of Tat as a transactivator of the HIV-1 promoter in the long-terminal-repeat (LTR) region and its genetic heterogeneity suggest a functional role for Tat in disease progression and outcome (Bratanich et al., 1998; Korber et al., 2001). Tat can

upregulate LTR-transcriptional activity and thus replication. Cellular and viral transcription increases several hundredfold (Irish et al., 2009) in the presence of Tat. Various Tat proteins isolated from different subtypes show differences in the ability to induce viral and cellular gene transcription (Roof et al., 2002).

The multifunctional Viral Protein R (Vpr), although labeled as an accessory protein, has a vital role in efficient replication of the virus in the non-dividing macrophages and monocytes (Balliet et al., 1994; Connor et al., 1995). Vpr interacts with the Gag protein for packaging into the virion. It remains bound to the preintegration complex (PIC) upon entry of the viral contents into the target cell during infection, most likely aiding PIC entry into the nucleus. It has a highly functional role in replication as a transactivator of the LTR, binding to a number of host transcription factors, such as Sp1 (Sawaya et al., 1998) and TFIIB (Agostini et al., 1999). Extensive studies have shown that coactivation of glucocorticoid receptor by Vpr, resulting in transcription of elements in the promoter regions of the HIV LTR and host genes, enhances viral replication and downregulates host immune responses (Ayyavoo et al., 1997; Hapgood and Tomasicchio, 2010). Vpr also drives apoptosis and G2 cell cycle arrest (Morellet et al., 2009; Pandey et al., 2009; Tungaturthi et al., 2004). The processes enhance immune escape and HIV-1 replication, respectively. As Vpr seems to be essential in the viral life cycle, it is likely that *vpr* gene sequences will exhibit signature changes among the categories of HIV-1 disease progression. Due to the importance of Vpr, this review focuses mainly on the polymorphisms in *vpr*, their effects on Vpr function, and their potential consequences on disease outcome. The review will also show how using an *in vitro* model to study polymorphisms may help us better understand the mechanisms and develop therapeutics for treatment.

5. Methods used to generate mutations to study gene regulation

Identifying genes responsible for different phenotypes previously relied on isolating naturally occurring polymorphic genes within the organisms. This process is time-consuming and requires sequencing several isolates to identify the functionally relevant genes. Today there are many different means of following the general path of altering the genotype to observe the phenotype.

5.1 Site-directed mutagenesis

The use of a point mutation to observe the change of function that occurs as a result of a change in structure underscores the majority of studies on the role of polymorphisms in viral and host gene function. One method of generating a targeted mutation in a plasmid construct carrying the gene of interest uses a polymerase chain reaction (PCR) technique. This particular PCR-based method has the advantage of using a high-fidelity polymerase and methylation of DNA by common strains of *Escherichia coli.* The protocol is described as part of the commercially available QuikChange II™ Site-Directed Mutagenesis Kit (Agilent, 2010). PCR results in the amplification of daughter DNA strands that carry the mutation of interest (Agilent, 2010). In the initial PCR amplification step, two primers containing the mutation of interest bind to the complementary strands of the plasmid. This requires a mismatch of at least one nucleotide on the primers and, furthermore, optimization of the length of the primer for a suitable T^m for the annealing step of the PCR. The mismatch limits

the number of mutations that can be made per reaction to about four bases, depending on the length and T^m of the primer.

PCR will copy the entire parental template. The high-fidelity *Pfu Ultra* polymerase has an error rate of one per 2.5×10^6 nucleotides (Agilent, 2011), making the errors introduced by PCR a non-issue. Once the amplification of the template is complete, the PCR product contains the original parental DNA strands and the mutated and amplified daughter DNA strands. To select against the parental strands, the protocol takes advantage of the fact that the mutated and amplified daughter strands are unmethylated. In contrast, the parental (template) strands are methylated, as they originated from methylation-capable *E. coli*. Digesting the PCR product with *Dpn*I endonuclease, which digests methylated and hemi-methylated DNA, will cleave the parental strands and leave the daughter strands intact. Since the amplified DNA has the plasmid sequence of the parent and since there is an overlap region to allow circularization by homologous recombination, plasmids with the mutation are then isolated by transformation. Any candidate must then be screened and sequenced to confirm the mutagenesis.

6. Structure and function relationship of HIV-1 Vpr

To arrive at the structure of Vpr, several NMR studies have used fragments representing different segments of the Vpr protein. Alternatively the entire Vpr molecule has been analyzed in an appropriate solvent (Morellet et al., 2003; Wecker et al., 2002). The analyses showed a flexible N-terminus with a turn, the first alpha helix, turn, the second alpha helix, turn, the third alpha helix, then a flexible C-terminal domain (Morellet et al., 2003; Wecker et al., 2002). Morellet et al. (2003) used a different solvent, allowing them to "see" the tertiary structure of Vpr. The solvent was better at revealing the hydrophobic parts that are implicated in dimerization and interaction with other proteins (Morellet et al., 2003). Ideally NMR studies of protein structure employ solvents and conditions that approximate physiological conditions. However, the oligomerization property of Vpr makes proper solvation extremely difficult. Several studies have used trifluoroethanol (TFE), a hydrophobic solvent, in different proportions to counteract the interaction of the hydrophobic domains of Vpr, some of which cause Vpr to oligomerize (Engler et al., 2001; Wecker et al., 2002). Across these studies the length of the alpha helices differed due to the proportion of TFE in the solvents used.

Morellet et al. (2003) used CD3CN, a solvent with little hydrophobicity that approaches physiological conditions. This solvent allowed the following structural analysis: alpha helices 1, 2, and 3 (17-33, 38-50, 54-77), a flexible N-terminus (1-16) and a basic C-terminus (78-96). Each helix is amphipathic, containing hydrophobic and hydrophilic residues. The hydrophobic residues can allow for interactions with other proteins, most likely other cellular factors. In the third helix, L60, L67, I74, and I81 can form a leucine zipper (Morellet et al., 2003). The N-terminus consists largely of acidic residues, while the C-terminus consists largely of basic residues, arginine being the most prominent. The 3-D structure established by Morellet et al. (2003; 2009) shows a globular structure, in which hydrophobic interactions between the helices form a lipophilic core. Each of the helices in its amphipathic portions has acidic/basic residues, which are on the external face of the modeled protein. They possibly provide contacts for interactions with other proteins. The folding of the alpha helices entails that hydrophobic portions of the protein face outwards in contact with the

solvent. These unfavorable conditions could be ameliorated by binding to cellular partners. Mutational analyses of the protein have found many residues throughout the length of Vpr that maintain the structural integrity (Morellet et al., 2009). Single deletions of Y15 (see Table 1 in the chapter by Figurski et al. for the amino acid codes), K27, and Q44 result in disruption of structure. The helical domains contain residues crucial to structure, especially those that form the hydrophobic core. The basic C-terminal residues also affect structure and stability of Vpr.

6.1 Oligomerization

The property of forming Vpr oligomers has recently come under scrutiny. Zhao et al. (1994) incorporated mutagenesis in their study to elucidate a rough map of residues in Vpr necessary for this function. Deletion of individual residues in the 36-76 region diminishes formation of oligomers. Mutagenesis of individual resides in the leucine-isoleucine motif ([60]LIRILQQLLFIHFR) in the third helix (amino acids 55-77) reduced the self-associative capacity of Vpr monomers. The residues at position 60, 61, 63, 64, 67, 68, 70, and 74 mutated from leucine or isoleucine to alanine or histidine disrupted binding between monomers (Zhao et al., 1994). The Q44 residue also plays an important role in Vpr oligomerization. Structural analysis of the second alpha helix (amino acids 38-50) reveals a hydrophilic glutamine at the 44th position (Morellet et al., 2009). Deletion of this residue by site-directed mutagenesis disrupted the secondary structure and abolished Vpr—Vpr interaction (Fritz et al., 2008). Fritz et al. (2008) revealed via 3-D modeling that the ΔQ44 mutation destabilizes the formation of the hydrophobic core and the self-interaction of the helices, thus providing an explanation for the inability of such mutants to oligomerize. Although Vpr oligomerization plays a role in other functions, this group did not find a relation between oligomerization and the ability of Vpr to induce apoptosis.

Oligomerization seems necessary for other functions, such as nuclear localization and virion incorporation (Fritz et al., 2008; Fritz et al., 2010; Venkatachari et al., 2010). Venkatachari et al. (2010) hypothesized from a structural model of the oligomerization of Vpr that residues at the predicted helical interfaces contribute to dimerization. Substitution mutagenesis of A30 to leucine abolished dimerization. The authors attributed this to the position of A30 on the external face of the tertiary structure of Vpr, where it likely affects protein-binding. Furthermore, elimination of dimerization of Vpr abolished the ability of the protein to be incorporated into HIV virions. It also diminished its nuclear localization. These results implicate a necessary role of Vpr oligomerization in its incorporation into virions and, by extension, possibly playing a role in the infection of non-replicating immune cells, important targets of HIV-1.

6.2 Nuclear import

The nuclear import of the preintegration complex (PIC) upon entry of the virus enables productive infection to occur in targeted host cells, particularly non-dividing immune cells (*e.g.*, macrophages/monocytes) (Bukrinsky et al., 1992). As Vpr is bound to the PIC, Vpr shuttles the viral contents into the nucleus, where integration occurs (Popov et al., 1998). An earlier study established the necessity of Vpr in productive viral replication in macrophages (Connor et al., 1995), eventually leading to the identification of importin-α as an essential host factor in this process (Nitahara-Kasahara et al., 2007). Sherman et al. (2001) mapped

non-canonical nuclear localization signals (NLS) throughout the helical domains and the C-terminus of Vpr. This group used site-directed mutagenesis of its various domains in order to identify particular residues that function in nuclear import. In the C-terminus of Vpr, mutagenesis to change several arginine residues (those at positions 73, 76, 77, 85, 87, 88, and 90) to alanine resulted in the distribution of Vpr throughout the cell or cytoplasm, suggesting the importance of this segment of Vpr. In the N-terminus, leucine motifs with the consensus sequence LXXLL in the first and third helices enable nuclear localization. The sequences for each motif are as follows: in the first helix, [22]LLEEL[26]; and in the third helix, [64]LQQLL[68]. The following mutations disrupted nuclear localization most drastically: L22A, L23A, L26A, L64A, and L68A. Via site-directed mutagenesis, the authors established the importance of residues in both helices that are needed for Vpr translocation.

7. G2 cell cycle arrest

The G2 phase of the cellular life cycle serves as a checkpoint for the cell. Factors can halt the cell cycle progression into mitosis in the presence of excessive DNA damage. If these factors do not detect damage, the cell divides; but detection of chromatin disruption will activate factors, such as ATM (ataxia telangiectasia mutated) and ATR (ataxia telangiectasia mutated and Rad-3 related). These factors have downstream effects. Ultimately they hyper-phosphorylate the Cdc2-Cyclin B1 complex, the major controller of cell progression into mitosis. Hyper-phosphorylation inactivates Cdc2-Cyclin B1 to prevent cell division. This process arrests the cell in the G2 phase (Morellet et al., 2009; Pandey et al., 2009; Sherman et al., 2002). Vpr expression in various cell types leads to G2 cell cycle arrest. It does so through increased expression of p21 through the p53 pathway, a major regulator of progression through the G2 and M phases (Chowdhury et al., 2003). Vpr seems to act synergistically with p53, perhaps inducing transcription of p21 through its own transactivation mechanisms. Although the purpose of the G2 arrest function of the virus is debatable, evidence suggests that it enhances transcription from the viral promoter, the long terminal repeat (LTR) of the HIV genome (Goh et al., 1998). One mechanism by which this occurs is through enhancing transactivation of the LTR in CD4+ T-cells via Vpr itself. Vpr binds to other host transcription factors, which bind to sites in the LTR, allowing viral transcription to occur. This effect is enhanced during G2 arrest (Gummuluru and Emerman, 1999).

The C-terminal portion of the protein seems to be essential for induction of cell cycle arrest. Zhou and Ratner (2000) showed that the phosphorylated S79 residue is necessary for this function. Their mutagenesis study of substituting alanine for serine eliminated phosphorylation at this residue and abolished arrest in the G2 phase, correlating the two functions with each other. Furthermore, the mutation of G75A impaired G2 cell cycle arrest, as shown by Mahalingam et al. (1997). DeHart et al. (2007) proposed that Vpr hijacks the ubiquitin/proteasome pathway. One of the functions of this pathway is to target proteins for degradation and to cause G2 cell cycle arrest when needed. In their model, Vpr binds DCAF1, a subunit of the Cullin-4 E3 ubiquitin ligase, to lead to ubiquitination of an unknown host factor involved in halting the cell cycle progression. Change of Q65 to arginine eliminated this binding and impaired G2 cell cycle arrest. However, the group also generated an R80A substitution that also disrupted G2 cell cycle arrest; but this mutant Vpr maintained the ability to bind DCAF1. The authors interpreted these results to indicate that binding DCAF1 is necessary, but not sufficient, to cause G2 cell cycle arrest. A recent study

revealed a highly conserved motif in the C-terminus ([79]SRIG[82]), in which mutations that change each amino acid prohibit G2 cell cycle arrest. This study corroborates previous evidence of the functional role of the R80 residue (DeHart et al., 2007; Maudet et al., 2011). Mutagenesis studies using the substitutions R73A and R80A eliminated the induction of p21 transcription, which is necessary for induction of G2 cell cycle arrest.

While the C-terminal portion of Vpr seems necessary for G2 cell cycle arrest, several of the previously cited studies and others revealed that certain N-terminal residues may be indispensable for this function. At least two studies indicated that mutagenesis of A30 to leucine also abolished cell cycle arrest and reduced the transcription of p21 (Chowdhury et al., 2003; Mahalingam et al., 1997). As noted above, A30L eliminated oligomerization, which suggests a correlation between the functions of oligomerization and G2 cell cycle arrest. Also K27M (methionine for lysine 27) another substitution in the first helix, disables the induction of G2 cell cycle arrest. Overall the evidence indicates that N-terminal and C-terminal moieties influence the conformational binding determinants of Vpr involved in G2 cell cycle arrest.

8. Apoptosis

Apoptosis (programmed cell death) functions as a primary means of maintaining homeostasis among cells. Apoptosis can occur as a result of irreparable DNA damage or by disruption of essential cellular processes, such as transcription or translation. Infection by HIV disrupts the normal induction of apoptosis. Despite disrupting apoptosis, infected CD4+ T-cells still die (Groux et al., 1992; Jamieson et al., 1997). However, bystander cells not directly infected by HIV-1, such as CD8+ T-cells, neurons and other cell types, undergo apoptosis (Finkel et al., 1995; Zhang et al., 2003). The Vpr protein was shown to play a key role in the induction of apoptosis in several *in vitro* studies. Vpr was able to permeabilize the mitochondrial membrane and activate caspase 9 via cytochrome C release (Chen et al., 1999; Jacotot et al., 2000; Macreadie et al., 1995; Stewart et al., 1997). Jacotot et al. (2000) established the necessity of several arginine residues inside and between a repeated H(S/F)RIG motif in the Vpr amino acid sequence, specifically [71]HFRIG[75] and [78]HSRIG[82]. Site-directed mutagenesis to generate the substitutions R73A, R77A, and R80A abolished the apoptotic effect of Vpr on Jurkat cells through the pathway involving the mitochondria.

Correlation of the function of Vpr that influences apoptosis with the function that promotes G2 cell cycle arrest remains a complicated affair. Jacquot et al. (2007) reported experiments that indicated a correlation. The group generated several Vpr substitution mutants that abolished both cell cycle arrest and apoptosis. Their results were consistent with the model of G2 cell cycle arrest leading to the induction of apoptosis. Mutants K27M and A30L in the N-terminus of Vpr and R80A and R90K in the C-terminus disabled cytostatic capacity and reduced apoptosis in a T-cell line. However, Bolton and Lenardo (2007) reported that the Vpr with the R80A substitution attenuated apoptotic effects in Jurkat cells, although it remained G2 arrest-capable. Also Maudet et al. (2011) showed that the apoptotic function of alanine mutants in the [79]SRIG[82] motif, including R80, and the K27M substitution mutant remained intact without G2 cell cycle arrest. The evidence from such mutagenesis studies may indicate that, although G2 cell cycle arrest leads to apoptosis, a G2 cell cycle arrest-independent apoptotic pathway exists.

Interestingly, Maudet et al. (2011) showed that the ability of Vpr to bind DCAF1 is not only necessary for induction of G2 cell cycle arrest, but for apoptosis as well. The substitution mutant Q65R eliminated DCAF1 binding, abolishing cell death. The S79A and K27M mutants retained their ability to cause apoptosis while losing G2 arrest function. Mutagenesis was done to produce the double mutants K27M/Q65R and S79A/Q65R. These mutants eliminated the G2 cell cycle arrest-independent induction of apoptosis. As DCAF1 is essential for the targeting of proteins to the proteasome, the authors propose a model in which Vpr binds DCAF1 at a region containing the Q65 residue and functions as an adaptor to the ubiquitin/proteasome complex. Their mutagenesis studies suggest that Vpr contains two different binding domains that interact with two separate, and as yet unidentified, host targets. This ostensibly leads to ubiquitination and subsequent degradation of these proteins. Degradation of target 1 leads to G2 cell cycle arrest, and ubiquitination of target 2 leads to apoptosis. However, these targets have yet to be identified. The model implies complex pathways.
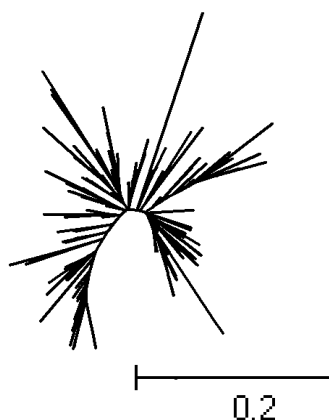
Fig. 2. Phylogeny of *vpr* Sequences Across Clades

9. Vpr polymorphisms across subtypes

A phylogenetic analysis of *vpr* sequences from clades A-D (including the highly prevalent subtype B shows a large diversity of sequences across the genetic lineages of HIV-1 subtypes **(Figure 2).** This indicates the existence of quasispecies of Vpr existing in the population and selective pressures acting on the *vpr* gene. Upon closer analyses of the sequences within the tree, several of the species between subtypes of Vpr show closer genetic relationships than other species within subtype. The variations in Vpr may correlate more to disease progression than to the clades. A comparative analysis of the frequencies of *vpr* alleles from long-term non-progressors (LTNP) and rapid progressors (RP) indicates the presence of mutations that are of interest **(Table 1)**.

Position	Consensus Sequence of LTNP alignment Residue	Residue conserved In the LTNP alignment (n=177)	Residue In Consensus Seq PR alignment	Residue conserved In the PR alignment (n=102)
2	E	174	E	101
3	Q	168	Q	96
4	A	175	A	96
5	P	176	P	101
6	E	166	E	100
7	D	170	D	91
8	Q	175	Q	101
9	G	165	G	102
10	P	176	P	102
11	Q	174	Q	98
12	R	167	R	102
13	E	168	E	100
14	P	177	P	100
15	Y	172	Y	96
16	N	158	N	96
17	E	169	E	96
18	W	175	W	102
19	T	153	T	93
20	L	173	L	102
21	E	174	E	102
22	L	170	L	95
23	L	170	L	102
24	E	169	E	100
25	E	172	E	102
26	L	173	L	101
27	K	172	K	102
28	N	74	S	40
29	E	171	E	101
30	A	170	A	102
31	V	171	V	101
32	R	167	R	100
33	H	174	H	101
34	F	174	F	101
35	P	175	P	101
36	R	171	R	88
37	I	50	V	51
38	W	174	W	101
39	L	175	L	99
40	H	175	H	88
41	S	89	G	71
42	L	171	L	101
43	G	171	G	101

A Mutagenesis Approach for the Study of the Structure-Function
Relationship of Human Immunodeficiency Virus Type 1 (HIV-1) Vpr

213

Position	Consensus Sequence of LTNP alignment Residue	Residue conserved In the LTNP alignment (n=177)	Residue In Consensus Seq PR alignment	Residue conserved In the PR alignment (n=102)
44	Q	175	Q	101
45	H	118	H	72
46	I	174	I	101
47	Y	173	Y	101
48	E	146	E	92
49	T	174	T	98
50	Y	175	Y	101
51	G	169	G	101
52	D	175	D	100
53	T	175	T	100
54	W	174	W	98
55	A	123	A	58
56	G	171	G	100
57	V	174	V	100
58	E	164	E	97
59	A	166	A	96
60	I	149	I	62
61	I	167	I	96
62	R	170	R	96
63	I	136	I	53
64	L	176	L	99
65	Q	173	Q	97
66	Q	174	Q	96
67	L	174	L	98
68	L	174	L	82
69	F	176	F	100
70	I	145	I	93
71	H	173	H	100
72	F	174	F	99
73	R	174	R	100
74	I	172	I	97
75	G	158	G	95
76	C	171	C	99
77	R	106	R	52
78	H	174	H	100
79	S	175	S	98
80	R	172	R	99
81	I	171	I	97
82	G	174	G	100
83	I	171	I	99
84	T	99	T	66

Position	Consensus Sequence of LTNP alignment Residue	Residue conserved In the LTNP alignment (n=177)	Residue In Consensus Seq PR alignment	Residue conserved In the PR alignment (n=102)
85	R	111	Q	40
86	Q	143	Q	83
87	R	168	R	94
88	R	168	R	97
89	A	140	A	83
90	R	169	R	95
91	N	170	N	56
92	G	171	G	57
93	A	153	A	55
94	S	162	S	51
95	R	173	R	58†

Table 1†. Frequency Analysis of Amino Acids Resulting from *vpr* Alleles Found in the LTNP and RP Groups

An approach using site-directed mutagenesis to study these substitutive polymorphisms will link these variants to possible effects on Vpr function and to the long-term non-progressor and rapid progressor statuses.

10. Mutagenesis studies in the context of disease progression and pathogenesis

Wang et al. (1996) analyzed sequences of the *vpr* genes from an HIV-infected mother-child pair who showed no sign of AIDS from initial infection in 1983 to the time of the study in 1995. These investigators found that samples from the mother and child had homogeneous and similar length polymorphisms in the C-terminal region of Vpr. However these polymorphisms were not present in samples from 30 patients who developed AIDS. Several other studies showed marked heterogeneity in *vpr* sequences derived from multiple patient samples (Ge et al., 1996; Kuiken et al., 1996). In addition to length polymorphisms, numerous studies have demonstrated an association between substitution mutants in the amino acid sequence of Vpr and disease (Caly et al., 2008; Lum et al., 2003; Tungaturthi et al., 2004).

Caly et al. (2008) found a mutation, F72L, in several *vpr* sequences derived from a single patient with LTNP status. This mutation seems to be correlated with the disruption of nuclear import. However only a limited conclusion about prognosis can be drawn from this result because of the small sample size. Yedavalli and Ahmad (2001) reported several mutations from *vpr* sequences extracted from HIV-infected mothers with LTNP status who did not transmit the infection to the child during labor. They found polymorphisms from two separate patients that led to the A30S and G75R substitutions in Vpr. They also found

---

† Amino acid abbreviations follow standard conventions from the International Union of Pure and Applied Chemistry (IUPAC) (Nomenclature, 1968).

C-terminal deletions at an abundant frequency in another sample. This finding implies the possibility of structural changes to Vpr that might alter function in the context of disease progression. The involvement of Vpr at all stages of the HIV-1 life cycle suggests that this protein influences the development of disease and the severity of the outcome. Our hypothesis is that with a sufficient sample size, assessment of *vpr* sequences derived from the LTNP and RP groups will reveal signature polymorphisms that may be linked to progression of AIDS *in vivo*. This may occur through the disruption or enhancement of hallmark Vpr functions. Already several groups have suggested that unique, signature polymorphisms in Vpr culled from LTNP patient samples are associated with the reduction of host cell apoptosis (Lum et al., 2003; Somasundaran et al., 2002).

However, given the complexities of the interrelated pathways by which Vpr induces apoptosis, as revealed by the mutagenesis studies detailed above, it is likely that the two published studies explain only a fraction of the causative factors behind the progression of HIV disease. Many questions need to be answered before the role of Vpr in disease induction is clear.

11. References

Agilent (2010). QuikChange II Site-Directed Mutagenesis Kit: Instruction Manual, Revision C.

Agilent (2011). *PfuUltra* High-Fidelity DNA Polymerase AD, Revision B.

Agostini, I., J. M. Navarro, M. Bouhamdan, K. Willetts, F. Rey, B. Spire, R. Vigne, R. Pomerantz and J. Sire (1999). "The HIV-1 Vpr Co-activator Induces a Conformational Change in TFIIB." *FEBS Letters* 450(3): 235-239.

Ayyavoo, V., A. Mahboubi, S. Mahalingam, R. Ramalingam, S. Kudchodkar, W. Williams, D. Green and D. Weiner (1997). "HIV-1 Vpr Suppresses Immune Activation and Apoptosis through Regulation of Nuclear Factor Kappa B." *Nature Medicine* 3(10): 1117-23.

Balliet, J. W., D. L. Kolson, G. Eiger, F. M. Kim, K. A. McGann, A. Srinivasan and R. Collman (1994). "Distinct Effects in Primary Macrophages and Lymphocytes of the Human Immunodeficiency Virus Type 1 Accessory Genes Vpr, Vpy, and Nef: Mutational Analysis of a Primary HIV-1 Isolate." *Virology* 200(2): 623-631.

Bolton, D. L. and M. J. Lenardo (2007). "Vpr Cytopathicity Independent of G2/M Cell Cycle Arrest in Human Immunodeficiency Virus Type 1-Infecteed CD4+ T Cells." *Journal of Virology* 81(17): 8878-8890.

Bratanich, A. C., C. Liu, J. C. McArthur, T. Fudyk, J. D. Glass, S. Mittoo, G. A. Klassen and C. Power (1998). "Brain-Derived HIV-1 *Tat* Sequences from AIDS patients with Dementia Show Increased Molecular Heterogeneity." *Journal of Neurovirology* 4(4): 387-393.

Bukrinsky, M. I., N. Sharova, M. P. Dempsey, T. L. Stanwick, A. G. Bukrinskaya, S. Haggerty and M. Stevenson (1992). "Active Nuclear Import of Human Immunodeficiency virus type 1 preintegration complexes." *Proceedings of the National Academy of Science* 89(14): 6580-6584.

Caly, L., N. Saksena, S. Piller and D. Jans (2008). "Impaired nuclear import and viral incorporation of Vpr derived from a HIV long-term non-progressor." *Retrovirology* 5(1): 67.

Casartelli, N., G. Di Matteo, M. Potesta, P. Rossi and M. Doria (2003). "CD4 and Major Histocompatibility Complex Class I Downregulation by the Human Immunodeficiency Virus Type 1 Nef Protein in Pediatric AIDS Progression." *J. Virol.* 77(21): 11536-11545.

Chen, M., R. T. Elder, M. G. O'Gorman, L. Selig, R. Benarous, A. Yamamoto and Y. Zhao (1999). "Mutational analysis of Vpr-induced G2 arrest, nuclear localization, and cell death in fission yeast." *Journal of Virology* 73(4): 3236-3245.

Chowdhury, I. H., X.-F. Wang, N. R. Landau, M. L. Robb, V. R. Polonis, D. L. Birx and J. H. Kim (2003). "HIV-1 Vpr Activates Cell Cycle Inhibitor p21/Waf1/Cip1: A Potential Mechanism of G2/M Cell Cycle Arrest." *Virology* 305(2): 371-377.

Cohen, E. A., G. Dehni, J. G. Sodroski and W. A. Haseltine (1990). "Human immunodeficiency virus vpr product is a virion-associated regulatory protein." *J. Virol.* 64(6): 3097-3099.

Connor, R. I., B. K. Chen, S. Choe and N. R. Landau (1995). "Vpr Is Required for Efficient Replication of Human Immunodeficiency Virus Type-1 in Mononuclear Phagocytes." *Virology* 206(2): 935-944.

DeHart, J., E. Zimmerman, O. Ardon, C. Monteiro-Filho, E. Arganaraz and V. Planelles (2007). "HIV-1 Vpr activates the G2 checkpoint through manipulation of the ubiquitin proteasome system." *Virology Journal* 4(1): 57.

Engler, A., T. Stangler and D. Willbold (2001). "Solution structure of human immunodeficiency virus type 1 Vpr(13-33) peptide in micelles." *European Journal of Biochemistry* 268(2): 389-395.

Finkel, T. H., G. Tudor-Williams, N. K. Banda, M. F. Cotton, T. Curiel, C. Monks, T. W. Baba, R. M. Ruprecht and A. Kupfer (1995). "Apoptosis occurs predominantly in bystander cells and not in productively infected cells of HIV- and SIV-infected lymph nodes." *Nature Medicine* 1(2): 129-134.

Fischer, W., V. V. Ganusov, E. E. Giorgi, P. T. Hraber, B. F. Keele, T. Leitner, C. S. Han, C. D. Gleasner, L. Green, C.-C. Lo, A. Nag, T. C. Wallstrom, S. Wang, A. J. McMichael, B. F. Haynes, B. H. Hahn, A. S. Perelson, P. Borrow, G. M. Shaw, T. Bhattacharya and B. T. Korber (2010). "Transmission of Single HIV-1 Genomes and Dynamics of Early Immune Escape Revealed by Ultra-Deep Sequencing." *PLoS ONE* 5(8): e12303.

Fischer, W., S. Perkins, J. Theiler, T. Bhattacharya, K. Yusim, R. Funkhouser, C. Kuiken, B. Haynes, N. L. Letvin, B. D. Walker, B. H. Hahn and B. T. Korber (2007). "Polyvalent Vaccines for Optimal Coverage of Potential T-Cell Epitopes in Global HIV-1 Variants." *Nature Medicine* 13(1): 100-106.

Fritz, J. V., P. Didier, J.-P. Clamme, E. Schaub, D. Muriaux, C. Cabanne, N. Morellet, S. Bouaziz, J.-L. Darlix, Y. Mely and H. d. Rocquigny (2008). "Direct Vpr-Vpr Interaction in Cells Monitored by two Photon Fluorescence Correlation Spectroscopy and Fluoescence Lifetime Imaging." *Retrovirology* 5(87).

Fritz, J. V., D. Dujardin, J. Godet, P. Didier, J. D. Mey, J.-L. Darlix, Y. Mely and H. d. Rocquigny (2010). "HIV-1 Vpr Oligomerization but Not That of Gag Directs the Interaction between Vpr and Gag." *Journal of Virology* 84(3): 1585-1596.

Gaschen, B., J. Taylor, K. Yusim, B. Foley, F. Gao, D. Lang, V. Novitsky, B. Haynes, B. H. Hahn, T. Bhattacharya and B. Korber (2002). "Diversity Considerations in HIV-1 Vaccine Selection." *Science* 296(5577): 2354-2360.

Ge, Y. C., B. Wang, D. E. Dwyer, S.-H. Xiang, A. L. Cunningham and N. Saksena (1996). "Sequence Note: Length Polymorphism of the Viral Protein R of Human Immunodeficiency Virus Type 1 Strains." *AIDS Research and Human Retroviruses* 12(4): 351-354.

Goh, W. C., M. E. Rogel, C. M. Kinsey, S. F. Michael, P. N. Fultz, M. A. Nowak, B. H. Hahn and M. Emerman (1998). "HIV-1 Vpr increases viral expression by manipulation of the cell cycle: A mechanism for selection of Vpr *in vivo*." *Nature Medicine* 4(1): 65-71.

Goulder, P. J. R., R. E. Phillips, R. A. Colbert, S. McAdam, G. Ogg, M. A. Nowak, P. Giangrande, G. Luzzi, B. Morgana, A. Edwards, A. J. McMichael and S. Rowland-Jones (1997). "Late Escape from an Immunodominant Cytotoxic T-Lymphocyte Response Associated with Progression to AIDS." *Nature Medicine* 3(2): 212-217.

Groux, H., G. Torpier, D. Monte, Y. Mouton, A. Capron and J. C. Ameisen (1992). "Activation-induced death by apoptosis in CD4+ T cells from human immunodeficiency virus-infected asymptomatic individuals." *Journal of Experimental Medicine* 175(2): 331-340.

Gummuluru, S. and M. Emerman (1999). "Cell Cycle- and Vpr-Mediated Regulation of Human Immunodeficiency Virus Type 1 Expression in Primary and Transformed T-Cell Lines." *J. Virol.* 73(7): 5422-5430.

Hapgood, J. P. and M. Tomasicchio (2010). "Modulation of HIV-1 Virulence Via the Host Glucocorticoid Receptor: Towards Further Understanding the Molecular Mechanism of HIV-1 Pathogenesis." *Archives of Virology* 155(7): 1009-1019.

HIV/AIDS, T. J. U. N. P. o. (2010). Global Report: UNAIDS Report on the Global AIDS Epidemic 2010, Joint United Nations Programme on HIV/AIDS (UNAIDS).

Irish, B. P., Z. K. Khan, P. Jain, M. R. Nonnemacher, V. Pirrone, S. Rahman, N. Rajagopalan, J. B. Suchitra, K. Mostoller and B. Wigdahl (2009). "Molecular Mechanisms of Neurodegenerative Diseases Induced by Human Retroviruses: A review." *American Journal of Infectious Diseases* 5(3): 231-258.

Jacotot, E., L. Ravagnan, M. Loeffler, K. F. Ferri, H. L. A. Vieira, N. Zamzami, P. Costantini, S. Druillennec, J. Hoebeke, J. P. Briand, T. Irinopoulou, E. Daugas, S. A. Susin, D. Cointe, Z. H. Xie, J. C. Reed, B. P. Roques and G. Kroemer (2000). "The HIV-1 Viral Protein R Induces Apoptosis via a Direct Effect on the Mitochondrial Permeability Transition Pore." *The Journal of Experimental Medicine* 191(1): 33-46.

Jacquot, G., E. Le Rouzic, A. David, J. Mazzolini, J. Bouchet, S. Bouaziz, F. Niedergang, G. Pancino and S. Benichou (2007). "Localization of HIV-1 Vpr to the nuclear envelope: Impact on Vpr functions and virus replication in macrophages." *Retrovirology* 4(1): 84.

Jamieson, B. D., C. H. Uittenbogaart, I. Schmid and J. A. Zack (1997). "High viral burden and rapid CD4+ cell depletion in human immunodeficiency virus type 1-infected SCID-hu mice suggest direct viral killing of thymocytes in vivo." *Journal of Virology* 71(11): 8245-8253.

Korber, B., B. Gaschen, K. Yusim, R. Thakallapally, C. Kesmir and V. Detours (2001). "Evolutionary and immunological implications of contemporary HIV-1 variation." *British Medical Bulletin* 58(1): 19-42.

Kuiken, C. L., M. T. E. Cornelissen, F. Zorgdrager, S. Hartman, A. J. Gibbs and J. Goudsmit (1996). "Consistent risk group-associated differences in human immunodeficiency virus type 1 vpr, vpu and V3 sequences despite independent evolution." *Journal of General Virology* 77(4): 783-792.

Levy, J. A. (2009). "HIV Pathogenesis: 25 Years of Progress and Persistent Challenges." *AIDS* 23(2): 147-160.

Lum, J. J., O. J. Cohen, Z. Nie, J. G. Weaver, T. S. Gomez, X.-J. Yao, D. Lynch, A. A. Pilon, N. Hawley, J. E. Kim, Z. Chen, M. Montpetit, J. Sanchez-Dardon, E. A. Cohen and A. D. Badley (2003). "Vpr R77Q is associated with long-term nonprogressive HIV infection and impaired induction of apoptosis." *The Journal of Clinical Investigation* 111(10): 1547-1554.

Macreadie, I. G., L. A. Castelli, D. R. Hewish, A. Kirkpatrick, A. C. Ward and A. A. Azad (1995). "A domain of human immunodeficiency virus type 1 Vpr containing repeated H(S/F)RIG amino acid motifs causes cell growth arrest and structural defects." *Proceedings of the National Academy of Sciences* 92(7): 2770-2774.

Mahalingam, S., V. Ayyavoo, M. Patel, T. Kieber-Emmons and D. B. Weinter (1997). "Nuclear Import, Virion Incorporation, and Cell Cycle Arrest/Differentiation Are Mediated by Distinct Functional Domains of Human Immunodeficiency Virus Type 1 Vpr." *Journal of Virology* 71(9): 6339-6347.

Maudet, C., M. Bertrand, E. Le Rouzic, H. Lahouassa, D. Ayinde, S. Nisole, C. Goujon, A. Cimarelli, F. Margottin-Goguet and C. Transy (2011). "Molecular Insight into How HIV-1 Vpr Protein Impairs Cell Growth through Two Genetically Distinct Pathways." *Journal of Biological Chemistry* 286(27): 23742-23752.

Morellet, N., S. Bouaziz, P. Petitjean and B. P. Rogues (2003). "NMR Structure of the HIV-1 Regulatory Protein VPR." *Journal of Molecular Biology* 327(1): 217-227.

Morellet, N., B. P. Roques and S. Bouaziz (2009). "Structure-Function Relationship of Vpr: Biological Implications." *Current HIV Research* 7(2): 184-210.

Nitahara-Kasahara, Y., M. Kamata, T. Yamamoto, X. Zhang, Y. Miyamoto, K. Muneta, S. Iijima, Y. Yoneda, Y. Tsunetsugu-Yokota and Y. Aida (2007). "Novel Nuclear Import of Vpr Promoted by Importin {alpha} Is Crucial for Human Immunodeficiency Virus Type 1 Replication in Macrophages." *J. Virol.* 81(10): 5284-5293.

Nomenclature, I.-I. C. o. B. (1968). "A one-letter notation for amino acid sequences. Tentative rules." *Biochemistry* 7(8): 2703-2705.

Pandey, R. C., D. Datta, R. Mukerjee, A. Srinivasan, S. Mahalingam and B. E. Sawaya (2009). "HIV-1 Vpr: A Closer Look at the Multifunctional Protein from the Structural Perspective." *Current HIV Research* 7(2): 114-128.

Pandori, M., N. Fitch, H. Craig, D. Richman, C. Spina and J. Guatelli (1996). "Producer-cell modification of human immunodeficiency virus type 1: Nef is a virion protein." *J. Virol.* 70(7): 4283-4290.

Peters, B. S. and K. Conway (2011). "Therapy for HIV: Past, Present, and Future." *Advances in Dental Research* 23(1): 23-27.

Popov, S., M. Rexach, L. Ratner, G. Blobel and M. Bukrinsky (1998). "Viral Protein R Regulates Docking of the HIV-1 Preintegration Complex to the Nuclear Pore Complex." *Journal of Biological Chemistry* 273(21): 13347-13352.

Poropatich, K. and D. J. Sullivan (2011). "Human immunodeficiency virus type 1 long-term non-progressors: the viral, genetic and immunological basis for disease non-progression." *Journal of General Virology* 92(2): 247-268.

Roof, P., M. Ricci, P. Genin, M. A. Montano, M. Essex, M. A. Wainberg, A. Gatignol and J. Hiscott (2002). "Differential Regulation of HIV-1 Clade-Specific B, C, and E Long Terminal Repeats by NF-κB and the Tat Transactivator." *Virology* 296(1): 77-83.

Sawaya, B. E., K. Khalili, W. E. Mercer, L. Denisova and S. Amini (1998). "Cooperative Actions of HIV-1 Vpr and p53 Modulate Viral Gene Transcription." *Journal of Biological Chemistry* 273(32): 20052-20057.

Sherman, M. P., C. M. C. de Noronha, M. I. Heusch, S. Greene and W. C. Greene (2001). "Nucleocytoplasmic Shuttling by Human Immunodeficiency Virus Type 1 Vpr." *J. Virol.* 75(3): 1522-1532.

Sherman, M. P., C. M. C. D. Noronha, S. A. Williams and W. C. Greene (2002). "Insights into the Biology of HIV-1 Viral Protein R." *DNA and Cell Biology* 21(9): 679-688.

Somasundaran, M., M. Sharkey, B. Brichacek, K. Luzuriaga, M. Emerman, J. L. Sullivan and M. Stevenson (2002). "Evidence for a cytopathogenicity determinant in HIV-1 Vpr." *Proceedings of the National Academy of Sciences* 99(14): 9503-9508.

Stewart, S. A., B. Poon, J. B. Jowett and I. S. Chen (1997). "Human immunodeficiency virus type 1 Vpr induces apoptosis following cell cycle arrest." *Journal of Virology* 71(7): 5579-5592.

Strebel, K. (2003). "Virus-host interactions: role of HIV proteins Vif, Tat, and Rev." *AIDS* 17(4): S25-S34.

Tolstrup, M., A. L. Laursen, J. Gerstoft, F. S. Pedersen, L. Ostergaard and M. Duch (2006). "Cysteine 138 Mutation in HIV-1 Nef from Patients with Delayed Disease Progression." *Sex Health* 3(4): 281-286.

Tungaturthi, P. K., B. E. Sawaya, V. Ayyavoo, R. Murali and A. Srinivasan (2004). "HIV-1 Vpr: Genetic Diversity and Functional Features from the Perspective of Structure." *DNA and Cell Biology* 23(4): 207-222.

Venkatachari, N. J., L. A. Walker, O. Tastan, T. Le, T. M. Dempsey, Y. Li, N. Yanamala, A. Srinivasan, J. Klein-Seetharaman, R. C. Montelaro and V. Ayyavoo (2010). "Human immunodeficiency virus type 1 Vpr: oligomerization is an essential feature for its incorporation into virus particles." *Virology Journal* 7(119).

Wang, B., Y. C. Ge, P. Palasanthiran, S.-H. Xiang, J. Ziegler, D. E. Dwyer, C. Randle, D. Dowton, A. Cunningham and N. K. Saksena (1996). "Gene Defects Clustered at the C-Terminus of the vpr Gene of HIV-1 in Long-Term Nonprogressing Mother and Child Pair:In VivoEvolution of vpr Quasispecies in Blood and Plasma." *Virology* 223(1): 224-232.

Wecker, K., N. Morellet, S. Bouaziz and B. P. Rogues (2002). "NMR Structure of the HIV-1 Regulatory Protein Vpr in H2O/trifluoroethanol. Comparison with the Vpr N-terminal (1-51) and C-terminal (52-96) domains." *European Journal of Biochemistry* 269(15): 3779-3788.

Wolinsky, S. M., B. T. M. Korber, A. U. Neumann, M. Daniels, K. J. Kunstman, A. J. Whetsell, M. R. Furtado, Y. Cao, D. D. Ho, J. T. Safrit and R. A. Koup (1996). "Adaptive Evolution of Human Immunodeficiency Virus-Type 1 During the Natural Course of Infection." *Science* 272(5261): 537-542.

Yedavalli, V. R. K. and N. Ahmad (2001). "Low Conservation of Functional Domains of HIV Type 1 vif and vpr Genes in Infected Mothers Correlates with Lack of Vertical Transmission." *AIDS Research and Human Retroviruses* 17(10): 911-923.

Zhang, K., F. Rana, C. Silva, J. Ethier, K. Wehrly, B. Chesebro and C. Power (2003). "Human immunodeficiency virus type 1 envelope-mediated neuronal death: uncoupling of viral replication and neurotoxicity." *Journal of Virology* 77(12): 6899-6912.

Zhao, L.-J., L. Wang, S. Mukherjee and O. Narayan (1994). "Biochemical Mechanism of HIV-1 Vpr Function: Oligomerization Mediated by the N-Terminal Domain." *The Journal of Biological Chemistry* 269(51): 32131-32137.

Zhou, Y. and L. Ratner (2000). "Phosphorylation of Human Immunodeficiency Virus Type 1 Vpr Regulates Cell Cycle Arrest." *J. Virol.* 74(14): 6520-6527.

New Insights into the Epithelial Sodium Channel Using Directed Mutagenesis

Ahmed Chraibi and Stéphane Renauld
University of Sherbrooke, Québec
Canada

1. Introduction

Directed mutagenesis is a fundamentally important DNA technology that seeks to change the base sequence of DNA and test the effect of the change on gene or DNA function. It can be accomplished using the polymerase chain reaction (PCR). For more than 20 years, many applications in both basic and clinical research have been revolutionized by PCR. The development of this technique allowed the substitution, addition or deletion of single or multiple nucleotides in DNA (Mullis and Faloona, 1987). Because of redundancy in the genetic code, such mutations do not always alter the primary structure of proteins. In this chapter, we will review the contribution of PCR-directed mutagenesis in the determination of the structure-function relationship of the epithelial sodium channel (ENaC), particularly with respect to the domains involved in proteolytic activation and ligand-induced stimulation of the channel.

2. Physiological role and structure of ENaC

ENaC is a key component of the transepithelial sodium transport. It is expressed at the apical membrane of a variety of tissues, such as the distal nephron of the kidney, lungs, exocrine glands (e.g., sweat and salivary glands) (Brouard et al., 1999; Duc et al., 1994; Perucca et al., 2008; Roudier-Pujol et al., 1996) and distal colon (Kunzelmann and Mall, 2002). In aldosterone-sensitive distal nephron (ASDN) and distal colon, this channel plays a major role in the control of sodium balance and blood pressure (Frindt and Palmer 2003; Garty and Palmer 1997). In lungs, ENaC regulates mucus secretion and aids in the protection of the airway surface (Randell and Boucher, 2006). Its role was clearly demonstrated in mice in which the ENaC gene was inactivated by homologous recombination (Hummler et al., 1996). ENaC belongs to a gene family with members found throughout the animal kingdom, the so-called ENaC/degenerin family, including the acid sensing ion channel (ACIC) and the Phe-Arg-Met-Phe amide-gated ion channel (FaNaCh), (Kellenberger and Schild. L, 2002). Using the *Xenopus* oocyte expression system and a distal colon cDNA library, the primary structure of ENaC was identified; and electrophysiologic characteristics of ENaC channel were determined (Canessa et al., 1993; Canessa et al., 1994; Lingueglia et al., 1993). ENaC is a heteromeric channel made of three subunits (α, β and γ) encoded by 3 different genes SCNN1a, SCNN1b and SCNN1g, respectively. Each subunit exhibits ~30% identity at the amino acid level and shares highly conserved domains. The

membrane topology of each subunit predicts the presence of two transmembrane domains (M1 and M2), a large extracellular loop (~70% of the size of the channel) and relatively short amino and carboxyl termini. The stoichiometry of ENaC was much discussed: several examples of biochemical and functional evidence are consistent with a heterotetrameric structure (2α, 1β, 1γ) (Anantharam A, 2007; Dijkink et al., 2002; Firsov et al., 1998), but octameric or nonameric structures have also been suggested (Eskandari et al., 1999; Snyder et al., 1998). Recent crystallographic data obtained on the related ASIC1 channel suggest ENaC most likely exists functionally as an αβγ heterotrimer complex (Jasti et al., 2007; Stockand et al., 2008). ENaC is characterized by high sodium selectivity ($P_{Na+}/P_{K+} > 100$), a low single-channel conductance (4-5 pS), gating kinetics characterized by long opening and closing times, and a specific block by amiloride (*Ki*: 100 -200 nM).

Sodium homeostasis requires that the entry of sodium through the apical membrane of epithelial cells is tightly controlled. This control may be realized by regulation of ENaC activity and expression. The role of different domains involved in this regulation has been determined by directed mutagenesis.

3. Mutations in ENaC subunits cause hereditary human disease

The role of ENaC in the regulation of blood pressure and regulation of extracellular fluid volume has been highlighted by the discovery of two severe human diseases. The diseases are due to loss or gain of function of ENaC. Homozygous inactivating mutations in the α, β or γ ENaC subunits cause pseudohypoaldosteronism type 1 (PHA-1), characterized by hypotension and severe hyperkalemic acidosis (Chang et al., 1996). Activating mutations in the genes for the β or γ ENaC subunits lead to Liddle's syndrome, characterized by autosomal-dominant hypertension accompanied by hypokalemic Alkalosis and volume expansion (Shimkets et al., 1994).

The mutations causing PHA-1 have been identified, and the mechanisms by which they led to a hypofunction of ENaC have been addressed. See (Kellenberger and Schild. L, 2002) for review.

In particular Chang et al. (1996) showed that a single point mutation (G37S) in the coding region for a highly conserved motif in the amino-terminal domain of the β subunit induces PHA-1. Grunder and co-authors (1997) showed that this domain is involved in the gating of ENaC. They identified that the mutation G37S in the gene for the β subunit and homologous mutations in the other subunit genes reduce channel function by changing the open probability.

Liddle syndrome has been linked genetically to mutations that delete or alter a conserved PY (proline-tyrosine) motif located in the carboxy-terminal domain of either β or γENaC (Hansson et al., 1995; Hansson et al., 1995; Shimkets et al., 1994; Tamura et al., 1996). Such deletions or point mutations lead to elevated channel function after expression in Xenopus oocytes, suggesting that the PY motif is involved in the regulation of activity and the density of ENaC channels at the cell surface (Firsov et al., 1996; Kellenberger et al., 1998; Schild et al., 1995; Schild et al., 1996; Shimkets et al., 1997). Mutations within the coding region for the PY motif were generated *in vitro* by directed mutagenesis. They have been widely studied to investigate the role of Nedd4-2 in the regulation of the number of ENaCs at the cell surface (Abriel and Horisberger, 1999; Debonneville et al., 2001; Kamynina and Staub, 2002;

Renauld et al., 2010; Staub et al., 1997). Thus, this has allowed the identification of tyrosine-carrying ubiquitin residues involved in Nedd4-2 dependent-internalization of the channel.

Intracellular C termini also harbor multiple phosphorylation sites and participate in the activity of the channel, suggesting that aldosterone, insulin, SGK1, PKA and PKC modulate the activity of ENaC by phosphorylation (Renauld et al., 2010; Shimkets et al., 1997).

4. Directed mutagenesis and regulation of ENaC by extracellular factors

Several members of the ENaC/degenerin family are clearly extracellular-ligand-gated channels. Numerous studies suggest that ENaC may also be a ligand-gated channel (Horisberger and Chraibi, 2004). A number of extracellular factors of various types have been shown to activate or inhibit ENaC. Amongst these factors, there are serine proteases, sodium itself, other inorganic cations, organic cations, and small molecules.

4.1 Activation by Serine proteases

In 1997 we cloned a serine protease that acts as a channel-activating protease, called CAP1 (Vallet et al., 1997); and we explored the mechanism by which it stimulates ENaC (Chraibi et al., 1998). We showed that the effect of CAP1 is done on the extracellular part of the channel, and it can be mimicked by trypsin or chymotrypsin. Ion selectivity, single channel conductance and channel density are not modified, which suggests that the serine proteases increase the open probability. During the last ten years, many studies showed that ENaC can be activated by other proteases, such as prostasin or furin (Hughey et al., 2004; Vuagniaux et al., 2000). Further progress in the understanding of the mechanism by which serine proteases activate ENaC has been made by functional investigation in heterologous expression systems combined with directed mutagenesis. Mutation of the CAP1 GPI-anchored consensus motif completely abolishes ENaC activation. However, catalytic mutants of CAP1 do not fully stimulate ENaC, suggesting that a noncatalytic mechanism is partly involved in this regulation pathway (Vallet et al., 2002). Thus, a putative site for CAP1 and trypsin action has been identified. However, there is no clear evidence of their role in the proteolytic activation of ENaC. Masilamani and co-authors (1999) first provided evidence for a possible cleavage of the γENaC subunit. These authors were able to show that the aldosterone infusion, or salt restriction, induced a shift in molecular weight of the gamma subunit from 85 to 70 KDa. Subsequently, it was shown that the serine proteases, including prostasin, plasmin, elastase and furin, cleave the extracellular domain of the α and γ subunits (Bruns et al., 2007; Caldwell et al., 2005; Hughey et al., 2004; Passero et al., 2008; Rossier, 2004; Vuagniaux et al., 2002). A basic motif (RKRK[186]) has been identified as a cleavage site for CAP1/Prostasin in the extracellular loop of γENaC (Bruns et al., 2007; Diakov et al., 2008). Additional cleavage sites within extracellular loop of α and γ subunits have been described (Garcia-Caballero et al., 2008; Myerburg et al., 2006). However, no site for furin was described in βENaC (see Figure 1).

4.2. Effects of extracellular sodium and other small molecules

4.2.1 Self-inhibition

We have shown that the external sodium exerts a fast inhibitory effect on ENaC activity, a phenomenon called sodium self-inhibition (Chraibi and Horisberger, 2002). We observed

that the apparent affinity constant for the site responsible for self-inhibition was significantly lower, with a K½ of 100-200 mM. The kinetics of this phenomenon strongly depended on temperature and the extent of proteolytic processing of the ENaC subunits. We demonstrated that the effect of temperature was due to a large decrease in the probability of channel opening at high temperatures, while the unitary current increased with temperature (Chraibi and Horisberger, 2003). Later Sheng et al. (2004, 2002) showed that the mutation of His282 in the α subunit or His239 in the γ subunit (these amino acids reside in close proximity to the defined sites for furin cleavage) enhanced and eliminated the sodium self-inhibition response, respectively.
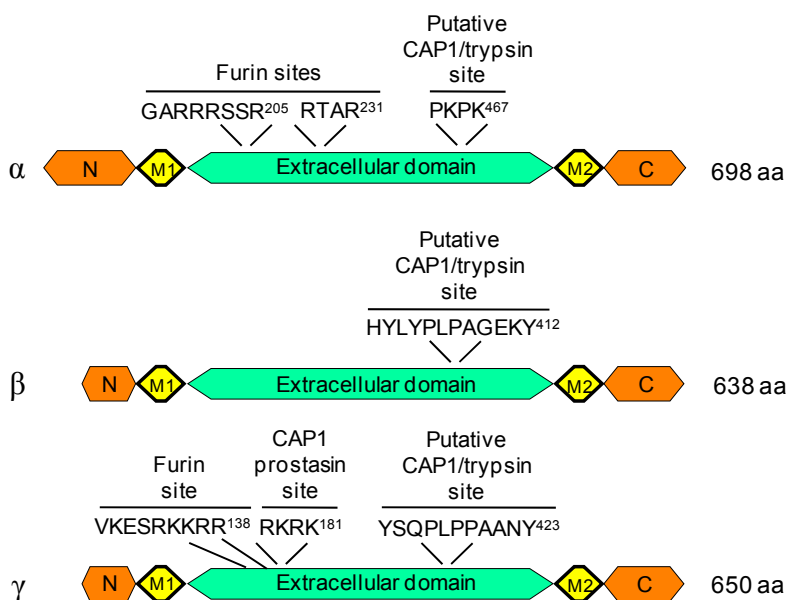
Fig. 1. Schematic representation of the rat ENaC subunits and their identified and putative sites for furin, CAP1, trypsin and prostasin. M1, M2: transmembrane domains; N, C: intracellular amino- and carboxy-termini, respectively.

4.2.2 Effects of cpt-cAMP and cpt-cGMP

cpt-cAMP, a membrane permeant cAMP analogue, has been described to be a species-dependent extracellular activator of ENaC. Rat and *Xenopus laevis* ENaC expressed in *Xenopus* oocytes are not sensitive to cpt-cAMP (Awayda et al., 1996). However, guinea pig (gp) channels could be activated by cpt-cAMP perfusion in the oocyte expression system (Liebold et al., 1996). The gp αENaC has been shown to be essential for this stimulation (Schnizler et al., 2000). However, the mechanism leading to ENaC stimulation did not exclude the possibility of an intracellular pathway involving protein kinase A (PKA). Further experiments demonstrated that PKA inhibitor PKI 6-22 did not prevent cpt-cAMP stimulation of gpENaC expressed in *Xenopus* oocytes. Furthermore, the α subunit containing the gp extracellular loop with rat intracellular C and/or N termini expressed in *Xenopus* oocytes together with rat βγ ENaC were

sensitive to cpt-cAMP (Chraibi et al., 2001). This chimeric channel demonstrated that the extracellular domain of the gp α subunit was the determinant for ENaC stimulation by cpt-cAMP. Thus, the molecule can be considered to be a ligand for the channel. Moreover, the outside-out configuration of the patch clamp showed an increase of the open probability and the number of open channels (N.Po) exposed to cpt-cAMP, confirming a direct interaction with the extracellular domain of the gpα, ratβγ chimera expressed in *Xenopus* oocytes. To determine which part of the extracellular domain of αENaC is involved in this regulation, we made four chimeric constructions of that subunit (Figure 2).
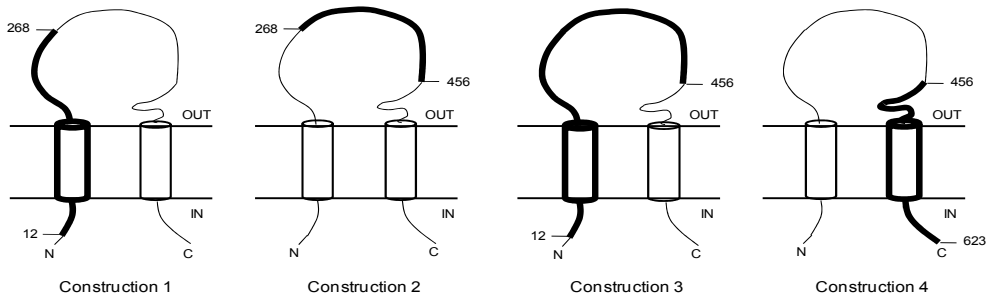
Fig. 2. Schematic representation of the ENaC subunits. Chimeric constructions of α subunits by fusion of the coding region for the αgp part (bold line) with the αrat part (thin line). Numbers indicate residues at corresponding positions on the αgp sequence.

To do so, two restriction sites were generated in guinea pig and rat αENaC cDNAs at homologous positions using a PCR technique. Then the appropriate fragment of gp cDNA was inserted into the rat cDNA between the restriction sites. Amiloride-sensitive current was measured in the presence and absence of 10 µM cpt-cAMP. We generated eleven swapping mutants of rat and gp αENaC using PCR-directed mutagenesis and expressed each of these mutants with the rat β and γ subunits in *Xenopus* oocytes. Among the eleven substitutions, Ile481 in the gp αENaC extracellular domain plays a major role in cpt-cAMP-induced ENaC activation. The *I481N* mutation in the gene for the αgp subunit completely abolished stimulation of ENaC. The *N510I* mutation in the gene for the αrat subunit caused intermediate sensitivity to cpt-cAMP. All other mutations or combination of mutations, including *N510I* in the αrat gene, did not increase the cpt-cAMP effect (Renauld et al., 2008).

Similarly to what we described with cpt-cAMP, Hong-Guan and coworkers (Nie et al., 2009) suggested that cpt-cGMP stimulates human, rat and mouse ENaC through direct interaction and not through the intracellular pathway. Indeed directed mutagenesis of the coding regions for potential phosphorylation sites for the cGMP-dependent kinases on ENaC did not affect cpt-cGMP-induced activation in *Xenopus* oocytes. Furthermore, knockdown of PKG isoforms did not prevent cpt-cGMP-dependent activation. Han and colleagues (2011) confirmed that cpt-cGMP-induced ENaC activation was mediated through direct interaction and an increases of N.Po. By directed mutagenesis, these authors were able to show that the mutations abolishing self-inhibition (βΔ*V348* and γ*H233R*) lost their responses to cpt-cGMP. The mutations augmenting this phenomenon (α*Y458A* and γ*M432G*) facilitated the stimulatory effects of this compound. Thus, these data suggest that the elimination of self-inhibition may be a novel mechanism for cpt-cGMP to stimulate ENaC.

α subunit	Icpt/Ictl	SEM	unpaired t-Test VS α rat wt
α gp wt	2.28	0.05	P<0.001
α rat wt	1.13	0.01	
construction 1	1.14	0.01	NS
construction 2	1.28	0.03	P<0.001
construction 3	1.34	0.03	P<0.001
construction 4	1.85	0.06	P<0.001
αr L493S	1.06	0.02	NS
αr S500N	1.08	0.03	NS
αr S507N	1.06	0.03	NS
αr I509T	1.05	0.02	NS
αr N510I	1.45	0.03	NS
αr K524T	1.02	0.03	NS
αr E531Q	1.12	0.02	NS
αr N542S	1.05	0.04	NS
αr K550N	1.12	0.02	NS
αr F554Y	1.16	0.02	NS
αr K561R	1.12	0.02	NS
αgp I481N	1.58	0.07	P<0.001

Table 1. Effect of cpt-cAMP on different constructions and mutants of the αENaC subunit expressed in *Xenopus* oocytes together with the β and γrat subunits. Results are presented as a ratio of amiloride-sensitive current measured after and before cpt-cAMP perfusion (Icpt/Ictl). gp, guinea pig; r, rat; wt, wild type; NS, not significant relative to αrat wt

4.2.3 Effects of glibenclamide

The same experimental approach was used to study the stimulation of ENaC by glibenclamide (Renauld and Chraibi, 2009). Glibenclamide, a high affinity-blocker of the K_{ATP} channel, has been shown to stimulate *Xenopus* ENaC (but not rat ENaC) expressed in *Xenopus* oocytes. The α subunit has been shown to be critical for this activation (Chraibi and Horisberger, 1999). As described with cpt-cAMP, patch clamp recordings in the outside-out configuration showed an increase of N.Po when *Xenopus* ENaC was exposed to glibenclamide. Another study has demonstrated that the αgp subunit, but not the αrat subunit, conferred sensitivity of ENaC to glibenclamide (Schnizler et al., 2003). Using mutagenesis, these authors were able to produce other chimeric rat/gp α subunits; and they suggested that the extracellular loop or the transmembrane domain of the αgp subunit is involved in the activation of the ENaC channel by glibenclamide. Thus, similarly to cpt-cAMP activation, channels composed of the αgp subunit and the β and γ subunits from rat are sensitive to glibenclamide, while channels composed of the α, β, and γ subunits from rat are resistant. We used the chimeras of the α subunit previously generated and found that construction 4 was also important for glibenclamide stimulation of the channel. Unlike cpt-cAMP, glibenclamide had no effect on the other constructions expressed with the β and γ subunits from rat. Moreover, directed mutagenesis did not reveal particular residues involved in this regulation.

α subunit	Iglib/Ictrl	SEM	unpaired t-Test VS α rat wt
α gp wt	1.63	0.02	P<0.001
α rat wt	0.96	0.01	
construction 1	1.03	0.04	NS
construction 2	1.04	0.05	NS
construction 3	1.03	0.01	NS
construction 4	1.27	0.04	P<0.001
αr L493S	0.92	0.03	NS
αr S500N	1.01	0.03	NS
αr S507N	0.98	0.02	NS
αr I509T	0.99	0.02	NS
αr N510I	1.00	0.01	NS
αr K524T	0.97	0.02	NS
αr E531Q	0.99	0.05	NS
αr N542S	0.98	0.02	NS
αr K550N	0.91	0.01	NS
αr F554Y	0.99	0.02	NS
αr K561R	0.96	0.02	NS

Table 2. Effect of glibenclamide on different constructions and mutants of αENaC expressed in *Xenopus* oocytes together with the β and γ subunits from rat. Results are presented as a ratio of amiloride-sensitive current measured after and before glibenclamide perfusion (Iglib/Ictl). gp, guinea pig; r, rat; wt, wild type; NS, not significant relative to αrat wt

4.2.4 Effects of other molecules

Capsazepine has been described as the first selective active activator for the δENaC subunit (Yamamura et al., 2004). Indeed capsazepine specifically stimulates human-made δ subunit, but not the α subunit expressed in *Xenopus* oocytes. Moreover, this molecule can stimulate the δENaC monomer, whereas no other vanilloid compound can produce changes in sodium amiloride-sensitive current. However, the authors did not determine any amino acids involved in this activation. Directed mutagenesis could be a powerful tool to understand differences between the α and δ subunits and resolve the structure-function relationships of both proteins.

S3969 is a small molecule described as a reversible activator of human, but not mouse, αβγ ENaC through direct interaction with the extracellular side of the channel by increasing N.Po (Lu et al., 2008). Interestingly, S3969 stimulates amiloride-sensitive current in oocytes expressing the δ subunit instead of α. The authors showed that βENaC was critical for this activation. Mouse-human chimeras of the β subunit confirmed the implication of the extracellular domain. More specifically, deletion of Val348 in βENaC completely abolished S3969 activation of ENaC. Maturation and optimal transport of ENaC to the plasma membrane requires furin cleavage of the β and γ subunits at a specific Arg. Mutations of the furin cleavage site in which Arg was replaced by Ala did not prevent ENaC activation by S3969, suggesting that proteolytic activation prior to S3969 stimulation is not necessary. Mutations producing pseudohypoaldosteronism type 1 (PHA1), resulting in salt-wasting, a genetic disease, have been generated in the α (*R508~STOP~*) and β (*G37S*) subunits. These

mutants decreased amiloride-sensitive current, but the S3969 compound was still able to stimulate ENaC activity.

5. Conclusion

The epithelial sodium channel (ENaC) has been used for decades as a therapeutic target against type 1 hypertension and Liddle syndrome. More recently, several studies pointed to ENaC as a potential target for cystic fibrosis (Zhou et al., 2011), a pathology characterized by an impaired Cl⁻ secretion through the cystic fibrosis transmembrane conductance regulator (CFTR) and an increase of Na⁺ reabsorption through ENaC. The studies of mutations involved in these diseases have been extremely helpful in determining the molecular mechanisms by which they lead to a dysfunction of ENaC. Furthermore, the experiments carried out on this topic have shown the contribution of the PCR-directed mutagenesis technique in the determination of the structure-function relationships of ENaC. These studies have led to a better understanding of the domains involved in ion selectivity, gating and expression of the channel at the cell membrane. Additional studies are needed to define other key domains of ENaC. They may provide a new strategy for the treatment of pathologies linked to dysfunction of this channel.

6. References

Abriel, H., Horisberger, J.D. 1999. Feedback inhibition of rat amiloride-sensitive epithelial sodium channels expressed in Xenopus laevis oocytes. *J Physiol* 516:31-43.

Anantharam A, P.L. 2007. Determination of epithelial na+ channel subunit stoichiometry from single-channel conductances. *J Gen Physiol.* 130:55-70

Awayda, M.S., Ismailov, II, Berdiev, B.K., Fuller, C.M., Benos, D.J. 1996. Protein kinase regulation of a cloned epithelial Na+ channel. *J Gen Physiol* 108:49-65

Brouard, M., Casado, M., Djelidi, S., Barrandon, Y., Farman, N. 1999. Epithelial sodium channel in human epidermal keratinocytes: expression of its subunits and relation to sodium transport and differentiation. *J Cell Sci* 112 (Pt 19):3343-52

Bruns, J.B., Carattino, M.D., Sheng, S., Maarouf, A.B., Weisz, O.A., Pilewski, J.M., Hughey, R.P., Kleyman, T.R. 2007. Epithelial Na+ channels are fully activated by furin- and prostasin-dependent release of an inhibitory peptide from the gamma-subunit. *J Biol Chem* 282:6153-60

Caldwell, R.A., Boucher, R.C., Stutts, M.J. 2005. Neutrophil elastase activates near-silent epithelial Na+ channels and increases airway epithelial Na+ transport. *Am J Physiol Lung Cell Mol Physiol* 288:L813-9

Canessa, C.M., Horisberger, J.D., Rossier, B.C. 1993. Epithelial sodium channel related to proteins involved in neurodegeneration. *Nature* 361:467-70.

Canessa, C.M., Schild, L., Buell, G., Thorens, B., Gautschi, I., Horisberger, J.D., Rossier, B.C. 1994. Amiloride-sensitive epithelial Na+ channel is made of three homologous subunits. *Nature* 367:463-7.

Chang, S.S., Grunder, S., Hanukoglu, A., Rosler, A., Mathew, P.M., Hanukoglu, I., Schild, L., Lu, Y., Shimkets, R.A., Nelson-Williams, C., Rossier, B.C., Lifton, R.P. 1996. Mutations in subunits of the epithelial sodium channel cause salt wasting with hyperkalaemic acidosis, pseudohypoaldosteronism type 1. *Nat Genet* 12:248-53.

Chraibi, A., Horisberger, J.D. 1999. Stimulation of epithelial sodium channel activity by the sulfonylurea glibenclamide. *J Pharmacol Exp Ther* 290:341-7.

Chraibi, A., Horisberger, J.D. 2002. Na self inhibition of human epithelial Na channel: temperature dependence and effect of extracellular proteases. *J Gen Physiol* 120:133-45.

Chraibi, A., Horisberger, J.D. 2003. Dual effect of temperature on the human epithelial Na+ channel. *Pflugers Arch* 447:316-320

Chraibi, A., Schnizler, M., Clauss, W., Horisberger, J.D. 2001. Effects of 8-cpt-cAMP on the epithelial sodium channel expressed in Xenopus oocytes. *J Membr Biol* 183:15-23.

Debonneville, C., Flores, S.Y., Kamynina, E., Plant, P.J., Tauxe, C., Thomas, M.A., Munster, C., Chraibi, A., Pratt, J.H., Horisberger, J.D., Pearce, D., Loffing, J., Staub, O. 2001. Phosphorylation of Nedd4-2 by Sgk1 regulates epithelial Na(+) channel cell surface expression. *Embo J* 20:7052-9.

Diakov, A., Bera, K., Mokrushina, M., Krueger, B., Korbmacher, C. 2008. Cleavage in the {gamma}-subunit of the epithelial sodium channel (ENaC) plays an important role in the proteolytic activation of near-silent channels. *J Physiol* 586:4587-608

Dijkink, L., Hartog, A., van Os, C.H., Bindels, R.J. 2002. The epithelial sodium channel (ENaC) is intracellularly located as a tetramer. *Pflugers Arch* 444:549-55

Duc, C., Farman, N., Canessa, C.M., Bonvalet, J.P., Rossier, B.C. 1994. Cell-specific expression of epithelial sodium channel alpha, beta, and gamma subunits in aldosterone-responsive epithelia from the rat: localization by in situ hybridization and immunocytochemistry. *J Cell Biol* 127:1907-21

Eskandari, S., Snyder, P.M., Kreman, M., Zampighi, G.A., Welsh, M.J., Wright, E.M. 1999. Number of subunits comprising the epithelial sodium channel. *J Biol Chem* 274:27281-6

Firsov, D., Gautschi, I., Merillat, A.M., Rossier, B.C., Schild, L. 1998. The heterotetrameric architecture of the epithelial sodium channel (ENaC). *Embo J* 17:344-52.

Firsov, D., Schild, L., Gautschi, I., Merillat, A.M., Schneeberger, E., Rossier, B.C. 1996. Cell surface expression of the epithelial Na channel and a mutant causing Liddle syndrome: a quantitative approach. *Proc Natl Acad Sci U S A* 93:15370-5.

Garcia-Caballero, A., Dang, Y., He, H., Stutts, M.J. 2008. ENaC proteolytic regulation by channel-activating protease 2. *J Gen Physiol* 132:521-35

Grunder, S., Firsov, D., Chang, S.S., Jaeger, N.F., Gautschi, I., Schild, L., Lifton, R.P., Rossier, B.C. 1997. A mutation causing pseudohypoaldosteronism type 1 identifies a conserved glycine that is involved in the gating of the epithelial sodium channel. *Embo J* 16:899-907.

Han, D.Y., Nie, H.G., Su, X.F., Shi, X.M., Bhattarai, D., Zhao, M., Zhao, R.Z., Landers, K., Tang, H., Zhang, L., Ji, H.L. 2011. CPT-cGMP Stimulates Human Alveolar Fluid Clearance by Releasing External Na+ Self-Inhibition of ENaC. *Am J Respir Cell Mol Biol*

Hansson, J.H., Nelson-Williams, C., Suzuki, H., Schild, L., Shimkets, R., Lu, Y., Canessa, C., Iwasaki, T., Rossier, B., Lifton, R.P. 1995. Hypertension caused by a truncated epithelial sodium channel gamma subunit: genetic heterogeneity of Liddle syndrome. *Nat Genet* 11:76-82.

Hansson, J.H., Schild, L., Lu, Y., Wilson, T.A., Gautschi, I., Shimkets, R., Nelson-Williams, C., Rossier, B.C., Lifton, R.P. 1995. A de novo missense mutation of the beta subunit

of the epithelial sodium channel causes hypertension and Liddle syndrome, identifying a proline-rich segment critical for regulation of channel activity. *Proc Natl Acad Sci U S A* 92:11495-9.

Horisberger, J.D., Chraibi, A. 2004. Epithelial Sodium Channel: A Ligand-Gated Channel? *Nephron Physiology* 96:37-41

Hughey, R., Bruns, J., Kinlough, C., Harkleroad, K., Tong, Q., Carattino, M., Johnson, J., Stockand, J., Kleyman, T. 2004. Epithelial sodium channels are activated by furin-dependent proteolysis. *J. Biol. Chem.* 30:18111-4

Hummler, E., Barker, P., Gatzy, J., Beermann, F., Verdumo, C., Schmidt, A., Boucher, R., Rossier, B.C. 1996. Early death due to defective neonatal lung liquid clearance in alpha- ENaC-deficient mice. *Nat Genet* 12:325-8.

Jasti, J., Furukawa, H., Gonzales, E.B., Gouaux, E. 2007. Structure of acid-sensing ion channel 1 at 1.9A° resolution and low pH. *nature* 449:316-324

Kamynina, E., Staub, O. 2002. Concerted action of ENaC, Nedd4-2, and Sgk1 in transepithelial Na(+) transport. *Am J Physiol Renal Physiol* 283:F377-87

Kellenberger, S., Gautschi, I., Rossier, B.C., Schild, L. 1998. Mutations causing Liddle syndrome reduce sodium-dependent downregulation of the epithelial sodium channel in the Xenopus oocyte expression system. *J Clin Invest* 101:2741-50.

Kellenberger, S., Schild. L. 2002. Epithelial sodium channel/degenerin family of ion channels: a variety of functions for a shared structure. *Physiol Rev.* 82:735-67

Kunzelmann, K., Mall, M. 2002. Electrolyte transport in the mammalian colon: mechanisms and implications for disease. *Physiol Rev* 82:245-89

Liebold, K.M., Reifarth, F.W., Clauss, W., Weber, W. 1996. cAMP-activation of amiloride-sensitive Na+ channels from guinea-pig colon expressed in Xenopus oocytes. *Pflugers Arch* 431:913-22

Lingueglia, E., Voilley, N., Waldmann, R., Lazdunski, M., Barbry, P. 1993. Expression cloning of an epithelial amiloride-sensitive Na+ channel. A new channel type with homologies to Caenorhabditis elegans degenerins. *FEBS Lett* 318:95-9

Lu M, Echeverri F, Kalabat D, Laita B, Dahan DS, Smith RD, Xu H, Staszewski L, Yamamoto J, Ling J, Hwang N, Kimmich R, Li P, Patron E, Keung W, Patron A, Moyer BD. 2008. Small molecule activator of the human epithelial sodium channel. *J Biol Chem* 283: 11981–11994

Masilamani, S., Kim, G.H., Mitchell, C., Wade, J.B., Knepper, M.A. 1999. Aldosterone-mediated regulation of ENaC alpha, beta, and gamma subunit proteins in rat kidney. *J Clin Invest* 104:R19-23.

Mullis, K.B., Faloona, F.A. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol* 155:335-50

Myerburg, M.M., Butterworth, M.B., McKenna, E.E., Peters, K.W., Frizzell, R.A., Kleyman, T.R., Pilewski, J.M. 2006. Airway surface liquid volume regulates ENaC by altering the serine protease-protease inhibitor balance: a mechanism for sodium hyperabsorption in cystic fibrosis. *J Biol Chem* 281:27942-9

Nie, H.G., Chen, L., Han, D.Y., Li, J., Song, W.F., Wei, S.P., Fang, X.H., Gu, X., Matalon, S., Ji, H.L. 2009. Regulation of epithelial sodium channels by cGMP/PKGII. *J Physiol* 587:2663-76

Passero, C.J., Mueller, G.M., Rondon-Berrios, H., Tofovic, S.P., Hughey, R.P., Kleyman, T.R. 2008. Plasmin activates epithelial Na+ channels by cleaving the gamma subunit. *J Biol Chem* 283:36586-91

Perucca, J., Bichet, D.G., Bardoux, P., Bouby, N., Bankir, L. 2008. Sodium excretion in response to vasopressin and selective vasopressin receptor antagonists. *J Am Soc Nephrol* 19:1721-31

Randell, S.H., Boucher, R.C. 2006. Effective mucus clearance is essential for respiratory health. *Am J Respir Cell Mol Biol* 35:20-8

Renauld S, Allache R, Chraibi A. 2008. Ile481 from the guinea pig alpha-subunit plays a major role in the activation of ENaC by cpt-cAMP. *Cell Physiol Biochem*. 22(1-4):101-8

Renauld, S., Chraibi, A. 2009. Role of the C-Terminal Part of the Extracellular Domain of the alpha-ENaC in Activation by Sulfonylurea Glibenclamide. *J Membr Biol*

Renauld, S., Tremblay, K., Ait-Benichou, S., Simoneau-Roy, M., Garneau, H., Staub, O., Chraibi, A. 2010. Stimulation of ENaC activity by rosiglitazone is PPARgamma-dependent and correlates with SGK1 expression increase. *J Membr Biol* 236:259-70

Rossier, B. 2004. The epithelial sodium channel: activation by membrane-bound serine proteases. *Proc Am Thorac Soc:* 1:4-9

Roudier-Pujol, C., Rochat, A., Escoubet, B., Eugene, E., Barrandon, Y., Bonvalet, J.P., Farman, N. 1996. Differential expression of epithelial sodium channel subunit mRNAs in rat skin. *J Cell Sci* 109 (Pt 2):379-85

Schild, L., Canessa, C.M., Shimkets, R.A., Gautschi, I., Lifton, R.P., Rossier, B.C. 1995. A mutation in the epithelial sodium channel causing Liddle disease increases channel activity in the Xenopus laevis oocyte expression system. *Proc Natl Acad Sci U S A* 92:5699-703

Schild, L., Lu, Y., Gautschi, I., Schneeberger, E., Lifton, R.P., Rossier, B.C. 1996. Identification of a PY motif in the epithelial Na channel subunits as a target sequence for mutations causing channel activation found in Liddle syndrome. *Embo J* 15:2381-7

Schnizler, M., Berk, A., Clauss, W. 2003. Sensitivity of oocyte-expressed epithelial Na+ channel to glibenclamide. *Biochim Biophys Acta* 1609:170-6

Schnizler, M., Mastroberardino, L., Reifarth, F., Weber, W.M., Verrey, F., Clauss, W. 2000. cAMP sensitivity conferred to the epithelial Na+ channel by alpha- subunit cloned from guinea-pig colon. *Pflugers Arch* 439:579-87.

Sheng, S., Bruns, J., Kleyman, T. 2004. Extracellular histidine residues crucial for Na+ self-inhibition of epithelial Na+ channels. *JBC* 279:9743-9

Sheng, S., Perry, C., Kleyman, T. 2002. External nickel inhibits epithelial sodium channel by binding to histidine residues within the extracellular domains of alpha and gamma subunits and reducing channel open probability. *J Biol. Chem.* 277:50098-111

Shimkets, R.A., Lifton, R.P., Canessa, C.M. 1997. The activity of the epithelial sodium channel is regulated by clathrin- mediated endocytosis. *J Biol Chem* 272:25537-41.

Shimkets, R.A., Warnock, D.G., Bositis, C.M., Nelson-Williams, C., Hansson, J.H., Schambelan, M., Gill, J.R., Jr., Ulick, S., Milora, R.V., Findling, J.W., et al. 1994. Liddle's syndrome: heritable human hypertension caused by mutations in the beta subunit of the epithelial sodium channel. *Cell* 79:407-14

Snyder, P.M., Cheng, C., Prince, L.S., Rogers, J.C., Welsh, M.J. 1998. Electrophysiological and biochemical evidence that DEG/ENaC cation channels are composed of nine subunits. *J Biol Chem* 273:681-4

Staub, O., Gautschi, I., Ishikawa, T., Breitschopf, K., Ciechanover, A., Schild, L., Rotin, D. 1997. Regulation of stability and function of the epithelial Na+ channel (ENaC) by ubiquitination. *EMBO J* 16:6325-6336

Stockand, J.D., Staruschenko, A., Pochynyuk, O., Booth, R.E., Silverthorn, D.U. 2008. Insight Toward Epithelial Na1 Channel Mechanism Revealed by the Acid-sensing Ion Channel 1 Structure. *IUBMB Life* 60: 620-628

Tamura, H., Schild, L., Enomoto, N., Matsui, N., Marumo, F., Rossier, B.C. 1996. Liddle disease caused by a missense mutation of beta subunit of the epithelial sodium channel gene. *J Clin Invest* 97:1780-4

Vallet, V., Chraibi, A., Gaeggeler, H.P., Horisberger, J.D., Rossier, B.C. 1997. An epithelial serine protease activates the amiloride-sensitive sodium channel. *Nature* 389:607-10.

Vallet, V., Pfister, C., Loffing, J., Rossier, B.C. 2002. Cell-surface expression of the channel activating protease xCAP-1 is required for activation of ENaC in the Xenopus oocyte. *J Am Soc Nephrol* 13:588-94

Vuagniaux, G., Vallet, V., Jaeger, N.F., Hummler, E., Rossier, B.C. 2002. Synergistic Activation of ENaC by Three Membrane-bound Channel- activating Serine Proteases (mCAP1, mCAP2, and mCAP3) and Serum- and Glucocorticoid-regulated Kinase (Sgk1) in Xenopus Oocytes. *J Gen Physiol* 120:191-201.

Vuagniaux, G., Vallet, V., Jaeger, N.F., Pfister, C., Bens, M., Farman, N., Courtois-Coutry, N., Vandewalle, A., Rossier, B.C., Hummler, E. 2000. Activation of the amiloride-sensitive epithelial sodium channel by the serine protease mCAP1 expressed in a mouse cortical collecting duct cell line. *J Am Soc Nephrol* 11:828-34.

Yamamura, H., Ugawa, S., Ueda, T., Nagao, M., Shimada, S. 2004. Protons activate the delta -subunit of epithelial Na+ channel in humans. *JBC* in preess (published on january 15, manuscript M400274200)

Zhou, Z., Duerr, J., Johannesson, B., Schubert, S.C., Treis, D., Harm, M., Graeber, S.Y., Dalpke, A., Schultz, C., Mall, M.A. 2011. The ENaC-overexpressing mouse as a model of cystic fibrosis lung disease. *J Cyst Fibros* 10 Suppl 2:S172-82

Use of Site-Directed Mutagenesis in the Diagnosis, Prognosis and Treatment of Galactosemia

M. Tang[1], K.J. Wierenga[2] and K. Lai[1]
[1]University of Utah School of Medicine,
[2]University of Oklahoma Health Sciences Center,
USA

1. Introduction

Site-directed mutagenesis (**SDM**) is undoubtedly one of the most powerful techniques in molecular biology. In this chapter, we will describe the use of SDM in the study of the human inherited metabolic disorder, Galactosemia (Type I, II, and III) and the development of novel therapies for the disease. This powerful technique not only helped confirm suspected *GAL* gene mutations in Galactosemia, but also played a significant role in unraveling the catalytic mechanisms of the GAL enzymes in the conserved Leloir pathway of galactose metabolism. To date, more than thirty disease-causing mutations in the human *GAL* genes have been characterized in great detail; and these findings have paved the way for innovative, state-of-the-art therapies, such as chaperone therapy. Recently, in order to optimize small molecule GALK inhibitors for the treatment of Type I Galactosemia, we have employed SDM to identify amino acids of the GALK enzyme that interact with its selective inhibitors. These studies exemplified the expanding roles of SDM in innovative drug design and in kinase inhibitor selectivity.

2. Background

2.1 What is galactosemia?

Galactose is a hexose that differs from glucose only by the configuration of the hydroxyl group at the carbon-4 position. Often present as an anomeric mixture of α-D-galactose and β-D-galactose, this monosaccharide exists abundantly in milk, dairy products and many other food types such as fruits and vegetables (Acosta and Gross, 1995; Berry et al., 1993). However, galactose can also be produced endogenously in human cells, mainly as products of glycoprotein and glycolipid turnover.(Berry et al., 1995, 2004). Once freely present inside the cells, β-D-galactose is epimerized to α-D-galactose through the action of a mutarotase (Beebe and Frey, 1998; Thoden and Holden, 2002a). α-D-galactose is then metabolized by the Leloir pathway (Leloir 1951), an evolutionarily conserved biochemical pathway which begins with the phosphorylation of galactose by the enzyme galactokinase (GALK) to form galactose-1 phosphate (gal-1P) (Cardini and Leloir, 1953). Gal-1P is subsequently, together with the substrate UDP-glucose, converted by galactose-1-phosphate uridylyltransferase

(GALT) to form UDP-galactose and glucose-1 phosphate (glu-1P) (Kalckar et al., 1953). The Leloir pathway is completed by reversibly forming UDP-glucose from UDP-galactose *via* UDP-galactose-4-epimerase (GALE) (Leloir 1953; Darrow and Rodstorm, 1968). Inherited deficiencies of GALK, GALT, and GALE activities in humans have all been observed, studied, and reviewed extensively (Bosch et al., 2002; Elsas 1993; Fridovich-Keil et al., 1993a). The clinical manifestations of each enzyme deficiency, however, differ markedly (Berry et al., 1995; Berry and Elsas, 2011; Fridovich-Keil et al., 1993a; Lai et al., 2009;). For instance, patients with GALK deficiency (MIM 230200) (Type II Galactosemia) have the mildest clinical consequences, as they may present only with cataracts (Bosch et al., 2002). On the other hand, GALT-deficiency (MIM 230400) (Type I or Classic Galactosemia) is potentially lethal in infancy, if undiagnosed and untreated, and is also associated with long-term, organ-specific complications (Berry et al., 1995). GALE-deficiency (MIM 230350) (Type III Galactosemia) has been somewhat controversial with regards to clinical manifestations, as this disorder is rare; and information is mostly derived from case reports (Fridovich-Keil et al., 1993a). Until newborn screening for GALE deficiency is available, the natural history will likely remain unknown. The differences in clinical outcome between GALT and GALK deficiencies reflect the differences in tissue response to the characteristic changes in the levels of galactose metabolites as a result of the respective enzyme deficiencies.

2.2 How are the different types of galactosemia detected and diagnosed?

Newborn screening programs worldwide have greatly facilitated the early detection of Galactosemia (Kaye et al., 2006; Levy 2010). The screening tests often involve the detection of elevated level of blood galactose and/or specific GAL enzyme in the dried blood spots on filter paper. Elevated galactose will detect GALK deficiency and GALT deficiency, but it may not detect GALE deficiency. Other states screen for GALT activity, and may therefore diagnose Type I Galactosemia. However, this screen will miss GALK and GALE deficiency. The final diagnosis is secured once the specific enzyme deficiency is confirmed by enzymatic assays or by DNA genotyping; these tests are available commercially in the USA (http://www.ncbi.nlm.nih.gov/sites/ GeneTests/, Tests #3437, #2229 and #53782).

2.3 What are the current treatments for galactosemia, and what is the outlook for patients?

The main aspect of management for all forms of Galactosemia is withdrawal of lactose/galactose from the diet as soon as the diagnosis is made, or even considered (Segal 1995). In infants, this means the replacement of breast/cow milk with soy-based formula. However, it has become clear that, despite early detection and (early) dietary intervention, there still is a significant burden of the disease, particularly for Classic Galactosemia where chronic problems persist through adulthood. The most common medical complications of Type I Galactosemia are speech dyspraxia, ataxia, premature ovarian insufficiency, and intellectual deficits, which are rarely seen in other forms of galactosemia (Waggoner et al., 1990; Waisbren et al., 2011). GALK deficiency (Type II Galactosemia) is managed also with lactose/galactose restriction, though the complications are mainly confined to the eye (cataracts) (Bosch et al., 2002). GALE deficiency is treated similarly, though complications of this deficiency may not be preventable with such restriction, as is GALT deficiency (Fridovich-Keil et al., 1993a).

3. Use of SDM to confirm disease-causing mutations in human GALT, GALK, and GALE genes identified clinically

3.1 The issues

Advances in federal and state newborn screening programs worldwide have resulted in the inclusion of the potentially lethal disorder, Galactosemia, in the list of diseases for which newborns are screened. Very often, once an affected newborn is identified by the biochemical assays, it is helpful to know the genotype of *GAL* gene involved because there appears to be a genotype-phenotype correlation for a few selected *GAL* gene mutations. The confirmation of the *GAL* genotypes in the affected patients will provide better prognosis. Additionally, a few well-characterized GAL enzyme variants have been shown to retain significant residual enzyme activities. Consequently, patients with selected mutations might benefit from novel therapies, such as chaperone therapies.

Unfortunately many patients with Galactosemia identified to-date have novel (private) nucleotide changes in their *GAL* genes. For instance, the *GALT* gene database set up by the ARUP Laboratories (Salt Lake City, USA) has recorded over 200 nucleotide changes of the *GALT* gene identified in patients with Type I Galactosemia (*www.arup.utah.edu/database/GALT/GALT_welcome.php*). Without clinical correlation, it is impossible to tell if any of these novel changes actually results in impaired GALT enzyme activities seen in the patients. Moreover, many patients are compound heterozygous for the *GAL* gene mutations. In other words, a single patient may have a unique nucleotide change in each of the two *GAL* alleles; and it is difficult to conclude which one is responsible for the reduction in total enzyme activity. Thus there is a real need to perform *in vitro* expression studies of the identified "variant" *GAL* genes.

3.2 Research design

Our laboratory and others have largely used similar strategies in confirming the suspected human *GAL* gene mutations in patients with Galactosemia (Fridovich-Keil et al., 1995a; Lai et al., 1999; Reichardt et al., 1992;). In almost all cases, we sub-cloned the cDNA of the respective *GAL* gene into expression plasmid vectors, before we performed SDM of the sub-cloned fragments to obtain "mutant" cDNAs with the same sequence changes observed in patients. We then expressed the wild-type and mutated cDNAs in heterologous expression systems, such as *Escherichia coli, Saccharomyces cerevisiae* or even mammalian expression systems. Subsequently, we tested for differences in kinetic parameters of the GAL enzymes, such as K_M and V_{max}, and the expression efficiencies, such as protein and/or mRNA abundances, between mutant and control cDNAs.

3.3 The results

The primary goal for expression analysis of the suspected disease-causing mutations in the *GAL* genes is to show that the nucleotide changes observed are causing impaired GAL enzyme activities and could therefore be the causes for the diseases. In addition, in the course of the analysis, kinetic parameters of the variant enzymes are often determined, which are expected to help advance the structural knowledge of the GAL enzymes.

3.3.1 Type I (GALT-deficiency) galactosemia

As mentioned above, more than 200 nucleotide changes in the *GALT* gene have been identified so far, mostly single nucleotide substitutions. The most common human *GALT* mutation, *Q188R*, is detected in over 70% of galactosemic patients in Europe and North America. The *Q188R* mutation is associated with a poor clinical outcome, even with a galactose-restricted diet (Guerrero et al., 2000; Murphy et al., 1999; Webb et al., 2003). *K285N* is the second most common mutation found in patients in Europe, especially in the countries of central and Eastern Europe, where it can account for up to 34% of *GALT* alleles (Greber-Platzer et al., 1997; Kozak et al., 2000). In the African-American population, the *S135L* mutation is predominant. The corresponding enzyme leads to a relatively benign outcome, if the mutation is identified and the patient is treated with a galactose-restricted diet in the newborn period (Lai et al., 1996, 2001; Landt et al., 1997). A more common mutation, *N314D*, occurs in all populations mentioned above and can lead to two different phenotypes, depending on the presence or absence of a 4-bp deletion in the coding region for the carbohydrate response element. When *N314D* is associated with a four-nucleotide deletion in the promoter region (the Duarte type 2), homozygosity for N314D and this altered promoter region causes a 50% decrease of GALT activity, with a mild or even undetected phenotype (Elsas et al., 1994). In the absence of this deletion in the promoter, homozygosity for the N314D missense mutation (the Los Angeles variant) results in normal GALT in erythrocytes (Shin et al., 1998). A 5-kb deletion is found so far exclusively in Ashkenazi Jewish patients (Coffee et al., 2006).

Due to its frequency among GALT-deficiency galactosemic patients and its association with a poor clinical outcome, the *Q188R* mutation has been extensively studied. The initial study using the COS cell expression system surprisingly showed that this mutation had about 10% of normal enzymatic activity (Reichardt et al., 1991). This result was not consistent with the clinical finding that patients homozygous for *Q188R* have no detectable enzyme activity in their red blood cells. Another study, carried out in a yeast model that was completely devoid of GALT activity, used a PCR-mediated SDM technique and clarified that the *Q188R* mutation did cause loss of function of both human and yeast GALT (Fridovich-Keil et al.,1993b). Interestingly, this study also showed that the mutant yeast, with its loss of GALT activity, could not survive in galactose media if the *Q188R* missense mutation was introduced, while reconstitution of wild-type GALT resulted in normal growth (Fridovich-Keil and Jinks-Roberson, 1993b). The confounding result of the first study is likely to be explained by the presence of endogenous GALT activity in the COS cells, highlighting the importance of studying mutations in a null background system, such as the *gal7*-deleted yeast model used in the second study. Alternatively, one should use purified mutant proteins in the analysis of the enzymatic activities. Subsequent studies further confirmed that the *Q188R* mutation not only totally abolishes GALT enzyme activity, but also acts as a partial dominant-negative mutation, as the heterodimer of *Q188R*/wild type has only 15% of wild-type activity (Fridovich-Keil et al., 1995a; Elsevier and Fridovich-Keil,1996). Kinetic analysis showed this mutation mainly causes impaired specific activity of the heterodimer without altering the K_M for both substrates. In order to further understand how mutation at this site could affect the enzyme, Lai and coworkers mutated glutamine-188 (Gln[188]) to arginine and asparagine, respectively, through SDM (Lai et al., 1999). More detailed kinetic measurement showed that mutating glutamine to arginine or asparagine did not affect the first step of the double-displacement action (UDP-Glu to glu-1p). In fact, Q188R-GALT even

had a better V_{max} as compared with the wild-type GALT. However, the *Q188R* mutation severely impaired the second step of the reaction. The crystal structures of *E. coli* GALT revealed that Gln[168] (equivalent to Gln[188] in human GALT) could stabilize the GALT-UMP intermediate through two hydrogen bonds formed between the amide side chain of Gln[188] and the phosphoryl oxygen of the UMP moiety (Wedekind et al., 1996). Through molecular modeling studies (or "virtual SDM"), Lai and coworkers changed glutamine to arginine and asparagine, respectively, and found that the number of hydrogen bonds formed between new amino acid residues and UMP moiety decreased to one, which could have destabilized the GALT-UMP intermediate required for the second displacement reaction (Lai et al., 1999). This destabilization was well manifested in the increased V_{max} in the Q188R mutant in the first displacement reaction, as the destabilization speeded up the recycling of the enzyme for the first reaction (Lai et al., 1999). To complete the double-displacement reaction, a stable GALT-UMP intermediate was required to bind gal-1P, which was better accomplished by the two hydrogen bonds from glutamine than by the single hydrogen bond from arginine or asparagine.

The *S135L* mutation was identified initially as a polymorphism with near normal enzymatic activity in the COS cell expression system (Reichardt et al., 1992). However, subsequent SDM studies in the yeast-expression system, defined this as a missense mutation that significantly impaired enzyme activity; but, unlike the *Q188R* mutation, it still had minor residual activity (Fridovich-Keil et al., 1995a). Later on, more detailed SDM and expression studies in yeast and *E. coli* heterologous expression systems revealed this mutation decreased the abundance of mutant protein about 2-fold compared with the wild type, as well as caused 10-fold decrease of specific activity with less than 2-fold of differences of K_M values for both substrates (Lai and Elsas, 2001; Wells and Fridovich-Keil, 1997). There was no apparent difference in releasing glu-1P between the wild type and this mutant (Lai and Elsas, 2001). Mutating this serine to alanine, cysteine, histidine, threonine or tyrosine by SDM confirmed that a hydroxyl group is required on the side chain of amino acid 135, since only the threonine substitution resulted in active enzyme (Lai and Elsas, 2001).

The *K285N* mutation compromises the activity of the enzyme, as well as its abundance, in the yeast expression system (Riehman et al., 2001). As for the *N314D* mutation, it was regarded as the reason of reduced enzymatic activity in Duarte 2 patients; but detailed enzymatic studies facilitated by SDM revealed that the mutation itself only causes isoelectric point shifting, without affecting protein abundance, subunit dimerization or activity (Fridovich-Keil et al., 1995b). The decrease in GALT activity observed in the Duarte type 2 patients is likely caused by the 4-bp deletion at the promoter region associated with the *N314D* mutation, which abolishes the binding sites of two transcription factors to the GALT gene promoter (Carney, et al., 2009). The fact that the Los Angeles variant has normal activity in the erythrocytes supports this conclusion (Carney et al., 2009).

