Structural insights into the basis and evolution of interactions in multi-subunit protein assemblies: Tryptophan Synthase and titin FNIII-repeats

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Rainer Michael Bucher

aus Efringen-Kirchen in Deutschland

Basel, 2007

i

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

Prof. Dr. Olga Mayans

Prof. Dr. Michael N. Hall
PD Dr. Michael Hennig

Basel, den 26. Juni 2007

<div style="text-align: right">

Prof. Dr. Hans-Peter Hauri
Dekan

</div>

Declaration

I declare that I wrote this thesis, **Structural insights into the basis and evolution of interactions in multi-subunit protein assemblies: Tryptophan Synthase and titin FNIII-repeats**, with the help indicated and only handed it into the Faculty of Science of the university of Basel and to no other faculty and no other university.

dedicated to my parents

für meine Eltern

Structural insights into the basis and evolution of interactions in multi-subunit protein assemblies: Tryptophan Synthase and titin FNIII-repeats

Table of Contents

List of Figures

List of Tables

List of Abbreviations

a.u.	Asymmetric Unit
Amp	Ampicillin
ARP	Automated Refinement Procedure
Cam	Chloramphenicol
CC	Coiled-coil
DMPK	Dystrophia myotonica protein kinase
H-bond	Hydrogen bond
Kan	Kanamycin
MALS	Multiangle light-scattering
MC	Main-chain
MPD	2-Methylpentane-2,4-diol
MR	Molecular Replacement
MRCK	DMPK related Cdc42-binding kinase
NCS	Noncrystallographic symmetry
ON	Overnight
PEG	Polyethylene glycol
rmsd	root mean square deviation
ROCK	Rho-associated kinase
RT	Room temperature
SC	Side-chain
SEC	Size exclusion chromatography
TrpS	Tryptophan Synthase
VR	Variable region across DMPK spliceoforms
wARP	Weighed Averaging of Multiple-Refined Dummy Atomic Models

Abstract

Cellular processes benefit from evolutionary shaping when optimized protein-protein interactions result in enhanced functionality. In fact, most cellular proteins are tightly embedded into biological networks that function following a modularity principle. Modularity, whether based on components as parts of stable protein complexes or as dynamic units that interact only transiently (as in signalling and metabolic cascades), facilitates the combinatorial generation of complexity in protein networks through the re-wiring of modules in addition to the diversification of individual proteins – thereby increasing the "evolvability" of the system. The mechanisms that drive the emergence and evolution of molecular recognition in protein networks remain unclear. It is difficult to justify such evolution on the basis of organismic advantage, since the latter might only be noticeable once full pathways and cascades have evolved. It is then likely that the evolution of protein-protein interactions is in the first instance driven by a molecular principle of local advantage to the protein system itself - for example, molecular stability. Unfortunately, it is difficult to gain insights into the evolution of protein-protein interactions since the pathways of evolutionary shaping normally let intermediates of evolution disappear. Subsequently, conclusions are more usually drawn from the comparison of proteins between different species and by mutagenesis probing.

In the current study, we aim at gaining an insight into the evolutionary shaping of proteins surfaces for hetero-complex formation by studying two systems at an early stage of development: Tryptophan Synthase B2b (TrpB2b) from *S. solfataricus* and the modular interfaces of the poly-FNIII tandems in the muscle filament titin. In the case of TrpB2b, the evolution of inter-subunit communication is addressed in addition. Both structures have been elucidated using X-ray crystallography and a comparative analysis of their surfaces has been carried out. The architectural elements subjected to evolutionary pressure have been identified and conclusions on their relation to function and evolution have been drawn.

1 Introduction

Interactions within and between proteins are influenced by the distinct properties (polarity and shape) of participating surfaces. The polarity and shape of involved amino acids depends on their atomic composition. Evolutionary shaping maximizes the compatibility of surfaces, where temporary and obligate interactions can be formed either directly between amino acids or mediated by water.
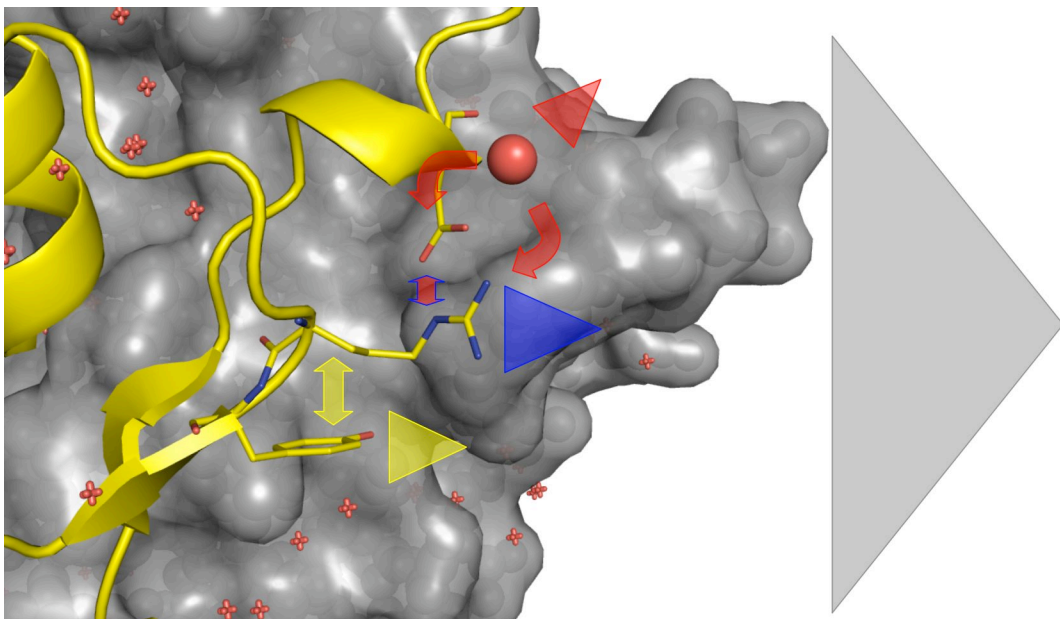
1.1   Polarity of bonds and molecules

When a covalent bond between two atoms is formed the aim is to create a noble gas like electron configuration by sharing their electrons. The electronegativity difference between the two elements involved in the bond defines the localization of the electrons (towards the nucleus of the atom at higher electronegativity, Figure 1.1). The asymmetric distribution of the electrons in the binding orbital results in polarity of the bond, which is proportional to the difference in electronegativity between the nuclei involved. The extreme case is the complete transfer of the electrons from one atom to the other in ionic species. Different intra- and intermolecular interactions, according to the polarity of bonds/molecules, can be formed:

(i)   Van der Waals interactions (London dispersity): result from temporary asymmetric electron distribution due to which molecules interact with each other.

(ii)   Dipole-dipole interactions: asymmetric electron distribution creates partial charges with a subsequent permanent dipole moment for bonds between atoms of medium electronegativity difference.

(iii) Hydrogen bonds (H-bonds): strong polarization of bonds resulting in permanent dipoles with highly asymmetric electron distribution and subsequent creation of partial charges. The strong type is found in bonds between hydrogen (H) and nitrogen (N), oxygen (O) or fluor (F). The weak type occurs when hydrogen is bonded to elements with lower electronegativity (*e.g.* sulfur or carbon with sp_2- or sp-hybridization).

(iv)   Ionic/electrostatic interactions: electron transfer from one atom to the other. A low percentage of covalent bond character remains.

Hydrogens involved in H-bonds can lose their covalent character (transfer of acidic protons). Acidity of these protons (hydrogen cation) depends on the nature of the covalent bond and the polarity of its environment. In general, the polarity of a molecule is related to the polarity of the bonds within this molecule. It can be expressed by the dipole moment, which is the vectorial addition of all the dipole moments between the bonds of the molecule. If a molecule is symmetrical, there is no overall dipole moment. Aromatic systems (*e.g.* benzene) are symmetric and have no net dipole moment. Aromatic molecules have a greater electron-density on the face of

totally covalent

nucleus

polar covalent

δ^+ δ^-

Δ**EN**

**bond
polarity**

ionic

+ **–**
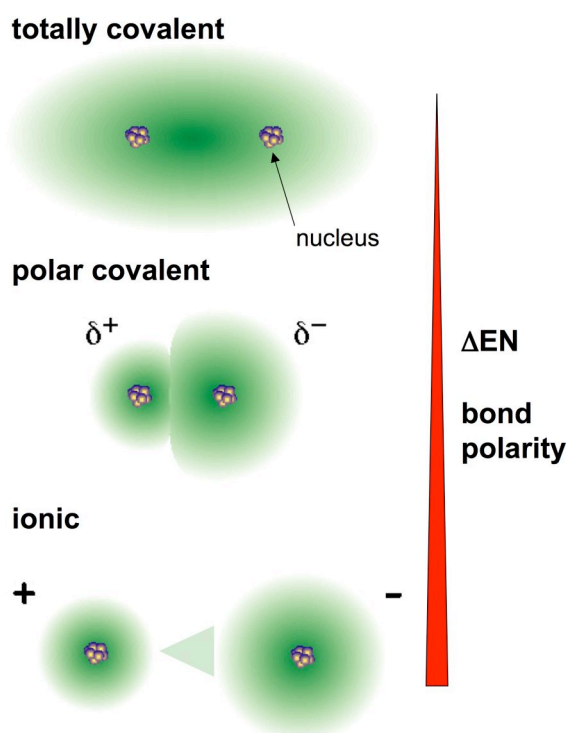
**Figure 1.1 Schematic representation of shared
(valence) electrons in a bond between nuclei with
different electronegativity difference (ΔEN).**
The red wedge indicates the change in the bond polarity
and the electronegativity difference of the nuclei. In
practice, even an ionic bond has at least a partial covalent
character (indicated by the green triangle). The image
was adapted from the homepage of Prof. Grandinetti from
Ohio-State University (www.chemistry.ohi-
state.edu/~grandi).

the ring and reduced on its edge, which gives rise to a quadrupole moment (Waters, 2002 and references therein). Aromatic interactions have been proposed to be composed of van der Waals, hydrophobic and electrostatic interactions (relative contribution is still under investigation). Several geometries between aromatic molecules are attractive. In the model of electrostatic ring interactions (Hunter & Sanders, 1991), the charge distribution in the π-system is represented by a set of point charges. The key feature of this model is that it considers π-electrons and σ-bonding systems separately and reveals π-π-repulsion and π-σ-attractions to be the governing factors in the interaction between two aromatic molecules. The T-shaped arrangement (edge-to-face, Figure 1.2A) is energetically favored (slightly) over the π-π-stacking arrangement (offset stacked, Figure 1.2B) in aqueous solutions. The electrostatic component is responsible for orientation of the two molecules respect to each other, while Van der Waals interactions seem to be the source of stabilization.

Not only interactions between aromatic systems can be observed. There are a variety of possible interactions from aromatic systems to different interaction partners, *e.g.* H-bonds (π-H-bonds, longer distance and considerably weaker) or interactions from the face of the aromatic ring to a cation (π-cation interactions, Mitchell *et al.,* 1994; Figure 1.2C).

Thus, the three aromatic amino acids (phenylalanine, tyrosine and tryptophan) can form a variety of interactions through their ring-moiety and can be seen as the global players in protein-protein interfaces (Meyer *et al.,* 2002 and references therein; see also section 1.4.4).
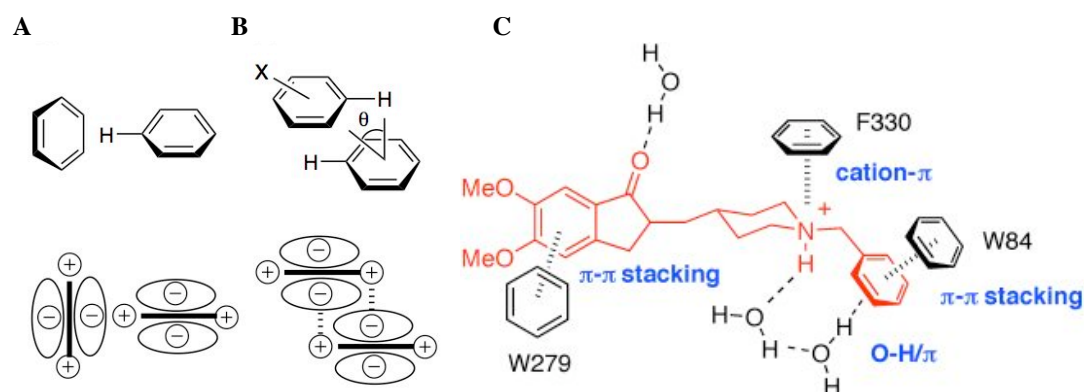
Figure 1.2 Geometries of aromatic interactions in proteins.
(**A**) edge-to-face (**B**) offset stacked (also observed in DNA base stacking). The two geometrical arrangements (**A/B**) are energetically almost equivalent (image from Waters, 2002). (**C**) Binding mode of the anti-Alzheimer drug E2020 in the active site of acetylcholinesterase from *Torpedo californica* (image adapted from Meyer *et al.,* 2002). Interactions between aromatic residues as well as from aromatic residues to other polar species can be observed.

1.2   Water

Water is composed of the most frequent (hydrogen) and the third most frequent (oxygen) elements in the universe and is the only pure substance found naturally in all three states of matter. It is implicated in many biomolecular processes, which is due to its polar properties and the fact that the majority of biomolecules act in an aqueous environment. As a result, it has been called the "molecule of life", the "matrix of life", "life's natural habitat" and similar denominations, that underline its biological importance (Finney, 2004).

A single water molecule is composed of a central oxygen atom (O) forming σ-bonds to two hydrogen atoms (H) (Figure 1.3A). Two electron pairs complete the distorted tetrahedral coordination sphere of the oxygen (104.5° HH-angle vs. 109.5° *e.g.* in methane). The molecule is a dipole (dipole moment: 1.85 D, Clough *et al.,* 1973) with partial negative charges located in the region of the oxygen atom and partial positive charges on each of the hydrogen atoms (Figure 1.3B).
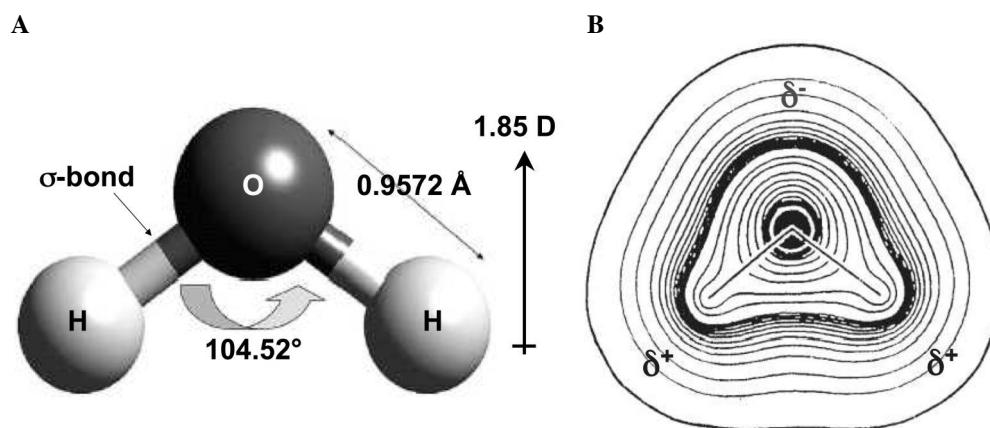
Figure 1.3 Geometry of water.
(**A**) Average geometry of a water molecule (adapted from Finney, 2004). (**B**) Contours of total electron density in the HOH plane, from quantum mechanical calculations by Hermansson (1984) with added partial charges.

The areas with different charges between water molecules are attracted to each other (H-bonding). This attraction explains many properties of water (*e.g.* its high melting point compared to binary hydrogen compounds of the direct neighbors of oxygen in the Periodic System of Elements). The energy of this H-bond is approximately 20 kJ/mol, which is in between that of an ionic and a Van der Waals interaction.

The preferred coordination geometry of water is tetrahedral (four-coordinated motif), which is central to liquid and solid water (Figure 1.4A). Neutron diffraction experiments indicated the existence of three-coordinated waters as well. The close location of the two lone-pairs results in a broad area with negative charge (Figure 1.4B) that even allows five-coordinated waters (Finney, 2004). This reflects the highly dynamic behavior and fluctuation of H-bonds in liquid water with average of four-coordinated water and a broad variety of OOO angles.

Water also shows an amphoteric character: dissociation to protons (H^+) and hydroxyl ions (OH^-). Other water molecules abstract protons to form hydroxonium ions (H_3O^+). The conduction of the excess proton in liquid water is very high. Electron and neutron Compton scattering experiments indicated an effect ascribed as attosecond quantum entanglement of protons (Chatzidimitriou-Dreismann *et al.*, 2003): neutrons and electrons colliding with water are scattered. The scattering indicates, that they are affected by a ratio of 1.5:1 of hydrogen to oxygen (in the attosecond timescale, 10^{-18} s). It has been suggested that in these timescales water reacts rather like $H_{3/2}O$ instead of H_2O.
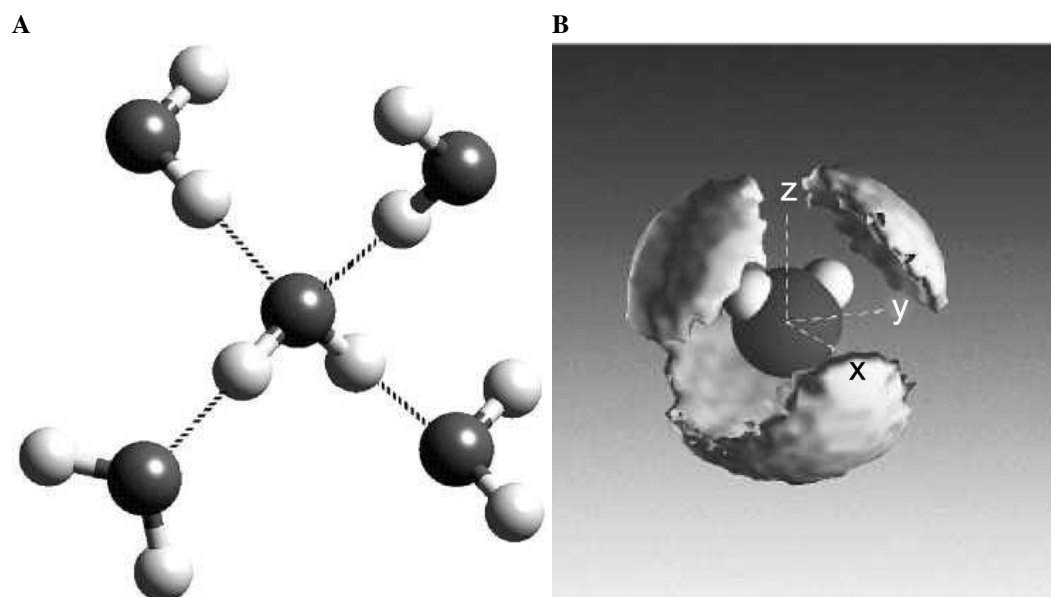
Figure 1.4 Coordination sphere of water.
(**A**) Representation of the tetrahedral coordination sphere of a water molecule. Each of the upper partners in the sphere donate a H-bond and each of the lower partners accept one, respectively. Dashed lines indicate H-bonds. The color code is as in Figure 1.3. (**B**) Spatial distribution function, which illustrates through lobes the average distribution of neighboring water molecules in liquid water (images adapted from Finney, 2004).

1.3   Solvation of molecules: hydrophobicity/hydrophility

The hydrophilic and/or hydrophobic properties of a given molecule dictate its behavior in a polar aqueous solution since, in general, favorable interactions are formed between groups of similar nature. A "hydrophile" is a molecule that has a certain polarizability or intrinsic polarity. The interaction between a hydrophile (functional group or molecule) and water obeys the requirements of a tetrahedral coordination and sufficient H-bonding.

The term "hydrophobicity" is commonly used to describe the low solubility of non-polar groups in water. A classical qualitative explanation of the hydrophobic effect is based on the weak interactions between water and non-polar substances versus the strong H-bonding interactions (high cohesive energy density) in the matrix of water (Lazaridis, 2001). Water adopts a clathrate-like arrangement around small non-polar solutes. These solutes do not aggregate but their clathrate cages share an interface that separates the cavities occupied by the solutes. Larger solutes are too big for clathrate cages. The resulting aggregation minimizes the exposed area of these molecules to water (Dill *et al.*, 2005). When a molecule contains polar as well as non-polar regions it is termed an amphiphile (*e.g.* lipids).

The organized folding behavior of proteins placed in aqueous environments is driven by the polar properties of water. It is important to understand protein three-dimensional structure as having evolved in water.

1.4   Proteins

Proteins are the most complex and abundant biomolecules in Nature. They must fold into complicated three-dimensional structures to become active (see section 1.4.1). The three-dimensional arrangement (fold) of the polypeptide chain is determined by the properties (*e.g.* size and polarity) of its amino acids. However, the mechanism underlying the folding event on the basis of sequence is currently not understood.

Protein sequences are shaped by a complex interplay of different selective pressures. The increasing number of sequenced genomes gives more and more insights into relationships between proteins of different organisms and their putative mechanisms of evolution. However, protein evolution is still poorly understood. It has been shown that a large number of known structures belong to a surprisingly small set of basic folds, so that the evolution of proteins is believed to be due to duplication and combination of a limited "toolkit" of domains (see sections 1.4.2 & 1.4.3). In addition, the native structure and functionality of a protein are often dependent on specific association states. This oligomerization can involve copies of a same protein (self-assembly) or can involve hetero-complexation with other proteins or biomolecules. It can also be of a transient or permanent (obligatory) nature. To date, little understanding exists of how the necessary macro-molecular recognition process as well as inter-subunit communication has evolved in the independent components to generate complex functional assemblies.

1.4.1   Protein folding

Protein folding is thought to be driven by hydrophobicity (Tanford, 1978; Rose *et al.,* 1985). Globular proteins are characterized by a tight packing of hydrophobic side-chains in the so-called "hydrophobic core". Formation of many favorable van der Waals interactions as well as exclusion of solvent molecules

maximize hydrophobic stabilization (Dill, 1990). Classical hypotheses describe the folding process of a polypeptidic chain as a nearly sequential series with discrete intermediates, where folding is initiated by the hydrophobic-collapse of the unfolded chain. In the next step, specific interactions are optimized; this fine-tuning shifts the equilibrium further towards the native state. Specific electrostatic interactions play a critical role in reaching the native fold as well as in stabilizing secondary structure elements, which assemble further by mutual conformational selection into folding units. These units contain a sufficiently large, buried, hydrophobic core and are capable of independent thermodynamic existence. One or more hydrophobic core units associate to form domains. Domains associate to subunits and subunits associate to form quaternary structures (Kumar & Nussinov, 2002 and references therein).

The energy landscape theory of folding considers multiple pathways of organization of partially folded intermediates towards the native fold. A funneled energy landscape (Figure 1.5) is responsible for the robust ability of proteins to fold (Onuchic & Wolynes, 2004).
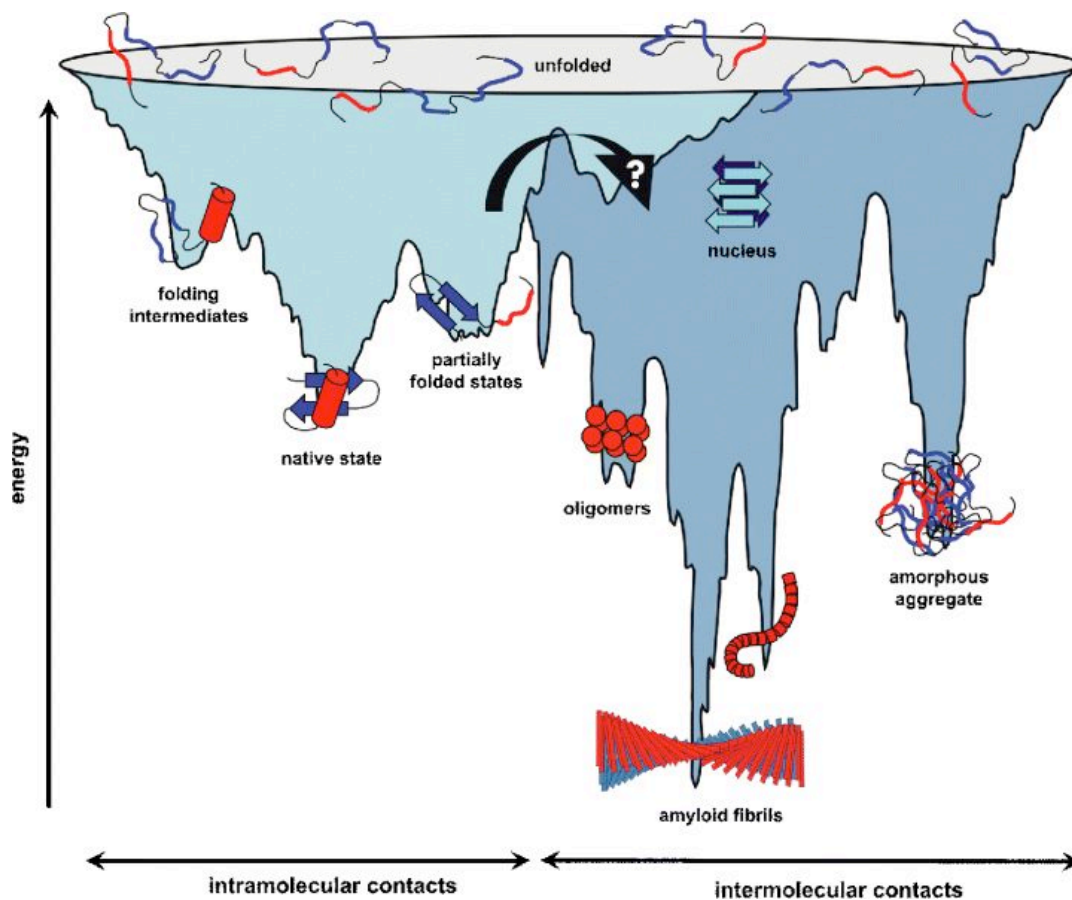
Figure 1.5 Schematic energy landscape for protein folding and aggregation.
The surface shows the multitude of conformations "funneling" towards the native state via intramolecular contacts (adapted from Jahn & Radford, 2005).

The surface of the folding funnel is specific of each polypeptide sequence and is determined by both kinetic and thermodynamic properties of the folding chain. In general, small single domain proteins (<100 residues) reach the native state in a sub-second timescale (smooth folding landscape with two species populated separated by a transition state barrier). Larger proteins (>100 residues) reveal a rougher energy surface with folding intermediates. In large multidomain proteins different regions fold in parallel, while a final step establishes all native intra- and interdomain contacts (Jahn & Radford, 2005 and references therein).

Partially folded states along the landscape may form favored intermolecular interactions resulting in aggregation. These aggregates can be the source of disease (Jahn & Radford, 2005). For example, amyloid fibrils are found to be insoluble deposits of aggregates involved in disorders like Alzheimer's and Parkinson's disease as well as type II diabetes and Creutzfeld Jacob disease.

Thermophilic adaptation of proteins can give insights into those factors that stabilize the fold of proteins without affecting their functionality. Stabilizing interactions have been suggested based on a comparison of structures of homologous pairs from mesophiles and thermophiles (Jaenicke & Böhm, 1998; Sterner & Liebl, 2000):

(i)   Increased H-bonding (higher side-chain involvement an more satisfied H-bonding pattern)

(ii)   Networks of improved electrostatic interactions

(iii)   Optimized hydrophobic interactions

(iv)   Higher packing density (lower number, volume and area of cavities)

(v)   Increased polar surface areas

(vi)   Higher α-helical content and stability

(vii)   Binding of metal ions

(viii) Fixation of the termini to the protein core

(ix)   Shortened solvent exposed loops

(x)   Residues with energetically unfavorable conformations mutated to Gly

(xi)   Higher Pro and β-branched residue content in loops

(xii)   Oligomer formation

(xiii) Reduced Asn, Gln, Cys and Met content (thermolabile amino acids).

Additionally, stabilizing effects were suggested on the basis of comparison of sequences in the genomes of mesophilic and thermophilic organisms:

(xiv)   Decreased polar uncharged residues (Asn, Gln, Ser, Thr; Haney *et al.,* 1999)

(xv)   Increased charged residues, especially Glu, Arg, Lys (Cambillau & Claverie, 2000).

(xvi)   Lower average size of thermophilic proteins (Chakravarty & Varadarajan, 2000).

A problem in the analysis of hydrophobic interactions is the temperature dependence of hydrophobic stabilization. The effect is enthalpic at room temperature and entropic at higher temperatures with the maximum at 75°C (Makhatadze & Privalov, 1995). On the contrary, the stabilizing effect of electrostatic interactions has been suggested to be proportional to temperature (Elcock, 1998; Xiao & Honig, 1999). Due to the lack of information on stabilizing effects of other interactions a general model for thermophilic stabilization is not possible at this time (Sterner & Liebl, 2001).

1.4.2   Protein fold families

The data retrieved by sequencing of genomes of organisms from all three kingdoms of life together with structural data derived from X-ray crystallography and NMR experiments is expanding our knowledge of structural families. Methods for the detection of sequence similarity have also advanced over the past years resulting in a gain of sensitivity. These advances make it possible to map domains from structural families onto genome sequences. The procedure reveals how protein families are distributed throughout the genomes and how they might influence functional repertoires and biological complexity.

There seems to be a universal set of domain families (approx. 140) throughout the three kingdoms of life. Families involved in protein biosynthesis are the largest conserved group in this universal set (Figure 1.6A & B). Approximately half of the domain sequences of organisms with known genome have been recognized as relatives of CATH families. CATH (Gerstein & Levitt, 1997) is a domain structure

database that hierarchically classifies protein domain structures and organizes them in a Dictionary of Homologous superfamilies (DHS; Pearl *et al.,* 2005). For a scheme showing the distribution of frequently assigned folds see Figure 1.6C.
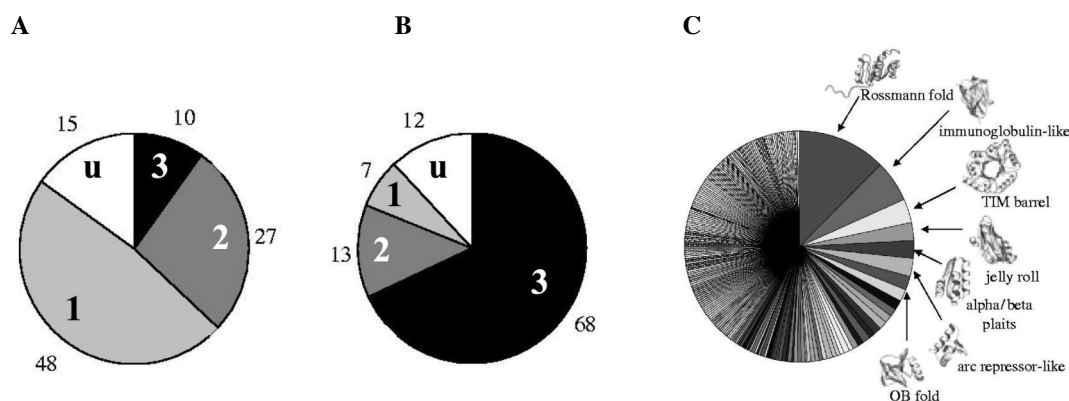
Figure 1.6 Protein families and folds.
Average proportions of genome sequences in common protein (**A**) and domain (**B**) families. (**A/B**) Numbers in the segments of the circles indicates whether a fold is present in one, two or three kingdoms of life or is unique (u) to one of them. The numbers around the diagrams represent the folds corresponding to the wedges in the diagram. (**C**) Distribution by fold in the CATH database. The angles subtended by the sectors correspond to the frequency of occurrence of the protein folds, measured by the number of close sequence families within each fold group (adapted from Marsden *et al.,* 2006).

1.4.3   Protein evolution

Proteins that share a common ancestor are called "homologous". They share a common three-dimensional fold (to a certain extent) and, often, also share common active sites or binding domains. Homologous sequences can be divided into orthologs (occur in different species) and paralogs (originating from a gene duplication event). The rate of divergence varies for each protein family investigated. Thus, the number of differences cannot be used to estimate the period over which the proteins diverged. This is particularly true for paralogs, since once a sequence has duplicated it may change very rapidly before selective pressure on its new function slows its rate of change.

Usually, sequence conservation occurs in the evolution of a protein when the selective pressure on different amino acids is high due to functionality. Conserved sequences normally have conserved functions. However, the conservation of function, not the conservation of sequence, is desired (Kirschner & Gerhart, 1998). Function

can be retained even with a low homology. Nevertheless, structural conservation is normally higher than sequence conservation and allows detection of remote homologies, for which all similarity has been washed out from the sequence (Marsden *et al.,* 2006).

To this date, efforts to understand protein evolution have mostly focused on domains (Figure 1.7), the independent folding units from which modern proteins are formed. Analyses have confirmed the extent to which some very common domains are duplicated, shuffled within a genome and combined in different ways (Vogel *et al.,* 2004a). This follows a mosaic model of constructing new proteins that can subsequently evolve modified functions, thereby expanding the functional repertoire of the organism (Vogel *et al.,* 2004b).
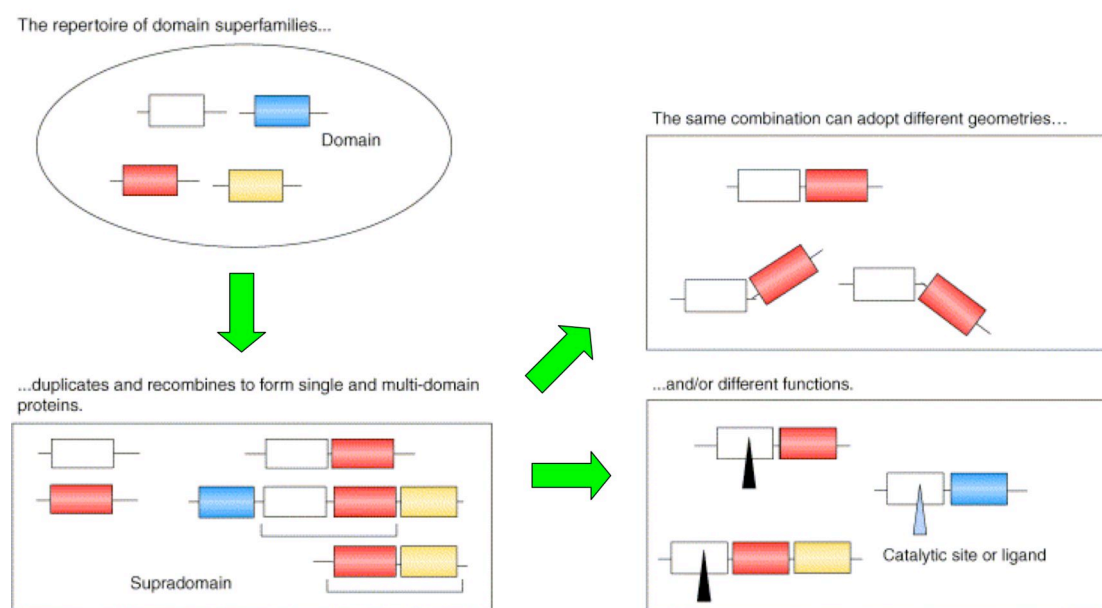
Figure 1.7 The role of domains in protein evolution.
Overview of different aspects of multidomain proteins: the repertoire of domain superfamilies and their role in the formation of multidomain proteins by duplication and recombination, and the geometry and functional relationship of domains within these combinations. Domains belonging to the same superfamily are represented by rectangles of the same colour. Supradomains are two- or three-domain combinations, which occur in different architectures with different N- and C-terminal neighbours, as shown in the second panel. These short series of domains form functional units that are reused in different protein contexts (adapted from Vogel *et al.,* 2004).