3.3.2 Type II (GALK-deficiency) galactosemia

More than 20 mutations associated with GALK deficiency have been reported to date. Through SDM studies, the majority of the mutations have been characterized. By expressing 10 variant GALK enzymes in GALK-less *E. coli*, Timson and Reece showed that five of mutant GALK enzymes (P28T, V32M, G36R, T288M and A384P), which are associated with more severe clinical phenotypes and near-zero blood galactokinase levels, are insoluble

(Timson and Reece, 2003). Further studies showed that these mutations disrupted the secondary structure of the enzymes, which could result in misfolding of the protein (Thoden et al., 2005). Four of the five soluble mutants (H44Y, R68C, G346S, and G349S, but not A198V) have impaired enzymatic properties, such as increased K_M for one or both substrates and decreased k_{cat}. All five are associated with low blood enzyme levels and milder symptoms. From the crystal structure of human GALK, it is clear that His[44], Gly[346] and Gly[349] are close to the active site. Additionally, these residues reside in the signature motif III of the GHMP kinase superfamily (Bork et al., 1993; Thoden et al., 2005). Therefore, it is not surprising that any changes in these resides would alter the kinetic parameters of the enzyme. As for A198V, its kinetic parameters are essentially indistinguishable from the wild-type enzyme. Compared to other mutations, from which patients will develop cataracts with high incidence within the first few years (without treatment), the A198V enzyme causes only a moderate incidence of cataracts in later life.

Similarly, Park and colleagues characterized another four missense mutations and one insertion (*G137R*, *R256W*, *R277Q*, *V281M* and *850_851insG*) by expressing the corresponding mutated genes in COS7 cells (Park et al., 2007). The steady-state expression level of R256W was lower than that of wild type. The stability of the mutant enzyme was significantly reduced, and it had no detectable activity. No protein was detected for the insertion variant. The other three mutations manifested enzymes with similar expression levels in the soluble fraction, as compared to the wild-type level. However, the *G137R* and *R277Q* enzymes had approximately 10%-15% of wild-type activity, and no activity was detected for the *V281M* enzyme.

3.3.3 Type III (GALE-deficiency) galactosemia

GALE deficiency exists in a continuum, from generalized to peripheral *via* intermediate (Openo et al., 2006). If GALE is deficient in all tissues, it is classified as generalized; and, if it is only deficient in red and white cells but normal in other tissues, it is known as peripheral deficiency. It is possible that the presence of bi-allelic amorphic mutations is incompatible with life (Sanders et al., 2010). Infants with generalized deficiency develop disease on a lactose-containing milk diet, while infants with peripheral disease remain well, at least in the newborn period. GALE deficiency has been extensively reviewed by Fridovich-Keil and coworkers (Fridovich-Keil et al., 1993a). Genomic *GALE* is about 5 kb in length, with multiple alternatively spliced transcripts. Some of the reported mutations are deposited in the HGMD database (http://www.hgmd.org/). Few case series have been reported, including a Korean study, reporting 37 patients with reduced GALE activity (Park et al., 2005), and two US-based studies, with one reporting 35 patients (Maceratesi et al., 1998) and the other, 10 patients (Openo et al.,2006). Others have reported a few cases (Alano et al., 1998; Wohlers et al., 1999). The *V94M* mutation has been reported in the homozygous state as being associated with generalized disease (Wohlers et al., 1999). In-depth studies of the *V94M* mutation through SDM in the yeast system showed that this mutation severely damages the specific activity of the enzyme predominantly at the level of V_M without affecting its abundance and thermal stability (Wohlers et al., 1999; Wohlers and Fridovich-Keil, 2000). In the same study, the *G90E* mutation was shown to have zero enzymatic activity, rendering the mutant enzyme to high temperature and protease (Wohlers et al., 1999). A more recent study further confirmed the impact of *V94M* and *G90E* on V_M (Timson 2005). Other missense mutations have not (yet) been

reported in patients, but they have been studied *in vitro* or in model systems. They are associated with severe enzyme deficiency; these include *G90E* and *L183P* (Quimby et al., 1997; Timson, 2005; Wohlers et al., 1999). Missense mutations associated with peripheral disease include *R169W*, *R239W* and *G302A* and have been described by Park and coworkers in individuals with peripheral GALE deficiency (Park et al., 2005). The *K257R* and *G319E* mutations have been described in African-Americans with peripheral deficiency (Alano et al., 1998). The *L183P* mutation encodes an enzyme that experiences severe proteolytic degradation during expression and purification. Also the authors showed that enzymes resulting from the *N34S*, *G90E* and *D103G* mutations exhibited increased susceptibility to digestion in limited proteolysis experiments (Timson 2005). An earlier study on *L183P* and *N34S* using SDM in a yeast model revealed that the L183P-hGALE mutant demonstrated 4% wild-type activity and 6% wild-type abundance, while N34S-hGALE demonstrated approximately 70% wild-type activity and normal abundance. However, yeast cells co-expressing both L183P-hGALE and N34S-hGALE exhibited only approximately 7% wild-type levels of activity, thereby confirming the functional impact of having both substitutions and raising the intriguing possibility that some form of dominant-negative interaction may exist between the mutant enzymes found in this patient (Quimby et al., 1997). Two other mutations, *D130G* and *L313M*, which are associated with intermediate epimerase deficiency, manifested enzymes with near normal GALE activity, but with compromised thermal stability and protease-sensitivity (Wohlers et al., 1999). Three other mutations associated with intermediate forms (*S81R*, *T150M* and *P293L*) were analyzed for their kinetic and structural properties *in vitro* and their effects on galactose-sensitivity of *S. cerevisiae* cells in the absence of Gal10p. All three mutations result in impairment of the kinetic parameters, principally the turnover number, k_{cat}, compared to the wild-type enzyme. However, the degree of impairment was mild compared with that seen with the mutation *V94M* (Chhay et al., 2008). Studies are limited by the fact the many patients are compound heterozygotes and by the observation that dominant-negative interactions may be involved in some of these cases.

4. Use of SDM in the understanding of catalytic mechanisms of the human GAL enzymes

4.1 The issue

Although the Leloir pathway is evolutionarily conserved and is indispensable for productive galactose metabolism, the catalytic mechanisms of the GAL enzymes are largely unknown.

4.2 Research design

Several groups have attempted to combine the techniques of SDM, analytical biochemistry and X-ray crystallography to advance the understanding of the catalytic mechanisms of the different GAL enzymes.

4.3 The results

4.3.1 GALK

GALK converts galactose to gal-1P by transferring γ–phosphate group of ATP to the O1 position of galactose. It belongs to a unique kinase superfamily – the GHMP kinase family,

which is named after four characteristic family members: galactokinase (GALK), homoserine kinase (HSK), mevalonate kinase (MVK) and phosphomevalonate kinase (PMVK) (Bork et al., 1993). This family of proteins was first identified by three highly conserved motifs among the four kinases mentioned above by sequence alignment and analysis. Motifs I and III are located at the N-terminal and C-terminal ends; and motif II, the most conserved, is located in the middle of the protein, with the consensus sequence of GLGSS(G/A/S) (Holden et al., 2004).

Interestingly, two different catalytic mechanisms have been proposed for this family. A common catalytic strategy to achieve nucleophilic attack is to use a negative charged residue, such as aspartate or glutamate, to act as a Brønsted base. This catalytic base can then abstract a proton from the hydroxyl group of the substrate converting the weakly nucleophilic hydroxyl group into the more strongly nucleophilic alkoxide ion, which then attacks the electron-deficient phosphorus atom in ATP (Fig. 1A). In such systems, it is common to find positively-charged lysine or arginine residues close to the catalytic site to help stabilize the negative charges on the enzyme and the substrates. Studies on MVK suggest this enzyme follows this mechanism. The crystal structure of MVK reveals an aspartate (residue 204 in the rat enzyme) positioned to act as an active site base. There is also a lysine (residue 13 in rat MVK), which is close to both the putative catalytic aspartate residue and the hydroxyl group of the substrate (Fu et al., 2002; Yang et al., 2002). Replacement of the lysine residue with a methionine by SDM resulted in a reduced, but non-zero, rate (V_{max} was reduced approximately 60-fold) (Potter et al., 1997). Similar results were observed when the equivalent lysine (residue 18) was changed to methionine in yeast mevalonate diphosphate decarboxylase (Krepkiy and Miziorko, 2004). These results are consistent with this positively-charged residue playing an assisting, but non-vital, role in catalysis. Crystal structures of GALK put it into this mechanism by revealing there are aspartate and arginine residues in the active center close to the galactose C1 hydroxyl group (Asp[186] and Arg[37] in the human structure, Asp[183] and Arg[36] in *Lactococcus lactis*) (Thoden and Holden, 2003; Thoden et al., 2005). Similarly, changing Arg[37] of human GALT to alanine resulted in a nearly inactive enzyme; and lysine resulted in compromised k_{cat} and K_M for galactose (Tang, et al., 2010).

In contrast, phosphoryl transfer in HSK has been suggested to occur by direct nucleophilic attack on the γ-phosphate group of ATP by the δ-hydroxyl of homoserine (Fig. 1B) (Krishna et al., 2001). In this mechanism, the latter is stabilized by the formation of a hydrogen bond to a neighboring asparagine residue (Asn[141]), which is not conserved in the superfamily. Catalysis is proposed to be assisted through activation of the γ-phosphate of ATP by the magnesium ion, which is coordinated by a conserved glutamate residue (Glu[130]) with the deprotonation of the δ-hydroxyl possibly involving the γ-phosphate (Krishna et al., 2001).

4.3.2 GALT

GALT catalyzes the transfer of the uridine monophosphate group (UMP) from uridine diphosphate-glucose (UDP-Glu) to gal-1p to form uridine diphosphate-galactose (UDP-Gal) and glucose-1-phosphate (glu-1P) (Kalckar et al., 1953). The reaction follows the double displacement mechanism as shown in Fig. 2 (Arabshahi et al., 1986). The most characteristic
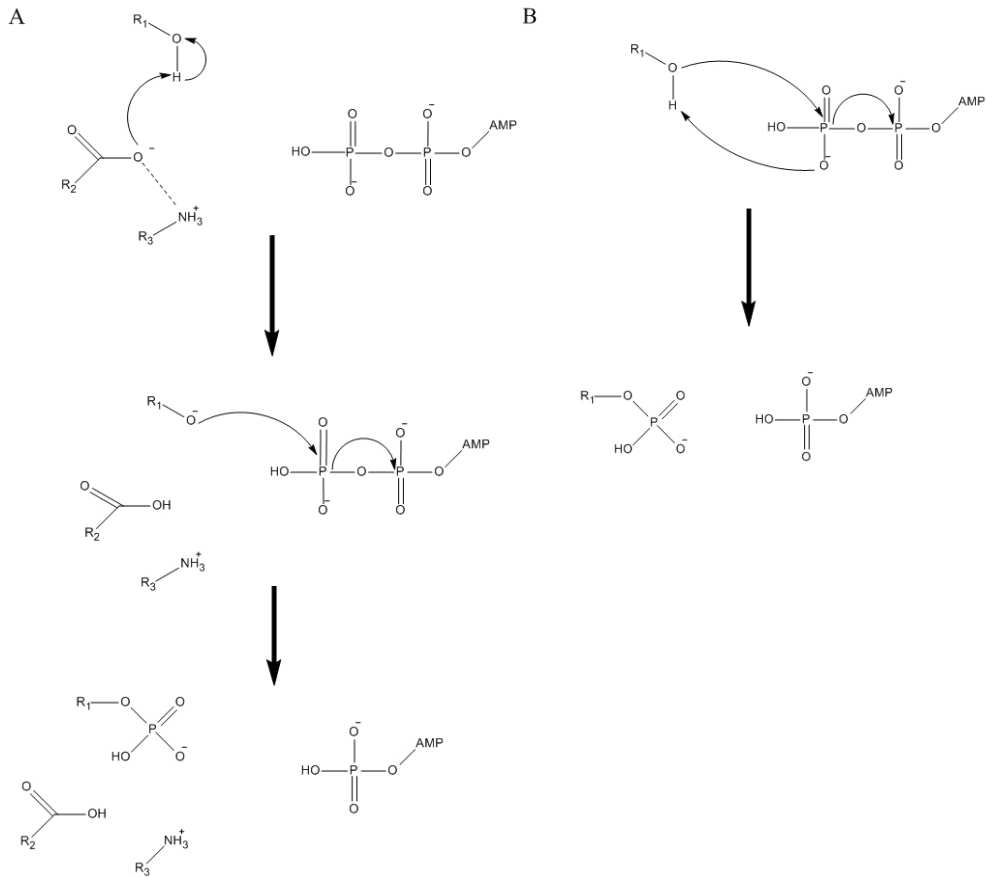
Fig. 1. Catalytic mechanisms proposed for GHMP kinase. **A**. The enzyme catalyzes the reaction through an active base residue R_1, which attracts a proton from the substrate R_3, converting the weakly nucleophilic hydroxyl to an alkoxide ion, which attacks the γ-phosphate of ATP. A positively charged residue R_2, sits close to the catalytic residue and stabilizes the alkoxide ion. **B**. There is no active base residue in the active center, the substrate directly attacks the γ–phosphate of ATP.

feature of the reaction is forming a covalent UMP-enzyme intermediate (Arabshahi et al., 1986). The intermediate was isolated by gel permeation chromatography in reaction mixtures containing the enzyme and radiolabeled UDP-Glu, and the radiolabeled intermediate could react with gal-1P or glu-1P to form the corresponding radiolabeled UDP sugar (Wong, et al., 1977a). This intermediate is very fragile in slightly acidic solutions but quite stable in strong basic solutions (Wong et al., 1977a; Yang and Frey, 1979), which indicates the intermediate is phosphoramides. Further degradation study of this intermediate confirmed that the nucleophile in GALT, to which the uridylyl group is bonded in the uridylyl-enzyme intermediate, is imidazole N3 of a histidine residue (Yang and Frey, 1979).
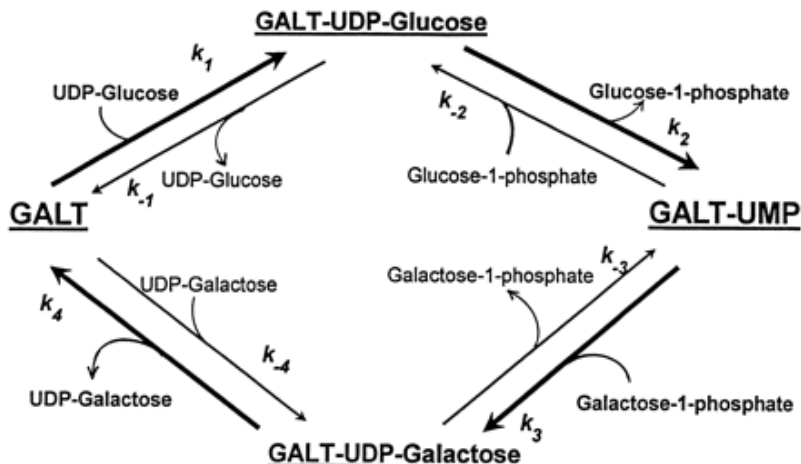
Fig. 2. Double displacement reactions of GALT. GALT binds to UDP-Glu to form a GALT-UDP-Glu intermediate. Glu-1-P is subsequently released, whereas the enzyme remains bound to UMP. Gal-1-P then reacts with the enzyme-UMP intermediate to form UDP-Gal, freeing the GALT enzyme for continued catalysis. k_n and k_{-n} denote rate constants of the forward and reverse reactions.

Substituting each of the 15 histidine residues in *E. coli* GALT with asparagines by SDM, proved that His[164] and His[166] were the only essential histidine residues in the enzyme (Field et al., 1989). In order to identify which of these two residues is the catalytic residue, two more specific mutations were introduced by SDM, *H164G* and *H166G*, which resulted in loss of function of the enzyme because of the missing imidazole ring of histidine, which might be filled and salvaged by adding exogenous imidazole ring. The experimental results showed that the activity of the H166G mutant could be recovered by adding exogenous imidazole ring, while mutant H164G could not. Therefore, His[166] provides the catalytic nucleophilic imidazole ring in the reaction (Kim et al., 1990).

Also, as mentioned earlier, by mutating Gln[188] of human GALT (equivalent to Gln[168] in *E. coli* GALT), the most common mutation found in Type I Galactosemia, to arginine and asparagine, respectively, we were able to determine that glutamine at position 188 stabilizes the UMP-GALT intermediate through hydrogen bonding and enables the double displacement of both glucose-1-phosphate (glu-1P) and UDP-galactose. The substitution of arginine or asparagine at position 188 reduces hydrogen bonding and destabilizes UMP-GALT. The unstable UMP-GALT allows single displacement of glu-1P with release of free GALT but impairs the subsequent binding of gal-1P and displacement of UDP-Gal (Lai, et al., 1999).

4.3.3 GALE

GALE catalyzes the inter-conversion of UDP-Glu and UDP-Gal to finish the Leloir pathway of galactose metabolism. There are four key steps for the reaction of GALE as shown in Fig. 3: (1) abstraction of the 4'-hydroxyl hydrogen of the sugar by an enzymatic base, (2) transfer of a hydride from C4 of the sugar to the C4 of NAD+ leading to a 4'-ketopyranose intermediate and NADH, (3) rotation of the resulting 4'-ketopyranose intermediate in the

active site, and (4) return of the hydride from NADH to the opposite face of the sugar (Maitra and Ankel, 1971). When purified, this enzyme contains tightly bound NAD+, which functions as an essential coenzyme to catalyze the reaction (Darrow and Rodstrom, 1968). The binding of the UDP group is strong, while binding with the galactosyl, glucosyl and 4-ketohexopyranosyl moieties is weak (Kang et al., 1975; Wong and Frey, 1977b). Early study on the catalytic mechanism of GALE focused on Lys^{153}, since it is close to the NAD+, and the positively-charged ammonium group of Lys^{153} may perturb the electron distribution in the nicotinamide ring of NAD+ through charge repulsion upon substrate binding (Swanson and Frey, 1993). Replacing this residue with alanine or methionine renders the inability of the mutant proteins to be reduced by the sugar in the presence or absence of UMP. As a result, the catalytic activities of the mutants decreased by a factor over 1000. Also the purified mutant contained much less NADH as compared with wild type (Swanson and Frey, 1993). These results indicate that Lys^{153} plays an important role in the UMP-dependent reduction of GALE-NAD+. Further studies identified two more important residues, Tyr^{149} and Ser^{124}, which are involved in glucose moiety binding (Thoden et al., 1996). SDM studies on the latter two residues revealed that that Tyr^{149} provides the driving force for general acid-base catalysis, while Ser^{124} plays an important role in mediating proton transfer (Liu et al., 1997). The crystal structure of human GALE confirmed that Tyr^{149} (Tyr^{157} for human GALE) sits at the proper position to interact directly with the 4'-hydroxyl group of the sugar and attracts the proton from the hydoxy group and transfers it to NAD+ (Thoden et al., 2000).

Unlike what was observed for the *E. coli* enzyme, the human enzyme can also convert UDP-*N*-acetylglucosamine (UDP-GlcNAc) to UDP-*N*-acetylgalactosamine (UDP-GalNAc) (Kingsley et al., 1986; Piller et al., 1983). Through structure analysis and alignment, investigators found that, when the human enzyme equivalent of Tyr^{299} in the *E. coli* protein is replaced with a cysteine residue (Cys^{307}), the active site volume for the human protein is calculated to be approximately 15% larger than that observed for the bacterial epimerase (Thoden 2001). Substituting Tyr^{299} of *E. coli* GALE with a cysteine residue by SDM confers UDP-GalNAc/UDP-GlcNAc converting activity to the bacterial enzyme with minimal changes in its three-dimensional structure. Specifically, although the Y299C mutation in the bacterial enzyme resulted in a loss of epimerase activity with regard to UDP-Gal by almost 5-fold, it resulted in a gain of activity against UDP-GalNAc by more than 230-fold (Thoden et al., 2002b).

5. Use of SDM in the development of novel treatment of Type I (classic or GALT-deficiency) galactosemia

5.1 The issues

Unlike Type II or the peripheral Type III Galactosemia, patients with Type I (GALT-deficiency) Galactosemia, also the most common type of Galactosemia, suffer a range of debilitating long-term complications, which include premature ovarian insufficiency, learning deficits, ataxia and speech dyspraxia (Lai et al., 2009; Berry and Elsas, 2011). The current galactose-restricted diet fails to prevent these complications, and the medical/ patient communities are yearning for a more effective therapy. The causes of these organ-specific complications remain unknown, but there is a strong association with the intracellular accumulation of gal-1P. But what is the source of gal-1P in these patients with Classic Galactosemia if they limit their galactose intake? Recent studies have shown that the

patients on a galactose-restricted diet are never really "galactose-free. A significant amount of galactose is found in non-dairy foodstuffs, such as vegetables and fruits (Berry et al., 1993; Acosta and Gross, 1995). More importantly, galactose is produced endogenously from the natural turnover of glycolipids and glycoproteins (Berry et al., 1995). Using isotopic labeling, Berry and coworkers demonstrated that a 50kg adult male could produce up to 2 grams of galactose per day (Berry et al., 1995, 2004). Once galactose is formed intracellularly, it is converted to gal-1P by GALK and in GALT-deficient patient cells. As a result, gal-1P is concentrated more than one order of magnitude above normal, even with strict adherence to a galactose-restricted diet. Accumulation of gal-1P is regarded as a major, if not sole, factor for the chronic complications seen in patients with Classic Galactosemia, as suggested by both clinical observation and experimental results from yeast models. Patients with inherited deficiency of GALK, who do not accumulate gal-1P, do not experience the brain and ovary complications seen in GALT-deficient patients (Gitzelmann et al., 1974; Gitzelmann 1975; Stambolian et al., 1986). While *gal7* (*i.e*, GALT-deficient) mutant yeast stops growing upon galactose challenge, a *ga17 ga11* double mutant strain (*i.e*, GALT- and GALK-deficient) is no longer sensitive to galactose (Douglas and Hawthorne, 1964, 1966). Based on these observations, in conjunction with dietary therapy, inhibiting GALK activity with a safe small-molecule inhibitor might prevent the squeals of chronic gal-1P exposure in patients with Classic Galactosemia.
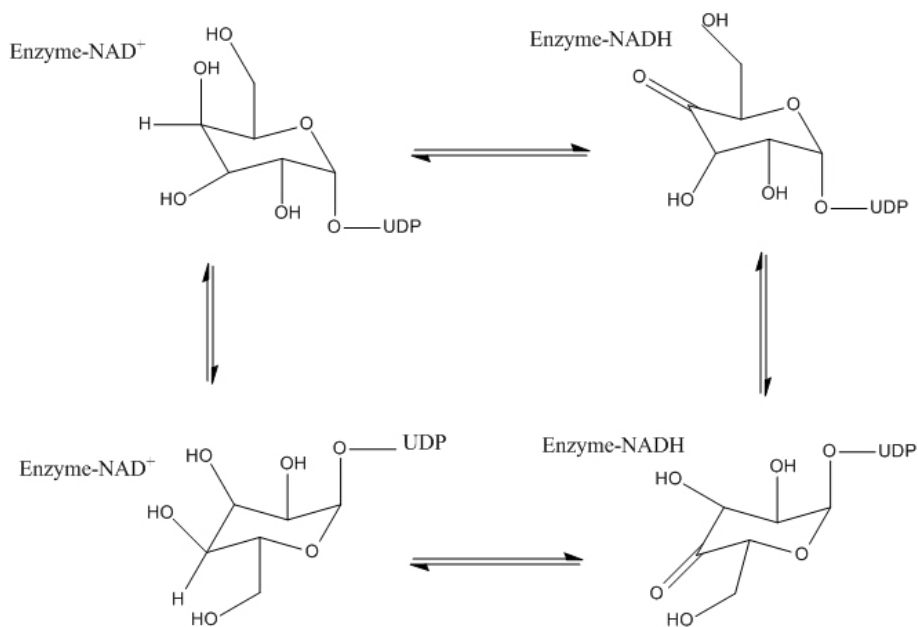
Fig. 3. Catalytic mechanism of GALE.

5.2 Research design

For the past few years, our group has conducted high-throughput screening (HTS) of small molecule compounds, which could inhibit human GALK enzyme *in vitro* (Tang et al., 2010;

Wierenga et al., 2008). To date, we have screened over 300,000 compounds of diverse chemical structures and identified a few promising hit compounds for further characterization. One of the characterization steps involved the use of SDM to change the respective amino acids of the GALK active site in order to confirm the predicted molecular interactions between the selected inhibitors and it target, GALK, through high-precision docking programs such as GLIDE (*Schrödinger*). Another characterization step that is noteworthy to mention is the assay for the kinase selectivity of the selected GALK inhibitors. As alluded above, GALK belongs to a unique small molecule kinase family, the GHMP kinase family (Bork et al., 1993). While the substrates of the GHMP kinases differ widely, the ATP-binding sites of the enzymes share a significant degree of structural homology (Tang et al., 2010). It is, therefore, important to ensure our selected GALK inhibitors did not cross-inhibit other GHMP kinases or other kinases in general.

5.3 The results

Selectivity is always one of the most important properties for developing therapeutic kinase inhibitors because of potential side-effects from unwanted inhibition of other kinases. During the characterization phase of our hit compounds, we found six compounds that selectively inhibit GALK but not any of the other GHMP kinases. These included MVK, which shares a high degree of structural similarity with GALK (Tang et al., 2010). In order to understand what structural elements conferred the specificity of these compounds, we aligned the crystal structure of human GALK and human MVK and focused on the ATP-binding site. Eight amino acid residues and the L1 loop were found to be different in these two kinases. SDM was employed to mutate each residue individually or the L1 loop, and the effects of the changes on the inhibitory capabilities of the compounds were tested. Two compounds were found to be affected by the mutation *S140G* (Table 1) (Tang et al., 2010). Ser[140] of GALK resides in the signature motif of the GHMP kinase family, Motif II; but this amino acid is not conserved among the GHMP kinases. GALK is the only member that has a serine at this site. This could explain the selectivity of these two compounds. Furthermore, computational molecular docking confirmed that these two compounds interacted with Ser[140] through hydrogen bonds; substituting serine with glycine abolished the hydrogen bonds and totally compromised the binding of the compounds to the enzymes.

Our use of SDM in the characterization of promising GALK inhibitors not only helped identify and confirm the amino acids of GALK with which these small molecules interact, but also exemplified a more rapid and cost-effective way to study the structural interactions between small molecule modifiers and their targets. This novel approach is particularly useful when large-scale co-crystallization projects are not feasible. These studies paved the way for more in-depth investigations to identify the structural determinants required for the inhibitor selectivity of GALK and GHMP kinases.

6. Concluding remarks

Using the disease Galactosemia as an example, we showed that site-directed mutagenesis (**SDM**) plays a vital role in biomedical research. As in the case of Galactosemia, in which the diagnosis begins at the bedside of the affected newborns, **SDM** can be employed in every step of basic and translational research in an attempt to improve the prognosis and

treatment of the patients. Further, not only did we show that **SDM** can be applied in traditional applications, such as expression analysis, we have also expanded its use in innovative drug design and the basic understanding of kinase inhibitor selectivity.

Mutations	k_{cat} (S^{-1})	K_M of ATP (μM)	K_M of Galactose (μM)	Effects on IC$_{50}$ of compound 1	Effects on IC$_{50}$ of compound 4	Effects on IC$_{50}$ of compound 24
T77L	4.1	218.4	1305.2	None	None	None
S79N	4.8	303.4	1227.3	None	None	None
L145Y	11.6	259.7	222.7	None	None	None
L145A	6.4	379.9	356.8	None	None	None
W106A	No protein expression	-	-	-	-	-
W106T	No protein expression	-	-	-	-	-
Y109L	43.2	70.2	963.2	None	None	None
Y109A	8.7	579.3	268.7	None	None	None
GALK Loop to MVK Loop	0.1	695.4	1857.3	None	None	None
S140G	2.1	8.2	141.9	None	Increased 10-fold	Increased 20-fold
L135P	13.3	51.1	544.9	None	None	None
R37K	0.4	6.4	623.8	None	None	None
R37A	No activity	-	-	-	-	-
WT	17.5	20.9	319	-	-	-

Table 1. Effect of amino acid changes in human GALK on their enzymatic properties and the IC$_{50}$ of selected inhibitors

7. Acknowledgement

8. References

Acosta, P. B. and Gross, K. C., 1995. Hidden sources of galactose in the environment. Eur. J. Pediatr. 154(7 Suppl 2): S87-92.

Alano, A., Almashanu, S., Chinsky, J. M., Costeas, P., Blitzer, M. G., Wulfsberg, E. A., and Cowan, T.M., 1998. Molecular characterization of a unique patient with epimerase-deficiency galactosaemia. J. Inherit. Metab. Dis. 21(4): 341-350.

Alano A, Almashanu, S., Maceratesi, P., Reichardt, J., Panny, S., and Cowan T.M., 1997. UDP-galactose-4-epimerase deficiency among African-Americans: evidence for multiple alleles. J. Invest. Med. 45: :191A.

Arabshahi, A., Brody, R. S., Smallwood, A., Tsai, T.C., and Frey, P.A., 1986. Galactose-1-phosphate uridylyltransferase. Purification of the enzyme and stereochemical course of each step of the double-displacement mechanism. Biochemistry. 25(19): 5583-5589.

Beebe, J. A. and Frey, P. A., 1998. Galactose mutarotase: purification, characterization, and investigations of two important histidine residues. Biochemistry. 37(42): 14989-14997.

Berry, G. T., Palmieri, M., Gross, K. C., Acosta, P. B., Henstenburg, J. A., Mazur, A., Reynolds, R., and Segal, S., 1993. The effect of dietary fruits and vegetables on urinary galactitol excretion in galactose-1-phosphate uridyltransferase deficiency. J. Inherit. Metab. Dis. 16(1): 91-100.

Berry, G. T., Nissim, I, Lin, Z., Mazur, A. T., Gibson, J. B., and Segal, S., 1995. Endogenous synthesis of galactose in normal men and patients with hereditary galactosaemia. Lancet. 346(8982): 1073-1074.

Berry, G. T., Moate, P. J., Reynolds, R.A., Yager, C. T., Ning, C., Boston, R. C., and Segal, S., 2004. The rate of de novo galactose synthesis in patients with galactose-1-phosphate uridyltransferase deficiency. Mol. Genet. Metab. 81(1): 22-30.

Berry, G. T. and Elsas, L. J., 2011. Introduction to the Maastricht workshop: lessons from the past and new directions in galactosemia. J. Inherit. Metab. Dis. 34(2): 249-255.

Bork, P., Sander, C., and Valencia, A., 1993. Convergent evolution of similar enzymatic function on different protein folds: the hexokinase, ribokinase, and galactokinase families of sugar kinases. Protein. Sci. 2(1): 31-40.

Bosch, A. M., Bakker, H. D., van Gennip, A. H., van Kempen, J. V., Wanders, R. J., and Wijburg, F. A., 2002. Clinical features of galactokinase deficiency: a review of the literature. J. Inherit. Metab. Dis. 25(8): 629-634.

Cardini, C. E. and Leloir, L. F., 1953. Enzymic phosphorylation of galactosamine and galactose. Arch. Biochem. Biophys. 45(1): 55-64.

Carney, A. E., Sanders, R. D.,, Garza, K. R., McGaha, L. A., Bean, L. J., Coffee, B. W., Thomas, J. W., Cutler, D. J., Kurtkaya, N. L., and Fridovich-Keil, J. L., 2009. Origins, distribution and expression of the Duarte-2 (D2) allele of galactose-1-phosphate uridylyltransferase. Hum. Mol. Genet. 18(9): 1624-1632.

Chhay, J. S., Vargas, C. A.,, McCorvie, T. J., Fridovich-Keil, and Timson, D. J., 2008. Analysis of UDP-galactose 4'-epimerase mutations associated with the intermediate form of type III galactosaemia. J. Inherit. Metab. Dis. 31(1): 108-116.

Coffee, B., Hjelm, L. N., DeLorenzo, A., Courtney, E. M., Yu, C., and Muralidharan, K., 2006. Characterization of an unusual deletion of the galactose-1-phosphate uridyl transferase (GALT) gene. Genet. Med. 8(10): 635-640.

Darrow, R. A. and Rodstrom, R., 1968. Purification and properties of uridine diphosphate galactose 4-epimerase from yeast. Biochemistry. 7(5): 1645-1654.

Douglas, H. C. and Hawthorne, D. C., 1964. Enzymatic Expression And Genetic Linkage Of Genes Controlling Galactose Utilization In Saccharomyces. Genetics 49: 837-844.

Douglas, H. C. and Hawthorne, D. C., 1966. Regulation of genes controlling synthesis of the galactose pathway enzymes in yeast. Genetics 54(3): 911-916.

Elsas, L. J., 1993. Galactosemia. (http://www.ncbi.nlm.nih.gov/books/NBK1518/)

Elsas, L. J., Dembure, P. P., Langley, S., Paulk, E. M., Hjelm, L. N., and Fridovich-Keil, J., 1994. A common mutation associated with the Duarte galactosemia allele. Am. J. Hum. Genet. 54(6): 1030-1036.

Elsevier, J. P. and Fridovich-Keil, J. L., 1996. The Q188R mutation in human galactose-1-phosphate uridylyltransferase acts as a partial dominant negative. J. Biol. Chem. 271(50): 32002-32007.

Field, T. L., Reznikoff, W. S., and Frey, P. A.., 1989. Galactose-1-phosphate uridylyltransferase: identification of histidine-164 and histidine-166 as critical residues by site-directed mutagenesis. Biochemistry 28(5): 2094-2099.

Fridovich-Keil, J., Bean, L., He, M., and Schroer, R., 1993a. Epimerase Deficiency Galactosemia. (http://www.ncbi.nlm.nih.gov/books/NBK51671/)

Fridovich-Keil, J. L. and Jinks-Robertson, S., 1993b. A yeast expression system for human galactose-1-phosphate uridylyltransferase. Proc. Natl. Acad. Sci. U. S. A. 90(2): 398-402.

Fridovich-Keil, J. L., Langley, S. D., Mazur, L. A., and Elsevier, J. P., 1995a. Identification and functional analysis of three distinct mutations in the human galactose-1-phosphate uridyltransferase gene associated with galactosemia in a single family. Am. J. Hum. Genet. 56(3): 640-646.

Fridovich-Keil, J. L., Quimby, B. B., Wells, L., Mazur, L. A., and Elsevier, J. P., 1995b. Characterization of the N314D allele of human galactose-1-phosphate uridylyltransferase using a yeast expression system. Biochem. Mol. Med. 56(2): 121-130.

Fu, Z., Wang, M., Potter, D., Miziorko, H. M., and Kim, J. J., 2002. The structure of a binary complex between a mammalian mevalonate kinase and ATP: insights into the reaction mechanism and human inherited disease. J. Biol. Chem. 277(20): 18134-18142.

Gitzelmann, R., Wells, H. J., and Segal, S., 1974. Galactose metabolism in a patient with hereditary galactokinase deficiency. Eur. J. Clin. Invest. 4(2): 79-84.

Gitzelmann, R., 1975. Letter: Additional findings in galactokinase deficiency. J. Pediatr. 87(6 Pt 1): 1007-1008.

Greber-Platzer, S., Guldberg, P., Scheibenreiter, S., Item., C., Schuller, E., Patel, N., and Strobl, W., 1997. Molecular heterogeneity of classical and Duarte galactosemia: mutation analysis by denaturing gradient gel electrophoresis. Hum. Mutat. 10(1): 49-57.

Guerrero, N. V., Singh, R. H., Manatunga, A., Berry, G. T., Steiner, R. D., Elsas, L. J., 2nd, 2000. Risk factors for premature ovarian failure in females with galactosemia. J. Pediatr. 137(6): 833-841.

Holden, H. M., Thoden, J. B., Timson, D. J., and Reece, R. J., 2004. Galactokinase: structure, function and role in type II galactosemia. Cell. Mol. Life. Sci. 61(19-20): 2471-2484.

Kalckar, H. M., Braganca, B., Munch-Petersen, H. M., 1953. Uridyl transferases and the formation of uridine diphosphogalactose. Nature. 172(4388): 1038.

Kang, U. G., Nolan, L. D., and Frey, P.A., 1975. Uridine diphosphate galactose-4-epimerase. Uridine monophosphate-dependent reduction by alpha- and beta-D-glucose. J. Biol. Chem. 250(18): 7099-7105.

Kaye, C. I., Accurso, F., La Franchi, S., Lane, P. A., Hope, N., Sonya, P., S, G. B., and Michele, A. L., 2006. Newborn screening fact sheets. Pediatrics. 118(3): e934-963.

Kim, J., Ruzicka, F., and Frey, P.A., 1990. Remodeling hexose-1-phosphate uridylyltransferase: mechanism-inspired mutation into a new enzyme, UDP-hexose synthase. Biochemistry. 29(47): 10590-10593.

Kingsley, D. M., Kozarsky, K. F., Hobbie, L., and Krieger, M., 1986. Reversible defects in O-linked glycosylation and LDL receptor expression in a UDP-Gal/UDP-GalNAc 4-epimerase deficient mutant. Cell. 44(5): 749-759.

Kozak, L., Francova, H., Fajkusova, L., Pijackova, A., Macku, J., Stastna, S., Peskovova, K., Martincova, O., Krijt, J., and Bzduch, V., 2000. Mutation analysis of the GALT gene in Czech and Slovak galactosemia populations: identification of six novel mutations, including a stop codon mutation (X380R). Hum. Mutat. 15(2): 206.

Krepkiy, D. and Miziorko, H. M., 2004. Identification of active site residues in mevalonate diphosphate decarboxylase: implications for a family of phosphotransferases. Protein Sci. 13(7): 1875-1881.

Krishna, S. S., Zhou, T., Daugherty, M., Osterman, A., and Zhang, H., 2001. Structural basis for the catalysis and substrate specificity of homoserine kinase. Biochemistry. 40(36): 10810-10818.

Lai, K., Langley, S. D., Singh, R. H., Dembure, P. P., Hjelm, L. N., and Elsas, L. J., 2nd, 1996. A prevalent mutation for galactosemia among black Americans. J. Pediatr. 128(1): 89-95.

Lai, K., Willis, A. C., and Elsas, L. J., 1999. The biochemical role of glutamine 188 in human galactose-1-phosphate uridyltransferase. J. Biol. Chem. 274(10): 6559-6566.

Lai, K. and Elsas, L. J., 2001. Structure-function analyses of a common mutation in blacks with transferase-deficiency galactosemia. Mol. Genet. Metab. 74(1-2): 264-272.

Lai, K., Elsas, L. J., and Wierenga, K. J., 2009. Galactose toxicity in animals. IUBMB Life. 61(11): 1063-1074.

Landt, M., Ritter, D., Lai, K., Benke, P. J., Elsas, L. J., and Steiner, R. D., 1997. Black children deficient in galactose 1-phosphate uridyltransferase: correlation of activity and immunoreactive protein in erythrocytes and leukocytes. J. Pediatr. 130(6): 972-980.

Leloir, L. F., 1951. The enzymatic transformation of uridine diphosphate glucose into a galactose derivative. Arch. Biochem. Biophys. 33(2): 186-190.

Leloir, L. F., 1953. Enzymic isomerization and related processes. Adv. Enzymol. Relat. Subj. Biochem. 14: 193-218.

Levy, H. L., 2010. Newborn screening conditions: What we know, what we do not know, and how we will know it. Genet. Med. 12(12 Suppl): S213-214.

Liu, Y., Thoden, J. B., Kim, J., Berger, E., Gulick, A. M., Ruzicka, F. J., Holden, H. M., and Frey, P. A., et al., 1997. Mechanistic roles of tyrosine 149 and serine 124 in UDP-galactose 4-epimerase from Escherichia coli. Biochemistry. 36(35): 10675-10684.

Maceratesi, P., Daude, N., Dallapiccola, B., Novelli, G., Allen, R., Okano, Y., and Reichardt, J., 1998. Human UDP-galactose 4' epimerase (GALE) gene and identification of five missense mutations in patients with epimerase-deficiency galactosemia. Mol. Genet. Metab. 63(1): 26-30.

Maitra, U. S. and Ankel, H., 1971. Uridine diphosphate-4-keto-glucose, an intermediate in the uridine diphosphate-galactose-4-epimerase reaction. Proc. Natl. Acad. Sci. U. S. A. 68(11): 2660-2663.

Murphy, M., McHugh, B., Tighe, O., Mayne, P., O'Neill, C., Naughten, E., Croke, D. T., 1999. Genetic basis of transferase-deficient galactosaemia in Ireland and the population history of the Irish Travellers. Eur. J. Hum. Genet. 7(5): 549-554.

Openo, K. K., Schulz, J. M., Vargas, C. A., Orton, C. S., Epstein, M. P., Schnur, R. E., Scaglia, F., Berry, G. T., Gottesman, G. S., Ficicioglu, C., Slonim AE, Schroer RJ, Yu C, Rangel VE, Keenan J, Lamance K, and Fridovich-Keil, J., 2006. Epimerase-deficiency galactosemia is not a binary condition. Am. J. Hum. Genet. 78(1): 89-102.

Park, H. D., Park, K. U., Kim, J. Q., Shin C. H., Yang, S. W., Lee, D. H., Song, Y. H., and Song, J , 2005. The molecular basis of UDP-galactose-4-epimerase (GALE) deficiency galactosemia in Korean patients. Genet. Med. 7(9): 646-649.

Park, H. D., Bang, Y. L., Park, K. U., Kim, J. Q., Jeong, B. H., Kim, Y.S., Song, Y. H., and Song, J., 2007. Molecular and biochemical characterization of the GALK1 gene in Korean patients with galactokinase deficiency. Mol. Genet. Metab. 91(3): 234-238.

Piller, F., Hanlon, M. H., and Hill, R. L., 1983. Co-purification and characterization of UDP-glucose 4-epimerase and UDP-N-acetylglucosamine 4-epimerase from porcine submaxillary glands. J. Biol. Chem. 258(17): 10774-10778.

Potter, D., Wojnar, J. M., Narasimhan, C., and Miziorko, H. M., 1997. Identification and functional characterization of an active-site lysine in mevalonate kinase. J. Biol. Chem. 272(9): 5741-5746.

Quimby, B. B., Alano, A., Almashanu S., DeSandro, A. M., Cowan, T. M., and Fridovich-keil, J. L., 1997. Characterization of two mutations associated with epimerase-deficiency galactosemia, by use of a yeast expression system for human UDP-galactose-4-epimerase. Am. J. Hum. Genet. 61(3): 590-598.

Reichardt, J. K., Packman, S., and Woo, S. L., 1991. Molecular characterization of two galactosemia mutations: correlation of mutations with highly conserved domains in galactose-1-phosphate uridyl transferase. Am. J. Hum. Genet. 49(4): 860-867.

Reichardt, J. K., Levy, H. L., and Woo, S. L., 1992. Molecular characterization of two galactosemia mutations and one polymorphism: implications for structure-function analysis of human galactose-1-phosphate uridyltransferase. Biochemistry. 31(24): 5430-5433.

Riehman, K., Crews, C., and Fridovich-Keil, J. L., 2001. Relationship between genotype, activity, and galactose sensitivity in yeast expressing patient alleles of human galactose-1-phosphate uridylyltransferase. J. Biol. Chem. 276(14): 10634-10640.

Sanders, R. D., Sefton, J. M., Moberg, K. H., and Fridovich-Keil, J. L., 2010. UDP-galactose 4' epimerase (GALE) is essential for development of Drosophila melanogaster. Dis. Model. Mech. 3(9-10): 628-638.

Segal, S., Berry, GT., 1995. Disorders of galactose metabolism. The Metabolic Basis of Inherited Diseases. B. A. Scriver D, Sly W, Valle D. New York, McGraw-Hill. I: 967-1000.

Shin, Y. S., Koch, H. G., Kohler, M., Hoffmann, G., Patsoura, A., Podskarbi, T., 1998. Duarte-1 (Los Angeles) and Duarte-2 (Duarte) variants in Germany: two new mutations in the GALT gene which cause a GALT activity decrease by 40-50% of normal in red cells. J. Inherit. Metab. Dis. 21(3): 232-235.

Stambolian, D., Scarpino-Myers, V., Eagle, R. C., Jr., Hodes, B., and Harris, H., 1986. Cataracts in patients heterozygous for galactokinase deficiency. Invest. Ophthalmol. Vis. Sci. 27(3): 429-433.

Swanson, B. A. and Frey, P. A., 1993. Identification of lysine 153 as a functionally important residue in UDP-galactose 4-epimerase from Escherichia coli. Biochemistry. 32(48): 13231-13236.

Tang, M., Wierenga, K., Elsas, L. J., and Lai, K., 2010. Molecular and biochemical characterization of human galactokinase and its small molecule inhibitors. Chem. Biol. Interact. 188(3): 376-385.

Thoden, J. B., Frey, P. A., and Holden, H. M., 1996. Molecular structure of the NADH/UDP-glucose abortive complex of UDP-galactose 4-epimerase from Escherichia coli: implications for the catalytic mechanism. Biochemistry. 35(16): 5137-5144.

Thoden, J. B., Wohlers, T. M., Fridovich-Keil, J. L., and Holden H. M., 2000. Crystallographic evidence for Tyr 157 functioning as the active site base in human UDP-galactose 4-epimerase. Biochemistry. 39(19): 5691-5701.

Thoden, J.B., Wholers. T., Fridovich-Keil, J.L., Holden, H.M., 2001. Human UDP-galactose 4-epimerase. Accommodation of UDP-N-acetylglucosamine within the active site. J. Biol. Chem. 4;276(18):: 15131-15136.

Thoden, J. B. and Holden, H. M., 2002a. High resolution X-ray structure of galactose mutarotase from Lactococcus lactis. J. Biol. Chem. 277(23): 20854-20861.

Thoden, J. B., Henderson, J. M., Fridovich-Keil, J. L., and Holden, H. M., 2002b. Structural analysis of the Y299C mutant of Escherichia coli UDP-galactose 4-epimerase. Teaching an old dog new tricks. J. Biol. Chem. 277(30): 27528-27534.

Thoden, J. B. and Holden, H. M., 2003. Molecular structure of galactokinase. J. Biol. Chem. 278(35): 33305-33311.

Thoden, J. B., Timson, D. J., Reece, R. J., and Holden, H. M., 2005. Molecular structure of human galactokinase: implications for type II galactosemia. J. Biol. Chem. 280(10): 9662-9670.

Timson, D. J. and Reece, R. J., 2003. Functional analysis of disease-causing mutations in human galactokinase. Eur. J. Biochem. 270(8): 1767-1774.

Timson, D. J., 2005. Functional analysis of disease-causing mutations in human UDP-galactose 4-epimerase. FEBS J 272(23): 6170-6177.

Waggoner, D. D., Buist, N. R., and Donnell, G. N., 1990. Long-term prognosis in galactosaemia: results of a survey of 350 cases."J. Inherit. Metab. Dis. 13(6): 802-818.

Waisbren, S. E., Potter, N. L., Gordon C. M., Green, R. C., Greenstein, P., Gubbels, C. S., Rubio-Gozalbo, E., Schomer, D., Welt, C., Anastasoaie, V., D'Anna, K, Gentile, J., Guo, C.Y., Hecht, L., Jackson, R., Jansma, B. M., Li, Y., Lip, V., Miller, D. T., Murray, M., Power, L., Quinn, N., Rohr, F., Shen, Y., Skinder-Meredith, A., Timmers, I., Tunick, R., Wessel, A., Wu, B. L, Levy, H., Berry, G. T., 2011. The adult galactosemic phenotype. J. Inherit. Metab. Dis.

Webb, A. L., Singh, R. H., Kennedy, M. J., and Elsas, L. J., 2003. Verbal dyspraxia and galactosemia. Pediatr. Res. 53(3): 396-402.

Wedekind, J. E., Frey, P. A., and Raymond, I., 1996. The structure of nucleotidylated histidine-166 of galactose-1-phosphate uridylyltransferase provides insight into phosphoryl group transfer. Biochemistry. 35(36): 11560-11569.

Wells, L. and Fridovich-Keil, J. L., 1997. Biochemical characterization of the S135L allele of galactose-1-phosphate uridylyltransferase associated with galactosaemia. J. Inherit. Metab. Dis. 20(5): 633-642.

Wierenga, K. J., Lai, K., Buchwald, P. and Tang, M., 2008. High-throughput screening for human galactokinase inhibitors. J. Biomol. Screen. 13(5): 415-423.

Wohlers, T. M., Christacos, N. C., Harreman, M. T., and Fridovich-Keil, J. L., 1999. Identification and characterization of a mutation, in the human UDP-galactose-4-epimerase gene, associated with generalized epimerase-deficiency galactosemia. Am. J. Hum. Genet. 64(2): 462-470.

Wohlers, T. M. and Fridovich-Keil, J. L. 2000. Studies of the V94M-substituted human UDPgalactose-4-epimerase enzyme associated with generalized epimerase-deficiency galactosaemia. J. Inherit. Metab. Dis. 23(7): 713-729.

Wong, L. J., Sheu, K. F., Lee, S. I. and Frey, P. A., 1977a. Galactose-1-phosphate uridylyltransferase: isolation and properties of a uridylyl-enzyme intermediate. Biochemistry. 16(5): 1010-1016.

Wong, S. S. and Frey, P. A., 1977b. Fluorescence and nucleotide binding properties of Escherichia coli uridine diphosphate galactose 4-epimerase: support for a model for nonstereospedific action. Biochemistry. 16(2): 298-305.

Yang, D., Shipman, L. W., Roessner, C. A., Scott, A. I., Sacchettini, J. C., 2002. Structure of the Methanococcus jannaschii mevalonate kinase, a member of the GHMP kinase superfamily. J. Biol. Chem. 277(11): 9462-9467.

Yang, S. L. and Frey, P. A., 1979. Nucleophile in the active site of Escherichia coli galactose-1-phosphate uridylyltransferase: degradation of the uridylyl-enzyme intermediate to N3-phosphohistidine. Biochemistry. 18(14): 2980-2984.

13

Inherited Connective Tissue Disorders of Collagens: Lessons from Targeted Mutagenesis

Christelle Bonod-Bidaud and Florence Ruggiero
Institut de Génomique Fonctionnelle de Lyon, ENS de Lyon, UMR CNRS 5242,
University Lyon 1,
France

1. Introduction

The extracellular matrix (ECM) is the cell structural environment in tissues and organs. The ECM is a dynamic structure that it is constantly remodelled. It contributes to tissue integrity and mechanical properties. It is also essential for maintaining tissue homeostasis, morphogenesis and differentiation, which it does, through specific interactions with cells. The ECM is composed of a mixture of water and macromolecules classified into four main categories: collagens, proteoglycans, elastic proteins, and non-collagenous glycoproteins (also called adhesive glycoproteins). The nature, concentration and ratio of the different ECM components are all important factors in the regulation of the assembly of complex tissue-specific networks tuned to meet mechanical and biological requirements of tissues.

Collagens form a superfamily of 28 trimeric proteins, distinguishable from the other ECM components by their particular abundance in tissues (collagens represent up to 80-90% of total proteins in skin, tendon and bones) and their capacity to self-assemble into supramolecular organized structures (the best known being the banded fibers). The collagen superfamily is highly complex and shows a remarkable diversity in structure, tissue distribution and function (Ricard-Blum and Ruggiero, 2005).

The importance of collagens has been illustrated by the wide range of mutations in collagen genes that result in minor and severe human diseases. Various mutations (point, null or structural mutations, insertions, exon skipping, deletions) in genes encoding collagens are known to be responsible for a large spectrum of human disorders (*e.g.*, Elhers-Danlos syndrome, epidermolysis bullosa, chondrodysplasia, osteogenesis imperfecta, Alport syndrome, Bethlem myopathy, Ulrich congenital muscular dystrophy, Fuchs' endothelial dystrophy, Knobloch syndrome) that affect different tissues and organs, such as skin, blood vessels, cartilage, bones, kidney, muscle, cornea and retina. Considering the variety of collagen-related diseases and the complexity of collagen biology, there is a clear need to understand how mutations alter collagen synthesis, cell trafficking, cell and molecular interactions to result in tissue dysfunction. In the eighties targeted mutagenesis emerged as a new approach to help establish the structure-function relationship of collagens. Along with the emergence of protein engineering and genetically modified mice, site-directed mutagenesis has become instrumental in understanding the physiopathology of diseases, as well as in developing new and specific therapies and drugs for the treatment of human

diseases. To date about 20 distinct genes encoding collagen chains have been ablated (by knock-out mutations) in mice or are involved in naturally occurring mutations. Only a few knock-in modified mice has been generated, in which a single point mutation or an exon deletion, for example, has been generated in a specific gene. This is likely due to the very large size of collagen genes. Site-directed knock-in mutations in mice have often proven to be more useful than knock-out mutations (which inactivate genes) for the analysis of the genotype-phenotype relationship, since small mutations represent the primary bases of inherited diseases.

The aim of this chapter is to describe the use of targeted mutagenesis in the understanding of the physiopathology of inherited connective tissue disorders. Specifically we are concerned with mutations in collagen genes. We will focus on the use of site-directed mutagenesis to analyze the causative effects of human-identified collagen gene mutations. Recombinant molecules were used to analyze the effects of these mutations on collagen structure, biosynthesis, posttranslational modifications and interactions with binding partners and cells. This work has considerably improved our knowledge in development and in human disorders. These results will then be compared with the limited information about the introduction of subtle targeted mutations into murine collagen genes.

2. The collagen superfamily at a glance

The 28 members of the collagen superfamily exhibit considerable complexity and diversity in structure, assembly and function. However, collagens also share common features. (i) All members are modular proteins composed of collagenous (COL) domains flanked by non collagenous (NC) domains or linker regions. (ii) They are trimeric molecules formed by the association of three identical or different α-chains, which are characterized by repetitions of the G-x-y tripeptide (with the x and y positions often occupied by proline and hydroxyproline, respectively). (Abbreviations and single-letter codes for amino acids are given in Table 1 of the chapter by Figurski *et al*.) (iii) They are able to assemble into supramolecular aggregates in the extracellular space, although this property has not been proven for all recently identified collagen members. Collagens also undergo various post-translational modifications, including proteolytic processing, fibril formation, reticulation, shedding of transmembrane collagens and production of functional domains (also called matricryptins) (Ricard-Blum and Ruggiero, 2005). The mechanisms of collagen biosynthesis are far from being completely understood. Our knowledge is primarily based on the biosynthesis of fibril-forming collagens. Triple-helix formation commonly starts at the C-terminus (C-propeptide) of the α-pro-chain and proceeds toward the N-terminus (N-propeptide) in a zipper-like fashion. Prior to and simultaneously with triple-helix formation, specific prolines and lysines are chemically modified by addition of hydroxyl group. These modifications play a pivotal role in stabilization and resistance to temperature. Completed trimeric procollagens are secreted from the cells, proteolytically processed and assemble into collagen fibrils (Ricard-Blum and Ruggiero, 2005).

Based on their structure and supramolecular organization, collagens have been divided into several subfamilies (Myllyharju and Kivirikko, 2001). They are (i) the fibril-forming collagens I, II, III, V, XI, XXIV and XVII, which share the capacity to assemble into organized fibrils; (ii) the network-forming collagens IV, VIII and X and the FACIT (Fibril-Associated Collagen with Interrupted Triple-helix collagens) collagens IX, XII, XIV, XVI, XIX, XX, XI

and XXII, which are known to mediate protein-protein interactions; (iii) the basement membrane multiplexin (multiple triple-helix domains and interruptions) collagens XV and XVIII; (iv) the transmembrane collagens, including the neuronal XXV collagen and types XIII, XVII, XXIII; and finally (v) other unconventional collagens, such as the anchoring fibrils collagen VII and the ubiquitous collagen VI, which assembles into characteristic beaded filaments (Table 1).

The length of the triple helical domains varies noticeably among different collagen types. Fibril-forming collagens consist of a long central COL domain with about 1000 amino acids (330 G-x-y tripeptide repeats), flanked by small terminal globular extensions (NC domains). After proteolytic processing of the N and C-terminal extensions, the mature molecules aggregate into highly ordered fibrils with a banded pattern observable by transmission electron microscopy. In other collagens, the COL domains are shorter and/or contain interruptions. The NC domains can represent the main part of the molecule, as for the FACIT collagen XII. Most, if not all, collagen types are recognized by specific cell receptors, such as the major ECM integrin receptors, collagen-specific discoidin domain receptors (DDR) and the transmembrane proteoglycan syndecans (Humphries *et a*l., 2006; Xian *et al.*, 2010; Leitinger *et al.*, 2007). Through various interactions with these cell receptors, collagens can induce intracellular pathways directly or indirectly and regulate cell functions, such as migration, proliferation and differentiation. Certain collagens can also bind to growth factors and control their bioavailability by acting as reservoirs. The controlled release of growth factors by proteolytic activity or expression of a splice variant that does not contain the binding site controls morphogenesis, as described for the cartilage collagen II (Zhu *et al.*, 1999).

3. A large spectrum of mutations in collagen genes causes inherited disorders

A myriad of mutations has been characterized in collagen genes (Table 1). The function of the gene product and its tissue localization are criteria that lead to a number of inherited connective tissues disorders (reviewed in Bruckner-Tuderman and Bruckner, 1998; Bateman *et al.*, 2009). Typically mutations in collagen genes are null-mutations, *i.e.*, those resulting in the translation of an α-chain that cannot assemble into a triple helix and is consequently degraded intracellularly. Null mutations reduce the overall quantity of collagen in tissue and generally cause a human disorder. Small deletions and base substitutions can lead to synthesis of a mutated α-chain that is able to form a triple helix. The molecule is secreted, but its structure is compromised for supramolecular assembly, which normally occurs in the extracellular space. *In fine* collagen gene mutations result in defective matrix assembly and organization that in turn can affect cell function (Figure 1). In cases of large multimeric molecules, such as collagens, dominant-negative mutations can be more deleterious than null mutations. However, a growing body of evidence shows that the synthesis of a large quantity of abnormal collagen molecules in cells during development can induce endoplasmic reticulum stress, with consequences ranging from cell recovery to death (Tsang *et al.*, 2010). The correlation between phenotype severity and the location of a point mutation in the gene is not clear. However, a mutation located in the coding region for the amino-terminus of the fibrillar collagen triple helix generally results in a mild phenotype, whereas a mutation in the coding region for the carboxy-terminus of the molecule is often lethal. This

observation may be related to the C- to N-terminus directional propagation of the triple helix and the role of the C-propeptides in α-chain registration and triple helix nucleation. The nature of glycine substitution in the G-x-y repeats and the neighboring amino-acid sequence may have different biochemical and clinical consequences. These consequences include (i) delay of the triple-helix formation and over-glycosylation (Raghunath *et al.*, 1994); (ii) alteration of procollagen processing (Lightfoot *et al.*, 1994); (iii) retention of unfolded abnormal proteins intracellularly, leading to ER stress; and (iv) formation of abnormal unstable trimeric molecules, leading to disrupted fibrillogenesis.

The presence of a glycine in every third position is critical for triple-helix formation, since only glycine, the smallest amino acid, fits into the center of the triple helix. The majority of dominant-negative mutations in collagen genes are due to replacements of one of the glycines in the collagenous domains of the α-chains with a larger amino acid. Glycine substitution mutations in collagen genes underlie heritable connective tissue diseases, such as osteogenesis imperfecta (OI), chondrodysplasias, certain subtypes of Ehlers-Danlos syndrome (EDS), or Alport's syndrome (reviewed in Bruckner-Tuderman and Bruckner, 1998; Bateman *et al.*, 2009). Since a non-glycine amino acid does not easily fit into the interior space of the triple helix, helix formation is distorted, thereby affecting its structure and stability and impeding fibrillogenesis. Delay in triple-helix formation can result in over-modification and may affect collagen function.

Osteogenesis imperfecta (OI), also known as brittle bone disease, is caused by mutations in genes for collagen I, the most abundant collagen in organisms. OI is characterized by fragile bones that break easily and reduced bone mass. Most OI cases are believed to be associated with glycine substitution mutations in the *COL1A1* or *COL1A2* genes. Over 200 mutations have been reported for the *COL1A1* (located on chromosome 17) and *COL1A2* (located on chromosome 7) genes, which code for the collagen I pro-α1 and pro-α2 chains, respectively. This fact may explain the wide range of clinical characteristics and degrees of severity that are seen in the disease (Kuivaniemi *et al.*, 1991, Byers and Steiner 1992, Dalgleish, 1998). Because collagen I is found in other tissues of the body, OI has non-skeletal manifestations as well. People with OI may also suffer from muscle weakness, hearing loss, fatigue, joint laxity, distensible skin, or dentinogenesis imperfecta. The fibril-forming collagen I is mostly synthesized as the $[\alpha1(I)]_2\alpha2(I)$ heterotrimer chain, though a minor form $[\alpha1(I)]_3$ is expressed in embryonic tissues. *COL1A1* and *COL2A1* are both susceptible to various mutations responsible for the production of quantitatively or qualitatively deficient fibrils. The clinical severity of OI relates to the extent of the conformational change in the collagen triple helix induced by the glycine substitution. These mutations result in altered fibrillogenesis. However, no general mechanism can be drawn from genotype/phenotype analyses.

Collagen VII, encoded by *COL7A1*, is the major component of the anchoring fibrils at the dermo-epidermal junction (Burgeson, 1993). *COL7A1* gene mutations cause dystrophic epidermolysis bullosa (DEB), a skin-blistering disorder (Bruckner-Tuderman, 1999). Approximately 200 mutations of *COL7A1* have been characterized, leading to a very high molecular heterogeneity of collagen VII defects (Dunnill *et al.*, 1996). Almost all cases of dominant DEB are caused by a glycine substitution in the triple helical region of collagen VII, and most of the mutations are unique to individual families. Some glycine substitutions in collagen VII interfere with biosynthesis of the protein in a dominant-negative manner, whereas others may lead to collagen VII retention within the rough endoplasmic reticulum.

Mutations in the *COL5A1* and *COL5A2* genes, encoding respectively the pro-α1 and pro-α2 chains of the fibril-forming collagen V, have been identified in approximately 50% of patients with a clinical diagnosis of classic Ehlers-Danlos syndrome (EDS) (Malfait *et al.*, 2010). Collagen V contains a third chain, proα3(V); but no mutation in *COL5A3* has been reported so far. Classic EDS is a heritable disorder of connective tissues characterized by skin hyperextensibility, fragile and soft skin, delayed wound healing with formation of atrophic scars, easy bruising, and generalized joint hypermobility. The majority of mutations lead to a non-functional *COL5A1* allele. One mutant *COL5A1* transcript showed a premature stop codon. A minority of mutations affect the structure of the central helical domain. In approximately one-third of patients, the disease is caused by a mutation leading to a non-functional *COL5A1* allele, resulting in collagen V haploinsufficiency. Structural mutations in *COL5A1* or *COL5A2*, resulting in the production of a functionally defective protein, account for a small proportion of patients.

Collagen V is a quantitatively minor fibril-forming collagen that co-polymerizes with collagen I to form heterotypic fibrils (Fichard *et al.*, 1995). Co-polymerisation has a critical role in the nucleation and growth of fibrils in tissues. A collagen V feature is to retain in the mature molecule a major part of the α1(V) N-propeptide which projects beyond the surface of collagen fibrils. This domain was proposed to limit heterotypic fibril growth by steric hindrance and electrostatic interactions (Linsenmayer *et al.*, 1993). Skin biopsies revealed abnormalities in fibril formation (altered diameter, contour, or shape of dermal fibrils). However, abnormalities of fibril structure affected less than 5% of fibrils (reviewed in Fichard *et al.*, 2003). Moreover, the clinical phenotype of classical EDS supports an important role of collagen V in the biomechanical integrity of the skin, tendon and ligaments, although collagen V is only a minor component of the affected tissues. Thus, collagen V may be involved in functions other than the control of fibril growth in classical EDS. A likely hypothesis is that collagen V might be involved in the physiopathology of EDS through interactions with other fibril-associated components and/or with cell receptors. Along this line, it has been shown that mutations in the genes for the collagen V-binding partners, tenascin-X (*TNXB* gene) and collagen I (*COL1A1* gene), resulted in EDS (Lindor and Bristow, 2005).

Although mutant gene products are thought to impair matrix structure and assembly that eventually alters tissue function, growing evidence links ER stress and the unfolded protein response (UPR) to the initiation and progression of a broad repertoire of connective tissue disorders, including those caused by collagen gene mutations. Some mutant chains cannot be incorporated into procollagen molecules, consequently causing protein degradation with important downstream effects. Misfolded or slowly folding collagens are retained within the endoplasmic reticulum (ER) and ultimately targeted for degradation by a mechanism initially called "protein suicide." Because connective-tissue cells typically produce large quantities of collagens, the contribution of ER stress induced by misfolded collagens in disease pathogenesis has certainly been underrated. The current knowledge on the implications of unfolded protein response and ER stress in connective tissue diseases has been recently reviewed, and readers are referred to these reviews for further reading (Boot-Handford and Briggs, 2010; Tsang *et al.*, 2010). Notably, mutations in genes encoding collagen I (*COL1A1* and *COL1A2*) (osteogenesis imperfecta), collagen II (*COL2A1*) (spondyloepiphyseal dysplasia), and collagen X (*COL10A1*) (metaphyseal

Collagen subfamily	Collagen type	Associated diseases	Gene	Modified mice	References (mouse models)
Fibril forming collagens	I	osteogenesis imperfecta	COL1A1	knock-out	Löhler et al., 1984
				single mutation	Forlino et al., 1999
			COL1A2	single mutation	Daley et al., 2010
	II	chondrodysplasias	COL2A1	knock-out	Aszodi et al., 1998
	III	vascular Ehlers-Danlos syndrome	COL3A1	knock-out	Liu et al., 1997
	V	classic Ehlers-Danlos syndrome	COL5A1	knock-out	Wenstrup et al., 2004
			COL5A2	knock-in	Andrikopoulos et al., 1995
		nd	COL5A3	knock-out	Huang et al., 2011
	XI	chondrodysplasias	COL11A1	knock-out	Wenstrup et al., 2011
			COL11A2	knock-out	Mc Guirt et al., 1999
FACIT	IX	Stickler syndrome	COL9A1	knock-out	Fässler et al., 1994
			COL9A2	functional knock-out	Hagg et al., 1997
			COL9A3		
	XII	nd	COL12A1	knock-out	Izu et al., 2011
	XIV	nd	COL14A1	knock-out	Ansorge et al., 2009
	XIX	nd	COL19A1	knock-out	Sumiyoshi et al., 2004
Multiplexins	XV	nd	COL15A1	knock-out	Eklund et al., 2001
	XVIII	Knobloch syndrome	COL18A1	knock-out	Seppinen et al., 2008
Network forming collagens	IV	nd	COL4A1	nd	
			COL4A2		
		Alport syndrome	COL4A3	knock-out	Cosgrove et al., 1996
			COL4A4	knock-out	Lu et al., 1999
			COL4A5	knock-out	Rheault et al., 2004
			COL4A6	nd	
	VIII	corneal dystrophia	COL8A1	knock-out	Hopfer et al., 2005
			COL8A2	knock-out	Hopfer et al., 2005
	X	chondrodysplasias	COL10A1	knock-out	Kwan et al., 1997
Beaded-filament forming collagen	VI	Bethlem myopathy and Ullrich congenital muscular dystrophy	COL6A1	knock-out	Bonaldo et al., 1998
			COL6A2	nd	
			COL6A3		
		nd	COL6A4	nd	
			COL6A5		
			COL6A6		
Anchoring collagens	VII	epidermolysis bullosa	COL7A1	knock-out	Heinonen et al., 1999
Transmembrane collagens	XVII	epidermolysis bullosa	COL17A1	knock-out	Tanimura et al., 2011

nd, not determined

Table 1. Collagen types, associated-diseases and mouse models.

chondrodysplasia) have been shown to induce ER dilatation in patient cells. Mutations that affect the triple helix, the C-propeptide for the fibril-forming collagens, and splice donor sites, as well as single amino-acid substitutions, were shown to cause ER stress. Recently, mutations that affected the signal peptide domain of the proα1(V)-collagen chain were shown to cause classic EDS. The signal peptides are the addresses of proteins destined for secretion. The mutant procollagen V is retained within the cell, leading to a collagen V haploinsufficiency and altered collagen fibril formation. It is probable that the signal peptide mutation also causes accumulation of the mutated protein within the ER and eventually to ER stress, as described for other collagen-related disorders (Symoens *et al.,* 2009).
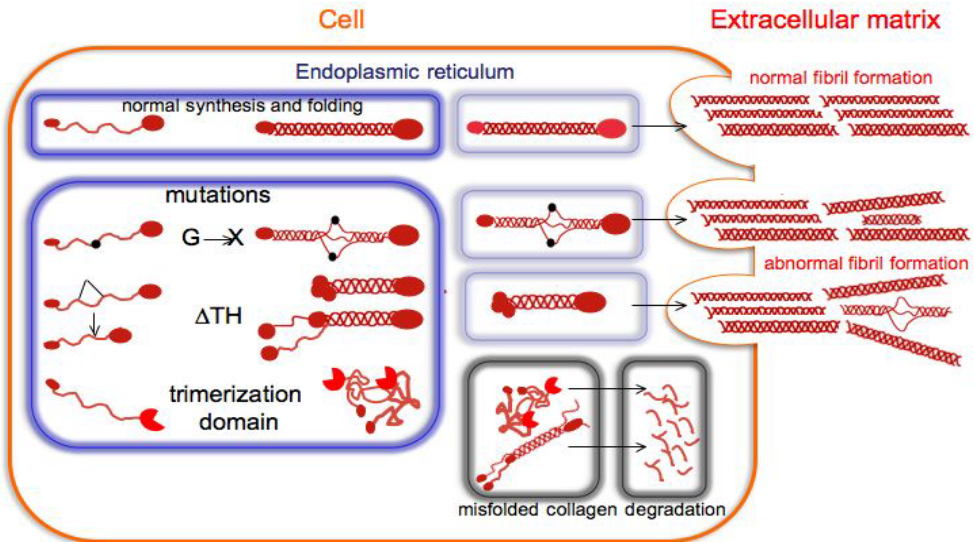
Fig. 1. Schematic diagram illustrating the biological consequences of point mutations or small deletions in collagen genes on chain synthesis, protein folding and subsequent fibril assembly in the extracellular matrix.

Mutations in the three major collagen VI genes (*COL6A1*, *COL6A2* and *COL6A3*) cause multiple muscle disorders, including the severe Ullrich congenital muscular dystrophy (UCMD) and the mild Bethlem myopathy, which is characterized by muscle weakness with striking joint laxity and progressive contractures. Three genetically distinct novel chains α4(VI), α5(VI), and α6(VI) have recently been identified; but very little is known about their molecular assembly and biosynthesis and their possible involvement in human diseases (Gara *et al*., 2011). Collagen VI biosynthesis is a complex multistep process. Monomer formation results from the heterotrimeric association of the three chains [α1(VI), α2(VI), and α3(VI)] encoded by the *COL6A1*, *COL6A2* and *COL6A3* genes. Monomers first assemble into antiparallel dimers that associate laterally to form tetramers stabilized by disulphide bonds. The tetramers associate linearly to form the unique beaded filaments, the ultimate step of collagen VI biosynthesis. Dominant and recessive autosomal mutations in *COL6A1*, *COL6A2,* and *COL6A3* primarily result in dysfunctional microfibrillar collagen VI in muscle extracellular matrix. However they also affect other connective tissues, such as skin and

tendons. Different mutations have been shown to have variable effects on protein assembly, secretion, and its ability to form a functioning extracellular network. As observed in other collagen-related diseases, glycine-substitution mutations in *COL6A1*, *COL6A2*, or *COL6A3* that disrupt the triple-helix motif constitute a frequent pathogenic mechanism. Triple-helix distortion may exert a dominant-negative effect by reducing the ability of mutated monomers to form beaded filaments. Interestingly, mitochondrial dysfunction was implicated in the pathogenesis of a myopathic phenotype. Muscles lacking collagen VI are characterized by the presence of a dilated sarcoplasmic reticulum and dysfunctional mitochondria. This condition triggers apoptosis and leads to myofiber degeneration. Recently, it was shown that the persistence of abnormal organelles and apoptosis observed in some congenital muscular dystrophies are caused by defective activation of the autophagic machinery. Autophagy has a key role in the clearance of damaged organelles and in the turnover of cell components and is thus essential for tissue homeostasis. Recently, 56 novel mutations have been described, allowing a clinical classification and revealing the complexity of genotype-phenotype relationships (Briñas *et al.*, 2010).

The paucity of evidence-based data regarding correlations of genotype and phenotype is in part due to the large spectrum of mutations reported for the collagen genes [*e.g.*, about 200 mutations for the collagen I genes responsible for OI (Dalgleish, 1998); 160 mutations in the *COL4A5* gene encoding collagen IV α5 chain responsible for Alport syndrome; 200 mutations in *COL7A1* responsible for EDB]. Things are not as simple as one gene-numerous mutations-one phenotype. Sometimes a combination of a mutation for a connective tissue disorder and a specific collagen gene mutation will result in another disease. Some patients with UCMD show clinical characteristics typical of classical disorders of connective tissue, such as EDS. Ultrastructure of skin biopsy samples from patients with UCMD showed alterations of collagen fibril morphology in skin that resemble those described in patients with EDS (Kirschner *et al.*, 2005). Recently, using the yeast two-hybrid approach, we showed a direct interaction between collagen V and collagen VI that may nicely explain the overlap of UCMD and classic EDS (Symoens *et al.*, 2011). Unexpectedly an arginine-->cysteine substitution localized at position 134 of the α1(I) collagen chain resulted in classical EDS (Nuyntick *et al.*, 2000). This finding is indicative of genetic heterogeneity in collagen-related disorders.

A powerful approach to study the biochemical consequences of mutation and the protein structure/function relationship is to engineer a specific mutation into a functional domain of the molecule. Targeted mutagenesis approaches, including the use of alanine-scanning mutagenesis techniques, have led to important insights into the effects of collagen mutations on protein structure and function. A major limitation of mutagenesis strategies to investigate collagens is the large number of collagen gene mutations to be investigated in order to have a better understanding of the molecular mechanisms of "collagenopathies." Knowledge about the impact of collagen mutations has also been hampered by the technical difficulty of introducing targeted mutations of very large collagen genes into mice.

4. Lessons from site-directed mutagenesis of recombinant collagen genes and derived fragments

Production of a recombinant collagen gene represents a powerful technique to introduce a human mutation into the gene of interest by site-directed mutagenesis. It allows one to

analyze the impact of the mutation on collagen assembly and secretion. Collagen biosynthesis is a complex multistep process that takes place in the intracellular and extracellular space and includes various post-translationnal modifications, such as prolyl- and lysyl-hydroxylation, glycosylation, trimerization, proteolytic processing, polymerization and cross-links. Because of recombinant technology, these large multimeric proteins have been produced in large amounts in almost all existing expression systems (Ruggiero and Koch, 2008). This technological breakthrough enabled researchers to analyze in detail the effects of collagen mutations on biosynthesis, molecular and cell interactions, processing and, in some cases, self-assembly. Researchers can also address the question of the correlation of genotype, protein structure and function.

Mutations occurring in collagen I genes are the most extensively studied mutations among all collagen types. A first set of experiments substituted glycine 859 of the proα1(I) chain with cysteine or arginine by site-directed mutagenesis to reproduce two mutations identified in OI patients. In order to study the expression of the mutant molecule in the presence or absence of the wild-type proα1(I) chain, the mutated constructs were transfected into normal fibroblasts to look for a dominant-negative effect in the presence of the wild-type gene or in fibroblasts isolated from Mov13 homozygous mice (referred to as Mov13 fibroblasts hereafter), whose cells carry a provirus that prevents transcription initiation of the natural proα1(I) gene (Schnieke *et al.*, 1987). In agreement with observations of collagen I in OI patients, the mutated collagens were poorly secreted from the cells and exhibited reduced thermal stability and increased sensitivity to degradation. This supported the idea that the strict preservation of the G-x-y triplets is absolutely required for proper formation of the triple helix.

The integrity of the C-propeptide is pivotal for the trimerization of all fibril-forming collagens. The C-propeptides of the proα1(I) and proα2(I) chains contain an Asn-Ile-Thr sequence.   That sequence fits a consensus sequence for the addition of N-linked oligosaccharides. To analyze the role of this post-translational modification, the asparagine residue of the proα1(I) chain was changed to glycine by site-directed mutagenesis. The expression of the corresponding molecule was analyzed in transfected normal and Mov13 fibroblasts (Lamandé and Bateman, 1995). The mutation did not impair heterotrimeric assembly and secretion of hybrid procollagen I into the extracelllular space. Only a slight effect on C-proteinase cleavage efficiency was observed with the unglycosylated molecule. To circumvent the difficulty of producing a large repertoire of full-length mutated collagens I in order to undertake a genotype/phenotype analysis, a recombinant trimeric mini-collagen I was recently expressed in an *Escherichia coli* system. Recombinant mini-collagens can be obtained by fusing the sequence encoding a fragment of the proα1(I) chain triple-helix to the sequence encoding the C-terminal domain (called "foldon") of the bacteriophage T4 fibritin, which is capable of trimerization (Xu *et al.*, 2008). Two mutations (G901S and G913S), corresponding to mild and severe types of OI, respectively, were introduced into the recombinant mini-collagen I. Biophysical measurements and protease cleavage analysis revealed that the G913S mutant chain resulted in the formation of an unstable collagen I triple helix by disrupting salt bridges important for maintaining the chains in a triple-helix conformation (Yang *et al.*, 1997; Xu *et al.*, 2008). A very recent study utilized a recombinant bacterial collagen to develop a mutagenesis scheme in which a glycine residue within the triple-helix sequence is substituted with arginine or serine. The purpose was to analyze the

positional effect of glycine mutations on triple-helix formation and stability (Cheng *et al.*, 2011). Interestingly, all glycine mutations provoked a significant delay in the triple-helix formation. However, a more severe defect was observed when the mutation was located near the trimerization domain of the triple-helix where folding is initiated.

COL7A1 mutations cause dystrophic epidermolysis bullosa (DEB), a skin blistering disorder. Woodley and collaborators (2008) have used site-directed mutagenesis to elucidate the effect of human mutations on the function of collagen VII, which is the major component of the epidermal anchoring fibrils. To undertake a comprehensive analysis of the impact of human mutations in the formation, folding and stability of collagen VII and, particularly relevant to the DEB phenotype, its effect on cell attachment and migration, four distinct substitutions occurring in collagen VII (G2049E, R2063W, G2569R, and G2575R) were introduced using *COL7A1* cDNA. The authors demonstrated that the G2049E and R2063W mutants caused local destabilization of the triple helix and reduced the capability of collagen VII to elicit cell adhesion and migration. The G2569R and G2575R mutants interfered with triple-helix formation and stability. Alterations of protein stability and/or cell attachment to collagen VII mutants help explain the fragility of the dermal-epidermal junction observed in DEB patients. Naturally occurring *COL7A1* mutations were investigated in a separate study (Hammami-Hauasli *et al.*, 1998). As commonly described for glycine-substitution mutants of collagens, the authors showed that three glycine substitutions located in the same triple-helix portion affected folding, stability and secretion of procollagen VII in a dominant-negative manner. However, the glycine substitution G1519D located in another segment of the triple helix had no effect on procollagen VII secretion or its ability to anchor fibril assembly. These data showed that the biological impact of glycine substitutions can depend on their position within the triple helix, as shown for collagen I (Cheng *et al.*, 2011).

Human collagen IV mutations, thought to affect the biosynthesis of this basement membrane collagen, were extensively investigated. These mutations were known to cause Alport syndrome, a severe renal disease leading eventually to kidney failure. Collagen IV chains, α1(IV)-α6(IV), are encoded by 6 genes, *COL4A1-COL4A6*, respectively. Although mutations have also been identified in *COL4A3 and COL4A4*, about 30% of known missense mutations occur in the *COL4A5* gene, which encodes the human α5(IV) chain. Most of them are glycine substitutions. One glycine-substitution mutation in *COL4A5* could prevent correct α-chain folding or/and the association with other α-chains to form a stable triple helix. To address this question, the authors took advantage of the bacterial system. A DNA encoding a 22-kDa recombinant domain of the α5(IV) triple helix in its wild-type form or harboring the G1015V or G1030S mutations was expressed in *E. coli* (Wang *et al.*, 2004). The recombinant wild-type and mutant proteins were purified and assayed for changes in triple-helix assembly and stability by circular dichroism. The two different glycine-substitution mutants displayed different defects in the secondary structures of their protein products that matched with the severity of the patient phenotypes. However, the use of a bacterial system to analyze the effects of specific human mutations on mini-collagen assembly and stability presents several disadvantages. Because collagens are large multimeric proteins, full-length molecules cannot be produced in a bacterial host. Most importantly, the bacterial system is limited. Not all post-translational modifications needed for the triple-helix formation and stability, such as hydroxylation, glycosylation, and disulfide-bond formation, are present in bacteria. A few years later, the bacterial limitations were bypassed by the

development of the production of full-length recombinant collagen molecules in mammalian cells (Fichard *et al.*, 1997; Ruggiero and Koch, 2008). No less than eighteen human mutations (11 substitutions and 7 deletions) were introduced into the sequence encoding the trimerization NC1 domain of the α5(IV) chain gene. The constructs were transfected into cells together with constructs containing the wild-type sequences of α3(IV) and α4(IV) chains to analyze the impact of the mutations in the NC1 domain on the formation of the α3α4α5 collagen IV heterotrimer. Twelve out of 15 mutant chains did loose their capacity to assemble into heterotrimeric molecules. The three remaining mutants formed heterotrimers, but the mutations prevented their secretion into the extracellular space (Kobayashi *et al.*, 2008). The authors nicely demonstrated, using site-directed mutagenesis, that amino acid substitutions in the α5(IV) NC1 trimerization domain are specifically responsible for impairment of collagen IV heterotrimer assembly. This defect may be a main molecular mechanism for the pathogenesis of Alport syndrome. Interestingly, an interactome (a map of known and predicted molecular interactions, as well as phenotypic and structural landmarks) of collagen IV was recently constructed to identify functional and disease-associated domains and genotype-phenotype relationships (Parkin *et al.*, 2011). Construction of such interactomes will greatly improve our capacity to integrate all data from different site-directed mutagenesis experiments. This advance will greatly help our understanding of the molecular mechanisms underlying "collagenopathies"; and, consequently, it may lead to the development of specific treatments.

Collagens undergo a great variety of proteolytic modifications. The fate and functions of the released fragments derived from collagens are still under intensive investigation, but the consequences of mutations in the coding regions for the cleavage sites on collagen structure, self-assembly and function have not been investigated in detail. A large repertoire of proteinases is responsible for these processing interactions. Included among such enzymes are the ADAMTS (a disintegrin and metalloprotease with thrombospondin motifs) and the BMP-1/tolloid families of metalloproteinases and more recently the furin-like proprotein convertases (Ricard-Blum and Ruggiero, 2005). To investigate collagen processing, fastidious extraction and purification steps were often necessary to obtain limited amounts of unprocessed proteins and enzymes with full activity in order to undertake *in vitro* enzymatic assays. To circumvent this problem, we recently described a new cell system allowing a rapid and straightforward analysis of processing interactions. Our system relies on the use of site-directed mutagenesis. This strategy was particularly instrumental in analyzing the complex procollagen V processing during maturation. We showed it to be unique among the fibril-forming collagens (Bonod-Bidaud *et al.*, 2007). Collagen V is a minor fibrillar collagen that can be distinguished from the others by its capacity to control fibrillogenesis (Fichard *et al.*, 1995). In addition this molecule undergoes a particular form of processing; and it is involved in fundamental processes, such as development and human connective tissues disorders. The proα1(V) N-terminus can be processed by the procollagen proteinases ADAMTS-2 and BMP-1 (Colige *et al.*, 2005; Bonod-Bidaud *et al.*, 2007), whereas the C-propeptide can be cleaved by furin and BMP-1 (Kessler *et al.*, 2001). The proα1(V) C-propeptide furin cleavage site, which occurs immediately downstream of the recognition sequence RTRR, was double-mutated to alanine residues (R1584A/R1585A) to abolish furin cleavage. All constructs were introduced into cells, along with a BMP-1-expressing construct; and the cleavage products were directly analyzed in conditioned medium of the

transiently transfected cells. We were able to show that BMP-1 is capable of processing the α1(V) C-propeptide in absence of furin activity (Bonod-Bidaud *et al.*, 2007). In the same way, the determinant for α1(V) N-propeptide processing by BMP-1 activity was identified by introducing in the coding region for the cleavage site (S254/Q255-D256) three single mutations (S254A, Q255A and D256A), two double mutations (S254A/Q255N and Q255A/D256A) and one triple mutation (S254A/Q255A/D256A). The data highlighted the unexpected importance of the aspartic acid in the P2′ position of the BMP-1 cleavage site (Bonod-Bidaud *et al.*, 2007). Processing, proteolytic release of functional domains and shedding of collagens are involved in fundamental processes. It is likely that substitutions located in the proteolytic cleavage sites may represent a molecular cause of connective tissues disorders. A reported mutation in the α1(V) N-propeptide in one patient with classic EDS resulted in a protein product missing the sequence of exon 5 that encompasses the BMP-1 cleavage site. The abnormal-sized N-propeptide present in the mutated collagen V caused dramatic alterations in fibril structure (Takahara *et al.*, 2002).

5. Lessons from site-directed mutagenesis in mice

In vitro studies are useful and necessary approaches to understand the mechanisms of collagen biosynthesis and to establish structure-function relationship. However, they do not always reflect the normal and pathological *in vivo* situations. Genetically modified mice appear to be a powerful technique to better understand the physiopathology of connective tissue disorders. Several different genetically modified mice have been created during the last 10 years (reviewed in Aszódi *et al.*, 2006). This clearly opened doors to better understand collagen function in developing tissues and provide reliable mouse models for inherited collagen diseases. Along this line, a targeted disruption of *Col4a3* gene led to renal failure and eventually to the death of mice at 3-4 months of age (Cosgrove *et al.*, 1996; Miner and Sanes, 1996). This result is consistent with defects described for Alport disease.

In most cases, the gene of interest was disrupted and knock-out mice were preferably generated. Few transgenic mice harbouring point mutations or small deletion in collagen genes have been generated (Table 1). Naturally occurring mutations in mice disrupting collagen genes have also been identified and characterized. The *oim* mice present a spontaneously acquired deletion in the *Col1a2* gene that leads to an accumulation of $[\alpha(I)]_3$ collagen homotrimer in the extracellular matrix. These mice develop a phenotype similar to moderate OI in humans, providing a good model for this collagen disorder (Chipman *et al.*, 1993). It was shown that homozygous Mov13 embryos harboring an inactivated proα1(I) chain (due to the insertion of the Moloney murine leukaemia virus into the first exon of the *Col1a1* gene) died in utero around day 12 because of vascular failure (Löhler *et al.*, 1984). However, in 1999 Forlino *et al.* developed the first knock-in mouse model for human OI by introducing a G349C mutation into the *Col1a1* gene. Along this line, a knock-in mouse model for OI, harboring a point mutation (G610C) in *Col1a2* was recently created (Daley *et al.*, 2010). These mice had reduced body mass and bone strength and exhibited bone fracture susceptibility consistent with the clinical features of human OI. Thus, the G610C knock-in mouse represents a novel model for the study of OI pathogenesis and also for testing potential therapies for OI.

Another example concerns collagen V deficiency/dysfunction, which is responsible for Ehlers-Danlos syndrome (EDS). In the absence of the *Col5a1* gene, the mice died at the onset of organogenesis at approximately embryonic day 10 (Wenstrup *et al*., 2004). Interestingly, a targeted deletion in the *Col5a2* gene, encoding the proα2(V) chain, recapitulated many of the clinical, biomechanical, morphologic, and biochemical features of the classical EDS. The deletion removes the sequence encoding the N-telopeptide (*pN*), a 20-residue region that confers flexibility to the N-terminal part of the molecule (Andrikopoulos *et al*., 1995). A detailed study of the skin at the morphological, histological, ultrastructural and biochemical levels indicated that the *Col5a2* deletion impairs assembly and/or secretion of the $[\alpha1(V)]_2\alpha2(V)$ heterotrimer. Consequently, the $[\alpha1(V)]_3$ homotrimer, and not the $[\alpha1(V)]_2\alpha2(V)$ heterotrimer, is the predominant species deposited into the matrix, which in turn severely impaired extracellular matrix organization (Chanut-Delalande *et al*., 2004). These data underscored the importance of the collagen V $[\alpha1(V)]_2\alpha2(V)$ heterotrimer in dermal fibrillogenesis and can explain defects observed in the dermis of EDS patients.

6. Concluding remarks

Site-directed mutagenesis has been extensively used in collagen engineering and has shed light on collagen structure, expression, folding, secretion, interactions and self-assembly in the extracellular space. It also opened the way for the analysis of specific functional domains. It allowed the study of the wide variety of collagen types, including those expressed in trace amounts in tissues but nevertheless display pivotal functions. While it is true that site-directed mutagenesis has yielded important information on the functional consequences of a range of collagen mutations responsible for human diseases, only few studies have approached the consequences of collagen gene mutations on cell adaptation to ER stress. Collagen gene mutations affect protein synthesis, folding and secretion imbalance, which eventually induces ER stress. *In vitro* studies have been done on transfected cells, in which expression and trafficking of mutant collagen can be easily manipulated and analysed at the cellular level. The effects of gene manipulation can be studied *in vivo* using mice. The effect of collagen gene mutations on induction of an ER stress response could be straightforwardly addressed in the near future. It may be a key factor in pathogenesis (Boot-Handford and Briggs, 2010).

Mouse models are particularly useful for analysing the biological significance of collagens in pathological situations. Knock-out mice often lead to embryonic lethality, which hampers in-depth analysis of the phenotype. A few knock-in mice have been created with subtle mutations or small deletions that reproduce human mutations. The major reason for the paucity of knock-in mice is certainly that collagen genes are very large. Thus, they are difficult to manipulate. The introduction of a small deletion or a single point mutation in murine collagen genes still represents a considerable challenge. Nevertheless, the few examples of knock-in mouse lines tend to prove that mouse models can bring new information about *in vivo* consequences of collagen dysfunction that cannot be predicted by *in vitro* approaches. Knock-in mice are also indispensable models for assessing the effects of subtle mutations on tissue function, development, and aging. They are also valuable for developing specific gene therapy approaches to combat collagen-related disorders. The combination of site-directed mutagenesis in transfected cells and knock-in approaches in mice to address the impact of specific mutations will enable us to identify mechanisms

underlying the vast repertoire of collagen-related diseases. The implications may lead to the development of a specific therapy.

7. References

Andrikopoulos, K., Liu, X., Keene, D.R., Jaenisch, R. & Ramirez, F. (1995). Targeted mutation in the *col5a2* gene reveals a regulatory role for type V collagen during matrix assembly. *Nat Genet.* 9(1):31-36.

Aszodi, A., Legate, K.R., Nakchbandi, I. & Fässler, R. (2006). What Mouse Mutants Teach Us About Extracellular Matrix Function. *Annu. Rev. Cell Dev. Biol.* 22:591–621.

Bateman, J.F., Boot-Handford, R.P. & Lamandé, S.R. (2009). Genetic diseases of connective tissues: cellular and extracellular effects of ECM mutations. *Nat Rev Genet.* 10(3):173-83.

Bonaldo, P., Braghetta, P., Zanetti, M., Piccolo, S., Volpin, D. & Bressan, G.M. (1998). Collagen VI deficiency induces early onset myopathy in the mouse: an animal model for Bethlem myopathy. *Hum Mol Genet.* 7(13):2135-2140.

Bonod-Bidaud, C., Beraud, M., Vaganay, E., Delacoux, F., Font, B., Hulmes, D.J. & Ruggiero, F. (2007). Enzymatic cleavage specificity of the proalpha1(V) chain processing analysed by site-directed mutagenesis. *Biochem J.* 405(2):299-306.

Boot-Handford, R.P. & Briggs, M.D. (2010). The unfolded protein response and its relevance to connective tissue diseases. *Cell Tissue Res.* 339(1):197-211.

Briñas, L., Richard, P., Quijano-Roy, S., Gartioux, C., Ledeuil, C., Lacène, E., Makri, S., Ferreiro, A., Maugenre, S., Topaloglu, H., Haliloglu, G., Pénisson-Besnier, I., Jeannet, P.Y., Merlini, L., Navarro, C., Toutain, A., Chaigne, D., Desguerre, I., de Die-Smulders, C., Dunand, M., Echenne, B., Eymard, B., Kuntzer, T., Maincent, K., Mayer, M., Plessis, G., Rivier, F., Roelens, F., Stojkovic, T., Taratuto, A.L., Lubieniecki, F., Monges, S., Tranchant, C., Viollet, L., Romero, N.B., Estournet, B., Guicheney, P., Allamand, V. (2010). Early onset collagen VI myopathies: Genetic and clinical correlations. *Ann Neurol.* 68(4):511-520.

Bruckner-Tuderman, L. & Bruckner, P. (1998). Genetic diseases of the extracellular matrix: more than just connective tissue disorders. *J Mol Med.* 76(3-4):226–237.

Bruckner-Tuderman, L., Höpfner, B. & Hammami-Hauasli, N. (1999). Biology of anchoring fibrils: lessons from dystrophic epidermolysis bullosa. *Matrix Biol.* 18(1):43-54.

Burgeson, R.E. (1993). Type VII collagen, anchoring fibrils, and epidermolysis bullosa. *J Invest Dermatol.* 101(3):252-255.

Byers, P.H. & Steiner, R.D. (1992). Osteogenesis imperfecta. *Annu Rev Med.* 43:269-282.

Chanut-Delalande, H., Bonod-Bidaud, C., Cogne, C., Malbouyres, M., Ramirez, F., Fichard, A. & Ruggiero, F. (2004). Development of a functional skin matrix requires deposition of collagen V heterotrimers. *Mol. Cell. Biol.,* 24(13):6049-6057

Cheng, H., Rashid, S., Yu, Z., Yoshizumi, A., Hwang, E. & Brodsky, B. (2011). Location of glycine mutations within a bacterial collagen protein affects degree of disruption of triple-helix folding and conformation. J Biol Chem. 286(3):2041-2046.

Chipman, S.D., Sweet, H.O., McBride, D.J. Jr, Davisson, M.T., Marks, S.C. Jr, Shuldiner, A.R., Wenstrup, R.J., Rowe, D.W. & Shapiro, J.R. (1993). Defective pro alpha 2(I) collagen synthesis in a recessive mutation in mice: a model of human osteogenesis imperfecta. *Proc Natl Acad Sci U S A.* 90(5):1701-1705.

Colige, A., Ruggiero, F., Vandenberghe, I., Dubail, J., Kesteloot, F., Van Beeumen, J., Beschin, A., Brys, L., Lapière, C.M. & Nusgens B. (2005). Domains and maturation processes that regulate the activity of ADAMTS-2, a metalloproteinase cleaving the aminopropeptide of fibrillar procollagens types I-III and V. *J Biol Chem.* 280(41):34397-34408.

Cosgrove, D., Meehan, D.T., Grunkemeyer, J.A., Kornak, J.M., Sayers, R., Hunter, W.J. & Samuelson, G.C. (1996).Collagen COL4A3 knockout: a mouse model for autosomal Alport syndrome. *Genes Dev.* 10(23):2981-92.

Daley, E., Streeten, E.A., Sorkin, J.D., Kuznetsova, N., Shapses, S.A., Carleton, S.M., Shuldiner, A.R., Marini, J.C., Phillips, C.L., Goldstein, S.A., Leikin, S. & McBride D.J. Jr. (2010). Variable bone fragility associated with an Amish COL1A2 variant and a knock-in mouse model. *J Bone Miner Res.* 25(2):247-261.

Dalgleish, R. (1998). The Human Collagen Mutation Database 1998. Nucleic Acids Res. 26(1):253-255.

Dunnill, M.G., McGrath, J.A., Richards, A.J., Christiano, A.M., Uitto, J., Pope, F.M. & Eady, R.A. (1996). Clinicopathological correlations of compound heterozygous COL7A1 mutations in recessive dystrophic epidermolysis bullosa. *J Invest Dermatol*. 107(2):171-177.

Eklund, L., Piuhola, J., Komulainen, J., Sormunen, R., Ongvarrasopone, C., Fässler, R., Muona, A., Ilves, M., Ruskoaho, H., Takala, T.E. & Pihlajaniemi, T. (2001). Lack of type XV collagen causes a skeletal myopathy and cardiovascular defects in mice. *Proc Natl Acad Sci U S A*. 98(3):1194-1199.

Fässler, R., Schnegelsberg, P.N., Dausman, J., Shinya, T., Muragaki, Y., McCarthy, M.T., Olsen, B.R. & Jaenisch, R. (1994). Mice lacking alpha 1 (IX) collagen develop noninflammatory degenerative joint disease. *Proc Natl Acad Sci U S A*. 91(11):5070-5074.

Fichard, A., Kleman, J.P. & Ruggiero F. (1995). Another look at collagen V and XI molecules. *Matrix Biol.* 14(7):515-531.

Fichard, A., Tillet, E., Delacoux, F., Garrone, R. & Ruggiero, F. (1997). Human recombinant alpha1(V) collagen chain. Homotrimeric assembly and subsequent processing. *J Biol Chem.* 272(48):30083-7.

Fichard, A., Chanut-Delalande, H. & Ruggiero, F. (2003). The Ehlers-Danlos syndrome: the extracellular matrix scaffold in question. *Med/Sci* (Paris). 19(4):443-452.

Forlino, A., Porter, F.D., Lee, E.J., Westphal, H. & Marini, J.C. (1999).Use of the Cre/lox recombination system to develop a non-lethal knock-in murine model for osteogenesis imperfecta with an alpha1(I) G349C substitution. Variability in phenotype in BrtlIV mice. *J Biol Chem.* 274(53):37923-37931.

Gara, S.K., Grumati, P., Squarzoni, S., Sabatelli, P., Urciuolo, A., Bonaldo, P., Paulsson, M. & Wagener, R. (2011). Differential and restricted expression of novel collagen VI chains in mouse. *Matrix Biol*. 30(4):248-257.

Hagg, R., Hedbom, E., Möllers, U., Aszódi, A., Fässler, R. & Bruckner, P. (1997). Absence of the alpha1(IX) chain leads to a functional knock-out of the entire collagen IX protein in mice. J Biol Chem. 272(33):20650-20654.

Hammami-Hauasli, N., Schumann, H., Raghunath, M., Kilgus, O., Lüthi, U., Luger, T. & Bruckner-Tuderman, L. (1998). Some, but not all, glycine substitution mutations in

COL7A1 result in intracellular accumulation of collagen VII, loss of anchoring fibrils, and skin blistering. *J Biol Chem*. 273(30):19228-19234.

Heinonen, S., Männikkö, M., Klement, J.F., Whitaker-Menezes, D., Murphy, G.F. & Uitto, J. (1999). Targeted inactivation of the type VII collagen gene (Col7 1) in mice results in severe blistering phenotype: a model for recessive dystrophic epidermolysis bullosa. *J Cell Sci*. 112 (Pt 21):3641-3648.

Humphries, J.D., Byron, A. & Humphries, M.J. (2006). Integrin ligands at a glance. *J Cell Sci*. 119(Pt 19):3901-3903.

Huang, G., Ge, G., Wang, D., Gopalakrishnan, B., Butz, D.H., Colman, R.J., Nagy, A. & Greenspan, D.S. (2011). α3(V) collagen is critical for glucose homeostasis in mice due to effects in pancreatic islets and peripheral tissues. *J Clin Invest*. 121(2):769-783.

Hopfer, U., Fukai, N., Hopfer, H., Wolf, G., Joyce, N., Li, E. & Olsen, B.R. (2005). Targeted disruption of Col8a1 and Col8a2 genes in mice leads to anterior segment abnormalities in the eye. *FASEB J*. 19(10):1232-1244.

Izu, Y., Sun, M., Zwolanek, D., Veit, G., Williams, V., Cha, B., Jepsen, K.J., Koch, M. & Birk D.E. (2011). Type XII collagen regulates osteoblast polarity and communication during bone formation. *J Cell Biol*. 193(6):1115-1130.

Kessler, E., Fichard, A., Chanut-Delalande, H., Brusel, M. & Ruggiero, F. (2001). Bone morphogenetic protein-1 (BMP-1) mediates C-terminal processing of procollagen V homotrimer. *J Biol Chem*. 276(29):27051-27057.

Kirschner, J., Hausser, I., Zou, Y., Schreiber, G., Christen, H.J., Brown, S.C., Anton-Lamprecht, I., Muntoni, F., Hanefeld, F. & Bönnemann, C.G. (2005). Ullrich congenital muscular dystrophy: connective tissue abnormalities in the skin support overlap with Ehlers-Danlos syndromes. *Am J Med Genet A*. 132A(3):296-301.

Kobayashi, T., Kakihara, T. & Uchiyama, M. (2008). Mutational analysis of type IV collagen alpha5 chain, with respect to heterotrimer formation. *Biochem Biophys Res Commun*. 366(1):60-65.

Kuivaniemi, H., Tromp, G. & Prockop, D.J. (1991). Mutations in collagen genes: causes of rare and some common diseases in humans. *FASEB J*. 5(7):2052-2060.

Kwan, K.M., Pang, M.K., Zhou, S., Cowan, S.K., Kong, R.Y., Pfordte, T., Olsen, B.R., Sillence, D.O., Tam, P.P. & Cheah, K.S. (1997). Abnormal compartmentalization of cartilage matrix components in mice lacking collagen X: implications for function. *J Cell Biol*. 136(2):459-471.

Lamandé, S.R & Bateman, J.F. (1995). The type I collagen pro alpha 1(I) COOH-terminal propeptide N-linked oligosaccharide. Functional analysis by site-directed mutagenesis. *J Biol Chem*. 270(30):17858-17865.

Leitinger, B. & Hohenester, E. (2007). Mammalian collagen receptors. *Matrix Biol*. 26(3):146-155.

Lightfoot, S.J., Atkinson, M.S., Murphy, G., Byers, P.H. & Kadler, K.E.(1994). Substitution of serine for glycine 883 in the triple helix of the pro alpha 1 (I) chain of type I procollagen produces osteogenesis imperfecta type IV and introduces a structural change in the triple helix that does not alter cleavage of the molecule by procollagen N-proteinase. *J Biol Chem*. 269(48):30352-30357.

Lindor, N.M. & Bristow, J. (2005). Tenascin-X deficiency in autosomal recessive Ehlers-Danlos syndrome. *Am J Med Genet A*. 135(1):75-80.

Linsenmayer, T.F., Gibney, E., Igoe, F., Gordon, M.K., Fitch, J.M., Fessler, L.I. & Birk, D.E. (1993). Type V collagen: molecular structure and fibrillar organization of the chicken alpha 1(V) NH2-terminal domain, a putative regulator of corneal fibrillogenesis. *J Cell Biol.* 121(5):1181-1189.

Liu, X., Wu, H., Byrne, M., Krane, S. & Jaenisch, R. (1997). Type III collagen is crucial for collagen I fibrillogenesis and for normal cardiovascular development. *Proc Natl Acad Sci U S A.* 94(5):1852-1856.

Löhler, J., Timpl, R. & Jaenisch, R. (1984). Embryonic lethal mutation in mouse collagen I gene causes rupture of blood vessels and is associated with erythropoietic and mesenchymal cell death. *Cell.* 38(2):597-607.

Lu, W., Phillips, C.L., Killen, P.D., Hlaing, T., Harrison, W.R., Elder, F.F., Miner, J.H., Overbeek, P.A. & Meisler, M.H. (1999). Insertional mutation of the collagen genes Col4a3 and Col4a4 in a mouse model of Alport syndrome. *Genomics.* 61(2):113-124.

Malfait, F., Wenstrup, R.J. & De Paepe, A. (2010). Clinical and genetic aspects of Ehlers-Danlos syndrome, classic type. Genet Med. 12(10):597-605.

Miner, J.H., Sanes, J.R. (1996). Molecular and functional defects in kidneys of mice lacking collagen alpha 3(IV): implications for Alport syndrome. *J Cell Biol.* 135(5):1403-1413.

Myllyharju, J. & Kivirikko, K.I. (2001). Collagens and collagen-related diseases. *Ann Med.* 33(1):7-21.

Nuytinck, L., Freund, M., Lagae, L., Pierard, G.E., Hermanns-Le, T. & De Paepe, A. (2000). Classical Ehlers-Danlos syndrome caused by a mutation in type I collagen. *Am J Hum Genet.* 66(4):1398-1402.

Parkin, J.D., San Antonio, J.D., Pedchenko, V., Hudson, B., Jensen, S.T. & Savige, J. (2011). Mapping structural landmarks, ligand binding sites, and missense mutations to the collagen IV heterotrimers predicts major functional domains, novel interactions, and variation in phenotypes in inherited diseases affecting basement membranes. *Hum Mutat.* 32(2):127-143.

Raghunath, M., Bruckner, P & Steinmann, B. (1994). Delayed triple helix formation of mutant collagen from patients with osteogenesis imperfecta. *J Mol Biol.* 236(3):940-949.

Rheault, M.N., Kren, S.M., Thielen, B.K., Mesa, H.A., Crosson, J.T., Thomas, W., Sado, Y., Kashtan, C.E. & Segal, Y. (2004). Mouse model of X-linked Alport syndrome. *J Am Soc Nephrol.* 15(6):1466-1474.

Ricard-Blum, S. & Ruggiero, F. (2005). The collagen superfamily: from the extracellular matrix to the cell membrane. *Pathol Biol.* 53(7):430-442.

Ruggiero, F. & Koch, M. (2008). Making recombinant extracellular matrix proteins. *Methods.* 45(1):75-85.

Schnieke, A., Dziadek, M., Bateman, J., Mascara, T., Harbers, K., Gelinas, R. & Jaenisch, R. (1987). Introduction of the human pro alpha 1(I) collagen gene into pro alpha 1(I)-deficient Mov-13 mouse cells leads to formation of functional mouse-human hybrid type I collagen. *Proc Natl Acad Sci U S A.* 84(3):764-768.

Seppinen, L., Sormunen, R., Soini, Y., Elamaa, H., Heljasvaara, R. & Pihlajaniemi, T. (2008). Lack of collagen XVIII accelerates cutaneous wound healing, while overexpression of its endostatin domain leads to delayed healing *Matrix Biol.* 27(6):535-546.

Sumiyoshi, H., Mor, N., Lee, S.Y., Doty, S., Henderson, S., Tanaka, S., Yoshioka, H., Rattan, S. & Ramirez, F. (2004). Esophageal muscle physiology and morphogenesis require

assembly of a collagen XIX-rich basement membrane zone. *J Cell Biol*. 166(4):591-600.

Symoens, S., Malfait, F., Renard, M., André, J., Hausser, I., Loeys, B., Coucke, P. & De Paepe, A. (2009). COL5A1 signal peptide mutations interfere with protein secretion and cause classic Ehlers-Danlos syndrome. *Hum Mutat*. 30(2):E395-403.

Symoens, S., Renard, M., Bonod-Bidaud, C., Syx, D., Vaganay, E., Malfait, F., Ricard-Blum, S., Kessler, E., Van Laer, L., Coucke, P., Ruggiero, F. & De Paepe, A. (2011). Identification of binding partners interacting with the 1-N-propeptide of type V collagen. *Biochem J*. 433(2):371-381.

Takahara, K., Schwarze, U., Imamura, Y., Hoffman, G.G., Toriello, H., Smith, L.T., Byers, P.H. & Greenspan, D.S. (2002). Order of intron removal influences multiple splice outcomes, including a two-exon skip, in a COL5A1 acceptor-site mutation that results in abnormal pro-alpha1(V) N-propeptides and Ehlers-Danlos syndrome type I. Am J Hum Genet. 71(3):451-465.

Tanimura, S., Tadokoro, Y., Inomata, K., Binh, N.T., Nishie, W., Yamazaki, S., Nakauchi, H., Tanaka, Y., McMillan, J.R., Sawamura, D., Yancey, K., Shimizu, H. & Nishimura, E.K. (2011). Hair follicle stem cells provide a functional niche for melanocyte stem cells. *Cell Stem Cell*. 8(2):177-187.

Tsang, K.Y., Chan, D., Bateman, J.F. & Cheah, K.S. (2010). In vivo cellular adaptation to ER stress: survival strategies with double-edged consequences. *J Cell Sci*. 123(Pt 13):2145-54.

Wang, Y.F., Ding, J., Wang, F. & Bu, D.F. (2004). Effect of glycine substitutions on alpha5(IV) chain structure and structure-phenotype correlations in Alport syndrome. *Biochem Biophys Res Commun*. 316(4):1143-1149.

Wenstrup, R.J., Florer, J.B., Brunskill, E.W., Bell, S.M., Chervoneva, I. & Birk, D.E. (2004). Type V collagen controls the initiation of collagen fibril assembly. *J Biol Chem*. 279(51):53331-53337.

Wenstrup, R.J., Smith, S.M., Florer, J.B., Zhang, G., Beason, D.P., Seegmiller, R.E., Soslowsky, L.J. & Birk, D.E. (2011). Regulation of collagen fibril nucleation and initial fibril assembly involves coordinate interactions with collagens V and XI in developing tendon. *J Biol Chem*. 286(23):20455-204565.

Woodley, D.T., Hou, Y., Martin, S., Li, W. & Chen, M. (2008). Characterization of molecular mechanisms underlying mutations in dystrophic epidermolysis bullosa using site-directed mutagenesis. *J Biol Chem*. 283(26):17838-17845.

Xian, X., Gopa, S. & Couchman, J.R. (2010). Syndecans as receptors and organizers of the extracellular matrix. *Cell Tissue Res*. 339(1):31-46.

Xu, K., Nowak, I., Kirchner, M. & Xu, Y. (2008). Recombinant collagen studies link the severe conformational changes induced by osteogenesis imperfecta mutations to the disruption of a set of interchain salt bridges. *J Biol Chem*. 283(49):34337-34344.

Yang, W., Battineni, M.L. & Brodsky B. (1997). Amino acid sequence environment modulates the disruption by osteogenesis imperfecta glycine substitutions in collagen-like peptides. *Biochemistry*. 36(23):6930-6935.

Zhu, Y., Oganesian, A., Keene, D.R. & Sandell, L.J. (1999). Type IIA procollagen containing the cysteine-rich amino propeptide is deposited in the extracellular matrix of prechondrogenic tissue and binds to TGF-beta1 and BMP-2. *J Cell Biol*. 144(5):1069-1080.

Molecular Genetics in Applied Research

Biological Activity of Insecticidal Toxins: Structural Basis, Site-Directed Mutagenesis and Perspectives

Silvio Alejandro López-Pazos[1,3] and Jairo Cerón[2]
[1]Facultad de Ciencias de la Salud, Universidad Colegio Mayor de Cundinamarca,
[2]Instituto de Biotecnología, Universidad Nacional de Colombia, Santafé de Bogotá DC,
[3]Biology & Rural Ecology research group, Corporación Ramsar Guamuez,
El Encano (Pasto-Nariño)
[1,2,3]Colombia

1. Introduction

Insect pests destroy about 18% of crop production each year and transmit disease agents (Oerke & Dehn, 2004). Beetles (order Coleoptera) are the largest and most diverse group of eukaryotes. They contain species of harvest pests that produce major losses around the world (Wang et al., 2007). Some examples of coleopteran pests follow: *Dectes texanus* [Coleoptera (order): Cerambycidae (family)], attacks soybeans; *Tribolium castaneum* (Coleoptera: Tenebrionidae), a biological problem of stored products; *Hypothenemus hampei* (Coleoptera: Scolytidae), an entomological problem of coffee crops; and *Premnotrypes vorax* (Coleoptera: Curculionidae), a potato pest in South America (Abdelghany et al., 2010; Tindall et al., 2010; López-Pazos et al., 2009b; Pai & Bernasconi, 2008; Damon, 2000). Lepidopteran species constitute an important group of harmful harvest pests that affect commercial agriculture. Among them are the following: the cotton bollworms, *Helicoverpa armigera* and *H. zea* (both Lepidoptera: Noctuidae); *Tecia solanivora* (Lepidoptera: Gelechiidae), a pest in potato crops of the Americas; *Plutella xylostella* (Lepidoptera: Plutellidae), of great importance in cruciferous crops; and the fall armyworm, *Spodoptera frugiperda* (Lepidoptera: Noctuidae), which causes losses in corn, cotton and rice (Keszthelyi et al., 2011; Du et al., 2011; Chagas et al., 2010; Suckling & Brockerhoff, 2010; Bosa et al., 2006; Monnerat et al., 2006).

The biological control of insect pests is an important alternative to the management of insects (or Integrated Pest Management-IPM). Unfortunately insect pests have been attacked primarily with chemical products, which cause huge environmental losses and adverse effects on human health. However, biological control and IPM-compatible chemicals can be used together [as outlined in a recent review by Gentz et al., (2010). Extensive research has centred on the search for an appropriate insecticidal peptide or polypeptide with toxicity to pest organisms, but not to flora and fauna. Researchers also hope to establish the most appropriate means of delivering the biological molecule to its site of action (De Lima et al., 2007). Recombinant DNA technology allows the exploitation of the insecticidal properties of

entomopathogenic organisms. It offers environmentally friendly options for the cost-effective control of insect pests (St Leger & Wang, 2010). Bioinsecticides include microbial agents, natural enemies, plant defences, metabolites, pheromones and genes that transcribe toxic peptides or proteins. The number and variety of toxins is extensive. For example, there are at least 0.5 million insecticidal toxins from arachnids, and evidence suggests that the use of novel toxic factors is likely to be extensive (Whetstone & Hammock 2007).

2. Typical anti-insect toxins

There are two classes of insecticidal toxins: (1) peptide-like toxins (3-10 kDa) from some scorpion and spider venoms and (2) the high molecular mass toxins (*i.e.*, about 1000 residues), such as the latrotoxins from the venom of the spider *Latrodectus* or the crystal proteins of *Bacillus thuringiensis* (De Lima et al., 2007; Schnepf et al., 1998). The toxins of the first group consist of one chain that contains many cysteine residues and intramolecular disulfide bridges. These peptides interact with ion channels (*i.e.*, those for Na^+, K^+, Ca^{2+} and Cl^-) on cellular membranes (De Lima et al., 2007). Recently a peptide-like toxin nomenclature has been proposed that takes into account the basis of activity, the biological source and the relationship with other toxins (King et al., 2008). The primary sources of entomopathogenic proteins in the second group of toxins are several organisms, including spiders, snakes, scorpions, anemones, snails, lacewings, insects, fungi and bacteria (De Lima et al., 2007; Schnepf et al., 1998).

Toxins from arthropod venoms consist of combinations of biologically active compounds (peptides, proteins, nucleotides, lipids and other molecules). They are used for paralysing insects and for defence against natural enemies. They interact with ion channels and/or receptors from neurological systems in the target organism (De Lima et al., 2007).Venom-derived peptide toxins target voltage-gated Na^+, K^+, Ca^{2+}, or Cl^- channels. Proteins, such as neuropeptides and hormones, are analogous. Their effects depend upon their specific activities (Whetstone & Hammock, 2007). Antagonists disrupt and interfere with development and behaviour. Spiders and scorpions maybe the most important arthropods having insecticidal toxins. Many spider venoms contain a complex mixture of both neurotoxic and cytolytic toxins (see: www.arachnoserver.org). Virtually all insecticidal spider toxins contain a cystine-knot motif that provides them with chemical and biological stability (King et al., 2002; Tedford et al., 2004). These types of venoms contain acylpolyamines (from the Araneidae family), cytolytic toxins (from the Zodariidae family) and neurotoxic peptides (J-atratoxins), and neurotoxins (>10 kDa) and enzymes (~35 kDa) in the Sicariidae and Theridiidae families respectively (Vassilevski et al., 2009; Gunning et al., 2008). *Agelenopsis aperta* employs venom that is very active against insects. It is composed of toxins (agatoxins) that attack transmitter-activated cation channels, voltage-activated sodium channels and voltage-activated calcium channels. The α-agatoxins, μ-agatoxins and ω-agatoxins alter insect ion channels (Adams, 2004). Australian funnel-web spiders [Mygalomorphae (order): Hexathelidae (family): Atracinae (subfamily)] have ω-atracotoxins (36–37 residues with six cysteines in a disulfide pattern), which slow insect cation voltage-dependent channels (Chong et al., 2007).

Scorpions are a special group of organisms that have interesting toxins. These toxins have 23-78 residues. Generally the conformation has an α-helix packed against a three-stranded β-sheet stabilized by four disulfide bonds. Scorpion toxins recognize the face of voltage-

dependent sodium channels and alter their gating. They are defined as α-or β-toxins, based on their mechanism of action (Rodríguez de la Vega et al., 2010; Gurevitz et al., 2007; Karbat et al., 2004). Anti-insect α-toxins bind to voltage-dependent sodium channels with high affinity (Gordon et al., 2007). Scorpion β-toxins change the voltage dependence of channel activation. The first class of entomopathogenic scorpion β-toxins is comprised of excitatory toxins. They are composed of 70-76 amino acids. These toxins may induce spastic paralysis by the activation of sodium flux at negative membrane potential. A second group consists of depress ant toxins, which induce flaccid paralysis by depolarization of the axonal membrane. A third set is composed of active toxins, which act on both insect and mammalian sodium channels, with typical depressant effects on insects (Gurevitz et al., 2007).

Surprisingly, some insects (such as the tobacco hornworm *Manduca sexta*) produce insecticidal peptides (each peptide has 23 amino acids) from haemolymph. These molecules can cause paralysis in the larvae of many insects (Skinner et al., 1991). For example, a dose of 105 plaque-forming units of baculovirus containing a poneratoxin DNA sequence from the ant, *Paraponera clavata*, was adequate for controlling lepidopteran individuals (*S. frugiperda*) (Szolajska et al., 2004).

Microorganisms possess toxins for the biological control of insects. Fungus is an entomopathogenic option. *Beauveria bassiana* has a long history in relation to the control of lepidopteran, coleopteran and dipteran species (Howard et al., 2010; Qin et al., 2010; Cruz et al., 2006; Shah & Pell, 2003). *Metarhizium anisopliae* has been used against ticks and insects, this fungus has a wide set of virulent factors, such as lipolytic enzymes, proteases, chitinases and toxins (destruxins) (Schrank & Vainstein, 2010; Pava-Ripoll et al., 2008). Ascomycota (genera *Cordyceps*, *Hypocrella* and *Torrubiella*), Zygomycota (genera *Conidiobolus* and *Entomophaga*), Deuteromycota (genus *Aschersonia*), Zygomycetes (genus *Entomophthora*) and Hyphomycetes (genus *Hirsutella*), which have activity against lepidopterans and coleopterans (Shah & Pell, 2003). Many bacteria, such as *Serratia marcescens*, *Photorhabdus luminescens, B. thuringiensis* and *Xenorhabdus nematophilus*, can produce entomopathogenic toxins (Roh et al., 2010; Whetstone & Hammock, 2007). Baculoviruses have been used as safe and effective biopesticides for the protection of crops and forests in the Americas, Europe and Asia. The *oryctes* virus has also demonstrated insecticidal activity against the rhinoceros beetle. The entomopathogenic parvoviruses are an insecticidal option. The *H. armigera* stunt virus (a tetravirus) has been isolated from pests and may be useful for the development of genetically modified plants (Whetstone & Hammock, 2007).

Plants produce a great variety of toxic compounds that are responsible for insect self-defense mechanisms. Plant cyclotides contain 30 amino acids with acyclic peptide backbone and a knotted alignment of three conserved disulphide bonds connected in a "cystine knot" motif. Members of Lepidoptera and Coleoptera are susceptible to plant cyclotides from the Violaceae, Rubiaceae and Cucurbitaceae families (Gruber et al., 2007). Plant cysteine proteases are accumulated after lepidopteran infestation affecting insect growth (Pechan et al., 2002). Plant defensins are antimicrobial proteins with eight conserved cysteines and four disulfide bridges. Defensins attack lepidopteran α-amylases, causing feeding inhibition (Kanchiswamy et al., 2010; Rayapuram & Baldwin, 2008). Plant glucanases, chitinases, lectins and dehydrins are induced after attack by lepidopteran and coleopteran pests (Ralph et al., 2006).

3. The phylogenetic relationship of insecticidal toxins and their comparison with lepidopteran- and coleopteran-specific molecules

Twenty-seven amino acid sequences from the RCSB Protein Data Bank (PDB) (http://www.pdb.org/pdb/home/home.do) were selected by a bibliographical revision, using the criteria of established insect-specific toxicity. Next a phylogenetic analysis of insect-specific toxins was performed (Figure 1) by means of Phylogeny.fr platform (http://www.phylogeny.fr/) (Dereeeper et al., 2008). The available data from a bibliographical search,show insecticidal protein sequences from a large variety of organisms with toxicity against several orders of targets, including 11 anti-lepidopteran toxins and five coleopteran-specific toxins (Table 1).
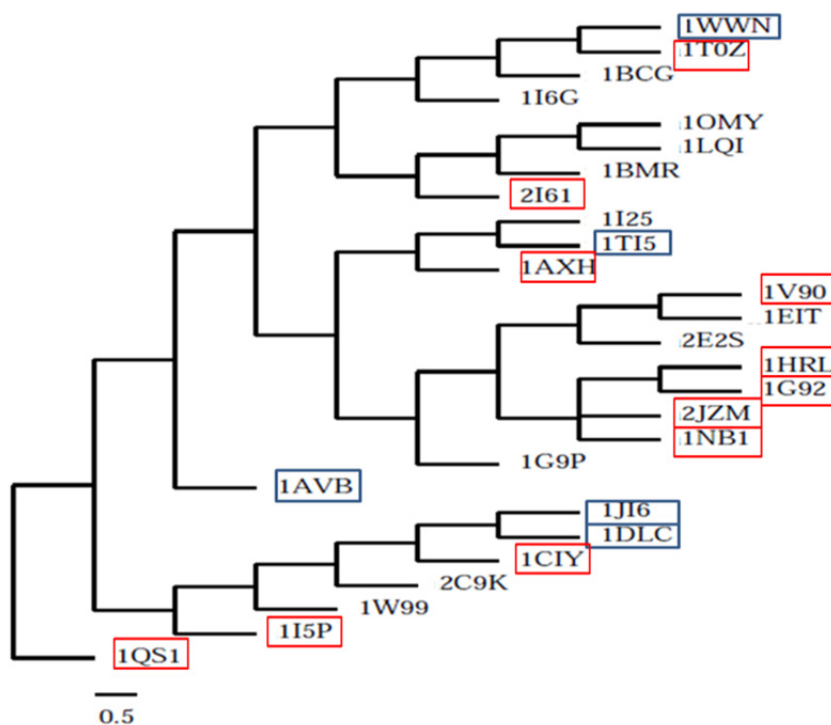
Fig. 1. Phylogenetic tree for insecticidal toxins. The blue squares indicate the coleopteran-specific amino acid sequences and the red squares show antilepidopteran toxins. The analysis of the toxins was done by the parsimony method with the TNT 1.1 program, using the alignment previously obtained with MUSCLE 3.7. The analysis was carried out 1000 times in order to obtain a strict consensus tree by using the bootstrapping tool. The consensus phylogenetic tree was computed by the TreeDyn 198.3. See the text for an analysis.

ID PDB	TOXIN	SOURCE	ORDER TARGET	REFERENCES
1AVB	Arcelin 1	*Phaseolus vulgaris*	Coleoptera	Fabre et al., 1998; Mourey et al., 1998
1AXH	ω-ACTX-HV1	*Hadronyche versuta*	Lepidoptera, Diptera, Ixodida	Chong et al., 2007; Fletcher et al., 1997
1BCG	Bjxtr-IT	*Buthotus judaicus*	Blattaria	Possani et al., 1999; Oren et al., 1998
1BMR	Lqh III	*Leiurus quinquestriatushebraeus*	Blattaria	Krimm et al., 1999
1CIY	Cry1Aa	*Bacillus thuringiensis*	Lepidoptera	Grochulski et al., 1995; López-Pazos & Cerón, 2007
1DLC	Cry3A	*Bacillus thuringiensis*	Coleoptera	Li et al., 1991; López-Pazos & Cerón, 2007
1EIT	μ-agatoxin	*Agelenopsis aperta*	Diptera	Adams, 2004; Omecinsky et al., 1996
1G92	Poneratoxin	*Paraponera clavata*	Lepidoptera	Szolajska et al., 2004
1G9P	ω-Atracotoxin-HV2A	*Hadronyche versuta*	Orthoptera	Chong et al., 2007; Wang et al., 2001
1HRL	PP1	*Manduca sexta*	Lepidoptera	Yu et al., 1999; Skinner et al., 1991
1I5P	Cry2Aa	*Bacillus thuringiensis*	Lepidoptera, Diptera	Morse et al., 2001; López-Pazos & Cerón, 2007
1I6G	CsE-v5	*Centruroides sculpturatus Ewing*	Blattaria	Jablonsky et al., 2001; Possani et al., 1999; Lee et al., 1994
1JI6	Cry3Bb1	*Bacillus thuringiensis*	Coleoptera	Galitsky et al., 2001; López-Pazos & Cerón, 2007
1LQI	Lqh(α)IT	*Leiurus quinquestriatus hebraeus*	Diptera	Tugarinov et al., 1997; Zilberberg et al., 1997
1I25	Huwentoxin-II	*Selenocosmia huwena*	Blattaria	Liang., 2004; Shu et al., 2002
1NB1	Kalata B1	*Oldenlandia affinis*	Lepidoptera	Rosengren et al., 2003; Gruber et al., 2007
1OMY	BmKaIT1	*Buthus martensii* Karsch	Diptera, Orthoptera	Ji et al., 1996

ID PDB	TOXIN	SOURCE	ORDER TARGET	REFERENCES
1QS1	VIP2	*Bacillus thuringiensis*	Lepidoptera	Han et al., 1999
1TI5	VrD1	*Vigna radiata*	Coleoptera	Liu et al., 2006
1T0Z	BmK IT-AP	*Buthus martensii* Karsch	Lepidoptera	Li et al., 2005; Hao et al., 2005
1V90	δ-palutoxin IT1	*Paracoelotes luctuosus*	Lepidoptera	De Lima et al., 2007; Ferrat et al., 2005
1WWN	BmK-βIT	*Buthus martensii* Karsch	It displays toxicity against Diptera and is related with AaIT from *Androctonus australis* Hector with activity against Blattaria, Orthoptera, Diptera and Coleoptera	Pava-Ripoll et al., 2008; Zlotkin et al., 2000
1W99	Cry4Ba	*Bacillus thuringiensis*	Diptera	Boonserm et al., 2005; López-Pazos & Cerón, 2007
2C9K	Cry4Aa	*Bacillus thuringiensis*	Diptera	van Frankenhuyzen, 2009; Boonserm et al., 2006
2E2S	Agelenin	*Agelena opulenta*	Orthoptera	Yamaji et al., 2007
2I61	LqhIT2	*Leiurus quinquestriatushebraeus*	Lepidoptera, Diptera	Karbat et al., 2007; De Lima et al., 2007
2JZM	Chymotrypsin inhibitor C1	*Nicotiana alata*	Lepidoptera	Schirra et al., 2008; Schirra et al., 2001; Miller et al., 2000

Table 1. Some toxins from several sources for which **experimentally determined structures** are available in the Protein Data Bank (PDB).

The observed toxin phylogenies - specifically active against lepidopteran species - have several relationships among them and are distributed along all of the branches (Figure 1). *B. thuringiensis* proteins (Cry and vegetative insecticidal protein (VIP)) are closely related in a separated branch, containing three lepidopteran-specific proteins (Cry1Aa, Cry2Aa and VIP2). BmK IT-AP is related with BmK-βIT, Bjxtr-IT and CsE-v5. The antilepidopteran structure 2I61 is in the same group as 1BMR, 1LQI and 1OMY. The Hadronyche versuta toxin (ω-ACTX-Hv1a) has proximity with Huwentoxin-II (*Ornithoctonus huwena*) and the coleopteran-specific VrD1 from the wild mung bean. 1V90 (a lepidopteran-specific toxin), 1EIT and 2E2S are close. The antilepidopteran toxic factors PP1, Poneratoxin, Kalata B1 and

chymotrypsin inhibitor C1, have proximity with ω-Atracotoxin-Hv2A from *H. versuta*. Only arcelin1 is in a different site. One might ask whether the amino acid sequences associated with antilepidopteran toxins could have the same biological role, such as 1G9P, 2E2S, 1EIT, 1I25 or 1WWN. Moreover, the phylogenetic tree showed no relationship among Coleopteran-specific sequences, except for 1DLC and 1JI6, which belong to the family of *B. thuringiensis* Cry toxins (Figure 1, Table 1). However, this analysis indicates that 1T0Z (from the Asian scorpion *Buthus martensi* Karsch) and 1I25 (from the Chinese bird spider *O. huwena*) may have anti-coleopteran properties due to the fact that they are in the same branch as 1WWN and 1TI5, respectively (Figure 1). Studies have shown that insecticidal toxins purified from arthropod venoms exert their effects via specific interactions with ion channels and receptors in the central or peripheral nervous system (De Lima et al., 2007; Bloomquist, 2003; Johnson et al., 1998; Fletcher et al., 1997). *B. martensi* Karsch venom has four peptides related to the excitatory insect toxin family and 10 related to the depressant insect toxin (Goudet et al., 2002). Huwentoxin-II (from the spider *O. huwena*) can paralyse cockroaches for hours (ED50 of 29 ± 12 nmol/g) and increase the activity of Huwentoxin-I (a toxin targeting ion channels) (Liang, 2004).

4. Insecticidal toxins and site-directed mutagenesis: case reports

Site-directed mutagenesis is a powerful methodology for studying function and protein structure through manipulation at the level of the DNA molecule. Advances in site-directed mutagenesis have allowed the transfer of new or improved gene roles between organisms, such as bacteria, plants and animals (Adair & Wallace, 1998; James & Dickinson, 1998). In this section, we describe several experiences of the application of site-directed mutagenesis on insecticidal toxin sequences.

4.1 Mutagenesis exposes essential residues in the anti-insect toxin Av2 from *Anemonia viridis*

Sea anemones (Metazoa, Cnidaria, Anthozoa, and Hexacorallia) are sessile predators that are highly dependent on their venom for prospering in a wide range of ecological environments. Venom analysis shows a significant collection of low molecular weight toxins: ~20 kDa pore-forming toxins, 3.5–6.5 kDa voltage-gated potassium channel-active toxins and 3–5 kDa polypeptide toxins active on voltage-gated sodium channels (Navs) (Moran et al., 2009). [A Nav has a central role in the excitability of animals. It functions in the initiation and propagation of action potentials (Goldin, 2002).]

The *Anemonia viridis* toxin 2 (Av2) is a lethal neurotoxin. Av2 has shown a clear preference for insect Nav from the assessment of toxin effects on the *Drosophila melanogaster* sodium channel (DmNav1) expressed in *Xenopus laevis* oocytes (Moran et al., 2009; Warmke et al., 1997). Hence, mutagenesis offers a means of examining residues thought to be important for Av2 activity on insect Navs. A synthetic gene coding for Av2 was designed. It was cloned into the expression vector pET-14b and used to transform appropriate *Escherichia coli* cells (strain BL21). Av2 point mutations (Note: amino acid abbreviations and single-letter designations are given in Table 1 of the chapter by Figurski et al.) [V2A (*i.e.*, residue 2 changed from V to A), P3A, L5A, D7A, S8A, D9A, G10A, G10P, S12A, V13A, R14A, G15A, G15P, N16A, T17A, L18A, G20P, I21A, P28A, S29A, W31A, H32A, N33A, K35A, K36A,

H37A, P39A, T40A, I41A, W43A and Q47A] were established by means of PCR (Polymerase Chain Reaction) using the appropriate primers and the synthetic Av2 gene as the DNA template. The mutant proteins were purified by reverse-phase high performance liquid chromatography. Toxicity assays were done on *Sarcophaga falculata* blowfly larvae. (They were scrutinized for immobilization and contraction). Competition binding assays were done with the neuronal membranes of adult cockroaches (*Periplaneta americana*). The toxicity correlated well with the results of the binding assays. This study indicated that N-terminal aliphatic residues (V2 and L5) play a role in such activity. The central region of the toxin is not involved in the toxic activity. W23 and L24 are important residues in toxin structure. At the C-terminus, it is noteworthy that residue I41 is involved in the bioactive surface of Av2. Residues V2, L5, D9, N16, L18 and I41 are pivotal amino acids for toxicity to blowfly larvae and for binding to cockroach neuronal membranes. The information from these mutants may be applicable to other insect orders (Moran et al., 2006).

4.2 Mutagenesis demonstrates that N183 is a key residue for the mode of action of the Cry4Ba protein

B. thuringiensis is a biopesticide bacterium. Its insecticidal properties are attributed (predominantly) to Cry toxins (a protein family), which are synthesized during the sporulation phase of the organism (Roh et al., 2007). The Cry protein is ingested by the susceptible insect, solubilized in the gut lumen, and cleaved by proteases to yield the activated 60 kDa toxin. Next Cry toxins are recognized by cadherin-like receptors (CADR) to assemble oligomeric forms of the toxin. The toxin oligomers have binding affinities to the secondary receptors: aminopeptidase N (APN), alkaline phosphatase (ALP), ADAM metalloprotease or glycosylphosphatidyl-inositol (GPI)-anchored proteins. The oligomers insert into the apical membrane of midgut-generating pores to cause osmotic lysis and insect death (Ochoa-Campuzano et al. 2007; Pigott & Ellar, 2007). Cry toxin is composed of three functional domains. Domain I comprises seven hydrophobic and amphipathic α-helices and is capable of forming pores in the apical membrane of the insect midgut. Domain II is made of three variable anti-parallel β-sheets, which are responsible for receptor recognition. Domain III has two anti-parallel β-strands involved in structural stability and receptor binding (Schnepf et al., 1998). Site-directed mutagenesis on Cry proteins revealed the function of each domain in the toxicity to the target insect. This fact provides a perspective on the generation of toxins with enhanced toxicity or new specificities.

A collection of Cry4Ba mutants (Figure 2), which are modified in polar uncharged residues (Y178, Q180, N183, N185, and N195) within α-helix 5, were developed to observe their effects on biological activity. All mutant toxins were generated using PCR-based site-directed mutagenesis, and each mutant was expressed from the *lac* promoter in *E. coli* upon IPTG (isopropyl β-D-thiogalactopyranoside) induction. The Cry4Ba-N183A mutant does not display lethality, while alanine substitutions for other residues (Y178, Q180, N185, and N195) still maintained more than 70% of the insect toxicity of the Cry4Ba standard (Figure 2). This result indicated that N183 plays an important role in the functionality of the Cry4Ba toxin (Likitvivatanavong et al., 2006).

Other studies indicated that N183 plays a crucial role in both toxic and structural properties. Mutants N183Q and N183K were made so as to be insoluble at alkaline pH. Mutations at N183 using several residues (with different structural characteristics) revealed that

substitutions with a polar amino acid still retained lethal activity similar to the Cry4Ba standard. Nevertheless, changes to charged or nonpolar residues suppressed biological activity (Figure 2). In conclusion, N183 polarity and α-helix 5 localization (in the middle of domain I) are very important to the toxicity of the Cry4Ba protein (Likitvivatanavong et al., 2006).
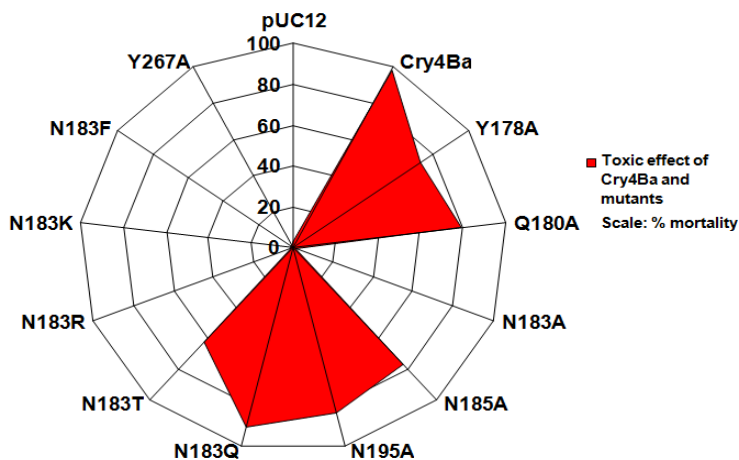
Fig. 2. Biological activity of Cry4Ba and mutants. The red colour indicates lethality and level. Bioassays for mosquito-larvicidal activity were performed using 2-day-old *Stegomyia* (*Aedes*) *aegypti* (mosquito) larvae. The altered residues in the mutant proteins are given on the outside of the graph. The gene for the mutant protein was inserted into the plasmid expression vector pUC12 and induced from the *lac* promoter. pUC12 on the graph depicts the toxicity of the vector alone.

4.3 A Juvenile hormone esterase with a mutated α helix shows improved insecticidal effects

Juvenile hormone (JH) regulates several physiological events in insects (development, metamorphosis, reproduction, diapause, migration, polyphenism and metabolism). JH esterase (JHE) is a hydrolytic enzyme from the α/β-hydrolase fold family, which metabolizes JH (Kamita et al., 2003). When JHE is injected into lepidopteran larval states, it causes a darkening and a decrease in feeding (Hammock et al., 1990; Philpott & Hammock, 1990). JHE is rapidly cleared from the haemolymph following inoculation, suggesting a discriminatory system for its elimination (El-Sayed et al., 2011). In testing, it was revealed that the double histidine mutated JHE [JHE K204H and R208H (in an amphipathic α helix)] is capable of blocking clearance from the haemolymph by reducing its binding to the JHE receptor. These experiments used *Autographa californica* NPV (AcMNPV, a baculovirus with pathogenic activity towards insect pests) as an expression vehicle. JHE shows enhanced insecticidal activity against the lepidopteran larvae of *M. sexta* (tobacco hornworm), *Heliothis virescens* (tobacco budworm) and *Agrotis ipsilon* (black cutworm) (El-Sayed et al., 2011).

Mutant and wild-type JHEs were produced and purified from insect cells, and their activities were found in the culture supernatants of insect cells. The specific activity of

mutant JHE was 6.5 nmol of JH III acid (a metabolism product of JH by JHE) formed min^{-1} mg^{-1}. The specific activity of wild-type JHE was 61.3 nmol of JH III acid formed min^{-1} mg^{-1}. The K204H and/or R208H alterations, although far-removed from the catalytic site of the protein, induced allosteric properties that led to a decrease in activity. No statistically significant differences were seen in the clearance of JH hydrolysis activity in the fourth instars of *H. virescens*, *A. ipsilon* and *M. sexta*. Bioassays (using the first instars of *H. virescens* and *A. ipsilon*) were done to establish the lethal concentration and the lethal time and to determine the result of the expression of mutant JHE on the insecticidal lethality of the baculovirus. The results showed that the median lethal concentration of mutant JHE was 3.2-fold lower in *H. virescens*, in contrast to the effect of AcMNPV. There is no effect on *A. ipsilon*, as observed by the bioassay (Table 2). The most notable difference between the esterases was the higher median lethal concentration (1.9-fold) of mutant JHE compared to a non-mutant JHE against *A. ipsilon* (Table 2). The median lethal concentration of mutant JHE in *H. virescens* was 3.5-fold lower than mutant JHE in *A. ipsilon*. The median lethal time of *H. virescens* and *A. ipsilon* treated with mutant JHE was about 4.8 and 5.3 days, respectively. It was about the same for non-mutant JHE. In addition, feeding assays were carried out using the first instars of *M. sexta* (for 4 days on an artificial diet or on a tomato leaf). The results showed 41–90% lower mass for the mutant than for the JHE wild type (non-mutant) at the end of the experiment. The study showed that point mutations of the amphipathic α-helix were sufficient for improving insecticidal activity (El-Sayed et al., 2011).

Insect	Esterase	Median lethal concentration (x10^5) (95% Confidence Limits)
H. virescens	Mutant JHE	1.8 (1.0-2.6)
	Wild type JHE	**2.7** (1.8-3.8)
A. ipsilon	Mutant JHE	6.3 (3.6-13)
	Wild type JHE	3.3 (2.3-4.6)

Table 2. Lethal concentrations of mutant and wild-type versions of JHE in the first instar larvae of *H. virescens* and *A. ipsilon*. Insects were inoculated with recombinant JHEs in a polyhedral virus vehicle. The median lethal concentration is expressed as polyhedra per ml (modified of El-Sayed et al., 2011).

4.4 Predicting important residues responsible for the capacity of scorpion α-toxins to discriminate between insect and mammalian voltage-gated sodium channels

Scorpion toxins are poison molecules (61–67 amino acids). Scorpion α-toxins recognize voltage-gated sodium channels (NaCh). NaChs mediate the temporary increase in sodium ion permeability thereby generating action potentials. The toxin expands the action potential by delaying the inactivation stage (Gordon et al., 2007). LqhαIT, from the scorpion *Leiurus quinquestriatus hebraeus*, is an α-toxin that is highly active on insect NaChs. A mutagenic analysis of LqhαIT was performed, revealing that the residues important for function are grouped into two different domains. A new toxin made by putting the efficient region of LqhαIT onto Aah2 (an anti-mammalian α-toxin from the scorpion *Androctonus australis* Hector) proved to be anti-insect (Karbat et al., 2004).

Mutations in the cDNAs of *L. quinquestriatus hebraeus* encoding LqhαIT were generated by PCR (Gurevitz et al., 1991). A CD (Circular Dichroism) Spectroscopy analysis was recorded at 25°C (Karbat et al., 2004). Some residues (Y14, E15, D19, Y21, E24, L25, K28, A39, N54 and P56) had no effect on the biological action or alteration of the CD spectrum. N44 and mutants F17G/A, R18A, W38A had decreased lethality and an unchanged CD spectrum. The F17W and W38Y mutants had activities similar to wild-type LqhαIT, so aromatic side chains affect toxin function. The substitutions I57A/T, R58K, V59A/G, R58K/V59A, K62A/L/R and R64N in the C-terminal region reduced biological activity. The substitution R58N had a marked negative effect on biological activity. This result implies that both charged amine groups and the aliphatic moiety in R58 are principal determinants in functionality. Biologically important residues appear in two domains. The first domain (core-domain) consists of F17, R18, W38 and N44. The second domain (NC-domain) is formed by residues K8, Y10, P56, I57, R58, V59, K62 and R64 (Karbat et al., 2004). LqhαIT and Aah2 have an overall similarity of 70%, although the similarity varies in the NC-domain. The core-domain and the NC-domain of Aah2 were replaced by the LqhαIT counterparts to generate four hybrids (Table 3). The constructs were evaluated with biological assays using *S. falculata* blowfly larvae. Immobilization and contraction were measured, and an effective dose of 50% (ED50) was calculated (Table 3) (Karbat et al., 2004).

Toxin	ED50/100 mg of *S. falculata* **body weight**
Parental	
LqhαIT	13 ng
Aah2	> 10 μg
Mutant toxin	
Aah2[LqhαIT(8–10)]	> 10 μg
Aah2[LqhαIT(56–64)]	> 10 μg
Aah2[LqhαIT(8–10, 56–64)]	64 ng
Aah2[LqhαIT(8–10, G17F, 56–64)]	37 ng

Table 3. Toxicity assays of Aah2 and its counterpart mutants (Karbat et al., 2004).

The similar activities of Aah2[LqhαIT(8–10, G17F, 56–64)] and LqhαIT indicate that their functional NC-domains are equally oriented. This indicates that the increase of insecticidal activity is related to the arrangement of the NC-domain in a structure that projects into the solvent. Remarkably this conformation is universal to all scorpion α-toxins with lethality on insects, in contrast with the flat face in α-toxins that are toxic to mammals (Karbat et al., 2004).

5. Final remarks

5.1 Novel sources?

Whole-genome sequencing projects are a resource of biological functions and their annotation allows for the detection of proteins through orthologous sequences (common ancestry), searches and primary and tertiary structure correlation - a process named "comparative genomics" (Lee et al. 2007; Ellegren, 2008). This theoretical approach makes it possible to find candidate toxins in sequenced genomes. An appropriate criterion for the identification of novel lepidopteran and coleopteran candidate toxins can be understood in

terms of the "guilt by association" principle (Gabaldon &Huynen, 2004; Aravind, 2000). For this reason, we applied a very basic protocol (Figure 3). BLAST (tblastn) searches from the National Centre for Biotechnology Information (NCBI) (http://blast.ncbi.nlm.nih.gov /Blast.cgi). Searches were done using each toxin (from Table 1) as a query. The iterative searches were done for proteins larger than 100 aminoacids with an inclusion threshold of 0.01 (the statistical significance limit for inclusion of a sequence in the process) and for proteins smaller than 100 aminoacids with an inclusion threshold of 0.1. The searches used the 881 completely sequenced bacterial and archaeal genomes available on the NCBI Microbial Genomes website at the time of this analysis (January 2011) and the entire NCBI environmental samples database (1.66 million Whole Genome Shotgun reads) (see http://www.ncbi.nlm.nih.gov/). The searches were done until either convergence was achieved or until the last iteration before the first known false positives appeared. Significant hits to proteins encoded in these genomes were further classified as possible insect-specific toxins. The BLAST analysis showed fourteen microbial sequences with a high similarity to insecticidal queries (Table 4). There is a version of Arcelin 1 encoded in the genome of the cyanobacterium *Acaryochloris marina* (Tables 1 and 4). Cry proteins from *B. thuringiensis* have a degree of correspondence to sequences in the genomes of four bacteria and one archaeon (Table 4). The VIP2 toxin from *B. thuringiensis* appears to be very diverse in nature. We found VIP-like toxins encoded by eleven bacterial genomes (Table 4). The identified lepidopteran-active toxins are associated with Cry1Aa, Cry2Aa and VIP2. Anti-coleopteran-like toxins were identified, and they are related to Arcelin 1 and Cry3A (Table 4). The search in the Environmental Sample Database showed seven most probable insecticidal sequences related with a Blattaria-active toxin, a coleopteran-specific toxin, four lepidopteran-active toxins and an anti-dipteran toxin (Table 4).
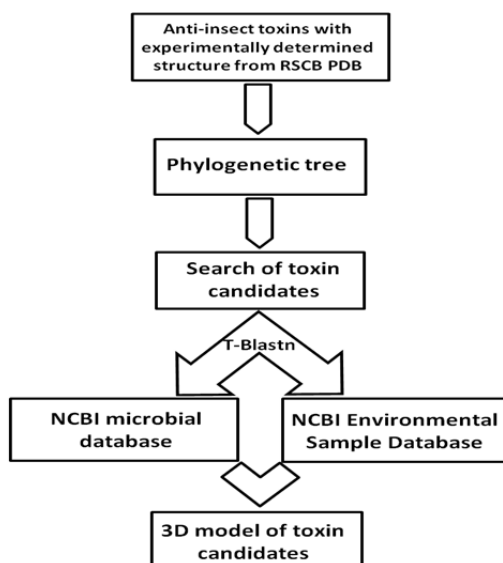
Fig. 3. Diagram of the work.The search for lepidopteran- and coleopteran-specific toxins was done through a basic strategy with the BLAST program on microbial and environmental genomes.

For our trial, the most important organisms harbouring lepidopteran- and coleopteran-active toxins are *A. marina*, *B. weihenstephanensis* and *Clostridium difficile*. First, *A. marina* is a unicellular cyanobacterium containing chlorophyll d as a major pigment (Ohashi et al., 2008). Second, *B. weihenstephanensis* is a Gram-positive, facultatively anaerobic, spore-forming bacterium. This organism has food poisoning potential and is able to grow aerobically at 7°C. *B. weihenstephanensis* has a 16s rDNA signature sequence [1003]TCTAGAGATAGA and the signature sequence [4]ACAGTT of the gene for CspA (a major cold shock protein) (Lechner et al., 1998). Third, *C. difficile* is a Gram-positive spore-forming anaerobic bacterium thought to be involved in diarrhoea and colitis. *C. difficile* codes for two potent toxins (A and B), which attach to specific receptors in the lumen of human colonic epithelium (Vaisnavi, 2010). It is interesting to note that "particular" organisms have versions of these kinds of toxins, such as *Methanosarcina acetivorans* (an acetate-using methanogen archaeon), *Dyadobacter fermentans* (a Gram-negative bacterium isolated from maize and related to *Runella slithyformis*), the marine bacterium *Microscilla furvescens*, and *Cupriavidus necator* - previously known as *Ralstonia eutropha*, a microorganism that can be isolated from several environmental sources, such as soil and water, and which is important in polyhydroxyalkanoate production and bioremediation by the degradation of chlorinated aromatic pollutants (Galagan et al., 2002; Chelius & Triplett, 2002; Lykidis et al., 2010). In addition, we detected other *Clostridium* and *Bacillus* species. The NCBI environmental samples database, a metagenome of the Sargasso Sea genetic diversity from the Venter et al. (2004) project, shows environmental sequences with anti-lepidopteran and anti-coleopteran potential (Table 4).

We built tertiary (3D) structures of some of the predicted toxins: a lepidopteran-active toxin, a coleopteran-specific toxin and a toxin from a metagenome sequence. Approximately 30% sequence identity in the primary sequence is required for the generation of useful structures (Forster, 2002; Paramasivan et al., 2006). Tertiary models of candidate insecticidal sequences were constructed by homology modelling using the crystal structure of homologous protein from the RCSB PDB database (http://www.pdb.org/pdb/home/home.do). We used SWISS-MODEL (http://swissmodel.expasy.org/) (Arnold et al., 2006) for the identification of templates (Table 4 footnotes). The structural alignments were generated with DeepView Swiss-PdbViewer 4.0 software (http://spdbv.vital-it.ch/) (Guex & Peitsch, 1997).

The final models (Figure 4) have a range of 33% to 37% identity with the templates. The toxins in Figure 4 correspond to the following (A) NCBI ID NC_009925.1 from the *A. marina* MBIC11017 genome (33% identity), (B) NCBI ID NC_010180 from the *B. weihenstephanensis* KBAB4 plasmid pBWB401 (37% identity) and (C) the hypothetical protein GOS_5670768 from the marine metagenome (33% identity) (Table 4). The most striking feature of the predicted structure of the candidate insect toxin from the *A. marina* genome consists of two large β-pleated sheets that form a scaffold on which is a possible a carbohydrate-binding region (Figure 4). These architectures and topologies are found in a wide variety of carbohydrate recognizing proteins, such as plant lectins, galactins and serum amyloid proteins (Loris et al., 1998). The model is structurally related to the jelly-roll topology, which facilitates viral entry into bacterial cells. Entry is mediated by interactions with sugar-modified proteins on the cell surface (Petrey & Honig, 2009). It has been postulated that the binding of the lectin to the sugar moiety of any of the glycosylated digestive enzymes is a potential factor of insecticidal activity (Peumans & Van Damme, 1995a, b). Based on the structural alignment of the aminoacid sequences of the toxin from *B. weihenstephanensis* with

INSECTICI DAL TOXIN (ID PDB)	ORGANISM TARGET GENOME/ENVIRONMEN TAL SOURCE	ID NCBI	E-VALUE	REGION
Microbial database				
1AVB[A]	_Acaryochloris marina MBIC11017_	NC_009925.1	3e-10	1669294-1669911
1CIY, 1DLC[B], 1I5P, 1JI6*, 1W99 and 2C9K**	_Bacillus weihenstephanensis KBAB4 plasmid pBWB401_	NC_010180	8e-97-2e-10	139296-138751
	Methanosarcina acetivorans C2A	NC_003552.1	4e-19-1e-04	3249335-3249832
	Dyadobacter fermentans DSM 18053	NC_013037.1	1e-15-4e-06	2869719-2870441
	Bacillus brevis NBRC 100599	NC_012491.1	5e-16-0.026[1]	4962833-4963585
	Ralstonia eutropha JMP134 Chromosome 1	NC_007347.1	1e-08-3.3[2]	411729-411409
1QS1	_Clostridium difficile_	ABHF02000033.1	2e-41	223624-224649
	Clostridium perfringens, E str. JGS1987	NZ_ABDW01000012.1	3e-39	66996-65971
	Clostridium botulinum, D str. 1873 plasmid pCLG1	NC_012946.1	1e-33	103322-104389
	Clostridium acetobutylicum ATCC 824	NC_003030.1	5e-17	398379-398876
	Bacillus cereus Rock4-18	NZ_ACMN01000162.1	1e-21	17703-17065
	Bacillus halodurans C-125	NC_002570.2	4e-15	3637460-3636978
	Streptomyces avermitilis MA-4680	NC_003155.4	5e-12	6590878-6591372
	Listeria monocytogenes FSL R2-561	AARS01000007.1	8e-12	72280-71786
	Lactobacillus brevis subsp. _gravesensis_	NZ_ACGG01000118.1	3e-11	220449-220006
	Aeromonas hydrophila subsp. _hydrophila_	NC_008570.1	2e-05	1214897-1215424
	Enterococcus faecalis V583	NC_004668.1	2e-05	311391-311870
Environmental database				
1BMR	hypothetical protein GOS_4202115 marine	gb|ECA60195.1	0.057	88-243

INSECTICIDAL TOXIN (ID PDB)	ORGANISM TARGET GENOME/ENVIRONMENTAL SOURCE	ID NCBI	E-VALUE	REGION
	metagenome			
1DLC[C]	hypothetical protein GOS_5670768 marine metagenome	gb|ECH33518.1	0.014	12-142
1QS1	hypothetical protein GOS_355881 marine metagenome	gb|EBA70908.1	6e-04	102-270
	hypothetical protein GOS_1734861 marine metagenome	gb|EDJ21677.1	8e-04	416-584
	hypothetical protein GOS_9568803 marine metagenome	gb|EBF61568.1	0.003	5-173
	hypothetical protein GOS_7854205 marine metagenome	gb|EBP79016.1	0.004	78-232
1W99	hypothetical protein GOS_6575573 marine metagenome	gb|EBX51304.1	0.010	29-95

Table 4. Results of the BLAST search in a microbial database (Blosum 62, E threshold 0.01) and Environmental Sample Database (Blosum 62, E threshold 0.01) (underlined by modelled sequences). * It is not compatible with *B .weihenstephanensis*. ** Only compatible with *B. weihenstephanensis* and *M. acetivorans*. [A] PDB template: 1G7Y chain C (lectin from the legume *Dolichos biflorus*). Model residues: 72-289.[B] PDB template: 3EB7 (Cry8Ea1). Model residues: 64-648.[C]PDB template: 2E58 (MnmC2 from *Aquifex aeolicus*). Model residues: 38-136. The ID PDB refers to code in Protein Data Bank; the ID NCBI refers to accession number in National Center for Biotechnology Information. The region column refers to the specific segment inside the DNA sequence from the ID NCBI column.

the Cry8Ea1 protein, a model of the toxin was obtained; and it corresponds to the general model for a Cry protein (Figure 4). The last structure corresponds to a sequence from the marine metagenome. It was built by homology to a possible transferase of *Aquifex aeolicus*, a hyperthermophilic microorganism that grows at 85-100°C. It has been suggested that this organism may be the earliest diverging eubacterium (Deckert et al., 1998). The model is composed of three α-helices and a large β-sheet, in which the first and second β-strands are arranged in parallel; and the third and fourth are anti-parallel. Interestingly, the model is somewhat similar to that of the aminoacyl-tRNA synthetase editing domain (Ribas de Pouplana & Schimmel, 2000; Naganuma et al., 2009). The phylogenetic relationships amongst these enzymes are clustered around substrate specificity (Guo et al., 2009). That the amino acid sequence from an ancient bacterium has identity with the Cry protein of *B. thuringiensis*, and that the toxin structure is similar to that of an aminoacyl-tRNA synthetase

editing domain and that it has a helix-sheet formation, hints at the origin of these toxins and their specificities.
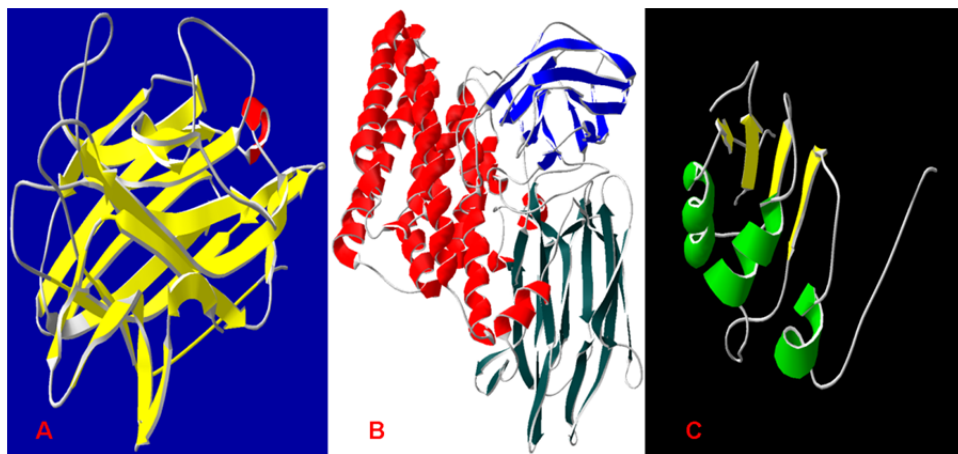
Fig. 4. Models of candidate toxins. (A) Insect toxin the from the *A. marina* genome (β-pleated sheets are in yellow); (B) Structure of the toxin from the *B. weihenstephanensis* genome (domain I is red; blue represents domain II; and domain III is green); and (C) model of the toxin from the marine metagenome (the helices are green, and the β-sheet is yellow). Also see the text.

5.2 *B. thuringiensis* vs. lepidopteran and coleopteran pests

The entomopathogenic bacterium *B. thuringiensis* has been used to help thwart the development of insect and plant resistance by using *cry* genes to construct lethal toxins against pest larvae. Some Cry proteins display biological activity against lepidopteran (Cry1, Cry2, Cry7, Cry8, Cry9, Cry15, Cry22, Cry32 and Cry51) and coleopteran (Cry1B, Cry1I, Cry3, Cry7, Cry8, Cry9, Cry14, Cry22, Cry23, Cry34, Cry35, Cry36, Cry37, Cry43 and Cry55) organisms (van Frankenhuyzen, 2009). Over the past fifteen years, research in our laboratory has focused on the study of the Cry proteins of the entomocidal bacterium *B. thuringiensis* for the biological control of insect pests in Colombia. This country is severely affected by lepidopteran and coleopteran pests, such as larvae of the potato tuber moth, *T. solanivora*; the armyworm, *S. frugiperda*; the Andean weevil, *Premnotrypes vorax* and the coffee berry borer (CBB), *Hypothenemus hampei.*

5.3 Our experience with lepidopterans

We worked with the tobacco budworm, *Heliothis virescens* (Lepidoptera: Noctuidae), an important pest in the Americas. This insect is susceptible to the Cry1Aa, Cry1Ab, Cry1Ac, Cry1Ae, Cry1B, Cry1F, Cry1I, Cry1J, Cry2, Cry8 and Cry9 toxins. Cry1Ac is the most active toxin against this pest (van Frankenhuyzen, 2009). In collaborative work, we tested chimeric Cry1 proteins (Cry1Ba, Cry1Ca, Cry1Da, Cry1Ea, and Cry1Fb) containing domain III of Cry1Ac, which shows higher toxicity in the Cry1Ba, Cry1Ca and Cry1Fb proteins. In addition, we considered an analysis for toxicity against *H. virescens* with the Cry1Ac domain

III triple-mutant toxin, named Tmut (N506D, Q509E, Y513A), supplied by Dr. Ellar (Burton et al., 1999). The test was done by means of a competition-binding assay using an immunoblotting method on nitrocellulose paper. Brush border membrane vesicles (BBMVs) from the *H. virescens* midgut were incubated with biotin-labelled toxin and with increasing concentrations of homologous (identical) or heterologous (mutant) toxin (Figure 5). The Tmut toxin was not able to compete with the Cry1Ac protein for binding to BBMVs (Figure 5). Also the mutant toxicity was 7-fold lower than the toxicity of the reference Cry1Ac. It indicates that at least one of the three residues (N506, Q509 and Y513) has an important role in the biological activity of the toxin (Karlova et al., 2005).
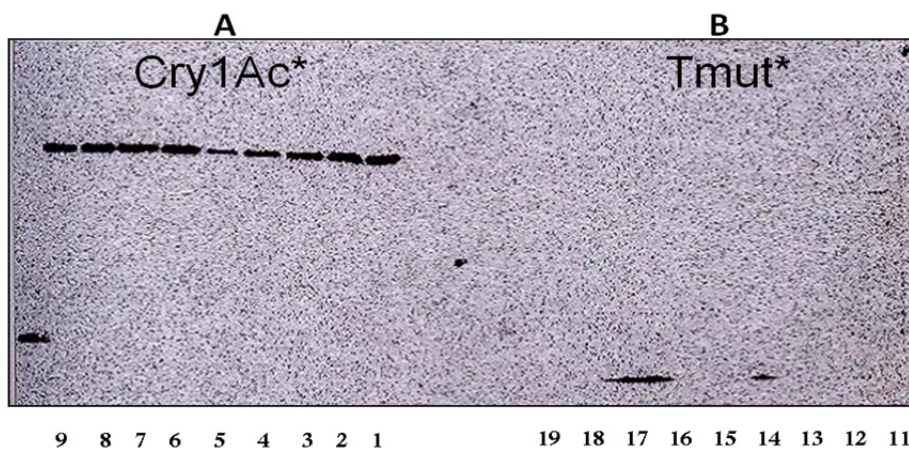
Fig. 5. The Cry1Ac binding reaction on *H. virescens* BBMVs. A. Lane 1, control with nothing added; lanes 2-5, homologous competition between parental Cry1Ac (10, 30, 90, 270 ng of the protein for each lane, respectively) and Cry1Ac labelled with biotin (10 ng); lanes 6-9, heterologous competition between the Cry1Ac domain III triple-mutant (named Tmut, which has the point mutations N506D, Q509E, and Y513A) toxin (10, 30, 90, 270 ng of the protein for each lane, respectively) and Cry1Ac labelled with biotin (10 ng). The first experiment (lanes 2-5) shows that the Cry1Ac wild-type protein (both the labelled and unlabelled proteins) binds to BBMVs (*i.e.*, competition was observed); the second experiment (lanes 6-9) indicates that the Cry1Ac domain III triple-mutant (Tmut) toxin was not able to bind to BBMVs and compete with the bound Cry1Ac wild type (labelled) protein (*i.e.*, competition was not visible). B was set up as follows: lane 11, a no-competitor control; lanes 12-15, heterologous competition between parental Cry1Ac (10, 30, 90, 270 ng of the protein for each lane, respectively) and the Cry1Ac domain III triple-mutant (Tmut) toxin labelled with biotin (10 ng); lanes 16-19, homologous competition between the Cry1Ac domain III triple-mutant (Tmut) toxin (10, 30, 90, 270 ng of protein for each line, respectively) and the Cry1Ac domain III triple-mutant (Tmut) toxin labelled with biotin (10 ng). However, the absence of bands in B confirmed that Tmut is unable to bind to BBMVs. The asterisk indicates the toxin labelled with biotin. Also see the text.

We collaborated in the genetic characterization of *S. frugiperda* (fall armyworm) strains from Brazil, Colombia and Mexico, all of which were correlated with vulnerability to the Latin American *B. thuringiensis* isolates and recombinant toxins (Monnerat et al., 2006). The recognition of genetic variability among insect strains is a decisive analysis for the development of improved pest control strategies, since the biological behaviour of Cry proteins on insect populations is dependent on the specific alleles (specially receptor related), the gene flow and fitness performance. Genetic analysis [molecular analysis for genetic variability was done with Random Amplification of polymorphic DNA (RAPD)] showed that these *S. frugiperda* populations had different levels of similarity among them (between 22% and 37%). *B. thuringiensis* isolates were found to have genes for Cry1 (Cry1Aa, Cry1Ab, Cry1Ac, Cry1B, Cry1C, Cry1D, Cry1E, Cry1G and Cry1I) and Cry2. The fall armyworm (*S. frugiperda*) groups differ in their susceptibilities to *B. thuringiensis*. The most toxic *B. thuringiensis* isolates for *S. frugiperda* had a mixture of genes for Cry1Aa, Cry1B and Cry1D. The Colombian population of this insect was the most susceptible to Latin American *B. thuringiensis* strains. The Mexican *S. frugiperda* was sensitive to recombinant Cry1Ca and Cry1Da. *S. frugiperda* from Brazil was highly susceptible to recombinant Cry1Ca, while the Colombian insects were susceptible to recombinant Cry1B, Cry1C and Cry1D proteins (Monnerat et al., 2006).

Recently we contributed to the determination of Cry1 toxicity against the first instar larvae of *T. solanivora*. We evaluated the products of the *cry1Aa*, *cry1Ab*, *cry1Ac*, *cry1Ca*, *cry1Da*, *cry1Ba*, *cry1Ea*, *cry1Fa* and *cry1Ia* genes and the gene for the hybrid protein SN1917 (encoding Cry1Ba and Cry1Ia in domain II) against the first instar larvae of this pest. We identified toxins with high activity relative to the Cry1Ba, Cry1Ac and SN1917 toxins (Martinez et al., 2003; López-Pazos et al., 2010).

5.4 Our experience with coleopterans

We researched the relationship between ecological niches of the Andean weevil, *P. vorax*, and the bacterium *B. thuringiensis*. We isolated and molecularly characterized *B. thuringiensis* native strains from potato areas (soil, store products and dead *P. vorax*). Bioassays were done using neonate larvae. In addition, the Cry3Aa recombinant toxin and its mutants (mutant 1: D354E; mutant 2: R345A, ΔY350, ΔY351; and mutant 3: Q482A, S484A, R485A) were constructed; and biological assays were performed. We found 300 strains (Bt index was 0.43, calculated as *B. thuringiensis* strains divided by the total amount of *Bacillus* strains) with 21 *cry* gene profiles. Unfortunately neither the isolates nor the recombinant Cry3Aa toxin were toxic against this coleopteran. However, a Cry3A triple mutant [R345A, ΔY350 (deletion), ΔY351 (deletion)] had a minor level of biological activity (mortality 21.87%), in contrast to wild-type Cry3Aa (<6%). This was probably due to site-directed modifications (López-Pazos et al., 2009b).

Coffee crops are severely affected by the CBB (coffee berry borer, *H. hampei*). Female insects drill fissures into the berry and lay their eggs, causing severe losses in production and quality. The entire metamorphosis takes place in the fruit (Damon 2000). This pest is currently present in more than 90% of the planted area (Bustillo 2006; Ramírez 2009). Recently, our research has been centred on the study of Cry toxins for the biological control of CBB, using recombinant proteins of Cry1B, Cry1I, Cry3A, Cry4, Cry9 and SN1917. Although the Cry1B and Cry3A proteins showed minor activity against the pest, the results

support the hypothesis that toxicity could be indirect and due to physiological factors of the insect rather than directly from the toxicity of dedicated toxin molecules. Unfortunately the Cry1I, Cry4, Cry9 and SN1917 hybrids were not toxic to CBB (López-Pazos et al. 2010, 2009a). We wanted to learn about the possible interaction between Cry toxins and the receptors in midgut CBB. Brush border membrane vesicles (BBMVs) from the midgut of *H. hampei* were prepared according to Wolfersberger et al. (1987). We used the Cry1B, Cry1I, Cry3A (López-Pazos et al. 2009a; López-Pazos et al. 2010), Cry4 and Cry9 proteins (Figure 6). BBMVs divided by protein electrophoresis showed bands between 20–220 kDa (Figure 6). A blotting test was prepared to determine the weight of Cry-binding proteins in CBB-BBMVs. Cry1B recognized proteins of ~190, 140, 80, 75, 60, 50 and 40 kDa (Figure 6). A signal for Cry1I was also visible at 140 kDa (Figure 6). Cry3A binding proteins were detected at ~140 kDa, 120 kDa and 70 kDa (Figure 6). Cry4 and Cry9 were not detected by any protein on BBMVs (Figure 6). There appeared to be several Cry1B and Cry3A toxin binding sites and/or receptors in the midgut epithelia of CBB.

5.4.1 The modes of action of Cry toxins in coleopterans: the case of CBB

The specific conditions in CBB gut physiology (acidic pH, types of proteases or high proportions of insecticide resistance alleles) are not favourable to the modes of action of the Cry proteins (López-Pazos et al. 2009a). The presence of candidate receptors for Cry proteins in CBB offers evidence for the potential of Cry protein use for the control of this pest. Cadherin-like receptors (CADR) have been studied in lepidopteran and dipteran insects. CADRs were isolated from the coleopterans *Diabrotica virgifera virgifera* (191 kDa) and *Tenebrio molitor* (179 kDa) (Sayed 2007; Fabrick et al. 2009). The CADR receptors are highly variable, with molecular weights ranging from 175 to 210 kDa. An important Cry protein binding site was found to be contained in CADR repeat number 12 (Pigott & Ellar 2007; Hua et al. 2004). It was possible improve the toxicity of Cry3 proteins against coleopterans by adding a CADR fragment containing Cry protein binding site (Park et al. 2009).

Aminopeptidase N (APN) is an N-acetyl-D-galactosamine (GalNAc)-bearing glycoprotein. APN is a receptor for Cry toxins. Different APNs have molecular weights of 90-170-kDa. It was proposed that the Cry-APN interaction has two steps: carbohydrate recognition and irreversible protein-protein interaction (Pigott & Ellar 2007). More than 60 different APNs have been registered in databases. They are from 26% to 65% similar (Herrero et al. 2005, Nakanishi et al. 1999). The 140 kDa protein (from BBMV analysis) is consistent with its being an APN. We do not know if the multiple Cry-binding polypeptides detected in CBB are different proteins or if they are one APN glycosylated differently.

It is also known that CADRs are susceptible to proteolytic digestion and for producing a ~120 kDa fraction. For this reason, CADRs can be confused with APNs in protein-protein interaction blots (Martínez-Rámirez et al. 1994). Cry proteins have multiple binding determinants, possibly specified independently by domains II and III. Moreover, Cry toxins interact with other classes of proteins in the Coleoptera order, such as ALP (molecular weight ~65 kDa), V-ATPase and the Heat-Shock Cognate protein (~ 80 kDa) and the ADAM metalloprotease (~30 kDa) (Hua et al. 2001; Ochoa-Campuzano et al. 2007; Martins et al. 2010; Nakasu et al. 2010). Any signals in the ligand blot for Cry1B and Cry3A would be related with these proteic groups. However, we identified the minor biological activity of Cry1B and Cry3A proteins on CBB larvae (López- Pazos et al. 2009a); and none was seen
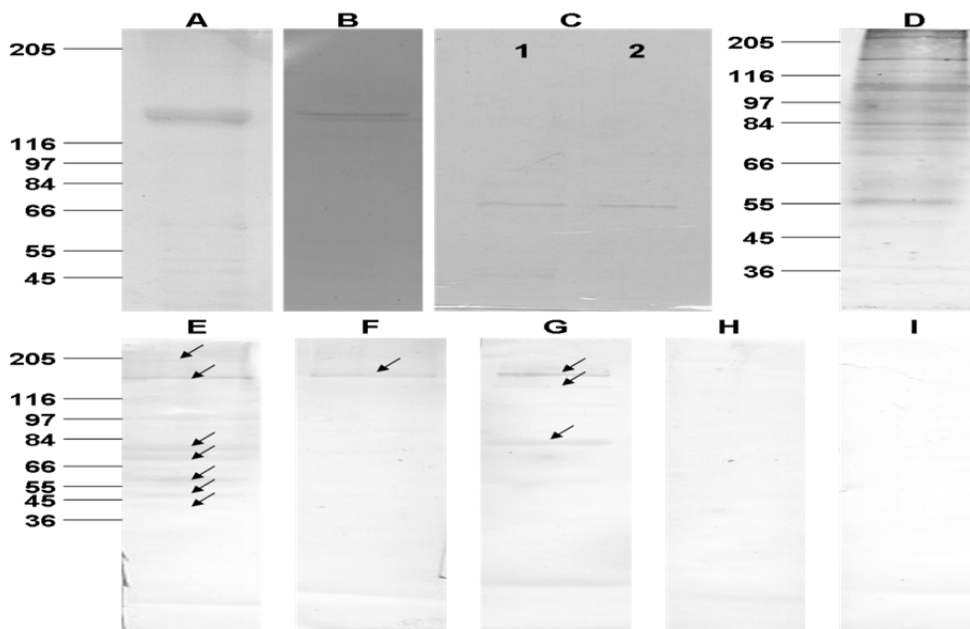
Fig. 6. SDS-polyacrylamide gel electrophoresis (SDS-PAGE) of recombinant toxins (A, B, and C) and ligand blots of Cry proteins on membrane vesicles from the midgut of the coffee berry borer (CBB-BBMVs) (E, F, G, H, and I). D shows SDS-PAGE of CBB-BBMV proteins. (A) Cry4 protoxin, (B) Cry9 protoxin, and (C) Cry4 (1) and Cry9 (2) protease-treated toxins; (D) brush-border-membrane-vesicle (BBMV) proteins from CBB. Cry-binding proteins (E-I) are indicated by the arrows. The biotin-labelled ligands (see below) are the following: (E) Cry1B, (F) Cry1I, (G) Cry3, (H) Cry4, and (I) Cry9. The numbers are molecular masses (kDa). Specifically, Cry4 and Cry9 were prepared for cloning by PCR amplification using the primers Cry4F (5′-ATGGGATCCTATCAAAATAAAAATGAATAT-3′) with Cry4R (5′-TCACTCGTTCATGCCTGCAGATTCAAT GCT-3′) and Cry9F (5′-ATGGGTACCAATAAACACGGAATTATTGGC-3′) with Cry9R (5′-TTACTGCAGTGTTTCAACGAA TTCAATACT-3′), respectively. *Bam*HI and *Kpn*I restriction sites were added to the sequences of the Cry4 and Cry9 forward primers (underlined), respectively. *Pst*I restriction sites were added to both the Cry4 and Cry9 reverse primers (underlined). The restriction sites were added to clone the amplified DNA fragment. The brush border membrane protein resolved on SDS-PAGE was transferred onto an Immobilon-P polyvinylidene difluoride (PVDF) membrane for blotting. The PVDF membrane was incubated with a biotin-labelled activated Cry toxin for binding, followed by washing with PBS/Tween (phosphate-buffered saline, pH7.4, containing 0.05% Tween-20) and incubation with streptavidin conjugated to peroxidise. The bands were visualized by peroxidase reacting with diaminobenzidine.

with Cry1I, Cry4, Cry9 and SN1917 hybrids. In this sense, there is a correlation between our data and ligand blot observations.

6. Conclusion

Insecticidal toxins are an important option for the biological control of lepidopteran and coleopteran insects. Their use in the genetic engineering of plants could provide a new generation of resistant crops. Such recombinant plants, thanks to their significant environmental and economic benefits, could help agricultural families in poor countries.

7. Acknowledgments

8. References

Abdelghany A. Y., Awadalla S. S., Abdel-Baky N. F., El-Syrafi H. A., Fields P. G. (2010). Effect of high and low temperatures on the drugstore beetle (Coleoptera: Anobiidae). *J. Econ. Entomol*. 103:1909-1914.

Adair J. R., Wallace T. P. (1998). Site-Directed Mutagenesis. *Molecular Biomethods Handbook*, Pages 347-360.

Adams M. E. (2004). Agatoxins: ion channel specific toxins from the American funnel web spider, *Agelenopsis aperta*. *Toxicon*. 43: 509–525.

Aravind L. (2000). Guilt by association: contextual information in genome analysis. *Genome Res*. 10: 1074–1077.

Arnold K., Bordoli L., Kopp J., Schwede T. (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics*. 22:195-201.

Bloomquist J.R. (2003). Mode of action of atracotoxin at central and peripheral synapses of insects. *Invertebr. Neurosci*. 5:45-50.

Boonserm P., Davis P., Ellar D. J., Li J. (2005). Crystal structure of the mosquito larvicidal Toxin Cry4Ba and its biological implications. *J. Mol. Biol*. 348: 363-382.

Boonserm P., Mo M., Angsuthanasombat C., Lescar J. (2006). Structure of the functional form of the mosquito larvicidal Cry4Aa toxin from *Bacillus thuringiensis* at a 2.8-angstrom resolution.*J. Bacteriol*.188: 3391-401.

Bosa C. F., Cotes A. M., Osorio P., Fukumoto T., Bengtsson M., Witzgall P. (2006). Disruption of pheromone communication in *Tecia solanivora* (Lepidoptera: Gelechiidae): flight tunnel and field studies. *J. Econ. Entomol*. 99:1245-1250.

Burton S.L., Ellar D.J., Li J., Derbyshire, D. J. (1999). N-Acetylgalactosamine on the putative insect receptor aminopeptidase N is recognised by a site on the domain III lectin-like fold of a *Bacillus thuringiensis* insecticidal toxin. *J. Mol. Biol*. 287:1011–1022.

Bustillo A. E. (2006). Una revisión sobre la broca del café, *Hypothenemus hampei* (Coleoptera: Curculionidae: Scolytinae), en Colombia. *Rev. Colomb. Entomol*.32: 101-116.

Chagas Filho N. R., Boiça A. L. Jr., Alonso T. F. (2010). Biology of *Plutella xylostella* L. (Lepidoptera: Plutellidae) reared on cauliflower genotypes. *Neotrop. Entomol*. 39:253-259.

Chelius M. K., Triplett E. W. (2000). *Dyadobacter fermentans* gen. nov., sp. nov., a novel gram-negative bacterium isolated from surface-sterilized Zea mays stems. *Int. J. Syst. Evol. Microbiol*. 50 Pt 2: 751-758.

Chong Y., Hayes J. L., Sollod B., Wen S., Wilson D. T., Hains P. G., Hodgson W. C., Broady K. W., King G. F., Nicholson G. M. (2007). The ω-atracotoxins: selective blockers of insect M-LVA and HVA calcium channels. *Biochem. Pharmacol*. 74:623-638.

Cruz L. P., Gaitan A. L., Gongora C. E. (2006). Exploiting the genetic diversity of *Beauveria bassiana* for improving the biological control of the coffee berry borer through the use of strain mixtures. *Appl. Microbiol. Biotechnol*. 71: 918-926.

Damon A. (2000). A review of the biology and control of the coffee berry borer, *Hypothenemus hampei* (Coleoptera: Scolytidae). *Bull. Entomol.Res*. 90:453-465.

Deckert G., Warren P. V., Gaasterland T., Young W. G., Lenox A. L., Graham D. E., Overbeek R., Snead M. A., Keller M., Aujay M., Huber R., Feldman R. A., Short J. M., Olsen G. J., Swanson R. V. (1998). The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature*.392: 353-358.

De Lima M. E., Figueiredo S. G., Pimenta AM. C., Santos D. M., Borges M. H., Cordeiro M. N., Richardson M., Oliveira L. C., Stankiewicz M., Pelhate M. (2007). Peptides of arachnid venoms with insecticidal activity targeting sodium channels. Comp.*Biochem.Physiol*. Part C. 146: 264–279.

Dereeper A., Guignon V., Blanc G., Audic S., Buffet S., Chevenet F., Dufayard J. F., Guindon S., Lefort V., Lescot M., Claverie J. M., Gascuel O. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic. Acids. Res*. 36 (Web Server issue):W465-W469.

Du E., Ni X., Zhao H., Li X. (2011). Natural history and intragenomic dynamics of the Transib transposon Hztransib in the cotton bollworm *Helicoverpa zea.Insect. Mol*. *Biol*.20:291-301.

Ellegren H. (2008).Comparative genomics and the study of evolution by natural selection. Mol. Ecol. 17:4586-4596.

El-Sayed A., El-Sheikh Shizuo G. Kamita, Kiem Vu , Bruce D. Hammock. (2011). Improved insecticidal efficacy of a recombinant baculovirus expressing mutated JH esterase from *Manduca sexta*. *Biological control.* 58: 354-361.

Fabre C., Causse H., Mourey L., Koninkx J., Rivière M., Hendriks H., Puzo G., Samama J. P., Rougé P. (1998). Characterization and sugar-binding properties of arcelin-1, an insecticidal lectin-like protein isolated from kidney bean (*Phaseolus vulgaris* L. cv. RAZ-2) seeds. *Biochem. J*. 329: 551-560.

Fabrick J., Oppert C., Lorenzen M. D., Morris K., Oppert B., Jurat-Fuentes J. L. (2009). A novel *Tenebrio molitor* cadherin is a functional receptor for *Bacillus thuringiensis* Cry3Aa toxin. *J. Biol. Chem*. 284: 18401-18410.

Ferrat G., Bosmans F., Tytgat J., Pimentel C., Chagot B., Gilles N., Nakajima T., Darbon H., Corzo G. (2005). Solution structure of two insect-specific spider toxins and their pharmacological interaction with the insect voltage-gated Na+ channel. *Proteins* 59:368-379.

Fletcher J. I., Smith R., O'Donoghue S. I., Nilges M., Connor M., Howden M. E., Christie M. J., King G. F. (1997). The structure of a novel insecticidal neurotoxin, ω-atracotoxin-HV1, from the venom of an Australian funnel web spider. *Nat.Struct.Biol*. 4: 559-566.

Forster M. J. (2002).Molecular modeling in structural biology.*Micron*.33: 365–384.

Gabaldon T., Huynen M. A. (2004).Prediction of protein function and pathways in the genomeera.*Cell. Mol. Life Sci*. 61: 930–944.

Galagan J. E., Nusbaum C., Roy A., Endrizzi M. G., Macdonald P., FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A., Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R., Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W., Zimmer A., Barber R. D., Cann I., Graham D. E, Grahame D. A., Guss A. M., Hedderich R., Ingram-Smith C., Kuettner H. C., Krzycki J. A., Leigh J. A., Li W., Liu J., Mukhopadhyay B., Reeve J. N., Smith K., Springer T. A., Umayam L. A., White O., White R. H., Conway de Macario E., Ferry J. G., Jarrell K. F., Jing H., Macario A. J., Paulsen I., Pritchett M., Sowers K. R., Swanson R. V., Zinder S. H., Lander E., Metcalf W. W., Birren B. (2002). The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity. *Genome Res*. 12: 532-542.

Galitsky N., Cody V., Wojtczak A., Ghosh D., Luft J. R. (2001). Structure of the insecticidal bacterial δ-endotoxin Cry3Bb1 of *Bacillus thuringiensis*. *Act Crystallogr* 2001; D57:1101-1109.

Gentz M. C., Murdoch G., King G. F. (2010). Tandem use of selective insecticides and natural enemies for effective, reduced -risk pest management. *Biol. Control*. 52: 208-215.

Goldin, A.L. (2002). Evolution of voltage-gated Naþ channels.*J. Exp. Biol*. 205: 575-584.

Gordon D., Karbat I., Ilan N., Cohen L., Kahn R., Gilles N., Dong K., Stühmerd W., Tytgat J., Gurevitz M..(2007). The differential preference of scorpion α-toxins for insect or mammalian sodium channels: Implications for improved insect control. *Toxicon*. 49: 452–472.

Goudet C., Chi C.-W., Tytgat J. (2002). An overview of toxins and genes from the venom of the Asian scorpion *Buthus martensi* Karsch. *Toxicon*. 40: 1239–1258.

Gunning S. J, Maggio F., Windley M. J., Valenzuela S. M., King G. F., Nicholson G. M. (2008). The Janusfaced atracotoxins are specific blockers of invertebrate KCa channels. *FEBS J*. 275: 4045-4059.

Grochulski P., Masson L., Borisova S., Pusztai-Carey M., Schwartz J. L., Brousseau R., Cygler M. (1995). *Bacillus thuringiensis* CryIA(a) Insecticidal Toxin: Crystal Structure and Channel Formation. *J. Mol. Biol*. 254: 447-464.

Gruber C. W., Cemazar M., Anderson M. A., Craik D. J. (2007). Insecticidal plant cyclotides and related cystine knot toxins. *Toxicon*. 49: 561-575.

Guex, N., Peitsch, M. C. (1997). Swiss-model and the Swiss-PdbViewer: An environment for comparative protein modelling. *Electrophoresis*. 18: 2714-2723.

Guo M., Chong Y. E., Beebe K., Shapiro R., Yang X. L., Schimmel P. (2009). The C-Ala domain brings together editing and aminoacylation functions on one tRNA. *Science*. 325:744-747.

Gurevitz M., Karbat I., Cohen L., Ilan N., Kahn R., Turkov M., Stankiewicz M., Stuhmerc W., Dong K., Gordon D. (2007). The insecticidal potential of scorpion β-toxins. *Toxicon*. 49: 473-489.

Gurevitz M., Urbach D., Zlotkin E., Zilberberg N. (1991) Nucleotide sequence and structure analysis of a cDNA encoding anα-insect toxin from the scorpion *Leiurus quinquestriatus hebraeus*. *Toxicon* 29:1270-1272.

Hammock B.D., Bonning B.C., Possee R.D., Hanzlik T.N., Maeda S. (1990). Expression and effects of the juvenile hormone esterase in a baculovirus vector. *Nature*.344: 458–461.

Han S., Craig J. A., Putnam C. D., Carozzi N. B., Tainer J. A. (1999) Evolution and mechanism from structures of an ADP ribosylating toxin and NAD complex. *Nat. Struct.Biol*. 6:932-936.

Hao C. J., Xu C. G., Wang W., Chai B. F., Liang A. H. (2005). Expression of an insect excitatory toxin, BmK IT, from the scorpion, *Buthus martensii* Karsch, and its biological activity. *Biotechnol. Lett*.27:1929-1934.

Herrero S., Gechev T., Bakker P. L., Moar W. J., de Maagd R. A. (2005).*Bacillus thuringiensis* Cry1Ca-resistant *Spodoptera exigua* lacks expression of one of four aminopeptidase N genes. *BMC Genomics*.6: 96.

Howard A. F., N'guessan R., Koenraadt C. J., Asidi A., Farenhorst M., Akogbéto M., Thomas M. B., Knols B. G., Takken W. (2010). The entomopathogenic fungus *Beauveria bassiana* reduces instantaneous blood feeding in wild multi-insecticide-resistant *Culex quinquefasciatus* mosquitoes in Benin, West Africa. *Parasit. Vectors*. 15: 87.

Hua G., Jurat-Fuentes J. L., Adang M. J. (2004). Bt-R1a extracellular cadherin repeat 12 mediates *Bacillus thuringiensis* Cry1Ab binding and toxicity. *J. Biol. Chem*. 279: 28051–28056.

Hua G., Masson L., Jurat-Fuentes J. L., Schwab G., Adang M. J. (2001). Binding analyses of *Bacillus thuringiensis* Cry δ-endotoxins using brush border membrane vesicles of *Ostrinia nubilalis*. *Appl. Environ. Microbiol*.67: 872-879.

James R. M., Dickinson P. (1998). Site-Directed Mutagenesis. *Molecular Biomethods Handbook*. Pages 361-381.

Ji Y. H., Mansuelle P., Terakawa S., Kopeyan C., Yanaihara N., Hsu K., Rochat H.(1996). Two neurotoxins (Bmk I and Bmk II) from the venom of the scorpion *Buthus martensi* Karsch: purification, amino acid sequences and assessment of specific activity. *Toxicon*.34: 987-1001.

Jablonsky M. J., Jackson P. L., Krishna N. R. (2001). Solution structure of an insect-specific neurotoxin from the new world scorpion*Centruroides sculpturatus* Ewing. *Biochemistry*. 40: 8273-8282.

Johnson J. H., Bloomquist J. R., Krapcho K. J., Kral R. M. Jr, Trovato R., Eppler K. G., Morgan T. K., DelMar E. G. (1998). Novel insecticidal peptides from *Tegenaria agrestis* spider venom may have a direct effect on the insect central nervous system. *Arch Insect Biochem Physiol*. 38: 19-31.

Kamita S. G., Hinton A. C., Wheelock C. G., Wogulis M. D., Wilson D. K., Wolf N. M., Stok J. E., Hock B., Hammock B. D. (2003). Juvenile hormone (JH) esterase: why are you so JH specific? *Insect Biochemistry and Molecular Biology*. 33: 1261–1273.

Kanchiswamy C. N., Takahashi H., Quadro S., Maffei M. E., Bossi S., Bertea C., Zebelo S. A., Muroi A., Ishihama N., Yoshioka H., Boland W., Takabayashi J., Endo Y., Sawasaki T., Arimura G. (2010). Regulation of *Arabidopsis* defense responses against Spodoptera littoralis by CPK-mediated calcium signaling.*BMC Plant.Biol*. 10: 97.

Karbat I., Frolow F., Froy O., Gilles N., Cohen L., Turkov M., Gordon D., Gurevitz M. (2004). Molecular Basis of the High Insecticidal Potency of Scorpion α toxins. *The Journal of Biological Chemistry*.279: 31679–31686.

Karbat I., Turkov M., Cohen L., Kahn R., Gordon D., Gurevitz M., Frolow F. (2007). X-ray structure and mutagenesis of the scorpion depressant toxin LqhIT2 reveals key determinants crucial for activity and anti-insect selectivity. *J. Mol. Biol*. 366: 586-601.

Karlova R., Weemen-Hendriks M., Naimov S., Ceron J., Dukiandjiev S., de Maagd R. (2005).*Bacillus thuringiensis*δ-endotoxin Cry1Ac domain III enhances activity against *Heliothis virescens* in some, but not all Cry1-Cry1Ac hybrids. *J. Invertebr. Pathol*. 88: 169–172.

Keszthelyi S., Pál-Fám F., Kerepesi I. (2011). Effect of cotton bollworm (*Helicoverpa armigera* Hübner) caused injury on maize grain content, especially regarding to the protein alteration. *Acta Biol. Hung*. 62: 57-64.

King G. F., Gentz M. C., Escoubas P., Nicholson G. M. (2008). A rational nomenclature for naming peptide toxins from spiders and other venomous animals. *Toxicon*. 52: 264-276.

King G. F., Tedford H. W., Maggio F. (2002). Structure and function of insecticidal neurotoxins from Australian funnel web spiders. *Toxin Reviews*. 21: 361-389.

Krimm I., Gilles N., Sautière P., Stankiewicz M., Pelhate M., Gordon D., Lancelin J. M. (1999). NMR structures and activity of a novel α-like toxin from the scorpion *Leiurus quinquestriatus* hebraeus. *J. Mol. Biol.* 285:1749-1763.

Lechner S., Mayr R., Francis K. P., Prüss BM, Kaplan T., Wiessner-Gunkel E., Stewart G. S., Scherer S. (1998).*Bacillus weihenstephanensis* sp. *nov*. is a new psychrotolerant species of the Bacillus cereus group. *Int. J. Syst. Bacteriol*.48: 1373-1 382.

Lee W., Moore C. H., Watt D. D., Krishna N. R. (1994). Solution structure of the variant-3 neurotoxin from *Centruroides sculpturatus* Ewing. *Eur. J. Biochem*.218: 89-95.

Lee D., Redfern O., Orengo C. (2007). Predicting protein function from sequence and structure. *Nat. Rev. Mol. Cell. Biol*. 8: 995-1005.

Li C., Guan R. J., Xiang Y., Zhang Y., Wang D. C. (2005). Structure of an excitatory insect-specific toxin with an analgesic effect on mammals from the scorpion *Buthus martensii* Karsch. *Acta Crystallogr. D Biol. Crystallogr*. 61(Pt 1):14-21.

Li J., Carroll J., Ellar D. J. (1991).Crystal Structure of Insecticidal δ-endotoxin from *Bacillus thuringiensis* at 2.5 Å Resolution.*Nature*.353:815-821.

Liang S. (2004). An overview of peptide toxins from the venom of the Chinese bird spider *Selenocosmia huwena* Wang [=*Ornithoctonus huwena* (Wang)] *Toxicon*.43: 575-585.

Likitvivatanavong S., Katzenmeier G., Angsuthanasombat C. (2006). Asn183 in α5 is essential for oligomerisation and toxicity of the *Bacillus thuringiensis* Cry4Ba toxin. *Archives of Biochemistry and Biophysics*.445: 46–55.

Liu Y. J., Cheng C. S., Lai S. M., Hsu M. P., Chen C. S., Lyu P. C. (2006). Solution structure of the plant defensin VrD1 from mung bean and its possible role in insecticidal activity against bruchids. *Proteins*.63:777-786.

López-Pazos S. A., Cerón J. A. (2007). Three-dimensional structure of *Bacillus thuringiensis* toxins: a review *Acta. Biol. Colomb*.12: 19-32.

López-Pazos S. A., Cortázar J. E., Cerón J. A. (2009a). Cry1B and Cry3A are active against *Hypothenemus hampei* Ferrari (Coleoptera: Scolytidae*). J. Invertebr. Pathol*.101: 242-245.