In the evolution of proteins and the complicated interactions between them, the coupling of processes can be separated into strong and weak linkages:

(i)   Strong linkage: strong dependence of one process on another occurs when two or more proteins aggregate into an active complex or when the product of an enzyme is the substrate of another enzyme. Steric requirements are high, and the complementary fit of surfaces of interacting components is precise. This formation of obliged, strictly interconnected complexes imposes a "forced proximity" that ensures a rapid information exchange between subunits (for example, substrate transfer) and eliminates delays that could result from cytosolic crowding and diffusion limitation. It can also provide a unique, local physicochemical environment leading to an increase in molecular stability, specificity, affinity and even enhanced activity through cooperativity.

(ii)   Weak linkage: activity of a process depends minimally on other components or processes. Weak linkage often occurs in regulatory pathways, *e.g.* signal transduction, neural relays or transcriptional control circuits. The components often have switch-like capacity to exist alternatively in active and inactive states, and signals release the innate activity. Weak linkage shows tolerance, flexibility and robustness.

Thus, the capacity of a lineage to evolve (evolutionary adaptability or "evolvability") depends on the linkage of the components. Evolvability is higher for proteins that are able to interact with common structural elements (Beltrao & Serrano, 2007). Based on the analysis of atomic structures and kinetic data, three pathways for the evolution of protein dimers (also applicable to other oligomers) have been suggested (Xu *et al.,* 1998). The first pathway involves formation of a functional dimer directly without going through an ancestor monomer. In the second pathway, residues in stable monomers mutate towards a complementary surface. Such dimers, evolved from ancestral monomers, are likely to have small interfaces with few key residues critical for binding. Third, the domain swapping mechanism replaces one segment in a monomer by an equivalent segment from an identical chain to form a dimer. These dimers contain at least one stable domain in a chain.

Units within a dimer can further evolve by gene duplication, which can initialize functional divergence (Pereira-Leal et al., 2007; Figure 1.8) to form large protein complexes. Oligomerization of paralogs and identical chains seems to be key to the evolution of protein-protein networks (self-interacting proteins have twice as many functionally related interaction partners as non-dimers; Ispolatov *et al.*, 2005). In fact, most cellular proteins are tightly embedded into biological networks that function following a modularity principle. The recognition of this principle has recently triggered the study of protein-protein interactions in complexes, pathways, and even complete organisms, that has become the new paradigm in protein biology. Modularity, whether based on components as parts of stable protein complexes or as dynamic units that interact only transiently (as in signaling and metabolic cascades), facilitates the combinatorial generation of complexity in protein networks by allowing the re-wiring of modules to add to the diversification of individual proteins. This increases the "evolvability" of the system and facilitates the accommodation to novelty through accumulation of non-lethal mutations.

The mechanisms that drive the emergence and subsequent evolution of modularity in cellular networks remain unclear to this date.

A

B
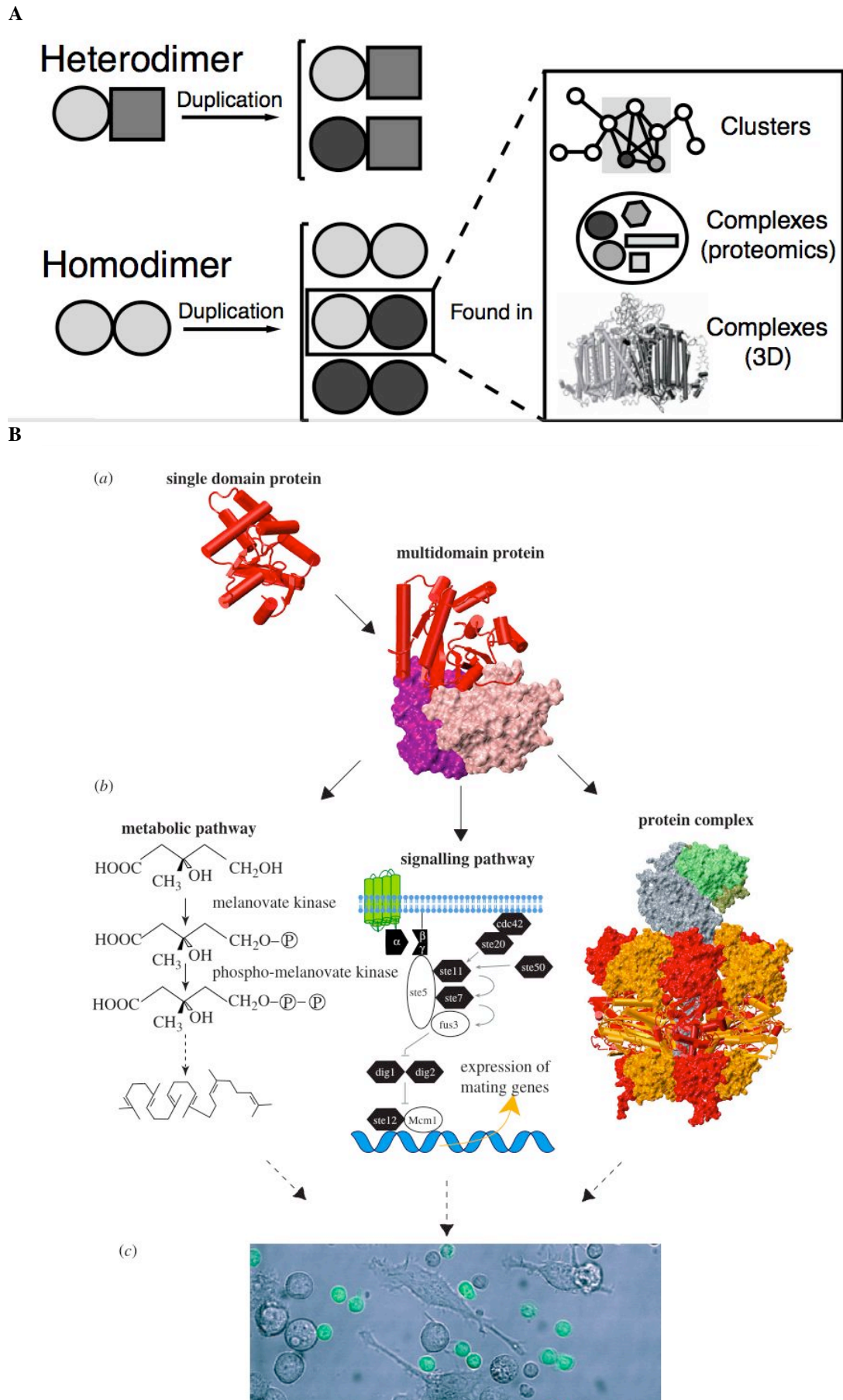
Figure 1.8 Hypothesis for origins and evolution of protein complexes.

Figure 1.8 Hypothesis for origins and evolution of protein complexes.
(**A**) Gene duplication with conservation of protein-protein interactions is frequent. Self-interactions have special structural properties that are conserved into the duplicated interaction between paralogous proteins (light-dark interaction). Interactions between paralogous proteins are more versatile functionally and structurally, and are systematically selected in the evolution of protein interaction networks and complexes. (**B**) Hierarchical modularity in biology. (a) Modularity at the protein level - proteins consist of modules formed by domains. The *S. cervisiae* uridylate kinase [1ukz] contains a single P-loop containing nucleotide triphosphate hydrolase domain as defined in the SCOP database (Murzin *et al.*, 1995; Andreeva *et al,.* 2004). A domain from the same superfamily is also found in the multi-domain protein EF-TU from *T. thermophilus* [1exm], also shown as a red cartoon. The two other domains belong to the SCOP superfamilies translation proteins (pink) and EF-TU/eEF-1alpha/eIF2-gamma C-terminal domain (violet). (b) Modularity at the cellular level - most proteins work in a cooperative manner with other proteins and form functional modules. Three distinct types of functional modules are shown: (i) a protein complex - the *B. taurus* ATP synthase [1e79]. Here six protein chains (three red and three orange) that all contain the P-loop containing nucleotide triphosphate hydrolase domain (cartoon representation) as well as several other domains, assemble and form a ring; (ii) a signalling pathway - mating response MAPK pathway in yeast (Schwartz & Madhani 2004); (iii) a metabolic pathway - mevalonate pathway. (c) The diversity of cellular types is a consequence of the distinct arrangement of modules at lower levels, and cells are themselves modules from which tissues are built, such as the dendritic cells (grey) and lymphocytes (green) shown here (images adapted from Pereira-Leal *et al.*, 2006/2007)

1.4.4   Protein-protein interactions

Protein-protein recognition is essential to a variety of cellular processes. It is implicated *e.g.* in signal transduction, antigen-antibody binding, regulation of gene expression, enhancement of structural stability and functionality. There is a huge variety of homo- and hetero-multimers, where the multimer represents the biologically active state. The strength of contacts at protein interfaces can even exceed that of the core of the proteins involved (Brinda & Vishveshwara, 2005). Several variables are involved in protein-protein interactions (Levy & Onuchic, 2006):

(i)   Protein side-chain and backbone flexibility
(ii) Solvent effects
(iii) Hydrophobic/non-polar interactions
(iv) Electrostatic interactions/H-bonds

The strength of non-covalent interactions is lowest for hydrophobic interactions ($\approx$10 kJ/mol) followed by H-bonds ($\approx$20 kJ/mol) and highest for salt bridges ($\approx$30 kJ/mol). The contribution of hydrophobic residues to the stabilization energy of interfaces is similar to the stabilization energy in protein folding (Guharoy & Chakrabarti, 2005). Several studies have reported that electrostatic

interactions can accelerate recognition by the long-range attraction that controls the formation of the complex. The interface gains stability by forming hydrophobic contacts and H-bonds (Levy & Onuchic, 2006 and references therein).

The biophysical characterization of protein-binding interfaces has been achieved through Ala-scanning mutagenesis (Clackson & Wells, 1995; DeLano *et al.,* 2000; DeLano, 2002). A hot spot has been defined as a residue that, when mutated to Ala, leads to a significant drop in the binding constant (typically 10-fold or higher) as determined by the change in the free energy of binding ($\Delta\Delta G$). Positioning of a residue in the core region of an interface and simultaneous H-bonding make it critical for binding (Guharoy & Chakrabarti, 2005). Systematic analysis identified Arg, Tyr and Trp as common hot spots, since these residues occur with high frequency and provide strong interface stabilization (usually ≥ 2 kcal/mol; Thorn & Bogan, 2001; Bogan & Thorn, 1998).

Every interface between two proteins can be divided into core (buried region) and rim (partially solvent accessible region surrounding the core) with an average area contribution of 75 % (core) and 25 % (rim). These results have been obtained from interfaces of homodimers (Bahadur *et al.,* 2003) and complexes (Chakrabarti & Janin, 2002). Composition of the core is similar to the protein interior and composition of the rim similar to the protein surface. The rim region shows preference for conserved Gly. This residue (together with Pro) has been suggested to have a structural role. Leu and aromatic residues are the highest ranked core residues. Arg, an interface hot spot residue, can be observed frequently in both regions (the rim and the core), which is likely due to its ability to form hydrophobic as well as polar interactions (Caffrey *et al.,* 2004). In general, residues involved in formation of interfaces are usually more conserved than the remainder of the protein surface (core- over rim-conservation, Guharoy & Chakrabarti, 2005). Interfaces of homodimers show a lower degree of conservation than other interfaces. Gaps in interface-forming sequence patches are more frequent in transient interfaces (Caffrey *et al.,* 2004).

Protein-protein interfaces have been investigated extensively, *e.g.* Cole & Warwicker (2002), Bahadur and coworkers (2003), Chakrabarti & Janin (2002), Guharoy & Chakrabarti (2005) and Ofran & Rost (2003). These studies differ in several aspects, *e.g.* the number of investigated structures or the groups of multimers defined. The challenge in this case is to intercorrelate their results, which

are in part contradictory, since results vary with the interface groups defined by the authors. Thus, a clear principle for the formation of interfaces is yet to be established. Nevertheless, some tendencies and preferences in residues and hot spots in interfaces have been found. The largest set of non-redundant interfaces has been investigated by Ofran & Rost (2003). The authors differentiated between interfaces within one chain (groups i and ii below), which are relevant to protein evolution, fold and functionality; and those formed between two chains (groups iii to vi) that are characterized by the chains involved (identical vs. different) and the duration of interface formation (transient vs. permanent). The data given in other publications have been added here to the classification by Ofran & Rost:

(i) Intra-domain: within one structural domain. Aliphatic and aromatic residues are overrepresented. The hydrophobic effect dominates the fold (this can also involve aliphatic moieties of charged residues).

(ii) Domain-domain: between different domains in one chain, which can be regarded as an intermediate between binding and folding surface. The amino acid composition of the interface resembles that of hetero-obligomers below (independently folded chains) since domain-domain (independently folded units) interfaces also show amino acid compositions similar to that of a protein surface (Jones *et al.,* 2000).

(iii) Homo-obligomer (from <u>oblig</u>ate oli<u>gomer</u>): permanently interacting identical chains show preference for interactions between identical amino acids. The average subunit interface is bigger than in complexes. Non-polar interactions dominate interface. The most abundant residues are Leu and Arg (with Asp, Glu and main chain carbonyl as acceptors for the satisfaction of polarity), followed by other aliphatic residues (Val, Ile, Met). Interfaces are enriched in aromatic residues (Bahadur *et al.,* 2003). Defined protein aggregation can be critical for the pathway to the native fold (coupled binding/folding event in homodimers).

(iv) Homo-complex: between transiently interacting identical chains. Depleted in salt bridges, under-represented in Trp (compared to other interfaces) and rich in contacts between identical residues. Binding of homodimers occurs between already folded monomers (Levy & Onuchic, 2006).

(v) Hetero-obligomer: between permanently interacting different protein chains. Depleted in hydrophobic residues compared to homo-multimers (groups iii and iv together).

(vi) Hetero-complex: between transiently interacting different protein chains. Preference for Arg over Thr as well as enriched in aromatic residues in the core region. Charged residues are depleted (except Arg). The rim region resembles the remainder of the surface. The aromatic residues Phe and Tyr have been suggested to be the main driving force for interface formation (Chakrabarti & Janin, 2002). Hydrophobic residues interact to the aliphatic moieties of charged residues.

The divergent composition of these protein-protein interfaces (i to vi) as investigated by Ofran & Rost (2003) is illustrated in (Figure 1.9).
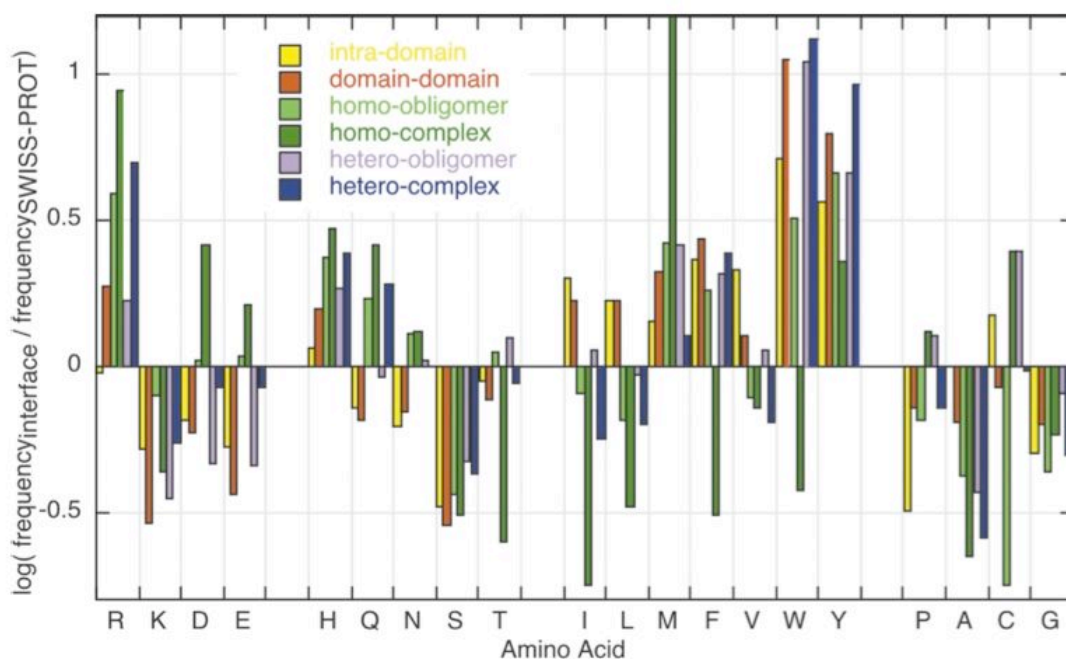
Figure 1.9 Composition of the six interface types defined by Ofran & Rost (2003).
The propensities of all residues found in SWISS-PROT were used as background. If the frequency of an amino acid is similar to its frequency in SWISS-PROT, the height of the bar is close to zero. Over-representation results in a positive bar, and under-representation results in a negative bar. The amino acid residues are identified by their one-letter code, sorted by biophysical features. Residue-residue preferences are shown in Figure 6.1 in appendix 6.2.

In general, protein-protein interfaces are depleted in Lys, Ser, Gly and Ala residues and show high preference for aromatic residues as well as Arg, His and Met. The Arg residues preferentially interact with the acidic residues across the interface. Aromatic and hydrophobic residues form interactions to each other and within their

groups. Biophysical similar residues (*e.g.* Leu and Ile or Asp and Glu) show similar trends across the different types of interfaces. Hydrophobic interactions are more frequent in permanent over transient interfaces (Jones & Thornton, 1997; Lo Conte *et al.*, 1999).

Water is also of importance in protein-protein interactions. A dominant role of water in binding is obviously expected especially in hydrophilic interface regions. The involvement of a water molecule is favorable when the enthalpy gain is higher than the entropic cost for immobilization (for protein folding and interface formation). On the other hand, bulk solvent exclusion, allowing direct protein-protein contacts, creates a favorable hydrophobic effect (Levy *et al.*, 2004). The aliphatic moieties of hot spot residues are usually surrounded by hydrophobic residues to occlude bulk water. The analysis of crystal structures revealed that a number of water molecules remain at interfaces and contribute to their stability (10-11 waters per 1000 Å^2 in oligomers and complexes, 15 in crystal contacts). Preferred amino acids with water mediated interactions are: Asp, Glu and Arg. Waters involved in interface formation can form rings around the interfaces as well as being buried within (Rodier *et al.*, 2005).

Determinants of protein-protein interfaces constitute hot spots, where a small fraction of residues cluster together resulting in strong stabilization. Interface hot spot residues are often those capable of forming the highest amount of different interactions: Arg (hydrophobic, H-bonds, electrostatic) and Tyr/Trp (hydrophobic, H-bonds, π-interactions). The folding of a protein can be seen as a hierarchical interface formation procedure: secondary structure elements form interfaces towards a domain-fold, domains form interfaces resulting in a pseudo-quarternary structure within one chain and different chains associate over interfaces to shape the quarternary structure of a protein. Relative contributions of hydrophobic and hydrophilic interactions differ in protein folding and association (Xu *et al.*, 1998). In this scenario, temporary interfaces represent either an evolutionary intermediate of process enhancement (see Figure 2.4) or interaction between units that have to be reversible (*e.g.* receptor-ligand complexes). Thus, the relationship of protein association between the interfaces of different hierarchical levels should be seen as a fluent process of evolutionary shaping, where increasing hydrophobicity is characteristic of a more obligate interface.

1.5   Aim of this work

The focus of this work was to investigate the structural basis of protein-protein interactions at different stages of evolutionary shaping using X-ray crystallography and homology modeling. Two protein components have been the primary target of this study: Tryptophan Synthase B2 from *S. solfataricus* and the poly-FNIII tandems from the A-band region of the human muscle filament titin.

Tryptophan Synthase (TrpS) is a metabolic enzyme from the tryptophan biosynthesis pathway that catalyzes the last two steps in the synthesis. It is a hetero-enzyme, whose quaternary structure is composed of a TrpA and TrpB subunits. This hetero-assembly has developed an intersubunit communication in the form of substrate channeling and mutual allosteric regulation that results in a direct catalytic advantage. However, ancestral TrpA and TrpB enzymes that either complex only transiently and have an impaired allosteric communication or even do not interact at all, have been retained in archae organisms. Structural changes in the enzyme products further correlate with gene translocation events within the genome, where genes coding for interacting forms of the enzymes have become incorporated into operon units subjected to a common regulation of their transcription. This makes of TrpS an ideal model system for the study of protein network evolution. In this work, we analyze an early snapshot of TrpB evolution, corresponding to a form of this enzyme (TrpB2b) that does not interact with TrpA and whose gene is not part of the *trp*-operon.

A frequent mechanism of protein evolution is the duplication, shuffling and combination in various ways of smaller independent folding units (domains). In this way functional complexity can be achieved by the re-wiring of a limited set of common building blocks. This mechanism often leads to the formation of the more complex multi-domain proteins encountered in higher organisms. In these, the evolution of domain-domain interactions is essential to the integration of the different modules into a global functional unit. In order to analyze whether the evolution of domain-domain interactions resembles that of inter-subunit recognition and communication in obligate hetero-assemblies, we have explored the structure of the FNIII-double tandem A77-A78 from the A-band of titin. This region of the filament is characterized by its composition of distinct super-repeats of IG- and FNIII-domains. The strictly regular domain pattern is thought to have arised by gene duplication

events. Contrary to the poly-IG fractions of the filament for which several tandem structures are now available, no atomic models of FNIII tandems of titin others than the one contained in this work exist to date. We have analyzed the structure of the A77-A78 tandem and extracted conclusions on evolutionary changes in tandem interfaces that modulate transition properties within the modular super-repeat structure. The building blocks of this modular system share a high similarity but a small subset of residues is responsible for their mutual arrangement. Here we speculate that changes at the linker regions might be the primary determinant of new discrete functions correlating with modular positions, for example relevant to myosin assembly.

2 *Ss*TrpB2b: an ancestral Tryptophan Synthase

Tryptophan biosynthesis takes place in free-living prokaryotes, lower eukaryotes and higher plants. It comprises seven catalytic activities (from chorismate) that are organized in a *trp*-operon (Yanofsky, 2003 and 2001, figure below). The structural inventory of all seven enzymes from this pathway has been completed and a wealth of data on their biochemistry and evolution exists – making this the second major metabolic route under study after the glycolysis. Further, the increasing availability of multiple genome sequence records makes of this pathway a unique model for the investigation of multi-level protein network modularity.

According to the organism under consideration, Trp enzymes can exist as individual components or as dual, fusion proteins with enhanced catalytic properties. Moreover, operon and non-operon variants – allegedly evolutionary remnants – remain available in several genomes. This is particularly the case for Tryptophan Synthase (TrpS), which catalyzes the last reaction in this biosynthesis. In the TrpS family, three enzyme classes exist with different association properties and allosteric communication. They represent different evolutionary states of macromolecular recognition and integration within a modular network. This work aims at the comparison of representatives from these families at the molecular level, where TrpS is to serve as a paradigm to determine the evolutionary shaping required to achieve specific recognition, assembly and inter-subunit communication as well as modulation of activity during network evolution.

Gln   Glu

TrpD

TrpE

N–(5'–phosphoribosyl)–
anthranilate

PRPP

(Mayans *et al.*, 2002)

TrpG

Anthranilate

(Knoechel et al, 1999)

TrpF
(Priestle et al, 1986)

Chorismate

1–(o–carboxyphenylamino)–1–
deoxyribulose–5–phosphate

TrpC

TrpA

Tryptophan

TrpB

glyceraldehyde–
3–phosphate   Ser

(Hyde et al., 1988)

indole–glycerolphosphate

25
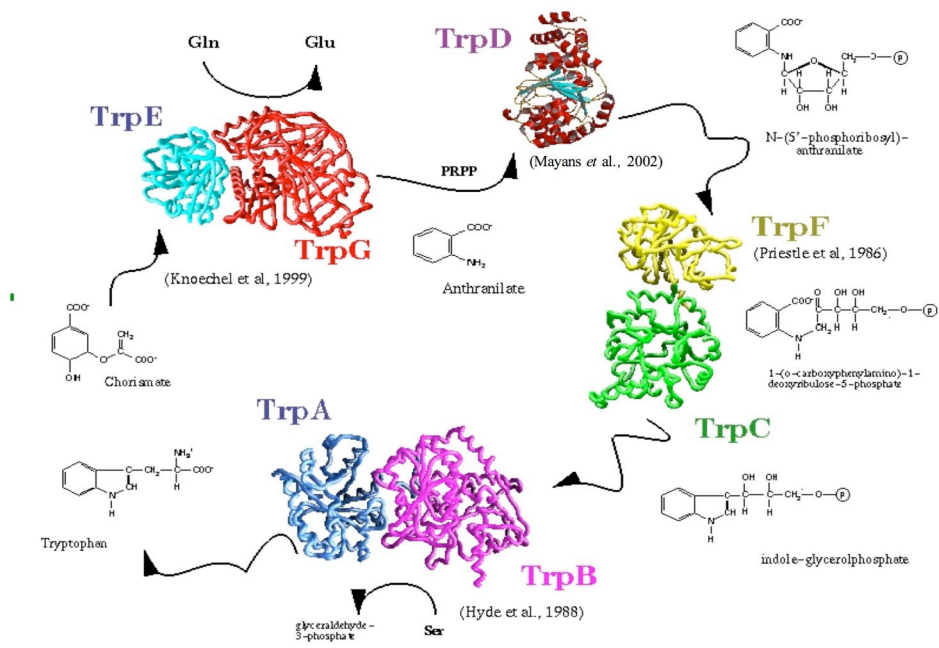
2.1   Abstract

There are two evolutionarily related families of the Tryptophan Synthase (TrpS) β-subunit, termed TrpB1 (gene always located inside the *trp*-operon) and TrpB2 (gene inside and/or outside the *trp*-operon). They show different association to TrpA and different allosteric communication depending on the location of the corresponding gene: transcripts encoded by genes inside the operon form hetero-tetramers with TrpA (permanent in the case of TrpB1 and transient in TrpB2), while transcripts (TrpB2 only) encoded by genes outside the operon lack these association properties. TrpB1/TrpA hetero-obligomers show mutual allosteric communication, but TrpB2/TrpA hetero-complexes accelerate only the TrpA reaction. While the TrpB1 family has already been structurally investigated, the atomic structure of a TrpB2 enzyme is yet to be determined.

We elucidated the first structure of a non-operon encoded Tryptophan Synthase from the hyperthermophilic Archeon *Sulfolobus solfataricus* (*Ss*TrpB2b) using X-ray crystallography. The structure of TrpB2b reveals unique structural features built upon a protein core common to both families. The main-differences between TrpB1 and TrpB2 family members cluster in the TrpA interface region indicating that this governs the association properties and allosteric communication of these enzymes. The different structural organization of *Ss*TrpB2b respect to TrpB1 suggests that segments common to both TrpB families as well as unique elements contribute to the *Ss*TrpB2/*Ss*TrpA interface with subsequent activation of TrpA. The different structural organization of TrpB2 family members and the altered sequence in common structural elements are responsible for the lack of allosteric communication from TrpA to TrpB2. The allosteric path of TrpB1/TrpA obligomers is likely to have evolved after establishment of a permanent interaction between these two proteins.

2.2 Introduction: Tryptophan Synthases

2.2.1 Pyridoxal-5'-phosphate (PLP, vitamin B$_6$)-dependent enzymes

Pyridoxal-5'-phosphate (PLP, vitamin B$_6$)-dependent enzymes are classified into four unrelated families (α, β, D-Ala aminotransferase and Ala racemase families; Table 2.1), characterized by independent folds (Christen & Mehta, 2001).

Table 2.1 Structural classification of PLP-dependent enzymes.

Family	Prototype
α	Asp aminotransferase
β	Trp Synthase (TrpB subunit)
D-Ala aminotransferase	D-Ala aminotransferase
Ala racemase	Ala racemase

The α family is the largest among the four families and shows the lowest internal structural homology. The β family, the second largest, is structurally rather uniform. The remaining D-Ala aminotransferase and Ala racemase families are small and comprise three and four members, respectively.

2.2.2 Tryptophan Synthases

Tryptophan Synthase (TrpS, EC 4.2.1.20) forms a hetero-obligomer type multienzyme complex composed of TrpA (α-subunit) and TrpB (β-subunit). The quarternary structure of the enzyme is heterotetrameric (Hyde *et al.,* 1988) with a central β_2-dimer and two α-subunits that interact with opposite faces of the dimer to form a linear $\alpha\beta\beta\alpha$ arrangement. Formation of this quaternary structure enhances the catalytic activity of both TrpA and TrpB components two orders of magnitudes (for *Salmonella typhimurium,* Miles, 1995). The enzyme catalyses the last two steps in the biosynthesis of Trp, also called $\alpha\beta$-reaction (Figure 2.1), where each reaction is catalyzed by one subunit of the enzyme.

28
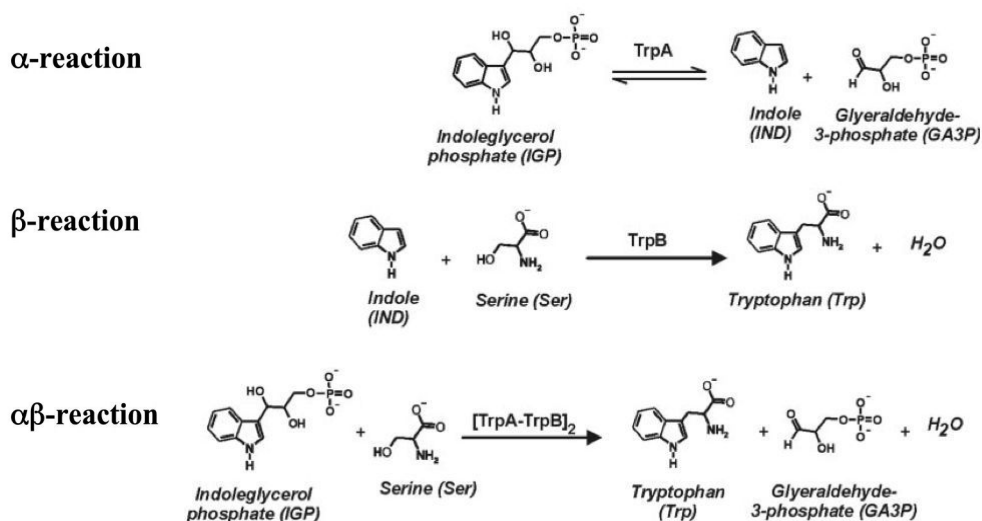
Figure 2.1 Reactions catalyzed by TrpS.
The physiological main-reaction is shown, side-reaction (elimination) yielding pyruvate is not shown (image from Hettwer & Sterner, 2003).

The active sites of TrpA and TrpB are connected via a tunnel (approx. 25 Å length, Figure 2.2). The hydrophobic character of the tunnel supports the diffusion of the product of the α-reaction (indole) to the active site of the β-subunit (Hyde *et al.*, 1988). A tunnel between active sites of two enzymes catalyzing distinct reactions was first observed in TrpS from *Salmonella typhimurium* by Hyde and coworkers (1988). Meanwhile similar channeling has also been observed in other enzymes (*e.g.* Carbamoyl Phosphate Synthetase or Glutamate Synthase; Raushel *et al.,* 2003).

The genes TrpA and TrpB, forming the multienzyme complex, are located on a same *trp*-operon. The genome sequencing of thermophilic microorganisms revealed an additional gene for a β-subunit outside the *trp*-operon. *Trp*B genes became then called *trp*B1 and *trp*B2, to refer to operon and non-operon copies respectively (Hettwer & Sterner, 2002). TrpB1 and TrpB2 protein products have distinct sequence features (as described below). To date *trp*B2 genes seem to be almost exclusive of *Archea*. They have been found in all but five species in the group of *Euryarchaeota* (*Methanococci*, *Methanopyri* and *Halobacteria*). More recently, it has become known that organisms can contain diverse combinations of *trp*B genes, including a *trp*B1/*trp*B2 combination (*e.g. Thermotoga maritima*) or two *trp*B2 genes (*e.g. Sulfolobus solfataricus*). While the *trp*B1 gene is always located in the

trp-operon, in organisms with two *trp*B2 genes one *trp*B2 gene (*trp*B2a) is in the *trp*-operon and the other outside (*trp*B2b). Transcripts from TrpB proteins from this operon have been found to bind to TrpA, where TrpB1 forms hetero-obligomers and TrpB2a hetero-complexes. Non-operon TrpB2b transcripts lack the capability to interact with TrpA, (Leopoldseder, 2005; details below).