López-Pazos S. A., Martínez J. W., Castillo A. X., Cerón Salamanca J. A.(2009b). Presence and significance of *Bacillus thuringiensis* Cry proteins associated with the Andean

weevil *Premnotrypes vorax* (Coleoptera: Curculionidae). *Rev. Biol. Trop*. 57: 1235-1243.

López-Pazos S. A., Rojas Arias A. C., Ospina S. A., Cerón J. (2010). Activity of *Bacillus thuringiensis* hybrid protein against a lepidopteran and a coleopteran pest. *FEMS Microbiol. Lett*. 302: 93-98.

Loris R., Hamelryck T., Bouckaert J., Wyns L. (1998). Legume lectin structure.*Biochim. Biophys*. Acta.1383: 9–36.

Lykidis A., Pérez-Pantoja D., Ledger T., Mavromatis K., Anderson I. J., Ivanova N. N., Hooper S. D., Lapidus A., Lucas S., González B., Kyrpides N. C. (2010). The complete multipartite genome sequence of *Cupriavidus necator* JMP134, a versatile pollutant degrader. *PLoS One*. 5: e9729.

Martínez-Ramírez A. C., González-Nebauer S., Escriche B., Real M. D. (1994). Ligand blot identification of a *Manduca sexta* midgut binding protein specific to three *Bacillus thuringiensis* CryIA-type ICPs. *Biochem. Biophys. Res. Commun*.201: 782–787.

Martínez W., Uribe D., Cerón J. (2003). Efecto tóxico de proteínas Cry1 de *Bacillus thuringiensis* sobre larvas de *Tecia solanivora* (Lepidoptera: Gelechiidae). *Rev. Colomb. Entomol*.29: 89–93.

Martins E. S., Monnerat R. G., Queiroz P. R., Dumas V. F., Braz S. V., de Souza Aguiar R. W., Gomes A. C., Sánchez J., Bravo A., Ribeiro B. M. (2010). Midgut GPI-anchored proteins with alkaline phosphatase activity from the cotton boll weevil (Anthonomus grandis) are putative receptors for the Cry1B protein of *Bacillus thuringiensis*. *Insect. Biochem. Mol. Biol*. 40: 138-145.

Miller E. A., Lee M. C. S., Atkinson A. H. O., Anderson M. A. (2000). Identification of a novel four-domain member of the proteinase inhibitor II family from the stigmas of *Nicotiana alata*. *Plant. Mol. Biol*. 42: 329-333.

Monnerat R., Martins E., Queiroz P., Ordúz S., Jaramillo G., Benintende G., Cozzi J., Real M. D., Martinez-Ramirez A., Rausell C., Cerón J., Ibarra J. E., Del Rincon-Castro M. C., Espinoza A. M., Meza-Basso L., Cabrera L., Sánchez J., Soberon M., Bravo A. (2006). Genetic variability of *Spodoptera frugiperda* Smith (Lepidoptera: Noctuidae) populations from Latin America is associated with variations in susceptibility to *Bacillus thuringiensis* Cry toxins. *Appl. Environ. Microbiol*. 72:7029-7035.

Moran Y., Cohen L., Kahn R., Karbat I., Gordon D., Gurevitz M. (2006). Expression and Mutagenesis of the Sea Anemone Toxin Av2 Reveals Key Amino Acid Residues Important for Activity on Voltage-Gated Sodium Channels.*Biochemistry*. 45: 8864-8873.

Moran Y., Gordon D., Gurevitz M. (2009). Sea anemone toxins affecting voltage-gated sodium channels-molecular and evolutionary features. *Toxicon*. 54: 1089–1101.

Morse R. J., Yamamoto T., Stroud R. M. (2001). Structure of Cry2Aa suggests an unexpected receptor binding epitope. *Structure*.9:409-417.

Mourey L., Pédelacq J. D., Birck C., Fabre C., Rougé P., Samama J. P. (1998). Crystal structure of the arcelin-1 dimer from *Phaseolus vulgaris* at 1.9-A resolution.*J. Biol. Chem*. 273:12914-12922.

Naganuma M., Sekine S., Fukunaga R., Yokoyama S. (2009). Unique protein architecture of alanyl-tRNA synthetase for aminoacylation, editing, and dimerization. *Proc. Natl. Acad. Sci*. USA.106: 8489-8494.

Nakanishi K., Yaoi K., Shimada N., Kadotani T., Sato R. (1999).*Bacillus thuringiensis* insecticidal Cry1Aa toxin binds to a highly conserved region of aminopeptidase N in the host insect leading to its evolutionary success. *Biochim. Biophy. Acta*.1432: 57-63.

Nakasu E. Y., Firmino A. A., Campos Dias S., Lima Rocha T., Batista Ramos H., Ramos de Oliveira G., Lucena W., Ribeiro da Silva Carlini C. R., Grossi de Sá M. F. (2010). Analysis of Cry8Ka5-binding proteins from *Anthonomus grandis* (Coleoptera: Curculionidae) midgut.*J Invertebr. Pathol*. 104: 227-230.

Ochoa-Campuzano C., Real M. D., Martinez-Ramirez A. C., Bravo A., Rausell C. (2007). An ADAM metalloprotease is a Cry3Aa *Bacillus thuringiensis* toxin receptor.*Biochem. Biophys. Res. Comm*. 362: 437-442.

Oerke E.-C., Dehne H.-W. (2004). Safeguarding production-losses in major crops and the role of crop protection. Crop Protection. 23: 275-285.

Ohashi S., Miyashita H., Okada N., Lemura T., Watanabe T., Kobayashi M. (2008). Unique photosystems in *Acaryochloris marina*. Photosynth.Res. 98:141-149.

Omecinsky D. O, Holub K. E., Adams M. E., Reily M. D. (1996). Three-dimensional structure analysis of μ-agatoxins: further evidence for common motifs among neurotoxins with diverse ion channel specificities. *Biochemistry*.35: 2836-2844.

Oren D. A., Froy O., Amit E., Kleinberger-Doron N., Gurevitz M., Shaanan B. (1998). An excitatory scorpion toxin with a distinctive feature: an additional α helix at the C terminus and its implications for interaction with insect sodium channels. *Structure*. 6:1095-1103.

Pai A., Bernasconi G. (2008). Polyandry and female control: the red flour beetle *Tribolium castaneum* as a case study. *J. Exp. Zool. B. Mol. Dev. Evol*. 310:148-159.

Paramasivan R., Sivaperumal R., Dhananjeyan K. J., Thenmozhi V., Tyagi B. K.(2006). Prediction of 3-dimensional structure of salivary odorant-binding protein-2 of the mosquito Culex quinquefasciatus, the vector of human lymphatic filariasis. *In Silico Biol*. 7: 1-6.

Park Y., Abdullah M. A., Taylor M. D., Rahman K., Adang M. J. (2009). Enhancement of *Bacillus thuringiensis* Cry3Aa and Cry3Bb Toxicities to Coleopteran Larvae by a Toxin-Binding Fragment of an Insect Cadherin.*Appl. Environ. Microbiol*.75: 3086-3092.

Pava-Ripoll M., Posada F. J., Momen B., Wang C., St Leger R. (2008). Increased pathogenicity against coffee berry borer, *Hypothenemus hampei* (Coleoptera: Curculionidae) by *Metarhizium anisopliae* expressing the scorpion toxin (AaIT) gene.*J. Invertebr. Pathol*.99: 220–226.

Pechan T., Cohen A., Williams W. P., Luthe D. S. (2002). Insect feeding mobilizes a unique plant defense protease that disrupts the peritrophic matrix of caterpillars. *Proc. Natl. Acad. Sci. U S A*. 99: 13319-13323.

Petrey D., Honig B. (2009). Is protein classification necessary? Toward alternative approaches to function annotation.*Curr Opin Struct Biol* 19:363–368.

Peumans W. J., Van Damme E. J. (1995a). Lectins as plant defence proteins. *Plant.Physiol*. 109: 347–352.

Peumans W. J., Van Damme E. J. (1995b).Role of lectins in plant defense.*Histochem.J*. 27: 253–271.

Philpott, M.L., Hammock, B.D. (1990). Juvenile hormone esterase is a biochemical anti-juvenile hormone agent. *Insect Biochemistry* 20: 451–459.

Pigott C.R., Ellar D.J. (2007). Role of Receptors in *Bacillus thuringiensis* Crystal Toxin Activity.*Microbiol. Mol. Biol. Rev*. 71: 255–281.

Possani L. D., Becerril B., Delepierre M., Tytgat J. (1999).Scorpion toxins specific for Na+-channels.*Eur. J. Biochem*. 264: 287-300.

Qin Y., Ying S. H., Chen Y., Shen Z. C., Feng M. G. (2010). Integration of insecticidal protein Vip3Aa1 into *Beauveria bassiana* enhances fungal virulence to *Spodoptera litura* larvae by cuticle and per Os infection. *Appl. Environ. Microbiol*.76:4611-4618.

Ralph S. G., Yueh H., Friedmann M., Aeschliman D., Zeznik J. A., Nelson C. C., Butterfield Y. S., Kirkpatrick R., Liu J., Jones S. J., Marra M. A., Douglas C. J., Ritland K., Bohlmann J. (2006). Conifer defence against insects: microarray gene expression profiling of Sitka spruce (*Picea sitchensis*) induced by mechanical wounding or feeding by spruce budworms (*Choristoneura occidentalis*) or white pine weevils (*Pissodes strobi*) reveals large-scale changes of the host transcriptome. *Plant. Cell. Environ*. 29: 1545-1570.

Ramírez R. (2009). La broca del café en Líbano. Impacto socioproductivo y cultural en los años 90.*Revista de Estudios Sociales*.32: 158-171.

Rayapuram C., Baldwin I. T. (2008). Host-plant-mediated effects of Nadefensin on herbivore and pathogen resistance in *Nicotiana attenuata*. *BMC Plant.Biol*. 8: 109.

Ribas de Pouplana L., Schimmel P. (2000). A view into the origin of life: aminoacyl-tRNA synthetases *Cell. Mol. Life Sci*. 57: 865–870.

Rodríguez de la Vega R. C., Schwartz E. F., Possani L. D. (2010). Mining on scorpion venom biodiversity. *Toxicon*. 56: 1155–1161.

Roh J. Y., Choi J. Y., Li M. S., Jin B. R., Je Y. H.(2007). *Bacillus thuringiensis* as a Specific, Safe, and Effective Tool for Insect Pest Control. *J. Microbiol. Biotechnol*. 17: 547–559.

Rosengren K. J., Daly N. L., Plan M. R., Waine C., Craik D. J. (2003). Twists, knots, and rings in proteins. Structural definition of the cyclotide framework. *J. Biol. Chem*. 278:8606-8616.

Sayed A., Nekl E. R., Siqueira H. A., Wang H. C., Ffrench-Constant R. H., Bagley M., Siegfried B. D. (2007). A novel cadherin-like gene from western corn rootworm, *Diabrotica virgifera virgifera* (Coleoptera: Chrysomelidae), larval midgut tissue. *Insect.Mol. Biol*. 16: 591–600.

Schirra H. J., Anderson M. A., Craik D. J. (2008). Structural refinement of insecticidal plant proteinase inhibitors from *Nicotiana alata*. *Protein. Pept. Lett*.15:903-909.

Schirra H. J., Scanlon M. J., Lee M. C., Anderson M. A., Craik D. J. (2001). The solution structure of C1-T1, a two-domain proteinase inhibitor derived from a circular precursor protein from *Nicotiana alata*. *J. Mol. Biol*. 306:69-79.

Schrank A., Vainstein M. H. (2010). *Metarhizium anisopliae* enzymes and toxins.*Toxicon*. 56: 1267-1274.

Schnepf E., Crickmore N., Van Rie J., Lereclus D., Baum J., Feitelson J., Zeigler D.R., Dean D.H. (1998). *Bacillus thuringiensis* and Its Pesticidal Crystal Proteins. *Microbiol. Mol. Biol. Rev*. 62: 775-806.

Shah P. A., Pell J. K. (2003). Entomopathogenic fungi as biological control agents.*Appl. Microbiol. Biotechnol*. 61: 413-423.

Shu Q., Lu S. Y., Gu X. C., Liang S. P. (2002). The structure of spider toxin huwentoxin-II with unique disulfide linkage: evidence for structural evolution. *Protein Sci*. 11:245-252.

Skinner W. S., Dennis P. A., Li J.P., Summerfelt R.M., Carney R. L., Quistad G. B. (1991). Isolation and Identification of Paralytic Peptides from Hemolymph of the Lepidopteran Insects Manduca sexta, Spodoptera exigua, and *Heliothis virescens*. *J. Biol. Chem.* 266: 12873-12877.

St. Leger R. J., Wang C. (2010). Genetic engineering of fungal biocontrol agents to achieve greater efficacy against insect pests.*Appl. Microbiol. Biotechnol*. 85:901–907.

Suckling D. M., Brockerhoff E. G. (2010). Invasion biology, ecology, and management of the light brown apple moth (Tortricidae). *Annu. Rev. Entomol*. 55:285-306.

Szolajska E., Poznanski J., López M., Michalik J., Gout E., Fender P., Bailly I., Dublet B., Chroboczek J. (2004). Poneratoxin, a neurotoxin from ant venom. Structure and expression in insect cells and construction of a bio-insecticide.*Eur. J. Biochem*.271: 2127–2136.

Tedford H. W., Sollod B. L., Maggio F., King G. F.(2004). Australian funnel-web spiders: master insecticide chemists. *Toxicon*. 43:601-618.

Tindall K. V., Stewart S., Musser F., Lorenz G., Bailey W., House J., Henry R., Hastings D., Wallace M., Fothergill K. (2010). Distribution of the long-horned beetle, *Dectes texanus*, in soybeans of Missouri, Western Tennessee, Mississippi, and Arkansas.*J. Insect.Sci*. 10:178.

Tugarinov V., Kustanovich I., Zilberberg N., Gurevitz M., Anglister J. (1997). Solution structures of a highly insecticidal recombinant scorpion α-toxin and a mutant with increased activity. *Biochemistry*.36: 2414-2424.

van Frankenhuyzen K. (2009).Insecticidal activity of *Bacillus thuringiensis* crystal proteins. *J. Invertebr. Pathol*.101: 1–16.

Venter, J.C., et.al.(2004). Environmental Genome Shotgun Sequencing of the Sargasso Sea.*Science*.304: 66-74.

Vaishnavi C. (2010). Clinical spectrum & pathogenesis of *Clostridium difficile* associated diseases. *Indian. J. Med. Res*. 131: 487-499.

Vassilevski A., Kozlov S. A., Grishin E. V. (2009). Molecular Diversity of Spider Venom. *Biochemistry (Moscow)*.74: 1505-1534.

Wang X. H., Connor M., Wilson D., Wilson H. I., Nicholson G. M., Smith R., Shaw D., Mackay J. P., Alewood P. F., Christie M. J., King G. F. (2001).Discovery and structure of a potent and highly specific blocker of insect calcium channels.*J. Biol. Chem*. 276: 40306-40312.

Wang L., Wang S., Li Y., Paradesi M. S. R., Brown S. J. (2007). BeetleBase: the model organism database for *Tribolium castaneum*. Nucleic. Acids. Res. 35: D476–D479.

Warmke, J. W., Reenan, A. G. R., Wang, P., Qian, S., Arena, J. P., Wang, J., Wunderler, D., Liu, K., Kaczorowski, G. J., Van der Ploeg, L. H. T., Ganetzky, B., and Cohen, C. J. (1997). Functional expression of Drosophila para sodium channels.Modulation by membrane protein TipE and toxin pharmacology.*J. Gen. Physiol*. 110: 119-133.

Whetstone P. A., Hammock B. D. (2007).Delivery methods for peptide and protein toxins in insect control.*Toxicon*. 49: 576–596.

Wolfersberger M. G., Luethy P., Maurer A., Parenti P., Sacchi F. V., Giordana B., Hanozet G. M. (1987). Preparation and partial characterization of amino acid transporting

brush border membrane vesicles from the larval midgut of the cabbage butterfly (*Pieris brassicae*). *Comp. Biochem.Physiol*. 86:301–308.

Yamaji N., Sugase K., Nakajima T., Miki T., Wakamori M., Mori Y., Iwashita T. (2007). Solution structure of agelenin, an insecticidal peptide isolated from the spider *Agelena opulenta*, and its structural similarities to insect-specific calcium channel inhibitors. *FEBS Lett*.581: 3789-3794.

Yu X. Q., Prakash O., Kanost M. R. (1999).Structure of a paralytic peptide from an insect, *Manduca sexta*.*J. Pept. Res*. 54:256-261.

Zilberberg N., Froy O., Loret E., Cestele S., Arad D., Gordon D., Gurevitz M. (1997). Identification of structural elements of a scorpion α-neurotoxin important for receptor site recognition.*J. Biol. Chem*. 272: 14810–14816.

Zlotkin E., Fishman Y., Elazar M. (2000).AaIT: from neurotoxin to insecticide.*Biochimie*.82: 869-881.

Site-Directed Mutagenesis as Applied to Biocatalysts

Juanita Yazmin Damián-Almazo and Gloria Saab-Rincón
Departamento de Ingeniería Celular y Biocatálisis,
Instituto de Biotecnología, Universidad Nacional Autónoma de México,
Cuernavaca, Morelos,
México

1. Introduction

Enzymes are biological catalysts responsible for supporting almost all of the chemical reactions in living organisms. Their *activities*, *specificities* and *selectivities* make them attractive as biocatalysts for a wide variety of industries. Examples are agrochemicals, detergents, starch, textiles, personal care, pulp and paper, food processing, and animal feed. The chemo-, enantio- and regioselectivities of biological catalysts are hallmarks that make them especially attractive for use in the synthesis of fine chemicals and pharmaceutical intermediates. They are a viable alternative to chemical synthesis, which is usually characterized by low yield and the accumulation of undesirable secondary products. The imminent decrease in the use of fossil fuels has turned attention to new enzyme-based developments for the production of biofuels (*e.g.*, biodiesel) that use renewable raw materials (Cherry & Fidantsef, 2003).

Nevertheless, natural enzymes are often not optimal for use in industrial conditions. It is usually necessary to change the conditions of the process or, most commonly, to alter one or more of the properties of the enzyme. Desirable changes in the enzyme are those that affect substrate specificity, expression level, solubility, stability, activity, selectivity, or thermal stability. Other desired effects could include tolerance to organic solvents or to extreme pH values (Hibbert et al., 2005; Turner, 2009).

Protein engineering usually involves the modification of amino acid sequences at the DNA sequence level by means of chemical or genetic techniques. The resultant protein is then tested for novel, optimal or improved physical and/or catalytic properties (Ulmer, 1983). There are two different basic approaches to engineer proteins, although it is common to combine both approaches for better results.

a. *Rational design*. Mutations are introduced at specific places in the protein-encoding gene. Positions to mutagenize are based on the knowledge of possible relationships of sequence, structure, function and/or the catalytic mechanism of the protein. Recently computational predictive algorithms have been developed and used to preselect promising target sites (Bolon & Mayo, 2001; Kaplan & DeGrado, 2004; Kuhlman et al., 2003; Meiler & Baker, 2006; Pavelka et al., 2009; Zanghellini et al., 2006). However, a

deep knowledge of the structure and energy functions is required in order to predict the changes required to modify some parameter of the enzyme. This is especially true if one wishes to change the reaction mechanism.

b. *Directed evolution*. This approach involves repeated cycles of random mutagenesis of and/or recombination with variants of the gene to create a library of genes with slightly different sequences. The enzyme variants thus obtained are submitted to genetic selection or to high-throughput screening to identify those enzyme variants with improvements in the desired property (Stemmer, 1994a; b; Zhao et al., 1998). Directed evolution has been demonstrated to be a very powerful technique, especially for increasing stability (Giver et al., 1998; Ladenstein & Antranikian, 1998; Song & Rhee, 2000; Uchiyama et al., 2000; Zhao & Arnold, 1999) or to change the specificity of an enzyme (Castle et al., 2004; Cohen et al., 2004; Christians et al., 1999; Joerger et al., 2003; Jurgens et al., 2000; Levy & Ellington, 2001; Matsumura & Ellington, 2001; Sakamoto et al., 2001; Song et al., 2002; Zhang et al., 1997). It is a particularly useful approach, since no structural or mechanistic information is required. In many cases, changes that contribute to the improved properties are far from the active sites. They would not have been targeted by a rational strategy.

Both techniques have strengths, but they also have some limitations. For this reason, it is often common to find rational design work combined with directed evolution. It may be desirable to tune some properties of a designed protein (Savile et al., 2010; Siegel et al., 2010) or randomization may be directed to specific regions of a protein that were identified in the design process.

This review summarizes the most common strategies used to identify possible targets for site-directed mutagenesis to enhance biocatalysis. We included sequence- and structure-based strategies for generating enzymes with desired properties. To illustrate a number of the points discussed above, special attention was paid to the site-directed mutagenesis of glycosyl hydrolases. We also used the modification of alpha amylases as a case study. We have described the sequence-based mutagenesis approach that was used to change the transglycosylation/hydrolysis ratio of alpha amylases. Residues involved in the hydrophobicity and electrostatic environment of the active site were identified by sequence and structural alignments with other glycosyltransferases. As a result, certain residues were targeted for mutagenesis. We also used a multiple sequence alignment and structural information in an approach to reduce the hydrolytic activity of the alpha amylase from *Thermotoga maritima*, while increasing its alcoholytic activity. Unlike the wild-type parent, the modified enzyme was able to synthesize alkyl-glucosides.

2. Approaches for selecting targets to mutagenize

Any biochemical, structural, or protein sequence information may be useful for identifying residues that may influence a desired enzyme property. The information may indicate changes that increase or decrease the overall fitness of the enzyme.

A common approach is to focus on regions or positions that may be directly related to the catalytic property. For example, amino acid residues that alter substrate specificity or selectivity are commonly non-conserved residues. They are often in close contact with catalytic residues in or near the active site, cofactors or substrates (Morley & Kazlauskas,

2005; Paramesvaran et al., 2009; Park et al., 2005). Another approach is the identification of sequence motifs that are thought to have been conserved during evolution (Saravanan et al., 2008). In contrast, residues thought to be involved in thermostabilization are spread throughout the entire sequence. Each such residue is thought to make a small contribution to thermostability. However, the additive effect can be significant. For this reason, random mutagenesis is a powerful tool for achieving protein stabilization. However, some features that are known to contribute to protein stability can be implemented by site-directed mutagenesis strategies. These include the introduction of additional disulfide bridges (Mansfeld et al., 1997); decrease of loop entropy by replacement of some amino acid residues to P or by the shortening of loops (Nagi & Regan, 1997); change of α-helix propensity by mutations to replace G residues (low α-helix propensity) to A residues (high α–helix proepnsity); or by the introduction of salt bridges to increase electrostatic interactions in the protein (Kumar et al., 2000; Lehmann & Wyss, 2001; Spector et al., 2000).

2.1 Sequence-based mutagenesis

2.1.1 Alanine scanning

Alanine scanning is a method used to determine the contribution of the side-chains of specific residues in a protein. Substitution of residues with alanine removes all side chain atoms past the β-carbon, without introducing additional conformational changes into the protein backbone. Although mutagenesis by alanine scanning can be a laborious method (because each alanine-mutated protein must be constructed, expressed and analyzed separately), it has nevertheless been useful for the study of interactions at protein-protein interfaces or for the identification of residues involved in substrate recognition (Gibbs & Zoller, 1991), protein stability (Blaber et al., 1995), or binding (Ashkenazi et al., 1990; Cunningham & Wells, 1989).

Alternatives to conventional alanine scanning are computational methods for modeling alanine-scanning mutants. This approach has proven to be useful for predicting active-site residues important for activity (Funke et al., 2005) and to identify amino acid residues important in protein-protein interactions (Kortemme et al., 2004).

2.1.2 Protein sequence alignment

Nature has had the opportunity to explore the protein sequence space through millions of years of evolution. Genetic drift is thought to be the driving force that is responsible for the sequence diversity observed today. However, residues that are indispensable to function and/or stability have been maintained by selective pressure. Multiple sequence alignments are useful tools for identifying positions that are unchangeable in a protein. They will also identify those regions with the plasticity to allow multiple changes. Briefly, when residues with a common evolutionary origin or having structural or functional equivalence are arranged so that the highly conserved residues are aligned, their alignment serves as an anchor for the alignment of the sequences in a set. Analysis of position-specific residue usage (residue profiles) gives information about amino acid conservation or variability at each position. When a multiple sequence alignment is combined with phylogenetic information, it is possible to explore ancestral relationships among groups of homologous protein sequences. It is also useful for identifying important amino acids that probably cannot be modified.

2.1.2.1 Correlating amino acid sequence patterns to specific properties

An approach for identifying residues that may be functionally relevant is to correlate an enzyme property with the amino acid patterns observed in a multiple sequence alignment. For example, comparison of more stable proteins with less stable ones is a strategy for identifying possible thermostabilizing residues (Ditursi et al., 2006; Gromiha et al., 1999; Kumar & Nussinov, 2001; Perl et al., 2000).

Sequence patterns can also be used to identify the determinants of specificity. Good examples are the attempts to change cofactor specificity in dehydrogenases to NAD^+, since NAD^+ is considerably less energetically demanding for the cell to make than is $NADP^+$ (Flores & Ellington, 2005; Kristan et al., 2007; Rodríguez-Zavala, 2008; Rosell et al., 2003).

Even distant mutations can significantly affect the properties of an active site. They may alter slightly the geometry, electrostatic properties or dynamics of amino acids in the active site. Distant residues that are important for their interactions with the active site may be seen as conserved in a multiple sequence alignment. Multiple sequence alignments have also revealed that some residues are infrequent in a sequence; but nevertheless are frequently adjacent. Such cluster-forming residues have probably coevolved. These "protein sectors" are often critical for specific functional roles, including substrate binding, stability, allosteric regulation or catalytic activity (Halabi et al., 2009).

2.1.2.2 Consensus sequence

The method of using a consensus sequence is based on the assumption that, in an amino acid sequence alignment of homologous proteins, the consensus amino acid at a given position contributes more to the stability or the function of the protein than does a non-consensus residue. This assumption is based on the belief that a consensus sequence may closely mimic the sequence of an ancestral protein. One hypothesis posits that many proteins were originally thermophilic or hyperthermophilic (Di Giulio, 2003). Under this premise, the consensus sequence has been used to improve the thermostability of several enzymes. This was achieved by mutation of several residues towards the consensus sequence obtained from a multiple sequence alignment. There are numerous examples in which this approach has been used to increase the thermostability of proteins. Some proteins reconstruct the complete consensus sequence (Lehmann et al., 2000; Sullivan et al., 2011). In others, point mutations were used to identify the residues that increased stability. They were then combined to increase the thermostability of the protein (Maxwell & Davidson, 1998; Nikolova et al., 1998; Yamashiro et al., 2010).

Similarly, Jochens et al. (2010) showed the improvement of an enzyme property by mutagenizing the codon for a residue to codons for those amino acids that appear frequently in natural enzymes at identical positions. Evolution probably selected these residues. They are unlikely to perturb the folding or the function of the protein. In contrast, absent and rarely occurring residues are the ones that are probably not allowed. Their rareness suggests that they may be deleterious to the protein. This approach was used to improve the activity and enantioselectivity of an esterase from *Pseudomonas fluorescens* (PFE). The amino acid distribution at four positions near the active site of PFE, previously reported to influence the enantioselectivity of the enzyme, was determined by a structure-guided multiple-sequence alignment of 171 esterases generated by the 3DM database (Kuipers et al., 2010). A library was created by site-directed mutagenesis of the coding regions for the four

active site positions in PFE. Substitutions were limited to frequently occurring residues. Almost all mutants in the library showed significantly improved activity towards a commonly used esterase substrate. Moreover, one mutant had its specific activity enhanced 240-fold relative to that of the wild-type enzyme. The mutant also exhibited substantially higher enantioselectivity in the hydrolysis of 3-phenyl butyric acid p-nitrophenyl ester (E=80) compared to that of the almost nonselective wild-type enzyme (E=3.2) (Jochens & Bornscheuer, 2010).

2.1.2.3 Design of ancestral proteins

As mentioned above, one hypothesis suggests that ancestral proteins were able to withstand the harsh conditions prevalent on earth at that time (Di Giulio, 2001). In addition to their thermostability, ancestral enzymes may have been promiscuous with respect to substrates. The evolution theory of proteins holds that current proteins evolved from low-specificity ancestral proteins. Because of their low specificity, the ancestral proteins evolved to become more efficient at using specific substrates. Thus, the reconstruction of ancestral sequences from multiple sequence alignments and phylogenetic trees may provide the opportunity to change enzyme specificity. Several methods based on this approach have been reported. Some of them are given below.

In the *Ancestral Library* method, all residues located close to or within the enzyme´s active site are mutated to residues predicted by phylogenetic analysis and ancestral inference. The substitutions are those residues found in the hypothetical proteins at various nodes and branches of the evolutionary trajectories of a given enzyme family. They do not reflect the entire diversity seen in existing family members (Alcolombri et al., 2011).

Alcolombri and coworkers (2011) used serum paraoxonases and cytosolic sulfotransferases (SULTs) as models. In order to promote changes in substrate specificity, they constructed ancestral libraries of enzymes. Their mutagenesis was directed to residues near or within the active site of an enzyme. From a phylogenetic tree, the most probable ancestral sequences were obtained for all nodes. Using these sequences as templates and the three-dimensional structure of the enzyme, residues in and near the active site were located. The ancestral residues were identified, and the relevant altered enzymes constituted a library of mutants. After activity screening, several variants with different activities and specificities were identified. Some mutants had up to 50-fold higher activity than the activity of the starting enzyme.

REAP (Reconstructing Evolutionary Adaptive Paths) analysis uses phylogeny to identify mutations in gene sequences that are thought to have emerged from a common universal ancestor during functional divergence. The findings are used to generate focused and functionally enriched enzyme libraries (Chen et al., 2010).

REAP was implemented to identify differences in the sequence of promiscuous viral polymerases and non-viral polymerases. The differences may be responsible for the functional divergence without loss of catalytic activity. Sequence alignments and a phylogenetic tree of 719 polymerases were constructed. Ancestral proteins sequences were inferred from the collection of sequences at the nodes of the tree. REAP identified sites that may have changed during the separation of viral and non-viral polymerases. In one example, mutations of the residues identified by REAP analysis for the DNA polymerase of *Thermus aquaticus* yielded 8 mutants that showed a change in the substrate specificity for unnatural dNTP´s (Chen et al., 2010).

Similarly, the *Evolutionary Trace* method correlates evolutionary variations within a gene of interest with divergence in the phylogenetic tree of that sequence family. This method has been shown to reveal the functional importance of residues (Lichtarge et al., 1996; Lichtarge et al., 2003).

2.1.2.4 SCA (Statistical Coupling Analysis)

Well-separated mutations can significantly affect the activity, specificity, or enantioselectivity of an enzyme by slightly altering the geometry, electrostatic properties or dynamics of amino acids in the active site. Moreover, it is known that physically contiguous residues form "protein sectors" that can be critical for specific functional roles, including substrate binding, stability, allosteric regulation and catalytic activity (Halabi et al., 2009).

Statistical Coupling Analysis (SCA) is a method that estimates the co-evolution between pairs of amino acids in the multiple sequence alignment of a protein family. SCA shows that proteins can be divided into "protein sectors." In several different proteins, the sectors correspond to amino acids that are physically contiguous. These amino acids often underlie various aspects of function, allosteric regulation, binding, catalytic specificity, and/or fold stability. An application of SCA revealed networks of small subsets of residues that link distant functional sites and cooperate in allosteric communication (Suel et al., 2003).

2.2 Structure-based mutagenesis

The evolution of proteins involves mutations of single residues, insertions, deletions (Pascarella & Argos, 1992), gene duplications, fusions, exon duplications and shuffling (Grishin, 2001). Such changes, which accumulate over time, make the identification of sequence similarities very difficult. However, structure is more preserved than sequence and can be used as an evidence of homology among proteins. Comparative analyses of protein sequences and structures are important approaches for the identification of structural, evolutionary and functional relationships between proteins.

The rapidly growing number of protein structures in the Protein Database (PDB) and advances in homology modeling are of great value for generating structural alignments. In general, these methods provide a measure of structural similarity between proteins. They also generate an alignment that defines the residues that have structurally equivalent positions in the proteins being compared. Homology modeling can be done even when no sequence similarity is detected. Based on structural alignments, it is possible to identify residues in direct contact with the substrate or near the active-site cavity. A more complex analysis can even look into enzyme locations that are far from the active site, but are part of a network of interactions that hold the active site together. The residues at these locations can be targeted for mutagenesis. There is a server called HotSpot Wizard that combines information from extensive sequence and structure database searches with functional data to create a mutability map for a target protein. This approach was validated by comparing "hot spot" predictions with mutations extracted from the literature (Pavelka et al., 2009).

2.2.1 Site-directed Saturation Mutagenesis (SDSM)

The Site-directed Saturation Mutagenesis (SDSM) approach consists of using all 20 amino acids at a position in a protein. Based on structural knowledge, it may be sufficient to target

the active-site residues (Park et al., 2005; Schmitzer et al., 2004; Wilming et al., 2002; Woodyer et al., 2003). SDSM can be used to complement error-prone PCR. One of the limitations of using error-prone PCR to generate variants of a protein is that the sequence exploration is limited to an average of seven amino acid substitutions per residue. Once positions that seem to be important for improving a property of a protein are identified by error-prone PCR, SDSM will tune the optimization by testing all 20 amino acids in those positions. Multiple positions can be mutagenized simultaneously by SDSM if only a few positions are being explored. However, the number of variants increases exponentially with the number of positions being explored. Therefore, if more than 3 positions are being randomized, it is better to carry out successive targeted randomizations at the positions.

2.2.2 Combinatorial Active-site Saturation Test (CAST)

The Combinatorial Active-site Saturation Test (CAST) was developed to increase the enantioselectivity and/or the substrate specificity of enzymes. The basis of the method is the generation of small libraries of mutant enzymes that are easy to screen for activity. The mutants are produced by simultaneous randomization of sets of two or three spatially close amino acids, whose side chains form part of the substrate-binding pocket.

Application of this methodology allowed the expansion of the substrate specificity of *Pseudomonas aeruginosa* lipase (PAL). Based on the crystal structure, pairs of amino acids surrounding the binding pocket were defined; and the corresponding libraries were created separately by simultaneous saturation mutagenesis at each pair. The libraries were screened for activity with different substrates. The best-performing variants were selected (Reetz et al., 2005). Further optimization was achieved by iterative cycles of CASTing (Reetz et al., 2006a). Mutants that enhanced a given catalytic property were selected. The residue positions thought to be responsible were organized into groups of two or three. Each group was randomized by saturation mutagenesis to create libraries that were subsequently screened. The best hit of those libraries was used as the template for the next round of mutagenesis. Variability at the other sites was introduced by another round of saturation mutagenesis. The process was continued until the desired degree of catalyst improvement had been achieved. Iterative screening has been applied to enhance very different catalytic properties, including thermostability (Reetz et al., 2006b), substrate acceptance and enantioselectivity (Clouthier et al., 2006; Reetz et al., 2006a).

2.2.2.1 B-factor iterative test

The B-factor iterative test is used to modify enzyme thermostability by increasing rigidity at sites to help prevent unfolding. The selection of target residues is made on the basis of crystallographic B-factor data. This value reflects the degree to which the measured electron density for a particular atom spreads out. It is strongly influenced by thermal fluctuations and the mobility of the atom. Residues with the highest B factors have high flexibility. Appropriate mutations lead to enhanced rigidity and, therefore, to higher thermostability. Target sites are chosen as sites for iterative saturation mutagenesis (ISM), in which each of the target residues is mutagenized to saturation. The best mutant of the first screening is then used as the template for a second round of saturation mutagenesis at one of the other selected sites. The cycle is repeated in an iterative manner (Reetz et al., 2006a).

This method has been used to enhance the thermostability and the tolerance to organic solvents of the mesophilic lipase (LipA) from *Bacillus subtilis*. After several rounds of ISM at

residues with high B-factors, a mutant was obtained with the inactivation temperature increased from 48°C to 93°C and with an improved robustness towards organic solvents without affecting the activity of the enzyme (Reetz et al., 2006b; Reetz et al., 2010). Similarly, Jochens and coworkers (Jochens et al., 2010) increased the Tm of the esterase from *Pseudomonas fluorescens* by 9 °C. Smart libraries were guided by the B-factor. By guiding ISM at residues displaying the lowest B-factors, Reetz and coworkers (Reetz et al., 2009) were able to create a lipase from *Pseudomonas aeruginosa* that has a decrease in the Tm value from 71.6 °C to 35.6 °C without affecting the catalytic profile of the enzyme.

2.3 Site-directed homologous recombination

Most proteins are only marginally stable. For this reason, the accumulation of a few mutations is sometimes sufficient to destabilize a protein. The introduction of variability by recombination with the structural gene for a homolog is often less perturbing for folding than is mutagenesis to introduce point mutations. The reason is that some amino acid changes introduced by recombination have already been selected by Nature to give a particular structure and function. Because recombination is more conservative than mutagenesis, several research groups have tried to introduce variability in a sequence by constructing chimeras with genes for homologous proteins. This can be done either randomly (Crameri et al., 1998; Minshull & Stemmer, 1999) or in a site-directed fashion (Landwehr et al., 2007; Li et al., 2007; Pantazes et al., 2007). By substituting a homologous segment, some interactions may be perturbed; and the protein might not be functional. The less perturbing sites for chimeragenesis are thus identified, and a library of recombinants is then constructed by recombining DNA fragments from different homologous genes. The resulting library is screened for a specific property. The properties have included thermostability, activity towards different natural and non-natural substrates, and/or specificity. The power of this technique, compared to random recombination strategies, is that the libraries constructed have a high percentage of folded proteins, thus making it easier to find interesting variants.

2.4 Site-directed loop exchange in proteins

With the same idea as the site-directed chimeragenesis described above, site-directed loop exchange is based on introducing variability only in the binding and/or catalytic loops; the rest of the structure is perturbed very little. The basis of this strategy is that a good portion of catalytic and molecular recognition sites are in loops, while the rest of the structure is the scaffold that maintains the residues and the network of interactions in place to create an environment suitable for catalysis. By exchanging loops, the sequence can be different from that of the parental protein; but the stability and the folding of the protein is maintained. The technique has been widely used, especially for antibodies. It has been recognized that the binding specificity of antibodies relies on the loops of the variable regions (Clark et al., 2009). More recently, import of loops from natural enzymes has been carried out to explore novel activities in a given scaffold (Park et al., 2006). In our laboratory, we have developed a strategy that allows systematic loop exchange from eight different proteins with a TIM barrel fold into a TIM barrel scaffold. (A TIM fold is characterized by a barrel formed by eight parallel β-strands surrounded by seven or eight α-helices. The loops that join the β-strands to the α-helices at the top of the barrel conform the active site in proteins sharing

this fold.*)* We demonstrated that the libraries generated had a high percentage of properly folded proteins (Ochoa-Leyva et al., 2009).

3. Glycosyl hydrolases

Glycosyl hydrolases (also called glycosidases) constitute a widespread group of enzymes that catalyze the hydrolysis of the glycosidic bond. Glycosyl hydrolases can hydrolyze the glycosidic linkage between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety to release smaller sugars. They are classified as exo- and endoglycosidases, depending on their ability to cleave a substrate at the end (the non-reducing end) or in the middle of an oligosaccharide or polysaccharide chain, respectively. Glycossyl hydrolases have been classified into more than 100 families based on amino acid sequence similarities (Davies & Henrissat, 1995; Henrissat, 1991; Henrissat & Bairoch, 1993; 1996; Henrissat & Davies, 1997). This classification system, available on the CAZy (CArbohydrate-Active EnZymes) web site (Cantarel et al., 2009), allows reliable prediction of evolutionary relationships, mechanism (retaining/inverting), active site residues and possible substrates. It is even a reasonable tool for newly sequenced enzymes for which function has not yet been biochemically demonstrated.

3.1 Reactions catalyzed by glycosyl hydrolases

In most cases, the hydrolysis of the glycosidic bond is catalyzed by two amino acid residues - a general acid (proton donor) and a nucleophile/base. Depending on the spatial positions of these catalytic residues, hydrolysis occurs via overall retention or overall inversion of the anomeric configuration (Davies & Henrissat, 1995; McCarter & Withers, 1994; Sinnott, 1990).

3.1.1 Inverting glycosyl hydrolases

Inverting enzymes act by a single-step, acid/base-catalyzed mechanism. Two residues, typically glutamic or aspartic acids located 6-11 Å apart, act as acid and base. The leaving group is directly displaced by the nucleophilic water with a single inversion at the anomeric centre (Fig. 1).
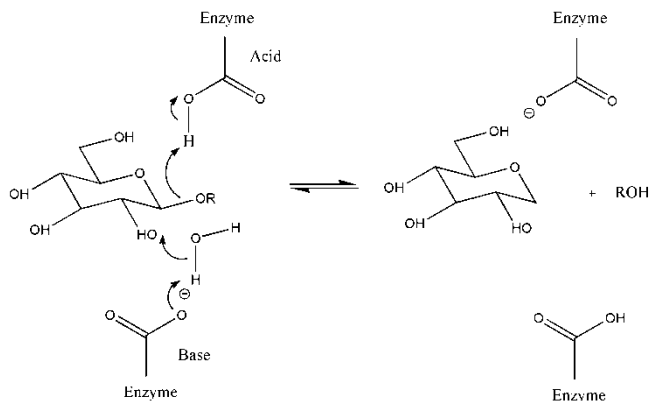
Fig. 1. Inversion hydrolysis mechanism of glycosyl hydrolases.

3.1.2 Retaining glycosyl hydrolases

Retaining glycosidases act through a double-displacement mechanism (each step resulting in inversion at the anomeric centre) involving a covalent glycosyl-enzyme intermediate (Fig. 2). The reaction is catalyzed with acid/base and nucleophilic assistance provided by two amino acid side chains, typically glutamate or aspartate, located 5.5 Å apart. In the first step (glycosylation), one residue plays the role of a nucleophile. It attacks the anomeric centre to displace the aglycon and form a glycosyl enzyme intermediate. At the same time, the other residue functions as an acid catalyst and protonates the glycosidic oxygen as the bond cleaves. In the second step (deglycosylation), the now deprotoned acid-base carboxylate functions as a base to activate the incoming nucleophile (water, saccharide or alcohol) to which the glycosyl is transferred from the enzyme intermediate to give the product.
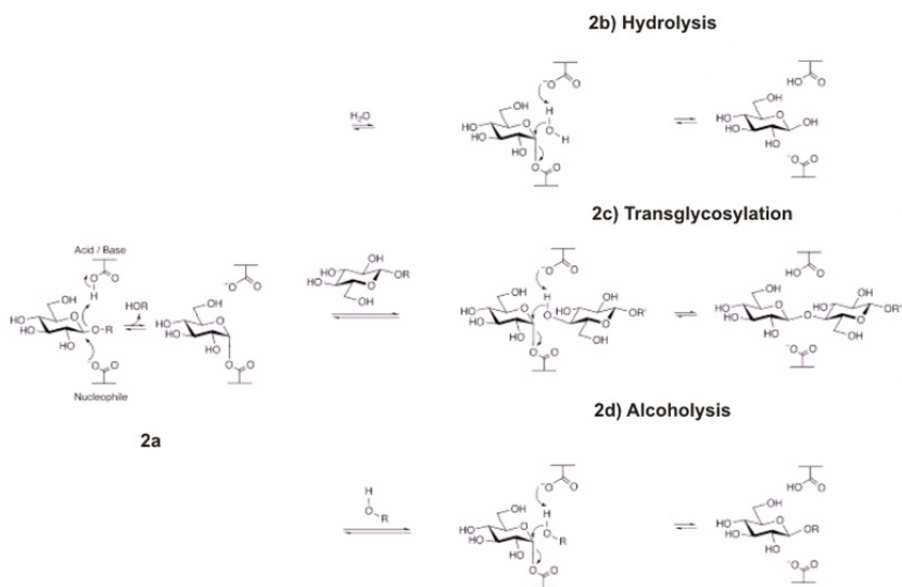
Fig. 2. Retaining mechanism of glycosyl hydrolases.
In a first reaction (a), glycosidic bond breakage of the donor saccharide is carried out and a glycosyl enzyme complex is formed. In (b), the incoming acceptor molecule (water) is activated to promote the release of sugar. If the incoming nucleophile is different from water, the enzyme carries out a transfer reaction, called transglycosylation if the incoming molecule is an oligosaccharide (c), or alcoholysis if the incoming nucleophile is an alcohol (d).

3.2 Industrial uses of glycosyl hydrolases

In nature, glycosyl hydrolases catalyze the degradation of diverse glycosylated polymers, like starch, glucogen, cellulose and hemicellulose. They also participate in anti-bacterial defense strategies (*e.g.*, lysozyme), in pathogenesis mechanisms (*e.g.*, viral neuraminidases)

and in normal cellular functions. Glycosyl hydrolases are also of great importance to industry. For example, in the food industry, enzymes like invertase and amylase are employed for the manufacture of invert sugar or maltodextrins; in the paper and pulp industry, xylanases are used to remove hemicelluloses from paper pulp; cellulases are widely used in the textile industry and in laundry detergents; and recently, cellulases and xylanases have been used in the conversion of lignocellulosic biomass into forms suitable for biofuel production.

However, in most cases, glycosyl hydrolases are not optimal in industrial conditions. It is often necessary to alter their stabilities, catalytic activities and/or substrate specificities by protein engineering methods. One of the most frequently altered enzyme properties is thermostability. It can be a limiting factor in the selection of enzymes for industrial applications due to the elevated temperatures or the extreme pH of many biotechnological processes. The stability of an enzyme can be improved by site-directed mutagenesis (Ben Mabrouk et al., 2011; Ghollasi et al., 2010; Leemhuis et al., 2004; Liu et al., 2008; Yin et al., 2011). One of the most exhaustive efforts was done by Palackal and coworkers (Palackal et al., 2004), who used saturation mutagenesis for each of the 189 amino acid residues of a xylanase. They generated a library of modified enzymes, each altered at single position. This library was then screened for variants with increased thermostabiltity, and nine single amino acid changes that contribute to increased stability were identified. These nine single substitutions were then combinatorially assembled to generate all 512 possible variants. Another round of screening identified eleven enzymes with melting temperatures up to 35°C higher than that of the wild-type enzyme.

Another enzyme property that is desirable to modify is the optimum pH. For example, in soybean β-amylase, the hydrogen bond networks around the catalytic base residue (E380) of the enzyme were removed by point mutations, raising the optimal pH from 5.4 to the more neutral pH range of between 6 and 6.6 (Hirata et al., 2004a; Hirata et al., 2004b).

In vivo, glycosidases catalyze the hydrolysis of glycosidic linkages. However, *in vitro*, they can be used as synthetic catalysts to form glycosidic bonds. This process is called the kinetic approach, and it can be accomplished by reverse hydrolysis or by transglycosylation (Fig. 2 c and d). The utility of glycosidases in the synthesis of glucosides through transglycosylation reactions has been employed to synthesize unusual products that are difficult to obtain by other methods. Several site-directed mutagenesis strategies have been used to increase the translycosylation activity of glycosidases or to change substrate specificity. For example, rational modification of the β-glycosidase from *Sulfolobus sulfataricus* to accept a wider range of substrates in transglycosylation reactions has been done. Site-directed mutagenesis was used to alter two key residues involved in substrate recognition to provide access to many different glycoside linkages, including the especially problematic β-mannosyl and β-xylosyl linkages (Hancock et al., 2005). We will focus our discussion on the protein engineering work on α-amylases carried out by us and others.

3.3 alpha-amylases

α-Amylases (EC number 3.2.1.1) are part of the family 13 of glycosyl hydrolases. They catalyze the hydrolysis of internal α-1,4-glycosidic linkages of starch, liberating poly- and oligosaccharides chains of varying lengths. They are found in both eubacteria and

eukaryotes. They have a large number of different substrate specificities, as well as huge variations in both temperature and pH optima (Vihinen & Mantsala, 1989).

α-Amylases are the starting enzymes in the industry of modification and conversion of starch. This is because of their capacity to catalyze reactions under environmentally friendly conditions and without the addition of expensive activated sugars (Buchholz & Seibel, 2008). In the sugar-producing industry (Nielsen & Borchert, 2000), bacterial and fungal α-amylases of family GH13, particularly those of the *Bacillus species*, play a vital role in the starch liquefaction process. Starch from wheat, maize and tapioca is hydrolyzed to produce oligosaccharides by the thermostable α-amylases from *Bacillus licheniformis*. The oligosaccharides are then saccharified to glucose by glucoamylase (Crabb & Shetty, 1999). According to the degree of hydrolysis of starch, α-amylases are divided into two categories: (1) saccharifying α-amylases, which hydrolyze 50 to 60% of the saccharide bonds and (2) liquefying enzymes, which process about 30 to 40% of starch hydrolysis (Fukumoto & Okada, 1963). The enzyme commonly used in the industrial process is the α-amylase from *Bacillus licheniformis*. It has the great advantage of being thermostable. This enzyme thus allows the fast hydrolysis of starch at the high temperatures required to dissolve it, with the consequent decrease in viscosity, before decreasing the temperature for the addition of the next enzyme in the process. Some of the disadvantages of using different enzymes are that the pH or temperature conditions may need to be adjusted during the process, with consequent increase in time, costs, and the introduction of salts (buffers) that will have to be removed from the final product. Thus, several research groups, including ours, have tried to engineer α-amylases to change their product profiles (*i.e.*, to make them more saccharifying) to increase their optimal temperatures and to widen their pH spectra.

All α-amylases consists of three domains, called A, B and C. Domain A contains the catalytic residues and has four conserved sequence regions (numbered I-IV) (Mackay et al., 1985; Nakajima et al., 1986; Rogers, 1985), which have been postulated to be essential for the function of α-amylase. Among α-amylase sequences, the four regions align and are spaced at similar intervals along the proteins. These regions presumably form the active site cleft, the substrate-binding site, and the site for binding the stabilizing calcium ion. Domain B forms a large part of the substrate binding cleft, and it is presumed to be important for the substrate specificity differences observed among α-amylases (MacGregor, 1988). It is the least conserved domain among α-amylases (Guzman-Maldonado & Paredes-Lopez, 1995). Finally, domain C constitutes the C-terminal part of the sequence and seems to be involved in substrate binding. All known α-amylases contain a conserved calcium ion located at the interface between domains A and B (Boel et al., 1990; Kadziola et al., 1998; Machius et al., 1998; Machius et al., 1995). The calcium ion is known to be essential for enzyme stability (Vallee et al., 1959).

Depending on the enzyme, the active site cleft can accommodate from four to ten glucose units, each one bound by amino acid residues that constitute the binding subsite for that glucose unit. Subsites are numbered according to the location of the scissile bond. In α-amylases there are two or three subsites on the reducing end of the scissile bond (subsites +1, +2 and +3). The number of subsites on the non-reducing side of scissile bond varies between two and eleven (subsites -1, -2, ... -11) (Brzozowski et al., 2000; Davies et al., 1997; MacGregor, 1988). The number of subsites and their affinities are some of the determinant factors of the final product profiles of α-amylases (Gyemant et al., 2002; Kandra et al., 2002; Matsui et al., 1992a; b).

3.3.1 Transglycosylation reactions in alpha-amylases

As other retaining glycosidases, α-amylases, particularly saccharifying amylases, can also catalyze transfer reactions, which are the result of employing molecules other than water (*e.g.*, carbohydrates or alcohols) as glucosyl acceptors (Fig. 2 c and b, respectively). When a high molecular-weight alcohol is used as an acceptor, the products are alkyl-glucosides. These molecules have a high surface tension activity that has important applications in several industries. Although various retaining glucosidases, like β-galactosidase (Moreno-Beltran et al., 1999; Svensson, 1994), β-xylosidase (Shinoyama et al., 1988), β-fructofuranosidase (Rodríguez et al., 1996; Straathof et al., 1988), and β-glucosidase (Chahid et al., 1992; Vulfson et al., 1990), have been used in alcoholysis reactions, the use of a readily available substrate, like starch, gives α-amylases great potential in the catalysis of this type of reaction.

We found a correlation between the efficiency of hydrolysis and the capacity of the enzymes to carry out transglycosylation reactions. A plausible hypothesis is that those enzymes that are able to transglycosylate can recycle intermediate size oligosaccharides produced during hydrolysis to generate longer ones that are better substrates. This would result in a more saccharifying pattern at equilibrium. Transglycosylation activity is not reported in the bacillary α-amylases used in the starch process industry. We decided to introduce this activity by engineering liquefying α-amylases from *Bacillus stearothermophilus* (Saab-Rincon et al., 1999) and *Bacillus licheniformis* (Rivera et al., 2003). We tried to identify residues that could be responsible for transferase activity. Kuriki and coworkers (Kuriki et al., 1996) suggested three residues that are likely to be responsible for controlling the water activity in the active site of the neopullulanase, a natural transferase from *Bacillus stearothermophilus*. When one of these residues (Y377) was mutated to a non-polar residue, the transglycosylation reaction was favored due to a change in the transglycosylation/ hydrolysis ratio. We carried out a multiple sequence alignment of α-amylases and cyclodextrin glycosyltransferases (CGTases) and identified a residue (A289 in the *Bacillus stearothermophilus* α-amylase) that is analogous to Y377 in the *Bacillus stearothermophilus* neopullulanase. The *Bacillus stearothermophilus* α-amylase is a liquefying enzyme unable to carry out transglycosylation reactions (Fig. 3). We used site-directed mutagenesis to change the A at residue 289 to Y and F, which are present in natural transferases, like neopollullanases and CGTases. The two mutants that were generated were able to carry out the transfer reaction not only to other saccharides (Fig. 4) but also to alcohols, like methanol, to produce methyl-glucosides. The A289Y mutant was more efficient at catalyzing transfer reactions than was A289F (Fig. 5) (Saab-Rincon et al., 1999). Apparently the hydrophobic nature of the mutant residues and the electrostatic interactions that may affect the geometry of the side chains in the active site are important for the transglycosylation reaction. In contrast, when the same mutations were introduced at the equivalent position (V286) in the α-amylase from *Bacillus licheniformis*, the V286Y mutant showed an increase of hydrolytic activity, whereas the V286F mutant had a higher translgycosylation/hydrolysis ratio (Rivera et al., 2003).

In contrast to bacterial liquefying α-amylases from *B. licheniformis* and *B. stearothermophilus*, several fungal amylases like those from *Aspegillus niger* and *Aspergillus oryzae* have the ability to carry out alcoholysis reactions. These two fungi amylases are responsible for saccharifying enzymes that produce maltose, maltotriose and some glucose. Santamaria and

coworkers (Santamaria et al., 1999) demonstrated that these enzymes were able to carry out alcoholysis reactions in the presence of methanol and starch as substrate, even at high methanol (20%) and starch (15%) concentrations. Although the alcoholysis reaction was reported in α-amylase from *A. oryzae* using aryl-maltoside and either, methanol, ethanol or butanol as substrates (Matsubara, 1961), the alcoholysis reaction with starch as substrate is less efficient (Santamaria et al., 1999).

```
   Enzyme          Region I           Region II           Region III            Region IV
                                          *                   *                      *
T. maritima    KVIMDLVINHT 129 DGFRIDAAKHI 223 VGEVFSGN 263 FNFALM 280 FLENHDL 311
A. oryzae      YLMVDVVANHM 123 DGLRIDTVKHV 211 IGEVLDGD 235 LNYPIY 255 FVENHDN 298
A. niger       YLMVDVVPDHM 123 DGLRIDSVLEV 211 VGEIDNGN 235 LNYPIY 255 FIENHDN 298
H. sapiens sal RIYVDAVINHM 102 AGFRIDASKHM 202 YQEVIDLG 238 TEFKYG 259 FVDNHDN 301
H. sapiens pan RIYVDAVINHM 102 AGFRIDASKHM 202 YQEVIDLG 238 TEFKYG 259 FVDNHDN 301
Pig Panc       RIYVDAVINHM 102 AGFRIDASKHM 202 FQEVIDLG 238 TEFKYG 259 FVNDHDN 301
Tenebrio       RIYVDAVINHM 100 AGFRVDAAKHM 190 YQEVIDLG 227 LEFQFG 248 FVDNHDN 288   α-Amylases
Alteromonas    DIYVDTLINHM  90 KGFRFDASKHV 179 FQEVIDQG 205 TEFKYS 226 FVDNHDN 265
Barley         KAIADIVINHR  93 DGWRFDFAKGY 184 VAEIWTSL 209 FDFTTK 249 FVDNHDT 286
B. lichen      NVYGDVVINHK 106 DGFRLDAVKHI 236 VAEYWQND 266 FDYPLH 289 FVDNHDT 329
B. amylo       QVYGDVVLNHK 104 DGFRIDAAKHI 236 VAEYWQNN 266 FDVPLH 289 FVDNHDT 329
B. stearo      QVYADVVFDHK 107 DGFRLDAVKHI 239 VGEYWSYD 269 FDAPLH 292 FVDNHDT 332
B. sub         KVIVDAVINHT 103 DGFRFDAAKHI 180 YGEILQDS 213 TASNYG 232 WVESHDT 270
B. circ1       KVIIDFAPNHT 141 DGIRMDAVKHM 234 FGEWFLGV 262 LDFRFA 286 FIDNHDM 329
B. circ2       KIVIDFAPNHT 141 DGIRVDAVKHM 234 FGEWFLGS 262 LDFRFN 286 FIDNHDM 329
B. specie      KVIIDFAPNHT 141 DGIRVDAVKHM 234 FGEWFLGV 262 LDFRFA 286 FIDNHDM 329   CGTases
B. stearo      KVIIDFAPNHT 137 DGIRMDAVKHM 230 FGEWFLSE 257 LDFRFG 281 FIDNHDM 325
Termo sulfu    KVIIDFAPNHT 142 DGIRLDAVKHM 235 FGEWFLGT 263 LDFRFS 283 FIDNHDM 330
Thermus sp     RVMLDAVFNHC 248 DGWRLDVANEI 233 LGEIWHDA 362 MNYPLA 380 LLGSHDT 422
B. lichen      KIMLDAVFNHI 252 DGWRLDVANEV 334 LGEIWHQA 363 MNYPFT 381 LLDSHDT 426   Maltogenic.
B. stearo      RVMLDAVFNHS 248 DGWRLDVANEV 333 LGEIWHDA 361 MNYPFT 380 LLGSHDT 425
```

Fig. 3. Multiple sequence alignment around the four characteristic regions observed in members of the glycoside hydrolase family 13.

The catalytic residues, conserved in all the sequences are marked with asterisks. Aromatic residues involved in the transglycosylation activity of the CGTases are indicated in grey. Residues mutagenized are shown in red. Sequences from α-amylases, GCTases and maltogenic amylases are taken from Damian-Almazo et al. (2008). The abbreviated terms follow: T. maritima, *Thermotoga maritima* α-amylase; A. oryzae, *Aspergillus oryzae* α-amylase; A. niger, *Aspergillus niger* α-amylase; H. sapiens sal, *Homo sapiens* salivary α-amylase; H. sapiens pan, *Homo sapiens* pancreatic α-amylase; Pig panc, *Sus scrofa* pancreatic α-amylase; Tenebrio, *Tenebrio molitor* (mealworm) α-amylase; Alteromonas, *Pseudoalteromonas haloplanktis* α-amylase; Barley, *Hordeun vulgare* α-amylase; B. lichen, *Bacillus licheniformis* α-amylase; B. amylo, *Bacillus amyloliquefaciens* chimera α-amylase; B. stearo, *Bacillus stearothermophilus* α-amylase; B. sub, *Bacillus subtilis* 2633 α-amylase; B. circ1, *Bacillus circulans* 251 cyclodextrin glycosyltransferase (CGTase); B. circ2, *Bacillus circulans* 8 CGTase; B. specie, Bacillus sp. 1011 CGTase; B. stearo, *B. stearothermophilus* CGTase; Termo sulfu, *Thermoanaerobacter thermosulfurogenes* CGTase; Thermus sp., Thermus sp.; B. lichen, *B. licheniformis* maltogenic amylase; B. stearo, *B. stearothermopilus* maltogenic amylase.
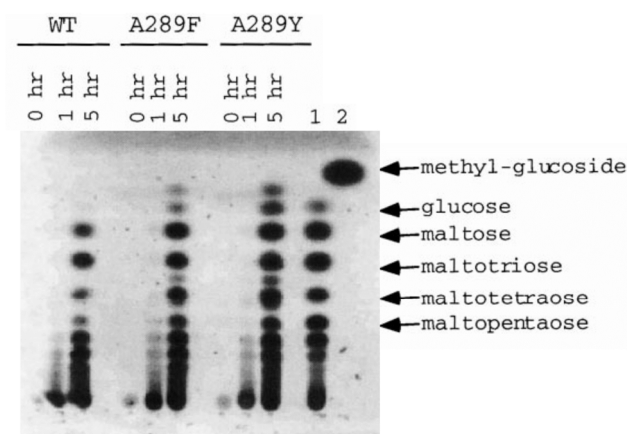
Fig. 4. Product profiles of the alcoholysis reaction with wild-type *Bacillus stearothermophilus* α-amylase and the A289F and A289Y variants at different times of reaction using starch and methanol as substrates.

The product profiles obtained with wild-type (WT) and mutant (A289F and A289Y) *Bacillus stearothermophilus* α-amylases are compared at 0, 1, 5 hours of reaction. We used as standards a mixture of oligosaccharides (1) and methyl-glucoside (2).

Although the wild-type enzyme and the A289F and A289Y mutants showed similar hydrolysis and transglycosylation patterns, the mutants showed products between the glucose and methyl-glucoside standards that could be attributed to alcoholysis reactions. Presumably, those spots for which there are no molecular weight markers correspond to alkyl-oligosaccharides.
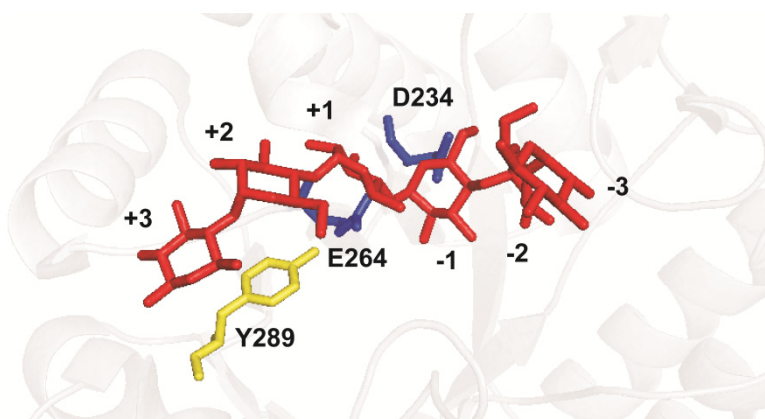
Fig. 5. Structural model of the A289Y mutant of *Bacillus stearothermophilus* α-amylase (PDB code 1HVX) crystallized with the substrate analogue acarbose. (The structure is based on the structure of the α-amylase from *Aspergillus oryzae*, PDB code 7TAA.) Catalytic residues are represented in blue sticks and the mutated residue (Y289) is represented in yellow.

However, the direct use of these enzymes for the production of alkyl-glucosides is precluded by the high temperature required for starch solubilization. The use of a thermophilic saccharifying α-amylase would be attractive, not only in the development of alcoholysis reactions, but also in the starch-processing industry.

Liebl et al. (1997) described an extracellular α-amylase (AmyA) produced by the hyperthermophillic bacterium *Thermotoga maritima* MSB8. The enzyme is a saccharifying amylase with an optimum temperature of 85°C. It can hydrolyze internal α-1-4-glycosidic bonds in various α-glucans, such as starch, amylose, amylopectin and glycogen, to yield mainly glucose and maltose as final products. Because AmyA has the advantage of being a saccharifying enzyme in a stable scaffold, we explored its properties in the transglycosylation and alcoholysis reactions (Damián-Almazo et al., 2008; Damian-Almazo et al., 2008; Moreno et al., 2010). In addition to the characterization reported by Lieb et al., we found that AmyA is capable of using small oligosaccharides (G2 to G7) as substrates for the transglycoslation reactions at short reaction times. This was followed by hydrolysis to yield glucose and maltose as final products. The ability of AmyA to use maltose as a substrate is unusual, as most α-amylases are not capable of using maltose to transfer glucosyl units to other oligosaccharides. Moreover, in the presence of various substrates, AmyA is able to form neotrehalose, a non-reducing disaccharide composed of two glucose molecules joined by α-1, β-1 linkage. It uses 6% maltose as a substrate. Like other saccharifying enzymes, AmyA is capable of transferring glycosyl units to methanol and butanol to produce alkyl-glucosides. When compared to other saccharifiyng α-amylases, AmyA has a high transfer capacity. The enzyme generates 7.5 mg/ml of methyl-glucoside (Moreno et al., 2010), almost three times the maximum amount found for the *A. niger* α-amylase and almost eight times the maximum amount found for the *A. oryzae* α-amylase (Santamaria et al., 1999).

In order to increase the alcoholytic activity present in AmyA, we constructed a structural homology model based on the structure of the α-amylase from *A. oryzae*. The low sequence identity between these enzymes precluded the use of the automatic modeler function in the Swiss Prot server (Sali et al., 1995; Sanchez & Sali, 1997). Therefore, the sequence alignment of the proteins had to be manually adjusted using as anchors the four highly conserved regions of the α-amylases, as shown in Fig. 3. Once a model was generated, the inhibitor molecule acarbose was placed in the active site using the coordinates of the *A. oryzae* α-amylase (PDB code 7TAA). A close-up of the active site model (Fig. 6) supported our hypothesis of the relationship between the presence of an aromatic residue at the position equivalent to Y377 in neopullulanase and the transglycosylation activity of the enzyme and a saccharifying profile. We identified other residues in subsite +1 that are involved in the transglycosylation activity of other glycosyl hydrolases (Kim et al., 2000; Leemhuis et al., 2004; van der Veen et al., 2001). One of these (H222) is part of the second highly conserved region among glycosyl hydrolases and has also been implicated in calcium ion coordination. In the AmyA model, this residue points toward the sugar moiety at subsite +1. Mutagenesis of the equivalent residue in other amylases has been shown to change transferase activity. In the case of the *B. stearothermophilus* α-amylase, the replacement of the equivalent H238 with aspartic acid generated an enzyme with a reduced hydrolysis rate and a modified final product profile (Vihinen & Mantsala, 1990). We constructed a small library by site-directed mutagenesis to explore the effects of replacing H222 with D, Q and E. All mutants showed a greater amount of methyl-glucoside than did the wild-type enzyme, as a result of a change in the alcoholysis/hydrolysis ratio. Mutant H222Q showed an increase in the alcoholysis

events as a consequence of an increase in alcoholysis and a reduction in hydrolytic activity of almost 30%. The same change was observed in mutants H222D and H222E. The instability of these mutants toward alcohols decreased the final yield of alkyl-glucoside, as shown in Fig. 7 (Damian-Almazo et al., 2008).
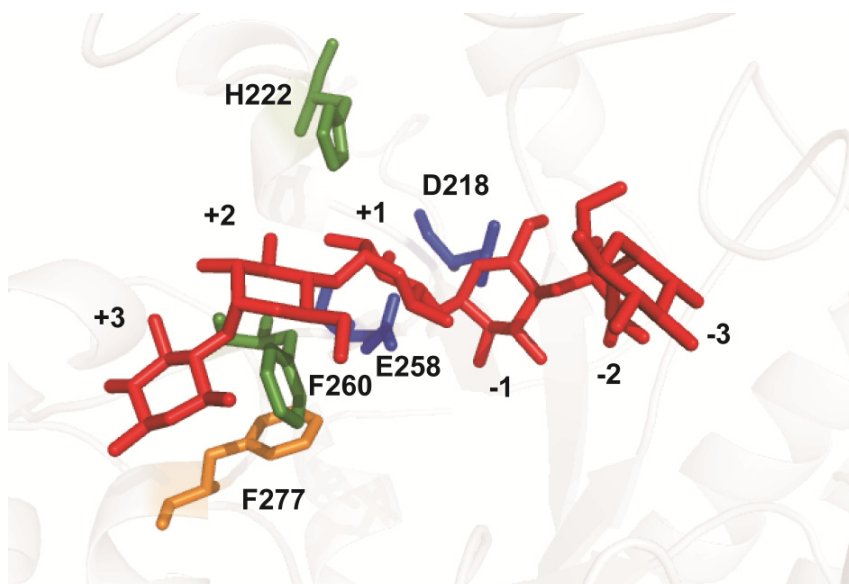
Fig. 6. Homology model obtained for AmyA active site.
The inhibitor acarbose (red) is surrounded by catalytic residues D218 and E258 (blue) and various mutated residues (green). The F277 residue, equivalent to Y377 of neopullulanase, is shown in orange.
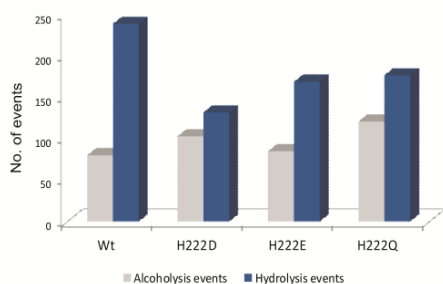
Enzyme	Alcoholysis/Hydrolysis ratio	Methyl glucoside (mg/ml)
Wt	33	7.5, ±0.6
H222D	78	10.6, ±0.4
H222E	51	9.7, ±1.8
H222Q	68	12.5, ±0.8

Fig. 7. Alcoholysis reaction yields of the wild-type α-amylase from *Thermotoga maritima* and some of the mutants generated
Quantification of alcoholysis reactions generated by wild-type α-amylase from *Thermotoga maritima* and the H222 residue mutants. (A) Alcoholysis and hydrolysis events from 6% starch - 20% methanol obtained with 20 U/ml of the enzymes shown; (B) alcoholysis/hydrolysis ratios and methyl-glucoside yields.

The comparison of liquefying and saccharifying α-amylases, neopollulanases, CGTases and maltogenic amylases through a multiple sequence alignment (Fig. 3) has also made possible the identification of other residues potentially involved in the transglycosylation activity (Fig. 6). In the CGTase from *Bacillus circulans*, residue F260 has been identified as part of a switch for the transglycosylation and hydrolysis reactions (van der Veen et al., 2001). Mutants formed by changing the equivalent residue in wild-type AmyA (F260) to W and G and the H222Q mutant showed opposite behaviors. In the presence of soluble starch as substrate, mutants H222Q and F260G leave higher amounts of high-molecular weight oligosaccharides, while the wild-type enzyme and mutant F260W show a higher proportion of glucose. These differences were seen as changes in the transglycolyslation/hydrolysis ratios. In the double mutant H222Q-F260W, the more transglycosidic pattern of H222Q was recessive, thus eliminating or reducing the presence of longer oligossacharides (Damián-Almazo et al., 2008).

4. Conclusions

Site-directed mutagenesis is a powerful tool for both the study of protein function and the design of novel proteins. Using several approaches to identify phylogenetically conserved residues or residues involved in binding, it has been possible to modify the properties of enzymes that have industrial and biotechnological applications. In order to increase the transglycosylation reactions carried out by α-amylases, we have applied site-directed mutagenesis to residues close to the active site. Based on multiple sequence alignments of natural transferases, like CGTases, we identified conserved residues involved in the transferase reactions of fungal and bacterial α-amylases. Changes to these residues in α-amylases that originally were unable to perform the translycosylation reactions altered the product profiles and increased the translgycosylation/hydrolysis ratios. Furthermore, it was possible to increase the alcoholysis reactions in the α-amylase from *Thermotoga maritima*, which was already capable of carrying out this kind of reaction at a low level.

5. Acknowledgements

6. References

Alcolombri, U., Elias, M. & Tawfik, D.S. (2011). Directed evolution of sulfotransferases and paraoxonases by ancestral libraries. *J Mol Biol*, Vol. 411, No. 4, pp. 837-853, 1089-8638 (Electronic) 0022-2836 (Linking).

Ashkenazi, A., Presta, L.G., Marsters, S.A., Camerato, T.R., Rosenthal, K.A., Fendly, B.M. & Capon, D.J. (1990). Mapping the CD4 binding site for human immunodeficiency virus by alanine-scanning mutagenesis. *Proc Natl Acad Sci U S A*, Vol. 87, No. 18, pp. 7150-7154, 0027-8424 (Print) 0027-8424 (Linking).

Ben Mabrouk, S., Aghajari, N., Ben Ali, M., Ben Messaoud, E., Juy, M., Haser, R. & Bejar, S. (2011). Enhancement of the thermostability of the maltogenic amylase MAUS149 by Gly312Ala and Lys436Arg substitutions. *Bioresour Technol*, Vol. 102, No. 2, pp. 1740-1746, 1873-2976 (Electronic) 0960-8524 (Linking).

Blaber, M., Baase, W.A., Gassner, N. & Matthews, B.W. (1995). Alanine scanning mutagenesis of the alpha-helix 115-123 of phage T4 lysozyme: effects on structure, stability and the binding of solvent. *J Mol Biol*, Vol. 246, No. 2, pp. 317-330, 0022-2836 (Print) 0022-2836 (Linking).

Boel, E., Brady, L., Brzozowski, A.M., Derewenda, Z., Dodson, G.G., Jensen, V.J., Petersen, S.B., Swift, H., Thim, L. & Woldike, H.F. (1990). Calcium binding in alpha-amylases: an X-ray diffraction study at 2.1-A resolution of two enzymes from Aspergillus. *Biochemistry*, Vol. 29, No. 26, pp. 6244-6249.

Bolon, D.N. & Mayo, S.L. (2001). Enzyme-like proteins by computational design. *Proc Natl Acad Sci U S A*, Vol. 98, No. 25, pp. 14274-14279.

Brzozowski, A.M., Lawson, D.M., Turkenburg, J.P., Bisgaard-Frantzen, H., Svendsen, A., Borchert, T.V., Dauter, Z., Wilson, K.S. & Davies, G.J. (2000). Structural analysis of a chimeric bacterial alpha-amylase. High-resolution analysis of native and ligand complexes. *Biochemistry*, Vol. 39, No. 31, pp. 9099-9107.

Buchholz, K. & Seibel, J. (2008). Industrial carbohydrate biotransformations. *Carbohydr Res*, Vol. 343, No. 12, pp. 1966-1979, 0008-6215 (Print) 0008-6215 (Linking).

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res*, Vol. 37, No. Database issue, pp. D233-238, 1362-4962 (Electronic) 0305-1048 (Linking).

Castle, L.A., Siehl, D.L., Gorton, R., Patten, P.A., Chen, Y.H., Bertain, S., Cho, H.J., Duck, N., Wong, J., Liu, D. & Lassner, M.W. (2004). Discovery and directed evolution of a glyphosate tolerance gene. *Science*, Vol. 304, No. 5674, pp. 1151-1154, 1095-9203 (Electronic) 0036-8075 (Linking).

Clark, L.A., Boriack-Sjodin, P.A., Day, E., Eldredge, J., Fitch, C., Jarpe, M., Miller, S., Li, Y., Simon, K. & van Vlijmen, H.W. (2009). An antibody loop replacement design feasibility study and a loop-swapped dimer structure. *Protein Eng Des Sel*, Vol. 22, No. 2, pp. 93-101, 1741-0134 (Electronic) 1741-0126 (Linking).

Clouthier, C.M., Kayser, M.M. & Reetz, M.T. (2006). Designing new Baeyer-Villiger monooxygenases using restricted CASTing. *J Org Chem*, Vol. 71, No. 22, pp. 8431-8437, 0022-3263 (Print) 0022-3263 (Linking).

Cohen, H.M., Tawfik, D.S. & Griffiths, A.D. (2004). Altering the sequence specificity of HaeIII methyltransferase by directed evolution using in vitro compartmentalization. *Protein Eng Des Sel*, Vol. 17, No. 1, pp. 3-11, 1741-0126 (Print).

Crabb, W.D. & Shetty, J.K. (1999). Commodity scale production of sugars from starches. *Curr Opin Microbiol*, Vol. 2, No. 3, pp. 252-256, 1369-5274 (Print) 1369-5274 (Linking).

Crameri, A., Raillard, S.A., Bermudez, E. & Stemmer, W.P. (1998). DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature*, Vol. 391, No. 6664, pp. 288-291.

Cunningham, B.C. & Wells, J.A. (1989). High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science*, Vol. 244, No. 4908, pp. 1081-1085, 0036-8075 (Print) 0036-8075 (Linking).

Chahid, Z., Montet, D., Pina, M. & Graille, J. (1992). Effect of water activity on enzymatic synthesis of alkylglycosides. *Biotechnol Lett*, Vol. 14, No., pp. 281-284.

Chen, F., Gaucher, E.A., Leal, N.A., Hutter, D., Havemann, S.A., Govindarajan, S., Ortlund, E.A. & Benner, S.A. (2010). Reconstructed evolutionary adaptive paths give polymerases accepting reversible terminators for sequencing and SNP detection. *Proc Natl Acad Sci U S A*, Vol. 107, No. 5, pp. 1948-1953, 1091-6490 (Electronic) 0027-8424 (Linking).

Cherry, J.R. & Fidantsef, A.L. (2003). Directed evolution of industrial enzymes: an update. *Curr Opin Biotechnol*, Vol. 14, No. 4, pp. 438-443, 0958-1669 (Print) 0958-1669 (Linking).

Christians, F.C., Scapozza, L., Crameri, A., Folkers, G. & Stemmer, W.P. (1999). Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling. *Nat Biotechnol*, Vol. 17, No. 3, pp. 259-264.

Damián-Almazo, J., López-Munguía, A., Soberón-Mainero, X. & Saab-Rincón, G. (2008). Role of the phenylalanine 260 residue in defining product profile and alcoholytic activity of the α-amylase AmyA from Thermotoga maritima. *Biologia*, Vol. 63, No. 6, pp. 1035-1043.

Damian-Almazo, J.Y., Moreno, A., Lopez-Munguia, A., Soberon, X., Gonzalez-Munoz, F. & Saab-Rincon, G. (2008). Enhancement of the alcoholytic activity of alpha-amylase AmyA from Thermotoga maritima MSB8 (DSM 3109) by site-directed mutagenesis. *Appl Environ Microbiol*, Vol. 74, No. 16, pp. 5168-5177, 1098-5336 (Electronic) 0099-2240 (Linking).

Davies, G. & Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure*, Vol. 3, No. 9, pp. 853-859, 0969-2126 (Print) 0969-2126 (Linking).

Davies, G.J., Wilson, K.S. & Henrissat, B. (1997). Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem J*, Vol. 321 (Pt 2), No., pp. 557-559, 0264-6021 (Print) 0264-6021 (Linking).

Di Giulio, M. (2001). The universal ancestor was a thermophile or a hyperthermophile. *Gene*, Vol. 281, No. 1-2, pp. 11-17.

Di Giulio, M. (2003). The Universal Ancestor was a Thermophile or a Hyperthermophile: Tests and Further Evidence. *Journal of Theoretical Biology*, Vol. 221, No. 3, pp. 425-436.

Ditursi, M.K., Kwon, S.J., Reeder, P.J. & Dordick, J.S. (2006). Bioinformatics-driven, rational engineering of protein thermostability. *Protein Eng Des Sel*, Vol. 19, No. 11, pp. 517-524, 1741-0126 (Print) 1741-0126 (Linking).

Flores, H. & Ellington, A.D. (2005). A modified consensus approach to mutagenesis inverts the cofactor specificity of Bacillus stearothermophilus lactate dehydrogenase. *Protein Eng Des Sel*, Vol. 18, No. 8, pp. 369-377, 1741-0126 (Print) 1741-0126 (Linking).

Fukumoto, J. & Okada, S. (1963). Studies on bacterial amylase. Amylase types of Bacillus subtilis species. *J Ferm Technol*, Vol. 41, No., pp. 427.

Funke, S.A., Otte, N., Eggert, T., Bocola, M., Jaeger, K.E. & Thiel, W. (2005). Combination of computational prescreening and experimental library construction can accelerate enzyme optimization by directed evolution. *Protein Eng Des Sel*, Vol. 18, No. 11, pp. 509-514, 1741-0126 (Print) 1741-0126 (Linking).

Ghollasi, M., Khajeh, K., Naderi-Manesh, H. & Ghasemi, A. (2010). Engineering of a Bacillus alpha-amylase with improved thermostability and calcium independency. *Appl Biochem Biotechnol*, Vol. 162, No. 2, pp. 444-459, 1559-0291 (Electronic) 0273-2289 (Linking).

Gibbs, C.S. & Zoller, M.J. (1991). Identification of electrostatic interactions that determine the phosphorylation site specificity of the cAMP-dependent protein kinase. *Biochemistry*, Vol. 30, No. 22, pp. 5329-5334, 0006-2960 (Print) 0006-2960 (Linking).

Giver, L., Gershenson, A., Freskgard, P.O. & Arnold, F.H. (1998). Directed evolution of a thermostable esterase. *Proc Natl Acad Sci U S A*, Vol. 95, No. 22, pp. 12809-12813.

Grishin, N.V. (2001). Fold change in evolution of protein structures. *J Struct Biol*, Vol. 134, No. 2-3, pp. 167-185, 1047-8477 (Print) 1047-8477 (Linking).

Gromiha, M.M., Oobatake, M. & Sarai, A. (1999). Important amino acid properties for enhanced thermostability from mesophilic to thermophilic proteins. *Biophysical Chemistry*, Vol. 82, No. 1, pp. 51-67.

Guzman-Maldonado, H. & Paredes-Lopez, O. (1995). Amylolytic enzymes and products derived from starch: a review. *Crit Rev Food Sci Nutr*, Vol. 35, No. 5, pp. 373-403.

Gyemant, G., Hovanszki, G. & Kandra, L. (2002). Subsite mapping of the binding region of alpha-amylases with a computer program. *Eur J Biochem*, Vol. 269, No. 21, pp. 5157-5162.

Halabi, N., Rivoire, O., Leibler, S. & Ranganathan, R. (2009). Protein sectors: evolutionary units of three-dimensional structure. *Cell*, Vol. 138, No. 4, pp. 774-786, 1097-4172 (Electronic) 0092-8674 (Linking).

Hancock, S.M., Corbett, K., Fordham-Skelton, A.P., Gatehouse, J.A. & Davis, B.G. (2005). Developing promiscuous glycosidases for glycoside synthesis: residues W433 and E432 in Sulfolobus solfataricus beta-glycosidase are important glucoside- and galactoside-specificity determinants. *Chembiochem*, Vol. 6, No. 5, pp. 866-875, 1439-4227 (Print) 1439-4227 (Linking).

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J*, Vol. 280 (Pt 2), No., pp. 309-316, 0264-6021 (Print) 0264-6021 (Linking).

Henrissat, B. & Bairoch, A. (1993). New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J*, Vol. 293 (Pt 3), No., pp. 781-788, 0264-6021 (Print) 0264-6021 (Linking).

Henrissat, B. & Bairoch, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. *Biochem J*, Vol. 316 (Pt 2), No., pp. 695-696, 0264-6021 (Print).

Henrissat, B. & Davies, G. (1997). Structural and sequence-based classification of glycoside hydrolases. *Curr Opin Struct Biol*, Vol. 7, No. 5, pp. 637-644.

Hibbert, E.G., Baganz, F., Hailes, H.C., Ward, J.M., Lye, G.J., Woodley, J.M. & Dalby, P.A. (2005). Directed evolution of biocatalytic processes. *Biomol Eng*, Vol. 22, No. 1-3, pp. 11-19, 1389-0344 (Print) 1389-0344 (Linking).

Hirata, A., Adachi, M., Sekine, A., Kang, Y.N., Utsumi, S. & Mikami, B. (2004a). Structural and enzymatic analysis of soybean beta-amylase mutants with increased pH optimum. *J Biol Chem*, Vol. 279, No. 8, pp. 7287-7295, 0021-9258 (Print) 0021-9258 (Linking).

Hirata, A., Adachi, M., Utsumi, S. & Mikami, B. (2004b). Engineering of the pH optimum of Bacillus cereus beta-amylase: conversion of the pH optimum from a bacterial type to a higher-plant type. *Biochemistry*, Vol. 43, No. 39, pp. 12523-12531, 0006-2960 (Print) 0006-2960 (Linking).

Jochens, H., Aerts, D. & Bornscheuer, U.T. (2010). Thermostabilization of an esterase by alignment-guided focussed directed evolution. *Protein Eng Des Sel*, Vol. 23, No. 12, pp. 903-909, 1741-0134 (Electronic) 1741-0126 (Linking).

Jochens, H. & Bornscheuer, U.T. (2010). Natural diversity to guide focused directed evolution. *Chembiochem*, Vol. 11, No. 13, pp. 1861-1866, 1439-7633 (Electronic) 1439-4227 (Linking).

Joerger, A.C., Mayer, S. & Fersht, A.R. (2003). Mimicking natural evolution in vitro: an N-acetylneuraminate lyase mutant with an increased dihydrodipicolinate synthase activity. *Proc Natl Acad Sci U S A*, Vol. 100, No. 10, pp. 5694-5699.

Jurgens, C., Strom, A., Wegener, D., Hettwer, S., Wilmanns, M. & Sterner, R. (2000). Directed evolution of a (beta alpha)8-barrel enzyme to catalyze related reactions in two different metabolic pathways. *Proc Natl Acad Sci U S A*, Vol. 97, No. 18, pp. 9925-9930.

Kadziola, A., Sogaard, M., Svensson, B. & Haser, R. (1998). Molecular structure of a barley alpha-amylase-inhibitor complex: implications for starch binding and catalysis. *J Mol Biol*, Vol. 278, No. 1, pp. 205-217.

Kandra, L., Gyemant, G., Remenyik, J., Hovanszki, G. & Liptak, A. (2002). Action pattern and subsite mapping of Bacillus licheniformis alpha- amylase (BLA) with modified maltooligosaccharide substrates. *FEBS Lett*, Vol. 518, No. 1-3, pp. 79-82.

Kaplan, J. & DeGrado, W.F. (2004). De novo design of catalytic proteins. *Proc Natl Acad Sci U S A*, Vol. 101, No. 32, pp. 11566-11570.

Kim, T.J., Park, C.S., Cho, H.Y., Cha, S.S., Kim, J.S., Lee, S.B., Moon, T.W., Kim, J.W., Oh, B.H. & Park, K.H. (2000). Role of the glutamate 332 residue in the transglycosylation activity of ThermusMaltogenic amylase. *Biochemistry*, Vol. 39, No. 23, pp. 6773-6780, 0006-2960 (Print).

Kortemme, T., Kim, D.E. & Baker, D. (2004). Computational alanine scanning of protein-protein interfaces. *Sci STKE*, Vol. 2004, No. 219, pp. pl2, 1525-8882 (Electronic) 1525-8882 (Linking).

Kristan, K., Stojan, J., Adamski, J. & Lanišnik Rižner, T. (2007). Rational design of novel mutants of fungal 17β-hydroxysteroid dehydrogenase. *J Biotechnol*, Vol. 129, No. 1, pp. 123-130.

Kuhlman, B., Dantas, G., Ireton, G.C., Varani, G., Stoddard, B.L. & Baker, D. (2003). Design of a novel globular protein fold with atomic-level accuracy. *Science*, Vol. 302, No. 5649, pp. 1364-1368, 1095-9203 (Electronic) 0036-8075 (Linking).

Kuipers, R.K., Joosten, H.J., van Berkel, W.J., Leferink, N.G., Rooijen, E., Ittmann, E., van Zimmeren, F., Jochens, H., Bornscheuer, U., Vriend, G., dos Santos, V.A. & Schaap, P.J. (2010). 3DM: systematic analysis of heterogeneous superfamily data to discover protein functionalities. *Proteins*, Vol. 78, No. 9, pp. 2101-2113, 1097-0134 (Electronic) 0887-3585 (Linking).

Kumar, S. & Nussinov, R. (2001). How do thermophilic proteins deal with heat? *Cell Mol Life Sci*, Vol. 58, No. 9, pp. 1216-1233, 1420-682X (Print) 1420-682X (Linking).

Kumar, S., Tsai, C.J. & Nussinov, R. (2000). Factors enhancing protein thermostability. *Protein Eng*, Vol. 13, No. 3, pp. 179-191, 0269-2139 (Print) 0269-2139 (Linking).

Kuriki, T., Kaneko, H., Yanase, M., Takata, H., Shimada, J., Handa, S., Takada, T., Umeyama, H. & Okada, S. (1996). Controlling substrate preference and transglycosylation activity of neopullulanase by manipulating steric constraint and hydrophobicity in active center. *J Biol Chem*, Vol. 271, No. 29, pp. 17321-17329.

Ladenstein, R. & Antranikian, G. (1998). Proteins from hyperthermophiles: stability and enzymatic catalysis close to the boiling point of water. *Adv Biochem Eng Biotechnol*, Vol. 61, No., pp. 37-85.

Landwehr, M., Carbone, M., Otey, C.R., Li, Y. & Arnold, F.H. (2007). Diversification of catalytic function in a synthetic family of chimeric cytochrome p450s. *Chem Biol*, Vol. 14, No. 3, pp. 269-278, 1074-5521 (Print).

Leemhuis, H., Rozeboom, H.J., Dijkstra, B.W. & Dijkhuizen, L. (2004). Improved thermostability of bacillus circulans cyclodextrin glycosyltransferase by the introduction of a salt bridge. *Proteins*, Vol. 54, No. 1, pp. 128-134, 1097-0134 (Electronic) 0887-3585 (Linking).

Lehmann, M., Kostrewa, D., Wyss, M., Brugger, R., D'Arcy, A., Pasamontes, L. & van, L.A. (2000). From DNA sequence to improved functionality: using protein sequence comparisons to rapidly design a thermostable consensus phytase. *Protein Eng*, Vol. 13, No. 1, pp. 49-57.

Lehmann, M. & Wyss, M. (2001). Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. *Curr Opin Biotechnol*, Vol. 12, No. 4, pp. 371-375.

Levy, M. & Ellington, A.D. (2001). Selection of deoxyribozyme ligases that catalyze the formation of an unnatural internucleotide linkage. *Bioorg Med Chem*, Vol. 9, No. 10, pp. 2581-2587.

Li, Y., Drummond, D.A., Sawayama, A.M., Snow, C.D., Bloom, J.D. & Arnold, F.H. (2007). A diverse family of thermostable cytochrome P450s created by recombination of stabilizing fragments. *Nat Biotechnol*, Vol. 25, No. 9, pp. 1051-1056, 1087-0156 (Print).

Lichtarge, O., Bourne, H.R. & Cohen, F.E. (1996). An evolutionary trace method defines binding surfaces common to protein families. *J Mol Biol*, Vol. 257, No. 2, pp. 342-358, 0022-2836 (Print) 0022-2836 (Linking).

Lichtarge, O., Yao, H., Kristensen, D.M., Madabushi, S. & Mihalek, I. (2003). Accurate and scalable identification of functional sites by evolutionary tracing. *J Struct Funct Genomics*, Vol. 4, No. 2-3, pp. 159-166, 1345-711X (Print) 1345-711X (Linking).

Liebl, W., Stemplinger, I. & Ruile, P. (1997). Properties and gene structure of the Thermotoga maritima alpha-amylase AmyA, a putative lipoprotein of a hyperthermophilic bacterium. *J Bacteriol*, Vol. 179, No. 3, pp. 941-948.

Liu, Y.H., Lu, F.P., Li, Y., Yin, X.B., Wang, Y. & Gao, C. (2008). Characterisation of mutagenised acid-resistant alpha-amylase expressed in Bacillus subtilis WB600. *Appl Microbiol Biotechnol*, Vol. 78, No. 1, pp. 85-94, 0175-7598 (Print) 0175-7598 (Linking).

MacGregor, E.A. (1988). Alpha-amylase structure and activity. *J Protein Chem*, Vol. 7, No. 4, pp. 399-415, 0277-8033 (Print) 0277-8033 (Linking).

Mackay, R.M., Baird, S., Dove, M.J., Erratt, J.A., Gines, M., Moranelli, F., Nasim, A., Willick, G.E., Yaguchi, M. & Seligy, V.L. (1985). Glucanase gene diversity in prokaryotic and eukaryotic organisms. *Biosystems*, Vol. 18, No. 3-4, pp. 279-292, 0303-2647 (Print) 0303-2647 (Linking).

Machius, M., Declerck, N., Huber, R. & Wiegand, G. (1998). Activation of Bacillus licheniformis alpha-amylase through a disorder-- >order transition of the substrate-binding site mediated by a calcium- sodium-calcium metal triad. *Structure*, Vol. 6, No. 3, pp. 281-292.

Machius, M., Wiegand, G. & Huber, R. (1995). Crystal structure of calcium-depleted *Bacillus licheniformis* alpha- amylase at 2.2 A resolution. *J Mol Biol*, Vol. 246, No. 4, pp. 545-559.

Mansfeld, J., Vriend, G., Dijkstra, B.W., Veltman, O.R., Van den Burg, B., Venema, G., Ulbrich-Hofmann, R. & Eijsink, V.G. (1997). Extreme stabilization of a thermolysin-like protease by an engineered disulfide bond. *J Biol Chem*, Vol. 272, No. 17, pp. 11152-11156.

Matsubara, S. (1961). Studies on Taka-amylase A. VII. Transmaltosidation by Taka-amylase A. *J Biochem*, Vol. 49, No. 3, pp. 226-231.

Matsui, I., Ishikawa, K., Miyairi, S., Fukui, S. & Honda, K. (1992a). Alteration of bond-cleavage pattern in the hydrolysis catalyzed by Saccharomycopsis alpha-amylase altered by site-directed mutagenesis. *Biochemistry*, Vol. 31, No. 22, pp. 5232-5236.

Matsui, I., Ishikawa, K., Miyairi, S., Fukui, S. & Honda, K. (1992b). A mutant alpha-amylase with enhanced activity specific for short substrates. *FEBS Lett*, Vol. 310, No. 3, pp. 216-218.

Matsumura, I. & Ellington, A.D. (2001). In vitro evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates. *J Mol Biol*, Vol. 305, No. 2, pp. 331-339.

Maxwell, K.L. & Davidson, A.R. (1998). Mutagenesis of a Buried Polar Interaction in an SH3 Domain:‰ Sequence Conservation Provides the Best Prediction of Stability Effects€ *Biochemistry*, Vol. 37, No. 46, pp. 16172-16182.

McCarter, J.D. & Withers, S.G. (1994). Mechanisms of enzymatic glycoside hydrolysis. *Curr Opin Struct Biol*, Vol. 4, No., pp. 885-892.

Meiler, J. & Baker, D. (2006). ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility. *Proteins*, Vol. 65, No. 3, pp. 538-548, 1097-0134 (Electronic) 0887-3585 (Linking).

Minshull, J. & Stemmer, W.P. (1999). Protein evolution by molecular breeding. *Curr Opin Chem Biol*, Vol. 3, No. 3, pp. 284-290.

Moreno-Beltran, A., Salgado, L., Vazquez-Duhalt, R. & López-Munguía, A. (1999). Modelling the alcoholysis reaction of beta-galactosidase in reverse micelles. *J Mol Catalysis B: Enzymatic*, Vol. 6, No. 1-2, pp. 1-10.

Moreno, A., Damian-Almazo, J.Y., Miranda, A., Saab-Rincon, G., Gonzalez, F. & Lopez-Munguia, A. (2010). Transglycosylation reactions of Thermotoga maritima [alpha]-amylase. *Enzyme and Microbial Technology*, Vol. 46, No. 5, pp. 331-337.

Morley, K.L. & Kazlauskas, R.J. (2005). Improving enzyme properties: when are closer mutations better? *Trends Biotechnol*, Vol. 23, No. 5, pp. 231-237, 0167-7799 (Print) 0167-7799 (Linking).

Nagi, A.D. & Regan, L. (1997). An inverse correlation between loop length and stability in a four-helix-bundle protein. *Fold Des*, Vol. 2, No. 1, pp. 67-75, 1359-0278 (Print) 1359-0278 (Linking).

Nakajima, R., Imanaka, T. & Aiba, S. (1986). Comparison of amino acid sequences of eleven different α-amylases. *Appl Microbiol Biotechnol*, Vol. 23, No. 5, pp. 355-360.

Nielsen, J.E. & Borchert, T.V. (2000). Protein engineering of bacterial alpha-amylases. *Biochim Biophys Acta*, Vol. 1543, No. 2, pp. 253-274.

Nikolova, P.V., Henckel, J., Lane, D.P. & Fersht, A.R. (1998). Semirational design of active tumor suppressor p53 DNA binding domain with enhanced stability. *Proc Natl Acad Sci U S A*, Vol. 95, No. 25, pp. 14675-14680, 0027-8424 (Print) 0027-8424 (Linking).

Ochoa-Leyva, A., Soberon, X., Sanchez, F., Arguello, M., Montero-Moran, G. & Saab-Rincon, G. (2009). Protein design through systematic catalytic loop exchange in the

(beta/alpha)8 fold. *J Mol Biol*, Vol. 387, No. 4, pp. 949-964, 1089-8638 (Electronic) 0022-2836 (Linking).

Palackal, N., Brennan, Y., Callen, W.N., Dupree, P., Frey, G., Goubet, F., Hazlewood, G.P., Healey, S., Kang, Y.E., Kretz, K.A., Lee, E., Tan, X., Tomlinson, G.L., Verruto, J., Wong, V.W., Mathur, E.J., Short, J.M., Robertson, D.E. & Steer, B.A. (2004). An evolutionary route to xylanase process fitness. *Protein Sci*, Vol. 13, No. 2, pp. 494-503, 0961-8368 (Print) 0961-8368 (Linking).

Pantazes, R.J., Saraf, M.C. & Maranas, C.D. (2007). Optimal protein library design using recombination or point mutations based on sequence-based scoring functions. *Protein Eng Des Sel*, Vol. 20, No. 8, pp. 361-373, 1741-0126 (Print).

Paramesvaran, J., Hibbert, E.G., Russell, A.J. & Dalby, P.A. (2009). Distributions of enzyme residues yielding mutants with improved substrate specificities from two different directed evolution strategies. *Protein Eng Des Sel*, Vol. 22, No. 7, pp. 401-411, 1741-0134 (Electronic) 1741-0126 (Linking).

Park, H.S., Nam, S.H., Lee, J.K., Yoon, C.N., Mannervik, B., Benkovic, S.J. & Kim, H.S. (2006). Design and evolution of new catalytic activity with an existing protein scaffold. *Science*, Vol. 311, No. 5760, pp. 535-538.

Park, S., Morley, K.L., Horsman, G.P., Holmquist, M., Hult, K. & Kazlauskas, R.J. (2005). Focusing mutations into the P. fluorescens esterase binding site increases enantioselectivity more effectively than distant mutations. *Chem Biol*, Vol. 12, No. 1, pp. 45-54, 1074-5521 (Print) 1074-5521 (Linking).

Pascarella, S. & Argos, P. (1992). Analysis of insertions/deletions in protein structures. *J Mol Biol*, Vol. 224, No. 2, pp. 461-471, 0022-2836 (Print) 0022-2836 (Linking).

Pavelka, A., Chovancova, E. & Damborsky, J. (2009). HotSpot Wizard: a web server for identification of hot spots in protein engineering. *Nucleic Acids Res*, Vol. 37, No. Web Server issue, pp. W376-383, 1362-4962 (Electronic) 0305-1048 (Linking).

Perl, D., Mueller, U., Heinemann, U. & Schmid, F.X. (2000). Two exposed amino acid residues confer thermostability on a cold shock protein. *Nat Struct Biol*, Vol. 7, No. 5, pp. 380-383, 1072-8368 (Print) 1072-8368 (Linking).

Reetz, M.T., Bocola, M., Carballeira, J.D., Zha, D. & Vogel, A. (2005). Expanding the range of substrate acceptance of enzymes: combinatorial active-site saturation test. *Angew Chem Int Ed Engl*, Vol. 44, No. 27, pp. 4192-4196, 1433-7851 (Print) 1433-7851 (Linking).

Reetz, M.T., Carballeira, J.D., Peyralans, J., Hobenreich, H., Maichele, A. & Vogel, A. (2006a). Expanding the substrate scope of enzymes: combining mutations obtained by CASTing. *Chemistry*, Vol. 12, No. 23, pp. 6031-6038, 0947-6539 (Print).

Reetz, M.T., Carballeira, J.D. & Vogel, A. (2006b). Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. *Angew Chem Int Ed Engl*, Vol. 45, No. 46, pp. 7745-7751, 1433-7851 (Print).

Reetz, M.T., Soni, P. & Fernandez, L. (2009). Knowledge-guided laboratory evolution of protein thermolability. *Biotechnol Bioeng*, Vol. 102, No. 6, pp. 1712-1717, 1097-0290 (Electronic) 0006-3592 (Linking).

Reetz, M.T., Soni, P., Fernandez, L., Gumulya, Y. & Carballeira, J.D. (2010). Increasing the stability of an enzyme toward hostile organic solvents by directed evolution based on iterative saturation mutagenesis using the B-FIT method. *Chem Commun (Camb)*, Vol. 46, No. 45, pp. 8657-8658, 1364-548X (Electronic) 1359-7345 (Linking).

Rivera, M.H., Lopez-Munguia, A., Soberon, X. & Saab-Rincon, G. (2003). Alpha-amylase from Bacillus licheniformis mutants near to the catalytic site: effects on hydrolytic and transglycosylation activity. *Protein Eng*, Vol. 16, No. 7, pp. 505-514.

Rodríguez-Zavala, J.S. (2008). Enhancement of coenzyme binding by a single point mutation at the coenzyme binding domain of E. coli lactaldehyde dehydrogenase. *Protein Science*, Vol. 17, No. 3, pp. 563-570.

Rodríguez, M., Gómez, A., González, F., Bárzana, E. & López-Munguía, A. (1996). Selectivity of methyl-fructoside synthesis with β-fructufuranosidase. *Appl Biochem Biotechnol*, Vol. 59, No., pp. 167-175.

Rogers, J.C. (1985). Conserved amino acid sequence domains in alpha-amylases from plants, mammals, and bacteria. *Biochemical and Biophysical Research Communications*, Vol. 128, No. 1, pp. 470-476.

Rosell, A., Valencia, E., Ochoa, W.F., Fita, I., Pares, X. & Farres, J. (2003). Complete Reversal of Coenzyme Specificity by Concerted Mutation of Three Consecutive Residues in Alcohol Dehydrogenase. *J. Biol. Chem.*, Vol. 278, No. 42, pp. 40573-40580.

Saab-Rincon, G., del-Rio, G., Santamaria, R.I., Lopez-Munguia, A. & Soberon, X. (1999). Introducing transglycosylation activity in a liquefying alpha-amylase. *FEBS Lett*, Vol. 453, No. 1-2, pp. 100-106.

Sakamoto, T., Joern, J.M., Arisawa, A. & Arnold, F.H. (2001). Laboratory evolution of toluene dioxygenase to accept 4-picoline as a substrate. *Appl Environ Microbiol*, Vol. 67, No. 9, pp. 3882-3887.

Sali, A., Potterton, L., Yuan, F., van, V.H. & Karplus, M. (1995). Evaluation of comparative protein modeling by MODELLER. *Proteins Struct Funct Genet*, Vol. 23, No. 3, pp. 318-326.

Sanchez, R. & Sali, A. (1997). Evaluation of comparative protein structure modeling by MODELLER-3. *Proteins Struct Funct Genet*, Vol. 1997, No. 50, pp. 50-58.

Santamaria, R.I., Del Rio, G., Saab, G., Rodriguez, M.E., Soberon, X. & Lopez-Manguia, A. (1999). Alcoholysis reactions from starch with alpha-amylases. *FEBS Lett*, Vol. 452, No. 3, pp. 346-350.

Saravanan, M., Vasu, K. & Nagaraja, V. (2008). Evolution of sequence specificity in a restriction endonuclease by a point mutation. *Proc Natl Acad Sci U S A*, Vol. 105, No. 30, pp. 10344-10347, 1091-6490 (Electronic) 0027-8424 (Linking).

Savile, C.K., Janey, J.M., Mundorff, E.C., Moore, J.C., Tam, S., Jarvis, W.R., Colbeck, J.C., Krebber, A., Fleitz, F.J., Brands, J., Devine, P.N., Huisman, G.W. & Hughes, G.J. (2010). Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science*, Vol. 329, No. 5989, pp. 305-309, 1095-9203 (Electronic) 0036-8075 (Linking).

Schmitzer, A.R., Lepine, F. & Pelletier, J.N. (2004). Combinatorial exploration of the catalytic site of a drug-resistant dihydrofolate reductase: creating alternative functional configurations. *Protein Eng Des Sel*, Vol. 17, No. 11, pp. 809-819, 1741-0126 (Print) 1741-0126 (Linking).

Shinoyama, H., Kamiyama, Y. & Yasui, T. (1988). Enzymatic synthesis of alkyl β–xylosides from xylobiose by application of the transxylosyl reaction of *Aspergillus niger* β-xylosidase. *Agric Biol Chem*, Vol. 52, No., pp. 2197-2202.

Siegel, J.B., Zanghellini, A., Lovick, H.M., Kiss, G., Lambert, A.R., St Clair, J.L., Gallaher, J.L., Hilvert, D., Gelb, M.H., Stoddard, B.L., Houk, K.N., Michael, F.E. & Baker, D. (2010). Computational design of an enzyme catalyst for a stereoselective

bimolecular Diels-Alder reaction. *Science*, Vol. 329, No. 5989, pp. 309-313, 1095-9203 (Electronic) 0036-8075 (Linking).

Sinnott, M.L. (1990). Catalytic Mechanisms of Enzymic Glycosyl Transfer. *Chem Rev*, Vol. 90, No., pp. 1171-1202.

Song, J.K., Chung, B., Oh, Y.H. & Rhee, J.S. (2002). Construction of DNA-shuffled and incrementally truncated libraries by a mutagenic and unidirectional reassembly method: changing from a substrate specificity of phospholipase to that of lipase. *Appl Environ Microbiol*, Vol. 68, No. 12, pp. 6146-6151.

Song, J.K. & Rhee, J.S. (2000). Simultaneous enhancement of thermostability and catalytic activity of phospholipase A(1) by evolutionary molecular engineering. *Appl Environ Microbiol*, Vol. 66, No. 3, pp. 890-894.

Spector, S., Wang, M., Carp, S.A., Robblee, J., Hendsch, Z.S., Fairman, R., Tidor, B. & Raleigh, D.P. (2000). Rational modification of protein stability by the mutation of charged surface residues. *Biochemistry*, Vol. 39, No. 5, pp. 872-879, 0006-2960 (Print) 0006-2960 (Linking).

Stemmer, W.P. (1994a). DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc Natl Acad Sci U S A*, Vol. 91, No. 22, pp. 10747-10751.

Stemmer, W.P. (1994b). Rapid evolution of a protein in vitro by DNA shuffling. *Nature*, Vol. 370, No. 6488, pp. 389-391.

Straathof, A.J.J., Vrijenhoef, J.P., Sprangers, E.P.A.T., Bekkum, H.v. & Kieboom, A.P.G. (1988). Enzymic formation of β-D-fructofuranosides from sucrose: activity and selectivity of invertase in mixtures of water and alcohol. *J Carbohydr Chem*, Vol. 7, No., pp. 223-238.

Suel, G.M., Lockless, S.W., Wall, M.A. & Ranganathan, R. (2003). Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat Struct Biol*, Vol. 10, No. 1, pp. 59-69, 1072-8368 (Print) 1072-8368 (Linking).

Sullivan, B.J., Durani, V. & Magliery, T.J. (2011). Triosephosphate isomerase by consensus design: dramatic differences in physical properties and activity of related variants. *J Mol Biol*, Vol. 413, No. 1, pp. 195-208, 1089-8638 (Electronic) 0022-2836 (Linking).

Svensson, B. (1994). Protein engineering in the alpha-amylase family: catalytic mechanism, substrate specificity, and stability. *Plant Mol Biol*, Vol. 25, No. 2, pp. 141-157.

Turner, N.J. (2009). Directed evolution drives the next generation of biocatalysts. *Nat Chem Biol*, Vol. 5, No. 8, pp. 567-573, 1552-4469 (Electronic) 1552-4450 (Linking).

Uchiyama, H., Inaoka, T., Ohkuma-Soyejima, T., Togame, H., Shibanaka, Y., Yoshimoto, T. & Kokubo, T. (2000). Directed evolution to improve the thermostability of prolyl endopeptidase. *J Biochem (Tokyo)*, Vol. 128, No. 3, pp. 441-447.

Ulmer, K.M. (1983). Protein engineering. *Science*, Vol. 219, No. 4585, pp. 666-671, 0036-8075 (Print) 0036-8075 (Linking).

Vallee, B.L., Stein, E.A., Sumerwell, W.N. & Fischer, E.H. (1959). Metal content of alpha-amylases of various origins. *J Biol Chem*, Vol. 234, No., pp. 2901-2905, 0021-9258 (Print) 0021-9258 (Linking).

van der Veen, B.A., Leemhuis, H., Kralj, S., Uitdehaag, J.C.M., Dijkstra, B.W. & Dijkhuizen, L. (2001). Hydrophobic Amino Acid Residues in the Acceptor Binding Site Are Main Determinants for Reaction Mechanism and Specificity of Cyclodextrin-glycosyltransferase. *J. Biol. Chem.*, Vol. 276, No. 48, pp. 44557-44562.

Vihinen, M. & Mantsala, P. (1989). Microbial amylolytic enzymes. *Crit Rev Biochem Mol Biol*, Vol. 24, No. 4, pp. 329-418.

Vihinen, M. & Mantsala, P. (1990). Conserved residues of liquefying alpha-amylases are concentrated in the vicinity of active site. *Biochem Biophys Res Commun*, Vol. 166, No. 1, pp. 61-65.

Vulfson, E.N., Patel, R., Beecher, J.E., Andrews, A.T. & Law, B.A. (1990). Glycosidases in organic solvents: I. Alkyl −β−glucoside synthesis in a water-organic two-phase system. *Enzyme Microb Technol*, Vol. 12, No., pp. 950-954.

Wilming, M., Iffland, A., Tafelmeyer, P., Arrivoli, C., Saudan, C. & Johnsson, K. (2002). Examining reactivity and specificity of cytochrome c peroxidase by using combinatorial mutagenesis. *Chembiochem*, Vol. 3, No. 11, pp. 1097-1104, 1439-4227 (Print) 1439-4227 (Linking).

Woodyer, R., van der Donk, W.A. & Zhao, H. (2003). Relaxing the nicotinamide cofactor specificity of phosphite dehydrogenase by rational design. *Biochemistry*, Vol. 42, No. 40, pp. 11604-11614, 0006-2960 (Print) 0006-2960 (Linking).

Yamashiro, K., Yokobori, S., Koikeda, S. & Yamagishi, A. (2010). Improvement of Bacillus circulans beta-amylase activity attained using the ancestral mutation method. *Protein Eng Des Sel*, Vol. 23, No. 7, pp. 519-528, 1741-0134 (Electronic) 1741-0126 (Linking).

Yin, Q., Teng, Y., Ding, M. & Zhao, F. (2011). Site-directed mutagenesis of aromatic residues in the carbohydrate-binding module of Bacillus endoglucanase EGA decreases enzyme thermostability. *Biotechnol Lett*, Vol. 33, No. 11, pp. 2209-2216, 1573-6776 (Electronic) 0141-5492 (Linking).

Zanghellini, A., Jiang, L., Wollacott, A.M., Cheng, G., Meiler, J., Althoff, E.A., Rothlisberger, D. & Baker, D. (2006). New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci*, Vol. 15, No. 12, pp. 2785-2794, 0961-8368 (Print) 0961-8368 (Linking).

Zhang, J.H., Dawes, G. & Stemmer, W.P. (1997). Directed evolution of a fucosidase from a galactosidase by DNA shuffling and screening. *Proc Natl Acad Sci U S A*, Vol. 94, No. 9, pp. 4504-4509.

Zhao, H. & Arnold, F.H. (1999). Directed evolution converts subtilisin E into a functional equivalent of thermitase. *Protein Eng*, Vol. 12, No. 1, pp. 47-53.

Zhao, H., Giver, L., Shao, Z., Affholter, J.A. & Arnold, F.H. (1998). Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat Biotechnol*, Vol. 16, No. 3, pp. 258-261.

New Tools or Approaches for Molecular Genetics

Using Cys-Scanning Analysis Data in the Study of Membrane Transport Proteins

Stathis Frillingos
University of Ioannina Medical School
Greece

1. Introduction

Membrane transport proteins represent a core group of gene products in all known genomes and play crucial roles in either human physiology and disease or diverse environmental adaptations of microorganisms. Despite their importance, analyses of such proteins have long been hindered by a lack of high-resolution models, which reflects the inherent problems of studying membrane transport proteins outside the membrane. Recently, however, progress with crystallography and structural modeling of membrane transport proteins and expanding information on new sequence entries from genome analyses have set the stage for more systematic mutagenesis approaches to link high-resolution structural data with functional evidence.

The current structural insight on secondary transporters and their classification raises new fundamental questions on the relationships of structure with function. A crucial aspect is that many of the functionally divergent homologs, or even separate families of secondary transporters, are evolutionarily and structurally related. Thus, the majority of known structures fall in only two common folds: lactose permease (LacY) and the neurotransmitter-sodium symporter prototype (LeuT). Other transporters with different folds still display the core feature of organization in structural repeats that coordinate to form the dynamic binding site (Boudker & Verdon, 2010). The binding site operates through the commonly accepted alternating access mechanism, which is now explained as involving both rocking movements of domains pivoting at the binding site and local motions of outside and inside gates flanking the binding site, leading to alterations between outward-facing, inward-facing and intermediate substrate-occluded conformation states (Forrest et al., 2011). Apart from the structural knowledge, systematic insight on the secondary transport mechanisms requires the concerted use of structural and functional approaches, as was achieved in the seminal case of lactose permease (Kaback et al., 2011). This is important to note, since the X-ray structures represent static snapshots of highly dynamic proteins outside their native membrane environment; and, with few exceptions (Weyand et al., 2011), interpretations have been based on compilations of different snapshots from different structural homologs.

In the context of the recent structural evidence, it is striking that transport protein families tend to display high evolutionary conservation in sequence and overall structure; but they also display high functional variations between homologs, implying that relatively few side-chain changes may account for key local effects on active-site conformation and function.

However, mutations responsible for such evolutionary plasticity are rarely discernible in the background of many optimizing, permissive or near-neutral mutations that have accumulated in the present-day sequences over evolutionary time (Harms & Thornton, 2010). Traditional structure-function analysis often fails to annotate such important residues because the mutations may be uninformative or lead to complete loss-of-function/structure phenotypes in the native background. Therefore, a rationally designed mutagenesis study of one particular homolog may yield important information on the overall active-site architecture and mechanism; but it is not enough to reveal the spectrum of molecular determinants dictating the different specificity trends within the family.

In practice, it is common that an evolutionarily-broad transporter family consists of several hundreds or even thousands of related members. They share the same structural fold and binding-site architecture based on X-ray crystallography of one, usually distal, homolog. Few of the members might have been characterized extensively with respect to function, and only one or two might have been studied rigorously with site-directed mutagenesis approaches. How then can we explore the whole spectrum of specificities in such families and derive more essential information on the molecular basis of the different functional profiles?

One approach to this research problem is based on data derived from Cys-scanning analysis. Cys-scanning mutagenesis has already proven to be an essential tool (often referred to as the "gold standard") for the study of structure-function relationships in membrane proteins (Frillingos et al., 1998). In several separate examples of membrane transporters, it has provided valuable insight on active site conformation and function, even before an X-ray structure has become available (Chen & Rudnick, 2000; Kaback et al., 2001; Sorgen et al., 2002; Zomot et al., 2002). Revisiting Cys-scanning evidence after elucidation of a corresponding high-resolution structure has yielded important novel implications on the mechanism (Crisman et al., 2009; Forrest et al., 2008; Guan & Kaback, 2006; Kaback et al., 2007, 2011). The site-specific knowledge derived from Cys-scanning analysis of a specific transporter can also be used, in principle, to design rapid and cost-effective mutagenesis strategies for the functional analysis of different, structurally-related homologs.

The rationale of this approach is to use data from a systematic Cys-scanning analysis of one homolog (the study prototype) in combination, when applicable, with homology-modeling to select new homologs for targeted mutagenesis studies. This approach is based on differences and the extent of conservation, not in the overall sequence, but in the subset of residues delineated as putatively important from the Cys-scanning data. These putatively important residues correspond to positions of the study prototype where a native amino acid is irreplaceable or replaceable with only few side chains or is sensitive to site-specific alkylation of a substituted Cys leading to inactivation. In several experimental paradigms, such positions have been shown to be (a) relatively few (less than 15% of the total number of residues, in general); (b) much more highly conserved among homologs than in the rest of the protein and often mapping within conserved motif sequences (for example, see Georgopoulou et al., 2010); and (c) linked either directly or indirectly with the substrate binding site conformation and function (Kaback et al., 2007). This latter property allows the use of this set of positions as targets of rationalized mutagenesis in new homologs, in order not only to provide a measure of the functional conservation between different homologs, but also to delineate determinants responsible for particular switches in substrate preference or specificity.

In this chapter, I first analytically explain the rationale, discrete steps and aims of the Cys-scanning-based approach. I then describe applications in two families of ion gradient-driven membrane transporters. The first is the oligosaccharide-proton symporter family (OHS), which includes the lactose permease LacY and other closely related sugar transporters. The second is the nucleobase-ascorbate transporter family (NAT/NCS2), which is evolutionarily ubiquitous and encompasses more distantly related homologs and specificities. It is important to emphasize that such an experimental approach, although conceptually sound, has seen limited applications in the field of membrane transport proteins to date. A more systematic and generalized application of this approach is expected to have a major impact on the field, since it should allow rapid and effective mutagenesis designs. The impact will occur even in the absence of a high-resolution model, provided only that Cys-scanning analysis has been performed for one of the transporter homologs.

2. Rationale, discrete steps, and aims of the approach

Cys-scanning mutagenesis is a well-established strategy for structure-function analysis of proteins. It has proven particularly useful and provided valuable insight for the analysis of polytopic membrane proteins and, in particular, membrane transporters. Cys-scanning protocols rely on the engineering and availability of functional protein variants that are devoid of all or part of the native Cys residues (Cys-less or Cys-depleted versions, respectively) and the use of these Cys-less or Cys-depleted versions as a background for site-specific mutagenesis to introduce new single-Cys replacements at selected positions. The term *scanning* derives from the common application of this strategy to individually replace each amino acid residue in a contiguous sequence portion or even in the whole sequence of a protein with Cys and create an extensive library of single-Cys replacement mutants for this protein. In addition, a battery of different site-directed techniques can be applied to probe specific features of each Cys-substituted position (accessibility to solvent, relevance to substrate binding, sensitivity to the conformational changes of turnover, proximity to other sites in the protein) with appropriate sulfhydryl-specific reagents. Thus, Cys-scanning analysis often yields a wealth of data that are used to build comprehensive structure-mechanistic models for the protein under study, even in the absence of high-resolution crystallographic evidence (Frillingos et al., 1998; Sorgen et al., 2002; Tamura et al., 2003).

Nevertheless, the research potential of Cys-scanning analysis is not limited to the systematic study of structure-function relationships of individual proteins. The evidence derived from Cys-scanning analysis of a study prototype can serve as a basis to materialize rapid and effective mutagenesis designs in new, previously unknown or unstudied proteins that are related to the study prototype by sequence or structure homology. In this way, evolutionarily broad families of related proteins can be studied with respect to their active site consensus architecture and function, the spectrum of different specificity trends and mechanistic deviations, and the molecular determinants responsible for these differences. New homologs and mutagenesis targets must be selected appropriately, on the basis of the prototypic Cys-scanning evidence. Such a Cys-scanning-based approach may prove extremely contributory in the field of membrane transporters. It can increase the low representation of experimentally characterized homologs in most transporter families and reduce the paucity of crystallographic structural models, as described in the *Introduction* section.

Five discrete steps of the approach are analyzed below, with emphasis on the rationale referring to membrane transport proteins. The probable outcomes from application of this strategy (as delineated in 2.5) are substantiated further by two research paradigms presented in section 3. The two paradigms are rather seen as pilot experimental studies which demonstrate the applicability and importance of such an approach for dissecting substrate recognition and selectivity determinants in new, *ab initio* studied transporters.

2.1 Capitalizing on the Cys-scanning mutagenesis legacy

Cys-scanning mutagenesis and site-directed cysteine modification has been widely used to elucidate structure-function relationships in membrane transport proteins. The reasons for this broad application are both practical and conceptual.

Practical reasons include:

a. the feasibility of engineering of many bacterial transporters devoid of native Cys residues (Cys-less versions) that are functionally equivalent to wild type (for example, Culham et al., 2003; Jung, H. et al., 1998; Sahin-Tóth et al., 2000; Slotboom et al., 2001; van Iwaarden et al., 1991; Weissborn et al., 1997) and represent an ideal substrate for Cys-scanning analyses, as well as the robust evidence that the vast majority of single-Cys replacement mutants do not affect dramatically the transporter expression, structural integrity or function (see Frillingos et al., 1998; Tamura et al., 2003);

b. the availability of a diverse compendium of thiol-specific reagents from established companies such as Molecular Probes, Toronto Research Chemicals, and others;

c. the range of thiol-specific reagents and strategies developed for membrane proteins, such as substituted-cysteine accessibility method (SCAM) using hydrophilic methanethiosulfonate (MTS) derivatives (Akabas et al., 1992) or other reagents (Yan & Maloney, 1993), cysteine-cysteine cross linking protocols (Wu & Kaback, 1996), site-directed fluorescence spectroscopy (Jung, K. et al., 1993; Wu et al., 1995), site-directed spin labeling (SDSL) (Voss et al., 1995), and site-directed alkylation in situ with radioactive (Frillingos & Kaback, 1996; Guan & Kaback, 2007) or fluorescent probes (Georgopoulou et al., 2010; Jiang et al., 2011);

d. the obvious advantage of Cys-scanning technologies over strategies, like Ala-scanning mutagenesis, which do not allow further site-specific derivatization of the substituted amino acid;

e. the fact that high-resolution crystallographic models did not appear for membrane transport proteins until the last two decades, *e.g.*, the ion gradient-driven transporters, for which the first X-ray structure appeared less than a decade ago (see Abramson et al., 2003), due to inherent difficulties with these hydrophobic, integral in the membrane and conformationally dynamic proteins. This delay allowed sufficient time for Cys-scanning applications to expand and provide alternative low-resolution approaches to the study of structure and mechanism (see Kaback & Wu, 1997).

Conceptual reasons include:

a. the success of systematic Cys-scanning analyses in revealing important residues of membrane transport proteins, including irreplaceable residues, binding-site residues or residues that are important conformationally for the mechanism of energy coupling. This is based on two parameters: (i) the utility of using single-Cys mutants in indicating

positions of low significance for the mechanism (active and alkylation-insensitive Cys mutants) and, at the same time, delineating the relatively few residues of potentially major significance (inactive or alkylation-sensitive Cys mutants) for more extensive study with site-directed mutagenesis; (ii) the diverse array of specific Cys modification reagents and protocols that have been developed and used to probe accessibility to solvent, relevance to substrate binding, sensitivity to the conformational changes of turnover, proximity to other sites or other functional properties for each Cys-substituted position (Frillingos et al., 1998);

b. the fact that low-resolution models derived for membrane transport proteins with Cys-scanning approaches continue to provide insight for this class of proteins, even in the post-crystallization era of research (see Kaback et al., 2007, 2011). Most characteristically, the information on the conformational dynamics of an active transport protein deduced from appropriate site-directed Cys modification assays is a valuable complement to the static crystal-structure images. Such information is always needed for an integrated insight on the transport mechanism (Kaback et al., 2011).

The wealth of data derived from the library of single-Cys, paired-Cys and other site-directed mutants produced for a particular transporter in the course of a Cys-scanning mutagenesis study can be used to design new approaches for the analysis of other homologs that might be poorly studied or even not characterized previously with respect to function. In many cases, the availability of at least one high-resolution structural prototype (from a solved X-ray structure for one homolog representing an evolutionarily broad family or group of families with structurally related transporters) might provide valuable additional information and allow the formulation of preliminary structural models. However, even in the absence of such models, the information from Cys-scanning analysis of a prototypic homolog *per se* is sufficient to guide selection of new homologs for study and of amino acid targets for effective site-directed mutagenesis designs in these homologs. The selection of new homologs depends, of course, on the research question asked. However, a common theme is to interrogate what is the structure-functional basis of particular differences in substrate selectivity or in the specificity profile for the recognition of ligands. These questions are highly relevant to the current state of the art in the field of membrane transporters, since many different transport proteins appear to be evolutionarily and structurally related; and rather small sequence changes at key residues are expected to dictate major functional differences (Boudker & Verdon, 2010; Forrest et al., 2011; Lu et al., 2011; Weyand et al., 2011; Yousef & Guan, 2009).

2.2 Selecting new homologs and mutagenesis targets to study

The first and "rate-determining" step in the Cys-scanning-based approach for the analysis of new transporter homologs is the selection of new homologs and mutagenesis targets to study. This selection is based essentially on the set of residues delineated as important from the Cys-scanning analysis of the prototypic homolog and depends, as a consequence, on the extent and results of analysis of the study prototype (Figure 1).

In principle, the Cys-scanning data refer to the functional properties of single-Cys, paired-Cys and other site-directed mutants engineered in the course of the relevant studies. A scanning experiment has two parts at minimum:

1. Analysis of Cys-replacement mutants: the scope is to delineate positions where a mutant is inactive, of very low activity or sensitive to inactivation by specific alkylating agents such as the relatively small and membrane permeable *N*-ethylmaleimide (NEM), which is commonly used to scan for alkylation-sensitive cysteines. Selected mutants are analyzed for site-directed alkylation in the presence or absence of substrate or in other conditions pertinent to appropriate mechanistic questions;

2. Further site-directed mutagenesis at positions where a single-Cys mutant presents with very low or negligible activity or high sensitivity to inactivation upon alkylation with NEM: the scope is to delineate positions where several replacements yield very low activity or different kinetics or specificity than wild type, and define a pattern of permissive and non-permissive replacements (taking into account the bulk, hydrophobicity, polarity, geometry or other properties of the side chain changes).

Overall, the two lines of experiments are expected to delineate a set of residues which are crucial for the transport mechanism of the study prototype in various respects. For example, positions where bulky replacements or alkylation of a substituted cysteine with the maleimidyl adduct lead to inactivation may reflect important conformational constraints and interactions with other parts of the protein that are essential for the permease turnover (Jiang et al., 2011; Tavoulari & Frillingos, 2008). On the other hand, residues which are replaceable with few or no other side chains and, at the same time, accommodate site-specific mutants of impaired affinity or distorted specificity for substrate may be crucial for substrate recognition and binding, while positions of Cys replacements which are protected from alkylation in the presence of substrate may be at the vicinity of the binding site. It is generally true that such important residues fall in one of the following three categories: (a) irreplaceable; (b) replaceable with few other side chains; (c) sensitive to inactivation of the Cys replacement by NEM. These three potential properties can be used to define the set of important residues of the study prototype, as deduced from Cys-scanning analysis data (Karena & Frillingos, 2011; Papakostas & Frillingos, 2012). It is also unequivocally true that this set of residues represents positions with a higher degree of side chain conservation than the rest of the protein and, in cases of transporters that have been studied thoroughly, correspond to a small percentage of the total amino acids in the sequence, usually 10-15% (Frillingos et al., 1998; Georgopoulou et al., 2010; Mermelekas et al., 2010; Tamura et al., 2003). These features make this clearly defined set of residues suitable for use as a basis to select (i) new homologs for *ab intio* study and (ii) amino acid targets for site-specific mutagenesis in these homologs (Figure 1). Homology modeling is also of great value in selecting mutagenesis targets for the new homologs provided that a prototypic crystal structure is available (for examples, see section 3). For reasons explained in the previous section, the Cys-scanning approaches yield additional dynamic information on the role of specific residues that cannot be provided by the structural models *per se*.

More explicitly, the first step of the approach involves a homology search referring not to the whole coding sequence but to the set of the important residues of the study prototype (as defined above). The aim of this search is to select new homologs (from the unknown or poorly-characterized pool of sequence entries) on the basis of specific differences in sequence, implying distinct conserved patterns that might correlate with shifts in specificity. This process resembles the search for characteristic sequence motifs that are conserved as a consensus throughout a transporter family. However, it is more effective in practice, as it is reinforced with experimental data (see Georgopoulou et al., 2010; Karatza et al., 2006; Kasho

et al., 2006; Papageorgiou et al., 2008). If the homolog under investigation is characterized functionally and turns out to be different in specificity from the prototype in an assayable manner, amino acid targets for site-directed mutagenesis of the new homolog are selected from the set of important residues of the prototype. The method is as follows:

1. Residues that are invariant between the new homolog and the study prototype are used as targets of conservative replacements or more extensive mutagenesis to provide a measure of the functional conservation for key conserved residues between the two transporters

2. Residues that differ between the new homolog and the study prototype are used as targets of replacements of each relevant position of the new homolog with the corresponding amino acid found in the prototype, in an attempt to modulate specificity or other substrate-recognition properties and draw comprehensive conclusions on the determinants of the different transporter preferences. Such mutagenesis designs may need to include combinatorial replacements and/or construction of cross-homolog chimeras, as described in the next section.
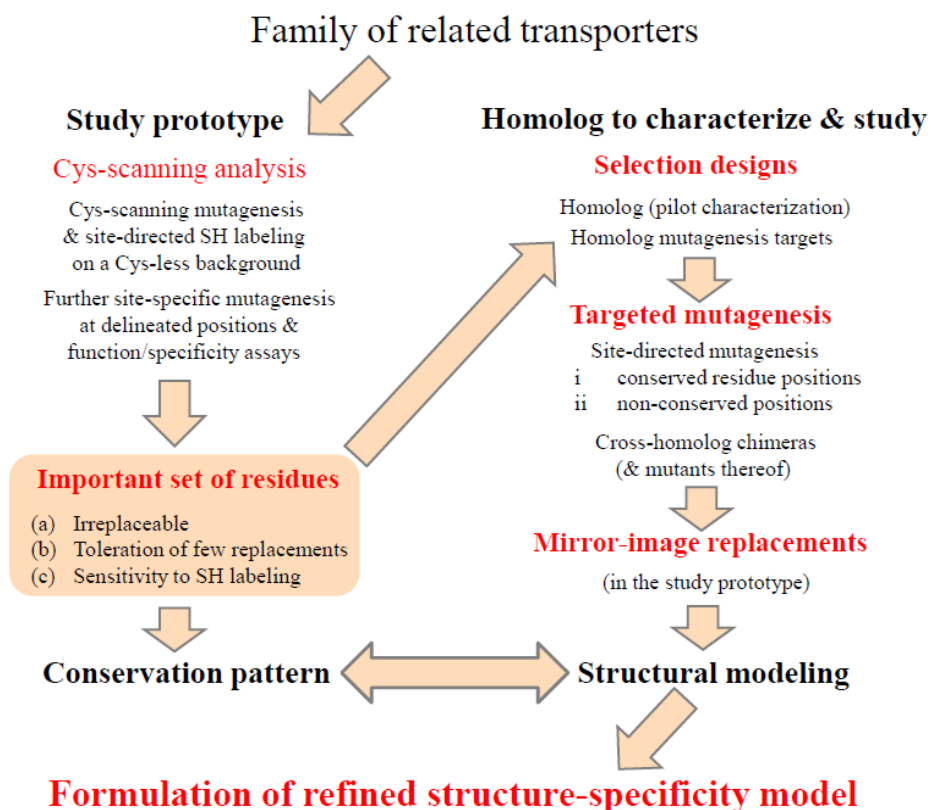

Fig. 1. Flowchart of the Cys-scanning analysis-based strategy for *ab initio* study of new transporter homologs. A detailed account of the individual steps and aims of this strategy is presented in section 2.

2.3 Exploring the use of cross-homolog chimeras and mutants

To dissect the molecular basis of different substrate selectivity trends between closely related transporters, the study of individual site-directed replacements is usually not sufficient; and more combinatorial approaches, involving multiple mutagenesis targets, are needed. The reason is that small contributions from relatively low-effect side chain changes may be crucial for the functional profile outcome, depending on the molecular background used for *in vitro* mutagenesis. In other words, the same replacements may have different effects when combined with different pre-existing mutations on the same transporter background.

Clearly, therefore, combinatorial replacements are important; and both site-specific replacement mutants and cross-homolog chimeras (replacing larger sequence regions and motifs that contain the single-amino acid targets with the homologous ones of the study prototype) can be used in this respect. In particular, the engineering of cross-homolog chimeras between transporters with different substrate profiles often yields variants that are relatively unstable or promiscuous with respect to recognition of substrates. The properties of a chimera may represent an advantageous starting point for exploring a series of mutational events leading to the evolution of new specificities (Tokuriki & Tawfik, 2009). For example, the engineered chimeras might lead to low expression, low activity or promiscuous specificity profiles (Papageorgiou et al., 2008; Papakostas et al., 2008). Chimeras with such properties can be subjected to further site-directed mutagenesis to re-introduce amino acids of the original set of important residues at specified positions. This further mutagenesis might reveal important determinants of uptake and specificity that are not evident in the native transporter backgrounds. The hypothesis is based on two pieces of evidence. (a) Mutations responsible for key functional switches in proteins often yield uninformative phenotypes. This is thought to be due to the effects of other, less important, mutations (a phenomenon known as conformational epistasis) (Harms & Thornton, 2010). (b) More promiscuous and conformationally dynamic proteins exhibit greater evolvability, or potential for divergence of new functions, as supported recently by studies involving molecular evolution, ancient gene resurrection and directed evolution (Tokuriki & Tawfik, 2009). The applicability of such combinatorial and evolution-directed studies is rather limited, at present, and has not been explored extensively in membrane transport proteins. However, this field is rapidly progressing and will probably have a significant impact on the rationale of related site-directed mutagenesis designs in the near future (Morange, 2010).

2.4 Mirror-image replacements in the prototypic transporter

Depending on the results from functional and specificity-profile analysis of the new homolog mutants and/or cross-homolog chimeras, mirror-image replacements can be designed and engineered at the corresponding positions of the study prototype to test whether a particular shift in specificity of the new homolog can be achieved in the inverse direction by replacements of the study prototype at the same residue sites. This line of experimentation is important because it reveals the extent to which a particular side chain influences the substrate recognition profile and may distinguish residues that have a major role on substrate preference from less influential ones. In addition, the comparison between mutants bearing replacements at the same site, but in different native (or chimeric) backgrounds, provides information on the functionally significant interactions of a key specificity mutation with other positions in the different related transporters.

2.5 Formulation of structure-specificity homology models

The sum of data from the comparative analysis of mutants and chimeras between the two related transporters with different specificities (*i.e.*, the new homolog and the initial study prototype) can be used to formulate refined structure-function models highlighting particular aspects of specificity. In addition, the conservation pattern of the residues involved in specificity can be taken into account to draw more generalized structure-specificity conclusions on a group of homologous transporters of the family. This process is facilitated most appropriately when an X-ray structure is available for at least one structural homolog of the transporter family under study, as is the case with the two research paradigms described in the following section.

3. Applications of the Cys-scanning-based approach

The use of a Cys-scanning analysis-based approach (as outlined in Figure 1) has seen limited application to membrane transport proteins to date. In this section, we present two paradigms of use of such an approach to address research questions on the differential substrate preference between closely related transporter homologs. The first paradigm (3.1, MelY) refers to homologs of the well known and thoroughly-studied lactose permease from *Escherichia coli* (LacY), which is a reference protein for all secondary (ion gradient-driven) active transporters. The lessons derived from the detailed site-directed analysis of LacY, which refers to frontline research contributions over almost three decades, are essential for understanding any transporter of this class (Guan & Kaback, 2006; Jiang et al., 2011). The X-ray structures solved for LacY and related homologs are also seminal, as the LacY structural fold typifies the Major Facilitator Superfamily (MFS), encompassing one fourth of all transporters, as well as other more distantly related families (Kaback et al., 2011). The second paradigm (3.2, UacT) refers to the evolutionarily ubiquitous family of nucleobase transporters NAT/NCS2, which has been studied with respect to structure-function relationships only in two members, UapA (Amillis et al., 2011) and XanQ (Karena & Frillingos, 2011). This family of transporters is modeled on a newly described, rather unusual, structural fold (Lu et al., 2011), which has important implications for the binding-site architecture and mechanisms of active transporters. It is also important with respect to potential biomedical applications concerning the development of pathogen-selective cytotoxic nucleobase analogs for targeted antimicrobial therapies (Köse & Schiedel, 2009).

3.1 Substrate selectivity of MelY (in the oligosaccharide-proton symporter family OHS)

The lactose permease from *Escherichia coli* (LacY) is a prototypic example for the study of secondary active transporters. It has been analyzed extensively with cysteine-scanning and site-directed mutagenesis leading to the delineation of residues that are crucial for the mechanism of β-galactoside:H$^+$ symport. X-ray structures of LacY solved in the presence or absence of substrate (Abramson et al., 2003; Chaptal et al., 2011; Guan et al., 2007; Mirza et al., 2006) have confirmed many of the conclusions derived from the biochemical and biophysical studies. The protein is composed of two domains, one N-terminal and one C-terminal, each representing a bundle of six transmembrane alpha-helices and designated N6 and C6, respectively (Abramson et al., 2003). Symmetrical movements between the two domains are associated with alternating opening and closure of the binding site to either

side of the membrane during the mechanism of active transport (the alternating access model) (Kaback et al., 2007). High-resolution structural evidence from three homologous transporters crystallized in the inward-facing conformation and one (the fucose permease FucP) captured in an outward-facing conformation provided additional strong support for the mechanism of alternating access in lactose permease and related transport proteins (Dang et al., 2010).

LacY belongs to the Oligosaccharide:H$^+$ Symporter (OHS) family, a member of the Major Facilitator Superfamily (MFS) (Kasho et al., 2006). The OHS family includes several functionally characterized bacterial proteins, such as the lactose permeases of *Citrobacter freundii* and of *Klebsiella pneumoniae*, the melibiose permease (MelY) of *Enterobacter cloacae*, the raffinose permease (RafB) of *E. coli*, and the sucrose permease (CscB) of *E. coli* (Kasho et al., 2006; Vadyvaloo et al., 2006). Although specificity profiles between members of this family are often closely related, mutagenesis studies to examine the basis of substrate selectivity in members other than LacY are rare. One such example refers to *E. cloacae* MelY, a symporter that had not been analyzed for structure-function relationships prior to application of a Cys-scanning based approach (Tavoulari & Frillingos, 2008).

3.1.1 The research question: subtle selectivity difference between LacY and MelY

MelY (GenBank BAA19154) exhibits 57% identity and 75% similarity in sequence with *E. coli* LacY (UniProtKB P02920). Both proteins transport lactose or melibiose or the monosaccharide galactose (with K_m values ranging from 0.2 mM to 0.6 mM), but MelY is unable to transport the analog methyl-1-thio-*β*,D-galactopyranoside (TMG), that is a very efficient substrate for LacY (K_m 0.54 mM). However, MelY recognizes TMG as a ligand and conserves Cys148 (of helix TM5) in the sugar binding site as a TMG-binding residue (Tavoulari & Frillingos, 2008). Therefore, there is a subtle difference in specificity between the two galactoside transporters. The difference concerns the inability of MelY to catalyze the active transport of TMG, although it can bind this substrate analog with high affinity, comparable to LacY. (The K_i values for competitive inhibition of the lactose uptake by TMG are in the range of 1-2 mM for both transporters.)

Homology alignment and threading of MelY into the known structure of LacY (Protein Data Bank ID: 1PV7) shows that the organization of residues in the putative MelY sugar-binding site is the same as in LacY, and residues irreplaceable for the symport mechanism are conserved. Moreover, MelY differs from LacY in only 15% of the subset of residues at which either a single-Cys mutant is inactivated by site-directed alkylation or few amino acid replacements are tolerated (Figure 2). These observations provide a basis for a systematic site-directed mutagenesis study of MelY aiming at identifying subtle-selectivity determinants within the set of important LacY residues (Frillingos et al., 1998; Kaback et al., 2001), which differ in MelY. Such an approach was taken recently (Tavoulari & Frillingos, 2008) for the dissection of side chain determinants of the substrate profile in MelY relative to the well known LacY.

The difference between the two transporters concerns the uptake of one particular analog (TMG). Since the difference is clearly not in binding per se but in the subsequent translocation reactions, that inability of MelY is probably a substrate-specific impairment of binding from the conformational changes needed to complete turnover. In other words, it

appears that, with TMG as a substrate, the conformational movements of MelY permease are inefficient. Thus, TMG binds but is not transported by wild-type MelY to any significant extent. To elucidate the structure-functional basis of this property in detail, both individual and combinatorial replacements were employed, involving rationally designed mutagenesis and analysis of several cross-ortholog chimeras, as described in the next sections.

3.1.2 Selection of mutagenesis targets

The initial targets for site-specific mutagenesis included residues of the important set of LacY (Figure 2) that are conserved or not conserved in MelY. Mutagenesis of MelY at conserved positions showed that irreplaceable residues of LacY are also irreplaceable in MelY (Asp-131, Arg-149, Glu-274, Arg-307, His-327, Glu-330), while Cys-153 (corresponding to Cys-148, which has been extensively utilized as a binding-site reporter in LacY) displays similar properties to those of Cys-148. Most notable is that Cys-153 of MelY, like Cys-148 of LacY, is highly sensitive to alkylation by *N*-ethylmaleimide (NEM) leading to inactivation of the single-Cys-153 mutant. Like in LacY, the presence of substrate fully reverses this inactivation. These initial observations established that the mechanism of galactoside transport is very similar between MelY and LacY and, in particular, that key residues of the binding site (like Cys-148/Cys-153) have the same functional role in both transporters. In addition, it was established that, although non-transportable, the analog TMG is specifically recognized as a ligand and is bound by MelY and that this binding involves the same functionally conserved residues as in LacY (Tavoulari & Frillingos, 2008).

In a second round of mutagenesis, non-conserved residues of the important set of LacY were replaced with the corresponding amino acid found in MelY or vice versa, aiming at modulating specificity to the counter homolog direction. These non-conserved positions are Leu-65/Val-70, Gly-96/Ala-101, Ala-122/Ser-127, Val-264/Als-269, Ala-279/Ser-284, Cys-355/Gln-360 and Val-367/Ala-372 (Figure 2). In the progress of studies, which involved functional analysis of mutants with respect to their efficiency for active transport of lactose, melibiose and TMG (Tavoulari & Frillingos, 2008), both combinatorial mutagenesis and chimera engineering were employed in an attempt to fully convert the one selectivity type (TMG-permissive or TMG-abortive) to the other (TMG-abortive or TMG-permissive, respectively) by replacing multiple selectivity-related targets simultaneously or larger sequence regions or domains of the transporters.

3.1.3 Cross-homolog chimeras

An interesting feature of transporters of the MFS superfamily, which highlights the conformational autonomy of the two domains (Figure 3), is that *in vivo* expression of the gene in two segments (after splitting the sequence at the central cytoplasmic loop between N6 domain and C6 domain) leads to functional complementation (Bibi & Kaback, 1990). Such functional complementation has also been observed with LacY splits at other loop sites, as depicted in Figure 2 (Kaback & Wu, 1997; Kaback et al., 2001). It was then not surprising that many of cross-homolog chimeras engineered between LacY and MelY at sites corresponding to active split junctions were also active with respect to galactoside transport (Figure 2; S. Frillingos, unpublished information). These observations emphasize the conformational flexibility allowed between the two domain repeats for establishing the dynamic binding site in LacY-type transporters. These observations are consistent with

structural and modeling information as well (Radestock & Forrest, 2011). In the course of the Cys-scanning analysis-based approach for studying the selectivity profile of MelY (Tavoulari & Frillingos, 2008), active MelY/LacY chimeras were used as a background for mutagenesis and proved crucial for implementing selectivity switches from the one profile to the other, as explained in the next section.
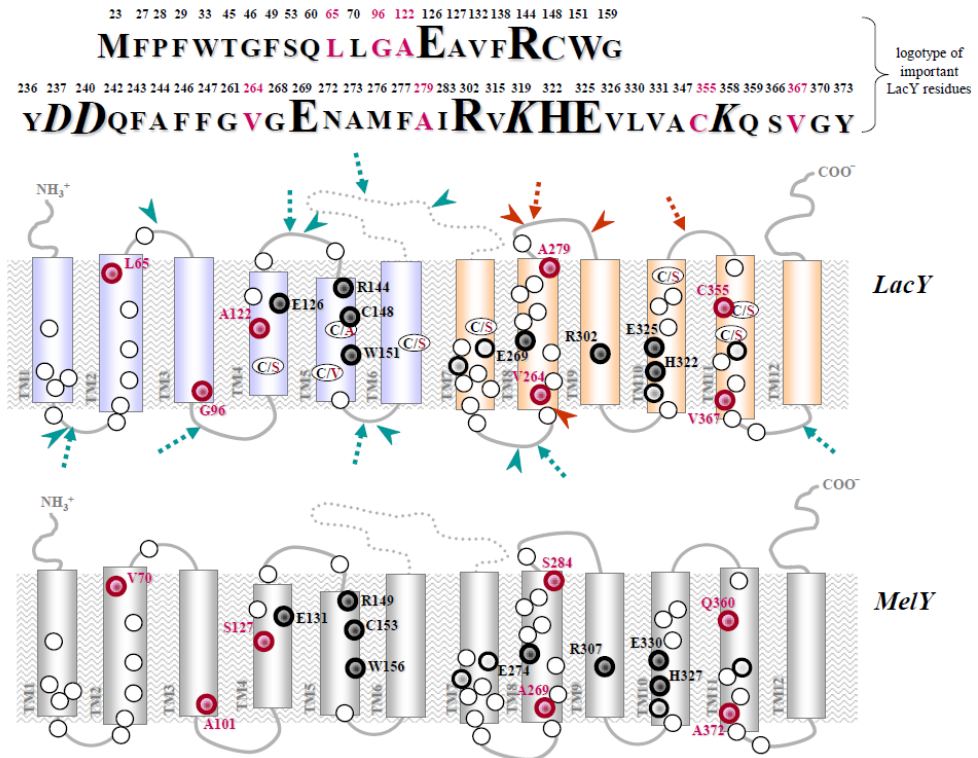
Fig. 2. **Topology models of LacY and MelY and the important set of residues in LacY**. A logotype of the important residues deduced from the Cys-scanning analysis of LacY is shown on top, with larger-size letters indicating functionally irreplaceable residues, medium size indicating residues that are replaceable with few alternative side chains and involved in binding, smaller size indicating residues where a Cys replacement is sensitive to inactivation by *N*-ethylmaleimide and *italics* denoting two pairs of Asp-Lys which are irreplaceable with respect to the charge-pair balance and/or orientation in the membrane (Abramson et al., 2003; Frillingos et al., 1998). Topology models are derived from the X-ray structure of LacY (PDB 1PV7) in combination with prediction algorithms and experimental data on the accessibility of loops to reagents or sequence insertions/deletions (Kaback et al., 2001) and homology threading of MelY (Tavoulari & Frillingos, 2008). The α-helical segments are indicated in rectangles (*blue* and *orange* in LacY denote helices of the N6 and the C6 domain, respectively) and the large intracellular loop between N6 and C6 is shown with a dashed line. *Arrowheads* and *broken arrows* in the LacY model denote positions of splits (the permease gene is split in two coding sequences which are expressed separately in

the same cell to test for functional complementation) or junctions of LacY/MelY chimeras, respectively; splits or chimeras are shown in *teal* (active constructs) and in *red* (inactive). The positions of the important LacY residues are shown in both models with *circles* and mutagenesis targets are *bolded* and shown in *black* (conserved residues) or in *red* (residues that differ in MelY). The eight native Cys residues of LacY and the amino acid replacing each Cys in the Cys-less permease version are shown in ellipses.

3.1.4 Site-directed mutagenesis of LacY and mirror-image replacements in MelY

The practical aim of mutagenesis studies targeted at positions of the non-conserved set of important LacY residues (Figure 2) was to establish side-chain requirements for converting the one transporter profile to the other, with focus on the criterion of whether a transporter variant can take up TMG. The switch between the two selectivity profiles was accomplished, to a major extent, by using combinations of site-specific replacements with cross-homolog chimeras interchanging the N6 and C6 domains of the two transporters (Tavoulari & Frillingos, 2008).

More analytically, the experimental strategy was applied as follows:

1. Switch from LacY to the transport profile of MelY
    *   Engineering and analysis of site-specific replacements of LacY residues:

        Site-directed mutagenesis was performed to replace each one of the important LacY residues that are not conserved in MelY (Figure 2) with the corresponding amino acid found in MelY and assay the mutants for active transport of lactose, melibiose and TMG. One of these mutants (V367A), as well as a double-replacement combining this mutation with another important-site mutation in TM11 (V367A/C355Q), showed negligible TMG uptake and a transport profile that resembles the one of MelY. The remaining six mutants (L65V, G96A, A122S, V264A, A279S and C355Q) transport lactose, melibiose or TMG at high rates and show no deviation from the LacY profile (Tavoulari & Frillingos, 2008).

    *   Analysis of chimeras that exchange domains N6 and C6 between LacY and MelY:

        The alternating access mechanism in LacY entails dynamic movements of domains N6 and C6 relative to each other to implement the conformational changes of turnover. N6-C6 chimeras, which exchange these two domains between LacY and its closely related homolog MelY, are highly active, implying a considerable flexibility in promoting such conformational changes between the two domains. Detailed functional analysis, however, revealed that both N6-C6 chimeras (with N6 from LacY and C6 from MelY or vice versa) were incapable of transporting TMG although they recognized TMG as a lactose-competitive ligand; on the other hand, the two chimeras were equally efficient to transport lactose or melibiose as LacY and MelY (Tavoulari & Frillingos, 2008). Thus, the interchange of the two domains in these chimeras represents a mutagenesis strategy different from the single-replacement mutations in helix TM11 to convert the LacY selectivity profile to that of MelY.

2. Switch from the transport profile of MelY to the profile of LacY
    *   Mirror-image replacements of residues in MelY:

Focusing on position Ala-372/Val-367 (see 1a), site-directed replacements involving mutation A372V in combination or not with Q360C (TM11) were made in the background of MelY. These mirror-image replacements failed to restore significant TMG uptake and showed no deviation from the transport profile of MelY, implying that a single change at position Ala-372/Val-367 is insufficient to yield the TMG-permissive phenotype.

- Mutagenesis in chimeric N6-C6 backgrounds:

   Combination of the mutations V367A (1a) or A372V (2a) with the corresponding N6-C6 background was examined to see whether the TMG-uptake activity can be restored by manipulating flexibility between the two domains. Strikingly, the N6(LacY)-C6(MelY/A372V) mutant showed high affinity and capacity for TMG uptake and a selectivity profile that is indistinguishable from the profile of LacY (Tavoulari & Frillingos, 2008), implying that the TMG transport cycle is restored by the interaction of Val-372(367) with residues of the N6 domain (Figure 3). The reverse mutant, N6(MelY)-C6(LacY/V369A), was indistinguishable in substrate selectivity from MelY and the parental N6-C6 chimeras.

3.1.5 The refined structure-function-selectivity model

An obvious conclusion from the above results would be that efficient transport of TMG requires fine-tuned coordination between the N6 domain and TM11 in C6 domain; this interaction is probably mediated through interactions of an alkylation-sensitive and solvent-accessible face of TM11 (Jiang et al., 2011) with residues of the N6 domain and Val-367 at the periplasmic end of TM11 might have an important contribution in this respect (Tavoulari & Frillingos, 2008).

Homology modeling of MelY in comparison with LacY (Figure 3) shows that Val-367 (TM11) is close to Ala-50 (TM2) in LacY and forms a hydrophobic network that might contribute to the functional inward-facing conformation along with other side chains from TM1, TM2 and TM5 of the N6 domain (Figure 3F). Such a network might be more crucial for TMG than for lactose or melibiose; the non-galactosyl moiety of TMG (which is small, hydrophobic and aglycon) is oriented differently in the binding pocket of LacY than are the non-galactosyl moieties of the disaccharides lactose or melibiose. The non-galactosyl moiety of TMG may promote a slightly different inward-facing conformation, in which hydrophobic interactions might play a pivotal role. On the other hand, orientation of the galactosyl moiety, which determines specificity and is bound by highly conserved and irreplaceable residues (Met-23, Glu-126, Arg-144, Cys-148, Trp-151, Glu-269), is the same for all substrates (Abramson et al., 2003). In MelY all galactosyl-binding residues are conserved; but the putative hydrophobic network at the periplasmic side is disrupted, with less bulky and/or less hydrophobic residues (V367A, A50S, I48V, F30L, A25T, V158G, I157T) (Figure 3E). This difference may account for the fact that MelY binds TMG, but fails to couple this binding to any significant transport. Interactions of the methyl group of TMG in the binding pocket of MelY might promote a tight closure of helices to the periplasmic side incompatible with active transport.

Formation of an efficient hydrophobic network will depend on the side chain contributions from the N6 half. Thus, despite the presence of a Val at position 367, packing of the helices

at the periplasmic side of mutant MelY(A367V) or chimera N6(MelY)-C6(LacY), which contribute less bulky and less hydrophobic side chains from the N6 domain, might still be refractory for an efficient coupling with TMG transport. On the other hand, optimal contacts between helices allow proper formation of the inward-facing conformation and progress of turnover for the TMG. Uptake appears to be restored when Ala-367 is replaced with Val in the N6(LacY)-C6(MelY) chimera, which reintroduces the side chains of LacY in TM1, TM2 and TM5 of the N6 domain. This might be due to reestablishment of hydrophobic interaction of Val-367 with residue(s) of the N6 domain at the periplasmic side and reconstitution of an efficient hydrophobic network.
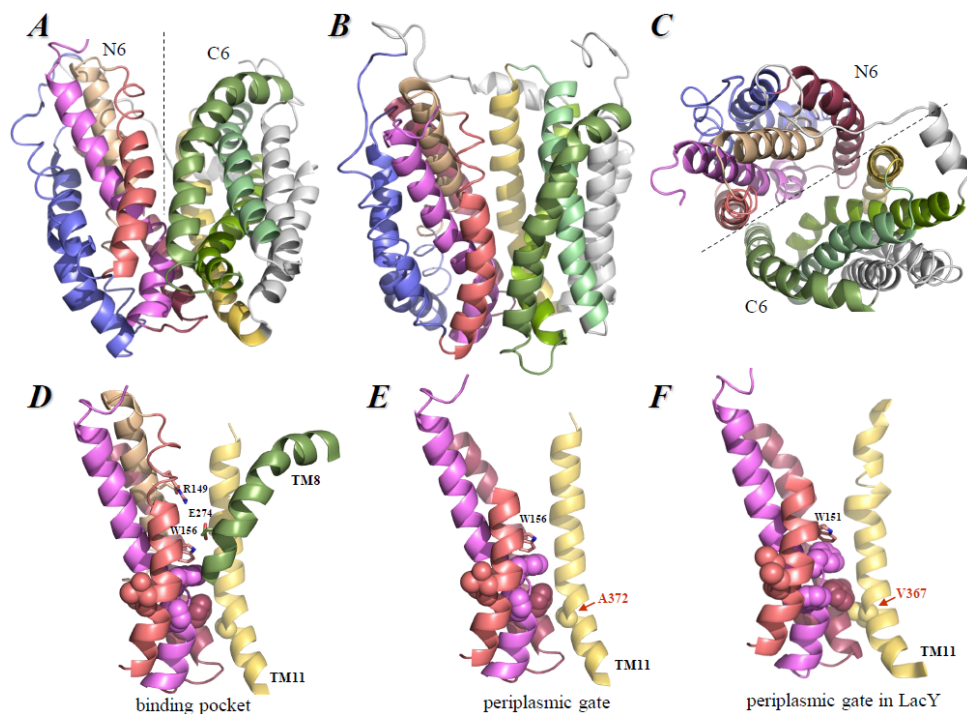
Fig. 3. **Structural models of MelY and comparison with the prototypic homolog (LacY)**. The sequence of MelY was threaded on the known X-ray structure of LacY (PDB 1PV7) (Abramson et al., 2003) using the SWISSPROT modeling server. The structural models were displayed with PyMOL v1.4. The overall helix packing model is shown in three different views (**A-C**). Views **A** (from the side of the membrane) and **C** (from the periplasm) highlight the axis of pseudosymmetry (broken line), which defines the two domains (N6 and C6). Domain N6 contains the bundle of transmembrane helices TM1 (*violet*), TM2 (*raspberry*), TM4 (*wheat*), TM5 (*salmon*) and two peripheral ones, TM3 and TM6 (*blue*). Domain C6 contains the bundle of helices TM7 (*split pea green*), TM8 (*smudge green*), TM10 (*pale green*), TM11 (*yellow orange*) and the peripheral TM9 and TM12 (*grey*). View **B** highlights the central position of TM11 at the interface between the N6 and C6 bundles. The arrangement of TM11 with respect to TM8 and to the N6 bundle of TM1, TM2, TM4 and TM5 is shown more clearly in **D-F**. Panel **D** highlights key galactoside-binding residues, which are invariant

between LacY and MelY. Panel *E* (and *F*, showing LacY) highlights the cluster of residues (at TM11, TM5, TM2 and TM1), which gate the periplasmic substrate pathway and differ between MelY and LacY (see text). The residue implicated in the change of specificity (Ala-372/Val-367) is indicated with an *arrow* and *red label*.

3.2 Ab initio analysis of UacT (in the nucleobase-cation symporter-2 family NAT/NCS2)

The Nucleobase-Ascorbate Transporter (NAT) or Nucleobase-Cation Symporter-2 (NCS2) family is evolutionarily ubiquitous and includes more than 2,000 putative members in all major taxa of organisms. Despite their relevance to the recognition and uptake of several frontline purine-related drugs, only 16 members have been characterized experimentally to date. These are specific for the cellular uptake of uracil, xanthine or uric acid (microbial, plant and non-primate mammalian genomes) or vitamin C (mammalian genomes) (Gournas et al., 2008; Yamamoto et al., 2010).

The NAT/NCS2 family is of particular interest in two respects. First, in an evolutionary perspective, it encompasses transporters with largely different substrate profiles that model on a novel, unprecedented structural fold, as revealed recently (Lu et al., 2011). Second, in a biomedical perspective, it offers important possibilities for translation of the structure-function knowledge to the rational design of targeted antimicrobial drugs, based on the fact that the human homologs do not recognize nucleobases or related cytotoxic compounds (Yamamoto et al., 2010). An additional research challenge is that only 15 of the ~2000 predicted members have been identified functionally and only two members have been studied rigorously with respect to analysis of structure-function relationships, namely the xanthine permease XanQ of *E. coli* (Georgopoulou et al., 2010; Karena & Frillingos, 2011) and the uric acid/xanthine permease UapA of *A. nidulans* (Amillis et al., 2011; Papageorgiou et al., 2008). It is notable that mutagenesis data from both lines of study have revealed striking similarities between the two transporters, reinforcing the idea that few residues conserved throughout the family may be invariably critical for function and underlie specificity differences. Most of these residues are also highlighted as active-site relevant in models built on the recently released X-ray structure of the uracil permease homolog UraA (Protein Data Bank ID: 3QE7) (Lu et al., 2011).

3.2.1 The research question: distinction between xanthine and uric acid (8-oxy-xanthine)

Most of the experimentally known members of the NAT/NCS2 family have been characterized as purine nucleobase transporters, which are specific for the proton gradient-driven uptake of xanthine, uric acid (8-oxy-xanthine) or both. This group of related transporters include 11 bacterial, fungal or plant homologs, namely the xanthine transporters XanQ (UniProtKB accession number P67444) and XanP (P0AGM9) from *Escherichia coli* and PbuX (P42086) from *Bacillus subtilis*, the uric-acid transporters UacT (or YgfU) (Q46821) from *E. coli* and PucK (O32140) and PucJ (O32139) from *B. subtilis*, and the dual-selectivity uric-acid/xanthine transporters UapA (Q07307) and UapC (P487777) from the filamentous fungus *Aspergillus nidulans*, AfUapA (XP748919) from its pathogenic relative *A. fumigatus*, Xut1 (AAX2221) from the yeast *Candida albicans*, and Lpe1 (AAB17501) from maize (*Zea mays*) (see Karena & Frillingos, 2011). Based on the spectrum of their known

specificities, a major research challenge is to understand the mechanism of differential recognition between xanthine and uric acid (8-oxy-xanthine) and between different binding-site preferences for xanthine analogs with variations at the imidazole moiety (8-methylxanthine, 8-azaxanthine, oxypurinol) (Goudela et al., 2005; Karatza & Frillingos, 2005).

Inspection of conserved sequence motifs and sequence alignment analysis of the different xanthine and/or uric acid-transporting homologs indicated interesting patterns of correlation with changes between xanthine-selective and xanthine/uric acid dual-selectivity NAT transporters, especially at a characteristic sequence region of transmembrane segment TM10 known as the NAT-signature motif (Georgopoulou et al., 2010). However, such sequence differences did not correlate with clear-cut changes in substrate selectivity of corresponding mutants that were made in either XanQ (Georgopoulou et al., 2010; Karatza et al., 2006) or UapA (Papageorgiou et al., 2008; Koukaki et al., 2005). In particular, the most pronounced change in XanQ was accomplished with replacement of Gly-333 to Arg (at the carboxyl-terminal end of the motif sequence) yielding aberrant recognition of 8-methylxanthine (which is not a wild-type ligand), but without affecting the selectivity preference for xanthine (Georgopoulou et al., 2010). Recognition of 8-methylxanthine has also been observed with a number of other single-replacement XanQ mutants and even with UapA/XanQ chimeras (Papakostas et al., 2008), implying that several changes at different sites in this xanthine-specific transporter can confer a degree of promiscuity for the recognition of analogs at the imidazole moiety of xanthine. However, since none of these changes resulted in a clear selectivity change (most notably, none allowed recognition or uptake of uric acid), it is evident that the strict preference of XanQ for xanthine is not easily modifiable and a more systematic approach is needed to address the basis of xanthine/8-oxy-xanthine selectivity differences. Such an approach is offered through the exploitation of evidence from a systematic Cys-scanning analysis of XanQ and the elucidation of the function of a new, uric-acid selective homolog (UacT), as described in the next section.

3.2.2 Selection of the homolog to study and the mutagenesis targets

The xanthine-specific permease XanQ has been subjected to a systematic Cys-scanning and site-directed mutagenesis study to address the role of each amino acid residue (Georgopoulou et al., 2010; Karatza et al., 2006; Karena & Frillingos, 2009, 2011; Mermelekas et al., 2010; Papakostas et al., 2008). Of more than 180 residues analyzed to date, a small set emerges as crucial for the mechanism at positions at which a native residue is functionally irreplaceable, replaceable with a limited number of side chains or sensitive to alkylation of a substituted Cys with N-ethylmaleimide leading to inactivation (Figure 4). Homology modeling showed that these functionally important residues could be implicated in substrate binding (Glu-272, Gln-324, Asp-276, Ala-323) or involved in crucial hydrogen bonding (Asn-325, His-31) or disposed to the cytoplasmic halves of TM10 and TM8, which contain key binding residues (Karena & Frillingos, 2011). Site-directed alkylation analysis of XanQ has suggested that Gln-324 and Asn-325 may participate directly in the XanQ binding site (Georgopoulou et al., 2010), while His-31 and Asn-93 are essential for the proper binding affinity and selectivity, as evidenced from ligand inhibition assays (Karena & Frillingos, 2009). In the light of the homologous UraA structure (Lu et al., 2011), it appears that the effect of His-31 (TM1) might be indirect through its interaction with Asn-325

(TM10), while Asn-93 (TM3) is at the binding pocket and may be involved in more direct interactions with substrate or substrate-binding residues (Karena & Frillingos, 2011).

The information derived from the Cys-scanning analysis of XanQ provides a basis to study structure-function relationships in other related members of the NAT/NCS2 family, which are not yet characterized or are poorly studied. In this respect, of particular interest are new homologs with distinct selectivity profiles relative to XanQ. One such homolog is UacT (more commonly known as YgfU), a low-affinity uric acid transporter from *E. coli* characterized recently (Papakostas & Frillingos, 2012). UacT is a proton-gradient-dependent, low-affinity (K_m 0.5 mM) and high-capacity transporter for uric acid that also transports xanthine, but with disproportionately low capacity. Although UacT shares low sequence homology with XanQ (28% identity of residues), it retains most of the residues of the important set identified in XanQ with Cys-scanning analysis (Figure 4). It thus offers a good substrate to apply the Cys-scanning-based approach for elucidation of changes involved in the switch of substrate preference from xanthine (XanQ) to uric acid (UacT).

To delineate targets of mutagenesis in UacT, we have taken into account residues of the important set of XanQ that are conserved or not conserved in the different-selectivity homolog, as depicted in Figure 4. The initial round of mutagenesis in UacT included (i) conservative replacements of residues that are invariant and functionally irreplaceable in XanQ (Glu-270, Asp-298, Gln-318, Asn-319); (ii) rationally designed replacements of side chains that correspond to affinity- or specificity-related residues in XanQ (His-37, Thr-100); (iii) replacements of non-conserved residues of the important set with the corresponding amino acid found in XanQ (T259V, M274D, L278T, V282S, S317A, V320N, R327G, S426N).

3.2.3 Cross-homolog chimeras and mutants

In their majority, the set of important residues identified for XanQ cluster at contiguous regions of transmembrane segments TM8 and TM10, as well as at specific sites in TM1, TM3 and TM14 (Figure 4). In a previous attempt to replace extended sequence portions containing multiple important residues of XanQ with the corresponding regions of the dual-selectivity UapA transporter from *A. nidulans* and search for deviations in substrate preference, it was striking that most of the engineered chimeric constructs were unstable and failed to express in the membrane (see Figure 4). Only one chimera of this set was expressible, but without displaying any transport activity. It was the one that replaced TM14 with the corresponding segment of UapA in the background of XanQ (Papakostas et al., 2008). Interestingly, this chimera could be rescued for active xanthine uptake with reintroduction of two residues from the important set (Asn-430, Ile-432) in the UapA-derived graft (Papakostas et al., 2008). In addition, further combinatorial replacements in this region progressively lead to restoration of full activity and the wild-type profile for xanthine selectivity and ligand specificity (Georgopoulou, K., Botou, M. & Frillingos, S., in preparation).

The difficulty in obtaining structurally stable and active chimeric constructs between the two different NAT transporters (XanQ, UapA) cannot be accounted for by the heterologous origin of the fungal UapA sequence. The engineered chimeras are transferred, induced for expression and tested in an *E. coli* K-12 host, yet similar difficulty is observed with cross-
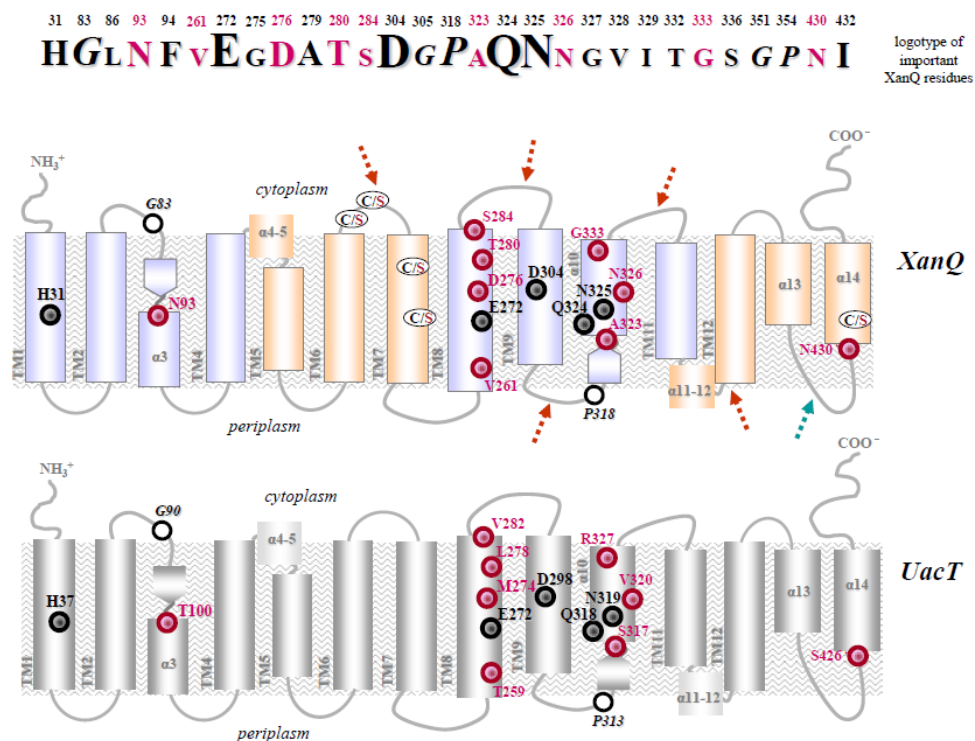
Fig. 4. **Topology models of XanQ and UacT and the important set of residues in XanQ**. A logotype of the important residues deduced from Cys-scanning analysis of XanQ is shown on top, with larger-size letters indicating irreplaceable residues, medium size indicating residues that are replaceable with few alternative side chains, smaller size indicating residues at which a Cys replacement is sensitive to inactivation by N-ethylmaleimide (IC$_{50}$ < 0.1 mM) and *italics* denoting residues that are crucial (smaller size) or irreplaceable (larger size) for expression in the membrane (Karena & Frillingos, 2009, 2011; Georgopoulou et al., 2010; Mermelekas et al., 2010). Topology models are derived from the X-ray structure of UraA (PDB 3QE7) in combination with prediction algorithms and experimental data on the accessibility of loops to hydrophilic reagents (Georgopoulou et al., 2010) and homology threading of XanQ and UacT (Karena & Frillingos, 2011; Lu et al., 2011). The α-helical segments are indicated in rectangles (*blue* and *orange* in XanQ denote helices of the core and the gate domain, respectively). *Broken arrows* in the XanQ model denote junctions of XanQ/UapA chimeras (Papakostas et al., 2008). [UapA is a fungal homolog with dual selectivity, *i.e.*, for both uric acid and xanthine.] The activities of the chimeras are denoted with *teal* (expressed in the membrane and activated upon reintroduction of particular residues; see text) and *red* (not expressed in the membrane). The positions of the important XanQ residues are shown in both models, with *circles* and mutagenesis targets *bolded* and shown in *black* (conserved residues) or in *red* (residues that differ in UacT). The five native Cys residues of XanQ and the amino acid replacing each Cys in the Cys-less permease version are shown in ellipses.

homolog chimeras between XanQ and its *E. coli* paralog UacT, although the phylogenetic distance between XanQ and UacT (28% sequence identity) is equivalent to the one between XanQ and UapA (30% sequence identity) (Georgopoulou, K. & Frillingos, S., in preparation). A more plausible interpretation stems from the intertwined-domain organization of the NAT transporters that was revealed recently from the crystal structure of the UraA homolog (Lu et al., 2011). These transporters are organized in a core and a gate domain, which are composed of two separate contiguous regions each (Figure 4). Although discontinuous in sequence, each domain represents a pair of internal repeats and forms a distinct fold in the structure (Figure 5). The core domain is thought to be pivotal for substrate binding and proton symport, and the gate domain is thought to be crucial for the conformational changes that allow access and release of substrate from the binding site (Lu et al., 2011). However, the relative arrangement of the helices of each domain, which interlace between the repeats to form the dynamic binding site, is highly sensitive to deregulation by changes at key sites. Deregulations leading to instability and loss of the protein expression can be introduced by discontinuities between TM8, TM9, TM10 and TM11 in the chimeric constructs or even by single amino acid changes at the beginning of the crucial TM10 (Pro-318) (Karatza et al., 2006) or TM3 (Gly-83) (Karena & Frillingos, 2011). Thus, sequence rearrangements within the gate domain, which is intimately associated with the binding site architecture, can be grossly deregulating or detrimental for the structural fold. The situation is different with the chimeras involving homologs of LacY (section 3.1) because each domain in the LacY fold is contiguous in sequence and the dynamic binding site is formed at the interface between the two bundles of helices, allowing more flexibility (Kaback et al., 2001).

3.2.4 Site-directed mutagenesis of UacT and mirror-image replacements in XanQ

The most significant conclusions from the analysis of individual site-directed replacements of UacT at the positions of putatively important residues (Figure 4) are that (a) functionally irreplaceable residues of XanQ (such as the substrate binding-relevant Glu-272 and Gln-324) are also irreplaceable in UacT, highlighting the functional conservation of the purine binding site in different-selectivity homologs, and (b) replacements lowering the bulk and polarity of the side chain at one position (Thr-100; TM3) allow conversion of the uric acid-selective UacT to a dual-selectivity variant (mutant with Ala in lieu of Thr-100) that transports both uric acid and xanthine (Papakostas & Frillingos, 2012). Thus, the side chain of Thr-100 at the middle of TM3 is associated directly with defining the purine substrate selectivity with respect to position 8 of the imidazole moiety. This conclusion is strengthened by mirror-image replacements made in XanQ, including an extensive site-directed mutagenesis at the corresponding amino acid found in TM3 (Asn-93) (Karena & Frillingos, 2009, 2011). Mutagenesis at Asn-93 revealed replacements that allow conversion of the xanthine-selective XanQ to dual-selectivity variants (mutants with Ala or Ser in lieu of Asn-93), even though these variants transported the non-wild-type substrate (uric acid) with very low capacity (Karena & Frillingos, 2011). The above considerations are reinforced by the fact that no other single-replacement mutants of either XanQ or UacT has been shown to convert the native transporter to a dual-selectivity one (Karena & Frillingos, 2011). In further support, a similar specificity-related effect is observed with mutants replacing the corresponding TM3 residue (Ser-154) in the fungal homolog UapA, in which introduction of an Ala in lieu of Ser-154 leads to higher affinity for xanthine relative to uric acid, thus shifting the dual-selectivity profile to the xanthine-selective direction (Amillis et al., 2011).

In summary, a major conclusion is that a polar side chain at positions Thr-100/Asn-93/Ser-154 at the middle of TM3 (Figure 4) is associated with the xanthine/uric acid selectivity. However, the selectivity changes observed with the relevant mutants are not sufficiently dramatic to emulate the properties of the other, different-selectivity homologs (Karena & Frillingos, 2011). In this respect, combinatorial replacements are needed to lead to more clear-cut shifts, involving, for example, other sites at which or sequence regions in which mutations have been shown to modify the specificity profile with respect to the imidazole moiety of the substrate (for example, recognition of 8-methylxanthine by XanQ mutants) to a lesser extent (Georgopoulou et al., 2010; Karatza et al., 2006; Karena & Frillingos, 2009).

3.2.5 The refined structure-function-selectivity models

The apparently unique selectivity-related role of Thr-100/Asn-93 in UacT (and XanQ) can be explained by taking into account the distinct conservation pattern of this residue in NAT transporters and homology modeling on the template of the UraA structure (Karena & Frillingos, 2011). In general, Asn-93 is poorly conserved as an amidic side chain even in close XanQ relatives (Karena & Frillingos, 2011); and the same is true of Thr-100 with respect to UacT relatives. However, the polar character of Thr-100/Asn-93 is conserved invariably in the known nucleobase-transporting NAT members (Asn, Thr or Ser), while the ascorbate-transporting SVCTs have an Ala at this position. Furthermore, all the dual-selectivity uric-acid/xanthine transporters (Xut1, UapA, UapC, AfUapA, Lpe1) have a Ser at the corresponding position. To understand the structural relevance of this difference, we have built structural models for the uric acid-selective UacT and the dual-selectivity fungal homologs and compared them with the one of the xanthine-selective XanQ (Figure 5 and data not shown). First of all, the models indicate that this position of TM3 is vicinal to the presumed substrate binding site formed between residues of the middle parts of TM3, TM8 and TM10 in NAT transporters (Figure 5). Strikingly, however, in UacT and all dual-selectivity NATs, the Thr or Ser replacing Asn-93 is distal from the conserved, substrate-relevant glutamate of TM8 (minimal distance between oxygen atoms, 6.0 Å), while Asn-93 in XanQ is significantly closer (distance between oxygen atoms, 4.5 Å). This difference is most prominent in the models of UacT (Figure 5) or UapA (Karena & Frillingos, 2011), which conserve nearly all the other side chains of functionally important residues of TM1, TM3, TM8 or TM10, except Asn-93.

In the dual-selectivity UapA, Ser-154 (corresponding to Asn-93 of TM3) is oriented away from the carboxyl group of Glu-356 (corresponding to Glu-272 of TM8) and leaves more space between TM3 and TM8 in the substrate binding pocket (Amillis et al., 2011; Karena & Frillingos, 2011). Thus, occupation of the Asn-93 position by Ser may relax a constraint for the recognition of analogs modified at position 8 of the imidazole moiety of xanthine and allow binding and transport of uric acid (8-oxy-xanthine), which modifies the NAT selectivity towards a less stringent, dual-substrate profile. Accordingly, the XanQ mutants replacing Asn-93 with Ser (or Ala) yield efficient recognition of 8-methylxanthine and low, but significant, uptake of uric acid, mimicking in part the fungal, dual-selectivity NATs (Karena & Frillingos, 2011).

In the uric acid transporter UacT, Thr-100 (corresponding to Asn-93) is oriented away from the carboxyl group of Glu-270 (corresponding to Glu-272 of TM8) leaving more space in the substrate binding pocket; but, at the same time, the pKa of Glu-270 may be distorted

significantly relative to the corresponding carboxylic acid in XanQ due to its proximity to hydrophobic groups from Thr-100 and Met-274 (Figure 5E). These changes on the substrate binding glutamate Glu-272/Glu-270 (Lu et al., 2011) might account for the selectivity difference between UacT (uric acid) and XanQ (xanthine). Interestingly, however, replacement of Asn-93 with Thr in XanQ cannot imitate the UacT profile, but leads to low affinity for all xanthine analogs, possibly due to interference of the methyl group of Thr-93 in the vicinity of the essential Glu-272 that is not counterbalanced by other permissive mutations (Karena & Frillingos, 2009, 2011). Based on this observation, it is evident that further combinatorial replacements are needed to promptly convert XanQ to the UacT selectivity profile or vice versa; and targets for such replacements have to be selected from the residues of the important set (Figure 4), which correspond to binding site-relevant positions (Karena, E., Papakostas, K. & Frillingos, S., in preparation).
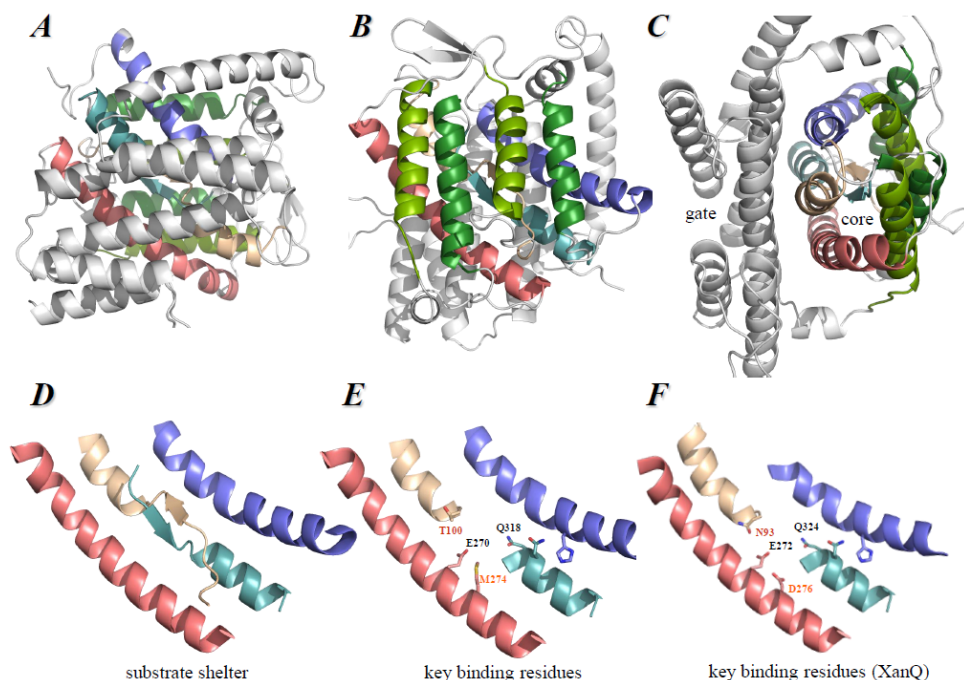
Fig. 5. **Structural models of UacT and comparison with the prototypic homolog (XanQ)**. The sequence of UacT (*A-E*) or XanQ (*F*) was threaded on the known X-ray structure of UraA (PDB 3QE7) (Lu et al., 2011) using the SWISSPROT modeling server, and the structural models were displayed with PyMOL v1.4. The overall helix packing model of UacT is shown in three different views (*A-C*). View *C* (from the periplasm) highlights the two domains (core and gate), which are interplexed and not readily discerned in the side views (*A* and *B*). Transmembrane segments of the core domain (associated with substrate binding) are shown in *blue* (TM1), *wheat* (TM3), *teal* (TM10), *salmon* (TM8), *pea green* (TM2, TM4) and *forest green* (TM9, TM11), while the gate domain (associated with the conformational changes allowing access and release of substrate from the binding site) is shown in *grey*. The arrangement of the four substrate-coordinating segments (TM1, TM3,

TM8 and TM10) is shown more clearly in *D-F*. Panel *D* highlights the central position of the two short antiparallel β-strands of TM3 and TM10, which provide a shelter for the nucleobase substrate (Lu et al., 2011). Panel *E* (and *F* showing XanQ) highlights key binding-site residues, with residues differing between UacT and XanQ indicated with a *red label* (Thr-100/Asn-93, implicated in the purine selectivity preference) and an *orange label* (Met-274/Asp-276). For clarity, only the helical segments of TM3 and TM10 are shown in *E* and *F*.

4. Conclusion and perspectives

The two paradigms described above highlight the applicability of approaches that employ Cys-scanning analysis data from a reference molecule (the study prototype) to guide the *ab initio* analysis of structure-function relationships of new transporters in evolutionarily conserved families of structurally related homologs. In particular, they provide a strategy for effective site-directed mutagenesis designs to dissect the molecular determinants underlying substrate selectivity shifts between closely related homologs. In parallel, they explore the mechanism of change to novel selectivity profiles through a combination of site-specific replacements in native and chimeric transporter backgrounds. Apart from the obvious contributions to the research of groups of transporters with high potential for translation to biomedical and other applications, a more systematic and generalized application of this strategy would certainly have a major methodological impact in the field. It should allow rapid and cost-effective mutagenesis designs on newly identified membrane transport proteins, even in cases in which a high-resolution model is unavailable.

5. Acknowledgments

6. References

Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H.R. & Iwata, S. (2003). Structure and mechanism of the lactose permease of *Escherichia coli*. *Science*, Vol.301, No.5633, (August 2003), pp. 610-615, ISSN 0036-8075

Akabas, M.H., Stauffer, D.A. & Karlin, A. (1992). Acetylcholine receptor channel structure probed in cysteine-substitution mutants. *Science*, Vol.258, No.5080, (October 1992), pp. 307-310, ISSN 0036-8075

Amillis, S., Kosti, V., Pantazopoulou, A., Mikros, E. & Diallinas, G. (2011). Mutational analysis and modeling reveal functionally critical residues in transmembrane segments 1 and 3 of the UapA transporter. *Journal of Molecular Biology*, Vol.411, No.3, (August 2011), pp. 567-580, ISSN 0022-2836

Bibi, E. & Kaback, H.R. (1990). In vivo expression of the *lacY* gene in two segments leads to functional lac permease. *Proceedings of the National Academy of Sciences of the U.S.A.*, Vol.87, No.11, (June 1990), pp. 4325-4329, ISSN 0027-8424

Boudker, O. & Verdon, G. (2010). Structural perspectives on secondary active transporters. *Trends in Pharmacological Sciences*, Vol.31, No.9, (September 2010), pp. 418-426, ISSN 0165-6147

Chaptal, V., Kwon, S., Sawaya, M.R., Guan, L., Kaback, H.R. & Abramson, J. (2011). Crystal structure of lactose permease in complex with an affinity inactivator yields unique insight into sugar recognition. *Proceedings of the National Academy of Sciences of the U.S.A.*, Vol.108, No.23, (June 2011), pp. 9361-9366, ISSN 0027-8424

Chen, J.G. & Rudnick, G. (2000). Permeation and gating residues in serotonin transporter. *Proceedings of the National Academy of Sciences of the U.S.A.*, Vol.97, No.3, (February 2000), pp. 1044-1049, ISSN 0027-8424

Crisman, T.J., Qu, S., Kanner, B.I. & Forrest, L.R. (2009). Inward-facing conformation of glutamate transporters as revealed by their inverted-topology structural repeats. *Proceedings of the National Academy of Sciences of the U.S.A.*, Vol.106, No.49, (December 2009), pp. 20752-20757, ISSN 0027-8424

Culham, D.E., Hillar, A., Henderson, J., Ly, A., Vernikovska, Y.I., Racher, K.I., Boggs, J.M. & Wood, J.M. (2003). Creation of a fully-functional cysteine-less variant of osmosensor and proton-osmoprotectant symporter ProP from *Escherichia coli* and its application to assess the transporter's membrane orientation. *Biochemistry*, Vol.42, No.40, (October 2003), pp. 11815-11823, ISSN 0006-2960

Dang, S., Sun, L., Huang, Y., Lu, F., Liu, Y., Gong, H., Wang, J. & Yan, N. (2010). Structure of a fucose transporter in an outward-open conformation. *Nature* Vol.467, No.7316, (October 2010), pp. 734-738, ISSN 0028-0836

Forrest, L.R., Zhang, Y.W., Jacobs, M.T., Gesmonde, J., Xie, L., Honig, B.H. & Rudnick, G. (2008). Mechanism for alternating access in neurotransmitter transporters. *Proceedings of the National Academy of Sciences of the U.S.A.*, Vol.105, No.30, (July 2008), pp. 10338-10343, ISSN 0027-8424

Forrest, L.R., Krämer, R. & Ziegler, C. (2011). The structural basis of secondary active transport mechanisms. *Biochimica et Biophysica Acta (BBA) – Bioenergetics*, Vol.1807, No.2, (February 2011), pp. 167-188, ISSN 0005-2728

Frillingos, S. & Kaback, H.R. (1996). Probing the conformation of the lactose permease of *Escherichia coli* by in situ site-directed sulfhydryl modification. *Biochemistry*, Vol.35, No.13, (April 1996), pp. 3950-3956, ISSN 0006-2960

Frillingos, S., Sahin-Tóth, M., Wu, J. & Kaback, H.R. (1998). Cys-scanning mutagenesis : a novel approach to a structure-function relationships in polytopic membrane proteins. *FASEB Journal*, Vol.12, No.13, (October 1998), pp. 1281-1299, ISSN 0892-6638

Georgopoulou, E., Mermelekas, G., Karena, E. & Frillingos, S. (2010). Purine substrate recognition by the nucleobase-ascorbate transporter motif in the YgfO xanthine permease: Asn-325 binds and Ala-323 senses substrate. *Journal of Biological Chemistry*, Vol.285, No.25, (June 2010), pp. 19422-19433, ISSN 0021-9258

Goudela, S., Karatza, P., Koukaki, M., Frillingos, S. & Diallinas, G. (2005). Comparative substrate recognition by bacterial and fungal purine transporters of the

NAT/NCS2 family. *Molecular Membrane Biology*, Vol.22, No.3, (May-June 2005), pp. 263-275, ISSN 0968-7688

Gournas, C., Papageorgiou, I. & Diallinas, G. (2008). The nucleobase-ascorbate transporter (NAT) family: genomics, evolution, structure-function relationships and physiological role. *Molecular Biosystems*, Vol.4, No.5, (May 2008), pp. 404-416, ISSN 1742-206X

Guan, L. & Kaback, H.R. (2006). Lessons from lactose permease. *Annual Reviews of Biophysics and Biomolecular Structure*, Vol.35, (June 2006), pp. 67-91, ISSN 1936-122X

Guan, L. & Kaback, H.R. (2007). Site-directed alkylation of cysteine to test solvent accessibility of membrane proteins. *Nature Protocols*, Vol.2, No.8, (August 2007), pp. 2012-2017, ISSN 1754-2189

Guan, L., Mirza, O., Verner, G., Iwata, S. & Kaback, H.R. (2007). Structural determination of wild-type lactose permease. *Proceedings of the National Academy of Sciences of the U.S.A.*, Vol.104, No.39, (September 2007), pp. 15294-15298, ISSN 0027-8424

Harms, M.J. & Thornton, J.W. (2010). Analyzing protein structure and function using ancestral gene resurrection. *Current Opinion in Structural Biology*, Vol.20, No.3, (June 2010), pp. 360-366, ISSN 0959-440X

Jiang, X., Nie, Y. & Kaback, H.R. (2011). Site-directed alkylation studies with LacY provide evidence for the alternating access model of transport. *Biochemistry*, Vol.50, No.10, (March 2011), pp. 1634-1640, ISSN 0006-2960

Jung, H., Rübenhagen, R., Tebbe, S., Leifker, K., Tholema, N., Quick, M. & Schmid R. (1998). Topology of the Na$^+$/proline transporter of *Escherichia coli*. *Journal of Biological Chemistry*, Vol.273, No.41, (October 1998), pp. 26400-26407, ISSN 0021-9258

Jung, K., Jung, H., Wu, J., Privé, G.G. & Kaback, H.R. (1993). Use of site-directed fluorescence labeling to study proximity relationships in the lactose permease of *Escherichia coli*. *Biochemistry*, Vol.32, No.46, (November 1993), pp. 12273-12278, ISSN 0006-2960

Kaback, H.R. & Wu, J. (1997). From membrane to molecule to the third amino acid from the left with a membrane transport protein. *Quarterly Reviews of Biophysics*, Vol.30, No.4, (November 1997), pp. 333-364, ISSN 0033-5835

Kaback, H.R., Sahin-Tóth, M. & Weinglass, A.B. (2001). The kamikaze approach to membrane transport. *Nature Reviews Molecular Cell Biology*, Vol.2, No.8, (August 2001), pp. 610-620, ISSN 1471-0072

Kaback, H.R., Dunten, R., Frillingos, S., Venkatesan, P., Kwaw, I., Zhang, W. & Ermolova, N. (2007). Site-directed alkylation and the alternating access model for LacY. *Proceedings of the National Academy of Sciences of the U.S.A.*, Vol.104, No.2, (January 2007), pp. 491-494, ISSN 0027-8424

Kaback, H.R., Smirnova, I., Kasho, V., Nie, Y. & Zhou, Y. (2011). The alternating access transport mechanism in LacY. *Journal of Membrane Biology*, Vol.239, No.1-2, (January 2011), pp. 85-93, ISSN 0022-2631

Karatza, P. & Frillingos, S. (2005). Cloning and functional characterization of two bacterial members of the NAT/NCS2 family in *Escherichia coli*. *Molecular Membrane Biology*, Vol.22, No.3, (May-June 2005), pp. 251-261, ISSN 0968-7688

Karatza, P., Panos, P., Georgopoulou, E. & Frillingos, S. (2006). Cysteine-scanning analysis of the nucleobase-ascorbate transporter signature motif in YgfO permease of *Escherichia coli*: Gln-324 and Asn-325 are essential and Ile-329–Val-339 form an

alpha-helix. *Journal of Biological Chemistry*, Vol.281, No.52, (December 2006), pp. 39881-39890, ISSN 0021-9258

Karena, E. & Frillingos, S. (2009). Role of intramembrane polar residues in the YgfO xanthine permease: His-31 and Asn-93 are crucial for affinity and specificity, and Asp-304 and Glu-272 are irreplaceable. *Journal of Biological Chemistry*, Vol.284, No.36, (September 2009), pp. 24257-24268, ISSN 0021-9258

Karena, E. & Frillingos, S. (2011). The role of transmembrane segment TM3 in the xanthine permease XanQ of *Escherichia coli*. *Journal of Biological Chemistry*, Vol.286, No.45, (November 2011), pp. 39595-39605, ISSN 0021-9258

Kasho, V.N., Smirnova, I.N. & Kaback, H.R. (2006). Sequence alignment and homology threading reveals prokaryotic and eukaryotic proteins homologous to lactose permease. *Journal of Molecular Biology*, Vol.358, No.4, (May 2006), pp. 1060-1070, ISSN 0022-2836

Köse, M. & Schiedel, A.C. (2009). Nucleoside/nucleobase transporters: Drug targets of the future? *Future Medicinal Chemistry*, Vol.1, No.2, (May 2009), pp. 303-326, ISSN 1756-8919.