The transcripts of these genes form phylogenetically separated groups, termed the TrpB1 and TrpB2 families (Hettwer & Sterner, 2002). Sequence identities within each of the families are approximately 60%, while they are lower between them (approx. 30%). Comparison of the TrpB1 structure from *Salmonella typhimurium* with sequence alignments of TrpB2 enzymes revealed clusters of conserved residues at the active site suggesting a common catalytic activity. However, a long N-terminal extension and two insertions in the vicinity of the TrpA interface suggested an altered association with TrpA (Hettwer & Sterner, 2002).
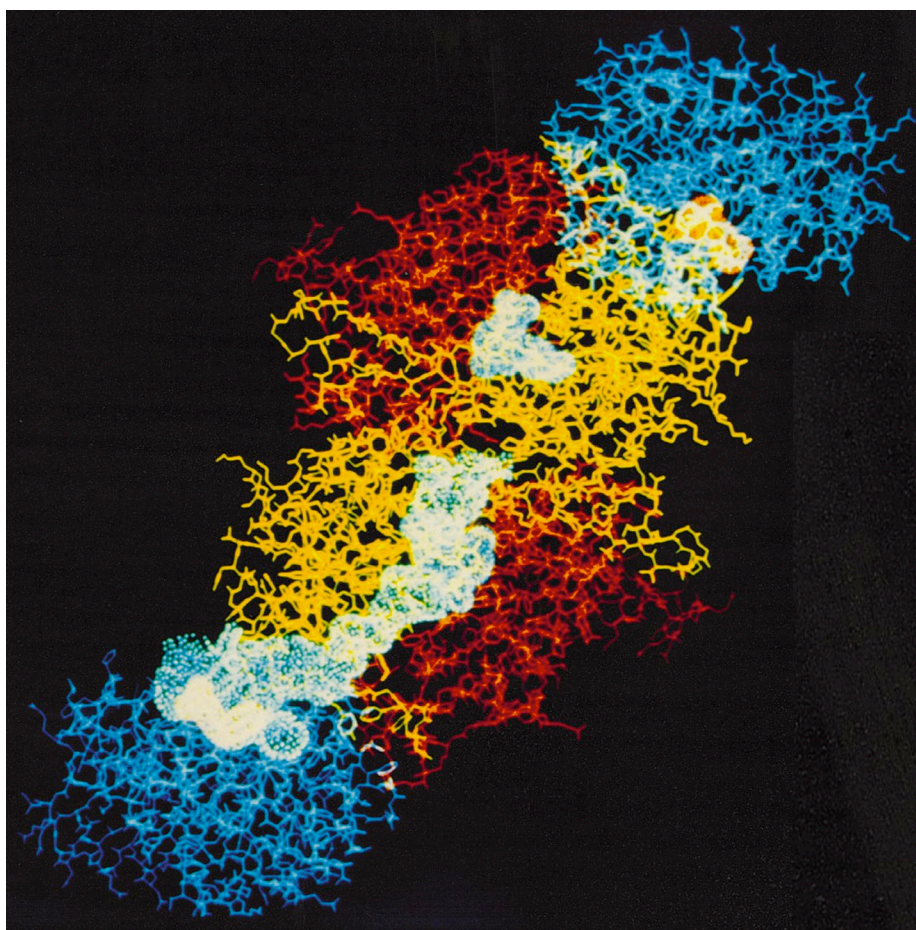
Figure 2.2 Three-dimensional structure of *St*TrpS (Hyde *et al.*, 1988).
The indole tunnel (shown in the lower moiety of the obligomer) passes from the active site of TrpA (blue) and the two domains TrpB1 (yellow/red) to PLP (also shown in the upper moiety of the second TrpB1 monomer) at the active site (modified from Miles, 2001).
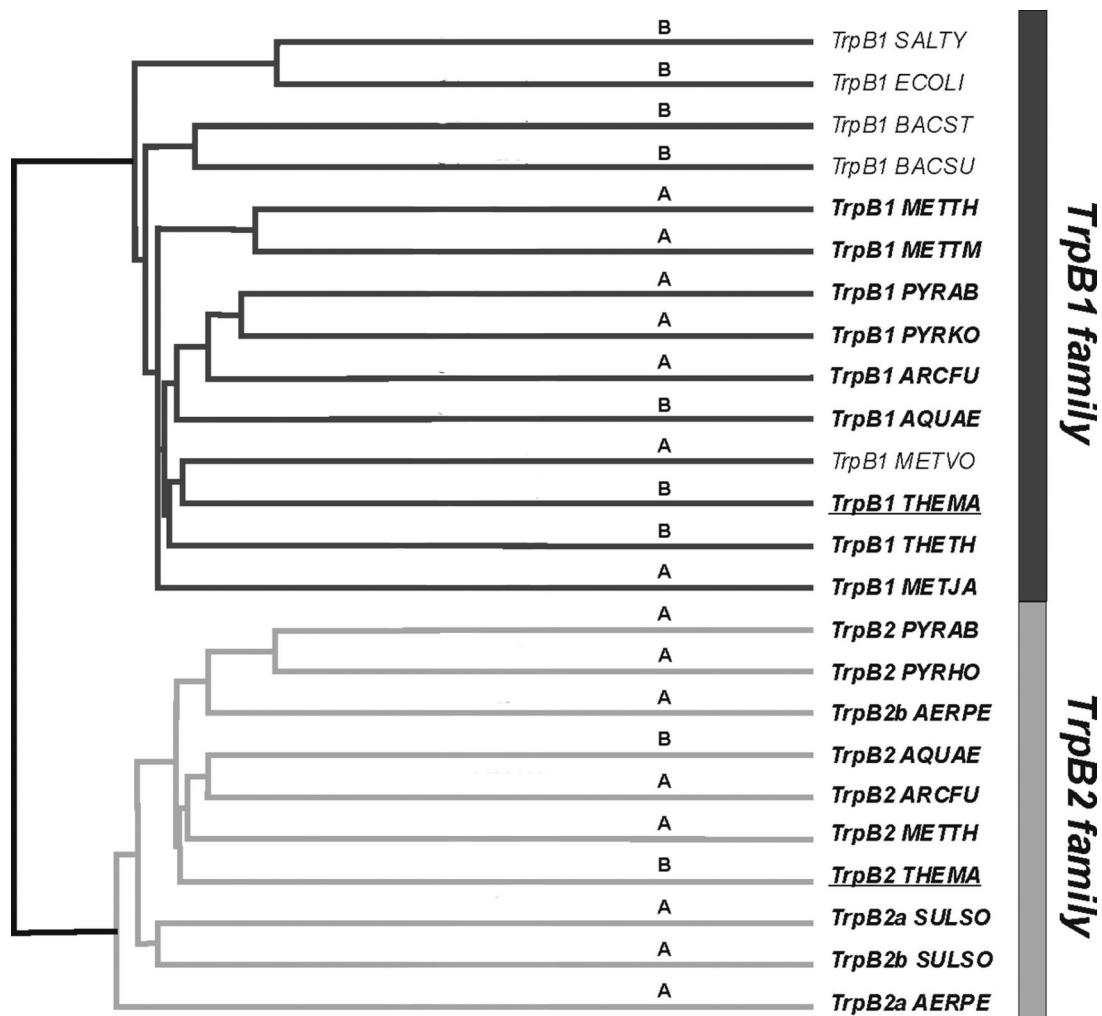
Figure 2.3 Phylogenetic tree of the two sequence families of TrpB proteins.
The TrpB1 family with representatives from mesophilic (regular type) and hyperthermophilic microorganism (in bold). The TrpB2 family has been found exclusively in hyperthermophiles. The two kingdoms where TrpB2 proteins have been found are indicated (A, archaeon; B, bacterium). The image was adapted from Hettwer & Sterner (2003). Abbreviations for organisms: *AERPE Aeropyrum pernix*; *ARCFU Archaeoglobus fulgidus*; *AQUAE Aquifex aeolicus*; *BACST Bacillus stearothermophilus*; *BACSU Bacillus subtilis*; *ECOLI Escherichia coli*; *METJA Methanococcus jannaschii*; *METTH Methanobacterium thermoautotrophicum*; *METTM Methanobacterium thermoautotrophicum* (strain Marburg / DSM 2133); *METVO Methanococcus voltae*; *PYRAB Pyrococcus abyssi*; *PYRKO Pyrococcus kodakaraensis*; *THEMA Thermotoga maritima*; *THETH Thermus thermophilus*; *SALTY Salmonella typhimurium*; *SULSO Sulfolobus solfataricus*

2.2.3   The Tryptophan Synthase of *Sulfolobus solfataricus*

Sulfolobus solfataricus is a hyperthermophilic crenarcheon with optimal growth conditions at high temperatures (75-80°C) and low pH values (2-4) (She *et al.*, 2001). The *trp* operon of the archeon comprises one *trp*A and one *trpB2* (TrpB2a transcript) gene. Additionally, a second *trp*B2 gene is located downstream of the operon (TrpB2b transcript). The two TrpB2 have high sequence identity (52 %). A similar occurrence was reported in four other crenarchea (*Pyrobaculum aerophilum, Aeropyrum pernix, Pirophilus torridus, Sulfolobus tokodaii*; Fitz-Gibbon *et al.*, 2002; Futterer *et al.*, 2004; Kawarabayasi *et al.*, 2001; Kawarabayasi *et al.*, 1999).

The TrpA/TrpB2 system from *Sulfolobus solfataricus* has been investigated by Leopoldseder (2005). Size exclusion chromatography and analytical ultracentrifugation experiments showed no evidence for the formation of a stable multienzyme complex between any of the β-subunits and the α-subunit. However, based on titration experiments a transient *Ss*TrpA:*Ss*TrpB2a complex formation during catalysis has been suggested (Leopoldseder *et al.*, 2006). The interaction of *Ss*TrpB2a and *Ss*TrpA was proven by the activation of the α-reaction (60-fold for addition of *Ss*TrpB2a and 270-fold for addition of *Ss*TrpB2a and L-Ser; no effect for *Ss*TrpB2b) in steady-state kinetics. Interestingly, experiments showed only upstream activation of the α-reaction by *Ss*TrpB2a but no activation of the β-reaction by *Ss*TrpA in the presence and absence of the substrate analogue IPP. Contrary to the reciprocal allosteric effect that results in mutual activation, as described for TrpB1, the effect in TrpB2a is only monodirectional, it activating *Ss*TrpA. The rate-limiting step in *S. solfataricus* is the TrpA reaction (2-fold lower turnover).

Steady-state kinetic experiments of the β-reaction revealed an order of magnitude higher catalytic efficiency (k_{cat}/K_M) for the operon coded *Ss*TrpB2a over the non-operon *Ss*TrpB2b. The higher efficiency was due to the higher turn-over number (k_{cat}) of *Ss*TrpB2a. Both the side reaction (Ser deaminase activity, Crawford *et al.*, 1964) and the main reaction (β-reaction) of TrpB1 proteins have been proposed to be the stand-alone functions of TrpB2s by Xie *et al.* (2002) and Leopoldseder *et al.* (2006), respectively. In fact, almost all organisms with a *trp*B2 gene lack a Ser deaminase gene.

However, the Ser deaminase activity (deamination of Ser yields pyruvate and ammonia) of TrpB2b shows lower catalytic efficiency than the β-reaction. The issue is controversial.

Mutagenesis studies based on the structure of *St*TrpB1 as well as sequence alignments of *Ss*TrpB2 family proteins have been performed by Leopoldseder (2005). Putative TrpA binding segments as well as the complete N-terminus were switched between the TrpB2 proteins of *S. solfataricus*. The significantly lower TrpA activation of two *Ss*TrpB2a mutants (TrpB2a_1 and TrpB2a_3 in Table 2.2) indicated the switched residues as candidates for interaction to *Ss*TrpA. The TrpB2b mutants had no effect on the TrpA reaction.

Table 2.2 Mutagenesis of *Ss*TrpB2a and *Ss*TrpB2b.

*Ss*TrpB2a_x	*Ss*TrpB2b_x	Modification	No[1]
Gln30-Ile37	Thr37-L40	Switched	1
Met1-Leu44[2]	Met1-Leu47	Switched	2
Asn274-Arg279	Asp277-Lys287	Switched	3
Lys190-Gln197	Gln193-Asn200	Switched	4

[1]in *Ss*TrpB2a_x/*Ss*TrpB2b_x
[2]insoluble

The non-operon coded TrpB2 family members have been suggested to be ancient β-subunit types, from which the operon coded TrpB1 proteins have evolved. In the model suggested by Leopoldseder and coworkers (2006), evolution of TrpB enzymes was initiated by gene duplication events followed by integration into the operon. The operon located enzyme evolved then the capability to form a permanent multienzyme complex. The TrpB2b class enzymes could have then taken a new function, *e.g.* indole salvage enzyme (Hettwer & Sterner, 2002) or Ser deaminase (Xie *et al.*, 2002). The final steps of TrpS evolution could have been loss of the gene outside the operon and fusion of TrpA and TrpB (Figure 2.4).
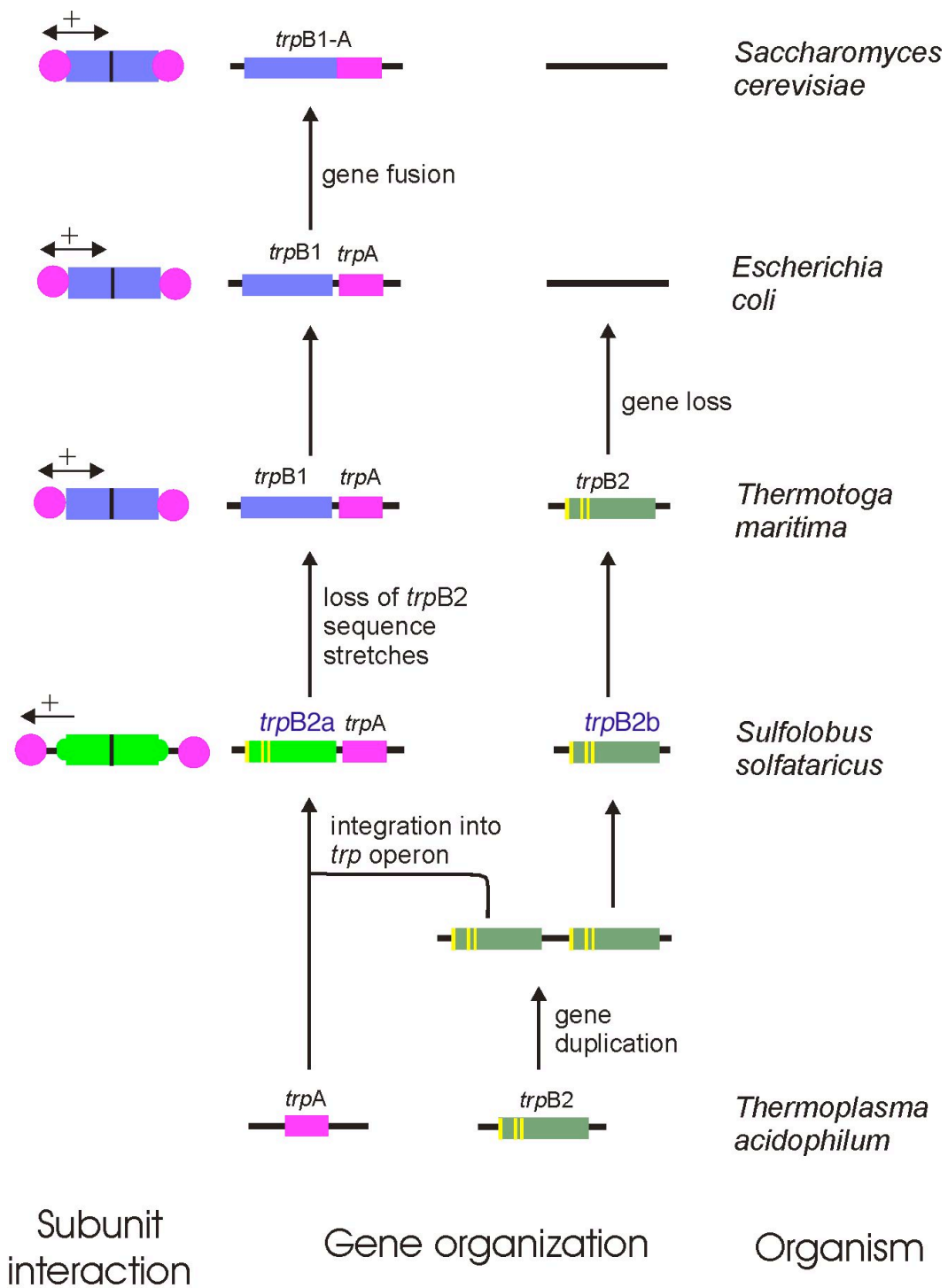
Figure 2.4 Hypothetical model for TrpB evolution.
The listed organisms provide examples of the given *trp*A and *trp*B gene organizations and the interactions between the TrpA and TrpB subunits. The horizontal arrows indicate the direction of subunit activation. The genes *trp*B2a from *S. solfataricus* and *trp*B2 from *T. maritima* might code for indole salvage proteins. Image adapted from Leopoldseder *et al.*, (2006).

2.2.4 Structure of TrpS

TrpS structures (composed of TrpA and TrpB1 subunits) have been determined for the thermophilic archeon *Pyrococcus furiosus* and the mesophilic bacterium *Salmonella typhimurium* (first structure of a TrpS; overview in Table 2.4). The structures of thermophiles and mesophiles show a similar fold. For *Pf*TrpS structures of the complex as well as its isolated components are known. The fold of the subunits is independent of multimer formation; only shifts of structural elements in respect to each other occur upon association (Lee *et al.*, 2005, details in section 2.2.6).

TrpA has a $(\alpha\beta)_8$-barrel (TIM-barrel) fold (Figure 2.5), which is found in approximately 10 % of all proteins with known structure (Miles, 2001; Sterner & Hocker, 2005). There are two additional helices (termed αH2' and αH8') to the $(\alpha\beta)_8$-barrel. Helix 8' is involved in coordination of the phosphate moiety of the substrate via its positive dipole moment (Weyand *et al.*, 2002).
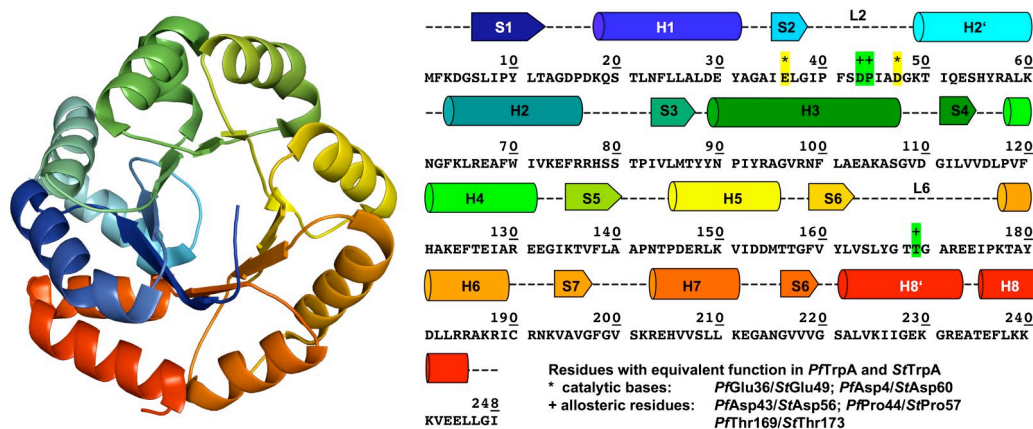
Figure 2.5 The α-subunit of *St*TrpS.
Gradient coloring from blue (N-terminus) to red (C-terminus). Two loops of TrpA have been identified to be involved in catalysis, αL2 and αL6. These are positioned downstream of the corresponding β-strands αS2 and αS6. The image was created from *St*TrpA (PDB code 1BKS, Hyde *et al.*, 1988) using PYMOL with secondary structure elements being mapped to the sequence according to Yamagata *et al.*, 2001.

The β-subunit shows two domains of approximately the same size (Table 2.3). The two domains can be superposed and show structural similarity (*Pf*TrpB1, *rmsd* = 2.7 Å over 73 C_α atoms, Hioki *et al.*, 2004). The N-terminal domain comprises a subdomain, the so-called COMM (from αβ-COMMunication) domain. During

catalysis the COMM domain undergoes a rigid body movement (10° rotation relative to the remainder of the β-subunit). Two Gly residues at the borders of this domain have been suggested to act as hinges during catalysis (Schneider *et al.*, 1998).

Table 2.3 Domain composition of *St*TrpB1 and *Pf* TrpB1.

	N-terminal	COMM	C-terminal	Reference
*St*TrpB1	1-52, 86-204	97-184[a]	53-85, 205-397	Hyde *et al.*, 1988
*Pf*TrpB1	1-46, 81-200	102-189[b]	47-80, 201-388	Hioki *et al.*, 2004

[a]subdomain of the N-terminal domain as defined by Schneider and coworkers (1998)
[b]defined by sequence comparisons

The interface between the β-subunits is formed by a broad, nearly flat surface with a dyad axis passing through it. Roughly, the subunits interact via cross-domain interactions to each other, where the N-terminal domain of one subunit interacts with the C-terminal domain of the complementary subunit (and vice versa). For a representation of the β-subunit and the biological relevant heterotetramer see Figure 2.6. Hydrophobic interactions are central to both interfaces (hetero-obligomer and homo-obligomer). The thermophilic TrpS gains stability by minimizing cavities and improving interactions across the interface (Hioki *et al.*, 2004). The αβ-dimer was suggested to be the functional unit in catalysis, while hetero-tetramer formation serves to stabilize this unit (Hyde *et al.*, 1988).

2.2.5 PLP-binding in TrpB1

PLP (vitamin B_6) dependent enzymes form a covalent bond between the ε-amino group of an active site Lys and the cofactor vitamin B_6 (Christen & Mehta, 2001). In TrpS, the cofactor binds to the β-subunit (between the N-terminal and C-terminal domains). The residues involved in PLP binding (Figure 2.7) are strictly conserved in both families of TrpB1 and TrpB2 proteins.

A
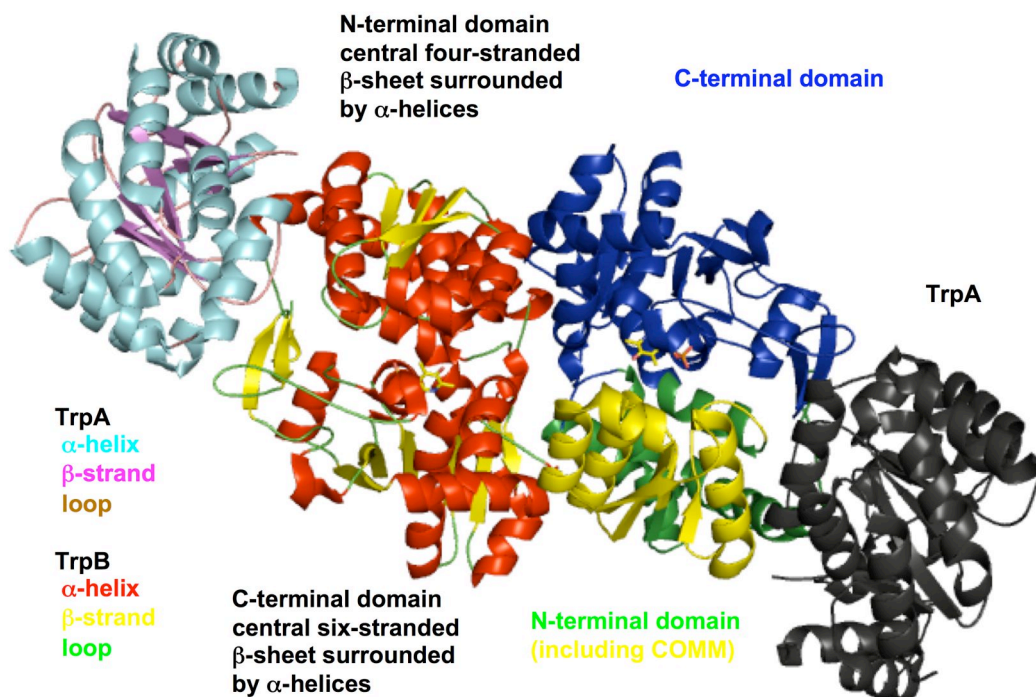
B
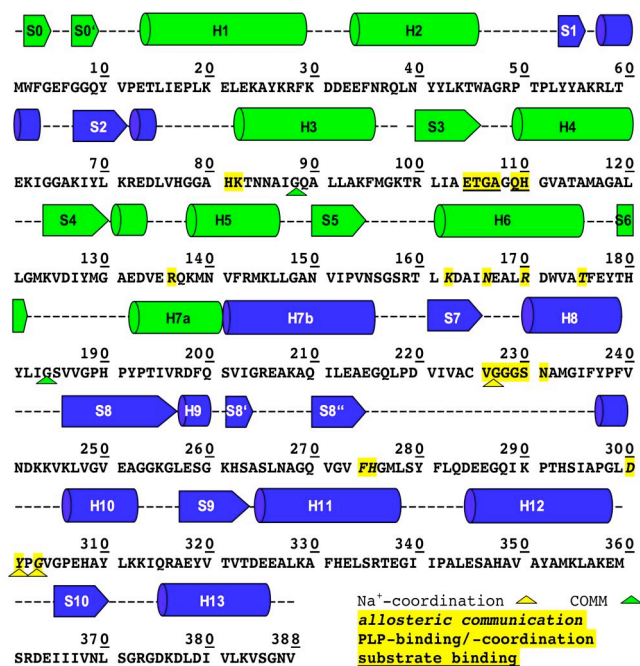
Figure 2.6 TrpS from *Pyrococcus furiosus* (PDB code 1WDW, Lee *et al.,* 2005).
(**A**) Biological heterotetramer (αββα). Left: TrpA and TrpB1 colored by secondary structures and the central fold of the domains Right: TrpA and TrpB1 colored by domains. The cofactor in the center between the two domains of TrpB1 is shown in stick representation. (**B**) Sequence of *Pf*TrpB1 with important residues highlighted by yellow background.
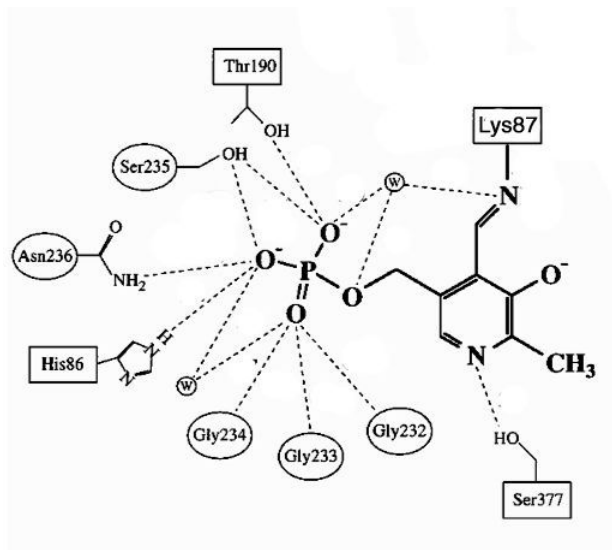
Figure 2.7 Coordination of the cofactor PLP (H-bonding; modified from Rhee *et al.,* 1997).
The P-loop is formed by residues Gly232-Asn236.

The active site Lys (upstream to H3; *St*Lys87, *Pf*Lys82) forms a covalent aldimine bond to PLP. This Lys residue is not only important to form the aldimine, but also acts as a catalytic base. The main contribution to coordination of the phosphate group is mediated by a motif on a turn between S7 and a H8 in the C-terminal domain. The sequence of this motif is Gly-Gly-Gly-Ser-Asn.

Gly-rich loops of analogous positioning between a strand and a helix are well known in phosphate binding proteins (mostly as part of nucleotide binding motifs). In these proteins, they are called "Gly-rich P-loop motif" and were first described for adenylate kinase (for reviews see Saraste *et al.,* 1990; Prasad 2001).

A His residue (directly preceding the catalytic Lys) interacts with the phosphate moiety and coordinates the pyridine ring of the cofactor (edge-to-face arrangement). Mutations in this residue (*St*βH86L) alter the catalytic and spectroscopic properties of the enzyme (Miles et al., 1989; Ro & Miles, 1999). Hence, this His was suggested to alter the pK_a value of the catalytic residue towards the physiological pH. Besides PLP coordination this His interacts with the adjacent catalytic Lys during the β-reaction. A Ser residue between S10 and H13 coordinates the pyridine nitrogen of PLP.

2.2.6   Formation of the TrpA/TrpB1 hetero-obligomer

Lee and coworkers (2005) investigated structural changes in the individual subunits of *P. furiosus* TrpS upon hetero-obligomer (αββα) formation. The α-subunit showed only small conformational changes (0.6 Å shifts for 36 of 247 residues, equal to 14.6%) in residues located in segments forming the hetero-obligomer interface, i.e. loops L2 and L6, which are partially ordered in *Pf*TrpA as well as residues in H4

(with subsequent hydrophobic rearrangement of *Pf*TrpA in that region). In *Pf*TrpB1, obligomer formation induced larger structural changes, which involved 140 of 388 residues (36.1 %) with C_α-displacement over 0.6 Å. Structural changes involved parts of the COMM subdomain (residues 100 to 170), which rearranged as a rigid body. In contrast, residues 255 to 303 (H9, S8', S8'' and H10) in the C-terminal domain move in an induced-fit conformational change. The authors concluded, that the COMM rearrangement transmits information between the active sites of the two enzymes upon obligomer formation (analogous to allosteric communication). These changes led to the formation of a long hydrophobic network with subsequent widening of the indole tunnel (Figure 2.8). The homo-obligomer interface, which is important for the functionality of *Pf*TrpB1, remained unchanged upon hetero-obligomer formation.
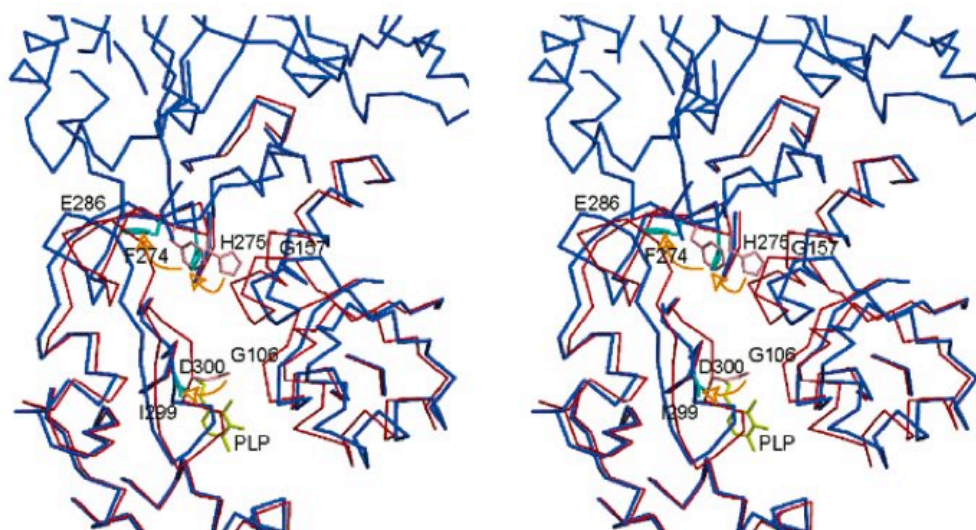
Figure 2.8 Stereo view of the tunnel-widening in *Pf*TrpB1.
The indole tunnel in *Pf*TrpB1 with blue and red lines indicating the *Pf*$\alpha_2\beta_2$ and *Pf*β_2 dimer, respectively. Distances between the residues Glu286 and Gly157 and between Ile299 and Gly106 expanded by 5.5 and 2.5 Å, respectively. Yellow arrows indicate conformational changes of Asp300 (opening the substrate entrance) and Phe274/His275 (gating residues). It should be mentioned that the conformational changes of the gating residues could have been induced by the presence of K^+, which was present in the crystallization buffer.

Structural changes during catalysis have been investigated for TrpS from *S. typhimurium* (Rhee *et al.*, 1997). The interaction pattern between the interface residues remained relatively unchanged, despite large positional changes.

2.2.7   Catalytic mechanism of TrpS

The elucidation of structures of complexes of TrpS with reaction intermediates has given detailed insights into conformational changes taking place during catalysis. This involved mainly structures from *S. typhimurium* (Table 2.4). The mechanism of catalysis itself has been examined by exploiting distinct spectral properties of transiently accumulating PLP intermediates (Miles, 1979).

Table 2.4 Structural data available for binding of substrates, intermediates and products in the β-subunit (after Miles, 2001, for *S. typhimurium* and *P. furiosus*[1]).

PDB code	$\alpha_2\beta_2$	Ligand[2] (α/β)	Intermediate[3]	Reference
1BEU	αD60N	IPP/Ser	IPP/E-Ser	Rhee *et al.*, 1998
1UBS	βK87T	*apo*/Ser	E-Ser	Rhee *et al.*, 1997
2TRS	βK87T	IPP/Ser	IPP/E-Ser	Rhee *et al.*, 1997
2TSY	βK87T	GP/Ser	GP/E-Ser	Rhee *et al.*, 1997
2TYS	βK87T	*apo*/Trp	E-Trp	Rhee *et al.*, 1997
1A50	WT	F-IPP/apo	IPP/E	Schneider *et al.*, 1998
1TTQ	WT	*apo/apo* (K$^+$)	E	Rhee *et al.*, 1996
1TTP	WT	*apo/apo* (Cs$^+$)	E	Rhee *et al.*, 1996
1BKS	WT	IPP/apo	E	Hyde *et al.*, 1988
1A5S	WT	F-IPP/Ser	F-IPP/E-AA	Schneider *et al.*, 1998
1A5A	αD60N	*apo/apo* (K$^+$)	E	Rhee *et al.*, 1998
1A5B	αD60N	IGP/*apo* (K$^+$)	IGP/E	Rhee *et al.*, 1998
2WSY	WT	*apo/apo*	E	Schneider *et al.*, 1998
1QOP	WT	IGP/*apo* (no ion)	IGP/E	Weyand & Schlichting,
1QOQ	WT	IPP/ *apo*	IPP/E	1999
1C29[4]	WT	TIA/ *apo*	TIA/E	Sachpatzidis *et al.*, 1999
1GEQ	WT α	*apo*/-	-	Yamagata *et al.*, 2001
1WDW	WT	*apo/apo*	E	Lee *et al.*, 2005
1V8Z	WT β$_2$	-/ *apo*	E	Hioki *et al.*, 2004

[1]*Pyrococcus furiosus* structures: 1WDW and 1V8Z; [2]monovalent cations Na$^+$ (if nothing else mentioned) and Trp/Ser are bound to the β-subunit. IPP (indole 3-propanol phosphate, substrate analogue), FIPP (5-Fluoro IPP), IGP (indole-3-glycerol phosphate), GP (DL-α-3-glycerol 3-phosphate), TIA (α transition state intermediate); [3]intermediates are illustrated in Figure 2.10; [4]representative; full: 1C29, 1C8V, 1C9D, 1CW2 and 1CX9

TrpA catalyzes a retroaldol reaction that results in the enzymatic cleavage of indole-3-glycerol phosphate (IGP) to indole and glyceraldehyde-3-phosphate (GA3P). This is termed the α-reaction. The catalytic mechanism of TrpA is believed to follow a general "push-pull" acid base mechanism:

(i) Indolenine tautomerization: the base B2 polarizes the N-H bond of indole and abstracts the proton. The acid B1H protonates the C3 atom and facilitates the electronic rearrangement that leads to the formation of the indolenine tautomer (first step in Figure 2.9).

(ii) C-C bond cleavage: the base B3 abstracts the proton from the C1′ hydroxyl of the glyceryl side chain of IGP. This event leads to the formation of indole and D-glyceraldehyde 3-phosphate. (GA3P, second step in Figure 2.9).