Koukaki, M., Vlanti, A., Goudela, S., Pantazopoulou, A., Gioule, H., Tournaviti, S. & Diallinas, G. (2005). The nucleobase-ascorbate transporter (NAT) signature motif in UapA defines the function of the purine translocation pathway. *Journal of Molecular Biology*, Vol.350, No.3, (July 2005), pp. 499-513, ISSN 0022-2836

Lu, F., Li, S., Jiang, Y., Jiang, J., Fan, H., Lu, G., Deng, D., Dang, S., Zhang, X., Wang, J. & Yan, N. (2011). Structure and mechanism of the uracil transporter UraA. *Nature* Vol.472, No.7342, (April 2011), pp. 243-246, ISSN 0028-0836

Mirza, O., Guan, L., Verner, G., Iwata, S. & Kaback, H.R. (2006). Structural determination of wild-type lactose permease. *EMBO Journal*, Vol.25, No.6, (March 2006), pp. 1177-1183, ISSN 0261-4189

Morange, M. (2010). How evolutionary biology presently pervades cell and molecular biology. *Journal for General Philosophy of Science*, Vol.41, No.1, (June 2010), pp. 113-120, ISSN 0925-4560

Papageorgiou, I., Gournas, C., Vlanti, A., Amillis, S., Pantazopoulou, A. & Diallinas, G. (2008). Specific interdomain synergy in the UapA transporter determines its unique specificity for uric acid among NAT carriers. *Journal of Molecular Biology*, vol.382, No.5, (October 2008), pp. 1121-1135, ISSN 0022-2836

Papakostas, K., Georgopoulou, E. & Frillingos, S. (2008). Cysteine-scanning analysis of putative helix XII in the YgfO xanthine permease: Ile-432 and Asn-430 are important. *Journal of Biological Chemistry*, Vol.283, No.20, (May 2008), pp. 13666-13678, ISSN 0021-9258

Papakostas, K., & Frillingos, S. (2012). Substrate selectivity of YgfU, a uric acid transporter from *Escherichia coli*. *Journal of Biological Chemistry,* vol.287, No.19, (May 2012), pp. 15684-15695, ISSN 0021-9258

Radestock, S. & Forrest, L.R. (2011). The alternating-access mechanism of MFS transporters arises from inverted-topology repeats. *Journal of Molecular Biology*, vol.407, No.5, (April 2011), pp. 698-715, ISSN 0022-2836

Mermelekas, G., Georgopoulou, E., Kallis, A., Botou, M., Vlantos, V. & Frillingos, S. (2010). Cysteine-scanning analysis of helices TM8, TM9a, TM9b and intervening loops in the YgfO xanthine permease: a carboxyl group is essential at position Asp-276.

Journal of Biological Chemistry, Vol.285, No.45, (November 2010), pp. 35011-35020, ISSN 0021-9258

Sahin-Tóth, M., Frillingos, S., Lawrence, M.C. & Kaback, H.R. (2000). The sucrose permease of Escherichia coli: functional significance of cysteine residues and properties of a cysteine-less transporter. *Biochemistry*, Vol.39, No.20, (May 2000), pp. 6164-6169, ISSN 0006-2960

Slotboom, D.J., Konings, W.N. & Lolkema, J.S. (2001). Cysteine-scanning mutagenesis reveals a highly amphipathic, pore-lining membrane-spanning helix in the glutamate transporter GltT. *Journal of Biological Chemistry*, Vol.276, No.14, (April 2001), pp. 10775-10781, ISSN 0021-9258

Sorgen, P.L., Hu, Y., Guan, L., Kaback, H.R. & Girvin, M.E. (2002). An approach to membrane protein structure without crystals. *Proceedings of the National Academy of Sciences of the U.S.A.*, Vol.99, No.22, (October 2002), pp. 14037-14040, ISSN 0027-8424

Tamura, N., Konishi, S. & Yamaguchi, A. (2003). Mechanisms of drug/H+ antiport: complete cysteine-scanning mutagenesis and the protein engineering approach. *Current Opinion in Chemical Biology*, Vol.7, No.5, (October 2003), pp. 570-579, ISSN 1367-5931

Tavoulari, S. & Frillingos, S. (2008). Substrate selectivity of the melibiose permease (MelY) from *Enterobacter cloacae*. *Journal of Molecular Biology*, vol.376, No.3, (February 2008), pp. 681-693, ISSN 0022-2836

Tokuriki, N. & Tawfik, D.S. (2009). Protein dynamism and evolvability. *Science*, Vol.324, No.5924, (April 2009), pp. 203-207, ISSN 0036-8075

van Iwaarden, P.R., Pastore, J.C., Konings, W.N. & Kaback, H.R. (1991). Construction of a functional lactose permease devoid of cysteine residues. *Biochemistry*, Vol.30, No.40, (October 1991), pp. 9595-9600, ISSN 0006-2960

Voss, J., Hubbell, W.L. & Kaback, H.R. (1995). Distance determination in proteins using designed metal ion binding sites and site-directed spin labeling: application to the lactose permease of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the U.S.A.*, Vol.92, No.26, (December 1995), pp. 12300-12303, ISSN 0027-8424

Weissborn, A.C., Botfield, M.C., Kuroda, M., Tsuchiya, T. & Wilson, T.H. (1997). The construction of a cysteine-less melibiose carrier from *E. coli*. *Biochimica et Biophysica Acta (BBA) – Biomembranes*, Vol.1329, No.2, (October 1997), pp. 237-244, ISSN 0005-2736

Weyand, S., Shimamura, T., Beckstein, O., Sansom, M.S., Iwata, S., Henderson, P.J. & Cameron, A.D. (2011). The alternating access mechanism of transport as observed in the sodium-hydantoin transporter Mhp1. *Journal of Synchrotron Radiation*, Vol.18, No.1, (January 2011), pp. 20-23, ISSN 0909-0495

Wu, J., Frillingos, S. & Kaback, H.R. (1995). Dynamics of lactose permease of Escherichia coli determined by site-directed chemical labeling and fluorescence spectroscopy. *Biochemistry*, Vol.34, No.26, (July 1995), pp. 8257-8263, ISSN 0006-2960

Wu, J. & Kaback, H.R. (1996). A general method for determining helix packing in membrane proteins in situ: helices I and II are close to helix VII in the lactose permease of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the U.S.A.*, Vol.93, No.25, (December 1996), pp. 14498-14502, ISSN 0027-8424

Vadyvaloo, V., Smirnova, I.N., Kasho, V.N. & Kaback, H.R. (2006). Conservation of residues involved in sugar/H+ symport by the sucrose permease of *Escherichia coli* relative to lactose permease. *Journal of Molecular Biology*, Vol.358, No.4, (May 2006), pp. 1051-1059, ISSN 0022-2836

Yamamoto, S., Inoue, K., Murata, T., Kamigaso, S., Yasujima, T., Maeda, J., Yoshida, Y., Ohta, K. & Yuasa, H. (2010). Identification and functional characterization of the first nucleobase transporter in mammals: implication in the species difference in the intestinal absorption mechanism of nucleobases and their analogs between higher primates and other mammals. *Journal of Biological Chemistry*, Vol.285, No.9, (February 2010), pp. 6522-6531, ISSN 0021-9258

Yan, R.T. & Maloney, P.C. (1993). Identification of a residue in the translocation pathway of a membrane carrier. *Cell*, Vol.75, No.1, (October 1993), pp. 37-44, ISSN 0092-8674

Yousef, M.S. & Guan, L. (2009). A 3D structure model of the melibiose permease of *Escherichia coli* represents a distinctive fold for Na+ symporters. *Proceedings of the National Academy of Sciences of the U.S.A.*, Vol.106, No.36, (September 2009), pp. 15291-15296, ISSN 0027-8424

Zomot, E., Zhou, Y. & Kanner, B.I. (2002). Proximity of transmembrane domains 1 and 3 of the gama-aminobutyric acid transporter GAT-1 inferred from paired cysteine mutagenesis. *Journal of Biological Chemistry*, Vol.280, No.27, (July 2005), pp. 25512-25516, ISSN 0021-9258

Site-Directed Mutagenesis Using Oligonucleotide-Based Recombineering

Roman G. Gerlach, Kathrin Blank and Thorsten Wille
Robert Koch-Institute Wernigerode Branch
Germany

1. Introduction

Methods enabling mutational analysis of distinct chromosomal locations, like site-directed mutagenesis, insertion of foreign sequences or in-frame deletions, have become of fast growing interest since complete bacterial genome sequences became available. Various approaches have been described to modify any nucleotide(s) in almost any manner. Some genetic engineering technologies do not rely on the *in vitro* reactions carried out by restriction enzymes and DNA ligases (Sawitzke et al., 2001). Complicated genetic constructs that seem to be impossible to generate *in vitro* can be created within one week using *in vivo* technologies (Sawitzke et al., 2001).

Over several decades, researchers developed and refined various strategies for genetic engineering that make use of the homologous recombination system. Its natural main functions are restoring collapsed replication forks, repairing damage-induced double-strand breaks and maintaining the integrity of the chromosome (Poteete, 2001).

We want to focus on a technique for recombination-mediated genetic engineering ("recombineering", Copeland et al., 2001). Recombineering requires only minimal *in vitro* effort. It has been applied to *Escherichia coli*, *Salmonella*, and a range of other Gram-negative bacteria, as well as to bacteriophages, cosmids and bacterial artificial chromosomes (BACs). It was demonstrated that single-stranded DNA (ssDNA) oligonucleotides can be used as substrates for recombineering in *E. coli* (Ellis et al., 2001, Heermann et al., 2008) and BACs (Swaminathan et al., 2001). However, most commonly linear, double-stranded DNA (dsDNA) has been used as the targeting construct (Maresca et al., 2010), e.g., for chromosomal gene replacement (Murphy, 1998), whole gene disruption (Datsenko et al., 2000) or the development of novel cloning strategies, including subcloning of BAC DNA (Lee et al., 2001).

In the early 1990s, the DNA double-strand break and repair recombination pathway proved to be very efficient for recombining incoming linear DNA with homologous DNA in the yeast *Saccharomyces cerevisiae* (Baudin et al., 1993). For generating *null* alleles of a distinct gene, a PCR-amplified *HIS3* selectable marker flanked by homologous sequences to the ORF (ranging from 35 to 51 nucleotides in length) was used to transform a recipient strain lacking the complete *HIS3* gene or a strain containing the His3Δ200 allele. Due to the auxotrophic selection marker, transformants bearing the expected mutation were among the His$^+$ clones, with up to 80% efficiency (Baudin et al., 1993).

In contrast to *S. cerevisiae*, *E. coli* fails to be readily transformable with linear DNA fragments because of rapid DNA degradation by the intracellular RecBCD exonuclease activity (Lorenz et al., 1994). Mutants defective in the RecBCD nuclease exhibit no degradation of linear DNA in *E. coli*, but unfortunately these strains are also deficient for any recombination events. This recombination defect can be partially rescued in strains with *recA+* background (Jasin & Schimmel, 1984). Other mutants defective in *recBC* (or either *recB* or *recC*) carrying an additional suppressor mutation, *sbc* (suppressor of *recB* and *recC*), possess activation of the RecET recombination pathway (*sbcA*) or enhanced recombination by the RecF pathway (*sbcB*). *recE* and *recT* are found on the defective lambdoid *E. coli* prophage Rac and encode an exonuclease and a ssDNA-binding/strand-exchange protein, respectively (Fouts et al., 1983, Poteete, 2001). Expression of *recET* is induced by few *cis*-acting point mutations, e. g. *sbcA6* (Clark et al., 1994).

One highly applicable RecET-mediated recombination reaction, termed 'ET-cloning', combines a homologous recombination reaction between linear DNA fragments and circular target molecules, like BAC episomes (Zhang et al., 1998). After co-transformation of linear and circular DNA molecules, only *recBC sbcA* recipient strains resulted in the intended recombination products (insertion or deletion). Recombination was more efficient with increasing length of the homology arms, and some constructs showed higher efficiency with increasing distance between the two homology sites (Zhang et al., 1998). For ET-cloning in *recBC+* strains, which are commonly used as hosts for P1, BAC or PAC episomes, a plasmid encoding the recombination functions was constructed. In pBAD-ETγ, *recE* is under the control of an inducible promoter; and *recT* is expressed from a strong constitutive promoter. To inhibit RecBC-mediated degradation of linear DNA, the λ *gam*, encoding the Redγ protein, was incorporated on the plasmid (Zhang et al., 1998). Later, *E. coli* hosts with chromosomally-encoded, inducible recombinases have been developed to allow easy manipulation of BAC DNA (Lee et al., 2001).

1.1 The bacteriophage λ Red recombination system

Besides the mutagenesis pathway described above, Red recombination is one of the most commonly exploited techniques to foster recombination between the bacterial chromosome and linear dsDNA introduced into the cell (Murphy, 1998). The Red recombination system of the bacteriophage λ leads to a precise and rapid approach with greatly enhanced rates of recombination, compared to those found in *recBC*, *sbcB* or *recD* mutants. Its ability to catalyze the incorporation of PCR-generated DNA species led to an immense spread of the system. Numerous groups have developed various methodologies tailored to a variety of scientific questions. Besides the high recombination rate, the biggest advantage of the λ Red system, compared to previously used recombination systems, is that it accepts very short regions of homologous DNA (stretches of less than 100 nucleotides) for recombination. Because fragments of such size can be readily synthesized, there is a high degree of freedom in designing targeting constructs for recombination.

Which components make up the λ Red system? The genes of the Red system, *exo*, *bet*, and *gam*, cluster together in the P_L operon, which is expressed in the early transcriptional program of bacteriophage λ (Poteete, 2001). The three resulting λ Red gene products, Redα, Redβ and Redγ, are necessary to carry out homologous recombination of dsDNA. Redα, whose monomer has a Mr of 24 kDa, is responsible for a dsDNA-dependent exonuclease

activity. It degrades one strand at the ends of a linear dsDNA molecule in the 5' to 3' direction. This generates 3' ssDNA overhangs, which are substrates for recombination (Little, 1967). The ring-shaped trimer of Redα passes dsDNA through its center at one end, but only ssDNA emerges from the other end (Kovall et al., 1997). The Redβ protein, whose monomer has a Mr of 28 kDa, acts as a ssDNA-binding protein that anneals complementary single strands and mediates strand exchange, thus finishing the recombination process (Li et al., 1998). When bound to ssDNA, Redβ forms large rings and makes a complex with Redα to promote recombination (Passy et al., 1999). Therefore, the Redα/Redβ pair has a function analogous to that of the RecE/RecT pair. Muyrers et al. (2000b) could highlight that there are specific interactions between the partners of each recombination system. The exonuclease RecE does not form complexes with the strand exchange protein Redβ or vice versa; Redα does not interact with RecT. Finally, the Redγ polypeptide, whose Mr is 16 kDa, protects linear dsDNA from degradation by binding the RecBCD protein and inhibiting its nuclease activities (Murphy, 1991).

1.2 Use of λ Red recombination for manipulation of bacterial genomes

The basic strategy of the λ Red system is the replacement of a chromosomal sequence with a (e.g., PCR-amplified) selectable antibiotic resistance gene flanked by homology extensions of distinct lengths. For genetic engineering in the *E. coli* chromosome, two efficient λ Red-mediated methods were developed by two independent groups. The first method utilized *E. coli* strains containing the P_L operon of a defective λ prophage (e.g., deletion of *cro* to *bioA* genes, which includes the lytic genes) under control of the temperature-sensitive λ cI-repressor (allele *cI857*, Yu et al., 2000). To express high levels of the P_L operon, cultures were shifted from repressing conditions at 32°C to inducing conditions at 42°C for 7.5 to 17.5 minutes. This was optimal for achieving maximal recombination levels. By shifting the cells back to 32°C, further expression of the P_L operon was repressed; and cell death was prevented. Furthermore, the absence of the Cro-repressor enabled P_L operon expression to be fully derepressed when the cI-repressor was inactivated under heat induction (Yu et al., 2000). Electroporation of this transiently induced λ lysogen with a short linear, PCR-generated dsDNA segment resulted in efficient recombination events (Poteete, 2001).

The second very efficient λ Red-mediated recombination approach, involved a low-copy plasmid with λ *gam*, *bet*, and *exo* under control of an arabinose-inducible promoter (pKD46, Datsenko et al., 2000). The plasmid pKD46 yielded greatly enhanced recombination events and was preferable to the similar pKD20. However, both harbor the Red system under a well-regulated promoter to avoid undesired reactions under non-inducing conditions and a temperature-sensitive replicon to allow for easy curing of the respective plasmid after recombination (Datsenko et al., 2000). For generating gene disruptions within the *E. coli* chromosome, the R6Kγ *ori*-based suicide plasmids pKD3 and pKD4 were used as templates to amplify the chloramphenicol (*cat*) or kanamycin (*aph*) resistance gene cassettes, respectively, in a PCR. Using primer pairs with site-specific homology extensions to amplify the resistance cassette, the resulting PCR product used to transform a freshly competent *E. coli* expressing λ Red from pKD46. After successful recombination, the resistance gene cassette can be excised by Flp recombinase supplied *in trans* (1.2.3, Fig. 1).

One example of a possible refinement of the λ Red procedure promotes high-frequency recombination using ssDNA substrates. It has been discovered that only λ Redβ is

absolutely required for ssDNA recombination (Ellis et al., 2001). Neither *exo* nor *gam* causes a dramatic effect on the recombination efficiency. Only a minor 5-fold reduction was observed in the *gam* deletion strain, most likely attributed to the single-stranded nuclease activity of the RecBCD protein complex, which is not inhibited in a *gam*-deficient strain (Ellis et al., 2001, Wang et al., 2000). λ Redβ bound to ssDNA is able to protect the DNA segment from nuclease attack, which might explain the recombination events that occured despite the *gam* deletion (Karakousis et al., 1998).

These methods offer a technology for studying bacterial gene functions or even for introducing mutations or markers in the chromosomes of eukaryotic cells, e.g., to provide special ''tags'' in the DNA of living cells (Ellis et al., 2001).

1.2.1 Gene deletion

λ Red recombination has been successfully used for convenient generation of gene deletions in *E. coli* (Datsenko et al., 2000), *Salmonella enterica* (Hansen-Wester et al., 2002), *Pseudomonas aeruginosa* (Liang et al., 2010, Quénée et al., 2005), *Streptomyces coelicolor* (Gust et al., 2004), *Shigella* spp. (Ranallo et al., 2006), *Yersinia pseudotuberculosis* (Derbise et al., 2003) and *Y. pestis* (Sun et al., 2008).

The first step in generating gene deletions is creating a linear targeting construct which consists usually of a resistance gene ("*res*" in Fig. 1) flanked by terminal extensions homologous to the sequences surrounding the deletion site. In Fig. 1, the homologous regions used for deletion of *orf*2 are depicted in blue and red. In the simplest approach, substrates for Red recombination can be generated by adding short homology extensions to PCR primers and using them to amplify a resistance cassette (Datsenko et al., 2000, Yu et al., 2000, Zhang et al., 2000). The length of the homology arms required for efficient recombination depends on the organism. For example, 36-50 bp (base pairs) were shown to work for *E. coli* or *Salmonella enterica* (Datsenko et al., 2000, Hansen-Wester et al., 2002) and 50-100 bp for *Pseudomonas aeruginosa* (Liang et al., 2010). Much longer homology arms were required for efficient Red recombination in *Y. pseudotuberculosis* and *Y. pestis* (~500 bp, Derbise et al., 2003, Sun et al., 2008) and *Vibrio cholerae* (100-1000 bp, Yamamoto et al., 2009). For the latter examples, multiple PCRs have to be done to generate functional targeting constructs.

In the next step, the PCR product is used to transform bacteria expressing λ Red proteins. Homologous recombination results in insertion of the cassette at the precise position determined by the homology extensions (Fig. 1). Transformants can be selected using their acquired antibiotic resistance. Target regions for site-specific recombinases (Fig. 1, yellow triangles) provide the option for subsequent removal of the resistance cassette (see also 1.2.3).

1.2.2 DNA insertion

In addition to removing DNA from bacterial genomes (1.2.1), λ Red recombination can also be applied to precisely insert any DNA within a genome. This approach has been widely used for analyzing bacterial gene expression via the generation of reporter gene fusions (Gerlach et al., 2007a, Lee et al., 2009, Yamamoto et al., 2009) or epitope tagging (Cho et al., 2006, Lee et al., 2009, Uzzau et al., 2001). In a similar approach, promoter sequences can be inserted or exchanged within the genome (Alper et al., 2005, Wang et al., 2009).
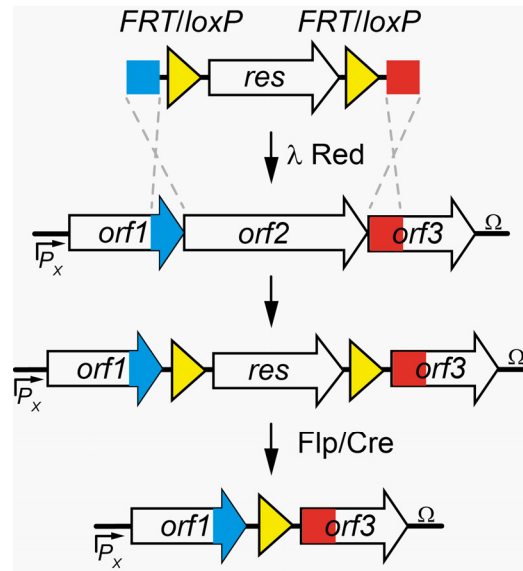
Fig. 1. Deletion of *orf2* using λ Red recombination and subsequent removal of resistance marker (*res*) via Flp or Cre recombinase. Red and blue regions denote regions of homology. The yellow region denotes the target for the site-specific recombinase.

In these cases, the targeting construct includes besides a selectable marker the DNA to be inserted. Using primers with homology extensions, these targeting constructs can be amplified by PCR from sets of template vectors available for different reporter genes (e.g., ß-galactosidase, luciferase, green fluorescent protein (*gfp*) and epitope tags (e.g., Flag®, haemagglutinin (HA), 8xmyc, 6xHis). An interesting alternative that obviates the need for multiple template vectors was developed by Gust et al. (2004). The resistance cassette and the reporter gene were amplified in separate PCRs, using in each reaction only one primer with the homology extension for Red recombination. The other primers included regions of homology to each other to allow them to anneal. The joint molecule was then used in a second round of PCR to generate a fusion cassette.

Depending on the scientific question to be answered, different integration strategies for reporter genes are available. For transcriptional fusions, a promoterless reporter gene is inserted downstream of a promoter of interest. The reporter gene may have optimized translational signals, including an optimized ribosome binding site (RBS) at the optimal distance from the start codon. If such a construct is inserted within an operon, hybrid operons are generated (Gerlach et al., 2007a). We have introduced so-called "start codon fusions," in which the reporter gene is inserted behind the native RBS and start codon of the gene under study, so that expression is assessed in the native genomic context (Gerlach et al., 2007a, Wille et al., 2012). This gene fusion strategy is closely related to translational fusions. The classical Red recombination protocol enables the easy generation of C-terminal fusion proteins, in which the reporter gene or epitope tag is inserted in-frame at any position in an open-reading-frame (ORF). For the generation of N-terminal fusions, a "scarless" recombination protocol (see 2.) has to be applied.

1.2.3 Site-specific recombination for removal of antibiotic resistance genes

Several methods, involving various site-specific recombination systems, have been developed for the removal of unwanted marker sequences from the chromosome. The most frequently used site-specific recombinases for subsequent excision of antibiotic resistance genes are Flp and Cre. Flp and Cre recombinases recognize 34-bp long sequences with palindromic elements (*FRT* and *loxP*, respectively). Both sites, *FRT* and *loxP*, consist of two 13-bp inverted repeats flanking an 8-bp non-palindromic core sequence. The asymmetry of the core sequence determines the polarity of the recombination site and has extensive consequences for the outcome of the recombination. Recombination between directly orientated recombination sites leads to excision of the intervening sequence (Fig. 1, yellow triangles), while recombination between inverted recombination sites results in an inversion of the intervening sequence. Recombination continuously takes place as long as the recombinase is present (Schweizer, 2003, Zhang et al., 2002). However, removal of resistance genes by either one of the site-specific recombinases leaves a *loxP* or *FRT* "scar" sequence within the genome (Datsenko et al., 2000, Lambert et al., 2007).

Although there is limited homology between the scar sequences themselves, they might serve as hotspots for recombination in successive recombination steps, representing a risk for unwanted deletions or chromosomal rearrangements (Datsenko et al., 2000). In addition, these scars might have influence on gene functions when operon structures or intragenic regions were modified (Blank et al., 2011). Mutations within the inverted repeats of *loxP*, resulting in the *lox66* and *lox71* alleles, can minimize genetic instability. Recombination using either generates *lox72*, which has a strongly reduced binding affinity for Cre (Lambert et al., 2007).

2. Site-directed mutagenesis using oligonucleotides

Precise insertion of chromosomal mutations has been established as the "gold standard" for analysis of bacterial gene function. Generation of point mutations, seamless deletions and in-frame gene fusions without leaving selectable markers or a recombination target site (e.g., *loxP* or *FRT*) requires reliable counterselectable markers. The published protocols are usually based on two successive rounds of recombination: (I) integration of a positively selectable marker (e.g., an antibiotic resistance gene) in the first step and (II) selection for marker loss (counterselection) in the final recombination step (Reyrat et al., 1998). Some of the counterselection methods have been shown to function together with the recombination of short synthetic DNA fragments. This circumvents the tedious and time-consuming requirement for PCR-based mutagenesis and cloning to generate mutant alleles for the second recombination step.

2.1 Counterselection with SacB

The *Bacillus subtilis sacB* gene is widely used as a genetic tool for counterselection puposes. SacB confers sucrose sensitivity to a wide range of Gram-negative bacteria (Blomfield et al., 1991, Kaniga et al., 1991). The levansucrase SacB catalyzes transfructorylation from sucrose to various acceptors, hydrolysis of sucrose and synthesis of levans (Gay et al., 1985). The molecular basis of its toxic effect on many Gram-negative bacteria is still not completely understood. It is thought that accumulation of levans in the periplasm or transfer of fructose

residues to inappropriate acceptors, subsequently generating toxic compounds, might play an important role (Pelicic et al., 1996).

Linear targeting constructs harboring *sacB* combined with a kanamycin resistance gene (*neo*, Zhang et al., 1998), an ampicillin resistance gene (*bla*, Liang et al., 2010) or a chloramphenicol resistance gene (*cat*, Sun et al., 2008) were used for the first recombination rounds. An interesting marker combining the ability for positive selection and counterselection in one fusion gene is *sacB-neo*. The resulting protein SBn confers kanamycin resistance, as well as sucrose sensitivity, to Gram-negative bacteria.

The latter two methods were used to generate gene deletions within the chromosomes of *P. aeruginosa* and *Y. pestis*, respectively. In these organisms, the λ Red system was used to recombine the targeting constructs with homology extensions of 50-100 bp (Liang et al., 2010) or ~500 bp (Sun et al., 2008) flanking the cassettes.

In the homologous recombination step, clones were selected for the respective antibiotic resistance. Recombinants were selected on medium plates containing 5-7% sucrose to select for loss of the cassette. Exact timing of counterselection is a critical issue when working with SacB or SBn, since *sacB*-inactivating mutations were shown to accumulate. For example, only 10-15% (not 100%) of *sacB*⁻clones lost their kanamycin resistance after counterselection of SBn on sucrose plates (Muyrers et al., 2000a). Nevertheless, Muyrers and coworkers successfully used this selection scheme to introduce a point mutation within a BAC. In these experiments, the altered base was included in one of the homology arms of both targeting constructs. Interestingly, the investigators observed that the homologous arms for the second targeting construct should be at least 500 bp in length. This excludes the use of a completely synthetic oligonucleotide at that step (Muyrers et al., 2000a).

2.2 Dual selection of recombinants with GalK or ThyA

Besides the fusion protein SBn, *E. coli* galactokinase (GalK) and thymidylate synthase A (ThyA) can each be used in the dual role of both a positive selective marker (selection for the gene) and a negative selective marker (selection for the absence of the gene). Depending on the substrates used for growing recombinants, gain or loss of the markers can be selected very efficiently. Both systems were established for the manipulation of BACs in appropriate *E. coli* host strains that are deficient for *galK* or *thyA* and have the λ Red recombination system present within their genomes (Wong et al., 2005, Warming et al., 2005). These methods cannot be used for site-directed mutagenesis of genomes of bacteria with functional *galK* or *thyA*.

The *galETKM* operon enables *E. coli* to metabolize D-galactose (Semsey et al., 2007). In the initial step, GalK catalyzes the phosphorylation of D-galactose to galactose-1-phosphate. Recombinants were selected after the first recombination step with D-galactose as the sole carbon source. A functional GalK is absolutely required for growth under these conditions. Besides D-galactose, GalK can efficiently phosphorylate a galactose analog, 2-deoxy-galactose (DOG). Because the product 2-deoxy-galactose-1-phosphate cannot be further metabolized, it is enriched to toxic concentrations (Alper et al., 1975). In the second recombination step, an oligonucleotide harboring the desired mutation was used as substrate for Red recombination. Loss of GalK was selected on plates containing glycerol as sole carbon source and DOG (Warming et al., 2005).

A *thyA* mutation results in thymine auxotrophy of *E. coli*, since the enzyme is required for dTTP and, therefore, DNA *de novo* synthesis. Integration of the *thyA*-containing targeting construct in the first recombination step was selected on minimal medium in the absence of thymine. For its function, ThyA requires tetrahydrofolate (THF) as a cofactor. During the process THF is oxidized to dihydrofolate (DHF). The pool of THF is replenished from DHF by dihydrofolate reductase (DHFR). The action of DHFR can be inhibited by the antibiotic trimethoprim. Using a growth medium supplemented with trimethoprim and thymine, loss of *thyA* was selected in the second round of recombination. At this step, a PCR product containing a mutated allele was applied as targeting construct (Wong et al., 2005).

2.3 Counterselection using streptomycin resistance

Several mutants of the ribosomal protein S12 (RpsL) were shown to confer streptomycin resistance (Sm^R, Springer et al., 2001). Strains harboring such an *rpsL* allele (e.g., *rpsL150*) can be used as hosts in another counterselection method. The main principle is based on the fact that streptomycin resistance-conferring mutations are recessive in merodiploid strains (Lederberg, 1951). In the first step, the wild type (wt) allele of *rpsL* was inserted within the gene of interest using λ Red or RecE/T recombination (1.1). A cassette containing an *rpsL-neo* fusion gene was used. Clones exhibited Km^R and Sm^S (Heermann et al., 2008). The desired point mutation for the gene of interest had already been introduced into one of the two 50-bp homology arms of the *rpsL-neo* cassette. The *rpsL-neo* cassette was then deleted in a second round of recombination. For this step, a single-stranded oligonucleotide containing no *rpsL* allele but containing arms of homology to the gene of interest were used. To maintain the mutation, one of the homology arms again harbored the desired mutation. Recombinants that deleted the wt *rpsL* could be selected by their resistance to streptomycin. After successful recombination, the wt copy of *rpsL* was removed, and the gene of interest was re-established, now containing the desired mutation (Heermann et al., 2008).

2.4 Selection with the fusaric acid sensitivity system

A counterselection technique developed by Bochner et al. (1980) enables direct selection of tetracycline sensitive (Tc^S) clones from a predominantly tetracycline resistant (Tc^R) bacterial population. The method is based on the hypersensitivity of lipophilic Tc^R cells to chelating agents, like fusaric acid or quinaldic acids. The precise mechanism of tetracycline exclusion is so far unknown and the subject of much speculation. The hypersensitivity seems to be caused by alterations of the host cell membrane, which are evoked from the expression of the tetracycline resistance gene. These alterations interfere, on one hand, with tetracycline permeation to confer tetracycline resistance, but, on the other hand, also to increase susceptibility to other toxic compounds (Bochner et al., 1980). This effect was exploited by using a medium that was effective for the selection of Tc^S revertants. The counterselection was successful in *Salmonella*; but it was much less effective with most, especially fast-growing, *E. coli* strains (Bochner et al., 1980). Decreasing the nutrient concentration of the selection plates significantly minimized the background of Tc^R colonies of fast-growing bacteria (Maloy et al., 1981).

The counterselection of Tc^R clones on Bochner-Maloy plates was sometimes used as the final step in recombineering protocols. Point mutations were inserted in BACs using a combination of λ *gam* (1.1) with RecE/T (1.) to integrate the gene for tetracycline resistance.

Fig. 2. Use of a *tet* cassette in conjunction with Bochner-Maloy Tc[S]-selective medium for counterselection. Blue and red regions denote regions of homology. "mut" indicates a mutation. The mutation is first in the targetting construct, then recombination transfers it to "*orfX*" ("*orfX*"mut). "P$_X$" designates the promoter for *orfX*; "Ω" indicates a transcriptional terminator.

A PCR product carrying the desired mutation was used in the second recombination step to exchange *tet*. Recombinants (Tc[S]) were selected on Bochner-Maloy plates (Nefedov et al., 2000). Similar two-step recombination approaches were also used to manipulate the genome of *Salmonella enterica* serovar Typhimurium ("*S*. Typhimurium") (Gerlach et al., 2009, Karlinsey, 2007). For this technique, a *tetAR* cassette, which encodes Tc[R], was inserted in the first step within a target gene ("*orfX*") with the help of homology extensions (blue and red, Fig. 2) resulting in tetracycline resistant clones. In the second recombination step, either PCR-derived mutant alleles (Karlinsey, 2007) or synthetic oligonucleotides (Gerlach et al., 2009) were used as targeting constructs to remove the *tetAR* cassette. For the latter rationale, no further cloning steps to generate the mutant allele (Fig. 2 "mut") were required. The applicability of this approach was proven with a *Salmonella* virulence-associated gene as an example. *siiF* encodes the putative ATPase of a type I secretion system (T1SS) located within

the *Salmonella* Pathogenicity Island 4 (SPI-4, Gerlach et al., 2007b). Previous work on a homologue demonstrated, that a single amino acid exchange within the Walker Box A of the ABC (ATP-binding cassette) motif disrupted the function of the transport ATPase (Koronakis, 1995). For our mutagenesis of *siiF*, we designed oligonucleotides to introduce a silent mutation resulting in a novel *Nla*IV site (2.6.1), as well as a change of Gly at position 500 to Glu (G500E) or a change of Lys at position 506 to Leu (K506L). Both amino acid positions are within the ABC motif. As a control, we introduced a silent mutation to generate a new *Sac*I restriction site (2.6.1) within *siiF*. After growth selection on Bochner-Maloy medium plates that favor the growth of Tc^S bacteria, clones were screened for the newly inserted restriction sites by the relevant restriction enzymes. Positive recombinants were subjected to functional analyses. The experiments showed the expected results: (I) no influence of the silent mutations and (II) loss of substrate secretion from the amino acid exchanges within Walker Box A (Gerlach et al., 2009).

The selection efficiency of Bochner-Maloy plates was reported not to exceed 50% (Podolsky et al., 1996). Therefore, the selection procedure was not very stringent. Exact timing of all incubation steps was necessary; but still high background might be observed, making purification of positive clones difficult. Highly increased selection efficiencies were obtained with plates containing 5-7 mM NiCl$_2$, which led to 80-100% positive Tc^S revertants (Podolsky et al., 1996).

2.5 Double-strand breaks introduced by I-*Sce*I can be used to select recombinants

The endonuclease I-*Sce*I of the yeast *Saccharomyces cerevisiae* is a novel tool for counter-selection. I-*Sce*I is an endonuclease with a long recognition sequence of 18 bp, thus ensuring the statistical absence of natural I-*Sce*I recognition sites within bacterial genomes (Monteilhet et al., 1990). Counterselection with I-*Sce*I is based on the induction of lethal double-strand breaks (DSB) within the genome or BAC, thus inhibiting DNA propagation. The Red system promotes enough recombination that recombinants can be obtained by screening survivors.

Several methods for site-directed mutagenesis of BACs and/or bacterial genomes utilizing I-*Sce*I expression have been published. Usually the methodologies are based on the insertion of an I-*Sce*I recognition sequence together with a positively selectable marker near the sequence to be modified. Furthermore, a system allowing for transient expression of the I-*Sce*I restriction enzyme in a coordinated fashion after expression of the λ Red recombination system is required. For the manipulation of BACs, a special *E. coli* host strain was developed to facilitate the independent expression of λ Red and I-*Sce*I. *E. coli* GS1783 harbors within its genome λ Red under control of a temperature-sensitive repressor and I-*Sce*I under control of an arabinose-inducible promoter (Tischer et al., 2010). However, for modification of bacterial genomes the components for recombination and I-*Sce*I have to be provided on one or two plasmid(s). The single plasmid solutions allow for independent inducible expression of both functions using arabinose and tetracycline (pWRG99 and pGETrec3.1; Blank et al., 2011, Jamsai et al., 2003) or arabinose and rhamnose (pREDI, Yu et al., 2008).

For mutagenesis of the genomes of *Salmonella enterica* servoar Enteritidis (*S.* Enteritidis, Cox et al., 2007) and *E. coli* (Kang et al., 2004), setups with the Red components and I-*Sce*I encoded on two separate plasmids were used. Targeting constructs consisting of an I-*Sce*I

recognition site, a Km^R cassette and long flanking homology regions (>200 bp) were constructed in a two-step PCR approach. After chromosomal integration of the linear DNA via λ Red-mediated recombination, clones were selected with kanamycin (Cox et al., 2007). A two-step PCR approach was used to combine long extensions homologous to *lamB* with sequences encoding antigen epitopes in the second targeting construct. For production of antibodies against foreign proteins, epitopes were inserted into the outer membrane protein LamB to facilitate surface presentation. This linear PCR product was used in a co-transformation with pBC-I-SceI at a molar ratio of 40:1 (PCR product:plasmid) into λ Red-expressing *S*. Enteritidis. I-*Sce*I was constitutively expressed from plasmid pBC-I-SceI (Kang et al., 2004). Screening for the desired recombinants was based on the inability to grow on kanamycin-containing plates (Cox et al., 2007). Because there is no convenient possibility for plasmid curing, the pBluescript (Stratagene)-based pBC-I-SceI remains in the host strains (Cox et al., 2007, Kang et al., 2004).

GET recombination is a method developed for the manipulation of BACs. It employs λ Gam and RecE/T for recombination (1.1) and I-*Sce*I for counterselection (Jamsai et al., 2003). In a study by Jamsai et al. (2003), the I-*Sce*I endonuclease gene downstream of a repressed promoter, together with a constitutive gene for Km^R and an I-*Sce*I recognition site, was inserted within the gene of interest in the first recombination step. As targeting construct for the second recombination step, a 1708-bp PCR product carrying the desired mutation was inserted in exchange for the I-*Sce*I-kanamycin resistance cassette. I-*Sce*I expression was induced for 30 minutes with addition of heat-treated chlorotetracycline. Expression was induced from both the inserted I-*Sce*I-kanamycin resistance cassette and plasmid pGETrec3.1 (Jamsai et al., 2003). The kanamycin resistance cassette, with its I-*Sce*I recognition site, was successfully removed from 23.6% of the colonies surviving expression of I-*Sce*I (Jamsai et al., 2003).

Because no specific mechanisms were implemented in pGETrec3.1 and pBC-I-SceI to promote convenient plasmid curing, it might be difficult to get plasmid-free host strains after site-directed mutagenesis. We solved that problem by integrating a tetracycline-inducible I-*Sce*I expression cassette from pST98-AS (Pósfai et al., 1999) in the temperature-sensitive λ Red expression plasmid pKD46 (Datsenko et al., 2000) to generate pWRG99 (Blank et al., 2011). In a first recombination step, a Cm^R cassette (*cat*) together with an I-*Sce*I recognition site (dark green, Fig. 3A) was integrated within *phoQ* in the genome of *S*. Typhimurium. For that, extensions of 40 nt length homologous to the regions surrounding the intended mutation site within *phoQ* were added to the primers (Fig. 3A, blue and red). PhoQ is the histidine sensor kinase of the *Salmonella* virulence-associated two-component signaling system PhoPQ. Successful recombinants were selected using chloramphenicol, and correct integration of the resistance cassette was checked by colony PCR and sequencing. For the second round of recombination, 80mer dsDNAs, derived from oligonucleotides, were introduced into pWRG99-harboring cells expressing λ Red recombination genes. These 80mer dsDNAs were designed (I) to delete the *phoQ* gene (not shown) or (II) to introduce a threonine to isoleucine exchange at position 48 (T48I) of PhoQ, together with a new *Sac*II restriction site (Fig. 3A). Expression of I-*Sce*I was induced with addition of anhydrotetracycline (AHT) leading to lethal DSBs in the clones still harboring the I-*Sce*I recognition site. Surviving clones were screened by PCR, restriction analysis using *Sac*II as well as phenotypically. The single amino acid exchange T48I within the periplasmic domain

of PhoQ results in constitutive activation of the response regulator PhoP (Miller, 1990). Besides virulence attenuation, constitutive PhoP activation leads to overexpression of the nonspecific acid phosphatase PhoN. Successful recombinants could therefore be screened phenotypically by forming blue colonies on 5-bromo-4-chloro-3-indolyl-phosphate toluidine salt (BCIP) plates due to increased PhoN activity. Correct recombination events could be further confirmed by a macrophage infection model, which showed the predicted virulence-attenuated phenotype (Blank et al., 2011).
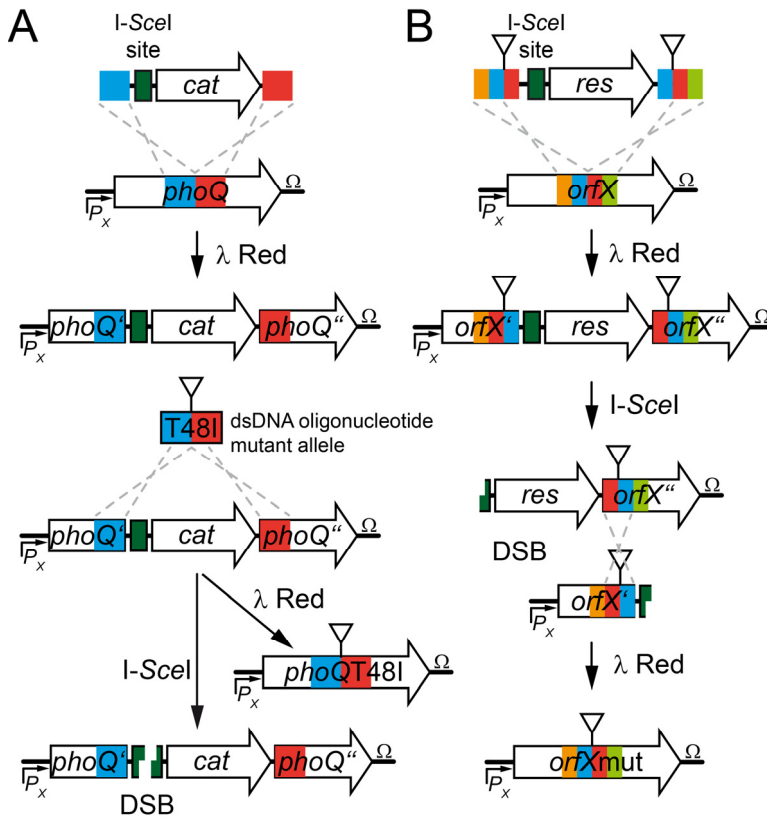
Fig. 3. Methods for single nucleotide exchange using I-*Sce*I counterselection. Details for A and B are in the text. The symbols and colors are the same as those described in Fig. 2. However in A *orfX* is *phoQ*. The antibiotic resistance is Cm^R; the resistance gene is *cat*. The mutation (inverted triangle) is a point mutation that changes threonine at position 48 in PhoQ to isoleucine (T48I). The green rectangle is the recognition site for the I-*Sce*I endonuclease; "DSB" is the "double-strand break" induced by I-*Sce*I. In B, the symbols are the same as those in A and Fig. 2. Unique to B are "*res*" (antibiotic resistance) and the modular extensions for the primers (orange, blue, red, and light green). All are homologous to *orfX*. The two internal primers (relative to the intact gene) have the blue and red extensions, which will become the duplicated region. The "left" primer has unique homology (orange); the "right" primer has unique homology (light green).

The Red recombination system can anneal single-stranded DNA derived from dsDNA substrates into replicating homologous target sequences (1.1). Usually homologous sequences for recombination are supplied with the homology extensions flanking the targeting construct. In contrast, homologous regions flanking a DSB generated by I-*Sce*I can also act as substrates for Red recombination. This strategy has been used for scarless site-directed mutagenesis of BACs (Tischer et al., 2010, Tischer et al., 2006) and the *E. coli* genome (Yu et al., 2008). These approaches require the integration of a duplicated sequence stretch in the first recombination round to serve as a substrate for recombination in the second round. For seamless deletions or insertion of point mutations, the duplications can be readily incorporated within the primers used to amplify the positive selection marker. In one study, modular sequence extensions, each ~20 nt in length (Fig. 3B; orange, blue, red and light green), were added to the primers used to amplify a resistance cassette (*res*). The two primers were unique but shared about 40 nt of sequence. The desired mutation (Fig. 3B, inverted triangles) was included in the duplicated sequence. This resulted in a ~40-bp sequence duplication (red and blue) after integration of the targeting construct into the gene of interest (*orfX*) via Red recombination (Fig. 3B, Tischer et al., 2010). The duplicated sequences were separated by the I-*Sce*I recognition site and after induction of DSB, the Red-mediated recombination between these sequences led to the reconstitution of *orfX* in its mutated form *orfX*mut (Fig. 3B). Figure 3B shows a generalization of this strategy. If longer DNA sequences need to be integrated (1.2.2), a preceding cloning step is required to insert the selectable marker together with the I-*Sce*I recognition site and a sequence duplication into the DNA to be inserted (Tischer et al., 2010). Combining I-*Sce*I-induced DSB with SacB-mediated sucrose sensitivity (2.1) was shown to improve selection for loss of resistance marker within the *E. coli* genome (Yu et al., 2008).

One major problem of the I-*Sce*I counterselection approach is the accumulation of point mutations within or deletion of the I-*Sce*I recognition site during the selection process. This effect demands tight regulation of I-*Sce*I expression to minimize selection pressure before the final (markerless) recombination takes place. It was important to optimize the procedures to maximize counterselection after the 2nd round of recombination.

2.6 Screening methods for recombinants

An underestimated problem is the screening effort needed to identify correct recombinants when using seamless recombination techniques. Although PCR fragment length polymorphism can be used in case of deletions and insert-specific PCRs in case of DNA insertion, successful single nucleotide exchanges are hard to detect. Direct phenotypical screening or the parallel introduction of novel restriction sites together with the nucleotide exchange are solutions of the problem.

2.6.1 Introduction of silent mutations to generate novel restriction sites

A screening problem arises if mutations introduced via recombineering have no direct or indirect impact on the phenotype or if the phenotypic test required is very time-consuming. Introduction of a novel restriction site adjacent to the mutation was proven to be very useful for colony screening. Designing the oligonucleotides for λ Red recombination offers the prospect of introducing silent mutations in the target region. Identification of silent

mutations generating novel restriction sites can be done *in silico* using WatCut [an online tool for restriction analysis, silent mutation scanning and SNP-RFLP analysis (http://watcut.uwaterloo.ca/)] or other DNA analysis software (e.g., Clone Manager). As mentioned before, this screening method has been used successfully [e.g., for screening *siiF* recombinants by introducing a new *Nla*IV and/or *Sac*II restriction site (2.4), as well as for the *phoQ*T48I mutation by generating a novel *Sac*II restriction site (2.5)].

2.6.2 Phenotypical screening

If available, phenotypical screening is the fastest way for selecting recombinants with the desired mutation. The screening is based on phenotypic differences between the mutant and the wt. In the simplest case, activity of an integrated reporter gene like *gfp* or *lacZ'* might be detected (Gerlach et al., 2007a). Another possibility for screening is the ability (gain of function) or inability (loss of function, auxotrophy) of the mutant to grow under specific conditions. In the latter case it is necessary to define permissive conditions in which the mutant can grow and restrictive conditions in which it cannot. Moreover, mutations might lead to differences in cell morphology that can be identified by microscopic examination. Last but not least, there may be a difference in the ability to utilize a chromogenic substrate, like BCIP, *p*-nitrophenyl-phosphate (pNPP) or 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal). These could be used for screening of positive recombinants. All of the phenotypes can originate directly from the activity of the mutated gene product or indirectly by influencing the activity and/or expression level of other proteins. One example is the PhoQ T48I mutant, which causes overexpression of PhoN. The overexpression can be monitored using the chromogenic substrate BCIP (2.5, Blank et al., 2011).

3. Conclusion

Two successive recombination steps catalyzed by the phage λ Red or phage Rac RecE/T recombination systems in combination with a negative selection procedure provide a venue for scarless mutagenesis within bacterial genomes and BACs. The outstanding ability of these enzymes to use homologous sequences as short as 35 bp as substrates for recombination allows the use of linear DNA derived from synthetic oligonucleotides as targeting constructs. The limiting step of this rationale is the availability of a reliable counterselection method. Here we gave an overview about recombination and the counterselection techniques successfully applied to the manipulation of bacterial genomes, as well as BACs.

4. References

Alper, H., Fischer, C., Nevoigt, E. & Stephanopoulos, G. (2005). Tuning genetic control through promoter engineering. *Proc Natl Acad Sci U S A*, Vol.102, No.36, pp. 12678-12683, ISSN 0027-8424

Alper, M.D. & Ames, B.N. (1975). Positive selection of mutants with deletions of the *gal-chl* region of the *Salmonella* chromosome as a screening procedure for mutagens that cause deletions. *J Bacteriol*, Vol.121, No.1, pp. 259-266, ISSN 0021-9193 (Print)

Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F. & Cullin, C. (1993). A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. *Nucleic Acids Res*, Vol.21, No.14, pp. 3329-3330, ISSN 0305-1048

Blank, K., Hensel, M. & Gerlach, R.G. (2011). Rapid and Highly Efficient Method for Scarless Mutagenesis within the *Salmonella enterica* Chromosome. *PLoS One*, Vol.6, No.1, pp. e15763, ISSN 1932-6203

Blomfield, I.C., Vaughn, V., Rest, R.F. & Eisenstein, B.I. (1991). Allelic exchange in *Escherichia coli* using the *Bacillus subtilis sacB* gene and a temperature-sensitive pSC101 replicon. *Mol Microbiol*, Vol.5, No.6, pp. 1447-1457, ISSN 0950-382X

Bochner, B.R., Huang, H.C., Schieven, G.L. & Ames, B.N. (1980). Positive selection for loss of tetracycline resistance. *J Bacteriol*, Vol.143, No.2, pp. 926-933, ISSN 0021-9193

Cho, B.K., Knight, E.M. & Palsson, B.O. (2006). PCR-based tandem epitope tagging system for *Escherichia coli* genome engineering. *Biotechniques*, Vol.40, No.1, pp. 67-72, ISSN 0736-6205

Clark, A.J., Satin, L. & Chu, C.C. (1994). Transcription of the *Escherichia coli recE* gene from a promoter in Tn*5* and IS*50*. *J Bacteriol*, Vol.176, No.22, pp. 7024-7031, ISSN 0021-9193

Copeland, N.G., Jenkins, N.A. & Court, D.L. (2001). Recombineering: a powerful new tool for mouse functional genomics. *Nat Rev Genet*, Vol.2, No.10, pp. 769-779, ISSN 1471-0056

Cox, M.M., Layton, S.L., Jiang, T., Cole, K., Hargis, B.M., Berghman, L.R., Bottje, W.G. & Kwon, Y.M. (2007). Scarless and site-directed mutagenesis in *Salmonella enteritidis* chromosome. *BMC Biotechnol*, Vol.7, pp. 59, ISSN 1472-6750

Datsenko, K.A. & Wanner, B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A*, Vol.97, No.12, pp. 6640-6645, ISSN 0027-8424

Derbise, A., Lesic, B., Dacheux, D., Ghigo, J.M. & Carniel, E. (2003). A rapid and simple method for inactivating chromosomal genes in *Yersinia*. *FEMS Immunol Med Microbiol*, Vol.38, No.2, pp. 113-116, ISSN 0928-8244

Ellis, H.M., Yu, D., DiTizio, T. & Court, D.L. (2001). High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc Natl Acad Sci U S A*, Vol.98, No.12, pp. 6742-6746, ISSN 0027-8424

Fouts, K.E., Wasie-Gilbert, T., Willis, D.K., Clark, A.J. & Barbour, S.D. (1983). Genetic analysis of transposon-induced mutations of the Rac prophage in *Escherichia coli* K-12 which affect expression and function of *recE*. *J Bacteriol*, Vol.156, No.2, pp. 718-726, ISSN 0021-9193

Gay, P., Le Coq, D., Steinmetz, M., Berkelman, T. & Kado, C.I. (1985). Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *J Bacteriol*, Vol.164, No.2, pp. 918-921, ISSN 0021-9193

Gerlach, R.G., Hölzer, S.U., Jäckel, D. & Hensel, M. (2007a). Rapid engineering of bacterial reporter gene fusions by using red recombination. *Appl Environ Microbiol*, Vol.73, No.13, pp. 4234-4242, ISSN 0099-2240

Gerlach, R.G., Jäckel, D., Hölzer, S.U. & Hensel, M. (2009). Rapid oligonucleotide-based recombineering of the chromosome of *Salmonella enterica*. *Appl Environ Microbiol*, Vol.75, No.6, pp. 1575-1580, ISSN 1098-5336

Gerlach, R.G., Jäckel, D., Stecher, B., Wagner, C., Lupas, A., Hardt, W.D. & Hensel, M. (2007b). *Salmonella* Pathogenicity Island 4 encodes a giant non-fimbrial adhesin and the cognate type 1 secretion system. *Cell Microbiol*, Vol.9, No.7, pp. 1834-1850, ISSN 1462-5814

Gust, B., Chandra, G., Jakimowicz, D., Yuqing, T., Bruton, C.J. & Chater, K.F. (2004). λ red-mediated genetic manipulation of antibiotic-producing *Streptomyces*. *Advances in applied microbiology*, Vol.54, pp. 107-128, ISSN 0065-2164

Hansen-Wester, I. & Hensel, M. (2002). Genome-based identification of chromosomal regions specific for *Salmonella* spp. *Infect Immun*, Vol.70, No.5, pp. 2351-2360, ISSN 0019-9567

Heermann, R., Zeppenfeld, T. & Jung, K. (2008). Simple generation of site-directed point mutations in the *Escherichia coli* chromosome using Red®/ET® Recombination. *Microbial cell factories*, Vol.7, pp. 14, ISSN 1475-2859

Jamsai, D., Orford, M., Nefedov, M., Fucharoen, S., Williamson, R. & Ioannou, P.A. (2003). Targeted modification of a human β-globin locus BAC clone using *GET Recombination* and an I-*Sce*I counterselection cassette. *Genomics*, Vol.82, No.1, pp. 68-77, ISSN 0888-7543

Jasin, M. & Schimmel, P. (1984). Deletion of an essential gene in *Escherichia coli* by site-specific recombination with linear DNA fragments. *J Bacteriol*, Vol.159, No.2, pp. 783-786, ISSN 0021-9193

Kang, Y., Durfee, T., Glasner, J.D., Qiu, Y., Frisch, D., Winterberg, K.M. & Blattner, F.R. (2004). Systematic mutagenesis of the *Escherichia coli* genome. *J Bacteriol*, Vol.186, No.15, pp. 4921-4930, ISSN 0021-9193

Kaniga, K., Delor, I. & Cornelis, G.R. (1991). A wide-host-range suicide vector for improving reverse genetics in gram-negative bacteria: inactivation of the *blaA* gene of *Yersinia enterocolitica*. *Gene*, Vol.109, No.1, pp. 137-141, ISSN 0378-1119

Karakousis, G., Ye, N., Li, Z., Chiu, S.K., Reddy, G. & Radding, C.M. (1998). The beta protein of phage λ binds preferentially to an intermediate in DNA renaturation. *Journal of molecular biology*, Vol.276, No.4, pp. 721-731, ISSN 0022-2836

Karlinsey, J.E. (2007) λ-Red genetic engineering in *Salmonella enterica* serovar Typhimurium. In: *Advanced Bacterial Genetics: Use of Transposons and Phage for Genomic Engineering*, K.T. Hughes, S.R. Maloy (eds.), pp. 199-209, Academic Press, ISBN 0076-6879, New York

Kovall, R. & Matthews, B.W. (1997). Toroidal structure of λ-exonuclease. *Science*, Vol.277, No.5333, pp. 1824-1827, ISSN 0036-8075

Lambert, J.M., Bongers, R.S. & Kleerebezem, M. (2007). Cre-*lox*-based system for multiple gene deletions and selectable-marker removal in *Lactobacillus plantarum*. *Appl Environ Microbiol*, Vol.73, No.4, pp. 1126-1135, ISSN 0099-2240

Lederberg, J. (1951). Streptomycin resistance: a genetically recessive mutation. *J Bacteriol*, Vol.61, No.5, pp. 549-550, ISSN 0021-9193)

Lee, D.J., Bingle, L.E., Heurlier, K., Pallen, M.J., Penn, C.W., Busby, S.J. & Hobman, J.L. (2009). Gene doctoring: a method for recombineering in laboratory and pathogenic *Escherichia coli* strains. *BMC Microbiol*, Vol.9, pp. 252, ISSN 1471-2180

Lee, E.C., Yu, D., Martinez de Velasco, J., Tessarollo, L., Swing, D.A., Court, D.L., Jenkins, N.A. & Copeland, N.G. (2001). A highly efficient *Escherichia coli*-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics*, Vol.73, No.1, pp. 56-65, ISSN 0888-7543

Li, Z., Karakousis, G., Chiu, S.K., Reddy, G. & Radding, C.M. (1998). The beta protein of phage λ promotes strand exchange. *Journal of molecular biology*, Vol.276, No.4, pp. 733-744, ISSN 0022-2836

Liang, R. & Liu, J. (2010). Scarless and sequential gene modification in *Pseudomonas* using PCR product flanked by short homology regions. *BMC Microbiol*, Vol.10, No.1, pp. 209, ISSN 1471-2180

Little, J.W. (1967). An exonuclease induced by bacteriophage λ. II. Nature of the enzymatic reaction. *J Biol Chem*, Vol.242, No.4, pp. 679-686, ISSN 0021-9258

Lorenz, M.G. & Wackernagel, W. (1994). Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol Rev*, Vol.58, No.3, pp. 563-602, ISSN 0146-0749

Maloy, S.R. & Nunn, W.D. (1981). Selection for loss of tetracycline resistance by *Escherichia coli*. *J Bacteriol*, Vol.145, No.2, pp. 1110-1111, ISSN 0021-9193

Maresca, M., Erler, A., Fu, J., Friedrich, A., Zhang, Y. & Stewart, A.F. (2010). Single-stranded heteroduplex intermediates in λ Red homologous recombination. *BMC Mol Biol*, Vol.11, No.1, pp. 54, ISSN 1471-2199

Monteilhet, C., Perrin, A., Thierry, A., Colleaux, L. & Dujon, B. (1990). Purification and characterization of the *in vitro* activity of I-*Sce* I, a novel and highly specific endonuclease encoded by a group I intron. *Nucleic Acids Res*, Vol.18, No.6, pp. 1407-1413, ISSN 0305-1048

Murphy, K.C. (1991). λ Gam protein inhibits the helicase and   -stimulated recombination activities of *Escherichia coli* RecBCD enzyme. *J Bacteriol*, Vol.173, No.18, pp. 5808-5821, ISSN 0021-9193

Murphy, K.C. (1998). Use of bacteriophage λ recombination functions to promote gene replacement in *Escherichia coli*. *J Bacteriol*, Vol.180, No.8, pp. 2063-2071, ISSN 0021-9193

Muyrers, J.P., Zhang, Y., Benes, V., Testa, G., Ansorge, W. & Stewart, A.F. (2000a). Point mutation of bacterial artificial chromosomes by ET recombination. *EMBO Rep*, Vol.1, No.3, pp. 239-243, ISSN 1469-221X

Muyrers, J.P., Zhang, Y., Buchholz, F. & Stewart, A.F. (2000b). RecE/RecT and Redα/Redβ initiate double-stranded break repair by specifically interacting with their respective partners. *Genes Dev*, Vol.14, No.15, pp. 1971-1982, ISSN 0890-9369

Nefedov, M., Williamson, R. & Ioannou, P.A. (2000). Insertion of disease-causing mutations in BACs by homologous recombination in *Escherichia coli*. *Nucleic Acids Res*, Vol.28, No.17, pp. E79, ISSN 1362-4962

Passy, S.I., Yu, X., Li, Z., Radding, C.M. & Egelman, E.H. (1999). Rings and filaments of β protein from bacteriophage λ suggest a superfamily of recombination proteins. *Proc Natl Acad Sci U S A*, Vol.96, No.8, pp. 4279-4284, ISSN 0027-8424

Pelicic, V., Reyrat, J.M. & Gicquel, B. (1996). Expression of the *Bacillus subtilis sacB* gene confers sucrose sensitivity on mycobacteria. *J Bacteriol*, Vol.178, No.4, pp. 1197-1199, ISSN 0021-9193

Podolsky, T., Fong, S.T. & Lee, B.T. (1996). Direct selection of tetracycline-sensitive *Escherichia coli* cells using nickel salts. *Plasmid*, Vol.36, No.2, pp. 112-115, ISSN 0147-619X

Pósfai, G., Kolisnychenko, V., Bereczki, Z. & Blattner, F.R. (1999). Markerless gene replacement in *Escherichia coli* stimulated by a double-strand break in the chromosome. *Nucleic Acids Res*, Vol.27, No.22, pp. 4409-4415, ISSN 1362-4962

Poteete, A.R. (2001). What makes the bacteriophage λ Red system useful for genetic engineering: molecular mechanism and biological function. *FEMS microbiology letters*, Vol.201, No.1, pp. 9-14, ISSN 0378-1097

Quénée, L., Lamotte, D. & Polack, B. (2005). Combined *sacB*-based negative selection and *cre-lox* antibiotic marker recycling for efficient gene deletion in *Pseudomonas aeruginosa*. *Biotechniques*, Vol.38, No.1, pp. 63-67, ISSN 0736-6205

Ranallo, R.T., Barnoy, S., Thakkar, S., Urick, T. & Venkatesan, M.M. (2006). Developing live *Shigella* vaccines using λ Red recombineering. *FEMS Immunol Med Microbiol*, Vol.47, No.3, pp. 462-469, ISSN 0928-8244

Reyrat, J.M., Pelicic, V., Gicquel, B. & Rappuoli, R. (1998). Counterselectable markers: untapped tools for bacterial genetics and pathogenesis. *Infect Immun*, Vol.66, No.9, pp. 4011-4017, ISSN 0019-9567

Sawitzke, J. & Austin, S. (2001). An analysis of the factory model for chromosome replication and segregation in bacteria. *Mol Microbiol*, Vol.40, No.4, pp. 786-794, ISSN 0950-382X

Schweizer, H.P. (2003). Applications of the *Saccharomyces cerevisiae* Flp-*FRT* system in bacterial genetics. *Journal of molecular microbiology and biotechnology*, Vol.5, No.2, pp. 67-77, ISSN 1464-1801

Semsey, S., Krishna, S., Sneppen, K. & Adhya, S. (2007). Signal integration in the galactose network of *Escherichia coli*. *Mol Microbiol*, Vol.65, No.2, pp. 465-476, ISSN 0950-382X

Springer, B., Kidan, Y.G., Prammananan, T., Ellrott, K., Böttger, E.C. & Sander, P. (2001). Mechanisms of streptomycin resistance: selection of mutations in the 16S rRNA gene conferring resistance. *Antimicrobial agents and chemotherapy*, Vol.45, No.10, pp. 2877-2884, ISSN 0066-4804

Sun, W., Wang, S. & Curtiss, R., 3rd (2008). Highly efficient method for introducing successive multiple scarless gene deletions and markerless gene insertions into the *Yersinia pestis* chromosome. *Appl Environ Microbiol*, Vol.74, No.13, pp. 4241-4245, ISSN 1098-5336

Swaminathan, S., Ellis, H.M., Waters, L.S., Yu, D., Lee, E.C., Court, D.L. & Sharan, S.K. (2001). Rapid engineering of bacterial artificial chromosomes using oligonucleotides. *Genesis*, Vol.29, No.1, pp. 14-21, ISSN 1526-954X

Tischer, B.K., Smith, G.A. & Osterrieder, N. (2010) *En passant* mutagenesis: a two step markerless red recombination system. In: *In Vitro Mutagenesis Protocols: Third Edition*, J. Braman (ed.), pp. 421-430, Humana Press, ISBN 1940-6029, New York

Tischer, B.K., von Einem, J., Kaufer, B. & Osterrieder, N. (2006). Two-step red-mediated recombination for versatile high-efficiency markerless DNA manipulation in *Escherichia coli*. *Biotechniques*, Vol.40, No.2, pp. 191-197, ISSN 0736-6205

Uzzau, S., Figueroa-Bossi, N., Rubino, S. & Bossi, L. (2001). Epitope tagging of chromosomal genes in *Salmonella*. *Proc Natl Acad Sci U S A*, Vol.98, No.26, pp. 15264-15269, ISSN 0027-8424

Wang, H.H., Isaacs, F.J., Carr, P.A., Sun, Z.Z., Xu, G., Forest, C.R. & Church, G.M. (2009). Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, Vol.460, No.7257, pp. 894-898, ISSN 1476-4687

Wang, J., Chen, R. & Julin, D.A. (2000). A single nuclease active site of the *Escherichia coli* RecBCD enzyme catalyzes single-stranded DNA degradation in both directions. *J Biol Chem*, Vol.275, No.1, pp. 507-513, ISSN 0021-9258

Warming, S., Costantino, N., Court, D.L., Jenkins, N.A. & Copeland, N.G. (2005). Simple and highly efficient BAC recombineering using *galK* selection. *Nucleic Acids Res*, Vol.33, No.4, pp. e36, ISSN 1362-4962

Wille, T., Blank, K., Schmidt, C., Vogt, V. & Gerlach, R.G. (2012). *Gaussia princeps* Luciferase as a Reporter for Transcriptional Activity, Protein Secretion, and Protein-Protein Interactions in *Salmonella enterica* Serovar Typhimurium. *Appl Environ Microbiol*, Vol.78, No.1, pp. 250-257, ISSN 1098-5336

Wong, Q.N., Ng, V.C., Lin, M.C., Kung, H.F., Chan, D. & Huang, J.D. (2005). Efficient and seamless DNA recombineering using a thymidylate synthase A selection system in *Escherichia coli*. *Nucleic Acids Res*, Vol.33, No.6, pp. e59, ISSN 1362-4962

Yamamoto, S., Izumiya, H., Morita, M., Arakawa, E. & Watanabe, H. (2009). Application of λ Red recombination system to *Vibrio cholerae* genetics: Simple methods for inactivation and modification of chromosomal genes. *Gene*, Vol.438, No.1-2, pp. 57-64, ISSN 1879-0038

Yu, B.J., Kang, K.H., Lee, J.H., Sung, B.H., Kim, M.S. & Kim, S.C. (2008). Rapid and efficient construction of markerless deletions in the *Escherichia coli* genome. *Nucleic Acids Res*, Vol.36, No.14, pp. e84, ISSN 1362-4962

Yu, D., Ellis, H.M., Lee, E.C., Jenkins, N.A., Copeland, N.G. & Court, D.L. (2000). An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc Natl Acad Sci U S A*, Vol.97, No.11, pp. 5978-5983, ISSN 0027-8424

Zhang, Y., Buchholz, F., Muyrers, J.P. & Stewart, A.F. (1998). A new logic for DNA engineering using recombination in *Escherichia coli*. *Nat Genet*, Vol.20, No.2, pp. 123-128, ISSN 1061-4036

Zhang, Y., Muyrers, J.P., Testa, G. & Stewart, A.F. (2000). DNA cloning by homologous recombination in *Escherichia coli*. *Nat Biotechnol*, Vol.18, No.12, pp. 1314-1317, ISSN 1087-0156

Zhang, Z. & Lutz, B. (2002). Cre recombinase-mediated inversion using lox66 and lox71: method to introduce conditional point mutations into the CREB-binding protein. *Nucleic Acids Res*, Vol.30, No.17, pp. e90, ISSN 1362-4962

Studying Cell Signal Transduction with Biomimetic Point Mutations

Nathan A. Sieracki and Yulia A. Komarova
University of Illinois – Chicago
USA

1. Introduction

Post-translational modification (PTM), the chemical modification of a protein after its translation, represents an evolutionarily conserved mechanism of regulation of protein function. (Deribe et al., 2010) Through modulation of conformational change, enzymatic activity or interaction with other proteins, it often provides a transient switch between various intracellular signals to orchestrate an integrated response of the cell to environmental cues.

Site-directed mutagenesis has proven to be an essential tool in studying the molecular and signaling mechanisms of complex cellular functions through the ability to mutate a specific residue in the sequence of native protein in order to change its function. In combination with proteomics and database analysis of primary sequences, it offers a tool for exploring the transient nature of protein-protein interactions along with activity of enzymes, the primary assemblies of signaling nodes.

This chapter will highlight the use of site-directed mutagenesis for studying signal transduction networks. It will first describe how the diversity of the natural 22 amino acid 'toolbelt' results in fortuitous isostructural and isoelectronic similarity to several key post-translationally modifications (e.g. phosphorylation, sulfoxidation and nitrosylation). This similarity has allowed researchers to introduce gain- or loss-of-function substitutions into primary sequences and thus produce a form of the protein that mimics its active or inactive state. The substitution of serine or threonine for phosphomimetic aspartic or glutamic acid or non-phosphorylatable residues, such as alanine, will be discussed, along with other permutations, including natural analogs of oxidized methionine and cysteine.

While discussing how site-directed mutagenesis offers to researchers a tool to engineer a constitutively active OFF and ON forms of a protein, we will also provide examples of how PTM translates into a chemical phenotype through the regulation of enzyme activity or modulation of binding affinity to partners. Also discussed in this chapter are limitations and considerations to be taken when interpreting results obtained with biomimetic tools.

With advanced computing analysis and algorithms operating on primary sequences, the predictability of PTM sites is rapidly becoming an automated process. Complex networks can be accessed and verified. Even in the age of high-resolution mass spectrometry, site-directed mutagenesis will continue to be a powerful tool for studying and validation of signaling networks.

2. Harnessing nature's toolbelt with biomimetic point mutations

Site-directed mutagenesis, a molecular biology technique used to introduce a point mutation into primary DNA sequence at a defined site, was first described by Michael Smith in 1978. (Hutchison et al, 1978) His group induced specific mutations into the viral DNA strand of the bacteriophage phiX174 *am3* using complementary oligodeoxyribonucleotides and specific mutagens. "It appears to us that general methods for the construction of any desired mutant sequence would greatly increase the ability of sequencing methods to yield biologically relevant information. Such methods should also make it possible to construct DNAs with specific modified biological functions" M. Smith and his colleagues proposed in the paper. This was proven to be correct by vast number of studies, which exploited site-directed mutagenesis as a tool for testing structure-function relationships over the course of three decades. Providing a constitutively activated or deactivated form of the protein or removing entirely the potential for protein modification, it has also been instrumental in dissecting signal transduction networks. It is instructive to understand the precise strengths and limitations of biomimetic approaches in order to extract the most meaningful data from experiments.

2.1 Layers of diversity

The quest to understand and control structure-function relationship has revealed that Nature is an exquisite chemist. The already significant chemical diversity of the amino acid pool is amplified profoundly when individual amino acids are condensed into peptide chains. Individual residues can act in a seemingly modular fashion, or residues can coordinate to bestow more complex and macroscopic functional and catalytic characteristics to the proteins and peptides on which they reside. This diversity, however, is even more dramatic than the genetic message may initially predict.

As summarized in an excellent review by Wash and co-authors, (Walsh et al., 2005), Nature has devised two mechanisms of 'proteome diversification' over the course of evolution. The amount of unique proteins per unit RNA transcript has been dramatically amplified to be far more diverse than Central Dogma of protein expression would have suggested. The first method lies at the level of the transcript. Alternate splicing of the same gene product can result in a variety of unique proteins with unique functions. The second is through targeted post-translational modification of existing proteins and peptides to chemically 'tag' them. These "tags" can directly influence the structure and function of the protein, or it may be interpreted by another docking small molecule, protein, or assembly of proteins as a biological ON/OFF switch. This largely reversible labelling offers the ability to tune and coordinate signalling networks on a timescale faster than is required for protein synthesis and proteolysis. A necessity for life, it is not surprising that many aspects of signalling cascades predate species divergence from unicellular organisms. (Tan, 2011)

While Nature has found a way to amplify the diversity of the pool of amino acid building blocks (in direct analogy to the diversification of DNA messages through arrangement of the nucleotide code), the selection pressures of evolution have still adhered to chemical intuition and energy conservation. In fact, evolution exerts less pressure on similar chemically structured residues. It seems the idea of a structural mimic is, ironically, bio-inspired.

2.2 Post-translational modification offers fast and dynamic response to stimulus

Post-translational modification of proteins offers another layer of complexity in the regulation of cellular processes. It provides a powerful tool for fast and reversible modification of protein structure, which effects dynamic changes in protein function, and ultimately translates into a cellular phenotype. The most easily deciphered systems are those in which there is a large chemical difference between the modified and non-modified group and systems that are controlled by a unimodal 'adding' or 'removing' of an effector moiety, as in the prototypical case of phosphorylation (*i.e.*, addition or removal of a phosphate group with a kinase or phosphatase, respectively). Chemical modification of an amino acid residue on an existing polypeptide is an energy-efficient means of reversibly and quickly altering the charge or steric properties of a peptide to result in a new molecular phenotype.

For example, a mere phosphorylation event induces a large structural change in the bacterial receiver domain of the nitrogen regulatory protein C (NtrC) (Kern et al., 1999) (Fig. 1). Phosphorylation of NtrC at this position Asp_{54} causes the movement of three alpha-helices, resulting in the formation of a hydrophobic patch thought to be necessary for downstream signal transduction.
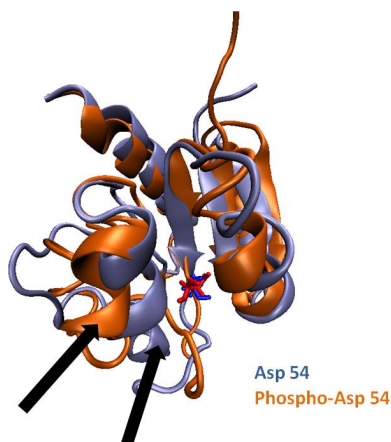
Fig. 1. Phosphorylation induces conformational changes**.** Crystal structure of NtrC before (blue) and after (orange) phosphorylation of Asp_{54}. A single residue modification from Asp_{54} (blue sticks) to the phosphorylated form (red sticks) results in major protein restructuring, including movement of α-helix 4 (black arrows) away from the active site to generate a hydrophobic patch. Structural coordinates were obtained from the Protein Databank PDB code: (1DC7, 1DC8).

To date, over 200 post-translation modifications have been identified; and they have been explored to varying extents, based on a mixture of propensity and technical accessibility. Phosphorylation, although overshadowed in propensity by glycosylation, is by far the most largely studied and characterized modification in the context of a signal relay in cellular systems. This is owed in large part to the availability of tools to efficiently mimic the isostructural property of phosphorylated residues, not the least of which is site-directed mutagenesis and the use of natural amino acids as constitutive substitutes.

2.3 Natural amino acids as substitutes for modified amino acids

Invention of site-directed mutagenesis offered scientists, for the first time, access to the entire amino acid regimen for a given residue location. Chemical similarity between certain post-translationally modified residues and some natural amino acids became central to the field of genetic-based manipulation of protein properties. Shown in Fig. 2 are examples of natural mimics of some well-studied post-translational modifications, highlighting structural similarities.
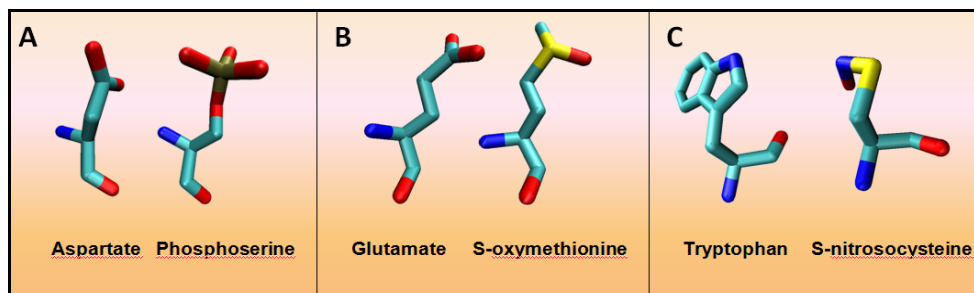

Fig. 2. Isostructural similarities between natural amino acids and post-translational modifications. Structural coordinates for natural amino acids and products of **(A)** phosphorylation **(B)**, sulfoxidation, and **(C)** nitrosylation. Carbon atoms are shown in cyan; oxygen, in red; nitrogen, in blue; and phosphorous in gold. Hydrogen atoms have been omitted for clarity.

While site-directed mutagenesis was first described in 1978 (Hutchison et al., 1978), the first examples of a PTM mimicking appeared in 1988 and 1989; the latter was described four years before the Nobel Prize in Chemistry was awarded to Michael Smith and Kary B. Mullis, for the development of site-directed mutagenesis and the polymerase chain reaction, respectively. When investigating the effect of phosphorylation of the large T antigen portion of Simian Virus 40 (SV40) on DNA-binding and replication, it was found that substitution of a known phosphorylatable Thr_{124} residue by glutamic acid, as opposed to alanine or cysteine, resulted in partial retention of native DNA replication activity. (Schneider et al., 1988) It was suggested that a negative charge, and not a phospho-group *per-se*, was required to promote DNA replication. Shortly thereafter, two papers revealed a remarkable structural and functional similarity between the phosphorylated form of serine and glutamic acid. In agreement with similar structural deviations of phosphorylated serine and glutamic acid, the substitution of the regulatory Ser_{46} for glutamic acid in HPr (Wittekind et al., 1989), a protein component of the bacterial PEP-dependent sugar-transport system, resulted in enzyme inactivation indistinguishable from the native phosphorylation. (Reizer et al., 1989) The crystal structure of unmodified and phospho-Ser_{46} would later be determined (Jia et al, 1994) (Audette et al., 2000), and an image of the electrostatic potential shows the large surface change that accompanies phosphorylation (Figure 3).

This provided the first direct evidence for the structural mimicking of a phosphate group using site-directed mutagenesis. It showed that an easily genetically encoded glutamic acid residue was a strong phosphoserine mimic. The result was an explosion in the usage of point mutations as surrogates for phosphorylated or phospho-defective forms of potential regulatory sites.
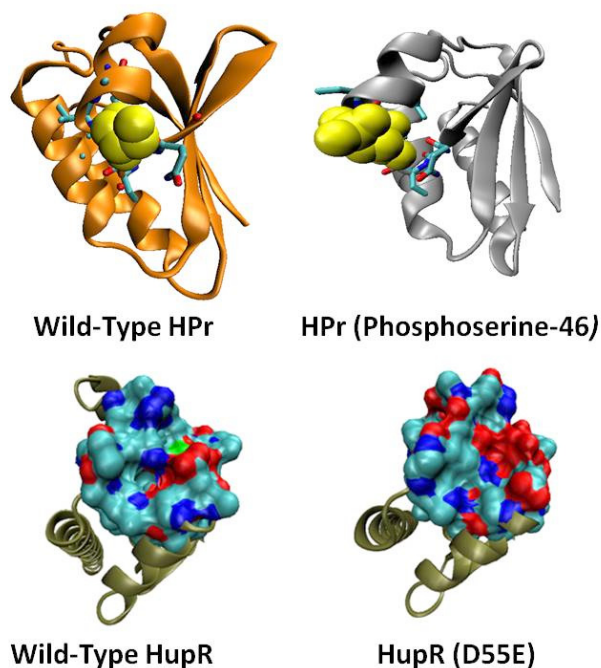
Wild-Type HPr   **HPr (Phosphoserine-46)**

Wild-Type HupR   **HupR (D55E)**

Fig. 3. Phosphorylation induces marked structural and surface-contour changes. (Top) Ribbon diagrams of crystal structures of native and phosophorylated forms of HPr; Ser_{46} is highlighted in yellow, and nearby residues are shown. (Bottom) Surface representation of the wild-type and the phosphomimetic $Asp_{55}E$ variants of Hydrogen Uptake Protein Regulator (*HupR*). A phosphomimetic mutation induced marked surface-contour changes that affected protein dimerization and function. Structural coordinates were obtained from the Protein Databank (PDB code: 1FU0, 1PTF, 2JK1, 2VUH). Highlighted residues and the surface layer proximal to residue 55 in HupR were color-coded by atom type: carbon atoms are cyan, nitrogen atoms are blue, oxygen atoms are red, and phosphorous atoms are green (inhibitor compound).

Another example of how substitution with phospho-mimetic glutamic acid can lead to a conformational change is shown in Fig. 3 (bottom). A surface-contour map of the receiver domain of nitrogen regulatory protein HupR (in which phosphorylation is actually inhibitory to transcription) before and after mutation of Asp_{55} to glutamic acid (Davies et al., 2009) illustrates a marked surface change due to mutation of a single residue, which leads to deactivation of the receiver domain in a manner similar to that observed for authentic phosphorylation.

Phosphorylation is not the only form of post-translation regulation that has been investigated with natural amino acid mimics. Biomimetic point mutations have been fundamental for studying some other means of protein modification including sulfoxidation, S-nitrosylation, tyrosine nitration, acetylation, methylation, etc. Several examples are illustrated in Table 1.

Modification	Protein	Substitution	Readout	Ref
Phosphorylation	CLIP-170	Ser→Ala Ser→Glu	Co-immunoprecipitation, protein kinetics at MT tip	(Lee et al., 2010)
	Inositol 1,4,5-Trisphosphate Receptor (InsP3R)	Ser→Ala Ser→Glu	Ca^{+2} release from ER	(Wagner et al., 2004)
	AMPA (GluA2 subunit)	Ser→Ala Ser→Glu	Conductance	(Kristensen et al., 2011)
	Itch (E3-Ubiquitin Ligase)	Ser→Ala Ser→Glu Thr→Ala Thr→Glu	Co-immunoprecipitation /Proteolysis/ Ubiquitinylation activity	(Gallagher et al., 2006)
Sulfinic acid	Peroxiredoxin	Cys→Glu	Crystallography	(Jonsson et al., 2009)
Sulfoxidation	Slo-1	Met→Leu Met→Glu	Ionic current (Inside-out patch clamp)	(Santarelli et al., 2006)
Acetylation	Heat shock protein- (HSP-90)	Lys→Arg Lys→Gln Cys→Ala	Chaperon activity	(Scroggins et al., 2007)
S-nitrosylation	Syntaxin-1	Cys→Ser Cys→Trp Cys→Met	Exocytosis (amperometry)	(Palmer et al., 2008)
	Heat shock protein- (HSP-90)	Cys→Ala Cys→Asn Cys→Asp	Protein dimerization (FRET) ATP-ase activity Reverse Transcriptase activity	(Retzlaff et al., 2009)
Methylation	Flap endonuclease 1 (FEN1)	Arg→Ala Arg→Phe Arg→Lys	Endonuclease activity	(Guo et al., 2010)

Table 1. Representative post-translational modifications mimicked by natural point mutations using site-directed mutagenesis.

Amino acid oxidation is a common mechanism for diversification of certain residues; and, like phosphorylation, there are both addition and removal effector proteins associated with the modification. The reactive sulfur group of the methionine residue can be oxidatively modified with a number of reagents to methionine sulfoxide or methionine sulfone, dramatically increasing the polarity and hydrophobicity of the residue. (Black and Mould, 1991) (Hoshi and Heinemann, 2001) A family of enzymes known as methionine sulfoxide reductases have been found competent to reduce the methionine sulfoxide back to a thioether, a necessary requirement for a signalling relay. Substitution of a leucine residue for methionine offers a suitably isostructural null mimic, removing the oxidizeable sulphur atom while maintaining similar geometric and hydrophobic characteristics. (Ciorba et al., 1997) (Chen et al., 2000) In one example, this modification, which adds electron density onto the long methionine residue, has been mimicked *in vitro* by glutamic acid substitution. Interestingly, it was recently found that calcium and voltage gated BK (Big Potassium) ion channels, responsible for regulating K^+ ion flux, are sensitive to oxidation. This has been accomplished through lining of ion-selective pore with the hydrophobic Met_{536}, Met_{712}, and Met_{739} residues, which can be oxidatively modified to the more polar methionine sulfoxide, offering tuning of the hydrophobicity of channel. Oxidation of these methionine residues results in a decrease in half-activation voltage, increasing channel activity. Interestingly, one must mutate all three methionine residues

within the channel to leucine to render the channel insensitive to oxidant treatment. If, however, one replaces any one of these three residues with glutamic acid, the channel behaves just like it does in the wild-type protein after treatment with the oxidant. (Santarelli et al., 2006) Thus, the technique of site-directed mutagenesis revealed that any of the three separate methionine residues can confer the oxidant-sensitivity phenotype to the protein, highlighting a robust selection pressure for that function.

Oxidized forms of cysteine, including nitrosyl, sulfenic acid, and sulfinic acid variants, have been implicated in reversible signalling arrays (Reddie and Carroll, 2008), and some have been mimicked with the use of natural amino acids. For example, cysteine sulfinic acid differs from aspartic acid by a single-atom replacement of sulphur atom with a carbon atom. This phosphomimicking strategy has been used in the investigation of the mechanism of peroxiredoxin (Pxr) (Jonsson et al., 2009), an enzyme involved in the ATP-dependent reduction of cysteine sulfinic acid back to a thiol residue. Reduction of an active site cysteine-sulfinic acid is first accomplished through ATP-dependent phospho-deficient of the sulfinic acid residue, followed by attack of a neighboring cysteine from sulfiredoxin (Srx) enzyme on the sulfinic phosphory ester intermediate. Remarkably, replacement of the catalytic cysteine in Pxr with an aspartic acid residue results in phosphorylation of the aspartic acid, but does not allow subsequent reactivity with the docked Srx. Thus, the single-atom replacement offered by native site-directed mutagenesis allowed for detailed mechanistic study of a trapped intermediate complex, inaccessible with other techniques. Substitution of a cysteine residue with a serine residue offers an effective single-atom replacement of a sulphur atom with an oxygen atom and eliminates sulphur-based chemistry from the site, while minimizing steric differences.

In several cases, tryptophan was reported to be an effective mimic of nitroso-cysteine (Palmer et al., 2008) (Wang et al., 2009) as computer simulations of the cysteine to tryptophan mutant suggest it contains steric bulk not offered by the next closest steric match, methionine, despite the absence of a sulphur atom in the former. In another case, nitroso-cysteine modification was mimicked by substitution with asparagine in studies investigating the effect of nitrosylation on ATPase activity of the 90 kDa heat shock protein (γHSP90). (Retzlaff et al., 2009) (Scroggins and Neckers, 2009) In this study an NO-insensitive γHSP90 was rendered sensitive to inactivation by NO treatment due to replacement of an alanine residue in the C-terminal domain with a cysteine, as this corresponding site is known to be a cysteine in all other HSP variants. Mutation of the engineered site to asparagine (γHSP90 A577N) recapitulates NO treatment of the engineered cysteine mutant, in contrast to a control isoleucine mutant (γHSP90 A577I), which serves as an unmodifiable mimick (shown in Fig 4). Although no citation or rationale was used for the mutant choices, these mimicking residues indeed recapitulated ATPase activity and chaperone activity of the NO-free and NO-bound forms of engineered γHSP90 protein.

As virtually every intracellular process is under dynamic regulation, it is no surprise that PTM-mimicking mutations have been used in conjunction with a myriad of analytical techniques. Shown in Table 1 are some of examples relevant to native site-directed mutagenesis - giving a flavour of the breadth of the strategy. In a manuscript that would bring the strategies of the last nearly 35 years full circle, it has recently been suggested that throughout evolution, phosphoserine residues on proteins have been mutated, through genetic drift. In one study they were more frequently converted into glutamic acid and
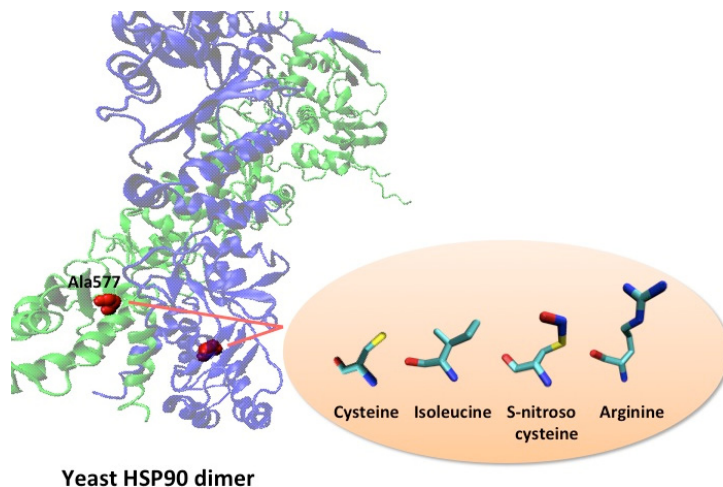
Yeast HSP90 dimer

Fig. 4. Mimicking S-nitrosylation in an engineered regulatory site. Ribbon representation of X-ray crystal structure of the Yeast HSP90 dimer; native Ala 577 (left) and engineered variants at position 577 (right) are highlighted in red. Note that there are structural similarities between cysteine and the non-modifiable isoleucine and between S-nitrosocysteine and arginine. Structural coordinates were obtained from the Protein Databank (PDB code: 2CG9) Stick-models of residues were color-coded by atom type: carbon atoms are cyan, nitrogen atoms are blue, and oxygen atoms are red.

aspartic acid than into any other residue, demonstrating evolution away from the switchable into permanent ON or OFF functions. (Kurmangaliyev et al., 2011) Although this work focused on non-structured regions and solely on phosphoserine evolution (the only post-translational modification for which there were sufficient data sets for statistical interpretation), these results might be viewed as an acknowledgement from biology of a mere requirement for a well-placed charge to induce the biological phenotype for which the switch evolved.

2.4 Site-directed mutagenesis in animal models

The ultimate verification of the experimental models obtained in *in vitro* assays or even in cultured cells comes with experiments in living organisms. Site-directed mutagenesis, in combination with other technologies for delivering a tailored plasmid to more complex genomes, has allowed the knock-in of mutations for individual engineered proteins in entire organisms, including mammals. (Yang et al., 1997) This technology has been used to investigate the role of individual amino acids in many biological processes and disease states, such as ASD (Autism Spectrum Disorder) (Bader et al., 2011), malignant hyperthermia (associated with heat stroke susceptibility) (Durham et al., 2008), and depression (Talbot et al., 2010), in the context of a living mammals. Phosphomimetic and phospho-deficient mutations have been knocked-in as well, leading to new understanding of downstream events to protein modification in the context of a living organism. In one prominent example, phosphorylation of ribosomal protein S6 was linked to control of cell size and glucose homeostasis in certain cell types through the use of a phospho-deficient

knock-in mouse model. (Ruvinsky et al., 2005) In this model, all five potentially phosphorylatable serine residues in the putative region of rS6 were mutated to alanine (rpS6$^{P-/-}$) to abolish the potential for their phosphorylation. Smaller cell size, increased protein synthesis, and a severe decrease of circulating insulin were among the phenotypic signatures of the phospho-deficient variant. The phenotype was later shown to include decreased muscle mass and energy availability.

In another salient example, both phosphomimetic and phospho-deficient mutations were generated in a transgenic mouse model of Huntington's disease - a neurodegenerative disorder characterized by motor impairments and various psychiatric symptoms. (Gu et al., 2009) In this model, two phosphorylatable serine residues (13 and 18) in the N-terminus of a disease-promoting mutant of the huntingtin protein (*mhtt*) were mutated to either glutamate (phosphomimetic) or alanine (phospho-deficient) residues to investigate effect of phosphorylation on modification of the *htt* protein with a polyglutamine (PolyQ) extension, and on the onset of Huntington's Disease. Both mutants were viable and successfully rescued the embryonic lethality of an *htt* knockout mutant. A remarkable difference was observed between the Ala and Glu variants. The mice that were phosphodeficient (Ala) in the two serine residues showed onset of disease pathogenesis from the mutant *mhtt* - as measured by neuropathological examination of brain tissue sections, as well as by various motor and psychological tests. Remarkably, the phosphomimetic mutant mice (Glu) showed resistance to *mhtt*-induced disease pathogenesis. This technique clearly demonstrated that loss–of-phosphorylation of these two serine sites is a key signature of the pathogenesis *in vivo*. These results further suggested that the N-terminus of *htt* may be a target for future drug design to treat disease, even after the initializing mutation is present.

Indeed these examples show the power of site-directed mutagenesis in studying the signaling events associated with the disease state. However, caution must always be taken in the interpretation of data obtained from studies using biomimetic constructs.

2.5 Limitations of site-directed mutagenesis for targeting signaling nodes

Despite vast precedent for the utilization of amino acid surrogates and mimics of post-translational modifications in the literature, it must be remembered that the data and interpretation are only as robust as the chemical match of the surrogate moiety to the target. For example, the carboxylic acid functionality of glutamic acid has differing fundamental chemical properties from a serine-phosphate group, including, but not limited to, access to a (-2) charge state (in the case of phosphate) and different acid dissociation constants. Despite this, the single mutation has proven a valuable tool in the mimicking of phosphorylated residues. In addition it has become a common and quite inexpensive laboratory procedure, which is qualified as a 'quick-check' protocol to produce constitutively active, inactive or null mutants of potential phosphorylation sites.

As mentioned, PTMs that result in a significant size or charge difference from the unmodified states have been among the earliest developed. For example, while phosphorylation is the most thoroughly studied form of modification, it has been predicted that protein methylation will be among the slowest developed signalling pathways. This is due to the subtle spatial and charge density differences between methylated and non-methylated forms of an amino acid residue. With mass spectrometry virtually alone in the struggle to address such modifications, the decoding of such signalling networks awaits the

development of sufficient chemical tools to recognize such subtle nodes and expedite interpretation and clarification of mass spectral data. (Huttlin et al., 2010)

It is probable that certain interactions, or classes of interactions, may be more stringent in which the electronic features of post-translationally modified residues are recognized and interpreted as signals. The importance of a cautious eye when utilizing signalling mimics was recently made salient when a phosphometic serine to glutamic acid mutant recapitulated one aspect of a signalling cascade, but not another. (Paleologou et al., 2010) While investigating potential phosphorylation sites in α-synuclein - a major component of the Alzheimer's disease amyloid - the authors generated serine to glutamate and serine to alanine mutants and compared them to wild-type and authentic phosphorylated material in ability to localize at the membrane and participate in fibril formation. In a model wherein phosphorylation of Ser_{87} blocks membrane recruitment and fibril formation, it was shown that substitution of native Ser_{87} with alanine resulted in wild-type-like membrane association and fibril formation, while authentic phospho-Ser_{87}, along with the $Ser_{87}E$ mutant, showed blunted formation of fibrils and poor membrane recruitment. Circular dichroism studies showed that, in a micellar environment, the authentic phospho-Ser_{87} modification resulted in an unstructuring of the random coiled protein, whereas the phosphomimic mutant could not recapitulate the membrane-associate phenotype. This unstructuring has been shown to be important in the interaction with membrane contents, and may further alter protein specificity. Thus, the phosphomimetic mutation introduced with site-directed mutagenesis could only reproduce one aspect of the biological phenotype. In this case, the authors were able to view the limits of the phosphomimic approach because they had access to constructs containing the authentic phosphoprotein for comparison. Nevertheless, it serves as an instructive case in which it is acknowledged that a given PTM can have multiple roles or be required in multiple steps of a signalling cascade or pathway.

3. Conclusion

We have demonstrated that site-directed mutagenesis has played an indispensable role in the unravelling of signal transduction pathways by offering access to constitutively active or inactive forms of several common PTMs. A critical eye must be used when utilizing surrogate residues that lack the capacity for a switchable context, and it is important to understand the limits of any biomimetic approach before results can be properly interpreted. Even with these limitations, it is no wonder the technique of site-directed mutagenesis has penetrated the field of signal transduction since the very conception of the technology. Indeed the utility of site-directed mutagenesis for the development and verification of models for biological processes and disease states will remain beneficial for years to come.

4. Acknowledgment

5. References

Audette G.F., Engelmann R., Hengstenberg W., Deutscher J., Hayakawa K., Quail J.W., Delbaere L.T.J., 2000. The 1.9 Å Resolution Structure of Phospho-Serine 46 HPr from Enterococcus Faecalis. J. Mol. Biol. 303, 545-553.