The crystal structure of *St*TrpS (Hyde *et al.*, 1988) and mutagenesis studies (Nagata *et al.*, 1989; Shirvanee *et al.*, 1990; Yutani *et al.*, 1987) led to propose that the proton-transfer steps during cleavage and synthesis of IGP are catalyzed by the bases *St*Glu49 and *St*Asp60 (Kirschner *et al.*, 1991; Nagata *et al.*, 1989). This was later supported by crystallographic studies (Rhee *et al.*, 1998; Rhee *et al.*, 1997).
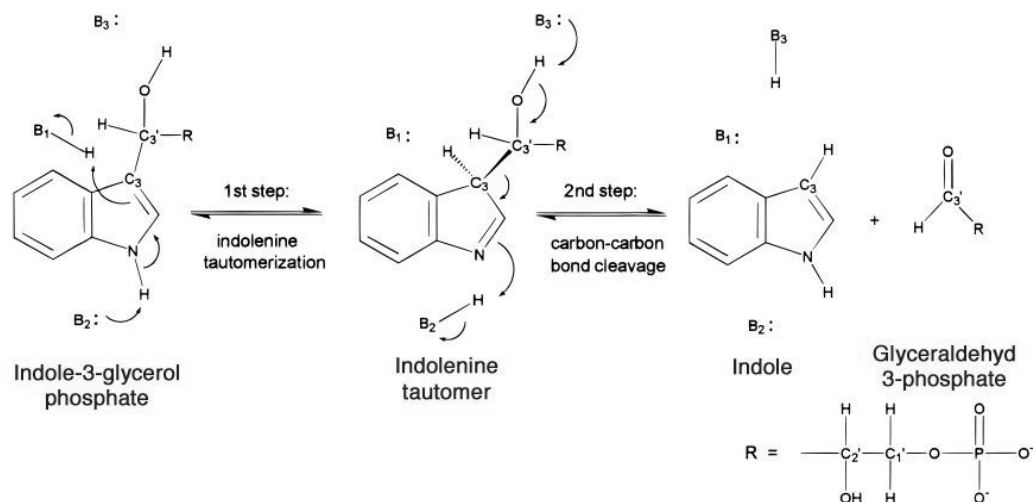
Figure 2.9 The α-reaction catalyzed by TrpA.
The catalytic bases B_1H and B_3 were identified to be *St*Glu49. B_2 was identified to be *St*Asp60 (Rhee *et al.*, 1998 and references therein). Figure adapted from Sachpatzidis *et al.*, 1999 (description: main text).

The TrpB reaction (or β-reaction) shows two stages (Figure 2.10). Stage I comprises the steps of transaldimination and subsequent condensation to form an amino acrylate intermediate:

(i)　Transaldimination: the internal aldimine bond (E) covalently bound βLys87 to PLP is replaced by the amino group of the substrate serine to form an external aldimine with PLP (E-Ser). The transaldimination proceeds over a *gem*-diamine intermediate.

(ii)　Condensation: βLys87 abstracts the acidic proton of the C_α of E-Ser. The hydroxy group of E-Ser leaves and a water molecule is formed by deprotonation of βLys87 to recover the catalytic base. This yields the formation of an amino acrylate intermediate (E-AA), a Michael-system with an electrophilic double bond.

The end of stage I with formation of E-AA can be regarded as a "hold modus" since the progress of the reaction is now dependent on the supply of indole from the α-reaction. Kinetic studies indicate that formation of the amino acrylate intermediate in the β-subunit significantly activates the α-reaction (25- to 30-fold). Conformational changes in this step are supposed to play roles in allosteric interactions between the α- and β-subunits (Anderson *et al.*, 1991; Banik *et al.*, 1995; Brzovic *et al.*, 1992a; Kawasaki *et al.*, 1987).

Stage II of the β-reaction starts when indole is transferred through the hydrophobic tunnel from the α-subunit to the β-subunit active site:

(i)　Michael-type addition: nucleophilic addition of indole to the electrophilic double bond on the *si* face of E-AA yielding a quinoide intermediate (E-Q) (Miles *et al.*, 1982). The nucleophilic addition of indole is believed to be facilitated by *St*Glu109, that can stabilize a positive charge at the indole NH (Anderson *et al.*, 1991; Brzovic *et al.*, 1992b). The formation of E-Q is an example of the electron storing properties (from the substrate bond) common to all pathways of PLP dependent enzymes (Christen & Mehta, 2001).

(ii) E-Q is protonated by the catalytic Lys to result in a second external aldimine (E-Trp). The reaction cycle is completed by the formation of the internal aldimine to the catalytic Lys (E) in a β-replacement reaction and release of Trp.

The formation of E-Trp in stage II is the trigger for deactivation of the α-site. The coordination of the external aldimines is mainly mediated via main chain atoms of the β-subunit and PLP coordination has been observed to be closely similar, with some minor differences in E-Trp (Rhee *et al.*, 1997; Leja *et al.*, 1995).

A

B   **C**
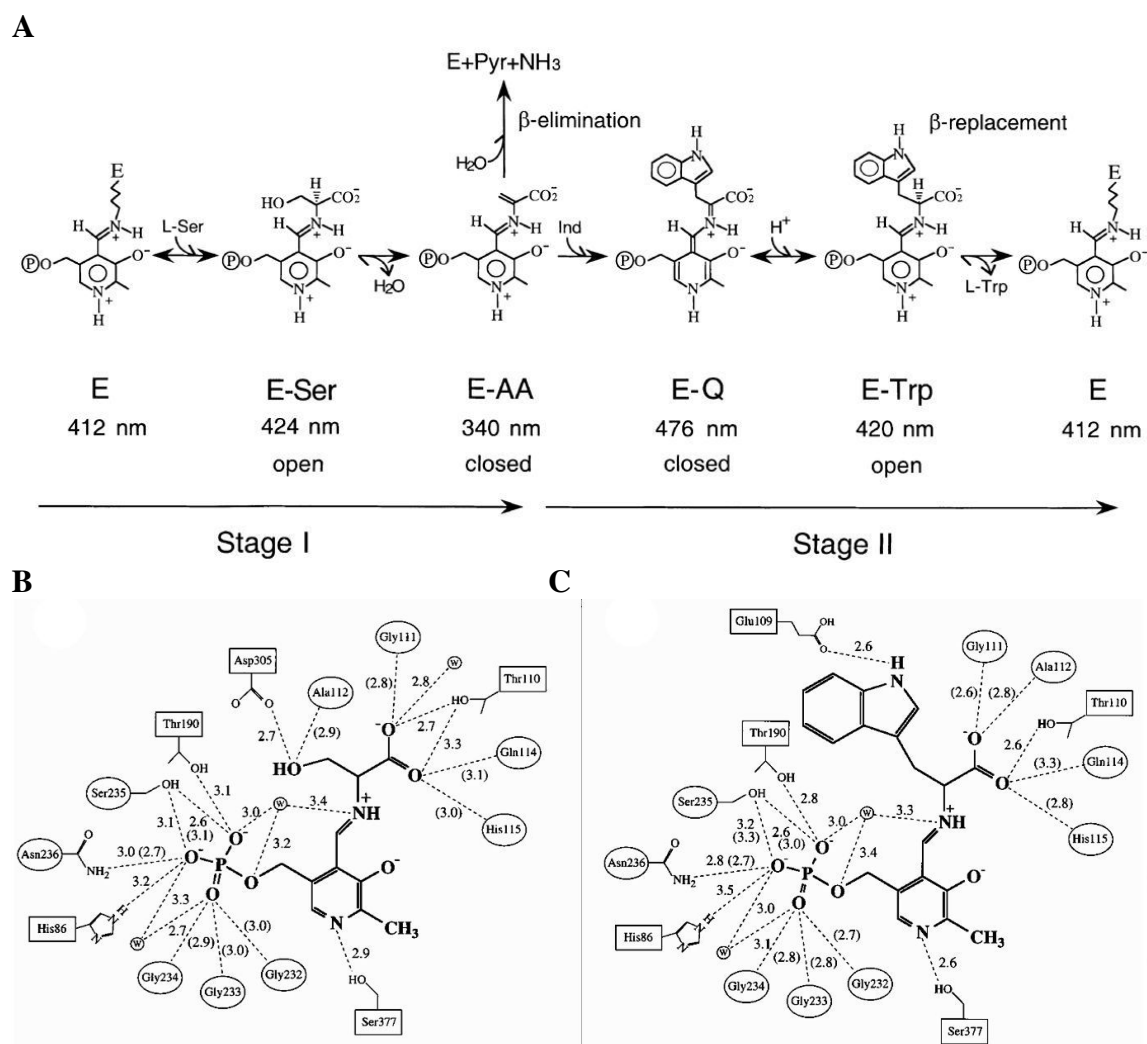
Figure 2.10 Catalytic mechanism of TrpB and coordination of external aldimines.
(A) Catalytic mechanism of TrpB1 from *S. typhimurium* divided in two stages (from Miles, 2001) At the end of stage I β-elimination to pyruvate is shown (Ser deaminase activity). Intermediates in catalysis and their λ_{Abs} is shown (for used abbreviations see main text) **(B)** Coordination of the external aldimine (E-Ser) *St*Asp305 is in "swing in" conformation, which is important for substrate specifity. During catalysis, it is in "swing out" conformation and not involved in binding of Ser. **(C)** Coordination of the aldimine formed with the product (E-Trp). **(B/C)** The interaction sphere taken was 3.5 Å for the *St*TrpB1K87T inactive mutant. The α-subunit was in the apo form in figures shown here. The figures were taken from Rhee *et al.*, 1997.

2.2.8   Effectors of TrpS catalysis

Catalysis of enzymes, especially enzymes in anabolic pathways, has to be regulated. This can happen at the genetic level (*e.g.* the coordinated expression of genes *trp*-operon; Yanofsky, 2001) or by affecting the enzymes itself (*e.g.* feedback-inhibition of Anthranilate Synthase; Zalkin, 1993). Repression and inhibition are extreme cases, where the systems are switched either "on" or "off". Besides these cases, there are many possibilities to negatively or positively influence an enzymatic reaction. Reactivity of TrpS containing a B1 subunit is affected by a complex system of allosteric regulation, including:

(i)   Monovalent cations

(ii) pH

(iii) α-subunit ligands (allosteric communication)

2.2.8a Monovalent cation binding

Monovalent cations play a role in catalysis of PLP-dependent enzymes (for a review see Woehl & Dunn, 1995). Structural studies revealed a binding-site for a cation (Na^+) 8 Å away from the phosphate moiety of PLP in the C-terminal domain of *St*TrpB1 (Rhee *et al.,* 1996). The binding-site is involved in allosteric communication between the active sites of TrpA and TrpB1 (Woehl & Dunn, 1999).

Effects mediated by the nature of alkali cations have been investigated by Rhee and coworkers (1996) using X-ray crystallography. Structural changes in *St*TrpS upon replacement of the natural Na^+ for K^+/Cs^+ are:

(i)   Higher flexibility of the α-subunit active site loop.

(ii) Displacement of cation binding loop and altered coordination sphere of the cation (Table 2.5).

(iii) Displacement of the side chain of Tyr279 and the putative tunnel gating residue Phe280 out of the tunnel of *St*TrpB1

(iv) Loss of interaction between Asp305 and Arg167 (both *St*TrpB1) with a putative new interaction partner Asp56 (*St*TrpA).

Table 2.5 Coordination of the monovalent cations in *St*TrpB1 after Rhee *et al.,* 1996[1].

	Na$^+$	K$^+$	Cs$^+$
Coordination no	5	4	6
Ligands[2]			Val231
	Gly232[3]	Gly232[3]	Gly232[3]
			Gly268
			Leu304
	Phe306	Phe306	Phe306
	Ser308	Ser308	Ser308
	2 x H$_2$O	1 x H$_2$O	
Coordination sphere	Tetragonal pyramidal	Like Na$^+$ (one H$_2$O less)	Octahedral

[1]sphere of maximum 4.0 Å around cation (for Cs$^+$)
[2]all amino acids are residues of the β-subunit, common residues are underlined
[3]Gly232 is also involved in PLP coordination by the P-loop.

Exclusively carbonyl oxygens of TrpB1 residues coordinate the cations in the binding-site. Variation in the monovalent cation affects the maximal rate of TrpS stage I in catalysis. The increased rate of the enzyme results from the altered ligand sphere of the given cations. Accordingly, the overall conformation of the enzyme is altered on a long-range scale favoring a more active state with increasing size of the cation. The presence of the physiological Na$^+$ ion increases the affinity of E-AA for indole, which favours the closed conformation. Since the closed conformation also stimulates the α-reaction, the cation binding-site is involved in allosteric communication (Woehl & Dunn, 1999).

2.2.8b pH-dependence

Given that catalytic acids and bases are involved in the reaction catalyzed by TrpS, a dependence on pH can be observed. Several catalytic properties of TrpS are affected by pH:

(i) Accumulation of the quinoid species (E-Q) in stage II of the β-reaction increases with pH (Goldberg & Baldwin, 1967; Mozzarelli *et al.*, 2000).

(ii) The rate-limiting step, carried out by *E. coli* TrpS, varies as a function of pH. At pH 6.5 it is the C_α proton removal from E-Ser to E-AA and at pH 7.6 it is the release of the product L-Trp (Lane & Kirschner, 1983).

(iii) The pH-dependence of β-replacement and β-elimination was found to be significantly altered by a *St*TrpB1H86L (PLP coordinating residue) mutation (Ro & Miles, 1999).

(iv) Accumulation of the external aldimine is favored at high pH, whereas the α-amino acrylate is stabilized at low pH (Peracchi *et al.*, 1996; Mozzarelli *et al.*, 1989). Replacement of the cation weakens dependence on pH with $Cs^+ > K^+ >$ no ions $> Na^+$ favouring formation of E-AA (Schiaretti *et al.*, 2004).

(v) The catalytic Lys residue (forming E) is involved in protonation initiating the β-elimination from E-Ser to E-AA (Hur *et al.*, 2002).

2.2.8c Allosteric communication

The regulation of enzyme activity by interaction with metabolites is key to cell function. A common event that accompanies ligand binding is the transition between open and closed states (Gutteridge &Thornton, 2004). For TrpS, the transition between open (low activity) and closed (high activity) states has a dual role: it increases the catalytic activity and prevents indole from escaping the enzyme.

Allosteric interactions occur when binding of one ligand at a specific site is influenced by binding of another ligand at a different (allosteric) site. A pathway for allosteric communication between the active sites of α- and β-subunits was proposed by Schneider and coworkers (1998). The model is based on kinetic, mutagenic and structural data. The pathway goes from α-subunit active site over β-subunit helix 6 (βH6) and the COMM domain to the β-subunit active site. Additionally, Rhee and coworkers (1997) investigated allosteric communication:

(i) *apo*: αL2 and αL6 are disordered and the α-subunit active site is accessible to solvent. Both subunits are in open conformation. The side chain of the gating residue (βPhe280) blocks the tunnel between the active sites of TrpA and TrpB.

A H-bond between the side chains of βAsn171 and βTyr279 supports the conformation of βPhe280 by sterical hindrance.

(ii) α-active site to βH6: IGP binds to the α-subunit. The mobile αL6 becomes ordered, closes the α-subunit active site and shields it from solvent exposure. Subsequently αThr183 (in αL6) orients the side chain of the catalytic base αAsp60 (in αL2), which coordinates the indole moiety of the substrate (IGP). The second catalytic base αGlu49 (in αL2) coordinates the glycerol moiety of IGP. A H-bond between main chain atoms of αGly184 (in αL6) and βSer178 (in βH6) is formed. Mutagenesis studies revealed βSer178Pro to lose intersubunit signaling (Raboni *et al.*, 2005).

(iii) βH6 to COMM domain: upstream along βH6, the H-bonding pattern is changed. Stabilization of αL2 is given by H-bonds from αAsp56 to βAsn171 and βLys167. The conformation of the side chains of βTyr279 and βPhe280 is changed since the H-bond from βTyr279 to βAsn171 remains. The tunnel is partially opened. Interactions from αAsp60 and αPro57 to βArg175 are formed and support the conformation of the α-site catalytic base, αAsp60. Since βH6 is included in the COMM domain, the altered interaction pattern induces a rigid body movement of this domain, which opens the β-subunit active site. An alternative conformation of βLys167 ("swing in" conformation; H-bonding to βAsp305; Rhee *et al.*, 1996) seems to be less important in allosteric regulation than in substrate specifity (Weber-Ban *et al.*, 2001, Ferrari *et al.*, 2003).

(iv) β-subunit closure: upon binding of Ser the COMM domain rotates 10° relative to the rest of the β-subunit moving the two regions closer together. The side chain of βAsp305 changes from the "swing in" conformation (E-Ser side chain coordination) to the "swing out" conformation. The carboxylate moiety of E-Ser is coordinated by residues of a turn in the COMM domain and βAsp305.

(v) Catalysis: substrate binding to the active sites of both subunits favors E-AA formation in the equilibrium of stage I during catalysis. Upon IGP cleavage the tunnel opens (displacement of αL2). Indole is transferred and stage II of catalysis yields E-Trp. The β-subunit is in "half open" conformational (similar to E-Ser without IGP bound to the α-subunit). With the transaldimination reaction the cycle is completed and products are released.

Allosteric effects in TrpS seem to be restricted to the communication between the active sites of the α- and β-subunits (overview in Figure 2.11). Communication between the active sites of the two adjacent β-subunits has not been reported so far. The formation of the homo-obligomer is considered to happen for the benefit of stability of the β-subunit (Hyde *et al.*, 1988).
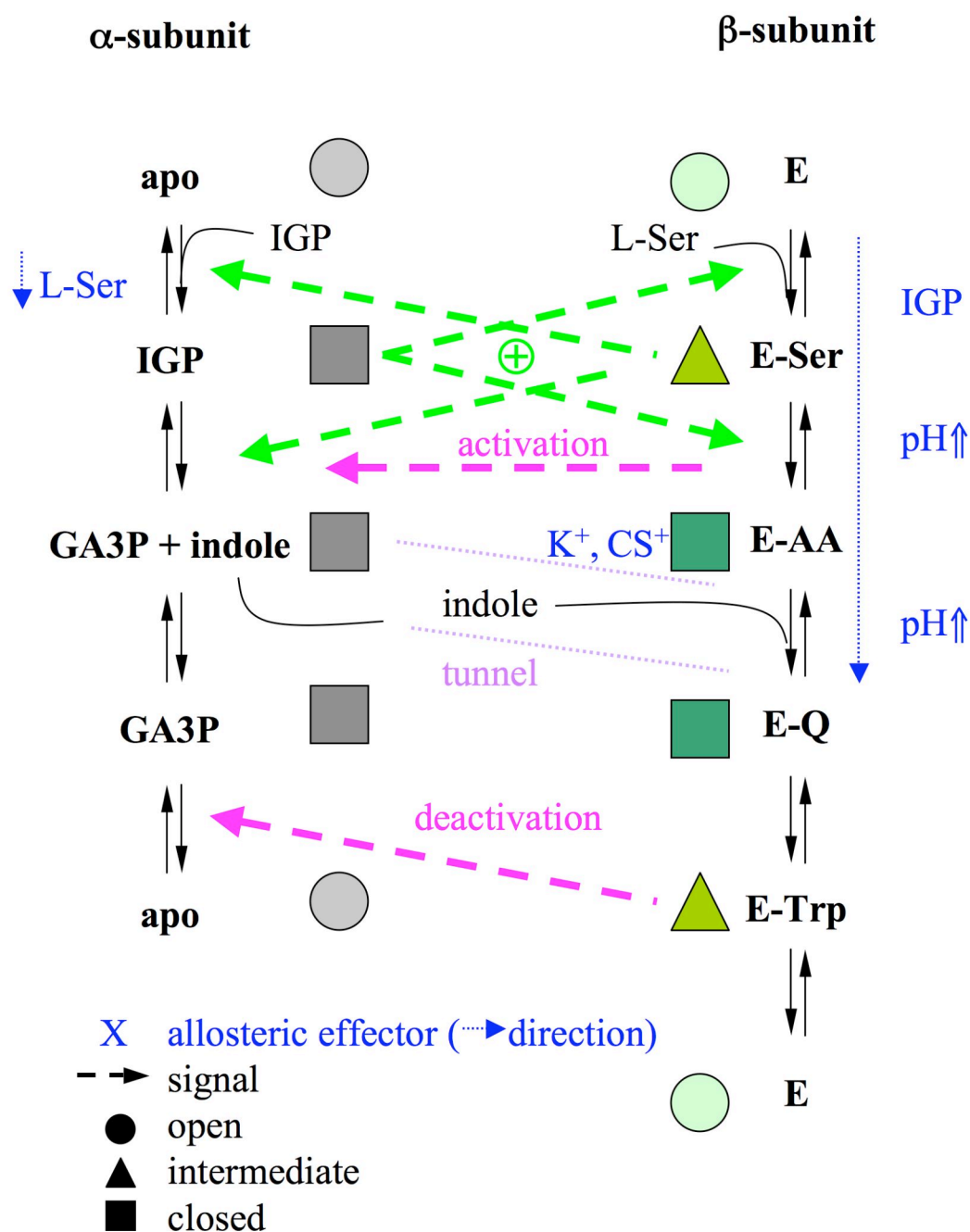
Figure 2.11 Overview of allosteric signalling and effects of TrpS during catalysis.
The states of the α-subunit are shown in different grays and of the β-subunit in different greens. Horizontal/diagonal arrows indicate a conformational change affecting the adjacent subunit (green/magenta) or the direction an allosteric effector alters the equilibrium distribution (blue). Allosteric effectors favoring a conformational state are shown next to the state, effectors altering an equilibrium distribution (blue) next to the reaction arrows (black) with direction of the positive effect.

2.3   Aim of the work

The evolution of TrpS is still under investigation. Initial models have been produced on the base of kinetic data and sequence comparisons between TrpB1 and TrpB2 family members. However, the structural base for the formation of the hetero-complex between *Ss*TrpB2a and *Ss*TrpA remains elusive. To date, *Ss*TrpB2a protein production has been a problematic task (extremely low yields). In contrast to that, *Ss*TrpB2b production yielded amounts sufficient for crystallographic investigation. The proteins share high sequence identity and therefore the structural homology should be also high. *Ss*TrpB2b lacks interaction to *Ss*TrpA. However, structural investigation of *Ss*TrpB2b can give detailed insight about the organization of the protein. This model system gives insights into the evolution of an enzyme, where evolved orthologues show different association properties. In particular, the structure is interesting since *S. solfataricus* houses two paralogues of TrpB2 that can be regarded as the initialization point of TrpB evolution towards higher complexation. Model systems with rudimentarily evolved properties are rare since we normally see only the result of evolutionary shaping.

A first goal was to elucidate the structure of *Ss*TrpB2b, which subsequently had to be set in context with TrpB1 family members as well as findings for *Ss*TrpB2a/*Ss*TrpB2b and *Ss*TrpA. Subsequently, the focus of the work was the comparative analysis of the structures of TrpB1 family members and *Ss*TrpB2b to reveal structural similarities and differences. A particular focus lay on catalysis, allosteric communication and interface formation in *Ss*TrpB2b compared to TrpB1 family members. Since sequence conservation point towards highly similar structures of *Ss*TrpB2b and *Ss*TrpB2a the lack of mutual allosteric communication and complexation properties of the TrpB2 family members in *S. solfataricus* could be investigated. The catalytic properties between TrpB1 and TrpB2 seem to be conserved, which allowed homology model based investigation of interface formation in TrpB2 members of *S. solfataricus*.

2.4   Methods

*Ss*TrpB2b from *Sulfolobus solfataricus* was expressed and purified in the laboratory of Prof. Reinhard Sterner at the University of Regensburg (Germany). The protein was received in frozen state (- 80°C) in 20 mM potassium phosphate buffer pH 7.5. The medium of the protein was changed to 25 mM NaCl and 10 mM HEPES pH 7.5 by dialysis overnight (approx 16 h) at 4°C. For crystallization trials concentrations of 10.0 ± 2.5 mg/mL (based on A_{280} values) were used. Defrosted aliquots of *Ss*TrpB2b were used in crystallization trials within 24 hours.

2.4.1   Crystallization

Crystallization trials were performed using the vapor diffusion method at room temperature, where 1 μL of *Ss*TrpB2b was mixed with 1 μL of reservoir solution to form a 2 μL drop. For initial screening, each droplet was equilibrated against 70 μL of reservoir solution in 96-well CrystalQuick protein crystallization plates (Greiner), using a sitting-drop setting.

Commercial kits (sparse matrix kits CS-I & II, Hampton research) yielded several initial crystallization conditions (Table 2.6). Crystals with a three-dimensional habit grew overnight. The conditions were reproduced using the hanging-drop vapor diffusion method with the same drop ratio and increased reservoir volume (500 μL) in VDX plates (Hampton Research).

Table 2.6 Conditions yielding crystals of *Ss*TrpB2b in an initial screening.

Preciptant[1]	Buffer	Additive
30% PEG-4000	100 mM Tris HCl pH 8.5	200 mM NaOAc x 3 H_2O
30% PEG-4000	100 mM Tris HCl pH 8.5	200 mM $MgCl_2$ x 6 H_2O
30% PEG-8000	100 mM Na-cacodylate pH 6.5	200 mM NaOAc x 3 H_2O

[1]PEG-X = poly ethylene glycol-(molecular mass)

Visual inspection of *Ss*TrpB2b crystals revealed that only few of them were single, while most showed multiple growth. Refinement of initial conditions required varying PEG content, pH and additive nature (variation of ionic species). This yielded crystals with improved size as well as a higher portion of single crystals (Figure 2.12).

Best crystals (grown from 30% PEG-4000, 100 mM Tris pH 8.5 and 100 mM NaCl) were frozen in mother liquor supplemented with additional 10% PEG-4000 as cryoprotectant and used in diffraction experiments.
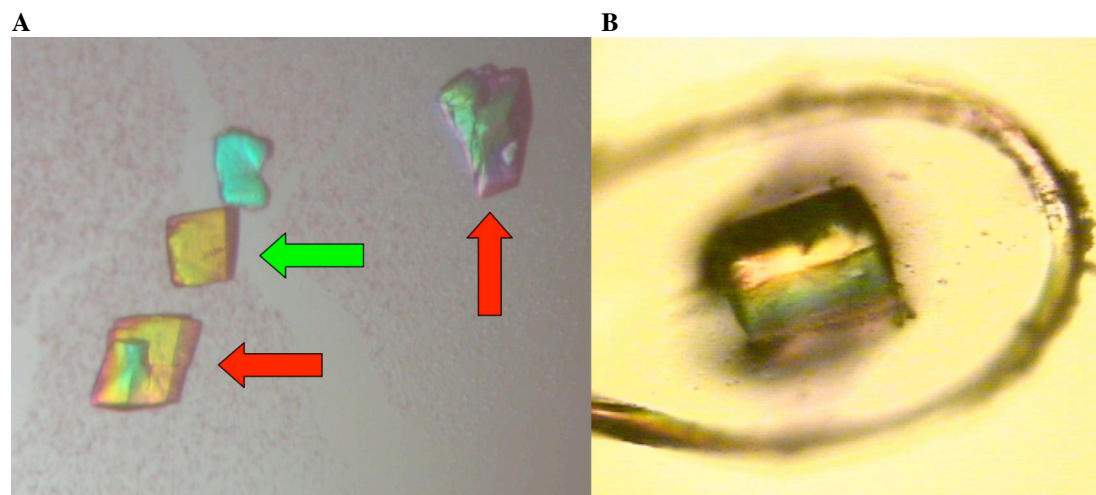
Figure 2.12 Crystals of *Ss*TrpB2b.
(**A**) Crystals in a hanging-drop grown from 25% PEG-4000, 100 mM Tris-HCl pH 8.0, 200 mM NaCl. A single crystal (indicated by the green arrow) with dimensions of approx. 140 μm x 140 μm x 140 μm and multiple crystals (indicated by red arrows) as well as precipitated protein. (**B**) Crystal (grown from 30% PEG-4000, 100 mM Tris pH 8.5 and 100 mM NaCl) used for native data collection mounted in a nylon cryo-loop (Hampton Research).

2.4.2 Collection of a native data set

X-ray data collection was carried out at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. A native data set of non-overlapping oscillations was collected at 100 K at the beamline ID-23 equipped with a marCCD 225 detector. The diffraction images (Figure 2.13) with a diffraction limit of 1.94 Å were processed with the XDS package (Kabsch, 1988). For data collection parameters and X-ray statistics see Table 2.7. The R_{free} (Brünger, 1993) was used as a cross-validation indicator in model building and refinement, for which reflections were partitioned into a working and a free set using FREERFLAG (CCP4, 1994).
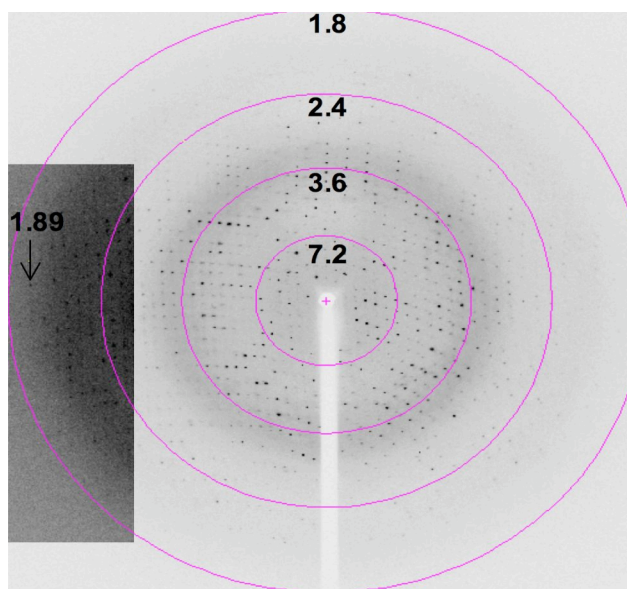
Figure 2.13 Diffraction pattern of *Ss*TrpB2b crystals.
High resolution reflections are shown in the inset. Resolution rings (Å) are displayed.

Table 2.7 Data collection parameters and X-ray statistics for *Ss*TrpB2b.

Parameter	Value
X-ray source	ESRF (Grenoble) ID-23
Detector	marCCD 225
Total rotation	180°
Oscillation range	0.5°
Exposure time	0.5 s
Detector to crystal distance	180 mm
Wavelength	0.9840 Å
Spacegroup	P 2_1
Unit cell [Å]	55.790, 61.590, 110.130, β = 80.73°
Resolution	19.97 - 1.94 (1.96 - 1.94)[1]
Unique reflections	54720 (1622)[1]
R_{sym} (I)	5.8 (38.3)[1]
I/σ (I)	16.7 (3.9)[1]
Multiplicity	3.76 (3.76)[1]
Completeness	99.8 (100.0)[1]

[1]outer resolution shell

2.4.3   Detection of noncrystallographic symmetry (NCS)

The Matthews coefficient V_m (Matthews, 1968) describes the packing density of a protein crystal given the volume of the unit cell V_{cell} [Å^3], the number of asymmetric units (a.u.) per unit cell z, the number of monomers per a.u. n and the molecular weight per monomer M_w [Da]. The Matthews coefficient V_m is then given by the relation:

$$V_m = (V/z)/(n*M_w)$$

Considering that the molecular weight of *Ss*TrpB2b is 47778 Da (as calculated from sequence data) the analysis of the Matthews coefficient revealed, that the a.u. of this crystal form must contain one to two molecular copies of *Ss*TrpB2b (Table 2.8).

Table 2.8 Results from Matthews coefficient calculations for *Ss*TrpB2b.

Molecules per a.u.[1]	Matthews coefficient	Solvent content [%]
1	3.9	68.3
2	2.0	36.6

[1]monomers of *Ss*TrpB2b

Usually, the Matthews coefficient varies between 1.6 and 3.6 Å^3/Da, which suggested that two molecular copies of *Ss*TrpB2b were present in the a.u of this crystal form. This agrees with the oligomeric state of the functional form of the protein that is a biological dimer. In the subsequent steps of structure elucidation it was confirmed that two *Ss*TrpB2b units formed this a.u. The non-crystallographic symmetry (NCS) relation between the two subunits was explored using self-rotation functions in AMORE (Navaza, 1994) and POLARRFN (CCP4, 1994). Despite exploring medium and low resolution ranges as well as diverse Patterson radii of integration, no convincing solution was obtained. Best solutions typically exhibited correlation coefficients of 25-45%, were relatively unstable and contrasted poorly against background peaks. An analysis of native Patterson maps also did not reveal translational NCS. Thus, no conclusions on NCS relations could be established at that time.

2.4.4 Phasing by Molecular Replacement (MR)

TrpB1 and TrpB2 proteins are evolutionarily related. Thus, phasing followed molecular replacement (MR) protocols using TrpB1 as search model. Homologous structures in the PDB database (www.rcsb.org/pdb) were identified by a sequence based homology search (BLAST, Altschul *et al.,* 1990). This revealed two TrpB1 possible templates: TrpB1 from *Salmolnella typhimurium* (*St*TrpB1) and *Pyrococcus furiosus* (*Pf*TrpB1), with *Pf*TrpB1 as the closest homologue (PDB entry 1V8Z, Hioki *et al.,* 2004) with a 32 % sequence identity. Based on this homology, a search model was created with the program O (Jones *et al.,* 1991), where strictly conserved residues and residues with similar shape (*e.g.* Asp and Asn) were included in the search model and all other non-Gly residues were exchanged for Ala (or to Gly if present in the *Ss*TrpB2b sequence; Figure 2.14).

```
TrpB2b     MAMRIRIDLPQDEIPAQWYNILPDLPEELPPPQDPTGKSLELLKEVLPSKVLELEFAKERYVKIPDEV--  68
1V8Z       -------------------------------MWFGEFGGQYVPETLIEPLKELEKAYKRF-KDDEEFNR  37

poly-Ala                              AAAAGAAAEAAAEALAAAAAELEAAAARA-KAAAEA--


TrpB2b     -LERYLQV--GRPTPIIRAKRLEEYLGNNIKIYLKMESYTYTGSHKINSALAHVYYAKLDNAKFVTTETG  135
1V8Z       QLNYYLKTWAGRPTPLYYAKRLTEKIGGA-KIYLKREDLVHGGAHKTNNAIGQALLAKFMGKTRLIAETG  107

poly-Ala   -LAAYAAA--GRPTPAAAAKRLAEEAAGAA-KIYLKAEAAAAAGAHKANAAAAAAAAAKAAAAAAAAAETG


TrpB2b     AGQWGSSVALASALFRMKAHIFMVRTSYYAKPYRKYMMQMYGAEVHPSPSDLTEFGRQLLAKDSNHPGSL  205
1V8Z       AGQHGVATAMAGALLGMKVDIYMGAEDVERQKMNVFRMKLLGANVIPVNS-----GSRTL-KDA-----I  167

poly-Ala   AGQAGAAAAAAALAAMKAAIFMAAAAAAAAAAAAAAMAAAGAAVAPAAS-----GAAAL-KDA-----A


TrpB2b     GIAISDAV---EYAHKNGGKYVVGSVVNSDIM------FKTIAGMEAKKQ-MELIGEDPDYIIGVVGGGS  265
1V8Z       NEALRDWVATFEYTH-----YLIGSVVGPHPYPTIVRDFQSVIGREAKAQILEAEGQLPDVIVACVGGGS  232

poly-Ala   GAAAADAV---EYAH-----YAVGSVVAAAAA------FAAAAGAEAKAQ-AEAAGEAPDAIAGAVGGGS


TrpB2b     NYAALAYPFLGDELRSGKVRRKYIASGSSEVPKMTKGVYKYDYPDTAKLLPMLKMYTIGSD--FVPPPVY  333
1V8Z       NAMGIFYPFVNDK----KVKLVGVEAGGKGLES-GKHSASLNAGQVGVFHGMLSYFLQDEEGQIKPTHSI  297

poly-Ala   NAAAAAYPFAADA----KVAAAAAAAGAAAAAA-AKGAAADAAAAAAAAAMLAAAAAAAA--AAPAAAA


TrpB2b     AGGLRYHGVAPTLSLLISKGIVQARDYSQEESFKWAKLFSELEGYIPAPETSHALPILAEIAEEAKKSGE  403
1V8Z       APGLDYPGVGPEHAYLKKIQRAEYVTVTDEEALKAFHELSRTEGIIPALESAHAVAYAMKLAKEMSRD--  365

poly-Ala   AGGLAYAGVAPAAAALAAAGAAQAAAAAAEEAAKAAAAASAAEGAIPAAEAAHAAAAAAAAAAEAAAA--


TrpB2b     RKTVLVSFSGHGLLDLGNYASVLFKE  429
1V8Z       -EIIIVNLSGRGDKDLDIVLKVSGNV  388

poly-Ala   -AAVAVAASGAGAADLGAAAAAAAAA
```

Figure 2.14 Sequence alignment of *Ss*TrpB2b, PfTrpB1 (1V8Z) and the search model.
*Ss*TrpB2b (TrpB2b) vs. 1V8Z (Tryptophan Synthase from *P. furiosus*) with the sequence of the model (trimmed to the minimal common sequence) below. Strictly conserved residues are highlighted in yellow, residues with similar shape in cyan and residues that have been replaced by Gly in magenta background, respectively.

MR was carried out using the program PHASER (McCoy *et al.,* 2005). Multiple trials were performed using the dimeric TrpB1 as search model as well as the monomeric form (Table 2.9). The MR solution was visually validated in O (Jones *et al.,* 1991) for packing and fitting in a calculated electron density map.

Table 2.9 Solution statistics for MR using PHASER.

	Z-score (rotation)	Z-score (translation)	log-likelihood gain[1]
Monomer 1	6.4	3.6	38
Monomer 2	6.7	9.3	114

[1]refined log-likelihood gain: 128

2.4.5   Model building and refinement

Refinement of atomic coordinates relies on the minimization of the discrepancy between observed and calculated structure factor amplitudes. Refinement of atomic coordinates is a cyclic process, which involves algorithms that can remove errors (*e.g.* chain tracing or conformations) combined with manual interpretation of electron density maps and subsequent model building using visualization software, in this case the programs O and COOT were employed (Emsley & Cowtan, 2004). The refinement parameters of the cofactor (PLP) were derived using the PRODRG2 server (Schuettelkopf & van Aalten, 2004).

The initial model was progressively edited in iterative cycles of model building and subsequently refined in CNS (Crystallography and NMR Suite CNS, Brünger *et al.,* 1998). The structure was refined against native data between 20.00 and 1.94 Å resolution. Model refinement included conjugate gradient minimization of atomic coordinates, bulk solvent correction and isotropic individual B-factor refinement with geometrical and B-factor NCS restraints applied throughout. Large parts of the initial model had to be truncated after it became evident that these disagreed with calculated maps leading to regions of non-interpretable electron density.

The program ARP/wARP (Perrakis *et al.,* 1999) was then used for model improvement (map enhancement mode). ARP/wARP is a program package for automated model building and refinement of protein structures. It combines reciprocal space structure-factor refinement with updating of the model (adding and/or removing

atoms) in real space to construct and improve protein models. Map enhancement mode uses a hybrid-model (combination of a partial protein model and a free-atom set), which allows a considerably better description of the map, since prominent features in electron density (unaccounted for by the current model) are described by free atoms. This procedure yielded significant phase improvement and subsequently improved electron density. The quality was sufficient for completion of model building (Figure 2.15). NCS restraints were removed at this point.

Since ARP/wARP uses REFMAC (CCP4 program suite; CCP4, 1994) for refinement, this program was then used for maximum likelihood based restrained model refinement without a translation/liberation/screw (TLS) tensor. This included sparse matrix minimization, overall individual isotropic B-factor refinement and bulk solvent correction. Solvent molecules were identified with ARP/wARP (solvent mode routine) and COOT. The solvent mode routine of ARP/wARP continuously updates the solvent structure, while the protein model remains unchanged. For refinement statistics see Table 2.10. Several residues were not identified as they were disordered in both NCS copies: the N-terminus (Met1 - Met3), the C-terminus (Glu429) and an N-terminal segment (Asp34 - Leu43; 10 residues). Additionally, there is a helix (H5', red arrow in Figure 2.15A), which showed interpretable electron density in one subunit and merely traceability for the backbone only in the other subunit.

A

B

44.2/46.7   40.5/41.7   18.6/25.4   R/R$_{free}$
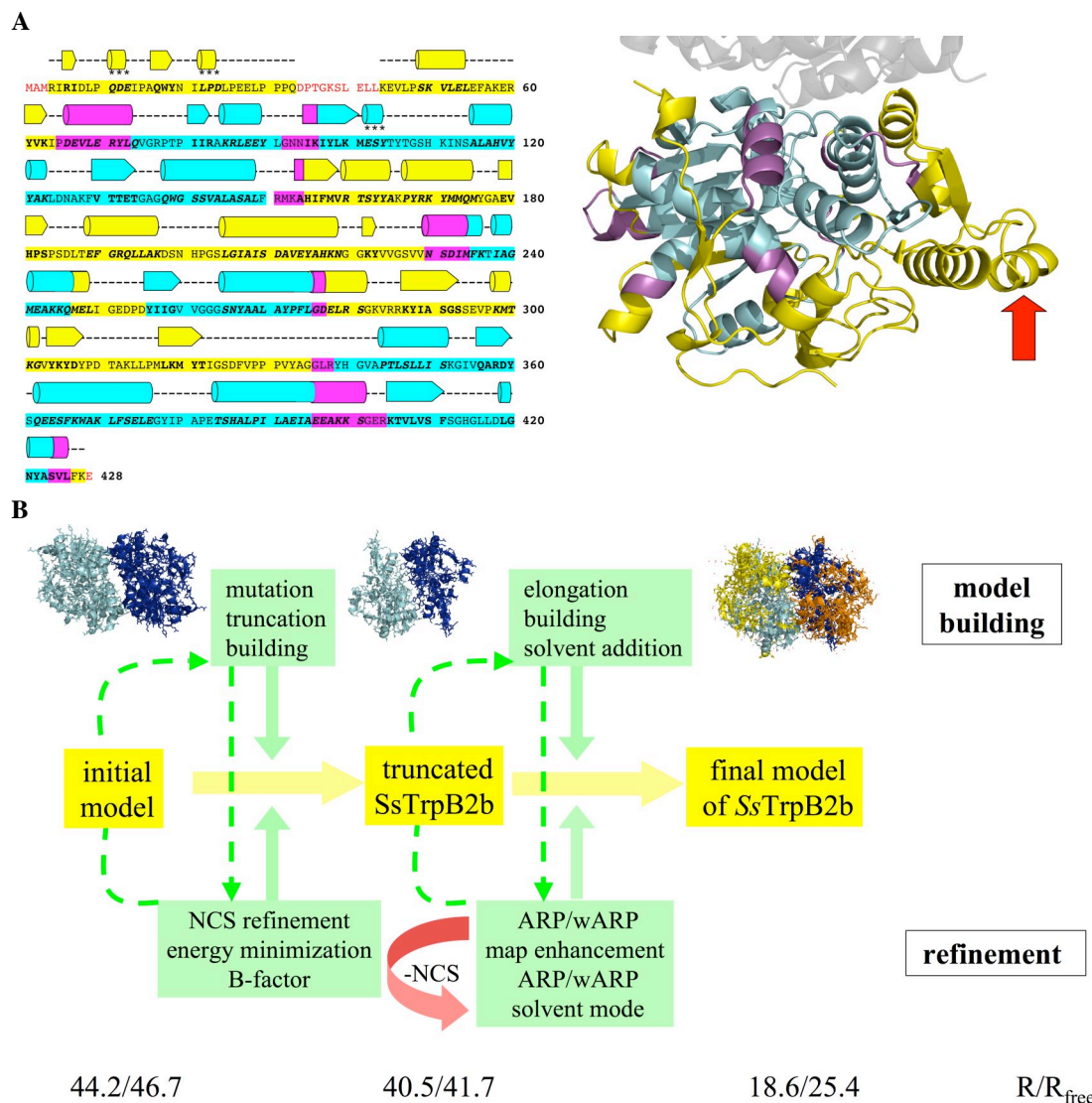
Figure 2.15 Model improvement in ARP/wARP and overview on model building/refinement.
(**A**) Sequence of *Ss*TrpB2b with secondary structure elements assigned on top. Residues with interpretable electron density in the truncated *Ss*TrpB2b model are colored in cyan, residues built after initial refinement rounds in magenta and ARP/wARP in yellow, respectively. The α-helix H5' is indicated by a red arrow. (**B**) The models in different stages of refinement are given in the yellow boxes and thumbnails above the boxes. The steps in order to improve the model done in building software are given in the green boxes above the models and done with refinement tools below the models. NCS restraints were excluded from the final stages of refinement. The dashed black double arrows indicate the cyclic nature of refinement.

Table 2.10 Model parameters of *Ss*TrpB2b.

R-factor/R_{free} [%][1]	18.6/24.5
Number of protein residues	828
Number of solvent molecules	362
Average B-factor (MC/SC/PLP)[2]	31.17/33.76/29.47 ‖ 24.35/27.11/22.63
rmsd (between NCS copies) [Å]	0.625 (all atoms)[3]
rmsd: bond length [Å]/bond angle [°]	0.019/1.77

[1]The R-free set comprised 894 reflections corresponding to 1.63% of the total data (54720 reflections).
[2]MC = main-chain, SC = side-chain for copy 1 ‖ copy 2
[3]calculated with LSQKAB

For a Ramachandran diagram of the two NCS-related copies of TrpB2b see Figure 2.16, for Ramachandran plot statistics Table 2.11 and for a summary of the steps towards the elucidation of the structure of *Ss*TrpB2 see Figure 2.17.

Table 2.11 Ramachandran plot[1] statistics for *Ss*TrpB2b.

	Number	Percentage
Residues in favored regions	636	91.4
Residues in additional allowed regions	52	7.5
Residues in generously allowed regions	3	0.4
Residues in disallowed regions	5	0.7

[1]for 696 non-Gly and non-Pro residues from analysis in PROCHECK

Residues in disallowed regions cluster in loop L8' (illustrated in Figure 2.28) in both copies. The electron density in this loop is poor and subsequently the residues have high B-factors indicating mobility. The carbonyl oxygens of these residues coordinate a single water molecule, which may influence the main-chain conformation in this region.
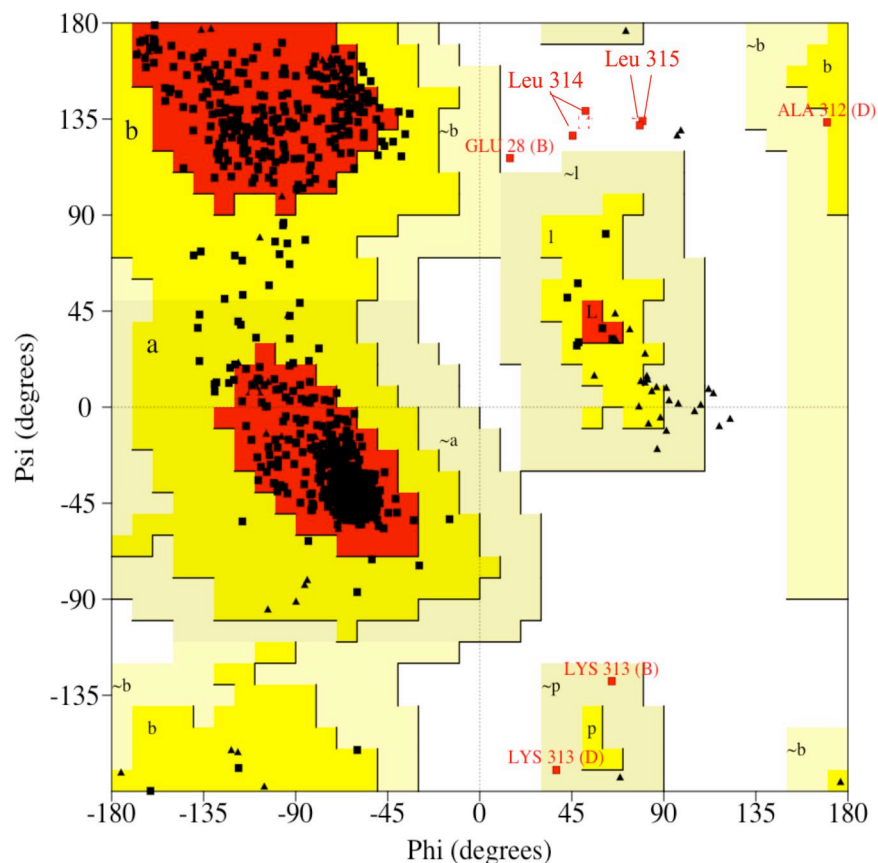
Figure 2.16 Ramachandran diagram for both NCS-related copies of *Ss*TrpB2b.
Favored regions are indicated by red, additional allowed regions by yellow, generously allowed regions by light yellow and disallowed regions by white background, respectively. All non-Gly and non-Pro residues (black squares) are in allowed regions. Most Gly and Pro residues (black triangles) are in allowed regions, as well. The plot was generated using PROCHECK (Laskowski *et al.*, 1993).

A
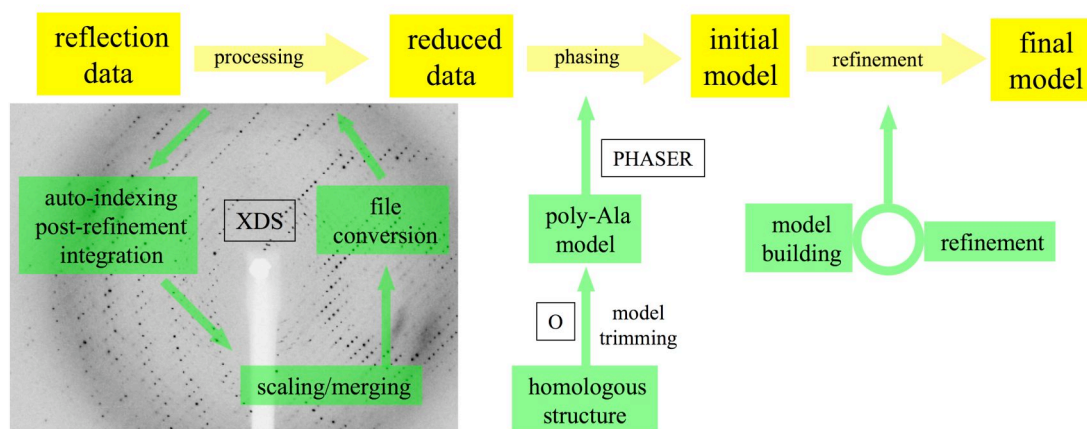
Figure 2.17 Summary of the steps during structure elucidation.
(**A**) The path to structure solution (yellow): linear steps during processing and phasing and cyclic iterative structure refinement (green side tracks, more details below).

2.5 Results and discussion

2.5.1 Overall structure of *Ss*TrpB2b

The crystal structure of *Ss*TrpB2b has been elucidated at 1.94 Å resolution using the MR method and is the first structure of a TrpB2 family member. The asymmetric unit (a.u.) includes two chains of *Ss*TrpB2b, corresponding to the biological dimer. The two molecular copies are related by a two-fold proper NCS (*rmsd* of 0.5 Å over 420 C_α-atoms calculated with TOP, CCP4, 1994). Both chains of *Ss*TrpB2b in the a.u. contain residues Arg4-Gln34 and Lys44-Lys428. Residues Asp186-Ser199, corresponding to H5', show clear density in one chain and weak density in the other.

The structure of *Ss*TrpB2b (Figure 2.18) reveals a fold similar to that of TrpB1 proteins reported so far (Lee *et al.*, 2005; Hyde *et al.*, 1988; Schneider *et al.*, 1998). For the assignment of secondary structure elements mapped onto the sequence of *Ss*TrpB2b see Figure 2.19. The secondary structure elements have been assigned using the software PROMOTIF (Hutchinson & Thornton, 1996) and their numbering is based on the comparison of *Ss*TrpB2b and *St*TrpB1 (Schneider *et al.*, 1998). Differences in the overall structure between *Ss*TrpB2b and the TrpB1 family members cluster in the hypothetical TrpA interface and are likely to be the effectors of the divergent interaction behavior of TrpB2 family members. These differences are as follows:

(i) The N-terminal extension: residues Arg4 to Asp66 adopt a different tertiary structure (details in section 2.5.5a, magenta bar in Figure 2.19).

(ii) The cation binding-site region (residues Val261 to Tyr339, green bar in Figure 2.19): shift of a loop segment and coordination of *Ss*Lys298 instead of the cation (details in section 2.5.5b).

(iii) The insertion helix H5' (residues Glu189 to Lys197, section 2.5.5c): this α-helix is unique to TrpB2 family members (yellow bar in Figure 2.19).

Additional small extensions of structural elements (H3 and H12) should participate into the local organization of *Ss*TrpB2b without any further effects on association.
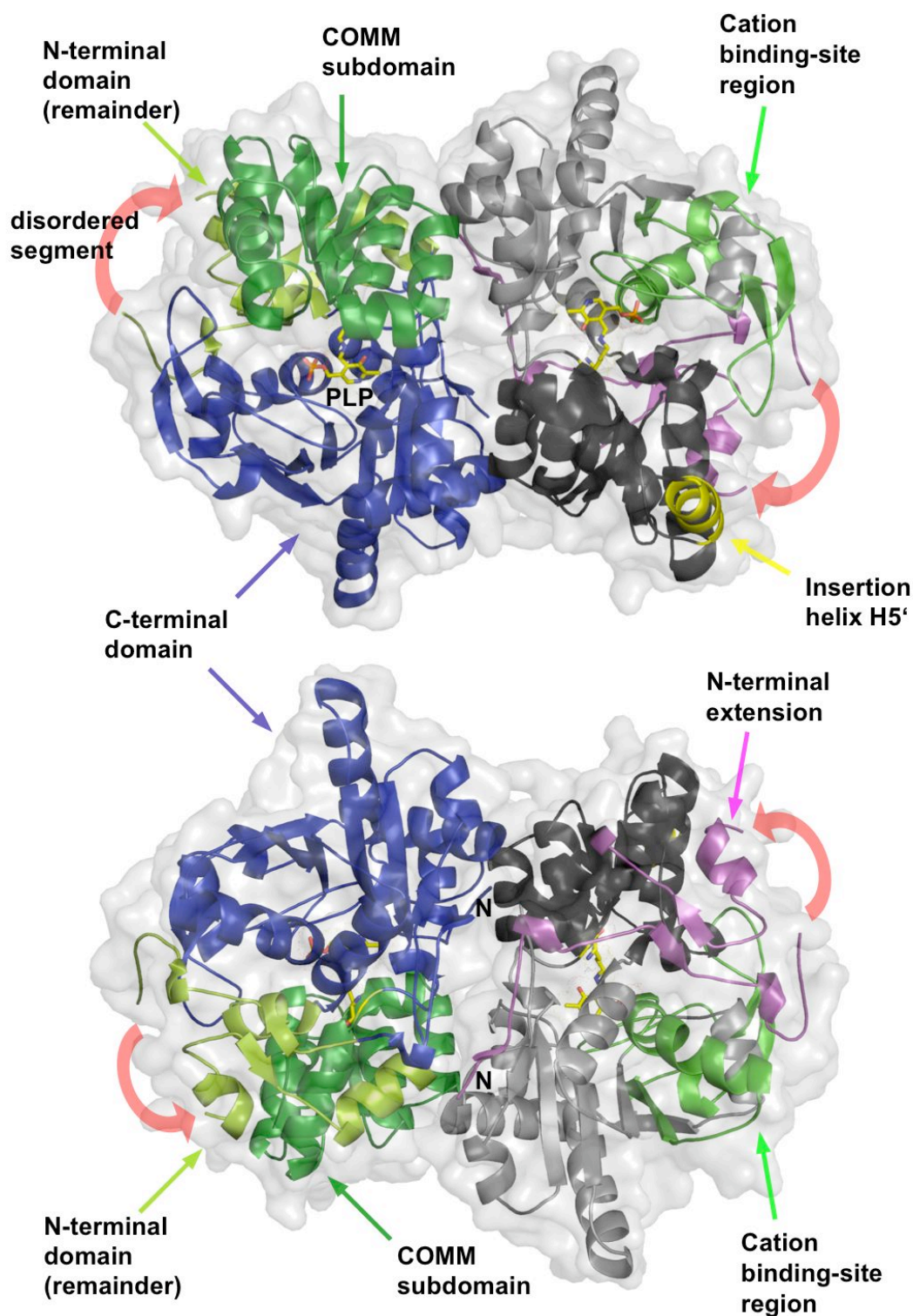
Figure 2.18 Overall structure of *Ss*TrpB2b.
The left subunit is colored according to the domain organization of *Ss*TrpB2b: the N-terminal domain is colored dark green (COMM subdomain) and lemon (remainder of N-terminal domain). The C-terminal domain is colored blue. The right subunit is colored according to differences between *Ss*TrpB2b, *Pf*TrpB1 and *St*TrpB1: the N-terminal domain is colored in black. In this domain there are the N-terminal extension (magenta) and the insertion helix H5' (yellow). The cation binding-site region (green) is in the C-terminal domain. The cofactor (including *Ss*Lys111) in both subunits is shown in stick representation (carbon: yellow, oxygen: red, nitrogen: blue and sulfur: orange), The upper and lower panel correspond to a 180° rotation around the horizontal axis in the paper plane. The N-termini are indicated in the lower panel (C-termini are not visible).
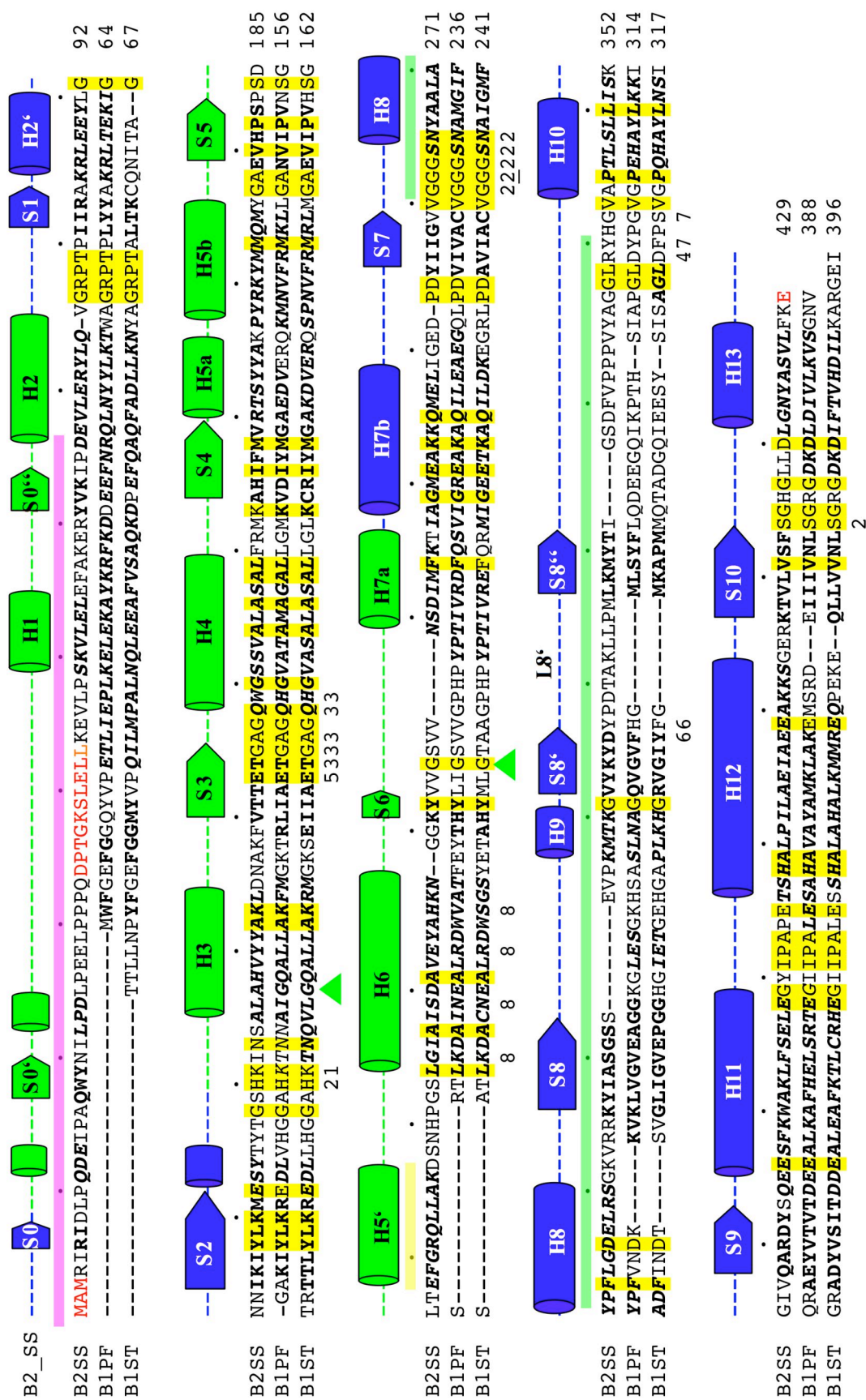
Figure 2.19 Structure based sequence alignment of *Ss*Tr-B2b, *Pf*TrpB1 and StTrpB1.

Figure 2.19 Structure based sequence alignment of *SsTrB2b, PfTrpB1* and StTrpB1.
The magenta bar indicates the N-terminal segment that shows altered tertiary structure between *Ss*TrpB2b (B2SS), *Pf*TrpB1 (B1PF) and *St*TrpB1 (*St*B1). The yellow and green bars indicate the insertion helix H5' and the cation binding-site region, respectively. The green triangles point to the start and end of the mobile subdomain (COMM) in the N-terminal domain. Every tenth amino acid in the *Ss*TrpB2b sequence is indicated by a dot above the sequence. A cartoon representation shows the secondary structure elements of *Ss*TrpB2b (N-terminal domain in green; C-terminal domain in blue). Disordered amino acids are indicated by red letters. Strictly conserved residues are shown with yellow background. Amino acids involved in secondary structure elements are written in bold letters (β-strands) and bold/italic (α-helices). Numbers below the sequences indicate residues in catalytic/structural features: 1 E-Lys, 2 E-Lys coordination, 3 E-Ser/E-Trp (Ser/Trp coordination) 4 E-Ser coordination and allosteric communication (side chain) 5 E-Trp coordination (side chain), 6 tunnel gating, 7 Na$^+$-binding site (note: involved *Ss*G262 is underlined), 8 allosteric regulation.

2.5.2 Domain organization of *Ss*TrpB2b

Each subunit of the β$_2$-dimer comprises the N- and C-terminal domains as reported for TrpB1 from *Pyrococcus furiosus* (Hioki *et al.,* 2004) and *Salmonella typhimurium* (Hyde *et al.,* 1988; Schneider *et al.,* 1998). Domain assignment (Table 2.12) has been done by comparative analysis of structure based sequence alignments of *St*TrpB1 (1BKS, *rmsd* = 1.41 Å over 273 C$_\alpha$ atoms, 33 % identity), *Pf*TrpB1 (1WDW, *rmsd* = 1.35 Å over 291 C$_\alpha$ atoms, 33 % identity) and *Ss*TrpB2b. Additionally, the very N-terminal residues (Arg4 to Asp8) contribute to the central β-sheet of the C-terminal domain (Figure 2.20) and, thus, have been assigned to the C-terminal domain. The main segment of each of the domains contains a 3-layer-(ABA) sandwich with Rossmann fold topology, which has been described in the CATH database for *St*TrpB1.

Table 2.12 Domain organization of *Ss*TrpB2b.

N-terminal domain	C-terminal domain
	Arg4 - Asp8
Leu9 - Gln74	Val75 - Ser109
His110 - Lys236[1]	Thr237 - Lys428

[1]residues Ala117 (H3) to Gly226 (downstream of S6) form the COMM domain
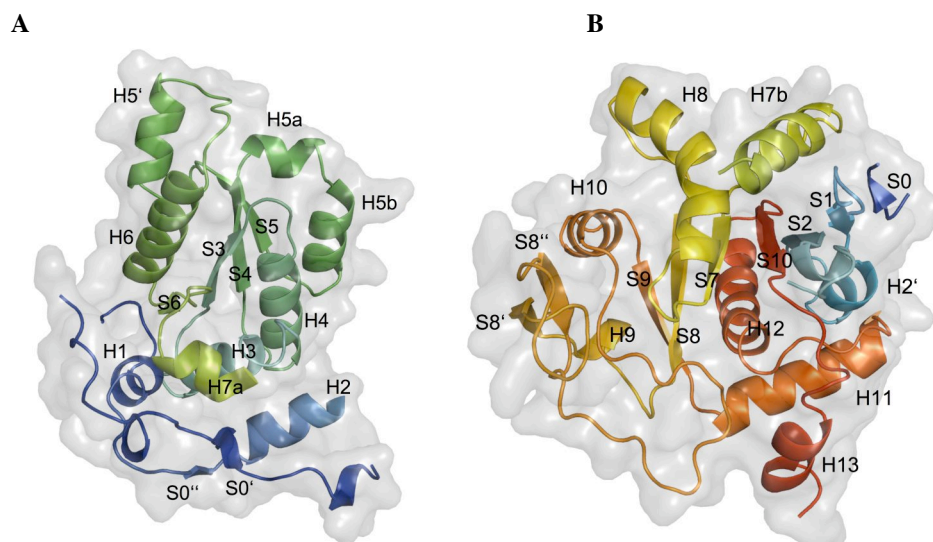
Figure 2.20 The domains of *Ss*TrpB2b.
*Ss*TrpB2b was colored from the N-terminus (blue) to the Cterminus (red) and split into its domains. (**A**) The N-terminal domain viewed from the Cterminal domain. The core of the domain comprises central four-stranded parallel β-sheet formed by β-strands S3, S4, S5 and S6. The β-sheet is flanked by α-helices H3, H4, H5a and H5b on one side as well as H6 and H5' on the other side. The COMM subdomain includes secondary structure elements from H3 (cyan) to S6 (lemon). The short segment comprises a two-stranded antiparallel β-sheet (strands S0' and S0''). The α-helices H1, H2 and H3 pack between this sheet and the core region. (**B**) The C-terminal domain viewed from the N-terminal domain. The C-terminal domain comprises a central seven-stranded mixed β-sheet formed by S0, S1, S2, S7, S8, S9 and S10. This β-sheet is flanked by α-helices H2', H11, H12 and H13 on one side and α-helices H7b, H8, H9 and H10 on the other side. Strands S8' and S8'' form another, two-stranded, antiparallel β-sheet, which is separated by α-helices H9 and H10 from the central sheet. (**A/B**) The adjacent subunit is located on the right side of the domains.

The core regions of the N- and C-terminal domains of TrpB1 proteins share mutual structural homology, primarily involving their core regions (Table 2.13; Figure 2.21).

Table 2.13 Comparison of the N- and C-terminal domains of *Ss*TrpB2b, *St*TrpB1 and *Pf*TrpB1.

Tryptophan Synthase	*rmsd* [Å]/ C_α-atoms	Reference
*Ss*TrpB2b	2.2/51[1]	
*St*TrpB1	2.2/73[2]	Hyde *et. al.*,1988
*Pf*TrpB1	2.7/69[3]	Hioki *et. al.*, 2004

superposed with [1]TOP, [2]ALIGN, [3]LSQKAB
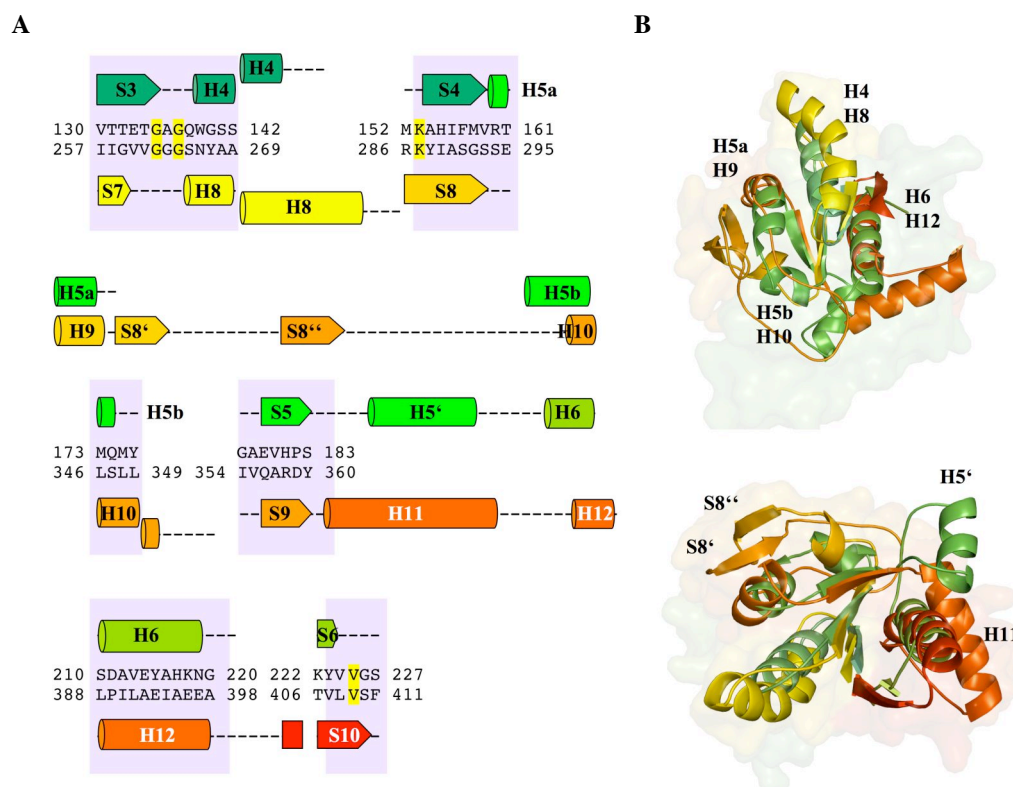
Figure 2.21 Superposition of the N- and C-terminal domains of *Ss*TrpB2b.
Structural elements are colored as in Figure 2.20. (**A**) The sequence of structurally equivalent residues is highlighted purple and strictly conserved residues by yellow background. (**B**) Upper panel: equivalent secondary elements are labeled. Lower panel: non-equivalent secondary structure elements are labeled.