Bader P.L., Faizi M., Kim L.H., Owen S.F., Tadross M.R., Alfa R.W., Bett G.C.L., Tsien R.W., Rasmusson R.L., Shamloo M. 2011. Mouse model of Timothy syndrome recapitulates triad of autistic traits. Proc. Natl Acad. Sci. USA 108, 15432-15437

Black S.D., Mould D.R., 1991. Development of Hydrophobicity Parameters to Analyze Proteins Which Bear Post- or Cotranslational Modifications. Anal. Biochem. 193, 72-82.

Chen J., Avdonin V., Ciorba M.A., Heinemann S.H., Hoshi T., 2000. Acceleration of P/C-Type Inactivation in Voltage-Gated K+ Channels by Methionine Oxidation. Biophys. J. 78, 174-187.

Ciorba M.A., Heinemann S.H., Weissbach H., Brot N., Hoshi T., 1997. Modulation of Potassium Channel Function by Methionine Oxidation and Reduction. Proc. Natl Acad. Sci. USA 94, 9932-9937.

Davies K.M., Lowe E.D., Vénien-Bryan C., Johnson L.N., 2009. The HupR Receiver Domain Crystal Structure in its Nonphospho and Inhibitory Phospho States. J. Mol. Biol. 385, 51-64.

Deribe Y.L., Pawson T., Dikic I., 2010. Post-Translational Modifications in Signal Integration. Nat. Struct. Mol. Biol. 17, 666-672.

Durham W.J., Aracena-Parks P., Long C., Rossi A.E., Goonasekera S.A., Boncompagni S., Galvan D.L., Gilman C.P., Baker M.R., Shirokova N., Protasi F., Dirksen R., Hamilton S.L. 2008. RyR1 S-Nitrosylation Underlies Environmental Heat Stroke and Sudden Death in Y522S RyR1 Knockin Mice. Cell 133, 53-65.

Gallagher E., Gao M., Liu Y.-C., Karin M., 2006. Activation of the E3 Ubiquitin Ligase Itch Through a Phosphorylation-Induced Conformational Change. Proc. Natl Acad. Sci. USA 103, 1717-1722.

Guo Z., Zheng L., Xu H., Dai H., Zhou M., Pascua M.R., Chen Q.M., Shen B., 2010. Methylation of FEN1 Suppresses Nearby Phosphorylation and Facilitates PCNA Binding. Nat. Chem. Biol. 6, 766-773.

Hoshi T., Heinemann S.H., 2001. Regulation of Cell Function by Methionine Oxidation and Reduction. J. Physiol. 531, 1-11.

Hutchison C.A., Phillips S., Edgell M.H., Gillam S., Jahnke P., Smith M., 1978. Mutagenesis at a Specific Position in a DNA Sequence. J. Biol. Chem. 253, 6551-6560.

Huttlin E.L., Jedrychowski M.P., Elias J.E., Goswami T., Rad R., Beausoleil S.A., Villén J., Haas W., Sowa M.E., Gygi S.P., 2010. A Tissue-Specific Atlas of Mouse Protein Phosphorylation and Expression. Cell 143, 1174-1189.

Jia Z., Vandonselaar M., Hengstenberg W., Wilson Quail J., Delbaere L.T.J., 1994. The 1.6 Å Structure of histidine-Containing Phosphotransfer Protein HPr from Streptococcus Faecalis. J. Mol. Biol. 236, 1341-1355.

Jonsson T.J., Johnson L.C., Lowther W.T., 2009. Protein Engineering of the Quaternary Sulfiredoxin-Peroxiredoxin Enzyme-Substrate Complex Reveals the Molecular Basis for Cysteine Sulfinic Acid Phosphorylation. J. Biol. Chem. 284, 33305-33310.

Kern D., Volkman B.F., Luginbuhl P., Nohaile M.J., Kustu S., Wemmer D.E., 1999. Structure of a Transiently Phosphorylated Switch in Bacterial Signal Transduction. Nature 402, 894-898.

Kristensen A.S., Jenkins M.A., Banke T.G., Schousboe A., Makino Y., Johnson R.C., Huganir R., Traynelis S.F., 2011. Mechanism of Ca2+/Calmodulin-Dependent Kinase II Regulation of AMPA Receptor Gating. Nature 14, 727-735.

Kurmangaliyev Y., Goland A., Gelfand M., 2011. Evolutionary Patterns of Phosphorylated Serines. Biol. Direct 6, 8.

Lee H-S., Komarova Y.A., Nadezhdina E.S., Anjum R., Peloquin J.G., Schober J.M., Danciu O., van Haren J., Galjart N., Gygi S.P., Akhmanova A, Borisy G.G., 2010.

Phosphorylation Controls Autoinhibition of Cytoplasmic Linker Protein-170. Mol. Biol. Cell 21, 2661-2673.

Paleologou K.E., Oueslati A., Shakked G., Rospigliosi C.C., Kim H.-Y., Lamberto G.R., Fernandez C.O., Schmid A., Chegini F., Gai W.P., Chiappe D., Moniatte M., Schneider B.L., Aebischer P., Eliezer D., Zweckstetter M., Masliah E., Lashuel H.A., 2010. Phosphorylation at S87 Is Enhanced in Synucleinopathies, Inhibits a-Synuclein Oligomerization, and Influences Synuclein-Membrane Interactions. J. Neurosci. 30, 3184-3198.

Palmer Z.J., Duncan R.R., Johnson J.R., Lian L.-Y., Mello L.V., Booth D., Barclay J.W., Graham M.E., Burgoyne R.D., Prior I.A., Morgan A., 2008. S-Nitrosylation of Syntaxin 1 at Cys145 is a Regulatory Switch Controlling Munc18-1 Binding. Biochem. J. 413, 479-491.

Reddie K.G., Carroll K.S., 2008. Expanding the Functional Diversity of Proteins Through Cysteine Oxidation. Curr. Opin. Chem. Biol. 12, 746-754.

Reizer J., Sutrina S.L., Saier M.H., Stewart G.C., Peterkofsky A., Reddy P., 1989. Mechanistic and Physiological Consequences of HPr(ser) Phosphorylation on the Activities of the Phosphoenolpyruvate:Sugar Phosphotransferase System in Gram-Positive Bacteria: Studies with Site-Specific Mutants of HPr. EMBO J. 8, 2111-2120.

Retzlaff M., Stahl M., Eberl H.C., Lagleder S., Beck J., Kessler H., Buchner J., 2009. Hsp90 is Regulated by a Switch Point in the C-terminal Domain. EMBO Rep. 10, 1147-1153.

Santarelli L.C., Wassef R., Heinemann S.H., Hoshi T., 2006. Three Methionine Residues Located Within the Regulator of Conductance for K+ (RCK) Domains Confer Oxidative Sensitivity to Large-Conductance Ca2+-Activated K+ Channels. J. Physiol. 571, 329-348.

Schneider J., Fanning E., 1988. Mutations in the Phosphorylation Sites of Simian Virus 40 (SV40) T Antigen Alter its Origin DNA-Binding Specificity for Sites I or II and Affect SV40 DNA Replication Activity. J. Virol. 62, 1598-1605.

Scroggins B.T., Neckers L., 2009. Just Say NO: Nitric Oxide Regulation of Hsp90. EMBO Rep. 10, 1093-1094.

Scroggins B.T., Robzyk K., Wang D., Marcu M.G., Tsutsumi S., Beebe K., Cotter R.J., Felts S., Toft D., Karnitz L., Rosen N., Neckers L., 2007. An Acetylation Site in the Middle Domain of Hsp90 Regulates Chaperone Function. Mol. Cell 25, 151-159.

Tan C.S.H., 2011. Sequence, Structure, and Network Evolution of Protein Phosphorylation. Sci. Signal. 4, 6.

Wagner L.E., Li W.-H., Joseph S.K., Yule D.I., 2004. Functional Consequences of Phosphomimetic Mutations at Key cAMP-Dependent Protein Kinase Phosphorylation Sites in the Type 1 Inositol 1,4,5-Trisphosphate Receptor. J. Biol. Chem. 279, 46242-46252.

Walsh C.T., Garneau-Tsodikova S., Gatto G.J., 2005. Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications. Angew. Chem. Int. Edit. 44, 7342-7372.

Wang P., Liu G.-H., Wu K., Qu J., Huang B., Zhang X., Zhou X., Gerace L., Chen C., 2009. Repression of Classical Nuclear Export by S-nitrosylation of CRM1. J. Cell Sci. 122, 3772-3779.

Wittekind M., Reizer J., Deutscher J., Saier M.H., Klevit R.E., 1989. Common Structural Changes Accompany the Functional Inactivation of HPr by Seryl Phosphorylation or by Serine to Aspartate Substitution. Biochemistry 28, 9908-9912.

Yang X.W., Model P., Heintz N., 1997. Homologous recombination based modification in Escherichia coli and germline transmission in transgenic mice of a bacterial artificial chromosome. Nat Biotechnol. 15(9):859-65.

Using Genetic Reporters to Assess Stability and Mutation of the Yeast Mitochondrial Genome

Shona A. Mookerjee[1] and Elaine A. Sia[2]
[1]The Buck Institute for Research on Aging
[2]Department of Biology, The University of Rochester
USA

1. Introduction

The mitochondrion has been an identified subcellular organelle since the late 1800s, though its function and importance remained relatively cryptic until almost a century later. With the molecular renaissance of the 1950s and 1960s came the tools and perspectives, including peptide sequencing, microscopy, and the chemiosmotic hypothesis, with which to formulate and test crucial questions about what mitochondria do and how they function. In addition, shortly after the structure of DNA was elucidated and its role as the genetic material of the cell established, the presence of extranuclear DNA was revealed by electron microscopy (Nass and Nass, 1963) and by density gradient fractionation (Haslbrunner et al., 1964), identifying this DNA as part of highly purified mitochondria.

The goal of targeted genetic manipulation within the mitochondrial genome is rapidly driving a second renaissance in mitochondrial biology. Current approaches to the mitochondrial uptake of DNA have recently been reviewed in detail (Mileshina et al., 2011a). Several model organisms are currently amenable to this technique although no vertebrate species are among them. In our view, successful mitochondrial transformation is hindered by our limited knowledge of the fundamental processes of mitochondrial DNA maintenance and repair. We are hopeful that a greater understanding of the replication, repair, and maintenance of the mitochondrial genome, afforded by the tools currently in existence and presented here, will soon allow the construction of mammalian mitochondrial disease models (Dunn et al., 2011), or lead directly to gene therapy for the treatment of mitochondrial diseases in patients (Schon and Gilkerson, 2010). In this review, we will discuss the current state of mitochondrial genetic reporters and their application toward understanding and manipulating mitochondrial genome maintenance.

In cells, mitochondrial DNA is organized into protein-associated structures called nucleoids. Quantitative PCR coupled with immunofluorescence microscopy revealed an estimated 2-8 mitochondrial genomes per nucleoid in human immortalized cell culture (Legros et al., 2004); higher-resolution microscopy identified more nucleoids per cell, bringing this estimate down to only 1-2 copies per nucleoid (Kukat et al., 2011). A growing number of the proteins associated with the mitochondrial nucleoid have been identified from eukaryotic model systems (Bogenhagen et al., 2003; Garrido et al., 2003; Kienhöfer et al., 2009). The conserved and abundant HMG protein, Abf2p in yeast, and mtTFA (or TFAM) in

vertebrates, is thought to organize and compact the mitochondrial genome, and is proposed to play a histone-like role in mitochondrial DNA organization (Kaufman et al., 2007; Pohjoismaki et al., 2006). In addition to these properties, TFAM has recently been reported to contribute to mitochondrial DNA replication (Pohjoismaki et al., 2006) and repair (Canugovi et al., 2010).

The mitochondrial genome is vitally important to the organelle's proper function. Its encoded proteins are almost all subunits of the mitochondrial respiratory complexes; mutations to these genes can disrupt bioenergetic function and control. There is also growing evidence that proper maintenance of the mitochondrial genome is tightly associated with normal mitochondrial behavior, including fusion, fission, and intracellular migration (Baker and Haynes, 2011; Chen et al., 2010; Gilkerson, 2009). The reasons for this are currently unknown, but we can speculate that nucleoids may help to define a minimal mitochondrial "unit" and that disruption of mitochondrial DNA resolution affects mitochondrial dynamics and transmission in dividing cells (Margineantu et al., 2002). Alternately, nucleoids may help to catalyze organization of cristae, respiratory complexes, or other submitochondrial structures, as has been suggested for the Complex V ATP synthase (Strauss et al., 2008), and faulty mitochondrial DNA maintenance disrupts this patterning.

Fission and fusion, coupled with autophagic degradation of mitochondrial material ("mitophagy"), are increasingly appreciated as a primary means of mitochondrial quality control. Instability of the mitochondrial genome is therefore not just a problem affecting its encoded products, but potentially the structural integrity of the entire organelle. Recent proposals suggest that the cellular dysfunction underlying neurodegeneration in Alzheimer's and Parkinson's diseases is caused or at least exacerbated by defective mitophagy and increased mitochondrial DNA instability (Chang, 2000; Corral-Debrinski et al., 1994; Coskun et al., 2011; Narendra et al., 2010; Narendra et al., 2008; Sasaki et al., 1998; Suen et al., 2010).

1.1 Yeast gene nomenclature

A comprehensive guide to yeast gene nomenclature is both published (*Trends in Genetics*, *Volume 14, Issue 11, Supplement 1, 1998, Pages S.10-S.11)* and available online (http://www.yeastgenome.org/help/yeastGeneNomenclature.shtml). We will summarize here the points that are most relevant to the following sections.

There are two nomenclature systems, reflecting the advent of genomic analysis. The ORF naming system, instituted in the post-genomic era, gives each predicted open reading frame a systematic designation reflecting its chromosomal position and strand orientation. This system will not be discussed here. We will focus on gene symbol nomenclature, which applies only to ORFs that express a known gene product and is in common use for most proteins.

In gene symbol nomenclature, yeast genes are given a three-letter "name" followed by a number. Gene names are italicized, with dominant (usually wild-type) alleles in capital and recessive (usually mutant) alleles in lowercase type. For example, *ABC1* denotes a wild-type, dominant gene; *abc1* a generic mutant. The names themselves correspond to characterized features, either phenotypic, biochemical, enzymatic, or genetic. Mutant alleles can carry a second number to specify a particular mutation, e.g., *abc1-1,* or a descriptive name. For

example, *abc1-G43V* describes a point mutant that bears a glycine-to-valine amino acid substitution at position 43. Gene knockouts are denoted with a delta symbol, e.g., *abc1-Δ*. Gene disruptions by insertion of another gene are denoted with a double colon, e.g., *abc1::URA3*. If a gene is fully replaced with the insertion, this is indicated as *abc1Δ::URA3*. In yeast, gene products may also carry the gene name, e.g., Abc1p or abc1p-G43V. Gene product names are interchangeable with a functional name, for example, Arg8p = acetylornithone transaminase.

Mitochondrial dysfunction can occur through single gene mutations in both nuclear and mitochondrial genomes, forming two classes of mutants. *Pet* mutants are those with nuclear DNA mutations, while *mit* mutants are those with mitochondrial DNA mutations. Additionally, mitochondrial genome status is also categorized. Respiring yeast strains carrying wild-type mitochondrial DNA are referred to as ρ^+. When yeast cells are grown on a fermentable carbon source such as glucose, spontaneous variants arise that cannot respire and form smaller colonies, termed "*petite*" mutants. The vast majority of spontaneous *petite* strains contain large-scale mitochondrial DNA deletions and rearrangements. Often these strains carry only a small fraction of their original mitochondrial DNA; retention of only 1% of the genome is commonly observed. The total mitochondrial DNA content of ρ^- and ρ^+ strains are equivalent, however, as the remaining fragments in ρ^- strains are amplified accordingly (Dujon, 1981).

Generation of mitochondrial DNA-free derivatives of any strain is achieved by culturing cells with ethidium bromide, which blocks mitochondrial DNA replication without affecting nuclear DNA (Meyer and Simpson, 1969). These ρ^0 strains can be studied directly or used as a tool for mitochondrial DNA manipulation.

1.2 The presence of mitochondrial DNA repair mechanisms

The assertion that mitochondrial DNA did not undergo repair was made as late as 1990 (Singh and Maniccia-Bozzo, 1990), in agreement with early reports (Clayton et al., 1974). Part of the appeal of this idea was the observation that mitochondrial DNA depletion can be induced by various insults, including oxidative stress (Shokolenko et al., 2009), ethanol (Ibeas and Jimenez, 1997) and zidovudine (AZT) (Arnaudo et al., 1991) treatment. This was interpreted as evidence that, when damaged, mitochondrial DNA was simply eliminated rather than repaired. However, in 1992, the first evidence for photolyase repair of UV-induced mitochondrial DNA damage in yeast was provided (Yasui et al., 1992), followed quickly by work from the Bohr and Campbell groups demonstrating uncharacterized repair activities and homologous recombination, respectively, in two mammalian mitochondrial systems (LeDoux et al., 1992; Thyagarajan et al., 1996). The fifteen years since have revealed an extensive set of mitochondrial DNA repair pathways, including base excision repair (BER), homologous recombination (HR), and non-homologous end joining (NHEJ). A recent review summarizes our current understanding of these pathways in mitochondria and other organellar genomes (Boesch et al., 2010).

1.3 Nuclear and mitochondrial DNA repair pathways share protein components

Many of the known mitochondrial DNA repair pathway proteins are mitochondrially-localized proteins initially characterized in nuclear repair. The first mitochondrial DNA repair protein identified, photolyase, was demonstrated to be one such dual-localized protein (Green and MacQuillan, 1980). Subsequent studies indicated localization of base

excision repair proteins to both subcellular compartments, including the yeast glycosylases Ntg1p (You et al., 1999), Ung1p (Chatterjee and Singh, 2001) and Ogg1p (Singh et al., 2001), the mammalian glycosylases UNG1 (Nilsen et al., 1997), MTH1 (Kang et al., 1995), OGG1 (Nishioka et al., 1999), and MYH (De Souza-Pinto et al., 2009; Nakabeppu et al., 2006), and the yeast AP endonuclease Apn1p (Ramotar et al., 1993). In human lymphoblasts, BER proteins were associated with the mitochondrial inner membrane fraction, where mitochondrial nucleoids are also found (Stuart and Brown, 2006). These findings illustrate the high evolutionary conservation of mitochondrial BER. Factors that regulate the subcellular localization of these proteins are not well understood; however, changes to localization in response to stress has recently been demonstrated (Griffiths et al., 2009; Swartzlander et al., 2010). This apparent recruitment of DNA repair proteins to the mitochondria may represent a DNA-specific communication pathway between the intramitochondrial and extramitochondrial environments.

Other DNA repair proteins have been shown to affect mitochondrial DNA maintenance in mammalian cells, including the BER flap endonuclease FEN1 (Liu et al., 2008), DNA double-strand break repair proteins, Rad51p (Sage et al., 2010), Mre11 (Dmitrieva et al., 2011) and Ku80 (Coffey et al., 1999), and the nucleotide excision repair protein CSA (Kamenisch et al., 2010). In addition, in yeast, DNA damage tolerance pathways that utilize the translesion polymerase complexes encoded by Rev1p, Rev3p, and Rev7p also impact mitochondrial mutagenesis (Kalifa and Sia, 2007; Zhang et al., 2006).

2. Manipulation of the yeast mitochondrial genome

2.1 Basic features of the mitochondrial genome

The yeast mitochondrial genome consists of 75-85 kb of double-stranded DNA, encoding seven protein products (Cox I, II, III, Atpase 6, 9, cyt b, Var1), 2 rRNAs and 24 tRNAs, while the human mitochondrial genome is a much smaller 16.5 kb and encodes 13 protein products (cox I, II, III, ND1-6, 4L, Atpase 6, 8, cyt b), 2 rRNAs and 22 tRNAs. Aside from the presence of complex I (NADH:ubiquinone oxidoreductase) subunit genes in the human mitochondrial genome and not yeast, and the presence of non-coding regions in the yeast genome, the two are remarkably similar in structure and encoded products, giving yeast mitochondrial genome manipulation great power to inform our understanding of mammalian mitochondrial DNA defects.

Mitochondrial and nuclear DNA in yeast are compositionally different; mitochondrial DNA is relatively AT-rich and highly repetitive, with G and C bases further segregated in coding regions. This repetition made initial sequencing of the entire yeast mitochondrial genome difficult (Foury et al., 1998) and is a continued challenge in targeted gene manipulation, particularly in the intergenic AT-rich regions. Yeast mitochondrial DNA has multiple regions of non-coding DNA, which are the primary contributors to the 83% AT bias of the genome. The size difference between human and yeast mitchondrial DNA is almost entirely due to the absence of these intergenic regions in the human mitochondrial genome.

2.2 Organisms with tractable mitochondrial genomes

To date, the number of organisms that have successfully undergone mitochondrial transformation remains small and is restricted to unicellular eukaryotes. The most widely used model remains the budding yeast *Saccharomyces cerevisiae* (Johnston et al., 1988), which

presents multiple biological advantages beyond the accessibility of its mitochondrial genome. Budding yeast are facultative anaerobes. This organism can survive by fermentation when oxygen is unavailable or when respiratory mechanisms are disrupted, as occurs during biolistic transformation (Section 2.3). Moreover, our comprehensive understanding and molecular tool base for the nuclear genome enables efficient analysis and powerful interpretation of mitochondrial phenotypes. Finally, homologous recombination in yeast is highly efficient in both nuclear and mitochondrial genomes, facilitating the introduction of targeted sequences to either genome.

The mitochondria of hyphal yeast *Candida glabrata* have also recently been reported to take up biolistically delivered DNA. Unlike wild-type *S. cerevisiae*, *C. glabrata* can maintain relatively stable heteroplasmy, maintaining mixed populations of transformed and endogenous DNA, which the authors propose as an ideal characteristic for studying the regulation of mitochondrial DNA transmission. Under selective pressure the exogenous genotype can be fixed to homoplasmy (Zhou et al., 2010). Heteroplasmy is the state of harboring multiple different mitochondrial genomes, while in homoplasmy only one type is present.

An important requirement of successful mitochondrial manipulation is the ability to select and purify a rare transformation event; in *S. cerevisiae*, the efficiency of transformation using microprojectile bombardment is on the order of one in 10^7 (Bonnefoy and Fox, 2007). As described in Section 2.3, this is the sole method currently known for successful mitochondrial incorporation of exogenous DNA into whole cells.

The only known algal species to date that can be biolistically transformed is the green alga *Chlamydomonas reinhardtii*. This species can incorporate exogenous DNA into both mitochondrial and chloroplast genomes, providing a unique model for studying the genetic and functional interactions of these two organelles (Randolph-Anderson et al., 1993). Here, the development of a selectable mutant facilitated the initial isolation of mitochondrial transformants. *C. reinhardtii* are normally able to grow in the dark if supplemented with acetate as a carbon source; due to a deletion disruption of the mitochondrial CYB gene encoding apocytochrome b, the *dum-1* mutant cannot. Transforming *dum-1* cells with mitochondrial DNA isolated from a *DUM-1* strain and growing cells in darkness plus acetate selects for restoration of a growth phenotype, indicating mitochondrial uptake of the wild-type CYB gene.

Biolistic transformation in multicellular organisms is hampered by an inability to select and amplify individual transformants. A promising recent attempt to transform mitochondria in a mouse embryonic fibroblast line with a "universal" neomycin marker via a bacterial conjugation-like mechanism was not successful in generating mitochondrial transformants (Yoon and Koob, 2011). However, mitochondria isolated from both mammalian and plant sources have been successfully transformed *in vitro* (Mileshina et al., 2011b; Yoon, 2005). Multiple methods, including DNA targeting with a protein localization signal, electroporation, and spontaneous mitochondrial uptake of linear DNA have all given rise to mitochondrial uptake of exogenous DNA, but these constructs could not be propagated in the mitochondrial genome of a viable and dividing cell (Mileshina et al., 2011a).

2.3 Biolistic transformation of yeast and selection of transformed clones

Microprojectile bombardment of DNA on a carrier is an effective method for delivering DNA past the plasma membrane and two mitochondrial membranes into the mitochondrial

matrix. This method was pioneered in plants by John Sanford (Sanford et al., 1987), and first demonstrated in yeast by Sanford, Butow and colleagues (Johnston et al., 1988). Described below is the general transformation procedure used for *S. cerevisiae* (Bonnefoy and Fox, 2007).

The linear or circular plasmid DNA to be transformed is alcohol-precipitated onto a carrier substrate, usually tungsten or gold particles <1µm. The bombardment itself occurs in a biolistic gun chamber (Sanford, 1988), where rising vacuum pressure ruptures a pressure sensitive disk holding the DNA-precipitated particles, driving the particles onto a freshly plated lawn of haploid yeast cells. The plate medium then selects for uptake of either the plasmid of interest or of a co-transformed marker if the target plasmid does not confer a selectable phenotype.

The target cells for mitochondrial transformation are typically ρ^0 (non-mitochondrial DNA-containing) derivatives of a chosen strain, to ensure that the only mitochondrial DNA present is transformation-derived. After selection for the co-transformed nuclear marker, transformants must be screened for the presence of the transforming mitochondrial DNA. Following mitochondrial uptake of the desired DNA, positive haploid clones are mated to a strain containing wild-type mitochondrial genomes, allowing mixing of the transformed DNA with the target mitchondrial DNA. Generally, the desired outcome is integration of the synthetic DNA construct into the mitochondrial genome, although mitochondrial plasmid maintenance can also occur. The specific example of *ARG8^m* integration is provided below in Section 3.1.

3. The *ARG8^m* auxotrophic mitochondrial reporter gene

3.1 Building an auxotrophic mitochondrial reporter: the *ARG8^m* gene

The phenotypic output of a genetic reporter system determines its strengths and weaknesses as an analytic tool. In yeast, multiple auxotrophic (factor-requiring) mutants have historically been used with great success as both selective markers and phenotypic reporters. The defined requirements of yeast grown in culture allow for synthetic reconstitution of growth media lacking a specific amino acid. Commonly used auxotrophic markers in yeast include growth status on media lacking uracil, histidine, leucine, arginine, methionine, and lysine. Many laboratory strains, including our wild-type strain, DFS188 and its derivatives, lack the ability to make these amino acids due to specific nuclear mutations. These strains are typically maintained in rich media, allowing unrestricted growth. Withdrawal of the amino acid in question results in cell death. Rescue of a growth phenotype occurs when the gene that complements the nuclear mutation for that amino acid's synthesis is supplied on a plasmid or as part of a conditional reversion construct, allowing cells to regain prototrophic (factor-independent) growth.

While an ideal system to assess mechanisms associated with nuclear gene expression, the mitochondrial genome has long been inaccessible to auxotrophic reporter manipulation because it does not encode any amino acid biosynthetic enzymes. Direct insertion of a nuclear gene is impossible, as the codon usage of the mitochondrial genome differs from the nuclear genome both in the preferred codon frequency and in some codon products. Multiple nuclear leucine codons encode threonine in mitochondria, and a nuclear UGA stop codon encodes tryptophan in the mitochondria of yeast (Bonitz et al., 1980). Generating a

mitochondrial auxotrophic reporter thus requires mutating a nuclear gene to enable its expression from the mitochondrial genome. Once constructed, this gene must be introduced into the mitochondrial genome with the appropriate transcriptional and translational cues. To date, the only such auxotrophic marker gene to be engineered in this way is the synthetic *ARG8^m* gene, made by Tom Fox and colleagues (Steele et al., 1996).

The nuclear *ARG8* gene encodes acetylornithine transaminase, which catalyzes an early step in the biosynthesis of ornithine, a precursor to both arginine and proline. The Arg8 protein is normally localized to the mitochondrial matrix and yields the active mitochondrial transaminase following cleavage of its N-terminal targeting sequence (Steele et al., 1996). It is therefore an ideal candidate for reconfiguration as a mitochondrial gene, as its product functions within the matrix and does not require mitochondrial export for phenotypic expression.

Fox and group began by synthetically generating a 1.3 kilobase fragment encoding the entire 423 amino acid acetylornithine transaminase enzyme. Substitutions were made at 12 CUN codons (n: Leu; mt: Thr) and 6 AUA codons (n: Ile; mt: Met) to maintain the Leu and Ile residues. In addition, each of the two Trp codons was changed to UGA (n: STOP; mt: Trp) ensuring *ARG8^m* expression from the mitochondrial genome only. This construct was introduced into a plasmid containing mitochondrial DNA sequence flanking the *COX3* gene, providing sequences for recombination-dependent integration and ensuring the presence of correct transcriptional and translational processing signals.

Several steps were then required to generate the desired end product of the *cox3::ARG8^m* sequence incorporated into the mitochondrial genome while maintaining isogenicity of both mitochondrial and nuclear genomes. These steps are a "shell game" of genetic manipulation, designed to shield various DNA pools from one another until the desired product is achieved (Fig. 1).

The plasmid containing *cox3::ARG8^m* was biolistically transformed into a ρ^0 haploid yeast strain, ensuring that plasmid DNA was the only DNA present in the mitochondrial compartment (Fig. 1). In addition, these yeast cells lack the nuclear *ARG8* gene and carry a mutation in the *KAR1* gene that prevents nuclear fusion during mating. A second plasmid carrying a functional *LEU2* allele was co-transformed to allow selection of successful DNA uptake by a Leu⁺ phenotype. Note that the mitochondrial DNA targeted construct, though present, is inactive; screening for its presence therefore requires selection of a DNA-dependent effect. Transformants were screened for the presence of the *ARG8^m* gene in the mitochondria by mating with a strain carrying a mitochondrial genome deletion in the 5′ untranslated region upstream of the *COX3* gene. Only Leu⁺ transformants that also carry the mitochondrial plasmid with the wild-type *COX3* 5′ sequence will complement the deletion to give rise to respiring recombinants. Positive mitochondrial transformants identified by this test mating were purified and used to generate the integrated reporters.

To allow mixing of the mitochondrial plasmid DNA with intact mitochondrial genomes, the biolistically transformed cells were mated with a second haploid strain bearing normal mitochondrial DNA (Fig. 1, Fig. 2A). Since one strain is karyogamy-deficient, the nuclear envelopes do not fuse. Cell division gives rise to haploid cells, and one haploid genome can be selected in subsequent divisions. The mitochondria, however, undergo rapid fusion, allowing interaction between plasmid and mitochondrial DNA. This process is known as
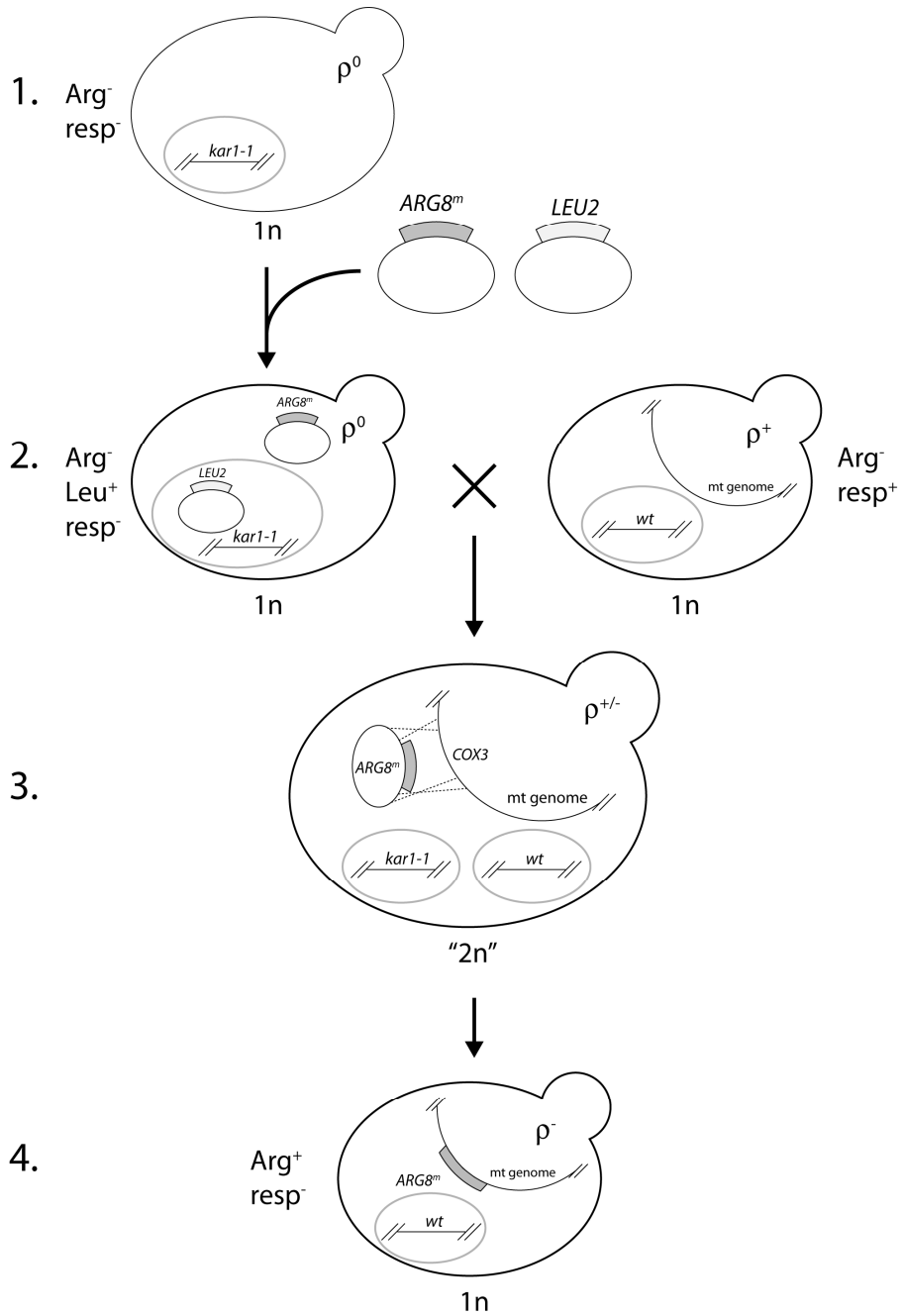
Fig. 1. Diagrammatic representation of biolistic transformation and integration of exogenous DNA into the yeast mitochondrial genome. Mitochondrial genome status is given in ρ

nomenclature. Slash marks indicate endogenous genomic DNA; ovals with bars for relevant genes represent plasmid DNA. Gray line denotes the nucleus. Selectable phenotypes of the genotype shown are written outside the cell. Ploidy is written below the cell. **Step 1.** A haploid, ρ^0, nuclear fusion-deficient mutant (*kar1-1*) is biolistically transformed with two plasmids, one bearing the *LEU2* allele and one the *ARG8^m* allele with *COX3* flanking sequence. **Step 2.** Transformants are selected by Leu+ growth, reflecting nuclear uptake of the *LEU2*-bearing plasmid (and possible uptake of the *ARG8^m*-bearing plasmid; confirmation of uptake is described in Sections 2.3 and 3.1). Leu+ transformants are crossed with wild-type cells. **Step 3.** Mixing of the *ARG8^m* plasmid with endogenous mitochondrial genomes; homologous regions of *COX3* non-coding sequence flanking *ARG8^m* mediate recombination at the mitochondrial *COX3* locus. **Step 4.** Replacement of endogenous *COX3* with *ARG8^m* in the mitochondrial genome, maintained by Arg+ selection. Loss of *COX3* confers respiration failure.

cytoduction. Homologous recombination between the *COX3* flanking sequences mediated replacement of the endogenous *COX3* gene with plasmid-derived *cox3::ARG8^m* sequence. By simultaneously selecting for both mitochondrial (Arg+) and nuclear markers, a haploid strain of the desired nuclear and mitochondrial backgrounds results that is phenotypically Arg+ and respiration-deficient, requiring a fermentable carbon source.

Mitochondrial genome incorporation of *cox3::ARG8^m* was confirmed by Southern blotting, while its requirement for the Arg+ phenotype was shown by curing yeast of their mitochondrial DNA and observing a reversion to an Arg- phenotype. The Arg+ phenotype was also dependent on the *COX3* translational activation complex, as deletion of the genes encoding the complex components result in Arg- cells. Finally, immunoblot analysis demonstrated that the protein product of the *ARG8^m* gene is identical in size to that of the nuclear *ARG8*, suggesting correct N-terminal processing. These controls elegantly demonstrate the correct location, expression and control of the *cox3::ARG8^m* reporter gene (Steele et al., 1996).

The *ARG8^m* reporter was initially used by Fox and group to examine mechanisms of mitochondrial translation (Bonnefoy and Fox, 2000; Dunstan et al., 1997), but its utility as a reporter extends to any assay in which reporter expression can be made a meaningful indicator of the function of interest. Fox and colleagues have also used the *ARG8^m* and other reporters (including a recoded mitochondrial GFP (Cohen and Fox, 2001)) to measure peptide import and export from the mitochondria (He and Fox, 1997; Torello et al., 1997), and to generate mutations in respiratory complex subunits (Ding et al., 2008). In addition, we and others have used *ARG8^m* expression to measure various aspects of mitochondrial DNA instability. The following sections will describe the construction, insertion, and use of these reporters in detail.

3.2 *ARG8^m* as a reporter of mitochondrial translation, DNA repair, recombination, and heteroplasmy

Prior to the construction of the *ARG8^m* mitochondrial reporter and the advent of qPCR and high-throughput sequencing, only two methods were available to easily assess mitochondrial DNA stability. First, ρ^- yeast cells with grossly defective mitochondrial DNA are respiration deficient and display slower growth on fermentable media, forming "petite"
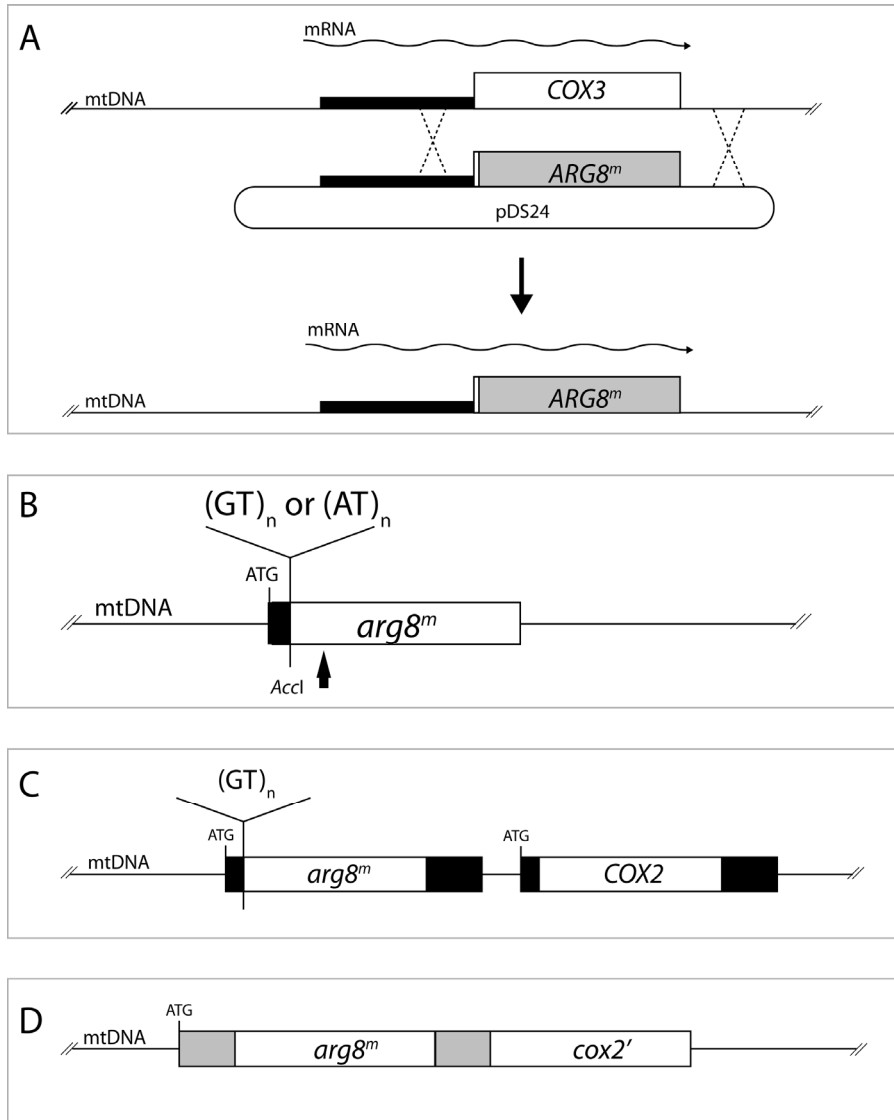
Fig. 2. *ARG8^m* construct diagrams. Slash marks indicate endogenous genomic DNA; ovals with bars for relevant genes represent plasmid DNA. **A.** Endogenous *COX3* locus (top) and replacement by *ARG8^m* borne on plasmid pDS24 to generate the mitochondrial genome-integrated *ARG8^m* allele (bottom). Thin black bar represents non-translated sequence; labeled white and gray bars denote the *COX3* and *ARG8^m* ORFs, respectively. Adapted from (Steele et al., 1996). **B.** Insertion of GT or AT dinucleotide repeats into a *cox3::ARG8^m* reporter fusion using an internal *Acc*I restriction endonuclease site. Black bar represents *COX3* coding sequence, ATG denotes *COX3* translational start, arrow represents post-translational cleavage site to yield functional Arg8p. Adapted from (Sia et al., 2000). **C.** Respiring

microsatellite reporter insertion upstream of the endogenous *COX2* locus, flanked by *COX2* non-coding sequence (black bars) including the translational start (ATG). Adapted from (Mookerjee and Sia, 2006). **D.** *Rep96::arg8m::cox2'* direct repeat-mediated deletion (DRMD) reporter. Gray bars represent the first 96 base pairs of *COX2* translated sequence, either containing (ATG, left of *arg8m* locus) or missing (right of *arg8m* locus) the ATG translational start codon. Adapted from (Phadnis et al., 2005).

colonies. Though petite formation is a somewhat useful indicator of gross mitochondrial DNA abnormality, its pleiotropic nature (the petite phenotype may result from nuclear *pet* as well as mitochondrial *mit* mutations), its origin via multiple types of mitochondrial DNA mutation, and the lack of a mechanistic explanation for the generation of ρ- genomes all limit its utility. Second, because mitochondria retain many prokaryotic properties, they are selectively sensitive to antibiotic drugs, including erythromycin. Erythromycin binds to the 21S mitochondrial ribosomal RNA, disrupting mitochondrial protein synthesis. Specific point mutations to the mitochondrial 21S rRNA gene prevent erythromycin from binding to the gene product, conferring an erythromycin resistant (Ery[R]) phenotype (Cui and Mason, 1989; Kalifa and Sia, 2007). In this way, the appearance of Ery[R] colonies can be used to estimate mitochondrial DNA point mutation accumulation rates using various calculation methods (Lea and Coulson, 1949; Luria and Delbruck, 1943). Ery[R] acquisition is useful because it is largely restricted to one mutation type. However, the spectrum of mutations that can be obtained is biased in that mutants must maintain mitochondrial ribosome function to preserve respiration competence. Since erythromycin does not affect yeast viability, only respiration, yeast strains are resistant to erythromycin in fermentable media. This makes it impossible to select point mutations under non-respiring conditions using Ery[R].

A spontaneous *cox3::arg8m* mutant, isolated by Fox and group and reported by Strand and Copeland, is an Arg- revertant that was shown to contain two nucleotide substitutions and a +1 frameshift insertion mutation (Strand and Copeland, 2002). Deletion of a single nucleotide restores the proper reading frame and allows reversion to an Arg+ phenotype (Zhang et al., 2006). This mutant was proposed as a replacement for the Ery[R] mutation assay, which measures true point mutations. While this reporter has some advantages over Ery[R] acquisition, this approach fails to distinguish between different types of mitochondrial DNA mutation and is therefore less helpful at elucidating the mechanisms responsible for their generation. Our subsequent work has demonstrated that the proteins that impact mitochondrial nucleotide substitutions and mitochondrial insertion/deletion mutations are not always the same, allowing the dissection of pathways that differentially affect point and frameshift mutation as described below.

4. Measuring mitochondrial microsatellite instability

Short, repetitive sequences consisting of di- and tri-nucleotide repeats are abundant in the nuclear genome, in both coding and non-coding regions. Their appearance and inherent instability in the coding regions of several proteins is the underlying cause of the polyglutamine diseases, including Huntington's disease and multiple types of spinocerebellar ataxia. These microsatellites are an important source of mutation in nuclear DNA through the internal repetition that facilitates repeat length changes by polymerase slippage, and through the ability of separate repeated regions to undergo homologous recombination in *trans*.

Yeast mitochondrial DNA is highly microsatellite-enriched, partially due to the abundance (83%) of A and T. Mammalian mitochondrial DNA lacks most of the non-coding AT-rich DNA (~44% A/T) that gives rise to this bias, but still contains AT-rich repetitive regions (Anderson et al., 1980). If this repetition confers higher mitochondrial DNA instability, similar to its effects on nuclear DNA (Wierdl et al., 1997), it could play an important role in mitochondrial dysfunction as it relates to aging and aging-related diseases. These findings provided the impetus for making a mitochondrial microsatellite reporter system.

In 2000, Petes, Fox, and colleagues published an analysis of mitochondrial DNA microsatellite instability using the *ARG8m* reporter as a marker of frameshift mutation within a microsatellite repeat (Sia et al., 2000). This approach built on a previous reporter of yeast nuclear microsatellite instability (Henderson and Petes, 1992). They generated *cox3Δ::arg8m*-bearing plasmids that contained poly-GT and poly-AT tracts 15-17 repeats in length 5′ to the *ARG8m* sequence, shifting the reading frame of the gene either 1 or 2 nucleotides out of frame (Fig 2B). Once incorporated and expressed, these *cox3Δ::arg8m*(G/A-T) constructs express a nonsense transcript. However, a frameshift mutation in the microsatellite tract that restores the correct reading frame restores a functional gene product, conferring an Arg+ phenotype. As a control, plasmids were also made bearing microsatellite inserts that did not result in an *ARG8m* frameshift, conferring a constitutive Arg+ phenotype, confirming that insertion of the repetitive sequence did not disrupt *ARG8m* function.

The experiments carried out with these strains were among the first to demonstrate fundamental differences between nuclear and mitochondrial DNA processing and maintenance. Unlike nuclear DNA, in which poly(AT) and poly(GT) tracts have similar levels of instability, with repeat addition favored, mitochondrial poly(AT) tracts are much more stable than poly(GT) tracts, and repeat deletion predominates. These and other differences suggest that assumptions of mitochondrial DNA behavior based on nuclear DNA may be inherently flawed, preventing a clear understanding of how to predictably manipulate mitochondrial DNA.

To allow microsatellite instability measurement in respiring cells, a state that imposes a respiration requirement on mitochondrial DNA maintenance, we developed and characterized a version of the microsatellite reporter that is respiration competent (Kalifa and Sia, 2007; Mookerjee and Sia, 2006). This ensures that the measured microsatellite instability occurs in an otherwise functional background. This new reporter serves two useful purposes, allowing us to determine the effect of an active respiratory chain on mitochondrial mutagenesis, and to assess microsatellite instability in mutant strains that only maintain mitochondrial DNA under constant respiratory selection. For this reporter, instead of replacing *COX3*, *ARG8m* was inserted with *COX2* flanking sequence upstream of the endogenous *COX2* locus (Fig. 2C). These untranslated flanking sequences ensure correct expression while avoiding disruption of cytochrome oxidase subunits. In this new genetic and functional context, poly(GT)$_{16}$ repeats in the +1 frame were approximately 10-fold more unstable than the original *cox3Δ::arg8m*(G/T)$_{16}$ (Kalifa and Sia, 2007). Further experiments will be required to determine whether this is due to the altered flanking sequence or respiration status of the cell.

5. Measuring direct repeat-mediated deletions

Accumulation of mitochondrial deletions is associated with multiple pathologies and with aging. These deletions are commonly flanked by direct repeats, raising speculation that they

are recombination-mediated. A detailed understanding of mitochondrial DNA recombination has lagged far behind equivalent nuclear recombination processes, which in turn limits our efforts to use recombination both as a molecular tool for genome manipulation and as a biological correlate of mitochondrial dysfunction. By manipulating the sequence context of the *ARG8^m* gene, we developed reporters to measure and characterize mitochondrial recombination in yeast. These studies have revealed detailed information about the requirements for and mechanisms of recombination in the yeast mitochondrial genome.

5.1 DFS188 *Rep96::ARG8^m::cox2'* reporter and variants

We generated a synthetic deletion substrate with the *ARG8^m* gene fused in frame to the first 99 bp of the mitochondrial *COX2* gene. This construct was followed by the entire *COX2* coding sequence, lacking the 5' start codon, giving rise to *ARG8^m* flanked by 96 bp of directly repeated *COX2* sequence (Fig. 2D). This deletion substrate expresses functional acetylornithine transaminase and confers an Arg+ phenotype. *COX2* is not expressed, as translation terminates with Arg8p. These cells are therefore non-respiring. However, deletion between the flanking 96-bp repeats excises *ARG8^m* and restores a functional *COX2* sequence with the appropriate initiation and termination signals. Cells that have undergone either sufficient recombination events, or homoplasmic fixation of one or a few recombination events, display a phenotypic shift to respiration competence and arginine auxotrophy. The work that followed development of the *Rep96::ARG8^m::cox2'* reporter used either this form or modified versions with changes to the flanking repeats.

Yeast bearing the *Rep96::ARG8^m::cox2'* reporter can be assayed by imposing different nutrient conditions that select for either the original construct or the deleted one. Cells are initially grown on medium lacking arginine to maintain the original reporter. Individual colonies are then separately diluted and plated on glycerol medium (YPG) to select for deletion events, which are scored after 3 days. The median number of colonies appearing on YPG plates is used to estimate the mutation rate using the method of the median (Lea and Coulson, 1949), which incorporates statistical assumptions about the number of mutational events vs. colony number.

5.2 Characterizing direct-repeat mediated deletion (DRMD) in yeast

Work by Phadnis *et al.* (2005) demonstrates the utility of this reporter in characterizing mitochondrial recombination. First, generation of deletion reporters containing different repeat lengths revealed that the rate of *ARG8^m* deletion is linearly dependent on repeat lengths between 33 and 96 bp; a 21-bp repeat did not facilitate a significant deletion rate. This establishes the minimal efficient processed sequence (MEPS) between 21 and 33 bp long. This MEPS is longer than the direct repeats flanking mitochondrial DNA deletions in mammalian systems, including the 13 bp flanking the 5-kb "common deletion" (Schon et al., 1989) and a 7-bp recombination-associated sequence (Myers et al., 2008). It should be noted that in humans, direct repeats mediate only some of the total detectable mitochondrial DNA deletions (Guo et al., 2010; Srivastava and Moraes, 2005).

Second, the effects of heterology on deletion efficiency were tested by introducing silent mutations into either the leading or following repeat (relative to the direction of

transcription), giving rise to ~2% heterology between the repeat sequences. This design was meant to allow comparison to similar work in the yeast nuclear genome, where a 3% heterology between 205 bp repeats decreased the rate of deletion formation 6-fold (Sugawara et al., 2004) These experiments revealed similar behavior in mitochondrial DNA, where a 3- to 4-fold reduction in deletion formation rate was observed. Interestingly, this effect was dependent on mutation placement in the leading repeat; the same mutations in the following repeat had no effect on deletion rate. One explanation for this is that the heteroduplex rejection mechanisms may act more stringently on particular mispair orientations. In successful deletion events, the final sequence was almost always that of the repeat closest to the remaining *COX2* sequence. These findings are likely to be related to the specific types of mechanisms that mediate DRMD, but at present are unexplained.

The possible mechanisms mediating repeat-dependent deletion can be partially distinguished based on the DNA products they generate, namely, whether detectable products are reciprocal or non-reciprocal. Unlike qPCR detection methods that are commonly used to measure mitochondrial DNA deletions *in vivo*, we used Southern blot analysis to examine the products of direct repeat dependent deletion. In theory, the reciprocal products, either a circular molecule or tandem duplication on another mitochondrial genome would both be detectable. We did not observe such species using this technique. However, we detected reciprocal products through PCR amplification and electrophoresis, indicating that reciprocal events do occur. This method cannot distinguish between circular "pop-outs" or tandem duplications. This analysis revealed that the majority of deletion events were non-reciprocal, suggesting a replication- or single strand annealing-based mechanism.

Third, genes believed to be involved in repeat-mediated deletion were tested to determine their effects on the rate of mutation. Mutations to the proof-reading domain of the mitochondrial DNA polymerase were shown to affect *Rep96::ARG8m::cox2'* deletion. The *mip1-D347A* exonuclease proofreading mutation, which confers a approximately 500-fold increase in point mutation rates as measured by Ery^R, actually decreases direct repeat-mediated deletion nearly 5-fold (Phadnis et al., 2005). Interestingly, several years later Vermulst *et al.* (2008) determined that the analogous Polγ mutation shifts deletion formation from a repeat-mediated to a non-repeat-mediated mechanism (Vermulst et al., 2008). That the analogous mutation appears to limit DRMD in both yeast and vertebrates suggests that at least some of the mechanisms for mitochondrial deletion may be conserved between these organisms.

5.3 Assaying heteroplasmy in yeast

The *Rep96::ARG8m::cox2'* reporter, by virtue of its construction, can also be used as a phenotypic marker of mitochondrial DNA heteroplasmy. Unlike mammalian cells, yeast containing multiple different mitochondrial DNA types will purify this population to fix one type within 6-10 budding cycles (Dujon, 1981). Several factors are thought to be involved in this process of homoplasmic sorting. First, compaction of multiple, possibly clonal genome copies into nucleoids reduces the number of heritable units. Second, a limited number of nucleoids are transmitted to daughter cells, and may undergo regulated sorting, further restricting mitochondrial DNA inheritance (Spelbrink, 2009). In mammaliam systems, a third factor may be that nucleoids do not appear to readily mix, even when given the opportunity to do so by cytoduction (Schon and Gilkerson, 2010).

As stated earlier, the *Rep96::ARG8ᵐ::cox2'* reporter confers respiration incompetence and prototrophy. This phenotype is mutually exclusive with that of the deleted reporter, which confers respiration competence and arginine auxotrophy. Therefore, growing yeast under conditions that select for both respiration competence and arginine prototrophy, representing one phenotype of each possible state of the reporter, should select for yeast that maintain heteroplasmy. This was shown successfully in an analysis of point mutants of *MSH1* a yeast homolog of the MutS mismatch repair initiation protein (Mookerjee and Sia, 2006). In contrast to wild-type yeast, in which heteroplasmic maintenance occurred at a frequency on the order of 10^{-5}, *msh1* alleles permitted frequencies of heteroplasmy up to 0.25, 4 to 5 orders of magnitude greater. The role of Msh1p in homoplasmic sorting remains undetermined.

6. Elucidation of mitochondrial DNA repair pathways

With specific reporters of microsatellite instability, point mutation, and direct repeat instability, coupled with direct sequencing and Southern blotting, the pathways of mitochondrial DNA repair become more readily accessible to quantitative analysis. This section will discuss some of the research findings resulting from use of the mitochondrial reporters described above.

6.1 Mismatch recognition combines with recombination and base excision repair pathways

Use of the *Rep96::ARG8ᵐ::cox2'* reporter allowed the initial characterization of mitochondrial DNA recombination requirements, including repeat size, degree of sequence identity, and directional/positional repair bias as described in Section 5.2 and in Phadnis et al. (2005). We have further applied the reporters to determining the total complement of mitochondrial DNA repair mechanisms and their interactions

Mismatch repair has been a predicted pathway of mitochondrial DNA repair since the identification of the mitochondrially-localized MutS homolog, Msh1p. However, there is currently no direct evidence in yeast for mismatch repair activity. Human mitochondria do have a putative mismatch repair mechanism (De Souza-Pinto et al., 2009), but do not possess a MutS homolog. Point mutation accumulation rates increase with Msh1p disruption, but evidence from multiple groups suggests that this is due to base excision repair (BER), rather than mismatch repair (MMR), defects. Further, no other characterized mismatch repair proteins are known to localize to the mitochondria.

Haploid yeast strains with deletions of the *MSH1* gene cannot maintain wild-type mitochondrial DNA, generating ρ⁻ petites at a high frequency even in the presence of selection on a non-fermentable carbon source (Chi and Kolodner, 1994). To explore this problem in detail, we characterized the mutagenic consequences of three point mutations to *MSH1*. These mutations are analogous to well-studied mutations in *E. coli* MutS and yeast nuclear MutS homologs and were chosen based on the biochemical functions of the mutant proteins and their ability to maintain mitochondrial function when under selection. The *msh1-F105A* substitution lies in the conserved DNA binding domain, and is predicted by its homology to MutS and Msh6p to impair DNA binding and mismatch recognition (Bowers et al., 1999; Schofield et al., 2001). The *msh1-G776D* and *msh1-R813W* substitutions both lie

within the highly conserved ATPase domain, although they are predicted to have different phenotypic and biochemical consequences. The yeast msh6p-G987D, analogous to our msh1p-G776D mutant, is significantly impaired in ATP-binding and displays an ability to bind mismatches, but is defective in further processing. The msh2p-R730W mutation, analogous to our msh1p-R813W, is able to bind ATP, but is defective in hydrolysis. While complexes containing this mutant form of Msh2p cannot perform mismatch repair, they remain at least partially functional for promoting deletions at directly repeated sequences (Kijas et al., 2003; Studamire et al., 1998).

By comparison with the known mutations, all three *msh1* mutations were predicted to result in loss of any mismatch repair activity. Consistent with this hypothesis, all three *msh1* mutants displayed increased point mutation rates. However, this increase was insufficient to explain the catastrophic loss of wild-type mitochondrial DNA and respiratory function (Mookerjee et al., 2005; Mookerjee and Sia, 2006), as mutations in the proof-reading domain of the mitochondrial replicative polymerase display more than ten-fold higher rates of point mutation, but can maintain ρ^+ DNA (Foury and Vanderstraeten, 1992).

We then characterized the effects of *MSH1* mutation on microsatellite instability, hypothesizing that since nuclear mismatch repair disruption greatly increases nuclear microsatellite instability, a similar observation in mitochondria would support a bona fide mismatch repair function. Examination of *msh1* alleles with disruptions in the DNA-binding (*msh1-F105A*) and ATPase (*msh1-R813W*) domains revealed no significant changes in GT microsatellite instability, suggesting that Msh1p initiates non-mismatch repair mechanisms. It is formally possible that mitochondrial mismatch repair is not equivalent with respect to microsatellites between mitochondria and the nucleus due to other differences between the two compartments. Still, this finding prompted us to search for other repair activities involving Msh1p.

Mismatch repair proteins have been shown to function in other DNA repair pathways, including BER, nucleotide excision repair (NER), and homologous recombination (Goldfarb and Alani, 2005; Polosina and Cupples, 2010). Using the mitochondrial reporters, we were able to examine the genetic interaction of *MSH1* with putative recombination (Mookerjee and Sia, 2006) and BER components (Pogorzala et al., 2009).

Though widely accepted as a functional mechanism in mitochondrial DNA, the proteins that carry out recombination, and the specific mechanisms themselves, are largely unknown. Due to differences in the available proteins, in the substrate, or in the presumably constant availability of a homologous template, mitochondria may combine existing repair components is ways not seen in the nucleus (Masuda et al., 2009). We speculated that Msh1p, like its nuclear homologs, may play a role in the generation of deletions at directly-repeated sequences, and therefore would be predicted to result in reduced DRMD. Unexpectedly, we found that all three *msh1* mutations increase deletion rate approximately 100-fold, revealing a novel role for Msh1p function in mitochondrial recombination suppression that is not mirrored by any of its nuclear homologs.

Previously, examination of the deletion junctions of ρ^- genomes in spontaneous *petite* strains had suggested that recombination utilizing repeated sequences may be the initiating event in their generation (Dujon, 1981). If so, all three of the *msh1* mutants would be predicted to give rise to similar, high levels of non-respiring cells. However, while the *msh1-*

F105A, msh1-G776D, and the *msh1-R813W* mutant strains all display high rates of point mutation and DRMD, the *msh1-R813W* mutant strain displays significantly lower rates of respiration loss than strains expressing the other two *MSH1* mutant alleles. This result calls into question several assumptions about how cells become ρ- *petite*. In these cells, respiration capability is lost with deletion of the majority of the mitochondrial genome, but it is clear that gene disruptions leading to increased rates of both point mutation accumulation and repeat-mediated deletion do not necessarily lead to loss of ρ+ mitochondrial DNA.

If Msh1p mutation allows it to bind but not release mismatched DNA, it is tempting to speculate that DNA molecules might become aberrantly linked. While not direct confirmation of this, we did find that the *msh1-G776D* and *msh1-R813W* alleles, and to a lesser extent, the *msh1-F105A* allele, conferred increases in the frequency of heteroplasmic maintenance of at least three or more orders of magnitude.

In addition to suppressing DRMD, Msh1p is also an important component of mitochondrial base excision repair. Through conventional epistasis analysis of *MSH1* and the BER genes *OGG1*, *NTG1*, and *APN1*, we found that Msh1p defects gave rise to different mutator phenotypes depending on the form of BER, and proposed that mismatch or lesion recognition contributes to short-patch BER, while DNA binding plays a stabilizing role during long-patch BER (Pogorzala et al., 2009).

6.2 Translesion polymerases facilitate frameshifts but suppress mitochondrial point mutation

The high fidelity of replicative DNA polymerases arises through extremely stringent requirements for nucleotide binding in the active site. While normally desirable, the inflexibility of these domains to accept alternate substrates causes replication fork stalling or collapse in the presence of cyclobutane-pyrimidine dimers, a common UV-induced product, and other damage resulting in large adducts.

To remedy this, a second class of polymerases exists that have much less stringent binding requirements for nucleotide incorporation. These translesion polymerases sacrifice fidelity for greater flexibility in template usage, favoring processivity but leading to mutagenic DNA synthesis that introduces both point and frameshift mutations. Consequently, their disruption gives rise to higher sensitivity to DNA damaging agents (e.g., UV) but lower nuclear DNA mutation rates in surviving cells.

Polγ is the only known mitochondrial replicative polymerase and, until recently, the only polymerase with known mitochondrial DNA activity. In yeast, *REV3*, encoding the catalytic subunit of the Polζ translesion polymerase, was previously implicated in mitochondrial DNA maintenance (Smolińska, 1987). Rev1p and Rev7p interact with Rev3p and are required for Rev3p-dependent synthesis opposite damaged templates (Lawrence et al., 2000). Subsequently, the Singh group observed that Rev1p, Rev3p, and Rev7p have N-terminal mitochondrial localization signals and demonstrated the ability of these N-terminal sequences to deliver GFP to mitochondria. They used the *cox3::arg8m* variant isolated by T. Fox to analyze Arg+ reversion as a measure of mutation, and found that single deletion of Rev1p, Rev3p, and Rev7p all result in a decrease in mutation frequency (Zhang et al., 2006). This reporter reverts to Arg+ via loss of a single base pair, indicating

that Polζ is important in the generation of mitochondrial insertion/deletion (indel) mutations.

Using the respiring (GT_{16}+1) and (GT_{16}+2) reporters (Fig 2C), (Mookerjee and Sia, 2006), we found that single deletions of *REV1*, *REV3*, and *REV7* all led to decreases in both spontaneous and UV-induced frameshift consistent with previously published work and supporting a role for all three gene products in the generation of mitochondrial indels. However, unlike nuclear DNA, which displays decreases in both frameshifts and point mutations, the *rev3-Δ* and *rev7-Δ* strains also showed unexpected increases in spontaneous mitochondrial DNA point mutation rate, and all three deletions gave rise to increases in UV-induced point mutation rates. Sequence analysis of these mutants revealed a UV-dependent shift favoring A to T transversion in the absence of Rev1p, consistent with repair of a thymine dimer by a mechanism biased towards adenine insertion (Kalifa and Sia, 2007).

These observations emphasize the importance of examining the effect of disrupting repair or damage tolerance pathways on multiple types of mutations, as they are often generated and repaired via distinct mechanisms. The employment of multiple mutagenesis reporters allows the proper dissection of these pathways.

7. Conclusion

Increasingly, proteins previously considered to be nuclear DNA repair factors are found to also display mitochondrial localization (Section 1.3), suggesting significant overlap between nuclear and mitochondrial DNA repair pathways with respect to their protein components. However, our analysis of the phenotypic consequences of mutating the relevant genes reveals significant differences in the contribution of these proteins to mitochondrial mutagenesis and repair (Sections 5 and 6). These differences likely result from compositional DNA differences, the different packaging and DNA-binding proteins, the different regulatory control of mitochondrial DNA replication and transmission, the exposure to certain kinds of damage, and the availability of other repair proteins, between nuclear and mitochondrial DNA. We should not expect that studies of nuclear DNA repair can simply be extrapolated to generate correct mitochondrial models.

Careful analysis of these pathways within the mitochondrion will require tools like those we have described here. These reporters provide us with the ability to differentiate between point substitutions, frameshift mutations, and deletion events and will be critical to elucidating specific pathways. While *Saccharomyces cerevisiae* is currently the only model system available for these studies, the conservation of DNA repair pathways among eukaryotes supports the hypothesis that some, if not most of these repair proteins will have conserved mitochondrial roles. Finally, understanding mitochondrial DNA repair serves two practical purposes. First, it allows us to understand an important genetic system whose failure is heavily implicated in aging. Second, understanding the mechanisms of repair may facilitate their exploitation for mitochondrial genome manipulation in other systems.

8. Acknowledgements

9. References

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J., Staden, R., Young, I. G., 1980. Sequence and organization of the human mitochondrial genome. Nature. 290, 457-465.

Arnaudo, E., Dalakas, M., Shanske, S., Moraes, C. T., DiMauro, S., Schon, E. A., 1991. Depletion of muscle mitochondrial DNA in AIDS patients with zidovudine-induced myopathy Lancet. 337, 508–510.

Baker, B. M., Haynes, C. M., 2011. Mitochondrial protein quality control during biogenesis and aging. Trends Biochem Sci. 1-8.

Boesch, P., Weber-Lotfi, F., Ibrahim, N., Tarasenko, V., Cosset, A., Paulus, F., Lightowlers, R. N., Dietrich, A., 2010. DNA repair in organelles: Pathways, organization, regulation, relevance in disease and aging. Biochim Biophysica Acta. 1813, 186-200.

Bogenhagen, D. F., Wang, Y., Shen, E. L., Kobayashi, R., 2003. Protein components of mitochondrial DNA nucleoids in higher eukaryotes. Mol Cell Proteomics. 2, 1205-16.

Bonitz, S. G., Berlani, R., Coruzzi, G., Li, M., Macino, G., Nobrega, F. G., Nobrega, M. P., Thalenfeld, B. E., Tzagoloff, A., 1980. Codon recognition rules in yeast mitochondria. Proc Natl Acad Sci USA. 77, 3167-70.

Bonnefoy, N., Fox, T. D., 2000. *In vivo* analysis of mutated initiation codons in the mitochondrial *COX2* gene of *Saccharomyces cerevisiae* fused to the reporter gene *ARG8m* reveals lack of downstream reinitiation. Mol Gen Genet. 262, 1036-46.

Bonnefoy, N., Fox, T. D., 2007. Directed alteration of *Saccharomyces cerevisiae* mitochondrial DNA by biolistic transformation and homologous recombination. Methods Mol Biol. 372, 153-66.

Bowers, J., Sokolsky, T., Quach, T., Alani, E., 1999. A mutation in the MSH6 subunit of the *Saccharomyces cerevisiae* MSH2-MSH6 complex disrupts mismatch recognition. J Biol Chem. 274, 16115-16125.

Canugovi, C., Maynard, S., Bayne, A.-C. V., Sykora, P., Tian, J., Souza-Pinto, N. C. d., Croteau, D. L., Bohr, V. A., 2010. The mitochondrial transcription factor A functions in mitochondrial base excision repair. DNA Repair. 9, 1080-1089.

Chang, S., 2000. The frequency of point mutations in mitochondrial DNA is elevated in the Alzheimer's brain. Biochem Biophys Res Comm. 273, 203-208.

Chatterjee, A., Singh, K. K., 2001. Uracil-DNA glycosylase-deficient yeast exhibit a mitochondrial mutator phenotype. Nucleic Acids Res. 29, 4935-4940.

Chen, H., Vermulst, M., Wang, Y. E., Chomyn, A., Prolla, T. A., Mccaffery, J. M., Chan, D. C., 2010. Mitochondrial fusion is required for mtDNA stability in skeletal muscle and tolerance of mtDNA mutations. Cell. 141, 280-9.

Chi, N. W., Kolodner, R. D., 1994. Purification and characterization of MSH1, a yeast mitochondrial protein that binds to DNA mismatches. J Biol Chem. 269, 29984-92.

Clayton, D. A., Doda, J. N., Friedberg, E. C., 1974. The absence of a pyrimidine dimer repair mechanism in mammalian mitochondria. Proc Natl Acad Sci USA 71, 2777-2781.

Coffey, G., Lakshmipathy, U., Campbell, C., 1999. Mammalian mitochondrial extracts possess DNA end-binding activity. Nucleic Acids Res. 27, 3348-3354.

Cohen, J. S., Fox, T. D., 2001. Expression of green fluorescent protein from a recoded gene inserted into *Saccharomyces cerevisiae* mitochondrial DNA. Mitochondrion. 1, 181-9.

Corral-Debrinski, M., Horton, T., Lott, M. T., Shoffner, J. M., McKee, A. C., Beal, M. F., Graham, B. H., Wallace, D. C., 1994. Marked changes in mitochondrial DNA deletion levels in Alzheimer brains. Genomics. 23, 471-6.

Coskun, P., Wyrembak, J., Schriner, S., Chen, H.-W., Marciniack, C., Laferla, F., Wallace, D. C., 2011. A mitochondrial etiology of Alzheimer and Parkinson disease. Biochim Biophys Acta. 1-33.

Cui, Z., Mason, T. L., 1989. A single nucleotide substitution at the *rib2* locus of the yeast mitochondrial gene for 21S rRNA confers resistance to erythromycin and cold-sensitive ribosome assembly. Curr Genet. 16, 273-279.

De Souza-Pinto, N. C., Mason, P. A., Hashiguchi, K., Weissman, L., Tian, J., Guay, D., Lebel, M., Stevnsner, T. V., Rasmussen, L. J., Bohr, V. A., 2009. Novel DNA mismatch-repair activity involving YB-1 in human mitochondria. DNA Repair. 1-16.

Ding, M. G., Butler, C. A., Saracco, S. A., Fox, T. D., Godard, F., di Rago, J.-P., Trumpower, B. L., 2008. Introduction of cytochrome b mutations in *Saccharomyces cerevisiae* by a method that allows selection for both functional and non-functional cytochrome b proteins. Biochim Biophys Acta. 1777, 1147-56.

Dmitrieva, N. I., Malide, D., Burg, M. B., 2011. Mre11 is expressed in mammalian mitochondria where it binds to mitochondrial DNA. Am J Phys. 301, R632-40.

Dujon, B., 1981. Mitochondrial genetics and functions. In: J. N. Strathern, et al., Eds.), Molecular biology of the yeast *Saccharomyces*: Life Cycle and Inheritance. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Dunn, D. A., Cannon, M. V., Irwin, M. H., Pinkert, C. A., 2011. Animal models of human mitochondrial DNA mutations. Biochim Biophys Acta. 1-30.

Dunstan, H. M., Green-Willms, N. S., Fox, T. D., 1997. In vivo analysis of *Saccharomyces cerevisiae COX2* mRNA 5'-untranslated leader functions in mitochondrial translation initiation and translational activation. Genetics. 147, 87-100.

Foury, F., Roganti, T., Lecrenier, N., Purnelle, B., 1998. The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. FEBS Lett. 440, 325-31.

Foury, F., Vanderstraeten, S., 1992. Yeast mitochondrial DNA mutators with deficient proofreading exonucleolytic activity. EMBO J. 11, 2717-26.

Garrido, N., Griparic, L., Jokitalo, E., Wartiovaara, J., van der Bliek, A. M., Spelbrink, J. N., 2003. Composition and dynamics of human mitochondrial nucleoids. Mol Bio Cell. 14, 1583-1596.

Gilkerson, R. W., 2009. Mitochondrial DNA nucleoids determine mitochondrial genetics and dysfunction. Int J Biochem Cell Bio. 41, 1899-906.

Goldfarb, T., Alani, E., 2005. Distinct roles for the *Saccharomyces cerevisiae* mismatch repair proteins in heteroduplex rejection, mismatch repair and nonhomologous tail removal. Genetics. 169, 563-74.

Green, G., MacQuillan, A., 1980. Photorepair of ultraviolet-induced petite mutational damage in *Saccharomyces cerevisiae* requires the product of the *PHR1* gene. J Bact. 144, 826-829.

Griffiths, L. M., Swartzlander, D., Meadows, K. L., Wilkinson, K. D., Corbett, A. H., Doetsch, P. W., 2009. Dynamic compartmentalization of base excision repair proteins in response to nuclear and mitochondrial oxidative stress. Mol Cell Biol. 29, 794-807.

Guo, X., Kudryavtseva, E., Bodyak, N., Nicholas, A., Dombrovsky, I., Yang, D., Kraytsberg, Y., Simon, D. K., Khrapko, K., 2010. Mitochondrial DNA deletions in mice in men: substantia nigra is much less affected in the mouse. Biochim Biophys Acta. 1-11.

Haslbrunner, E., Tuppy, H., Schatz, G., 1964. Deoxyribonucleic acid associated with yeast mitochondria. Biochem Biophys Res Comm. 15, 127-132.

He, S., Fox, T. D., 1997. Membrane translocation of mitochondrially coded Cox2p: distinct requirements for export of N and C termini and dependence on the conserved protein Oxa1p. Mol Bio Cell 8, 1449-1460.

Henderson, S. T., Petes, T. D., 1992. Instability of simple sequence DNA in *Saccharomyces cerevisiae*. Mol Cell Biol. 12, 2749-57.

Ibeas, J. I., Jimenez, J., 1997. Mitochondrial DNA loss caused by ethanol in *Saccharomyces* flor yeasts. Appl Envt Microbiol. 63, 7-12.

Johnston, S. A., Anziano, P. Q., Shark, K., Sanford, J. C., Butow, R. A., 1988. Mitochondrial transformation in yeast by bombardment with microprojectiles. Science. 240, 1538-1541.

Kalifa, L., Sia, E. A., 2007. Analysis of Rev1p and Pol zeta in mitochondrial mutagenesis suggests an alternative pathway of damage tolerance. DNA Repair 6, 1732-9.

Kamenisch, Y., Fousteri, M., Knoch, J., von Thaler, A.-K., Fehrenbacher, B., Kato, H., Becker, T., Dollé, M. E., Kuiper, R., Majora, M., Schaller, M., van der Horst, G. T., van Steeg, H., Röcken, M., Rapaport, D., Krutmann, J., Mullenders, L. H., Berneburg, M., 2010. Proteins of nucleotide and base excision repair pathways interact in mitochondria to protect from loss of subcutaneous fat, a hallmark of aging. J Exp Med. 207, 379-90.

Kang, D., Nishida, J., Iyama, A., Nakabeppu, Y., Furuichi, M., Fujiwara, T., Sekiguchi, M., Takeshige, K., 1995. Intracellular localization of 8-oxo-dGTPase in human cells, with special reference to the role of the enzyme in mitochondria. J Biol Chem. 270, 14659-14665.

Kaufman, B. A., Durisic, N., Mativetsky, J. M., Costantino, S., Hancock, M. A., Grutter, P., Shoubridge, E. A., 2007. The mitochondrial transcription factor TFAM coordinates the assembly of multiple DNA molecules into nucleoid-like structures. Mol Biol Cell. 18, 3225-36.

Kienhöfer, J., Häussler, D. J. F., Ruckelshausen, F., Muessig, E., Weber, K., Pimentel, D., Ullrich, V., Bürkle, A., Bachschmid, M. M., 2009. Association of mitochondrial antioxidant enzymes with mitochondrial DNA as integral nucleoid constituents. FASEB J. 23, 2034-44.

Kijas, A. W., Studamire, B., Alani, E. A., 2003. Msh2 separation of function mutations confer defects in the initiation steps of mismatch repair. J Mol Bio. 331, 128-138.

Kukat, C., Wurm, C. A., Spåhr, H., Falkenberg, M., Larsson, N.-G., Jakobs, S., 2011. Super-resolution microscopy reveals that mammalian mitochondrial nucleoids have a uniform size and frequently contain a single copy of mtDNA. Proc Natl Acad Sci USA. 108, 13534-9.

Lawrence, C. W., Gibbs, P. E., Murante, R. S., Wang, X. D., Li, Z., McManus, T. P., McGregor, W. G., Nelson, J. R., Hinkle, D. C., Maher, V. M., 2000. Roles of DNA polymerase zeta and Rev1 protein in eukaryotic mutagenesis and translesion replication Cold Spring Harb Symp Quant Biol. 65, 61-69.

Lea, D. E., Coulson, C. A., 1949. The distribution of the number of mutants in bacterial populations. J Genet. 49, 264-285.

LeDoux, S. P., Wilson, G. L., Beecham, E. J., Stevnsner, T., Wassermann, K., Bohr, V. A., 1992. Repair of mitochondrial DNA after various types of DNA damage in Chinese hamster ovary cells. Carcinogenesis. 13, 1967-73.

Legros, F., Malka, F., Frachon, P., Lombès, A., Rojo, M., 2004. Organization and dynamics of human mitochondrial DNA. J Cell Sci. 117, 2653-62.

Liu, P., Qian, L., Sung, J. S., De Souza-Pinto, N. C., Zheng, L., Bogenhagen, D. F., Bohr, V. A., Wilson, D. M., Shen, B., Demple, B., 2008. Removal of oxidative DNA damage via FEN1-dependent long-patch base excision repair in human cell mitochondria. Mol Cell Biol. 44.

Luria, S. E., Delbruck, M., 1943. Mutations of bacteria from virus sensitivity to virus resistance. Genetics 28, 491-511.

Margineantu, D. H., Gregory Cox, W., Sundell, L., Sherwood, S. W., Beechem, J. M., Capaldi, R. A., 2002. Cell cycle dependent morphology changes and associated mitochondrial DNA redistribution in mitochondria of human cell lines. Mitochondrion. 1, 425-435.

Masuda, T., Ito, Y., Terada, T., Shibata, T., Mikawa, T., 2009. A non-canonical DNA structure enables homologous recombination in various genetic systems. J Biological Chem. 284, 30230-9.

Meyer, R. R., Simpson, M. V., 1969 DNA biosynthesis in mitochondria. Differential inhibition of mitochondrial and nuclear DNA polymerases by the mutagenic dyes ethidium bromide and acriflavin. Biochem Biophys Res Comm 34, 238-244.

Mileshina, D., Ibrahim, N., Boesch, P., Lightowlers, R. N., Dietrich, A., Weber-Lotfi, F., 2011a. Mitochondrial transfection for studying organellar DNA repair, genome maintenance and aging. Mech Ageing Dev. 1-12.

Mileshina, D., Koulintchenko, M., Konstantinov, Y., Dietrich, A., 2011b. Transfection of plant mitochondria and *in organello* gene integration. Nucleic Acids Res. 10.1093/nar/gkr517.

Mookerjee, S. A., Lyon, H. D., Sia, E. A., 2005. Analysis of the functional domains of the mismatch repair homologue Msh1p and its role in mitochondrial genome maintenance. Curr Genet. 47, 84-99.

Mookerjee, S. A., Sia, E. A., 2006. Overlapping contributions of Msh1p and putative recombination proteins Cce1p, Din7p, and Mhr1p in large-scale recombination and genome sorting events in the mitochondrial genome of Saccharomyces cerevisiae. Mut Res. 595, 91-106.

Myers, S., Freeman, C., Auton, A., Donnelly, P., Mcvean, G., 2008. A common sequence motif associated with recombination hot spots and genome instability in humans. Nat Genetics. 40, 1124-1129.

Nakabeppu, Y., Kajitani, K., Sakamoto, K., Yamaguchi, H., Tsuchimoto, D., 2006. MTH1, an oxidized purine nucleoside triphosphatase, prevents the cytotoxicity and neurotoxicity of oxidized purine nucleotides. DNA Repair. 5, 761-772.

Narendra, D. P., Jin, S. M., Tanaka, A., Suen, D.-F., Gautier, C. A., Shen, J., Cookson, M. R., Youle, R. J., 2010. PINK1 is selectively stabilized on impaired mitochondria to activate Parkin. PLoS Bio. 8, e1000298.

Narendra, D. P., Tanaka, A., Suen, D.-F., Youle, R. J., 2008. Parkin is recruited selectively to impaired mitochondria and promotes their autophagy. J Cell Bio. 183, 795-803.

Nass, M. M., Nass, S., 1963. Intramitochondrial fibers with DNA characteriztics. I. Fixation and electron staining reactions. J Cell Bio. 19, 593-611.

Nilsen, H., Otterlei, M., Haug, T., Solum, K., Nagelhus, T. A., Skorpen, F., Krokan, H. E., 1997. Nuclear and mitochondrial uracil-DNA glycosylases are generated by alternative splicing and transcription from different positions in the UNG gene. Nucleic Acids Res. 25, 750-755.

Nishioka, K., Ohtsubo, T., Oda, H., Fujiwara, T., Kang, D., Sugimachi, K., Nakabeppu, Y., 1999. Expression and differential intracellular localization of two major forms of human 8-oxoguanine DNA glycosylase encoded by alternatively spliced OGG1 mRNAs. Mol Bio Cell. 10, 1637-1652.

Phadnis, N., Sia, R. A., Sia, E. A., 2005. Analysis of repeat-mediated deletions in the mitochondrial genome of *Saccharomyces cerevisiae*. Genetics. 171, 1549-59.

Pogorzala, L., Mookerjee, S., Sia, E. A., 2009. Evidence that msh1p plays multiple roles in mitochondrial base excision repair. Genetics. 182, 699-709.

Pohjoismaki, J. L. O., Wanrooij, S., Hyvarinen, A. K., Goffart, S., Holt, I. J., Spelbrink, J. N., Jacobs, H. T., 2006. Alterations to the expression level of mitochondrial transcription factor A, TFAM, modify the mode of mitochondrial DNA replication in cultured human cells. Nucleic Acids Res. 34, 5815-5828.

Polosina, Y. Y., Cupples, C. G., 2010. Wot the 'L-Does MutL do? Mut Res. 705, 228-38.

Ramotar, D., Kim, C., Lillis, R., Demple, B., 1993. Intracellular localization of the Apn1 DNA repair enzyme of *Saccharomyces cerevisiae*: Nuclear transport signals and biological role. J Biol Chem. 268, 20533-9.

Randolph-Anderson, B. L., Boynton, J. E., Gillham, N. W., Harris, E. H., Johnson, A. M., Dorthu, M. P., Matagne, R. F., 1993. Further characterization of the respiratory deficient *dum-1* mutation of *Chlamydomonas reinhardtii* and its use as a recipient for mitochondrial transformation. Mol Gen Genet. 236, 235-44.

Sage, J. M., Gildemeister, O. S., Knight, K. L., 2010. Discovery of a novel function for human Rad51: maintenance of the mitochondrial genome J Biol Chem. 285, 18984-18990.

Sanford, J. C., 1988. The biolistic process. Trends Biotech. 6, 299-302.

Sanford, J. C., Klein, T. M., Wolf, E. D., Allen, N., 1987. Delivery of substances into cells and tissues using a particle bombardment process. J Particulate Sci Tech. 5, 27-37.

Sasaki, N., Fukatsu, R., Tsuzuki, K., Hayashi, Y., Yoshida, T., Fujii, N., Koike, T., Wakayama, I., Yanagihara, R., Garruto, R., Amano, N., Makita, Z., 1998. Advanced glycation end products in Alzheimer's disease and other neurodegenerative diseases. Am J Path. 153, 1149-55.

Schofield, M. J., Brownewell, F. J., Nayak, S., Du, C., Kool, E. T., Hsieh, P., 2001. The Phe-X-Glu DNA binding motif of MutS. J Biol Chem. 276, 45505-45508.

Schon, E. A., Gilkerson, R. W., 2010. Functional complementation of mitochondrial DNAs: Mobilizing mitochondrial genetics against dysfunction. Biochim Biophys Acta. 1800, 245-249.

Schon, E. A., Rizzuto, R., Moraes, C. T., Nakase, H., Zeviani, M., DiMauro, S., 1989. A direct repeat is a hotspot for large-scale deletion of human mitochondrial DNA. Science. 244, 346-349.

Shokolenko, I., Venediktova, N., Bochkareva, A., Wilson, G. L., Alexeyev, M. F., 2009. Oxidative stress induces degradation of mitochondrial DNA. Nucleic Acids Res. 37, 2539-2548.

Sia, E. A., Butler, C. A., Dominska, M., Greenwell, P., Fox, T. D., Petes, T. D., 2000. Analysis of microsatellite mutations in the mitochondrial DNA of Saccharomyces cerevisiae. Proc Natl Acad Sci USA. 97, 250-5.

Singh, G., Maniccia-Bozzo, E., 1990. Evidence for lack of mitochondrial DNA repair following cis-dichlorodiammineplatinum treatment. Cancer Chemother Pharmacol. 26, 97-100.

Singh, K. K., Sigala, B., Sikder, H. A., Schwimmer, C., 2001. Inactivation of *Saccharomyces cerevisiae OGG1* DNA repair gene leads to an increased frequenncy of mitochondrial mutants. Nucleic Acids Res. 29, 1381-1388.

Smolińska, U., 1987. Mitochondrial mutagenesis in yeast: mutagenic specificity of EMS and the effects of *RAD9* and *REV3* gene products. Mut Res. 179, 167-74.

Spelbrink, J. N., 2009. Functional organization of mammalian mitochondrial DNA in nucleoids: History, recent developments, and future challenges. IUBMB Life. 62, 19-32.

Srivastava, S., Moraes, C. T., 2005. Double-strand breaks of mouse muscle mtDNA promote large deletions similar to multiple mtDNA deletions in humans. Hum Mol Gen. 14, 893-902.

Steele, D. F., Butler, C. A., Fox, T. D., 1996. Expression of a recoded nuclear gene inserted into yeast mitochondrial DNA is limited by mRNA-specific translational activation. Proc Natl Acad Sci USA. 93, 5253-7.

Strand, M. K., Copeland, W. C., 2002. Measuring mtDNA mutation rates in *Saccharomyces cerevisiae* using the mtArg8 assay. Methods Mol Biol. 197, 151-7.

Strauss, M., Hofhaus, G., Schröder, R. R., Kühlbrandt, W., 2008. Dimer ribbons of ATP synthase shape the inner mitochondrial membrane. EMBO J. 27, 1154-60.

Stuart, J., Brown, M., 2006. Mitochondrial DNA maintenance and bioenergetics. Biochimica et Biophysica Acta 1757, 79-89.

Studamire, B., Quach, T., Alani, E., 1998. *Saccharomyces cerevisiae* Msh2p and Msh6p ATPase activities are both required during mismatch repair. Mol Cell Bio. 18, 7590-7601.

Suen, D.-F., Narendra, D. P., Tanaka, A., Manfredi, G., Youle, R. J., 2010. Parkin overexpression selects against a deleterious mtDNA mutation in heteroplasmic cybrid cells. Proc Natl Acad Sci USA. 107, 11835-40.

Sugawara, N., Goldfarb, T., Studamire, B., Alani, E. A., Haber, J. E., 2004. Heteroduplex rejection during single-strand annealing requires Sgs1 helicase and mismatch repair proteins Msh2 and Msh6 but not Pms1. Proc Natl Acad Sci USA. 101, 9315-9320.

Swartzlander, D. B., Griffiths, L. M., Lee, J., Degtyareva, N. P., Doetsch, P. W., Corbett, A. H., 2010. Regulation of base excision repair: Ntg1 nuclear and mitochondrial dynamic localization in response to genotoxic stress. Nucleic Acids Res. 38, 3963-3974.

Thyagarajan, B., Padua, R. A., Campbell, C., 1996. Mammalian mitochondria possess homologous DNA recombination activity. J Biol Chem. 271, 27536-43.

Torello, A. T., Overholtzer, M. H., Cameron, V. L., Bonnefoy, N., Fox, T. D., 1997. Deletion of the leader peptide of the mitochondrially encoded precursor of *Saccharomyces cerevisiae* cytochrome c oxidase subunit II. Genetics. 145, 903-10.

Vermulst, M., Wanagat, J., Kujoth, G. C., Bielas, J. H., Rabinovitch, P. S., Prolla, T. A., Loeb, L. A., 2008. DNA deletions and clonal mutations drive premature aging in mitochondrial mutator mice. Nat Gen. 40, 392-4.

Wierdl, M., Dominska, M., Petes, T. D., 1997. Microsatellite instability in yeast: dependence on the length of the microsatellite. Genetics. 146, 769-79.

Yasui, A., Yajima, H., Kobayashi, T., Eker, A. P., Oikawa, A., 1992. Mitochondrial DNA repair by photolyase. Mut Res. 273, 231-236.

Yoon, Y. G., 2005. Transformation of isolated mammalian mitochondria by bacterial conjugation. Nucleic Acids Res. 33, e139-e139.

Yoon, Y. G., Koob, M. D., 2011. Toward genetic transformation of mitochondria in mammalian cells using a recoded drug-resistant selection marker. J Genet Genomics. 38, 173-179.

You, H. J., Swanson, R. L., Harrington, C., Corbett, A. H., Jinks-Robertson, S., Senturker, S., Wallace, S. S., Boiteux, S., Dizdaroglu, M., Doetsch, P. W., 1999. *Saccharomyces cerevisiae* Ntg1p and Ntg2p: broad specificity *N*-glycosylases for the repair of oxidative DNA damage in the nucleus and mitochondria. Biochemistry. 38, 11298-11306.

Zhang, H., Chatterjee, A., Singh, K. K., 2006. *Saccharomyces cerevisiae* polymerase zeta functions in mitochondria. Genetics. 172, 2683-8.

Zhou, J., Liu, L., Chen, J., 2010. Mitochondrial DNA Heteroplasmy in *Candida glabrata* after Mitochondrial Transformation. Eukaryotic Cell. 9, 806-814.

Site-Directed and Random Insertional Mutagenesis in Medically Important Fungi

Joy Sturtevant
LSUHSC School of Medicine
USA

1. Introduction

Site-directed and random mutagenesis have been useful tools in molecular biology. The application of directed mutagenesis in medically important fungi has been limited by the availability of molecular genetic techniques. Even species in which efficient genetic transformation methodologies exist, mutagenesis approaches were sparsely used due to diploidism. Lack of genetic tools hindered understanding of virulence mechanisms of medically important fungi. With the arrival of whole-genome sequencing, as well as improved techniques of genetic manipulation, the ability to address these questions is improving. A comprehensive review of mutagenesis in pathogenic fungi is outside the scope of this review, so not all studies were included. The intent of this review is to educate the reader on applications of site-directed and random insertional mutagenesis in medically important fungi in order to provide ideas for novel approaches to address major issues in pathogenic fungal research.

2. Site-directed mutagenesis

Site-directed mutagenesis has been exploited to understand signaling pathways, mechanisms of drug resistance, and identification of promoter DNA binding sites. Applications used less frequently have included protein localization and function of specific genes. In most instances, the site of the mutation was selected due to homology to model species or mammalian genes.

2.1 Signaling pathways

The most commonly reported application of site-directed mutants is the construction of dominant-negative and dominant-active alleles. The ability to make dominant-active alleles is particularly useful in diploid strains, since both endogenous alleles do not have to be disrupted. The amino acids chosen for mutation are often based on homology to *Saccharomyces cerevisiae* or *Aspergillus nidulans*. Although the roles of genes in signaling pathways were identified in model fungi, the regulation and downstream effects of these pathways are often very different in medically important fungi.

2.1.1 Phosphomimetics

The introduction of an amino acid substitution so the residue acts as constitutively phosphorylated or non-phosphorylated is a common technique to study cellular processes. Phosphomimetics have been used to study MAPK, cAMP-PKA, calcineurin, and two–component signaling, as well as cytokinesis and the heat shock response, in medically important fungi (Bockmühl and Ernst, 2001; Fox and Heitman, 2005; Hicks et al., 2005; Li et al., 2008; Menon et al., 2006; Nicholls et al., 2011).

The cAMP-PKA pathway regulates multiple cellular processes in eukaryotes. cAMP levels are regulated by phosphodiesterases (*PDE1*, *PDE2*), which in turn are regulated by protein kinase A (*PKA*) in some species. In *Cryptococcus neoformans*, the cAMP pathway is involved in multiple cellular processes, including virulence factor expression (melanin and capsule formation) (Hicks et al., 2005). Since the role in cAMP degradation and regulation by *PKA* of *PDE1* and *PDE2* differ among species, the goal of this study was to learn the functions of the PDEs in *C. neoformans*. In order to identify if *PDE1* was regulated by *PKA* through phosphorylation, a site-directed mutation in the *PDE1* at a putative *PKA* phosphorylation site was introduced based on work in *Saccharomyces*. Site-directed mutagenesis was performed by overlap PCR (Section 2.7), and the product was ligated into a *C. neoformans* transformation vector. The predicted outcome was an inactive *PDE1* and, thus, increased activation of the PKA pathway. In this way, the use of site-directed mutagenesis validated that PKA directly regulated the activation of *PDE1* in *C. neoformans* (Hicks et al., 2005).

Putative phosphorylated residues have not always been identified previously in model fungi. Consequently, *in silico* analysis can be utilized to identify putative phosphorylation sites (Bockmühl and Ernst, 2001; Li et al., 2008). *In silico* analysis predicted certain threonine residues as phosphorylation sites in the *Candida albicans* APSES protein Efg1p. These sites were mutated. Phenotypic analysis demonstrated that the mutations differentially affected morphogenesis, an important virulence attribute of *C. albicans* (Bockmühl and Ernst, 2001*).

Unlike previous studies, target residues in the two-component response regulator, Ssk1p, were identified by sequence comparison to a bacterial response regulator (Menon et al., 2006). Invariant aspartic acid residues were substituted using site-directed mutagenesis. This study demonstrated that phosphorylation of two different residues affects regulation of different cellular processes involved in virulence (Menon et al., 2006).

2.1.2 G protein signaling

Another common application for site-directed mutagenesis has been G protein signaling. In *Aspergillus fumigatus,* asexual sporulation results in release of spores that are inhaled by man, which can lead to serious manifestations. In the non-pathogen, *Aspergillus nidulans*, it was known that G protein signaling pathways were responsible for both vegetative growth and conidiation. Activation of *flbA* is required for conidiation, and this was probably through the activation of the GTPase activity of the G alpha protein FadA. Mah et al. (2006) used this framework to determine if similar regulation occurred in the pathogen *A. fumigatus* (Mah and Yu, 2006). They were able to confirm that Af*flb* regulated the G protein signaling through GpaA (homolog of FadA). Gene disruption and random chemical mutagenesis confirmed the role of Af*flb* in conidiation. Dominant-active and dominant-negative mutant alleles of *gpaA* (made by overlap PCR) demonstrated that it is a

downstream target in this pathway. Interestingly conidiation appeared even in the absence of Af*flb* (Mah and Yu, 2006), which may aid in dissemination.

As in *Aspergillus*, G protein signaling is also responsible for cellular differentiation in *C. albicans*. Much of the initial molecular dissection of the signaling pathways involved in morphogenesis was deciphered by constructing dominant-active and dominant-negative alleles of the G signaling proteins. Site-specific mutations were introduced in *CDC42, RAC1, GPA2, RAS1* and *RAS2* based on homology to *Saccharomyces* and mammalian G proteins (Bassilana and Arkowitz, 2006; Feng et al., 1999; Sanchez-Martinez and Perez-Martin, 2002; vandenBerg et al., 2004). The mutant alleles were introduced into exogenous loci under the expression of constitutive or regulatable promoters (Bassilana and Arkowitz, 2006; Feng et al., 1999; Sanchez-Martinez and Perez-Martin, 2002). However, since *CDC42* is essential, the mutated alleles were introduced at the endogenous locus in a *CDC42/cdc42* heterozygote (VandenBerg et al., 2004). These studies demonstrated the existence of different hyphal induction pathways, cross-talk between the MAPK and cAMP pathways, and distinction between growth and morphogenesis.

2.2 Mechanisms of drug resistance

Many fungal species present antifungal drug resistance *in vivo*. Studies have enhanced our understanding of this resistance. In most species, drug resistance is due to increased expression of export channels and/or mutations in target genes of the antifungal agent. Further studies confirmed the importance of the mutations.

The azoles interfere with ergosterol biosynthesis by targeting lanosterol 14 alpha-demethylase (*ERG11, CYP51*). Mutations in the gene resulted in reduced binding by the azole compound. Mutational hotspots were identified by sequencing the gene of interest from fungal strains isolated from patients or strains that have been passaged in the presence of the drug *in vitro*. It was then necessary to confirm that the mutation correlated with reduced susceptibility to the drug. Site-directed mutagenesis is an ideal method for validation. Due to homology of the *ERG11* gene among fungal species, many studies were performed in the more genetically malleable yeasts, *S. cerevisiae* or *Pichea pastoris* (Alvarez-Rueda et al., 2011). These studies confirmed that the mutated expressed protein is more or less susceptible to drug, but they do not definitively prove that the mutation was the reason for the clinical resistance. With the advent of genetic transformation techniques in medically important fungi, it is now possible to perform these experiments in the appropriate fungal host.

The most studied gene is *ERG11* in *C. albicans*, reviewed in Morio et al. (2010). Over 144 amino acid substitutions have been identified. It is less clear how many of these contribute to *in vivo* resistance. Additionally, some mutations may result from *in vitro* manipulations. A recent screen of azole-susceptible and resistant clinical isolates demonstrated that mutations are associated with both susceptibility and reduced susceptibility. Only 18% of isolates had no polymorphisms (Morio et al., 2010). These results highlighted the need to confirm that a specific mutation correlated with acquisition of resistance. In *C. albicans,* this has been approached by site-directed mutagenesis. Initial studies cloned the *ERG11* open reading frame (ORF) in a plasmid and then used PCR to introduce site-directed mutations. The PCR products were ligated into a *S. cerevisiae* expression vector. *S. cerevisiae* (azole

susceptible strain) was transformed with plasmids containing the mutated *C. albicans ERG11* gene and tested in a series of assays for reduced azole susceptibility. In this manner, azole resistance was correlated with specific amino acid substitutions (Kakeya et al., 2000; Lamb et al., 2000; Lamb et al., 1997; Sanglard et al., 1998; Sheng et al., 2010). Direct mutagenesis of the *ERG11* gene in *C. albicans* has not been reported. However, direct mutagenesis of the azole-target gene *cyp51A*, was performed in *A. fumigatus* (Mellado et al., 2007; Snelders et al., 2011). Two approaches were used. In the first study, mutated sequences were amplified by PCR from clinical isolates that demonstrated reduced susceptibility to itraconazole (Mellado et al., 2007). Previously it was known that itraconazole resistance correlated with specific amino acid mutations at G54 and M220 (See Table 1 in the chapter by Figurski et al. for the amino acid codes). However, this study identified a new mutation site (L98H) in conjunction with a duplication in the promoter sequence. In order to confirm the importance of these mutations, an azole-susceptible strain was transformed with the mutated allele; and transformants were plated on itraconazole (Mellado et al., 2007). Although the importance of the mutation sites were confirmed, the transformation selection criteria were not efficient. A second study expanded upon this approach and used 3-D modeling to determine a mechanistic reason for the azole resistance conferred by the mutations (Snelders et al., 2011). Specific amino acids were substituted in the *cyp51A* using the QuickChange XL Site-Directed Mutagenesis Kit (Stratagene) (Section 2.7). The appropriate PCR products were cloned into a vector that contained a hygromycin resistance marker and flanking sequences for introduction into the endogenous *cyp51A* site. Therefore, positive transformants were selected on hygromycin and further tested for azole resistance/susceptibility phenotypes. Consequently, the inclusion of a dominant selective marker improved the efficiency of screening transformants (Snelders et al., 2011). To further expand identification of potential mutation sites, the same experimental approach could be used by subjecting the *cyp51A* gene to random PCR-directed mutagenesis (Palmer and Sturtevant, 2004) and thereby identify new mutation sites that confer altered susceptibility.

2.3 Promoter response elements

Mutagenesis is a common approach to identify DNA binding sites in promoters. Nested deletions are probably the most commonly reported method used in the medically important fungi. Site-directed mutagenesis has been used to introduce point mutations in the *hapB* promoter in *A. nidulans*; so this approach may be used in *A. fumigatus* in the future (Brakhage and Langfelder, 2002). Site-directed mutagenesis has been used to identify putative promoter elements in chitin synthases (*CHS2, CHS8*) in *C. albicans* (Lenardon et al., 2009). The significance of this study is that chitin synthases are up-regulated in response to cell wall stress and thus are important for fungal survival. In this study, the promoters of *CHS2* and *CHS8* were mutated by site-directed mutagenesis or nested deletions. The selected sites were chosen due to previous studies or by *in silico* analysis. The mutated *Candida* promoter sequences were ligated upstream of the *Streptococcus thremophilus lacZ* gene. If the mutated site in the promoter element were important for a specific cell wall stress, *lacZ* would not be induced; and colonies would be white instead of blue on X-gal-containing medium. (X-gal is 5-bromo-4-chloro-indolyl-β-D-galactopyranoside.) The chosen potential sites reflected induction of several pathways and included known binding motifs. Mutations of individual promoter elements selected by *in silico* analysis had no effect on expression. Consequently, they performed nested deletions using exonuclease III digestion

of restriction-digested plasmids containing the *CHS2* and *CHS8* genes. These mutated promoters were introduced into *Candida* by transformation, and induction of the *lacZ* reporter was assayed after stresses. In this manner they were able to identify regions, but not specific regulatory elements, in the *CHS2* and *CHS8* promoters that responded to cell wall stressors (Lenardon et al., 2009). In order to determine which signaling pathways acted upon these promoters, the mutated *CHS2* and *CHS8* constructs were introduced into *C. albicans* signaling pathway deletion mutants. These studies demonstrated that the cell wall integrity, calcineurin, and HOG (osmotic sensing) pathways mediated expression through the *CHS2* promoter; only the cell wall integrity pathway affected *CHS8*-mediated expression (Lenardon et al., 2009). A pitfall of the deletion method is that it does not identify exact residues. It is possible to lose structural consistency, and it may delete other regulatory elements. In order to identify appropriate binding sites, random mutagenesis of *CHS2* and *CHS8* promoters by XL1-Red (a mutator strain of *Escherichia coli* useful for mutating cloned fragments) could have been performed (Palmer and Sturtevant, 2004). The ensuing transformants could be quickly screened on cell wall stressor media.

In another study, the authors wanted to identify the promoter binding sites in the pH responsive gene, *PHR1*, in *C. albicans* (Ramon and Fonzi, 2003). The pH response pathway had been well researched in *A. nidulans*. However, the promoter binding elements in the *Aspergillus* pH responsive gene (*pacC*) could not be translated to *PHR1*. Therefore, regions of DNA binding were identified *in vitro* by ChIP (Chromatin Immunoprecipitation Assay) and then confirmed by site-directed mutagenesis (Ramon and Fonzi, 2003).

Site-directed mutagenesis and ChIP have also been used to identify the genes that a specific transcription factor binds. A good illustration is the gain-of-function allele of the transcription factor *CAP1* that was constructed by site-directed mutagenesis and then analyzed in ChIP assays (Znaidi et al., 2009).

2.4 Gene function-essential genes

Surprisingly, site-directed mutagenesis has not been used extensively to determine the function of a gene. Gene disruption is routinely the method of choice to study gene function. However, this is not possible when studying the function of essential genes. The use of conditional promoters is often used. Results can sometimes be misleading, since phenotypic testing is performed under suboptimal growth conditions due to promoter-dependent nutritional constraints. Even so, expression of an essential gene under a conditional reporter does not allow complete analysis of multifunctional genes. Very few studies have taken advantage of directed mutagenesis of a specific gene.

The first report of the use of site-directed mutagenesis of an essential and multifunctional gene was the signaling regulatory gene, *BMH1* (14-3-3 gene). There is only one 14-3-3 protein (Bmh1p) in *C. albicans*, and it is essential (Cognetti et al., 2002). Multiple approaches were attempted to express the gene under a regulatable promoter, but they were unsuccessful (Palmer et al., 2004). This may be because *BMH1* regulates multiple cellular processes involved in growth, and the phenotypic studies were performed under suboptimal growth conditions due to promoter-dependent nutritional constraints. Therefore, *BMH1* appeared to be an excellent candidate to test the feasibility of both site-directed and random mutagenesis. Amino acid residues in the 14-3-3 allele required for

ligand binding, dimerization, and growth were reported for other eukaryotic species. Due to the high degree of conservation between 14-3-3 proteins, the same residues were selected for substitution in the *C. albicans BMH1* allele. Six sites were chosen. Transformants were screened for filamentation and growth defects (Palmer et al., 2004). Two approaches tested the applicability of random mutagenesis of the *BMH1* allele (Palmer and Sturtevant, 2004). A plasmid containing the *BMH1* allele was propagated in the *E. coli* XL-1 Red strain (Stratagene), which is deficient in multiple primary DNA repair pathways and thus introduces random mutations in the plasmid. Mutagenized plasmids were isolated after 11 to 44 divisions and introduced into the remaining *BMH1* locus in a *BMH1* heterozygote strain. The second random mutagenesis approach was PCR-mediated. (DNA polymerases used for PCR can be mutagenic under certain conditions.) The *BMH1* allele was subject to PCR amplification with an unbalanced nucleoside pool. The PCR products were ligated into a *Candida* transformation vector, and pools were introduced into the *BMH1* heterozygote by transformation, as above. The *E. coli*-mediated mutagenesis resulted in a higher efficiency of correct integration of the mutated allele than did the PCR-mediated method. Around 1400 (1000 – *E. coli*-mediated; 368 – PCR-mediated) *C. albicans* transformants containing randomly mutagenized *BMH1* alleles were screened under a variety of phenotypic stresses. These tests were rapid and easily visible; thus, they translated easily into a screen. Mutant alleles were isolated from transformants that demonstrated altered phenotypes and were sequenced. In the end, from 1000 *E. coli*- and 368 PCR-mutated colonies, 2 and 4 alleles, respectively (0.4%), were identified with altered coding sequences. That these mutations were responsible for the altered phenotypes was validated by constructing *C. albicans* strains isogenic for the site-directed mutations, as described previously. While the efficiency of the random mutagenesis methods was lower than reported for bacteria, non-lethal mutants were identified. Thus, this is a valid approach to study gene function in fungi (Palmer and Sturtevant, 2004). The outcome of the site-directed and random mutagenesis approaches was a set of isogenic strains in which *BMH1* or a mutant *BMH1* allele was expressed under its own promoter at an exogenous locus. These strains were analyzed under a variety of environmental conditions reflecting stresses in the host. It was possible to discriminate between separate pathways involved in filamentation, growth, and survival in the host (Kelly et al., 2009; Palmer et al., 2004; Palmer and Sturtevant, 2004). Additional mutants may have been identified if transformants were screened in additional tests or in *in vivo* models. On the other hand, since *BMH1* is an essential gene, there may be a limited number of amino acids that can be mutated and still result in a non-lethal allele.

Site-directed mutagenesis was also used to decipher the role of the hemoglobin response gene (*HBR1*) in vegetative growth. It was known that *HBR1* induced mating type genes, but mating is not an essential process in *C. albicans* (Peterson et al., 2011*)*. Sites required for optimal growth and the oxidative stress response in the homologous gene in *Saccharomyces* were targeted in the *C. albicans* gene. The mutant alleles were introduced into a *HBR1* heterozygote and were regulated by the *MET3* promoter. This study identified amino acid residues important for mating locus regulation, but not for vegetative growth. Thus, amino acids identified to be important in model fungal species do not always translate to related pathogenic fungi (Peterson et al., 2011).

Essential genes are often prospective drug targets. One such gene is *MET6*. In *C. albicans*, Prasannan et al. (2009) constructed GST fusions of mutated *C. albicans MET6* and expressed the fusion protein in a *MET6 Saccharomyces* mutant. A 3-D model that was modified from

the known crystallized structure of the *Arabidopsis* enzyme was used to select sites for mutation in *MET6*. Eight residues were chosen based on conservation across species and probability of being catalytic sites. Site-directed mutagenesis was introduced by the Quikchange kit from Stratagene. The mutant GST-Met6p fusion proteins demonstrated varied enzymatic activity validating the use of this approach in the design of new antifungal drugs (Prasannan et al., 2009).

2.5 Gene function – genes with multiple functions

The transcriptional regulator, *EFG1*, regulates multiple cellular processes in *C. albicans*. *EFG1* is a member of the APSES protein family. Although the domain that defines this family is known, the actual structure-function relationships were not understood. Thus, defined regions within and flanking the APSES domain were deleted. This was mediated by PCR. Instead of amino acid substitution, 15 – 103 nucleotides were deleted from within the *EFG1* gene, similar to what is done for promoter bashing (mutating promoters) (Noffz et al., 2008). A disadvantage to this approach is that it is not possible to discriminate if an altered phenotype is due to the compromise of protein structure or to the absence of protein expression. However, immunoblotting confirmed that the mutated protein was expressed in all mutants. Two mutants did express lower levels of Efg1p that could account for altered phenotypes (Noffz et al., 2008). Thus, it is important to confirm protein expression of mutants. The authors were able to associate specific regions of Efg1p with distinct cellular processes that it regulates. The deletion alleles were also used in over-expression and one-hybrid (gene-fusion technology to identify a DNA-binding domain) experiments. Thus, this approach was successful in determining structure – function relationships of an APSES protein (Noffz et al., 2008).

2.6 Other applications

Site-directed mutagenesis has been used to determine how GPI-tagged proteins discriminate between localization to the plasma membrane and cell wall (Mao et al., 2008). N and C termini of cell wall or plasma proteins were fused to GFP. The termini were subjected to truncation and mutagenesis. Localization of mutant alleles was examined by microscopy. One potential pitfall, however, is that the GFP tag itself can cause protein mislocalization. These experiments identified the omega cleavage site. Further domain exchange and mutagenesis studies identified which residues dictated cell wall or plasma membrane localization (Mao et al., 2008).

2.7 Methodology

Site-directed mutagenesis (*i.e.*, targeted substitution of one or more nucleotides) in a gene was normally performed via overlap PCR and/or the QuikChange Site-Directed Mutagenesis Kit (Stratagene/Agilent Technologies). The principle of these methods is the same. Complementary primers are designed with the nucleotide substitution at the desired site of the mutation. The primers are complementary to the region of the template with the wild-type residue. The template is a double-stranded DNA vector (usually a plasmid) containing a DNA clone of the region of interest. PCR with a high fidelity polymerase results in a plasmid with the mutation of the primer. The product is digested with *Dpn*I,

which cleaves only the parental plasmid (template) because *Dpn*I requires fully or hemi-methylated DNA. (The parental plasmid is methylated by the *E. coli* host; DNA amplified by PCR is unmethylated.) The resulting DNA is then introduced into competent cells by transformation. Resulting plasmids are sequenced to confirm the mutation. It is also important to confirm that the mutation does not affect gene expression. Single-site mutagenesis has also been used to introduce silent mutations that result in construction of a restriction enzyme site in order to facilitate genetic manipulation (Cognetti et al., 2002; Schmalhorst et al., 2008).

3. Insertional mutagenesis

Insertional mutagenesis methods are commonly used in model fungi species. Although genomes are similar between model and medically important fungal species, there are still significant differences. Forward screens (screens for new genes that are involved in a phenotype, often using homologs) in model fungi will not identify genes important for pathogenesis, since these species are usually attenuated in virulence or are avirulent. Signaling pathways are shared among fungi, but downstream targets and regulation vary. It is estimated that only 61% of the essential genes in *S. cerevisiae* are also essential in *C. albicans*. There may be even more differences in filamentous fungi (Carr et al., 2010). The advent of improved genetic techniques and whole-genome sequencing has dramatically improved the ability to perform forward screens in the medically important fungi. One major drawback has been diploidism. Ways to circumvent the problem of diploidy have included parasexual genetics (non-meiotic conversion of a diploid to a haploid) (Carr et al., 2010; Firon et al., 2003) and haploid insufficiency (a phenotype resulting from the loss of one allele in a diploid) (Uhl et al., 2003). Additional requirements that are species–specific include a 'mutagen' and an appropriate screen/phenotype. Insertional mutagenesis is normally now facilitated by transposons, but it is still necessary to identify transposons that work efficiently in the fungal species of choice. Much of the initial work demonstrated a bias for insertions, including a bias of non-coding regions. A recent analysis of three transposons has identified Tn7 as having the least insertion bias in *Candida glabrata* (Green et al., 2012). This would probably translate to other fungal species whose genomes are also rich in A/T sequences. Certainly, in the post-genomics era, utilization of forward genetics approaches have increased due to the improved ability to identify the site of insertion.

3.1 Selection of insertion mutants by complementation of auxotrophy

Initial studies used complementation of auxotrophy as a 'mutagen.' Auxotrophic strains were transformed with plasmids carrying an auxotrophic marker (*e.g.*, *URA3/5*). For example, in *C. neoformans*, capsule formation is associated with virulence. Laccase is required for capsule formation. To identify the laccase gene, a *ura*-deficient mutant was transformed multiple times (to obtain independent mutants) with a *URA* construct that has an *E. coli*-specific replicon. When expressed in *C. neoformans,* the construct integrates randomly into the genome and complements the uracil auxotrophy. Transformants were selected for growth on medium lacking uracil. They were then screened on differential media that would identify strains with laccase deficiency due to a pigment change. Out of 1000 transformants, nine strains with an altered phenotype were identified. Plasmid rescue was performed to identify the insertion point. (Plasmid rescue results from cleaving

genomic DNA with the appropriate restriction enzyme. The inserted fragment, along with a piece of the interrupted gene, is released. The released DNA can circularize in the presence of ligase and form a plasmid that replicates in *E. coli.* Sequencing of the piece of interrupted gene is easily done and identifies the gene, which can then be cloned intact.). In this manner, a novel virulence attribute was identified, the vacuolar (H+) – ATPase subunit (*VPH1*) (Erickson et al., 2001). In general, the drawbacks to the auxotrophic approach were inefficient integration, integration via homologous rather than non-homologous recombination, and difficulty in identification of the insertion site.

3.2 Signature-tagged mutagenesis (STM)

Signature-tagged mutagenesis is a method originally designed to identify genes required for pathogenesis (Hensel et al., 1995). A large number of mutants were created by insertional mutagenesis. The inserted DNA includes a unique oligonucleotide tag that resembles a 'barcode.' In principle, up to 96 mutants can be inoculated into one host; strains not recovered are thought to harbor a mutation specific for *in vivo* growth (Hensel et al., 1995). This method was first used in *Salmonella* and was modified for *C. glabrata, A. fumigatus,* and *C. neoformans* (Brown et al., 2000; Cormack et al., 1999; Nelson et al., 2001). There were certain considerations in translating this approach to fungi, including larger genomes, non-coding DNA, inefficient methods for insertion, selection of the appropriate host environment (Brown et al., 2000), and inoculation parameters (Nelson et al., 2001). These issues were addressed in the studies below (Brown et al., 2000; Nelson et al., 2001).

The first studies were performed prior to the identification of useful transposons. In order to identify virulence factors in *A. fumigatus,* two approaches were used to address random insertion of signature tags (Brown et al., 2000). The first used restriction-mediated integration (REMI). Protoplasts of the recipient strain were transformed with clones with tags in the presence of the restriction enzyme *Kpn*I (96 transformations). The rationale was that these clones would integrate into *Kpn*I sites randomly situated in the genome. The construction of the second library relied on ectopic integration, and *Aspergillus* was transformed with linearized clones (84 transformations). The tags for the transformation constructs for both approaches were generated by PCR using templates developed for *Salmonella typhimurium* and cloned into a fungal transformation vector that carries a gene for hygromycin resistance (Brown et al., 2000). A similar approach was used for *Cryptococcus neoformans*, and the selection of insertions was based on ectopic integration of a linear plasmid conveying hygromycin resistance (Nelson et al., 2001). Further analysis demonstrated that integration was mostly random, except for one hotspot that was the actin/*RPN10* promoter. In both cases, integration efficiency was lower than reported for bacteria.

Many of the medically important fungi can cause different types of infections and/or colonize and infect multiple organs. Unlike bacteria, they do not have true 'virulence factors'; but they do have virulence "attributes." Since *in vivo* murine models are involved, it is important to limit the number of mice used; and thus it is necessary to predetermine the appropriate model, the time points and the organs to harvest. For *Aspergillus fumigatus*, the STM libraries were tested in an immunosuppressed murine inhalation model (Brown et al., 2000). For *C. neoformans*, Nelson et al. (2001) carefully determined the course of infection in a murine model and chose a time point that reflected attenuated or increased virulence based

on cfu (colony forming units) counts in the brain (Nelson et al., 2001). They also asked an interesting question: Would a virulent strain allow survival of an attenuated strain? For instance, if the virulent strain damaged endothelium, normally avirulent strains might theoretically have increased abilities to disseminate. They tested this by co-infecting with acapsular (avirulent) and capsular (virulent) strains. The avirulent acapsular strains were not recovered, and they concluded that virulent strains would not help avirulent strains (Nelson et al., 2001). However, this may not be true for all attenuated strains; and a strain's ability to piggyback upon another will depend on its defect. This is a general drawback of STM and confirms that virulence tests with single strains have to be performed.

Another important parameter is the number of strains that can be injected into a mouse and have an equal opportunity to survive. Nelson et al. (2001) did a prescreen with hygromycin and G418 resistant strains (100:1) and ascertained that it was possible to inoculate 100 strains. However, studies with hybridization signals showed that they could not reliably detect more than 80 strains. Experiments were performed with pools of 48 strains. Six hundred seventy-two mutants were screened, and 39 gave different output signals. Twenty-four of the mutants were tested singly in the mouse, and 6 of these had significant changes in virulence (Nelson et al., 2001). Brown et al. (2000) determined that subsequent hybridization efficiency was 80%, so, although they used pools of 96, they always inoculated 2 mice per pool (Brown et al., 2000). In total 4648 tagged strains were screened, and 35 strains (0.8%) gave weak signals in the output pool after two rounds of STM. These strains were tested in a competitive inhibition infection, in which the attenuated strain was present as 50% of the inoculum. Nine strains showed a competitive disadvantage, and two of these demonstrated significantly reduced virulence. The site of the mutation of one strain was not identifiable; the second mutation was upstream of the PABA synthetase gene. Further analysis confirmed that *pabaA* is required for virulence.

Cormack et al. (1999) exploited STM to construct a mutant library in *C. glabrata* (Cormack et al., 1999). Each strain could be easily identified by a distinct tag. Ninety-six unique strains were generated by integrating 96 different tags, flanked by identical primer sites, into the already disrupted *URA3* locus. Since *C. glabrata* has an efficient system of non-homologous recombination, the *Saccharomyces URA3* gene was used for random mutagenesis. Transformants were selected on media minus uracil. Pools of the 96 tagged strains were screened for adherence to human cultured epithelial cells. Out of 4800 mutants (50 pools of 96), 31 mutants demonstrated aberrant adherence. Sixteen of these were non-adherent. Interestingly, 14/16 of these integrated into non-coding sequence upstream of the same gene, *EPA1*. This led to the identification of subtelomeric transcriptional silencing (Cormack et al., 1999). However, this method would not have identified *EPA1* by traditional STM, since *EPA1* null mutants are virulent *in vivo*.

3.3 Transposon-mediated insertional mutagenesis

Transposon technology has been used in pathogenic fungi to construct libraries, add epitope tags, and understand cellular processes. The technology has been adapted for diploid organisms using the parasexual cycle, haploid insufficiency, and homologous recombination (Carr et al., 2010; Davis et al., 2002; Firon et al., 2003; Juarez-Reyes et al., 2011; Spreghini et al., 2003; Uhl et al., 2003). The use of transposons has superseded auxotrophic and STM approaches.

Essential genes are often considered good drug targets. Firon et al. (2003) exploited the parasexual cycle to develop a transposon-mediated insertional mutagenesis protocol to identify essential genes in *A. fumigatus* (Firon et al., 2003). A diploid strain, homozygous auxotrophic for pyrimidines and heterozygous for a spore color marker, was randomly mutagenized with an *imp160*::*pyrG* transposon. The candidate mutant strains were induced to become haploid by the mitochondrial destabilizer, benomyl. The genotype of the parent strain allowed haploid progeny to be identified by pigmentation. Diploid strains were grey-green, but haploid progenies were white or reddish. Replica plating identified the haploid progeny that harbored transposons. If haploid strains carried a transposon-inactivated allele, they expressed pyrG and grew on both selective (without uridine/uracil) and non-selective media. Conversely, strains without a transposon grew only on non-selective media. If the transposon inactivated an essential gene, the haploid strain did not grow on either medium. With this approach, 3% of the haploid progeny of 2,386 diploid strains were found to be unable to grow on either medium and, therefore, possibly had mutations in essential genes. These strains were propagated further on selective media and haploid progeny could not be obtained from 1.2% of the resultant diploid revertants. The sites of insertion were determined by 2-step PCR using semi-random primers and 5'-end transposon-specific primers (see Section 3.5.1). Ninety percent of insertion sites were identified (Firon et al., 2003). Since the insertion rate of the transposon into essential loci was lower than expected, additional transposon insertion sites were analyzed. Although an insertion site did not depend on genome sequence or chromosomal location, there did appear to be a bias toward noncoding regions (34%) (Firon et al., 2003). Carr et al. (2010), who observed that transposon mobilization could be induced at 10 °C, improved upon this approach. Therefore, using the same screen, 96 additional essential loci were identified. They found no obvious bias of insertion in noncoding regions. Interestingly, only half of the genes had essential homologs in *Saccharomyces,* confirming the necessity for species-specific screening.

Uhl et al. (2003) developed a transposon mutant library in *C. albicans*. Restriction enzyme-digested *C. albicans* gDNA (genomic DNA) was mixed with a linearized donor transposon Tn7–containing plasmid. This plasmid harbored elements for replication in *E. coli*, for selection in both *E. coli* and *C. albicans* and a fungal *lacZ* reporter system. The fragments were ligated and introduced into *E. coli* by transformation. Plasmids were isolated from over 200,000 transformants and batch isolated. Transposon–gDNA junctions were sequenced in plasmids to confirm random integration. *C. albicans* was transformed with the Tn7-gDNA plasmids to give an 18,000-strain transposon mutant library. It was assumed that each strain had an independent insertion. That would mean there was a transposon approximately every 2.5 kb. However, only one allele of a gene was disrupted in these strains. (*C. albicans* is diploid, so one allele remains non-disrupted.) Uhl et al. (2003) exploited haploid insufficiency to screen for filamentation mutants, since heterozygote strains in genes involved in morphogenesis exhibit reduced filamentation. This screen was rapid and successful for identifying processes that required genes sensitive to dosage effects. However, this certainly will not be the case for all genes involved in pathogenesis.

Davis et al. (2002) constructed a transposon mutant library in *C. albicans*, but these strains harbored insertions in both alleles. This approach was based on a homologous recombination model that allowed the disruption of both alleles of *C. albicans* in one transformation step (Enloe et al., 2000). A cassette (UAU), which contains the *URA3* gene

disrupted with a functional *ARG4* gene, was inserted into transposon Tn7. the Tn7–UAU transposon was inserted randomly into a *C. albicans* library. Digestion with the appropriate restriction enzyme released DNA fragments that contained *C. albicans* DNA interrupted with the Tn7-UAU transposon. These fragments were used to transform *C. albicans*. Homology from the interrupted DNA allowed replacement of the chromosomal wild-type version by homologous recombination. The chromosomal version was then mutated because it carried the Tn7-UAU transposon. Recombinants could be selected because transformation into the recipient ura- arg- *Candida* strain will confer arginine prototrophy. Occasionally the other intact copy of the gene acquired the transposon. Thus, both copies of the gene were mutated. Using arginine selection, homozygous mutants could not be distinguished from the heterozygotes. However, in a small percentage of ARG+ transformants, the *ARG4* gene is spontaneously looped out. If there were two copies of the transposon and if looping out occurred in one, it gave an ARG+, URA+ strain. Thus, both alleles were disrupted. This allowed for the construction of a large set of mutants, though it was still not as efficient as it would be for a haploid strain. This library is widely used by the *Candida* community (Davis et al., 2002; Norice et al., 2007; Park et al., 2009).

Spreghini et al. (2003) exploited transposon mutagenesis to add an epitope to the putative cell wall protein, Dfg5p. Since conventional epitope tagging of amino and carboxyl termini was not an option, they wanted to identify an internal site which, when disrupted with a tag, did not compromise function. The Tn7 transposon was used to mutagenize the *DFG5* insert in a plasmid and insertions within *DFG5* coding region were confirmed by sequencing. Then the mutagenized plasmid was redigested to get rid of the majority of Tn7, leaving only a 15-bp (base pair) insertion, which resulted in an insertion of 5 amino acids that did not disrupt function and could be recognized by an available antibody. The internally tagged *DFG5* insert was then ligated into *C. albicans* vectors for further study (Spreghini et al., 2003).

In the transposon examples above, mutagenesis was performed *in vitro*, and then mutagenized DNA was introduced into recipient strains. Magrini and Goldman (2001) took a different approach by directly mutagenizing *Histoplasma capsulatum in vivo*. The transformation cassette was a linear telomere vector (because the presence of a telomeric sequence is required for efficient homologous recombination in *Histoplasma*) containing the selection marker *URA5*, the *MOS1* transposase gene regulated by a strong promoter, and the hygromycin resistance gene flanked by *MOS1* terminal repeats to create a synthetic transposon. *Histoplasma* transformants were selected in presence of 5-FOA (5-Fluoroorotic acid, which selects against URA5) to select for loss of the donor plasmid and on hygromycin for the presence of the synthetic transposon, which encodes hygromycin resistance. It is not known if this library has been utilized because T-DNA appears to be more commonly used in *Histoplasma* (see below).

A novel use of random insertion was the analysis of subtelomeric silencing of *C. glabrata* adhesin genes. Learning where silencing occurred was accomplished by randomly placing a *URA3* reporter at different distances from a telomere and examining where *URA3* was silenced. The transposon Tn7–*URA3* was introduced into a subtelomeric sequence of *C. glabrata* cloned on an *E. coli* plasmid. Resulting constructs were integrated into subtelomeric regions of *C. glabrata* by homologous recombination. It was possible to select for 'silenced' *URA3* on 5-FOA media (Juarez-Reyes et al., 2011).

3.4 *Agrobacterium* T-DNA

Agrobacterium tumefaciens carries an approximately 200-kbp (kilobase pair) tumor-inducing (Ti) plasmid. A portion of this plasmid is called T-DNA. In plants, the T-DNA randomly inserts in the genome; and the outcome is a tumorous growth. This plasmid has been modified for genetic manipulation purposes to retain the insertional DNA (T-DNA). The plasmid vector can also replicate in *E. coli* and has cloning sites for additional DNA. T-DNA has been used to construct mutants with increased, reduced, or no expression of genes, depending on the plasmid used (Krysan et al., 1999). In the last decade, insertional mutagenesis via T-DNA has been successfully adapted for medically important fungi. In general, a fungal selectable marker is ligated into the *Agrobacterium* Ti plasmid within the T-DNA region and introduced into *A. tumefaciens* by electroporation. Equal concentrations of *A. tumefaciens* carrying the delivery plasmid and target fungal strain are incubated together for varying lengths of time under conditions that mimic plant wound conditions, which are accomplished by low pH and the addition of acetosyringone. The T-DNA is transferred to the target organism by a conjugation-like mechanism. A mutant that contains an insertion of T-DNA is selected with the appropriate fungal selective marker.

Prior to T-DNA mutagenesis, insertional mutagenesis was attempted by electroporation or biolistic transformation of naked DNA. Researchers have developed protocols that have improved the efficiency of transformation using T-DNA in *C. neoformans* (Idnurm et al., 2004), *Histoplasma* and *Blastomyces* (Brandhorst et al., 2002; Edwards et al., 2011; Gauthier et al., 2010; Laskowski and Smulian, 2010; Marion et al., 2006; Smulian et al., 2007; Sullivan et al., 2002). In addition, T-DNA mutagenesis protocols have been developed for *Coccidioides* (Abuodeh et al., 2000), *Trichoderma* spp. (Cardoza et al., 2006; Dobrowolska and Staczek, 2009; Yamada et al., 2009), and *Penicillium marneffei* (Kummasook et al., 2010; Zhang et al., 2008). In *C. neoformans*, the use of T-DNA improved both the efficiency and the stability of transformation events. The resulting transformants also demonstrated less complicated integrations and less additional gene rearrangements. There did seem, however, to be a bias for promoter sequences. In one study, some of the integration events were not linked to *NAT*, the gene for the Nourseothricin resistance marker on the inserted DNA (Idnurm et al., 2004).

Blastomyces, in particular, is a challenge to transform, since it is multinucleate. Transforming DNA often integrates at multiple sites (Brandhorst et al., 2002). This is usually bypassed by transforming conidia or performing multiple rounds of selection to enrich for homokaryons (all the nuclei are genetically identical). Sullivan et al. (2002) developed a protocol for both *Histoplasma* and *Blastomyces*. Many conditions were tested, including bacteria:yeast ratios, life stage of the recipient strain, and the choice of selectable marker. Interestingly, the efficiency of transformation was 5–10 times higher with uracil selection than with hygromycin selection. Southern analysis confirmed that integration was random, but there were often direct repeat concatemers in *Blastomyces.* There were clear improvements over electroporation, including increased efficiency, ability to use spores as the recipient, and single-site integrations (Sullivan et al., 2002). Additional studies in *Blastomyces* using T-DNA have identified genes involved in phase transition (Gauthier et al., 2010).

T-DNA was used to identify genes in *Histoplasma* involved in pathogenesis in a novel high-throughput macrophage-killing screen (Edwards et al., 2011). Transgenic (a novel gene was introduced) macrophage lines were constructed that constitutively expressed bacterial *lacZ*.

The activity of β- galactosidase, the product of *lacZ*, directly correlated to the number of macrophages. Thus, this line was used as a readout for macrophage killing. Over 2000 *Histoplasma* transformants made from *A. tumefaciens*-treated *Histoplasma* cells were incubated with macrophages and screened for killing activity after 7 days. Three strains were less efficient in killing, and one was significantly inefficient in killing both transgenic and primary macrophages. Flanking sequences were identified by PCR and sequencing. The authors identified a new virulence gene in *Histoplasma*, a homolog of Hsp82 (Edwards et al., 2011).

Marion et al. (2006) performed a more comprehensive analysis of insertional mutagenesis in *Histoplasma capsulatum* using *Agrobacterium*-mediated transformation. Optimal co-incubation times, bacteria:yeast ratios and temperature were determined. Southern hybridization analysis showed that approximately 90% of the insertions were random and at a single site. Inverse PCR and plasmid rescue were used to identify the flanking sequences. Their results indicated that mutagenesis by T-DNA resulted in the absence of chromosomal rearrangements and deletions. The biological relevance of the T-DNA mutants was approached by screening for genes involved in the biosynthesis of α-(1, 3)–glucan, which is posited to be a virulence attribute. The absence of α-(1, 3)–glucan was easily visualized since colonies have a smooth, rather than rough, morphology. Approximately 50,000 insertional mutants were screened, and 25 had smooth morphology. Eighty-eight percent had single insertions and reduced α-(1, 3)–glucan. Five of twenty-two had distinct insertions in the α-(1, 3)–glucan synthase gene (AGS1), which validated their screen. RNAi technology (synthetic inhibitory RNA) was used to confirm the insertion mutant phenotype with the wild-type allele. The phenotypes of the two other mutants were confirmed. One mutation was in *UGP1* (previously reported to play a role in glucan synthesis). The other mutation was in the amylase gene, which was previously unreported to play a role (Marion et al., 2006).

The use of T-DNA in *Histoplasma* has provided additional information. As with all genetic manipulations, it is important to confirm that the mutation is responsible for the ensuing phenotype. Smulian et al. (2007) wanted to make GFP-expressing strains and used hygromycin resistance as a marker and T-DNA as the tool for integration. It turned out that all the transformants were hypervirulent. Site-directed mutagenesis of the hygromycin resistant gene, *hph*, confirmed that the increased virulence was due to the acquisition of hygromycin resistance. One mutant actually gained the ability to form cleistothecia, a mating structure that was not present in the parent strain. This phenotypic trait was not due to the *hph* gene; and, thus, the strain may be used as a tool to study mating in *Histoplasma* (Laskowski and Smulian, 2010).

3.5 Methodologies to identify the site of insertion

3.5.1 Two-step PCR

In the first step of two-step PCR (Chun et al., 1997), sequence on one side of the insertion site is amplified with a degenerate primer and a primer homologous to the sequence in one of the ends of the inserted DNA. (There are two end-specific primers. A primer specific for only one end is used. Note that a tranposon can insert in either orientation.) The degenerate primer contains 20 nucleotides of defined sequence at the 5'-end, 10 nucleotides of

degenerate sequence (*i.e.*, all 4 nucleotides are used at each position for synthesis) + GATAT at the 3'-end. The sequence GATAT is predicted to occur every 600 bp in the yeast genome. The second step amplifies the first PCR product with two non-degenerate primers. The forward primer contains the 20 nt (nucleotides) of defined sequence in the degenerate primer. The reverse primer is immediately 3' (antisense strand) to the insertion-specific primer used in the first PCR reaction to guarantee that the desired DNA is amplified (Chun et al., 1997). This method was originally defined in *Saccharomyces* and was successfully used to identify transposon insertions in *A. fumigatus* (Carr et al., 2010; Firon et al., 2003).

3.5.2 Thermal asymmetric interlaced PCR (TAIL PCR)

TAIL PCR (Liu and Whittier, 1995) is another method to identify sequences flanking insertions. It is a modified version of hemispecific (one-sided) PCR. The purpose is to favor amplification of the desired product. It uses specific primers homologous to DNA in the integrating cassette or plasmid and a degenerate primer that can anneal to the gDNA flanking the insertion. The strategy is that the specific primers are long, nested, and have a high Tm; the degenerate primer is short and has a low Tm. The first five cycles are high stringency cycles to favor annealing to and linear amplification from the specific primer. Then there is one low stringency cycle to allow the degenerate primer to anneal. Because there are now several copies of the gDNA adjacent to the insertion, the chance of the degenerate primer annealing to the desired product is increased. However, other products might form from the primers finding additional annealing sites in the genome. Using a second and a third primer completely homologous to the inserted DNA will favor the desired product that is made from both the specific and degenerate primers instead of either one alone. This is accomplished by interlacing reduced stringency and high stringency cycles.

4. Closing remarks

Site-directed and insertional mutagenesis are techniques that can be used to advance our understanding of the pathogenesis of medically important fungi. The exploitation of these tools has resulted in a better understanding of drug-resistant mechanisms, transcription factors, signaling pathways and vital cellular processes. Site-directed mutagenesis could be better utilized to decipher the functions of essential and multi-functional genes. While all approaches cannot be used in the always-diploid strains, transposon-mediated insertional mutagenesis can be used to construct libraries. Additionally, T-DNA can be used to improve transformation efficiency in dimorphic fungi and in *C. neoformans.*

5. References

Abuodeh, R. O., Orbach, M. J., Mandel, M. A., Das, A., Galgiani, J. N., 2000. Genetic Transformation of *Coccidioides immitis* Facilitated by *Agrobacterium tumefaciens*. Journal of Infectious Diseases. 181,6: 2106-2110.

Alvarez-Rueda, N., Fleury, A., Morio, F., Pagniez, F., Gastinel, L., Le Pape, P., 2011. Amino Acid Substitutions at the Major Insertion Loop of *Candida albicans* Sterol 14alpha-Demethylase Are Involved in Fluconazole Resistance. PLoS ONE. 6,6: e21239.

Bassilana, M., Arkowitz, R. A., 2006. Rac1 and Cdc42 Have Different Roles in *Candida albicans* Development. Eukaryotic Cell. 5,2: 321-329.

Bockmühl, D. P., Ernst, J. F., 2001. A Potential Phosphorylation Site for an A-Type Kinase in the Efg1 Regulator Protein Contributes to Hyphal Morphogenesis of *Candida albicans*. Genetics. 157,4: 1523-1530.

Brakhage, A. A., Langfelder, K., 2002. MENACING MOLD: The Molecular Biology of *Aspergillus fumigatus*. Annual Review of Microbiology. 56,1: 433-455.

Brandhorst, T. T., Rooney, P. J., Sullivan, T. D., Klein, B. S., 2002. Using new genetic tools to study the pathogenesis of *Blastomyces dermatitidis*. Trends in Microbiology. 10,1: 25-30.

Brown, J. S., Aufauvre-Brown, A., Brown, J., Jennings, J. M., Arst, H., Jr., Holden, D. W., 2000. Signature-tagged and directed mutagenesis identify PABA synthetase as essential for *Aspergillus fumigatus* pathogenicity. Mol Microbiol. 36,6: 1371-1380.

Cardoza, R. E., Vizcaino, J. A., Hermosa, M. R., Monte, E., Gutierrez, S., 2006. A comparison of the phenotypic and genetic stability of recombinant *Trichoderma* spp. generated by protoplast- and *Agrobacterium*-mediated transformation. J Microbiol. 44,4: 383-395.

Carr, P. D., Tuckwell, D., Hey, P. M., Simon, L., d'Enfert, C., Birch, M., Oliver, J. D., Bromley, M. J., 2010. The Transposon impala Is Activated by Low Temperatures: Use of a Controlled Transposition System To Identify Genes Critical for Viability of *Aspergillus fumigatus*. Eukaryotic Cell. 9,3: 438-448.

Chun, K. T., Edenberg, H. J., Kelley, M. R., Goebl, M. G., 1997. Rapid Amplification of Uncharacterized Transposon-tagged DNA Sequences from Genomic DNA. Yeast. 13,3: 233-240.

Cognetti, D., Davis, D., Sturtevant, J., 2002. The *Candida albicans* 14-3-3 gene, BMH1, is essential for growth. Yeast. 19,1: 55-67.

Cormack, B. P., Ghori, N., Falkow, S., 1999. An adhesin of the yeast pathogen *Candida glabrata* mediating adherence to human epithelial cells. Science. 285,5427: 578-582.

Davis, D. A., Bruno, V. M., Loza, L., Filler, S. G., Mitchell, A. P., 2002. *Candida albicans* Mds3p, a Conserved Regulator of pH Responses and Virulence Identified Through Insertional Mutagenesis. Genetics. 162,4: 1573-1581.

Dobrowolska, A., Staczek, P., 2009. Development of transformation system for *Trichophyton rubrum* by electroporation of germinated conidia. Curr Genet. 55,5: 537-542.

Edwards, J. A., Zemska, O., Rappleye, C. A., 2011. Discovery of a Role for Hsp82 in *Histoplasma* Virulence through a Quantitative Screen for Macrophage Lethality. Infect. Immun. 79,8: 3348-3357.

Enloe, B., Diamond, A., Mitchell, A. P., 2000. A single-transformation gene function test in diploid *Candida albicans*. J Bacteriol. 182,20: 5730-5736.

Erickson, T., Liu, L., Gueyikian, A., Zhu, X., Gibbons, J., Williamson, P. R., 2001. Multiple virulence factors of *Cryptococcus neoformans* are dependent on VPH1. Mol Microbiol. 42,4: 1121-1131.

Feng, Q., Summers, E., Guo, B., Fink, G., 1999. Ras Signaling Is Required for Serum-Induced Hyphal Differentiation in *Candida albicans*. Journal of Bacteriology. 181,20: 6339-6346.

Firon, A., Villalba, F., Beffa, R., d'Enfert, C., 2003. Identification of Essential Genes in the Human Fungal Pathogen *Aspergillus fumigatus* by Transposon Mutagenesis. Eukaryotic Cell. 2,2: 247-255.

Fox, D. S., Heitman, J., 2005. Calcineurin-Binding Protein Cbp1 Directs the Specificity of Calcineurin-Dependent Hyphal Elongation during Mating in *Cryptococcus neoformans*. Eukaryotic Cell. 4,9: 1526-1538.

Gauthier, G. M., Sullivan, T. D., Gallardo, S. S., Brandhorst, T. T., Wymelenberg, A. J. V., Cuomo, C. A., Suen, G., Currie, C. R., Klein, B. S., 2010. SREB, a GATA transcription factor that directs disparate fates in *Blastomyces dermatitidis* including morphogenesis and siderophore biosynthesis. PLoS Pathogens. 6,4: 1-16.

Green, B., Bouchier, C., Fairhead, C., Craig, N., Cormack, B., 2012. Insertion site preference of Mu, Tn*5*, and Tn*7* transposons. Mobile DNA. 3,1: 3.

Hensel, M., Shea, J. E., Gleeson, C., Jones, M. D., Dalton, E., Holden, D. W., 1995. Simultaneous identification of bacterial virulence genes by negative selection. Science. 269,5222: 400-403.

Hicks, J. K., Bahn, Y.-S., Heitman, J., 2005. Pde1 Phosphodiesterase Modulates Cyclic AMP Levels through a Protein Kinase A-Mediated Negative Feedback Loop in *Cryptococcus neoformans*. Eukaryotic Cell. 4,12: 1971-1981.

Idnurm, A., Reedy, J. L., Nussbaum, J. C., Heitman, J., 2004. *Cryptococcus neoformans* Virulence Gene Discovery through Insertional Mutagenesis. Eukaryotic Cell. 3,2: 420-429.

Juarez-Reyes, A., De Las Penas, A., Castano, I., 2011. Analysis of subtelomeric silencing in *Candida glabrata*. Methods Mol Biol. 734: 279-301.

Kakeya, H., Miyazaki, Y., Miyazaki, H., Nyswaner, K., Grimberg, B., Bennett, J. E., 2000. Genetic analysis of azole resistance in the Darlington strain of *Candida albicans*. Antimicrob Agents Chemother. 44,11: 2985-2990.

Kelly, M. N., Johnston, D. A., Peel, B. A., Morgan, T. W., Palmer, G. E., Sturtevant, J. E., 2009. Bmh1p (14-3-3) mediates pathways associated with virulence in *Candida albicans*. Microbiology. 155,Pt 5: 1536-1546.

Krysan, P., Young, J., Sussman, M., 1999. T-DNA as an insertionalmutagen in *Arabidopsis*. Plant Cell. 11,12: 2283 - 2290.

Kummasook, A., Cooper, C. R., Vanittanakom, N., 2010. An improved *Agrobacterium*-mediated transformation system for the functional genetic analysis of *Penicillium marneffei*. Medical Mycology. 48,8: 1066-1074.

Lamb, D. C., Kelly, D. E., Schunck, W. H., Shyadehi, A. Z., Akhtar, M., Lowe, D. J., Baldwin, B. C., Kelly, S. L., 1997. The mutation T315A in *Candida albicans* sterol 14alpha-demethylase causes reduced enzyme activity and fluconazole resistance through reduced affinity. J Biol Chem. 272,9: 5682-5688.

Lamb, D. C., Kelly, D. E., Baldwin, B. C., Kelly, S. L., 2000. Differential inhibition of human CYP3A4 and *Candida albicans* CYP51 with azole antifungal agents. Chem Biol Interact. 125,3: 165-175.

Laskowski, M. C., Smulian, A. G., 2010. Insertional mutagenesis enables cleistothecial formation in a non-mating strain of *Histoplasma capsulatum*. BMC Microbiology. 10.

Lenardon, M., Lesiak, I., Munro, C., Gow, N., 2009. Dissection of the *Candida albicans* class I chitin synthase promoters. Molecular Genetics and Genomics. 281,4: 459-471.

Li, C. R., Wang, Y. M., Wang, Y., 2008. The IQGAP Iqg1 is a regulatory target of CDK for cytokinesis in *Candida albicans*. EMBO J. 27,22: 2998-3010.

Liu, Y. G., Whittier, R. F., 1995. Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. Genomics. 25,3: 674-681.

Magrini, V., Goldman, W. E., 2001. Molecular mycology: a genetic toolbox for *Histoplasma capsulatum*. Trends in Microbiology. 9,11: 541-546.

Mah, J.-H., Yu, J.-H., 2006. Upstream and Downstream Regulation of Asexual Development in *Aspergillus fumigatus*. Eukaryotic Cell. 5,10: 1585-1595.

Mao, Y., Zhang, Z., Gast, C., Wong, B., 2008. C-Terminal Signals Regulate Targeting of Glycosylphosphatidylinositol-Anchored Proteins to the Cell Wall or Plasma Membrane in *Candida albicans*. Eukaryotic Cell. 7,11: 1906-1915.

Marion, C. L., Rappleye, C. A., Engle, J. T., Goldman, W. E., 2006. An alpha-(1,4)-amylase is essential for alpha-(1,3)-glucan production and virulence in *Histoplasma capsulatum*. Mol Microbiol. 62,4: 970-983.

Mellado, E., Garcia-Effron, G., Alcazar-Fuoli, L., Melchers, W. J., Verweij, P. E., Cuenca-Estrella, M., Rodriguez-Tudela, J. L., 2007. A new *Aspergillus fumigatus* resistance mechanism conferring *in vitro* cross-resistance to azole antifungals involves a combination of cyp51A alterations. Antimicrob Agents Chemother. 51,6: 1897-1904.

Menon, V., Li, D., Chauhan, N., Rajnarayanan, R., Dubrovska, A., West, A. H., Calderone, R., 2006. Functional studies of the Ssk1p response regulator protein of *Candida albicans* as determined by phenotypic analysis of receiver domain point mutants. Molecular Microbiology. 62,4: 997-1013.

Morio, F., Loge, C., Besse, B., Hennequin, C., Le Pape, P., 2010. Screening for amino acid substitutions in the *Candida albicans* Erg11 protein of azole-susceptible and azole-resistant clinical isolates: new substitutions and a review of the literature. Diagn Microbiol Infect Dis. 66,4: 373-384.

Nelson, R. T., Hua, J., Pryor, B., Lodge, J. K., 2001. Identification of virulence mutants of the fungal pathogen *Cryptococcus neoformans* using signature-tagged mutagenesis. Genetics. 157,3: 935-947.

Nicholls, S., MacCallum, D. M., Kaffarnik, F. A., Selway, L., Peck, S. C., Brown, A. J., 2011. Activation of the heat shock transcription factor Hsf1 is essential for the full virulence of the fungal pathogen *Candida albicans*. Fungal Genet Biol. 48,3: 297-305.

Noffz, C. S., Liedschulte, V., Lengeler, K., Ernst, J. F., 2008. Functional Mapping of the *Candida albicans* Efg1 Regulator. Eukaryotic Cell. 7,5: 881-893.

Norice, C. T., Smith, F. J., Jr., Solis, N., Filler, S. G., Mitchell, A. P., 2007. Requirement for *Candida albicans* Sun41 in Biofilm Formation and Virulence. Eukaryotic Cell. 6,11: 2046-2055.

Palmer, G. E., Johnson, K. J., Ghosh, S., Sturtevant, J., 2004. Mutant alleles of the essential 14-3-3 gene in *Candida albicans* distinguish between growth and filamentation. Microbiology. 150,Pt 6: 1911-1924.

Palmer, G. E., Sturtevant, J. E., 2004. Random mutagenesis of an essential *Candida albicans* gene. Curr Genet. 46,6: 343-356.

Park, H., Liu, Y., Solis, N., Spotkov, J., Hamaker, J., Blankenship, J. R., Yeaman, M. R., Mitchell, A. P., Liu, H., Filler, S. G., 2009. Transcriptional Responses of *Candida albicans* to Epithelial and Endothelial Cells. Eukaryotic Cell. 8,10: 1498-1510.

Peterson, A. W., Pendrak, M. L., Roberts, D. D., 2011. ATP Binding to Hemoglobin Response Gene 1 Protein Is Necessary for Regulation of the Mating Type Locus in *Candida albicans*. Journal of Biological Chemistry. 286,16: 13914-13924.

Prasannan, P., Suliman, H. S., Robertus, J. D., 2009. Kinetic analysis of site-directed mutants of methionine synthase from *Candida albicans*. Biochemical and Biophysical Research Communications. 382,4: 730-734.

Ramon, A. M., Fonzi, W. A., 2003. Diverged Binding Specificity of Rim101p, the *Candida albicans* Ortholog of PacC. Eukaryotic Cell. 2,4: 718-728.

Sanchez-Martinez, C., Perez-Martin, J., 2002. Gpa2, a G-Protein alpha Subunit Required for Hyphal Development in *Candida albicans*. Eukaryotic Cell. 1,6: 865-874.

Sanglard, D., Ischer, F., Koymans, L., Bille, J., 1998. Amino acid substitutions in the cytochrome P-450 lanosterol 14alpha-demethylase (CYP51A1) from azole-resistant *Candida albicans* clinical isolates contribute to resistance to azole antifungal agents. Antimicrob Agents Chemother. 42,2: 241-253.

Schmalhorst, P. S., Krappmann, S., Vervecken, W., Rohde, M., Muller, M., Braus, G. H., Contreras, R., Braun, A., Bakker, H., Routier, F. H., 2008. Contribution of Galactofuranose to the Virulence of the Opportunistic Pathogen *Aspergillus fumigatus*. Eukaryotic Cell. 7,8: 1268-1277.

Sheng, C., Chen, S., Ji, H., Dong, G., Che, X., Wang, W., Miao, Z., Yao, J., Lü, J., Guo, W., Zhang, W., 2010. Evolutionary trace analysis of CYP51 family: implication for site-directed mutagenesis and novel antifungal drug design. Journal of Molecular Modeling. 16,2: 279-284.

Smulian, A. G., Gibbons, R. S., Demland, J. A., Spaulding, D. T., Deepe, G. S., Jr., 2007. Expression of Hygromycin Phosphotransferase Alters Virulence of *Histoplasma capsulatum*. Eukaryotic Cell. 6,11: 2066-2071.

Snelders, E., Karawajczyk, A., Verhoeven, R. J. A., Venselaar, H., Schaftenaar, G., Verweij, P. E., Melchers, W. J. G., 2011. The structure-function relationship of the *Aspergillus fumigatus* cyp51A L98H conversion by site-directed mutagenesis: The mechanism of L98H azole resistance. Fungal Genetics and Biology. 48,11.

Spreghini, E., Davis, D. A., Subaran, R., Kim, M., Mitchell, A. P., 2003. Roles of *Candida albicans* Dfg5p and Dcw1p Cell Surface Proteins in Growth and Hypha Formation. Eukaryotic Cell. 2,4: 746-755.

Sullivan, T., Rooney, P., Klein, B., 2002. *Agrobacterium tumefaciens* integrates transfer DNA into single chromosomal sites of dimorphic fungi and yields homokaryotic progeny from multinucleate yeast. Eukaryot Cell. 1,6: 895 - 905.

Uhl, M. A., Biery, M., Craig, N., Johnson, A. D., 2003. Haploinsufficiency-based large-scale forward genetic analysis of filamentous growth in the diploid human fungal pathogen *C. albicans*. EMBO J. 22,11: 2668-2678.

VandenBerg, A. L., Ibrahim, A. S., Edwards, J. E., Jr., Toenjes, K. A., Johnson, D. I., 2004. Cdc42p GTPase Regulates the Budded-to-Hyphal-Form Transition and Expression of Hypha-Specific Transcripts in *Candida albicans*. Eukaryotic Cell. 3,3: 724-734.

Yamada, T., Makimura, K., Satoh, K., Umeda, Y., Ishihara, Y., Abe, S., 2009. *Agrobacterium tumefaciens*-mediated transformation of the dermatophyte, *Trichophyton mentagrophytes*: an efficient tool for gene transfer. Med Mycol. 47,5: 485-494.

Zhang, P., Xu, B., Wang, Y., Li, Y., Qian, Z., Tang, S., Huan, S., Ren, S., 2008. *Agrobacterium tumefaciens*-mediated transformation as a tool for insertional mutagenesis in the fungus *Penicillium marneffei*. Mycological Research. 112,8: 943-949.

Znaidi, S., Barker, K. S., Weber, S., Alarco, A.-M., Liu, T. T., Boucher, G., Rogers, P. D., Raymond, M., 2009. Identification of the *Candida albicans* Cap1p Regulon. Eukaryotic Cell. 8,6: 806-820.

Recombineering and Conjugation as Tools for Targeted Genomic Cloning

James W. Wilson[1], Clayton P. Santiago[1],
Jacquelyn Serfecz[1] and Laura N. Quick[2]
[1]Villanova University,
[2]Children's Hospital of Philadelphia,
USA

1. Introduction

The ability to obtain DNA clones of genes that normally reside in microbial genomes was a huge technical advance in molecular biology. At first, cloning genes utilized approaches involving the complementation of mutants or the screening of genomic libraries to find sequences that hybridized to homologous DNA probes. Typically, this involved using restriction enzymes to clone random genomic fragments followed by subcloning of a smaller piece of the original clone. Then the development of PCR and genomic sequencing allowed specific genomic sequences to be amplified and cloned with more convenience. Now genes are able to be synthesized "from scratch" and ordered from various companies or institutions. However, if many genes contained on a contiguous large genomic segment are required to be cloned, significant technical barriers exist. For the purposes of this discussion, we will establish that a "large" genomic segment constitutes greater than 10 kilobases, since PCR and man-made DNA synthesis become technically challenging and/or costly above this DNA size. Therefore, a convenient, reproducible, and cost-efficient technique to clone large sections of microbial genomes would be highly advantageous.

Frequently bacteria organize genes that work together for a common function as a continuous, physically-linked series across a genome. Large genomic fragments containing many genes that work together for a specific function are very useful for the following reasons: (1) bacteria are able to be engineered for specific purposes in a "quantum leap" using such DNA clones; and (2) basic evolutionary questions are able to be answered using large genomic clones, such as: "Can the cloned gene set be expressed and functional outside of the context of the original genome/species?" These approaches extend the study of genomics by identifying potentially interesting parts of genomes identified via sequencing and studying them in different strain backgrounds. A clear example of this approach is the cloning of protein secretion systems and the subsequent study of these clones (Blondel et al. 2010; Ham et al. 1998; Hansen-Wester, Chakravortty, and Hensel 2004; McDaniel and Kaper 1997; Wilson, Coleman, and Nickerson 2007; Wilson and Nickerson 2006). However, many other gene systems can be studied in this way, with examples including polysaccharide secretion pathways (for capsule and LPS synthesis) and metabolic pathways (anabolism and/or catabolism of key molecules, such as those used in bioremediation). Our ability to extend genomics beyond sequencing to the

utilization of newly-identified multi-gene pathways to engineer bacteria will depend upon our ability to clone, manipulate, and transfer large genomic fragments.

A recent strategy that exploits recombineering and conjugation provides a convenient approach to cloning large bacterial genomic fragments (Blondel et al. 2010; Santiago, Quick, and Wilson 2011; Wilson, Figurski, and Nickerson 2004; Wilson and Nickerson 2007). This approach involves insertion of recombinase sites (*e.g.*, FRT, *loxP*) at positions flanking a targeted genomic region, followed by subsequent recombinase-mediated excision of the region as a non-replicating circular molecule (Fig. 1). Then the excised region is "captured" via either site-specific or homologous recombination onto a conjugative plasmid (such as the broad-host-range IncP plasmid R995) that allows the transfer and isolation of the desired construct in a fresh recipient strain (Fig. 1). The advantages of this approach are (1) the highly specific targeting of exact cloning endpoints using recombineering and (2) the use of conjugation to allow the desired construct to be isolated away from the donor strain (in which the recombination events take place). In addition, except for the synthesis of recombineering PCR products, this protocol takes place entirely in bacterial cells, using basic, low-cost microbiological techniques. Though early approaches used subcloned DNA fragments to allow homologous recombination, the use of recombineering for both the introduction of target flanking sites and the capture on R995 alleviates the need for this subcloning.

2. Targeted cloning of large bacterial genomic fragments

2.1 The VEX-Capture technique

The original technique using this approach is termed VEX-Capture (Wilson, Coleman, and Nickerson 2007; Wilson, Figurski, and Nickerson 2004; Wilson and Nickerson 2006, 2006, 2007). The pVEX series of suicide plasmids was used to introduce *loxP* sites into regions flanking targeted genomic regions via homologous recombination (Fig. 2) (Ayres et al. 1993). Cre recombinase (expressed from a plasmid) was used to excise the targeted region and homologous recombination was used to capture the excised circle (Fig. 2). Note that the homologous recombination is driven by the endogenous bacterial RecA-mediated mechanism. A series of *Salmonella typhimurium* genomic islands ranging from 26 to 50 kilobases in size were targeted for cloning using this technique (Wilson, Coleman, and Nickerson 2007; Wilson, Figurski, and Nickerson 2004; Wilson and Nickerson 2006, 2006). Since these islands contain genes that are unique to *S. typhimurium*, one of the initial basic applications of these clones was to study their gene expression patterns in different bacteria (Wilson, Figurski, and Nickerson 2004; Wilson and Nickerson 2006). Though some *S. typhimurium* genes on the tested genomic island were expressed in all bacteria, several genes displayed genus-specific expression patterns (Fig. 3). This indicated that the mechanisms used to express these genes are absent or function differently in certain bacterial genera. These mechanisms could be the focus of study to understand gene expression functions that work only in certain bacterial groups, such as pathogens or environmental bacteria.

Two separate *S. typhimurium* type III secretion systems were cloned using the VEX-Capture approach (Wilson, Coleman, and Nickerson 2007; Wilson and Nickerson 2006). These systems are encoded at the *Salmonella* pathogenicity island 1 and 2 regions (SPI-1 and SPI-2, respectively) of the *S. typhimurium* genome (McClelland et al. 2001). Both clones are
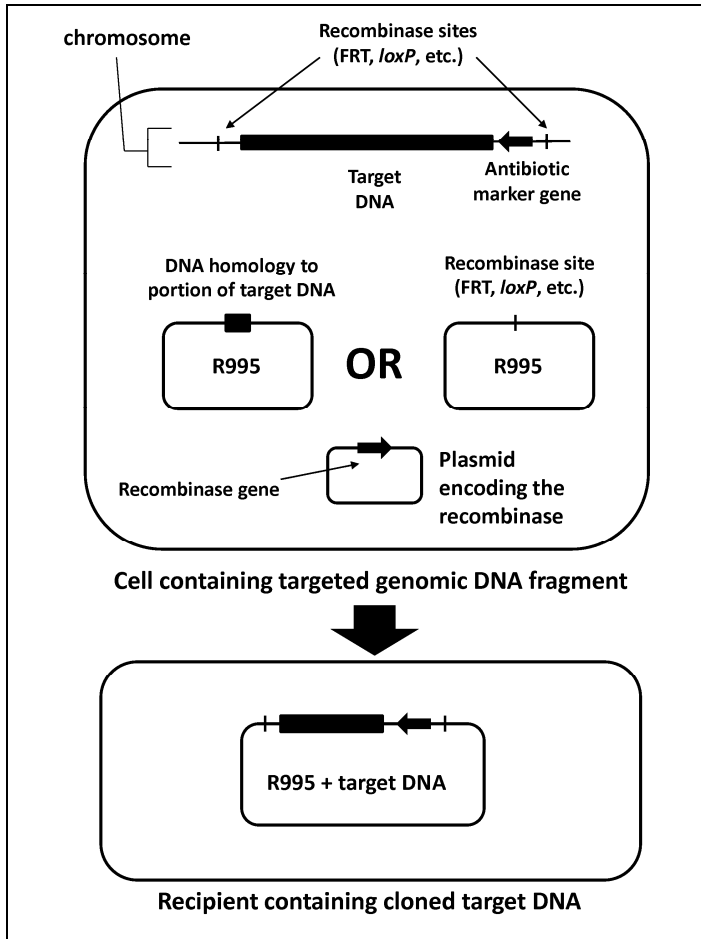
Fig. 1. General outline of VEX-Capture to clone large genomic fragments. A large fragment of a bacterial genome (generally considered as greater than 10 kilobases) is targeted for excision and cloning by inserting recombinase sites at flanking positions. At least one antibiotic marker gene is required to be associated with the target DNA for subsequent selection. The self-transmissible IncP plasmid R995 serves as a cloning vector that will capture the excised genomic fragment using either a small region of DNA homologous to the excised fragment or a corresponding recombinase site. Also co-resident in the same cell is a plasmid expressing the recombinase that recognizes the recombinase sites. Expression of the recombinase results in excision of the target DNA as a non-replicating circular molecule. This circular molecule will be inserted into R995 via homologous recombination or via the recombinase activity. This construct is conveniently isolated away from the target strain via conjugation to a differentially-marked recipient strain and selection for the appropriate markers. In the recipient strain, structural confirmation of the construct and testing for gene expression and function can occur. In addition, transfer to new bacterial recipients can be performed.

functional and serve to complement protein secretion defects in *S. typhimurium* mutants that are deleted for each SPI-1 and SPI-2 island (Fig. 4). However, the authors found remarkably different results between R995 + SPI-1 and R995 + SPI-2 when tested for expression in other Gram-negative bacteria (Fig. 5). The R995 + SPI-2 clone readily displays expression of SPI-2 (indicated using Western blot analysis of the SseB protein) in other Gram-negative genera, while the R995 + SPI-1 clone displays an expression defect outside of *S. typhimurium* (assayed using Western blot analysis of the SipA and SipC proteins). This result suggests that the regulatory mechanisms controlling SPI-1 and SPI-2 expression have evolved differently and in such a way that manifests itself upon transfer to new bacterial backgrounds.

2.2 VEX-Capture modified

A modification of VEX-Capture was used to clone the type VI secretion system encoded at *Salmonella* pathogenicity island 19 (SPI-19) in the *S. gallinarum* genome (Blondel et al. 2010). In this approach, the *loxP* sites and markers (for chloramphenicol and spectinomycin resistance) were PCR-amplified from the pVEX vectors and inserted into flanking positions using phage λ Red recombination (Fig. 6). The SPI-19 region was excised via Cre recombinase and captured onto R995 using homologous recombination (Fig. 6). The resulting R995 + SPI-19 clone was used to complement the colonization defect of the *S. gallinarum* SPI-19 deletion strain in a chicken infection model (Blondel et al. 2010). In addition, the authors transferred the R995 + SPI-19 clone into *S. enteriditis*, a species that contains significant sequence deviation in SPI-19 relative to *S. gallinarum*, to test if the presence of the *S. gallinarum* SPI-19 would increase *S. enteriditis* chicken colonization.

Interestingly, the presence of SPI-19 decreased the ability of *S. enteriditis* to colonize in this infection model (Blondel et al. 2010). This is consistent with the observations described above that demonstrate genomic island phenotypes can differ greatly, depending on the bacterial background.

2.3 New R995 derivatives allow an "all recombinase" approach

Recently an entirely recombinase-based approach for this techninque has been described using modified R995 plasmids (Santiago, Quick, and Wilson 2011). The new series of R995 derivatives encode a range of different marker combinations to increase utility in situations where several markers are used or are already present in the strain background. In addition, these R995 derivatives contain FRT sites that can facilitate the capture of genomic regions that have been excised using the Flp/FRT system (Fig. 7). A major advantage to this approach is that no regions of homology are needed to be cloned into any plasmids. Thus, the only step that takes place outside of cells is the amplification of the PCR products used for λ Red insertion of FRT sites into the flanking positions in the genome. This technique was demonstrated by cloning 20-kilobase regions from the *S. typhimurium* and *Escherichia coli* genomes (Santiago, Quick, and Wilson 2011).

2.4 Catalogue of reagents

Table 1 serves as a summary list of reagents used for the recombinase/conjugation-based cloning of genomic fragments. The PCR template plasmids are suicide plasmids and can
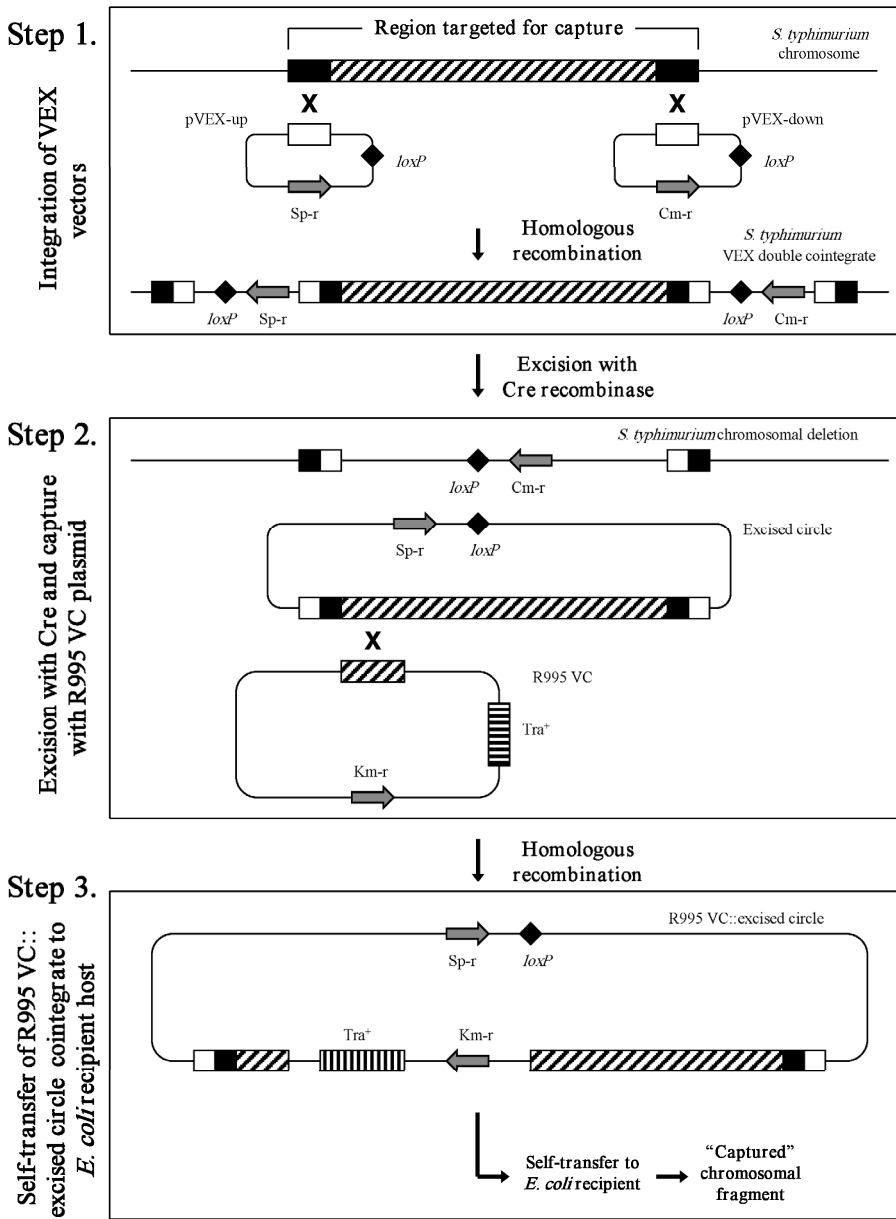
Fig. 2. The VEX-Capture system. Excision and capture of a section of the *S. typhimurium* genome is depicted to illustrate the functioning of the VEX-Capture system. In step one, differentially-marked pVEX vectors containing DNA fragments homologous to the ends of the targeted genomic region are integrated at the desired locations to form a double

cointegrate. In this structure, single *loxP* sites are located on either side of the targeted region. In step two, the targeted region is excised from the genome by the Cre recombinase, and the excised circle is ''captured'' via homologous recombination with the R995 VC plasmid. Note that the capture fragment on R995 VC is shown as targeted to one end of the excised genomic region, but it can be targeted to any location on the excised region. In step 3, the R995 VC-excised circle plasmid is transferred to an *E. coli* recipient to create a strain containing the captured genomic fragment. Diagram not drawn to scale. Reprinted from (Wilson and Nickerson 2007).
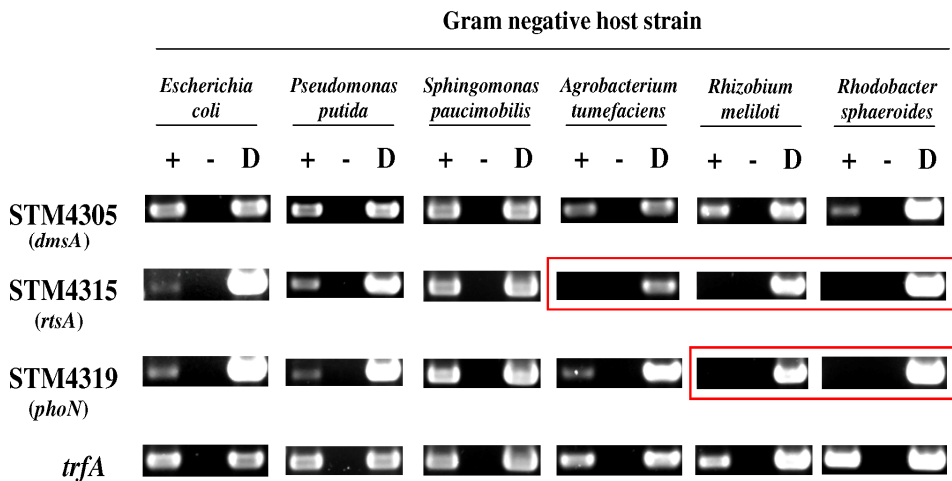

Gram negative host strain

Fig. 3. RT-PCR analysis of *S. typhimurium* island 4305 after transfer to different Gram-negative hosts. The indicated Gram-negative strains containing R995 + *S. typhimurium* island 4305 were analyzed for expression of island genes STM4305, STM4315, STM4319 and the R995 replication gene *trfA* (which serves a positive control). Total RNA from each strain was isolated and reversed transcribed, and the samples were PCR-amplified using primers against the indicated genes. The (+) and (-) lanes indicate samples with and without the reverse transcriptase step, respectively, and the (D) lane indicates where R995 + island 4305 DNA isolated from each was used as template. PCR samples were run on agarose gels and stained with ethidium bromide. The boxed pictures indicate where expression of the gene is not detectable. This figure demonstrates genus-specific expression patterns for those island genes. Reprinted from (Wilson and Nickerson 2006).
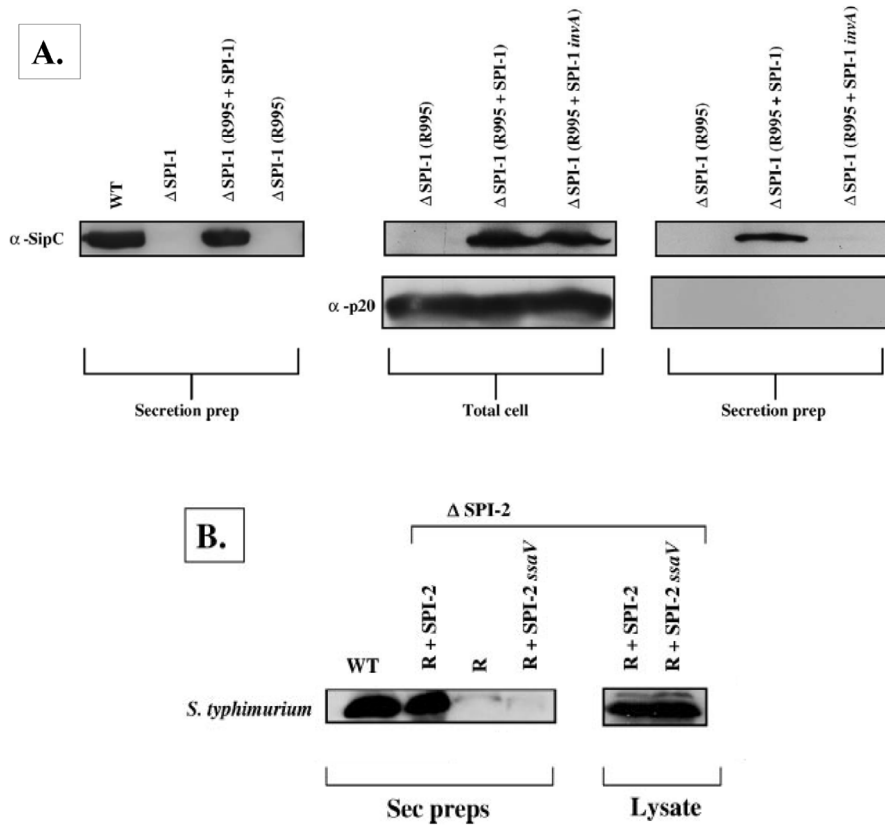
Fig. 4. R995 + SPI-1 and R995 + SPI-2 clones complement corresponding *S. typhimurium* SPI-1 and SPI-2 deletion mutants for substrate protein secretion. Panel A: Western blot analysis of protein secretion preparations and total cell lysates from *S. typhimurium* delta SPI-1 strains containing either R995, R995 + SPI-1, or R995 + SPI-1 *invA* plasmids. The last plasmid contains a mutation in the *invA* gene encoding a SPI-1 type III system protein that is essential for SPI-1-mediated secretion. Antibodies against the SPI-1 secreted substate SipC and the non-secreted bacterial cellular protein p20 are used. Panel B: Western blot analysis as in Panel A but using *S. typhimurium* delta SPI-2 strains containing either R995, R995 + SPI-2, or R995 + SPI-2 *ssaV* (mutation for the *ssaV* gene essential for SPI-2 secretion activity). Antibodies against the SPI-2 protein substrate SseB are used. The results of both panels demonstrate that the cloned SPI-1 and SPI-2 regions on R995 are functional and complement deleted SPI-1 and SPI-2 secretion systems. Reprinted from (Wilson, Coleman, and Nickerson 2007; Wilson and Nickerson 2006).
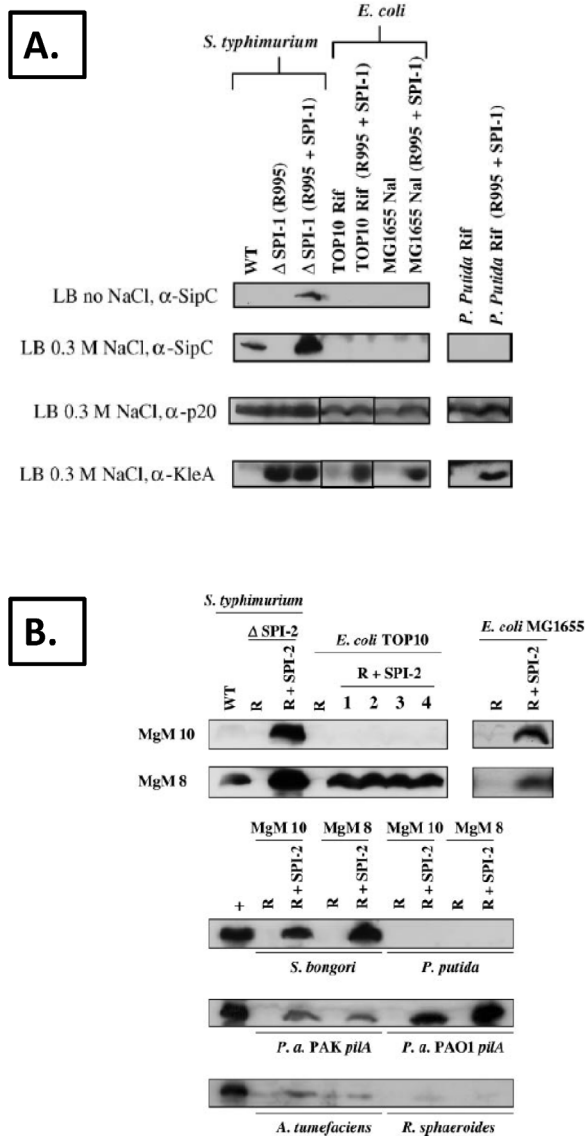
Fig. 5. Different expression patterns for SPI-1 and SPI-2 in different Gram-negative bacterial genera. Panel A: Plasmid R995 + SPI-1 was analyzed for expression of the SPI-1 protein SipC via Western blot analysis in *S. typhimurium, E. coli,* and *Pseudomonas putida*. In addition, the samples were also probed for the bacterial housekeeping p20 protein and the R995-encoded protein KleA as controls. The samples shown are total cell lysates of each strain. SipC expression is not detectable in *E. coli, P. putida,* attentuated in *P. aeruginosa* and *Agrobacterium tumefaciens* (the last two species not shown). Panel B: Plasmid R995 + SPI-2 expression was analyzed via Western blot assay against the SPI-2 protein SseB in various Gram-negative

bacteria. In contrast to SPI-1, expression of SPI-2 was readily detected in a range of bacterial backgrounds. Two points are of particular note: (1) In *S. typhimurium*, SPI-2 expression is regulated by growth media conditions, such that 10 mM MgCl$_2$ and pH 7.5 repress expression (MgM 10 media) and 8 μM MgCl$_2$ and pH 5.5 activate expression (MgM 8 media). However, expression from R995 + SPI-2 does not follow this regulation, except in the *E. coli* strain TOP10. R995 + SPI-1 expression shows a similar result in *S. typhimurium* in relation to its regulation by sodium chloride; and (2) *P. putida* appears to be recalcitrant to both SPI-1 and SPI-2 expression. Reprinted from (Wilson, Coleman, and Nickerson 2007; Wilson and Nickerson 2006).
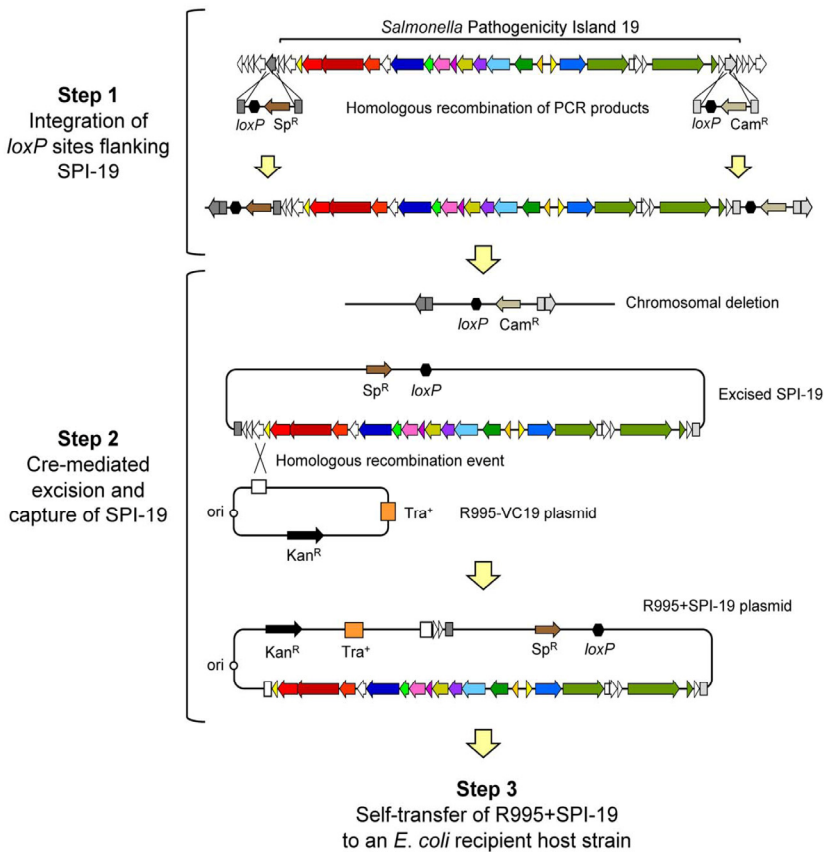

Fig. 6. Schematic representation of the capture of SPI-19 from *S. gallinarum* 287/91 using a modified VEX-Capture method. To clone the type VI secretion system from the *S. gallinarum* genome, Blondel *et. al.* PCR-amplified markers and *loxP* sites from pVEX vectors and inserted them into flanking positions using phage λ Red recombination. The Cre-excised circular molecule was captured by R995 via homologous recombination, and the construct was isolated upon conjugation to an *E. coli* recipient. This construct was used for complementation analysis in a chicken model of infection using *S. gallinarum* and *S. enteriditis* strains and demonstrates the utility of R995 capture plasmids for *in vivo* pathogenesis studies. Reprinted from (Blondel et al. 2010).
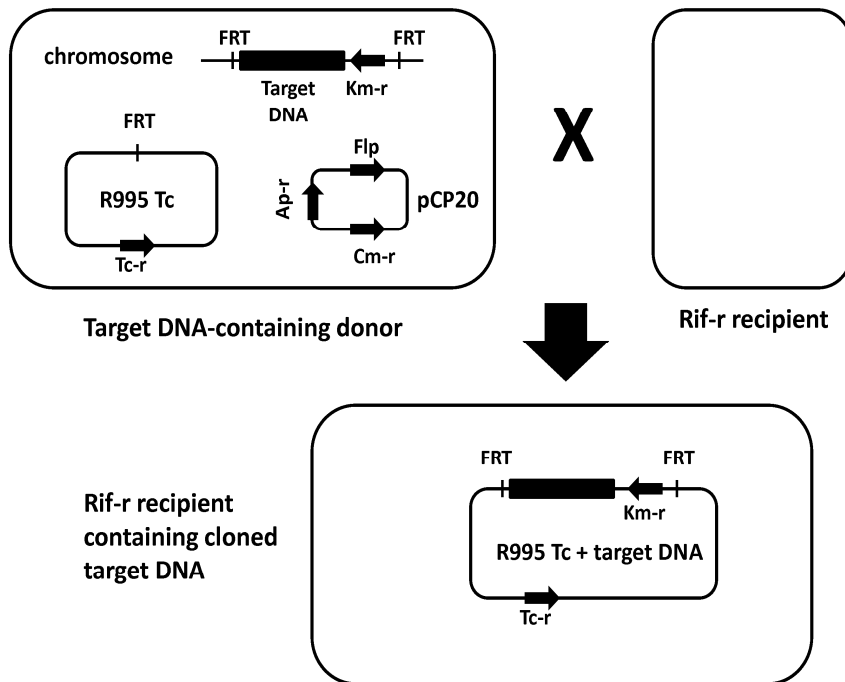
Fig. 7. An "all recombinase" approach to cloning large genomic DNA fragments to R995. This procedure utilizes specially designed R995 derivatives containing FRT sites that can be used as insertion points for a genomic fragment excised using the Flp/FRT system. A targeted DNA region in a bacterial genome is flanked by FRT sites and an antibiotic resistance marker as diagrammed using λ Red recombination. To accomplish this, the "unmarked" FRT site (to the left of the target DNA in the chromosome) is introduced via standard λ Red recombination markers (in Table 1) followed by Flp-mediated deletion of the marker to leave the single, unmarked FRT site. Next, the second flanking FRT site is introduced using a PCR fragment designed with a marker and single FRT site, such that the marker is located between the FRT site and the target DNA. In this example, the marker encodes kanamycin resistance. An R995 derivative containing an FRT site (and encoding tetracycline resistance in this example) is transferred to this strain via conjugation, and then the Flp-expressing plasmid pCP20 is introduced via electroporation. The electroporation outgrowth culture can be used directly as a donor for conjugation with a rifampicin (Rif)-resistant recipient strain. Alternatively, the electroporation can be plated on media containing tetracycline (Tc) and kanamycin (Km) and the colonies can be used as donor. The conjugation is plated on media containing Rif, Tc, and Km to select recipients that have obtained the cloned target DNA on R995. The transconjugants can be used to confirm the clone. A transconjugant can also be used as a donor for transfer of the clone to other bacterial strains for subsequent studies. Reprinted from (Santiago, Quick, and Wilson 2011).

Plasmid or strain	Chacteristics	Reference
Template plasmids		
pKD3	FRT:Cm-r:FRT, R6K ori	Datsenko, 2000
pKD4	FRT:Km-r:FRT, R6K ori	Datsenko, 2000
pJW101	FRT:Tp-r:FRT, R6K ori	Quick, 2010
pJW102	FRT:Sp-r:FRT, R6K ori	Quick, 2010
pJW107	FRT:Tc-r:FRT, R6K ori	This chapter
pVEX1212	*loxP*:Sp-r, P1 ori	Ayres, 1993
pVEX2212	*loxP*:Cm-r, R6K ori	Ayres, 1993
Lambda Red plasmids		
pKD46	Ap-r	Datsenko, 2000
pJW103	Km-r	Quick, 2010
pJW104	Cm-r	Quick, 2010
pJW105	Tp-r	Quick, 2010
pJW106	Sp-r	Quick, 2010
Recombinase plasmids		
pCP20	Flp, Ap-r, Cm-r	Datsenko, 2000
pEKA30	Cre, Ap-r, Tc-r	Wilson, 2004
pEKA16	IPTG-inducible Cre, Ap-r, Tc-r	Wilson, 2004
R995 derivatives		
R995	Km-r, Tc-r	Pansegrau, 1994
R995 Km Cm	*ΔtetRA*::FRT-Cm-FRT	Santiago, 2011
R995 Km Tp	*ΔtetRA*::FRT-Tp-FRT	Santiago, 2011
R995 Km Sp	*ΔtetRA*::FRT-Sp-FRT	Santiago, 2011
R995 Tc Cm	*ΔaphA*::FRT-Cm-FRT	Santiago, 2011
R995 Tc Tp	*ΔaphA*::FRT-Cm-FRT	Santiago, 2011
R995 Tc Sp	*ΔaphA*::FRT-Cm-FRT	Santiago, 2011
R995 Km	*ΔtetRA*::FRT	Santiago, 2011
R995 Tc	*ΔaphA*::FRT	Santiago, 2011
R995 Cm	*ΔtetRA*::FRT, *ΔaphA*::Cm-r	Santiago, 2011
R995 Tp	*ΔtetRA*::FRT, *ΔaphA*::Tp-r	Santiago, 2011
R995 Sp	*ΔtetRA*::FRT, *ΔaphA*::Sp-r	Santiago, 2011
Strains		
AS11	*pir*+ (for R6K ori plasmids)	Ayres, 1993
EKA260	*repA*+ (for P1 ori plasmids)	Ayres, 1993
TOP10 Rif	Rif-r recipient	Wilson, 2004

Table 1. Catalogue of reagents for recombinase/conjugation cloning. Please note that the template plasmids are suicide plasmids that require either AS11 or EKA260 for replication and that the λ Red plasmids and pCP20 are temperature-sensitive for replication (requiring 30 degrees C). The pJW plasmids are derived from either pKD3 (pJW101 and pJW102) or pKD46 (pJW103, pJW104, pJW105, and pJW106) (Quick, Shah, and Wilson 2010).

only replicate in corresponding strains that encode either the R6K Pir protein or P1 RepA protein (Ayres et al. 1993; Datsenko and Wanner 2000). This allows the PCR reaction to be directly electroporated into target cells with no background problems caused by the replication of the templates. It is worthwhile to note the PCR template plasmids with FRT sites contain two such sites flanking a given antibiotic resistance marker. Thus, care must be taken to amplify products containing only one FRT site for the second flanking insertion into the genome to avoid marker loss problems upon Flp expression (please refer to Fig. 7 for more details). It is also worthwhile to note that the self-transmissible IncP plasmid R995 displays a remarkably broad-host-range for both its conjugation and replication system (Adamczyk and Jagura-Burdzy 2003; Pansegrau et al. 1994; Thorsted et al. 1998). This facilitates R995 conjugative transfer to a wide variety of Gram-negative and Gram-positive bacteria and replication in almost all Gram-negative bacteria. Any other conjugative plasmid could be used for this procedure. However, IncP plasmid R995 and related plasmids are excellent options due to their broad-host-range, fully sequenced genomes, and high degree of characterization (especially for the IncPα plasmids R995, RK2, RP4, etc.).

3. Conclusion

Recombineering and conjugation can be exploited to provide a convenient, reproducible, and cost-effective technique for cloning large bacterial genomic fragments. This technique can be performed using easily obtained PCR products, readily available plasmids and strains, and simple, basic microbiology protocols. One question regarding the use of this system is: how large a genomic fragment can be accommodated by R995? So far, the biggest fragment cloned using this technique has been about 50 kilobases, but the upper limits of size have not yet been tested in any systematic way. To make genomic clones more amenable to medical or environmental applications, removal of antibiotic resistance markers and the conjugative transfer system would need to be accomplished. We are currently pursuing the development of alternative selection schemes and removable conjugation systems to address this issue. Overall, the use of the recombinase/conjugation cloning approach is currently underdeveloped as a technique and could expand the field of genomics by providing experiment-based strategies to answer important evolutionary questions.

4. Acknowledgment

5. References

Adamczyk, M., and G. Jagura-Burdzy. 2003. Spread and survival of promiscuous IncP-1 plasmids. *Acta Biochim Pol* 50 (2):425-53.

Ayres, E. K., V. J. Thomson, G. Merino, D. Balderes, and D. H. Figurski. 1993. Precise deletions in large bacterial genomes by vector-mediated excision (VEX). The *trfA* gene of promiscuous plasmid RK2 is essential for replication in several gram-negative hosts. *J Mol Biol* 230 (1):174-85.

Blondel, C. J., H. J. Yang, B. Castro, S. Chiang, C. S. Toro, M. Zaldivar, I. Contreras, H. L. Andrews-Polymenis, and C. A. Santiviago. 2010. Contribution of the type VI secretion system encoded in SPI-19 to chicken colonization by *Salmonella enterica* serotypes Gallinarum and Enteritidis. *PLoS One* 5 (7):e11724.

Datsenko, K. A., and B. L. Wanner. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* 97 (12): 6640-5.

Ham, J. H., D. W. Bauer, D. E. Fouts, and A. Collmer. 1998. A cloned *Erwinia chrysanthemi* Hrp (type III protein secretion) system functions in *Escherichia coli* to deliver *Pseudomonas syringae* Avr signals to plant cells and to secrete Avr proteins in culture. *Proc Natl Acad Sci U S A* 95 (17):10206-11.

Hansen-Wester, I., D. Chakravortty, and M. Hensel. 2004. Functional transfer of Salmonella pathogenicity island 2 to *Salmonella bongori* and *Escherichia coli*. *Infect Immun* 72 (5):2879-88.

McClelland, M., K. E. Sanderson, J. Spieth, S. W. Clifton, P. Latreille, L. Courtney, S. Porwollik, J. Ali, M. Dante, F. Du, S. Hou, D. Layman, S. Leonard, C. Nguyen, K. Scott, A. Holmes, N. Grewal, E. Mulvaney, E. Ryan, H. Sun, L. Florea, W. Miller, T. Stoneking, M. Nhan, R. Waterston, and R. K. Wilson. 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* 413 (6858): 852-6.

McDaniel, T. K., and J. B. Kaper. 1997. A cloned pathogenicity island from enteropathogenic *Escherichia coli* confers the attaching and effacing phenotype on *E. coli* K-12. *Mol Microbiol* 23 (2):399-407.

Pansegrau, W., E. Lanka, P. T. Barth, D. H. Figurski, D. G. Guiney, D. Haas, D. R. Helinski, H. Schwab, V. A. Stanisich, and C. M. Thomas. 1994. Complete nucleotide sequence of Birmingham IncP alpha plasmids. Compilation and comparative analysis. *J Mol Biol* 239 (5):623-63.

Quick, L. N., A. Shah, and J. W. Wilson. 2010. A series of vectors with alternative antibiotic resistance markers for use in lambda Red recombination. *J Microbiol Biotechnol* 20 (4):666-9.

Santiago, CP, LN Quick, and JW Wilson. 2011. Self-Transmissible IncP R995 Plasmids with Alternative Markers and Utility for Flp/FRT Cloning Strategies. *J Microbiol Biotechnol* 21 (11).

Thorsted, P. B., D. P. Macartney, P. Akhtar, A. S. Haines, N. Ali, P. Davidson, T. Stafford, M. J. Pocklington, W. Pansegrau, B. M. Wilkins, E. Lanka, and C. M. Thomas. 1998. Complete sequence of the IncPbeta plasmid R751: implications for evolution and organisation of the IncP backbone. *J Mol Biol* 282 (5):969-90.

Wilson, J. W., C. Coleman, and C. A. Nickerson. 2007. Cloning and transfer of the Salmonella pathogenicity island 2 type III secretion system for studies of a range of gram-negative genera. *Appl Environ Microbiol* 73 (18):5911-8.

Wilson, J. W., D. H. Figurski, and C. A. Nickerson. 2004. VEX-capture: a new technique that allows *in vivo* excision, cloning, and broad-host-range transfer of large bacterial genomic DNA segments. *J Microbiol Methods* 57 (3):297-308.

Wilson, J. W., and C. A. Nickerson. 2006. Cloning of a functional Salmonella SPI-1 type III secretion system and development of a method to create mutations and epitope fusions in the cloned genes. *J Biotechnol* 122 (2):147-60.

Wilson, J. W., and C. A. Nickerson. 2006. A new experimental approach for studying bacterial genomic island evolution identifies island genes with bacterial host-specific expression patterns. *BMC Evol Biol* 6:2.

Wilson, J. W., and C. A. Nickerson. 2007. *In vivo* excision, cloning, and broad-host-range transfer of large bacterial DNA segments using VEX-Capture. *Methods Mol Biol* 394:105-18.