The central β-sheets of the N-terminal (S3 - S6) and C-terminal (S7 - S10) domains as well as their flanking α-helices (H4 and H6 in the N-t and H8 and H12 in the C-terminal domain) are structurally equivalent. The α-helices H5' (N-terminal domain) and H11 (C-terminal domain) deviate in their positions. H5' is not present in TrpB1 family members. This indicates that *Ss*TrpB2b may have evolved from two ancestral proteins corresponding to each of the domains present. It is likely that theses ancestral proteins have been created by gene duplication. In fact, a structure based PDB search (using the DALI server, Holm & Sander, 1993) revealed several bacterial/archeal proteins with structural similarity to the COMM (Table 2.14) and the C-terminal domain (Table 2.15). These homologs, mostly of unknown function, could be ancestral proteins, which may still remain in the genomes but an evolved protein with enhanced functionality may have replaced them functionally.

Table 2.14 The COMM subdomain and similar domains (from DALI query)

PDB ID	*rmsd* [Å] over C_α	Z^1	#aa	Description
1L1S	3.7 (83)	5.7	111	Protein of unknown function MTH1491[2]
1T57	3.2 (91)	5.4	179	Protein of pyruvate kinase activity MTH1675[2]
1ID1	3.4 (73)	5.1	153	RCK domain from *E. coli* potassium channel

[1]higher Z-scores mean higher similarity (dissimilar for Z < 2)
[2]from *Methanobacterium thermoautotrophicum* (structural genomics)

Table 2.15 The C-terminal domain and similar domains/proteins (from DALI query)

PDB ID	*rmsd* [Å] over C_α	Z^1	#aa	Description
1T57	4.9 (114)	4.5	179	Protein with pyruvate kinase activity MTH1675[2]
2F46	4.2 (94)	3.6	39	Protein of unknown function 7380613[3]
1PDO				Putative phosphoprotein phosphatase[4]

[1]higher Z-scores mean higher similarity (dissimilar for Z < 2)
[2]from *Methanobacterium thermoautotrophicum* (structural genomics)
[3]from *Neisseria meningitis* FAM18 (structural genomics)
[4]from *Arabidopsis thaliana gene* AT1G05000 (structural genomics

Thus, evolution of the *Ss*TrpB2b monomer (or in general TrpB2 proteins) could have included recombination of paralogues. Then, a hypothetical pathway of interface formation resulting in an ancient *Ss*TrpB2 version has probably gone over hetero-complex to hetero-obligomer and consequential domain-domain interface formation by fusion of the two chains.

2.5.3   Active site of SsTrpB2b

The active sites of TrpB1 and TrpB2 family members are almost identical. The residues equivalent to those reported by Rhee and coworkers (1997) are involved in PLP-binding/-coordination (Figure 2.22 and Table 2.16) and are strictly conserved between the two families. In both families, the distance between the PLP molecules in the active sites of each of the subunits across the homo-obligomer interface is approximately 30 Å.
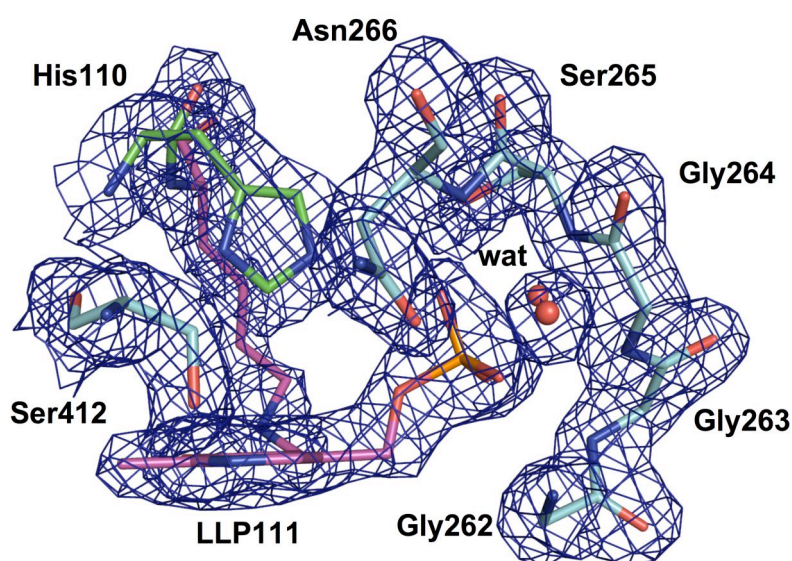
Figure 2.22 Active site residues and the corresponding electron density (contoured at 1 σ).
Residues/waters in a 3.2 Å sphere of the cofactor vitamin B_6 are shown in stick/cpk representation. The carbon atoms of the covalent bound Lys111 and the cofactor are shown in magenta (residue called LLP111). The remaining residues are colored by domain (N-terminal green, C-terminal cyan). Oxygen atoms are colored red, Nitrogen blue and Sulfur orange, respectively.

Table 2.16 Binding of the cofactor in *Ss*TrpB2.

*Ss*TrpB2b[1]	Binding/coordination of
Lys111	PLP, covalent binding
His110	Phosphate binding
Gly262-Gly-Gly-Ser-Asn266	P-loop, phosphate binding
Ser412	Pyridine nitrogen
2 x H_2O	Phosphate

[1]residues are strictly conserved in both TrpB families

Based on the assumption of identical substrate coordination properties between the TrpB family members, potential catalytic residues can be assigned in TrpB2. Comparison of *Ss*TrpB2b with *St*TrpB1 structures suggests that the segment Glu133-Thr-Gly-Ala-Glu-Trp139 (equivalent to Glu109-Thr-Gly-Ala-Glu-His115 in *St*TrpB1; S3 to H4 in Figure 2.19) fulfill identical roles in both TrpB families. *Ss*Trp139 (equivalent to *St*His115) is the only residue in this segment that is not strictly conserved between TrpB1 and TrpB2 family members. *Ss*Arg338 (equivalent to *St*Asp305, which coordinates E-Ser) is another residue without conservation.

However, both *Ss*Trp139 and *Ss*Arg338 are strictly conserved within the TrpB2 family while the equivalent *St*His115 and *St*Asp305 are strictly conserved among TrpB1 proteins. (Figure 6.2; appendix 6.2). Superposition of putative catalytic residues in PyMOL (DeLano, 2002) reflects the high similarity of *Ss*TrpB2b to *Pf*TrpB1 (*rmsd* = 0.76 Å, *apo*, PDB code 1WDW) as well as *St*TrpB1 (*rmsd* = 0.51 Å, *apo*, PDB code 2WSY). Even the side chains show almost identical conformations (Figure 2.23).
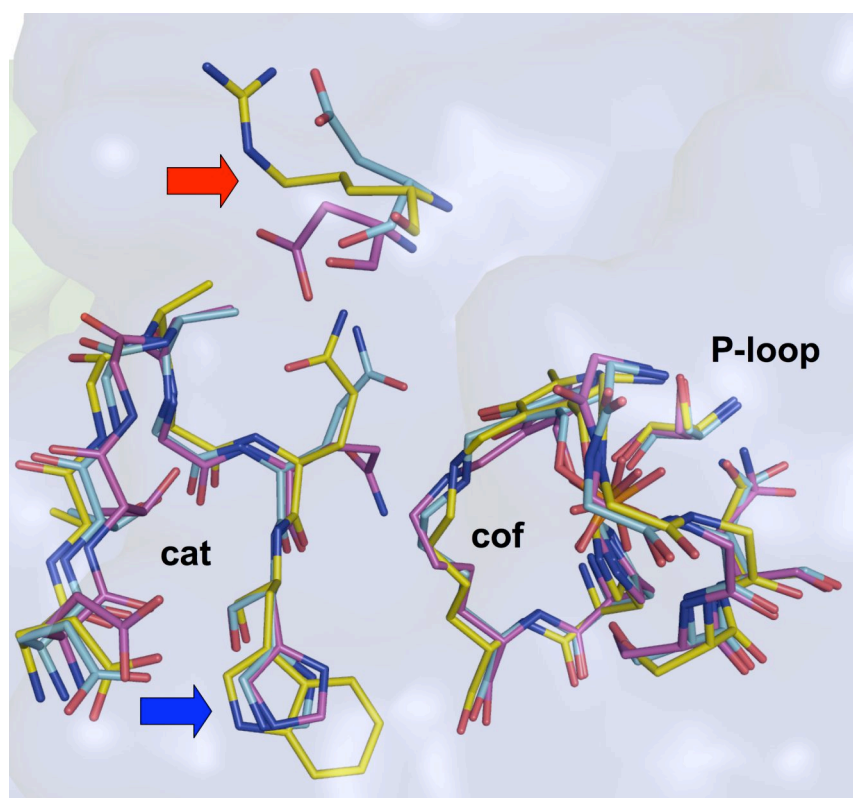
Figure 2.23 Superposition of some catalytic residues between TrpB family representatives. *Ss*TrpB2 (yellow carbon atoms), *St*TrpB1 (magenta carbon atoms) and *Pf*TrpB1 (green carbon atoms) superpose with very low *rmsd*s. Residues without conservation between the families are indicated by arrows: *Ss*Trp139 (blue arrow) superposes better than *Ss*Arg338 (red arrow). The catalytic turn (cat) and the P-loop with the internal aldimine and the adjacent *Ss*His110 ("cof" for cofactor binding) show an excellent superposition.

The side chain of *Ss*Trp139 adopts a conformation almost identical to that of *St*His115. The nitrogen atoms of the ring systems of the two amino acids are in the same position. This effect results from a H-bond of the indole side chain to the side chain oxygen of *Ss*Thr134. This polar interaction can be equally formed by His and Trp residues. The alternating properties of the His/Trp side-chains do not appear to affect the catalytic properties of the TrpB1 and TrpB2 families, which are equivalent

according to biochemical studies (Leopoldseder *et al.,* 2006) since only the main-chains of *Ss*Trp138 and *St*His115 are involved in E-Ser/E-Trp coordination.

On the contrary, the mutation of a residue with function in allosteric communication and/or substrate specificity can have severe effects on the functionality of an enzyme. *St*Asp305 has been shown to have an effect on substrate specificity (Ferrari *et al.,* 2003). Based on the analysis of a *St*Asp305Ala mutated variant, this role has been assigned to the carboxylate group of the side-chain. Due to the significant difference in size and shape, *Ss*Arg338 is unlikely to be involved in catalysis like *St*Asp305. The guanidine group of *Ss*Arg338 only forms water-bridged H-bonds to *Ss*TrpB2b residues (illustrated in Figure 2.28 in section 2.5.5b). Allosteric communication, which involves H-bonding between the side-chains of *St*Asp305 and *St*Lys167 in *St*TrpB1, is unlikely in this case. Residues involved in cofactor binding and coordination as well as residues with putative catalytic function complete the active site of *Ss*TrpB2b (Figure 2.24).
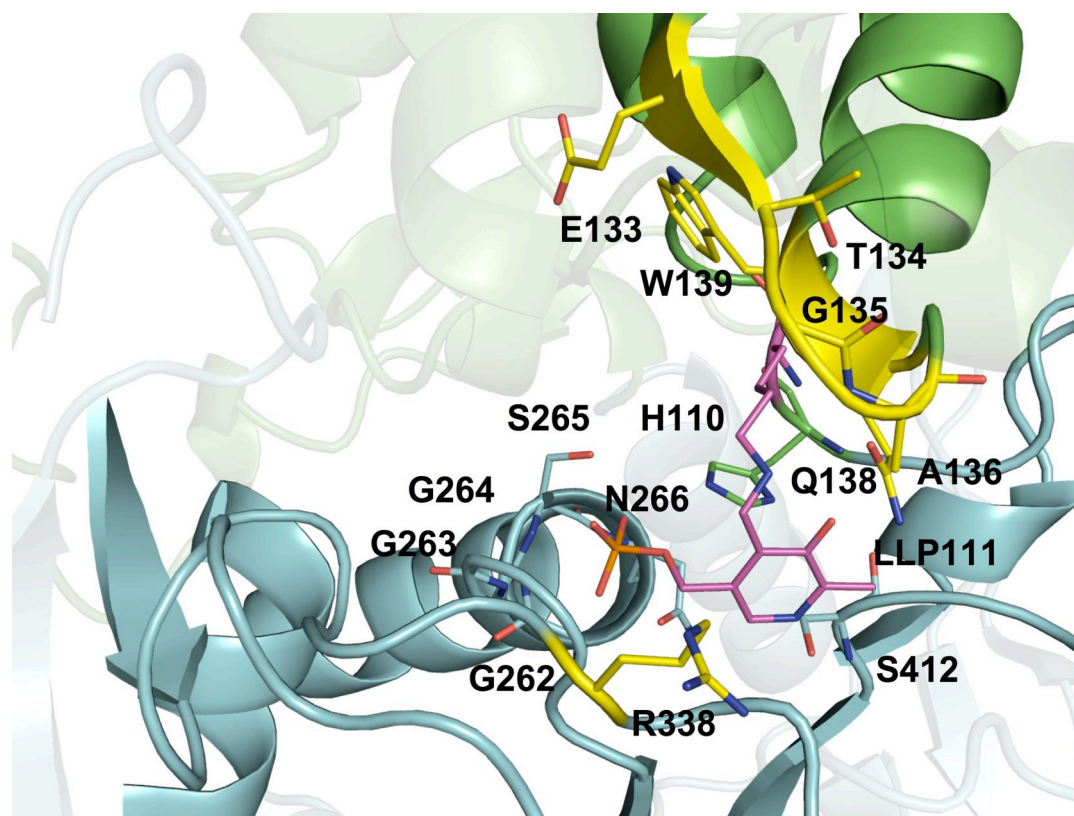
Figure 2.24 The active site of *Ss*TrpB2b.
View from the active site of a hypothetical α-subunit into the *Ss*TrpB2b active site. The cofactor and the internal aldimine to Lys111 (LLP111) are colored in magenta. Residues that have a role during catalysis are indicated by stick representation (green: N-terminal domain; blue: C-terminal domain; exception are putative residues reported for substrate/product binding: yellow). Secondary structure elements upstream and downstream of residues with catalytic function are represented as solid cartoons. For a better view, the remaining secondary structure elements are represented as transparent cartoons.

2.5.4   Protein-protein interfaces of *Ss*TrpB2b

*Ss*TrpB2b comprises two protein-protein interfaces: a domain-domain interface between the N- and C-terminal domains and a homo-obligomer interface between the subunits of *Ss*TrpB2b. These interfaces have been analyzed with LIGPLOT (Wallace *et al.,* 1995) to elucidate their interaction pattern, which has been mapped to the sequences of *Ss*TrpB2b and *Ss*TrpB2a (Figure 2.25).

The domain-domain interface between the N- and C-terminal domains shows a predominantly hydrophobic character. Aliphatic residues are most abundant in this interface and mainly form interactions between the non-COMM segment and the C-terminal domain. There is a polar contribution, which clusters in the N-terminal sequence segment characteristic of the TrpB2 family. These results reveal the permitted mobility of the COMM domain during catalysis in an otherwise tightly packed enzyme resembling intra-domain interfaces. The residues involved in the domain-domain interface show high conservation between *Ss*TrpB2b and *Ss*TrpB2a, while two residues (*Ss*TrpB2b Glu27 and Lys352) are swapped between the two proteins. The high conservation (54% sequence identity) suggests a similar fold of these two TrpB2 family members, while the disordered segment upstream to H1 could show differences between *Ss*TrpB2a and *Ss*TrpB2b.

The homodimer interface shows high similarity between TrpB1 and TrpB2 family proteins. The N-terminal domain of one subunit predominantly interacts with the C-terminal domain of the other subunit (and vice versa). The amino acid composition of the interface is predominantly hydrophobic, confirming the trend described in section 1.4.4. In *Ss*TrpB2b, the most abundant residues in the interface are Arg and Tyr, which in some cases use the full spectrum of possible interactions, emphasizing their role as global players in protein-protein interfaces. Additionally, several Leu residues contribute to the interface. The segment of the N-terminal extension in the rim region of the dimer interface, which contributes roughly 10 % to the interface (Table 2.17), is an exception to the remaining hydrophobic dimer interface. However, its contribution to a gain in stability should be higher than expected by pure comparison of interface areas, since in contains mostly H-bonds.

Table 2.17 Comparison of dimer interfaces of *Ss*TrpB2b and TrpB1 family proteins.

Tryptophan Synthase (PDB code)	Buried surface area [Å²][1]
*Ss*TrpB2b	4437.8
*St*TrpB1 (1BKS)	4021.5
PfTrpB1 (1WDW)	4057.6

[1]the buried surface area has been calculated with GRASP (Nicholls *et al.,* 1991).
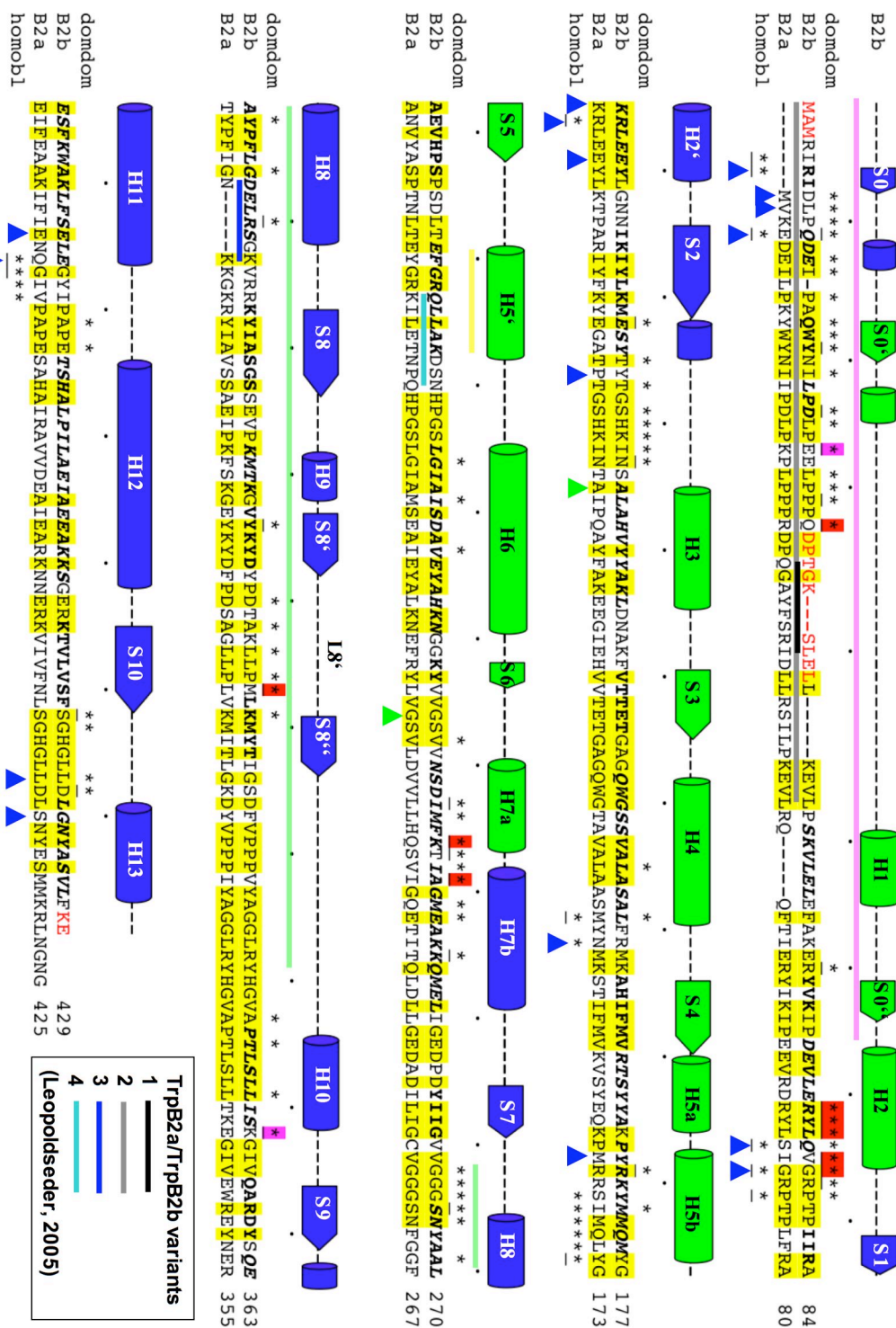
Figure 2.25 The interfaces of *Ss*TrpB2b (B2b) in comparison with *Ss*TrpB2a (B2a).

Figure 2.25 The interfaces of *SsTrpB2b* (B2b) in comparison *with* *SsTrpB2a* (B2a).
The sequences are aligned as in Figure 6.2 in the appendix. The composition of this figure is analogous to Figure 2.19. Asterisks (underlined for formation of H-bonds) indicate residues forming interactions in the domain-domain (domdom, above *Ss*TrpB2b) and homo-obligomer (homobl). In the domain-domain interface (above *Ss*TrpB2b sequence), red background indicates residues with mainchain-mainchain interactions involved in the local organization of structural elements (*e.g.* H2) and magenta background the swapped residues (reflecting a pseudo-conservation), respectively. In the homo-obligomer interface, residues forming an H-bond to one of the 16 bridging waters are indicated by blue triangles (there are no waters involved in the domain-domain interface.).

Surprisingly, the polar segments of the N-terminal extension contribute to both, the domain-domain and the homo-obligomer interface indicating a role as an interconnecting stabilization element between the domains and subunits in the context of thermophilic adaptation. Between the TrpS families, the fixation of the N-terminus in the domain-domain and homo-obligomer interface is unique to the TrpB2 family. In TrpB1 proteins, hetero-obligomer formation may have made the N-terminal extension segment obsolete, since the N-terminus of TrpB1 family members is anchored in the hetero-obligomer interface.

2.5.5   Differences between the TrpB1 and TrpB2 families

2.5.5a The N-terminal extension

The N-terminal extension (residues Arg4 to Asp66) interconnects all interfaces in TrpB2 family members: the N-terminus folds into the homo-obligomer interface and it progresses along the domain-domain interface towards the putative hetero-complex (*Ss*TrpB2a only) interface (Figure 2.26). The α-helices H1 in TrpB1 and TrpB2 are the first co-locating structural elements (perpendicular orientation), while H1 in *Ss*TrpB2b packs into a cavity not present in TrpB1 family members (upstream to H7a in Figure 2.19). In TrpB1 family members, H1 is a central element in the hetero-obligomer TrpA interface, while it is distant to the interface in *ss*TrpB2b. This suggests, that H1 is not involved in the hetero-complex interface between *Ss*TrpB2a and *Ss*TrpA. In contrast, the disordered segment in *Ss*TrpB2b (Asp33 to Leu43, upstream to H1) could be involved in interactions between *Ss*TrpB2a and *Ss*TrpA. Two insertions in this segment could influence the association properties of *Ss*TrpB2b and *Ss*TrpB2a.
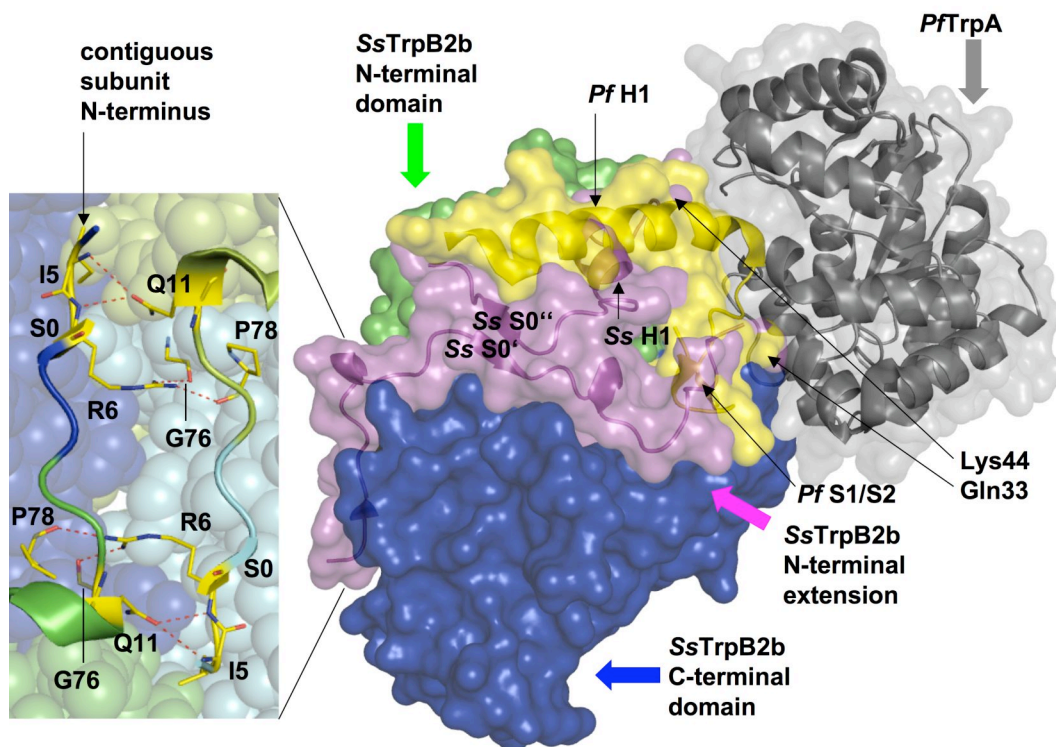
Figure 2.26 The N-terminal extension.
Right: the N-terminal extension in *Ss*TrpB2b (magenta) compared to the structure of the N-terminus in *Pf*TrpB1 (yellow). The N-terminus of *Ss*TrpB2b reveals a completely altered structure compared to the N-terminus of *Pf*TrpB1. *Pf*TrpA is shown in gray. The borders of the disordered segment (Gln33 to Lys44) in the N-terminal extension are indicated by arrows. Only one subunit is shown. Left: zoom into the homo-obligomer interface. Residues forming H-bonds (indicated by red dashes) across the interface are shown in stick representation and labeled. The N-terminus contributes with β-strand S0 (labeled) to the C-terminal domain.

The N-terminal extension is truncated in *Ss*TrpB2a suggesting that S0 is missing in the central β-sheet of the C-terminal domain. This may be a first step towards anchorage of the N-terminus segment in the TrpA interface analogous to TrpB1 family members.

2.5.5b The cation binding-site region

Proteins of the TrpB1 family show coordination of a Na^+ ion, while in *Ss*TrpB2b the amino group of *Ss*Lys298 replaces this cation. The coordination sphere of Na^+ and *Ss*Lys298 show high similarity (Table 2.18 and Figure 2.27). *Ss*Lys298 anchors the α-helix H9 in the "cation-binding site" of *Ss*TrpB2b. Thus, *Ss*Lys298 links three elements in the "cation binding-site" (instead of two in *St*TrpB1). This may provide additional stabilization to the corresponding region.

Table 2.18 Cation binding site in proteins of the TrpB1 and TrpB2 families.

Tryptophan Synthase	Ligand	Coordination sphere[1]	Reference
*Ss*TrpB2b	K298	V261, G262, Y339, H$_2$O[2]	
*St*TrpB1	Na$^+$	G232, F306, S308, 2 H$_2$O[3]	Rhee *et al.,* 1996
*Pf*TrpB1	Na$^+$	G227, Y301, G303, 3 H$_2$O[4]	unpublished, 1V8Z used, only partial occupancy observed

[1]conserved residues (position) in the coordination sphere are underlined
coordination: [2]tetragonal planar, [3]tetragonal pyramidal, [4]octahedral (in 1V8Z)
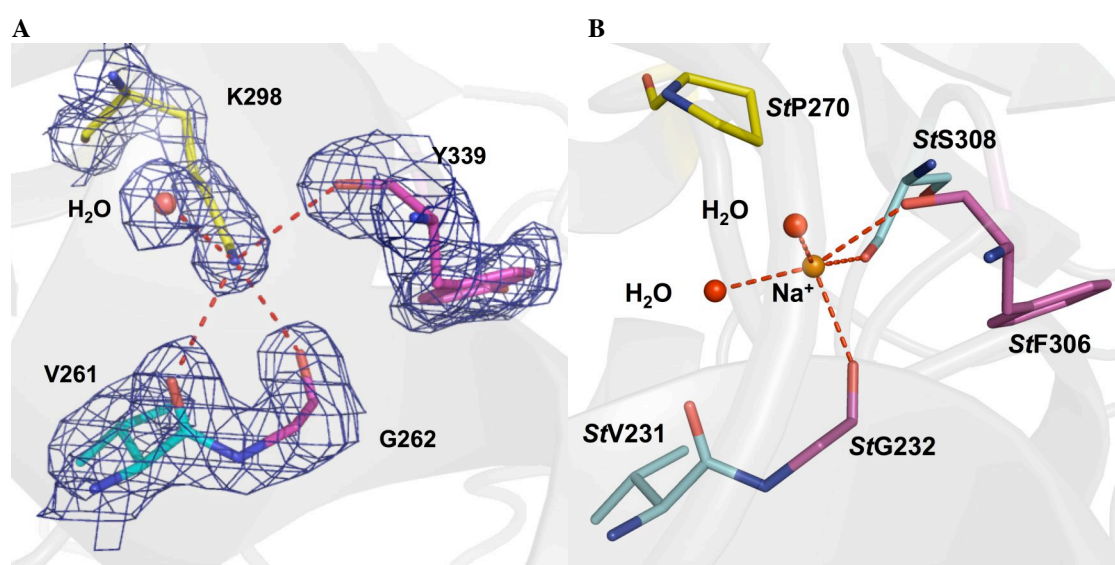
Figure 2.27 The coordination sphere of *Ss*Lys298 and the cation binding-site of *St*TrpB1.
Residues corresponding to structurally equivalent positions in both coordination spheres are shown in magenta and residues only present in one sphere in cyan, respectively. The tetragonal plane of the coordination sphere is shifted between *Ss*TrpB2b and *St*TrpB1. Water molecules are shown as red spheres. (**A**) The coordination sphere of *Ss*Lys298 (yellow) and the corresponding electron density (dark blue, contoured at 1 σ). (**B**) Na$^+$ (orange) coordination in *St*TrpB1. Pro270, which is structurally equivalent to *Ss*Lys298, has been added. The atoms forming the coordination sphere are connected with red dashes.

Due to the lacking sodium ion, the E-AA intermediate and subsequently the closed conformation of *Ss*TrpB2b cannot be stabilized in the same way as in TrpB1. Further, the region flanked by the N- and C-terminal residues coordinating *Ss*Lys298 reveals significant structural differences between TrpB2 and TrpB1 members (Figure 2.28). Structure based sequence alignments (Figure 6.2 in appendix) suggest that these differences constitute a second structural feature separating TrpB2 and TrpB1 family members.
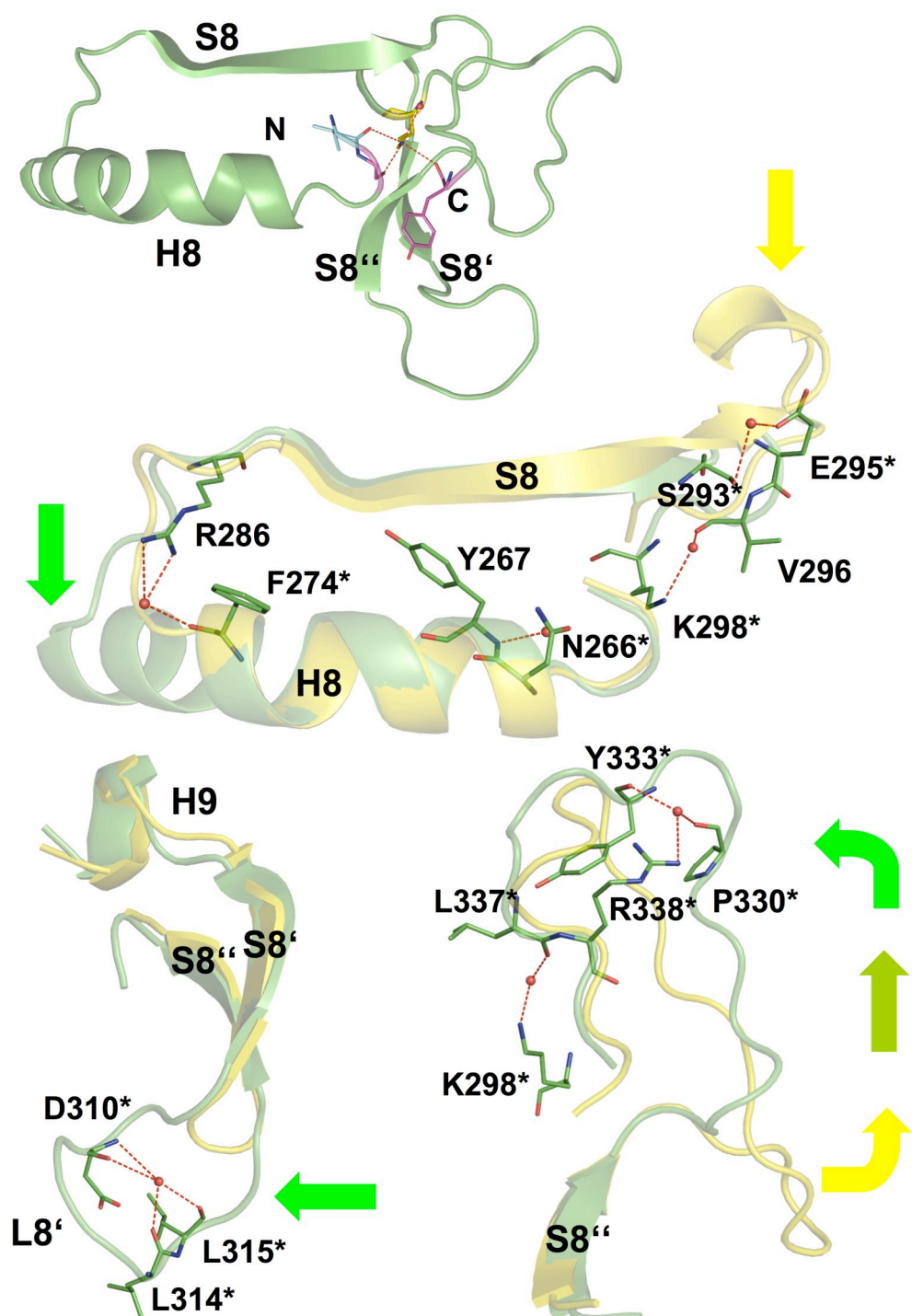
Figure 2.28 Comparison of the cation binding-site regions of *Ss*TrpB2b and *St*TrpB1.
(**A**) Overview of the cation binding site of *Ss*TrpB2b (green; also for B, C and D). The N- and C-terminal residues of the region are indicated with N and C, respectively. (**B** to **D**) The cation binding-site region of *Ss*TrpB2b superposed to the corresponding region in *St*TrpB1 (yellow). For a better overview the region has been split into three fragments. Residues forming water-bridged interactions are labeled (asterisks indicate strict conservation between *Ss*TrpB2b and *Ss*TrpB2a) and shown as sticks. Green and yellow arrows indicate insertions in TrpB2 and TrpB1 family members, respectively. Secondary structure elements are labeled. (**B**) Segment representing the N-terminal moiety of the cation binding-site region. The α-helix H8 shows variable length in both, TrpB2 and TrpB1 family members. The water coordinating Lys298 forms a bridge to Val296. (**C**) The loop L8' is significantly elongated in *Ss*TrpB2b. Residues of L8' coordinate a water molecule, which should further stabilize this segment. (**D**) The segment downstream to S8'' reveals a structural shift of a loop segment.

There are two insertions in the structure of *Ss*TrpB2b that may affect its functionality: the structurally altered loop segment downstream of S8'' and the elongation of L8'. The loop segment downstream of S8'' carries a residue (*St*Asp305, equivalent to *Ss*Arg338), which is involved in allosteric communication in *St*TrpB1. Due to the interaction pattern of *Ss*Arg338, this should not be able to participate in a TrpB1-analogous allosteric communication. Additionally, a putative acceptor residue of the allosteric communication (*St*Lys167) would be a Gly in *Ss*TrpB2b, which lacks a side chain critical to this communication. The residues downstream of *Ss*Arg338 form interactions to L8'. Thus, an element involved in allosteric communication in TrpB1 family members is missing in *Ss*TrpB2b, which should break the allosteric pathway between the active sites of TrpA and TrpB2 family members.

The insertion in L8' is equivalent to a plug in the tunnel entrance, since gating residues in TrpB1 members are located there. The residues *Ss*Asp307 and *Ss*Tyr309 (equivalent to the gating residues *St*Tyr279 and *St*Phe280) are in different conformations in *Ss*TrpB2b and *St*TrpB1. Instead of an interaction pattern between *St*Tyr279 in L8' and H6, *Ss*Asp307 (strictly conserved in TrpB2 family members) forms an interaction to the strictly conserved *Ss*Thr322. The main-chain of *Ss*Tyr309 interacts to the strictly conserved *Ss*His340. Thus, an interaction pattern critical to allosteric communication in *St*TrpB1 cannot be formed and the conformation of L8' should be decoupled from COMM subdomain movement analogous to TrpB1. The high conservation of L8' in TrpB2 family members suggests *Ss*Lys313 to be a candidate involved in the different association behaviour of *Ss*TrpB2b and *Ss*TrpB2a, if this structural element is involved. In *Ss*TrpB2a, a Gly in the position equivalent to *Ss*Lys313 should give L8' higher conformational freedom and minimize volume requirements in this position. Additional flexibility may be critical to an alternative tunnel opening mechanism, which is likely to happen in *Ss*TrpB2a due to decoupling of L8' and H6.

2.5.5c The insertion helix H5'

The insertion helix H5' is vicinal of the hypothetical TrpA interface of *Ss*TrpB2b. It is present in all members of the TrpB2 family. H5' forms exclusively interactions to H6 in the COMM domain. Thus, it shields H6, which is involved in allosteric communication, from accessibility. The interaction between H5' and H6 is

mostly hydrophobic. Only one H-bond is formed by the main chain of H5' Glu214 (strictly conserved in TrpB2 family members). The helix has a putative intrinsic mobility, suggested by the high B-factors of the crystallographic model. Due to this mobility, H5' may undergo structural displacement during catalysis, with subsequent solvent exposure of H6. The low conservation of H5', together with its putative mobility, makes it hard to estimate its role in the association behavior of *Ss*TrpB2b/*Ss*TrpB2a to TrpA. However, it is likely that in *Ss*TrpB2b H5' blocks allosteric communication (if analogous to TrpB1), since the accessibility of H6 is critical to this process.
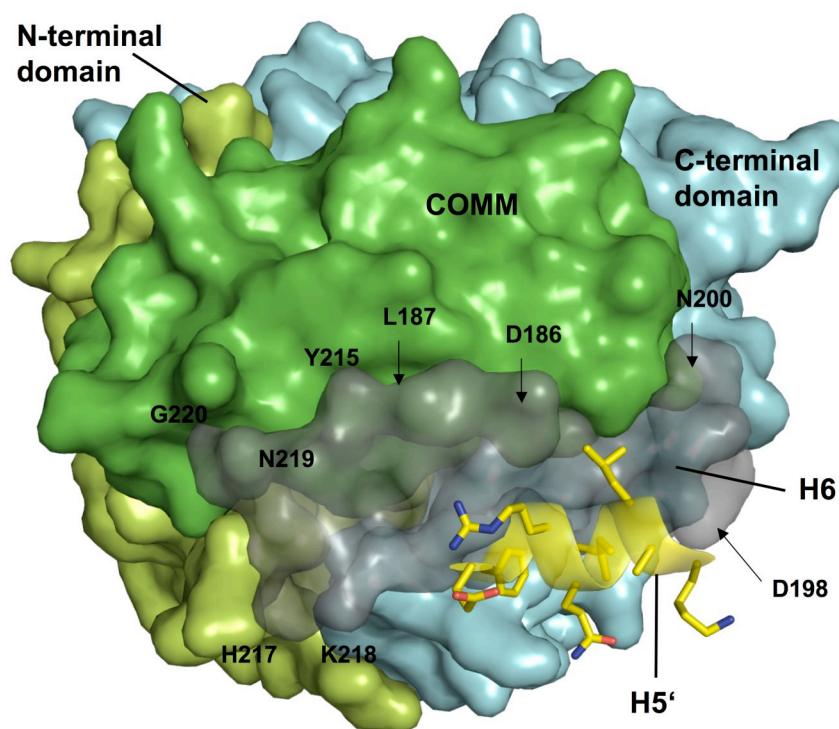
Figure 2.29 The insertion helix H5' (displayed as sticks) and its environment.

Three regions with altered structure cluster in the hypothetical TrpA interface of *Ss*TrpB2b: L8', H5' and the N-terminal extension (borders of the disordered segment). It is likely, that the composition of these elements dictates the association behavior of *Ss*TrpA and *Ss*TrpB2b (lack of complex formation) as well as *Ss*TrpB2a (hetero-complex formation).

2.5.6   The hetero-complex between *Ss*TrpB2a and *Ss*TrpA

The activation of the α-reaction of *Ss*TrpA by *Ss*TrpB2a constitutes indirect evidence for hetero-complex formation during catalysis. It seems that *Ss*TrpB2a and/or *Ss*TrpA create a conformational state during catalysis, which is capable to form the binding interface. The catalysis of *St*TrpB1 and *St*TrpA has been investigated intensively on a structural level over the past two decades (Figure 2.30).
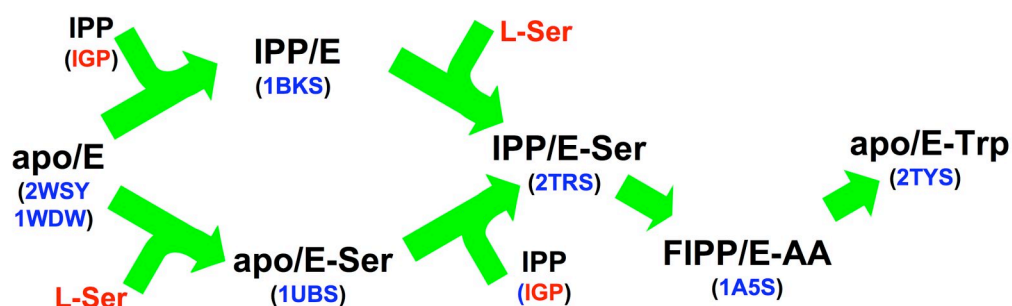
Figure 2.30 States of catalysis for which structural information exists for TrpB1/TrpA.
All structures are from *S. typhimurium* TrpB1/TrpA (except 1WDW from *P. furiosus*). Active site complexes are shown in black (α-active site/β-active site) with PDB codes below (blue). Natural substrates entering the active sites during catalysis in red (used analogues are colored in black, abbreviations are given in Table 2.4).

2.5.6a *Homology models*

Crystal structures of *St*TrpB1 at different points during catalytic were used to produce homology models of active states of *Ss*TrpB2b. In catalysis, the COMM subdomain moves as a rigid body, while the rest of the enzyme remains unchanged (*rmsd* = 0.24 Å, Schneider *et al.*, 1998). Due to this and the significantly altered tertiary structure between the TrpB1 and TrpB2 family, homology models were produced by splitting *Ss*TrpB2b into COMM subdomain and remainder of the protein. These two segments were independently superposed to the corresponding regions in TrpB1 (Table 2.19). To produce the final models, the superposed segments were subsequently idealized with REFMAC (Murshudov *et al.*, 1997; Figure 2.31A).

Table 2.19 Statistics of superposition of the COMM domains of *Ss*TrpB2b and *St*TrpB1.

Template	*rmsd* [Å] (COMM)	Over C_α (COMM)	*rmsd* [Å] (rest[1])	Over C_α (rest[1])	Homology model
1UBS	0.9	91	1.1	196	*Ss*TrpB2b_ESer
2TYS	1.0	89	1.1	195	*Ss*TrpB2b_ETrp

[1]rest defines *Ss*TrpB2b with COMM domain removed, namely residues 4-115 and 227-428

However, clashes in *Ss*TrpB2b models (prior to idealization) involved residues upstream to *Ss*H1, the active site loop (N-terminal COMM subdomain border), *Ss*H6, the C-terminal COMM subdomain border and L8' indicating that these regions may have to undergo conformational changes during catalysis. Despite, the external aldimines (E-Ser and E-Trp) could be modeled within the active sites of the homology models. The resulting complexes revealed protein:ligand interactions analogous to those in TrpB1, indicating a good structural correlation of the *Ss*TrpB2b_ESer and *Ss*TrpB2b_ETrp conformational states.

Due to the high sequence conservation between *Ss*TrpB2b and *Ss*TrpB2a, additional homology models (SWISS MODEL, Guex & Peitsch, 1997; Schwede *et al.*, 2003) could be generated for *Ss*TrpB2a (Figure 2.31B). Additionally, the homology model *Ss*TrpB2b_ESer was used as template for homology modeling (SWISS MODEL) of the corresponding catalytic state of *Ss*TrpB2a.
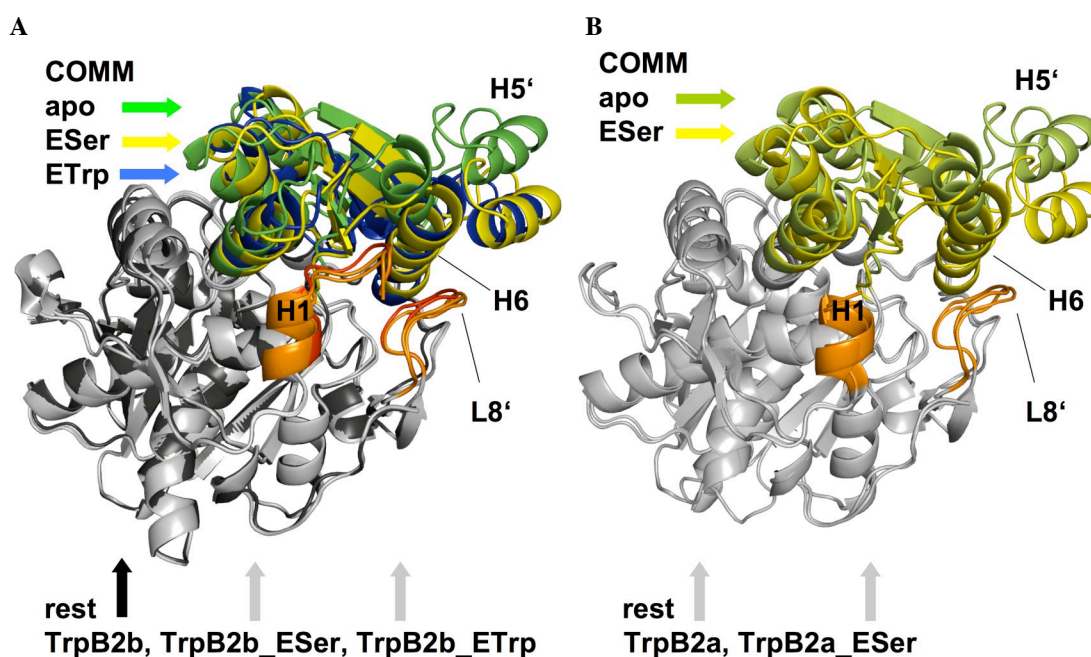
Figure 2.31 *Ss*TrpB2b and homology models.
Crystal structure and homology models of *Ss*TrpB2b (**A**) and homology models of *Ss*TrpB2a (**B**). The COMM subdomain movement is best visible in H5' (labeled). H1 and its upstream region as well as L8' (both orange and labeled) move away from H6 (labeled) due to the COMM movement.

The homology model *Ss*TrpB2a is essentially identical to *Ss*TrpB2b (*rmsd* = 0.1 Å over 398 C$_\alpha$-atoms using TOP). The same is valid for the superposition of *Ss*TrpB2b_ESer and *Ss*TrpB2a_ESer (*rmsd* = 0.2 Å over 399 C$_\alpha$-atoms using TOP). The corresponding structure based sequence alignments are analogous to Figure 2.25, while H1 is also present in both *Ss*TrpB2b and *Ss*TrpB2a .

2.5.6b   The interface between TrpB1/TrpB2 and TrpA

A detailed description of interfaces in the Tryptophan Synthases of *P. furiosus* and *S. typhimurium* is not available in the bibliography. A single comparison of the homo-obligomer interface has been reported that focuses on the stabilization energy of H-bonds (Hioki *et al.,* 2004). Hence, the hetero-obligomer interface between TrpA and TrpB1 was analyzed here using the program LIGPLOT (Wallace *et al.,* 1995). A structure-based sequence alignment (Swiss PDB Viewer; statistics in Table 2.20) was produced to correlate involved segments between the TrpB family members. This revealed that segments with obligatory interactions in TrpB1 belong to the N-terminal extension and the cation binding-site region in *Ss*TrpB2b. Due to the COMM subdomain movement in TrpB1, this element is only temporarily involved in interactions to TrpA with H6, which is shielded by H5' in *Ss*TrpB2b/*Ss*TrpB2a (Figure 6.2 in Appendix).

Table 2.20 Superposition of *Ss*TrpB2b and *Pf*TrbpB1 (1WDW) as well as *St*TrpB1 (rest).

PDB code	*rmsd* [Å]	Over C$_\alpha$-atoms
1WDW	1.34	291
2WSY	1.46	263
1BKS	1.42	272
1UBS	1.44	261
2TRS	1.40	248
1A5S	1.46	259
2TYS	1.41	248

Analogous to TrpB1, TrpA structures were superposed using Swiss PDB Viewer (statistics in Table 2.21) and a structure-based sequence alignment of TrpA representatives (MyHits, Figure 6.3 in appendix) created. In TrpA, residues forming

interactions to TrpB1 proteins cluster in three regions involving the active site loop L2 and its downstream region (H2') as well as the α-helices H3 and H4. *Ss*TrpA reveals a significantly altered L2 sequence and the truncation of H2', which seems to be a feature unique to *Ss*TrpA. This may affect the catalytic efficiency of *Ss*TrpA compared to other TrpA proteins. The loop L6 has a variable length between different organisms. L6 is the initiation point of allosteric communication and has a supportive role in TrpA catalysis (stabilization of L2 and closure of the active site). *Ss*TrpA has the shortest L6 indicating that both, allosteric communication and L2 stabilization may be affected.

Table 2.21 Statistics of the superposition of *Pf*TrpA (1WDW) to TrpAs of *S. typhimurium*.

PDB code	*rmsd* [Å]	Over C_α-atoms
2WSY	1.17	204
1BKS	1.18	216
1UBS	1.19	218
2TRS	1.25	211
1A5S	1.19	215
2TYS	1.24	213

2.5.6c *Putative role of SsTrpB2b segments in the hetero-complex interface*

It can be expected that the relative position of the TrpA interface in *Ss*TrpB2a has been retained and evolutionary shaping has optimized the interaction pattern in the hetero-complex towards a hetero-obligomer interface. However, the TrpA/TrpB interface reveals various differences between the TrpB1 and TrpB2 family, which are likely to affect the interaction pattern to TrpA. Due to the increased catalytic efficiency of *Ss*TrpA in presence of *Ss*TrpB2a (60-fold for *apo* and 270-fold for addition of L-Ser, Leopoldseder *et al.,* 2006) it is likely that conformational changes in *Ss*TrpB2a during catalysis create a surface with improved interactions to *Ss*TrpA.

The structures of *Ss*TrpB2a_ESer and *St*TrpB1 (2TRS, rmsd = 1.5 Å over 283 C_α-atoms) have been superposed and the coordinates of *St*TrpA (with a closed active site) extracted. Comparison of *St*TrpA and *Ss*TrpB2b/*Ss*TrpB2a reveals H5', H6 and L8' as structural elements with a putative role in the hetero-complex interface (Figure 2.32). It is likely, that H1 and the disordered segment upstream of H1 are also

involved in this interface as indicated by a *Ss*TrpB2a variant in the studies of Leopoldseder (2005). Additionally, this showed lower activation of *Ss*TrpA by a second *Ss*TrpB2 variant with "switched" structural residues in H8 (variants indicated in Figure 2.25), which is distant to the interface. This modification is likely to affect the structure of H1 and its upstream region. This, together with structure based sequence alignments of TrpA proteins and interactions across the hetero-obligomer interface of *St*TrpS, can be used to create a hypothetical model of the role of TrpB2 elements in hetero-complex formation and allosteric communication:

(i)   L8' in *Ss*TrpB2a stabilizes the active site loop L2 of *Ss*TrpA. In *Ss*TrpB2b, this loop cannot interact with L2 because of steric hindrance. The best candidate is the side-chain of *Ss*Lys313 in the core of the interface, which is equivalent to a Gly in *Ss*TrpB2a. This would explain the moderate increase in catalytic efficiency of *Ss*TrpA upon addition of *Ss*TrpB2a. This stabilization may be provided by an induced fit conformational change of L8', which might open the tunnel entrance.

(ii) The mobile α-helix H5' in *Ss*TrpB2a provides additional stabilization to *Ss*TrpA L2. In *Ss*TrpA, H2' (downstream of L2) is truncated but H5' in *Ss*TrpB2a may fulfill the role of the truncated segment instead. The E-Ser conformational state moves H5' closer to *Ss*TrpA which is likely to provide further stabilization to *Ss*TrpA L2. However, this could be an element stabilizing the hetero-complex interface without increasing catalytic activity (as indicated by the switch of H5' between TrpB2b and TrpB2a; Leopoldseder, 2005)[1]. The COMM domain adopts semi-closed conformation, L8' has to obey this movement by a conformational change favoring interactions to *Ss*TrpA. Significantly altered sequences of H5' between *Ss*TrpB2a and *Ss*TrpB2b are likely to affect a possible interaction of this helix to *Ss*TrpA.

(iii) Analogous to TrpB1, H6 interacts with *Ss*TrpA L6 providing stabilization.

(iv) H1 and the disordered segment upstream are in the vicinity of *Ss*TrpA, which makes the involvement of this segment likely (analogous to the involvement of the N-terminus of *St*TrpB1 in the hetero-obligomer interface).

---

[1] If this stabilization would be provided by the conserved N-terminal moiety of H5' TrpB2 variants would have no effect on interface formation/activation.
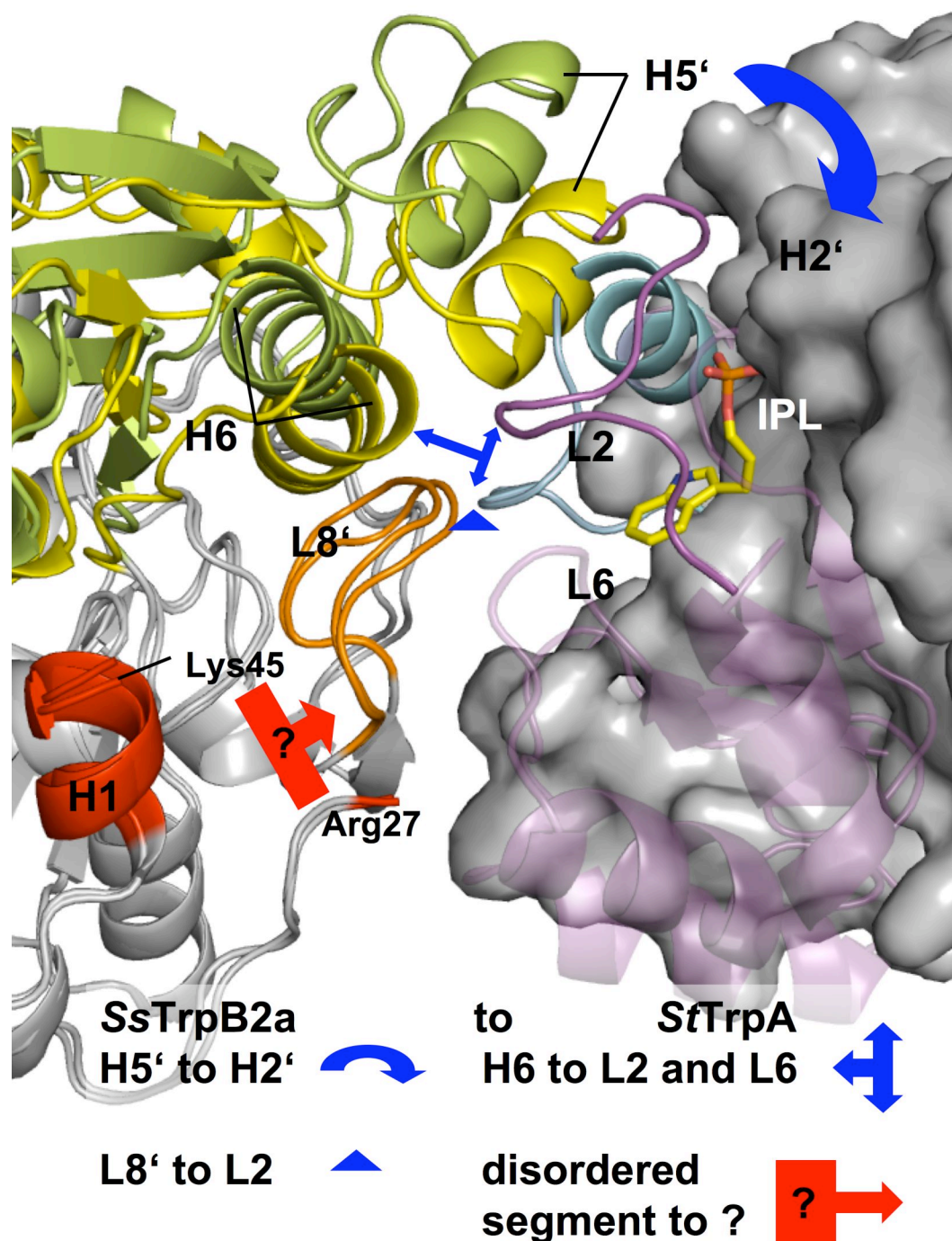
Figure 2.32 Hypothetical hetero-complex interface between *Ss*TrpB2a and *St*TrpA.
*Ss*TrpB2a is coloured as in Figure 2.31, while H1 and the first residue upstream to the disordered segment are coloured red. Different signs indicate structural elements in *Ss*TrpB2a and *St*TrpA (L2 and its downstream region shown solid, cyan; L6 solid, magenta). The remainder of *St*TrpA is shown in gray surface representation as well as transparent magenta cartoon representation (transparent to show the active site). The disordered segment in *Ss*TrpB2a may interact with the region shown as transparent magenta cartoon. The substrate analogue (IPL, yellow) is shown in the active site of *St*TrpA.

The differences in structural elements with putative involvement in TrpA binding between *Ss*TrpB2a and *Ss*TrpB2b are likely to be the determinant of hetero-complexation. In L8' (*Ss*TrpB2b), Lys313 (Gly in *Ss*TrpB2a) is a residue that

is underrepresented in protein-protein interfaces in general. In *Ss*TrpB2a, a Tyr in H5' is an interface hot spot residue, while in *Ss*TrpB2b a Phe is in the equivalent position. The disordered segment carries several hot spot residues in *Ss*TrpB2a (two Arg and one Tyr), which are missing in *Ss*TrpB2b. This segment may be able to form interactions to *Ss*TrpA providing additional stabilization to the interface. Alternatively, the truncation of H8 could allow different packing of H1 and its upstream region including the disordered segment.

2.5.6d   Allosteric communication

The suggested path of mutual allosteric communication between the active sites of TrpA and TrpB1 involves regions that have been altered across TrpB1 and TrpB2 family members (N-terminal extension, cation binding site region, insertion helix H5' and the active site). This path shows several potential point breaks:

(i)   The residue *St*Asn171 (strictly conserved in the B1 family) is equivalent to *Ss*Ser210 (strictly conserved in B2 members). The gating residues, which interact with this Asn in TrpB1, adopt different conformations and are shielded by L8' in TrpB2 family members.

(ii) Truncation of H6 in the COMM subdomain of TrpB2 enzymes could prevent an interaction pattern as that of TrpB1 (H6 in TrpB1 interacts with L6 in TrpA). Additionally, this loop is truncated in *Ss*TrpA.

(iii) The interaction of βH6 to the gating residues is not present in *Ss*TrpB2b and can subsequently not be changed. For the formation of these interactions, L8' would have to undergo conformational changes

(iv) *Ss*Arg339 adopts a different conformation than the equivalent *St*Asp305 (neither "swing-in" nor "swing-out" conformation).

A mutual allosteric communication path between *Ss*TrpB2a and *Ss*TrpA cannot be formed since several critical elements are either not present or shielded from the required interaction. However, *Ss*TrpB2a shows a higher activation of the α-reaction (between 60 and 300-fold) than *St*TrpB1 (25-fold; Miles, 2001). Since the α-reaction is the rate-limiting step in the αβ-reaction of *S. solfataricus*, *Ss*TrpB2a also

benefits from increased indole availability. Another role of the allosteric communication is to prevent the loss of indole during channelling between active sites. The lack of the cation decreases the stability of the E-AA state, which has higher affinity for indole. An indole molecule in the tunnel may get lost if *Ss*TrpA dissociated from *Ss*TrpB2a (because of a conformational change induced by reaction of E-AA in E-Ser). This emphasizes a role of *Ss*TrpB2b as an indole salvage protein in a system, which is in a nascent phase of evolution towards formation of protein-protein interfaces.

2.5.7   Summary and conclusions

2.5.7a Summary

During structure elucidation it became obvious that *Ss*TrpB2b and TrpB1 family members comprise a common core and differ mainly in their interface to TrpA. Structural features of *Ss*TrpB2b have been analyzed and set into the context of the TrpB2 and TrpB1 families. The extensive studies of TrpB1 family members over the last two decades made it possible to model conformational states of *Ss*TrpB2b during catalysis. The high sequence conservation between *Ss*TrpB2b and *Ss*TrpB2a allowed the production of homology models to draw conclusion on the molecular determinants of the complexation behavior of *Ss*TrpB2b and *Ss*TrpB2a.

The elucidation of the structure of TrpB2b from *Sulfolobus solfataricus* (non-operon located *trp*B2 transcript) gave detailed insight into similarities and differences between the TrpB1 and TrpB2 families. The domain organization is highly similar and the active site composition indicates identical catalytic properties in both families. Differences between these families include three regions (N-terminal extension, insertion helix H5' and the cation binding-site region) clustering in the hypothetical hetero-complex interface of TrpB2 family members. The first residues of the N-terminal extension of *Ss*TrpB2b fold back onto the dimer interface and contribute to the central β-sheet in the C-terminal domain. This segment represents a polar extension to otherwise predominantly hydrophobic homo-obligomer interfaces in TrpB1 and TrpB2 enzymes. The monovalent cation is not present in TrpB2 members and is replaced by the side-chain of a Lys residue (*Ss*Lys298), which may

stabilize the corresponding region. The insertion helix (H5'), exclusively found in TrpB2, shields H6 (an element involved in allosteric communication in the TrpB1 family) from access.

Homology models suggest a *Ss*TrpA:*Ss*TrpB2a interaction pattern, which is similar to that of TrpB1:TrpA. The key to activation of *Ss*TrpA seems to lie on the contributions of H5' and L8, elements likely to prevent the formation of a *Ss*TrpB2b:*Ss*TrpA hetero-complex. The allosteric communication path is potentially disrupted as several points in TrpB2 family members, which results in monodirectional activation of TrpA.

2.5.7b Conclusion

The model system *Ss*TrpB2 is an ancestral initiation point of TrpB protein evolution, likely to correspond to a nascent phase. Oligomerization is "tested" by formation of a hetero-complex between *Ss*TrpB2a and *Ss*TrpA, while a second copy (*Ss*TrpB2b) guarantees retained functionality (catalysis). The evolution towards modern hetero-obligomers seems to arise from altered stabilization (by fixation of the N-termini) and enhanced functionality (allosteric communication). The fixation of the N-terminus in the homo-obligomer interface has become obsolete upon formation of hetero-obligomers. In this context, an initial evolutionary step towards reorganization of the N-terminus seems to be the truncation of H8 in *Ss*TrpB2a, which may have allowed different packing of H1 and its upstream region with subsequent minimization of sterical clashes. The enhanced catalytic efficiency of *Ss*TrpA favours the evolutionary pathway from the observed dimer over hetero-complexation towards formation of a hetero-obligomer. Subsequently, sterical clashes in the hetero-complex interface between *Ss*TrpA and *Ss*TrpB2a have been minimized to stabilize it.

An initial stability gain may have loosened evolutionary pressure on the cation binding-site. The obligate interaction between TrpB1 and TrpA also loosens evolutionary pressure on the length of L8' since indole cannot be lost due to complex dissociation and the role of gating residues can then be established. A final step in the evolution of mutual allosteric communication could have been the complementary shaping of helix H6, which is the central element in the path. Even "shifting" of structural elements between TrpB2 and TrpA (*e.g.* H5' in *Ss*TrpB2b may have

become H2' in TrpA) seems to be involved in the evolution from two independent enzymes towards a multienzyme complex.

After the completion of this work, the structure of SsTrpA was completed in our lab. This will allow now furthering the evolutionary analysis of hetero-complexation in this group of enzymes.

3 A77-A78: insights into the A-band organization in human muscle

The sarcomere – contractile unit of striated muscle – is one of the most organized and highly synchronized bio-machinery known to date. The titin filament plays multiple roles both in its complex function and ordered ultrastructure. In particular, titin is believed to act as a scaffold or template for the assembly of multiple other sarcomeric proteins, supporting numerous protein-protein interactions. For example, the poly-FnIII super-repeat structure of titin's A-band region is thought to serve as an assembly matrix for the myosin-based thick filament. In agreement with this, the periodicity of the latter matches the periodicity of long super-repeats in titin. To date, no structural information exists on the conformational arrangement of domains within the poly-FNIII arrays of A-band titin. To this effect, we have elucidated the crystal structure of the two-FnIII tandem A77-A78, which provides an insight into the intramolecular FnIII-FnIII interfaces of titin and their potential involvement in thick filament assembly.
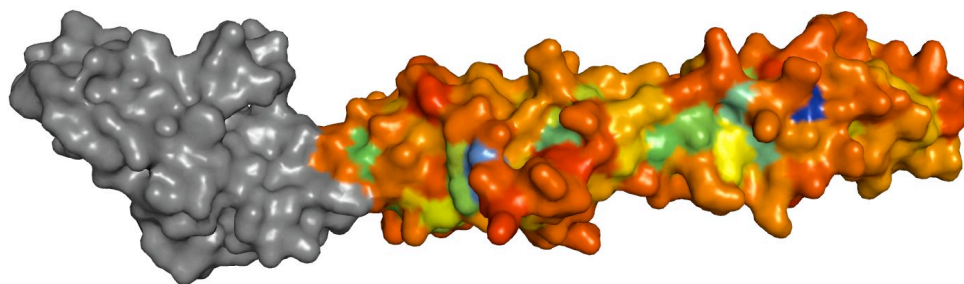
3.1 Abstract

Titin is a gigantic protein that spans half of the sarcomere. Its A-band fraction is believed to act as a matrix for myosin filament assembly. This part of the titin chain is composed of fibronectin type III (FNIII) and Immunoglobulin-like (IG) modules, organized in two types of super-repeats containing seven and eleven domains (short and long type, respectively). To date, there are no atomic models of FNIII tandems from titin and subsequently the architecture of its A-band region remains elusive.

We elucidated the first structure of an intracellular FNIII tandem, A77-A78 from a long super-repeat of human titin, using X-ray crystallography. A77-A78 correspond to positions two and three of the fourth super-repeat tandem. A77-A78 adopts a semi-extended conformation, defined by an unusually polar domain interface crowded with tight H-bond contacts. This suggests that the structure of this tandem is well-defined and stiff. The modular interface is determined by the properties of end loop clusters as well as by linker composition, which are mostly conserved across equivalent domain positions across super-repeats. In titin's FNIII transitions, A77-A78 seems to represent the tightest transition with the lowest conformational freedom. In other titin FNIII tandems, longer linkers and a higher hydrophobic content of the interface seems to allow them higher conformational freedom. This suggests a super-repeat dependent rigidity of FNIIIs transitions, which creates tightly associated tandems with a possible myosin assembly matrix function as well as transitions that may act as hinges for fine-tuning of this assembly.

Based on the recent structure of A168-A170 (IG-IG-FNIII) from the M-line of titin, a homology model of the tandem A76-A78 was produced. This show a zig-zag arrangement where highly conserved patches co-align along the filament in the filament surface with the best accessibility. This leads to speculate on the potential role of these regions in higher assembly within the sarcomere.

3.2 Introduction

3.2.1 The sarcomere and its major components

Muscle ultrastructure is characterized by a complex arrangement of multi-protein arrays. The sarcomere (Figure 3.1) is the basic contractile unit of striated muscle, being the smallest self-contained functional unit of myofibrils. Sarcomeric features resolved by electron microscopy (EM) are the transverse thin and thick filaments and the longitudinal Z- and M-lines (the lines are also referred to as discs or bands) as well as the P-zone.
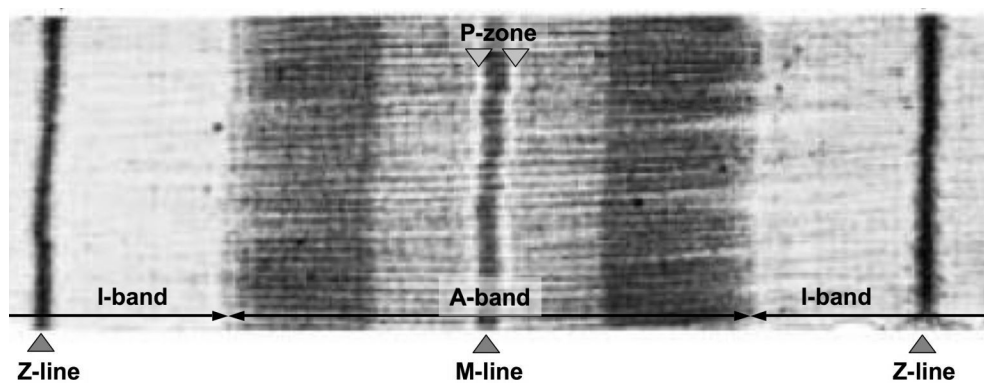
Figure 3.1 The sarcomere.
The sarcomere comprises the region from Z-line to Z-line (indicated by vertical arrows). For a description of the regions in the sarcomere see upcoming chapters. Image adapted from Stryer, 1975, Biochemistry.

The I-band includes the thin filament and the Z-disc. The thin filament – which is actin-based - is anchored in the Z-disc and interacts with the thick, myosin-based filament during active force generation according to the "sliding filament" model (Huxley & Niedergerke, 1954) following the "Lymn-Taylor-cycle" (Lymn & Taylor, 1971) of ATP hydrolysis by myosin during contraction. The A-band comprises the area where thick and thin filaments interdigitate. Its central region is occupied only by the thick filament, with the M-line denoting the geometrical center of the sarcomere where the thick filament switches its polarity, so that half sarcomeres are in fact repeats with opposite direction.

3.2.2 Myosin and the thick filament

The central component of the thick filament is myosin, which is composed of a light chain (rod domain, oligomerization motif) and a heavy chain (head domain, active force generation). Myosin oligomerizes over its rod domains from initial dimeric building blocks to a bipolar filament (parallel arrangement in each half sarcomere). Each thick filament spans the whole A-band region and switches its polarity at the M-line, where myosin molecules are arranged antiparallel. This creates a region free of myosin heads, the P-zone (also referred to as bare-zone), where various other proteins interconnect the filaments into hexagonal arrangements. Low-angle X-ray diffraction (Huxley & Brown, 1967) and EM (Luther *et al.*, 1981) indicated periodicity of the myosin heads in the thick filament of vertebrates, where 6 myosin molecules (3 dimers[1]) form a crown with 429 Å periodicity and 143 Å sub-periodicity resulting in three-fold rotational symmetry along the filament (Figure 3.2A). The spacing between the crowns shows periodic perturbation within a 429 Å repeat (Cantino *et al.*, 2000).

The second principal component of the thick filament is Myosin binding protein-C (MyBP-C, also called C-Protein), exclusively found in the C-zone of the A-band. There it binds myosin and titin's IG domains at intervals of myosin crown periodicity (Rome *et al.,* 1973). Roles in filament assembly during myofibrillogenesis and the regulation of contraction have been proposed (Kenny *et al.,* 1999 and references therein). MyBP-C is largely composed of domains with IG and FNIII fold and additional unique sequences. The last four domains (C7-C10) show a pattern FNIII-IG-FNIII-IG and are associated with titin and myosin binding (Freiburg & Gautel, 1996) (Figure 3.2B). C10 is the main myosin binding-site while C7-C9 are supposed to give MyBP-C its proper location by interactions with myosin as well as each first IG domain of titin's long super-repeat[2] (MyBP-C domains C8-C10).

---

[1] 4 dimers per crown in invertebrates and 6 dimers in scallop; same periodicity
[2] the composition of titin's super-repeats is shown in section 3.2.3a
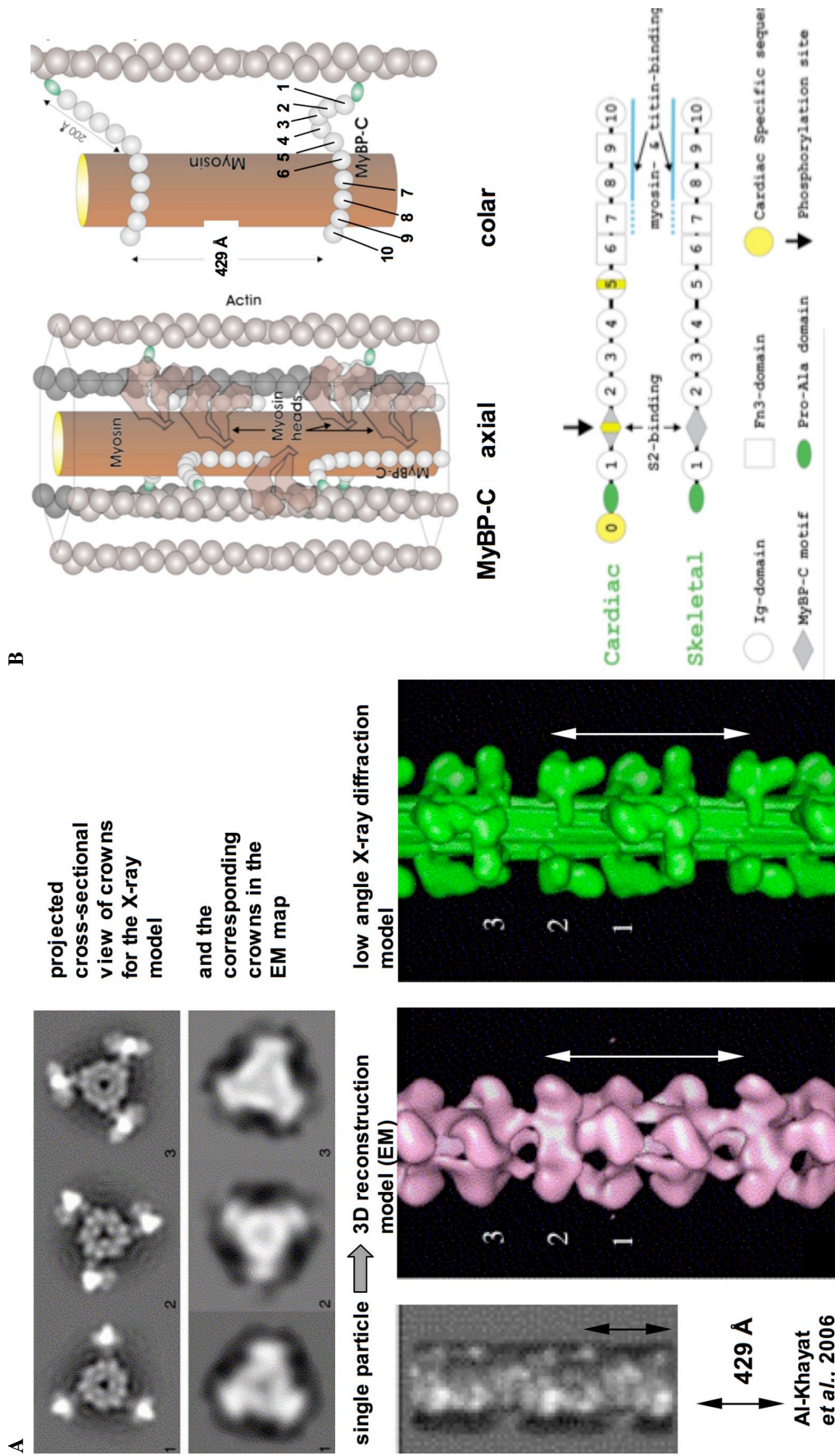
Figure 3.2 Vertebrate myosin and MyBP-C.

Figure 3.2 Vertebrate myosin and MyBP-C.
(**A**) Myosin models from low angle X-ray diffraction and electron microscopy (EM) single particle 3D reconstruction. The cross sectional view (top) of X-ray and EM models showing three symmetry-related crowns (sub-periodicity 143 Å) to result in a global periodicity of 429 Å (both indicated in the bottom panel). (**B**) The domain composition of myosin binding protein C (MyBP-C; lower panel) and the interaction models between myosin and MyBP-C. Rome *et al.* (1973) suggested that three MyBP-C molecules arrange in collar fashion around the myosin filament (domains C5 to C10 of three chains wrap around the myosin filament) while domains C1 to C4 point away from myosin. The second model by Squire & Knupp (2005) shows an axial orientation of MyBP-C along the myosin filament. The three C-terminal domains (C8-C10) interact with the first IG modules in titin's long super-repeat (Freiburg & Gautel, 1996). MyBP-C molecules can also interact with each other (over C7 and C10 as well as C5 and C8; Moolman-Smook *et al.*, 2002; Squire *et al.*, 2003).

The A-band region of titin constitutes an additional component of the thick filament, which has been proposed to act as a molecular ruler and template in thick filament assembly (Whiting *et al.,* 1989). This involves six titin molecules per thick filament in each half of the sarcomere (estimated from mass measurements of the titin content of myosin; Liversage *et al.,* 2001). The mechanism underlying myosin assembly *in vivo* is yet to be understood.

3.2.3 Titin

The gigantic protein titin (also known as connectin) spans half of the sarcomere ($>1.3\ \mu$m). It is anchored to the Z-disc, spans the I-band, penetrates the A-band and attaches to an adjacent titin molecule in the M-line. With up to 3.7 MDa it is the largest protein known to date (Maruyama *et al.,* 1977; Wang *et al.,* 1979). The building blocks of titin were initially identified in 1995 (Labeit and Kolmerer, 1995), when sequence data first became available, and were completed in 2001 once the domain composition of the different isoforms was established (Bang et al, 2001). This revealed that most of the 38138 amino acids encoded by the human titin gene fold into immunoglobulin (Ig) and fibronectin-3 (Fn3) domains, which add up to ~300 modules in the larger isoforms (approx. 90% of its molecular mass). They form mostly linear tandems distributed along the length of the filament. Other important components of titin include the so-called PEVK region in the I-band (rich in proline, glutamate, valine and lysine residues), unique sequences - yet to be characterized - and one single catalytic domain, a kinase (TK), located at the M-line. TK is related to myosin light chain kinase (MLCK)-like Ser/Thr kinase and phosphorylates telethonin in developing muscle (Mayans *et al.*, 1998). This phosphorylation is thought to be important for initiating or regulating myofibrillogenesis. Titin plays different roles

across the diverse functional regions of the sarcomere, participating in sarcomere assembly and myofibrillogenesis, maintenance of ultrastructure, elasticity, myofibril turn-over and mechanotransduction in muscle (for reviews see Granzier & Labeit, 2005; Squire *et al.*, 2005).

3.2.3a Titin's A-band organization

Titin's role of molecular template and ruler is particularly pronounced in the A-band, where allegedly contribute to regulate the assembly of the myosin-based thick filament. A-band titin is composed of serially linked IG and FNIII domains, organized in super-repeats (Figure 3.3) composed of either seven (short repeat) or eleven (long repeat) domains. The short super-repeats consist of a [IG-(FNIII)$_2$-IG-(FNIII)$_3$] domain arrangement and are located in the D-zone (from distal to the M-line) at the tip of the myosin filament. The contiguous long super-repeats have a [IG-(FNIII)$_2$-IG-(FNIII)$_3$-IG-(FNIII)$_3$] composition and are located in the C-zone (from central in the A-band), which ends at the edge of the bare zone. Long repeats are believed to have evolved from the short form by gene duplication (Higgins *et al.,* 1994).
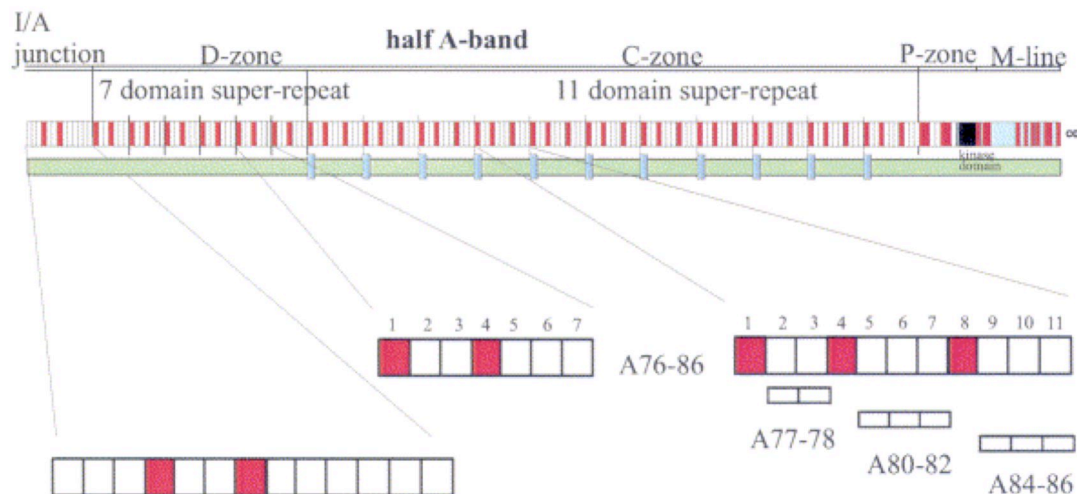
Figure 3.3 The modular organization of titin's A-band region.
Top panel: overview of titin's domain organization. IG-like modules are depicted as red boxes, FNIII-like modules are shown in white. The titin kinase domain is colored in black and unique sequences are colored in light blue. Underneath, the myosin filament is sketched in green with the attached MyBP-C shown as blue boxes. Bottom panel: blow-up of the I/A-junction, a short and a long (FNIII module numbering shown underneath) super-repeat domain (adapted from Muhle-Goll *et al.*, 2001).

Low-angle X-ray diffraction (Squire *et al.,* 2004) suggested periodicities of 40 Å for titin's modules along the myosin filament (38 Å optical diffraction; Hanson *et al.,* 1971) and subsequently the involvement of the long super-repeat[1] in myosin periodicity (freeze-fractions of muscle A-bands by Cantino *et al.,* 2002; low angle X-ray diffraction by Oshima *et al.,* 2003 and Squire *et al.,* 2005). In a hypothetical model by Squire and coworkers (2005; Figure 3.4), three equivalent titin strands (dimeric titin) arrange in either a coil or in zigzag around the myosin filament. Axial arrangement of titin on the myosin filament was suggested by Cantino and coworkers (2002). There is little evidence favoring one over the other arrangement.
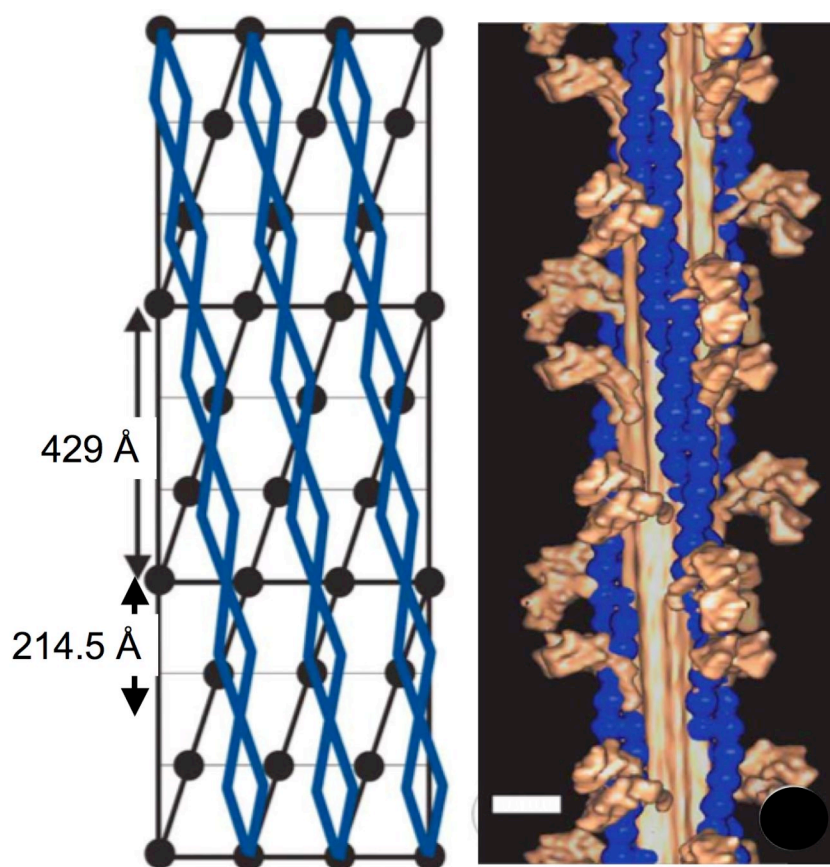
Figure 3.4 Hypothetical titin-myosin model.
Left: radial net of the myosin head positions on the surface of a thick filament viewed from outside the filament, where each black circle represents the location of a pair of myosin heads. The myosin head arrangement is represented as right-handed, three-start, nine head pair per turn helix of pitch 3x429 Å (true repeat 429 Å). On the surface of the thick filament six titin molecules (blue) three equivalent pairs of two titin molecules (blue; organized into three equivalent strands) intertwine to form a coil with a pitch of 214.5 Å. This titin arrangement possesses a 1287 Å true repeat and interactions with three successive 429 Å myosin repeats are not equivalent. Right: Three-dimensional representation of the thick filament with twisting titin strands whose radial net is depicted left. Adapted from Squire *et al.,* 2005.

---

[1] eleven domains in extended conformation span 418-440 Å

Binding studies (cosedimentation and solid phase binding assays, Muhle-Goll *et al.,* 2001) of FNIII tandems (A77-A78, A80-A82 and A84-A86) revealed binding of titin's FNIIIs to myosin, heavy meromyosin (globular heads with a section of the rods) and to sub-fragment S1 (myosin head domain). Reliable models of the organization in the myosin filament are still not available.

3.2.3b Titin's main building blocks

IG and FNIII modules (~100 aa length) belong to the superfamily of immunoglobulin-like proteins, which exhibit a two-layer β-sandwich architecture following a greek-key topology (CATH domain database; Figure 3.5).
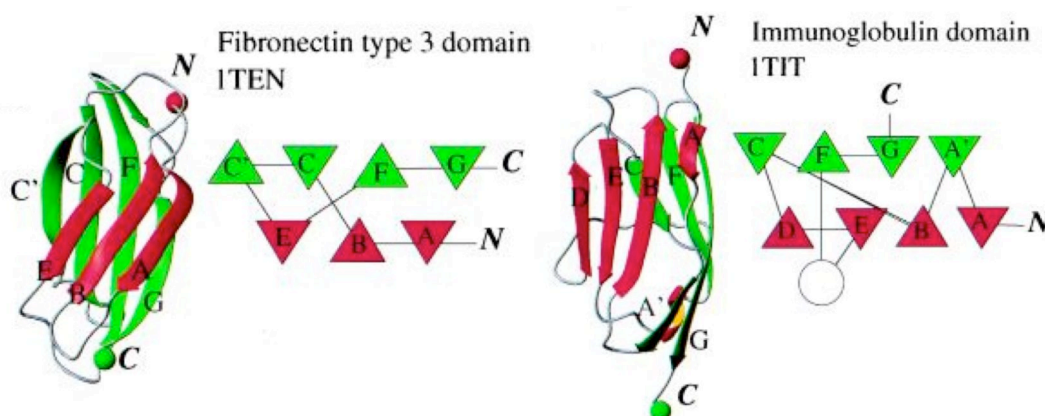
Figure 3.5 IG and FNIII modules and their topology.
The strands of the two β-sheets in FNIII (left) and IG (right) modules are shown in red and green. The topology of FNIII and IG domains differs mainly by a switch of a single β-strand between the sheets (Bork *et al.*, 1994), called C' in FNIII domains and D in IG domains (image adapted from Paci & Karplus, 2000).

Titin's IG modules have been classified to the intermediate (I) set of domains as they contain features intermediate between the variable (V) and constant (C1 and C2) set of immunoglobulin frames (Harpaz & Chothia, 1994). Titin's FNIII modules are found exclusively in the A-band region and are the only intracellular FNIII modules known to date.

There are features common to both IG (Improta *et al.*, 1996; Witt *et al.*, 1998) and FNIII (Muhle-Goll *et al.,* 1998; Muhle-Goll *et al.,* 2001) domains of titin. These features comprise a Pro-rich motif at the N-terminus, an Asn-x-x-Gly-motif in the FG β-hairpin and the fact that N- and C-termini are located in opposite ends of the modules. Additionally, there is a so-called Tyr-corner, characterized by the interaction

of the side-chain of a conserved Tyr residue (from strand E) to the main-chain of the i-5 residue (Hamill *et al.*, 2000). To date, there are only a few atomic models of modular components of titin (Table 3.1). Most models correspond to domains in isolation, although most recently two tandem structures have been reported, namely the IG-tandem Z1-Z2 (Zou *et al.*, 2006; Marino *et al.*, 2006) and the IG-IG-FNIII tandem A168-A170 (Mrosek *et al.*, 2007). It should be noted that, to this date, a single structure of a FNIII module from human titin, A71, has been elucidated (Muhle-Goll *et al.*, 1998).

Table 3.1 Structures of IG/FNIII modules from human titin (updated from Marino *et al.*, 2005).

Module[1]	Location	PDB ID	Method	Reference
Z1Z2	Z-disc	2A38	X-ray Cryst.	Marino *et al.*, 2006
Z1Z2[2]	Z-disc	1YA5	X-ray Cryst.	Zou *et al.*, 2006
I1	I-band[3]	1G1C	X-ray Cryst.	Mayans *et al.*, 2001
I91[4]	I-band[5]	1TIT[6]	NMR	Improta *et al.*, 1996
A71[7]	A-band	1BPV	NMR	Muhle-Goll *et al.*, 1998
A168-A170	P-zone	2NZI	X-ray Cryst.	Mrosek *et al.*, 2007
M1	M-line	2BK8	X-ray Cryst.	Müller *et al.*,
M5	M-line	1NCT[8]	NMR	Pfuhl & Pastore, 1995

[1] all modules (except A71 and A170) have IG fold, [2] in complex with telethonin, [3] constitutive, proximal I-band [4] former name: I27, [5] constitutive, distal I-band [6] also 1TIU, [7] position seven in the third long super-repeat domain, [8] also 1NCU, 1TNM and 1TNN

To date, the architecture and conformational dynamics of the titin chain remains mostly unknown. One single complete study exists that addresses the properties of a two-IG tandem. This carried out by Marino and coworkers (2006) investigated Z1-Z2 from the N-terminus of the titin filament using X-ray crystallography, small angle X-ray scattering (SAXS), NMR relaxation and residual dipolar couplings. This revealed that Z1-Z2 preferentially adopts an extended conformation at equilibrium in solution despite a lack of interactions across the IG modules and the presence of a three-residue, hydrophilic linker. The restricted nature of the modular dynamics in Z1-Z2 appears mediated by a minor hydrophobic contribution from the aliphatic fraction of linker residues. The putative role of this preferential conformation is the selective recruitment of the Z1-Z2 binding partner telethonin by steric means. No analogous studies have been carried out so far for titin's FNIII tandems.

3.2.3c The architecture of titin's FNIII modules

The NMR solution structure of A71 (Muhle-Goll *et al.,* 1998) gave detailed insights into the structure of intracellular FNIII modules. Based on this structure, the conservation of titin FNIII modules was estimated by sequence alignments and mapped to A71 (Figure 3.6), which showed repeat position dependent conservation (Muhle-Goll *et al.,* 2001).
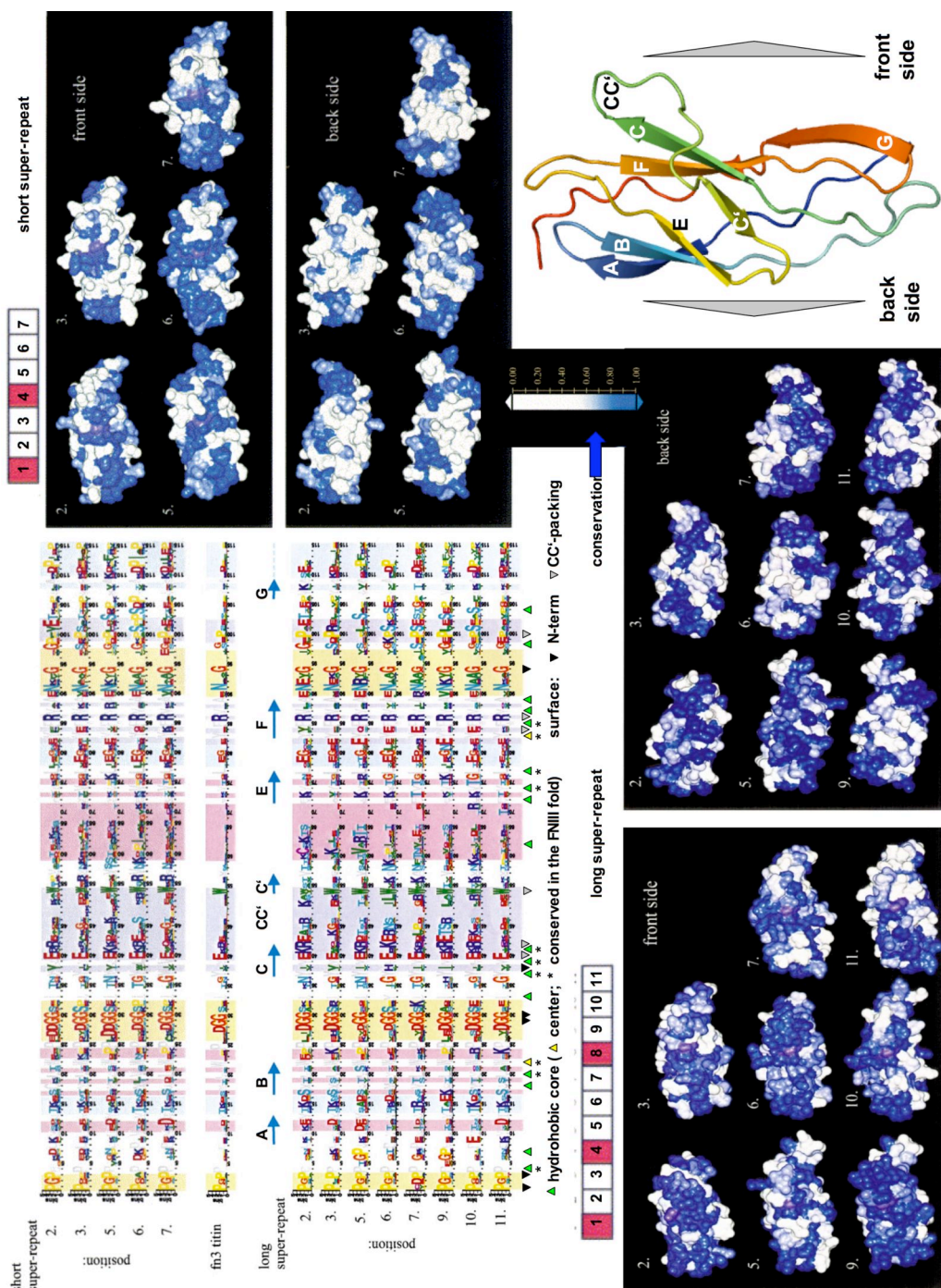
Figure 3.6 Conservation in titin's FNIII modules.

Figure 3.6 Conservation in titin's FNIII modules.
Sequence logos of the consensus sequences of the FNIII domain families in titin (top left). Each alignment column corresponds to a stack of letters where the total height of the stack equals the relative entropy of the column. Letters that make up one stack are sorted according to their height, which corresponds to their contribution to the relative entropy of the column. Letters in gray correspond to conserved positions in the interior of the domains. The underlying color code corresponds to the location of the residues on the surface of the domains (yellow: N-terminal side, cyan: C-terminal side, magenta: back side and purple: front side/CC'-loop packing). Residues involved in the hydrophobic core as well as highly conserved surface patches are indicated below the sequences (as given in Muhle-Goll *et al.,* 1998). The cartoon representation of A71 (lower right) shows the front and back sides of the module. The position-dependent surface conservation scores are shown for each position of the short (upper right) and the long super-repeat (lower left) from white (no conservation) to dark blue (>85% conserved). The almost invariant Trp54 (center of the CC'-loop packing) is shown in magenta. Domain positions in the super-repeat are indicated above the structures. Image adapted from Muhle-Goll *et al.,* 2001.

In individual FNIII modules, the conserved residues can be classified into two classes (Muhle-Goll *et al.,* 2001): residues belonging to the hydrophobic core (maintaining the module fold) and residues on the domain surface (required for specific functions). The hydrophobic core of A71 is dominated by aromatic and aliphatic residues, which group in two centers, with Trp22 at the N-terminal side and Tyr72 (also involved in the Tyr-corner) as the C-terminal center. The surface conservation of titin's FNIII modules depends on their position within the repeat, with the highest conservation in FNIII modules downstream to IG modules. In general, the front-side (side of the four-stranded β-sheet) has a higher degree of surface conservation, which mainly involves the CC'-loop as a hot spot of conservation (Muhle-Goll *et al.,* 1998). There, a highly conserved Trp residue (Trp54 in A71) packs between surface-exposed residues on the neighboring β-strands. Its interaction partners are preferentially Arg or Lys residues surrounded by a second layer of Glu residues.

The loops of titin's FNIII modules have a conserved length, contrary to extracellular FNIIIs that show variable loop lengths with low conservation (*e.g.* FNIII modules of fibronectin, Leahy *et al.,* 1995, Copie *et al.,* 1998 and Sharma et al., 1999; integrin, de Pereda *et al.,* 1999; neuroglian, Huber *et al.,* 1994 and tenascin, Bisig et al., 1999). Key residues present in the loop regions of all titin FNIII domains suggest limited structural freedom. The tight packing of conserved residues in the N-terminal loop region contains a central poly-Pro (type II) helix, which packs into a groove between the BC-loop and FG-turn. The central element in the C-terminus is the Tyr-corner, which packs against the AB-hairpin and the EF-loop. The C-terminal domain border is a residue two positions downstream of the last β-strand (G-strand).

3.2.3d Evolution of titin's FNIII-modules

Titin's FNIII modules have evolved by gene duplication events. A phylogenetic tree of titin's FNIII modules (Figure 3.7A) reveals branches tightly packed at the center of the tree indicating early and rapid duplication events. In general, modules in equivalent positions across repeat cluster and, thus, they share higher similarity than domains within a same repeat. These findings led to a model, in which whole super-repeat tandems duplicated during the evolution of titin. The long super-repeat evolved in an additional duplication event from the short super-repeat (Figure 3.7B).
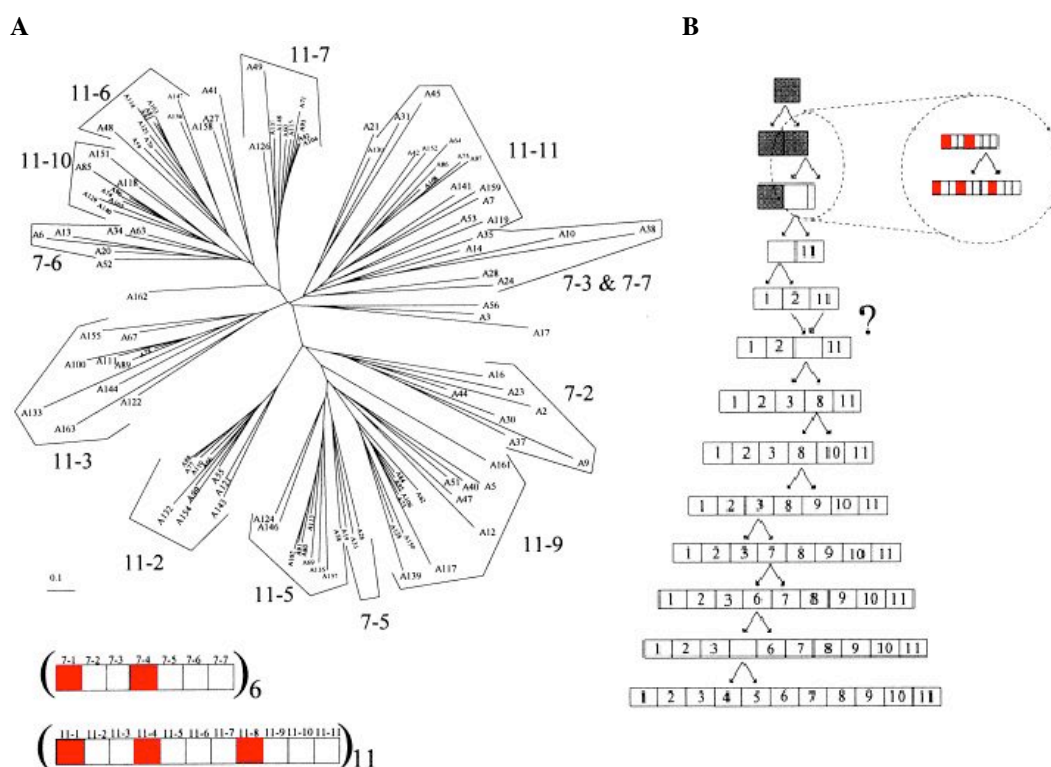
Figure 3.7 The evolution of titin's super-repeats and FNIII modules.
(**A**) Neighbor-joining tree of titin's FNIII modules and the corresponding nomenclature (indicated above the repeat structure) below. IG modules (red) and FNIII modules (white) cluster according to their super-repeat domain position. The branches tightly packed at the center of the tree indicate early and rapid duplication events. (**B**) Proposed order of duplication based on analysis of the branching pattern: the gene of the short super-repeat (gray) has duplicated to form the long super-repeat, which was extended by another gene duplication event (partial duplication of the C-terminal moiety of the short super-repeat; adapted from Kenny *et al.*, 1999).

3.3   Methods

The cloning (for cDNA sequence see EMBL X90568), expression and purification of the FNIII tandem A77-78 from titin's A-band have been described by Muhle-Goll and coworkers (2001). The plasmid (pETM11) with the expression clone of the construct A77-78 was kindly provided by Claudia Muhle-Goll (EMBL, Heidelberg).

3.3.1   Expression and purification

Expression and purification protocols were adapted and optimized from Muhle-Goll *et al.,* (2001). Briefly: the recombinant protein was expressed in *E. coli* strain BL21(DE3) *Rosetta* using pETM11 as expression vector. The cells were grown at 37°C up to an OD_{600} of approximately 0.6 in Luria Bertani medium supplemented with 25 μg/mL Kanamycin and 34 μg/mL Chloramphenicol. Expression was induced by addition of *iso*-Propyl-β-D-thiogalactopyranosid (IPTG) to a final concentration of 1 mM. At the induction point the temperature was lowered to 30°C and cells grown further for approximately 16 hours. Cells were harvested by centrifugation (3400 x g, 4°C, 15 min). Bacterial pellets were resuspended in lysis buffer (100 mM NaCl, 50 mM Tris HCl pH 7.2, 2 mM DTT) in the presence of a protease inhibitor cocktail (Boehringer), DNase (2 μg/mL) and lysozyme (10 mg/mL). The resulting suspension was lysed by sonication (Branson Sonifier, ice cooling, 3 s pulse, 6 s pause, 3 min overall) and the homogenate centrifuged (43'000 x g, 4°C, 30 min). The supernatant was applied to a Ni^{2+}-chelating HisTrap column (GE Healthcare) equilibrated in lysis buffer. Elution used lysis buffer supplemented with 250 mM imidazol. The eluate was dialyzed against 150 mM NaCl, 50 mM Tris-HCl pH 8.0, 2 mM DTT in the presence of TEV-protease (ratio TEV:protein = 1:50) at 4°C for 16 h. This resulted in efficient tag removal (>90 %). Traces of uncut protein and the His-tagged TEV protease were removed by subtractive affinity chromatography (Ni^{2+}-chelating HisTrap column). The flow-through fractions were concentrated and applied to a Superdex 75 HiLoad 16/60 prep grade column (GE Healthcare) equilibrated in lysis buffer. The elution profile of this size exclusion chromatography showed a single peak (red bar in Figure 3.8). A schematic overview of the expression

and purification procedure is shown in Figure 3.8. The purified protein was concentrated to 20-80 mg/mL (determined by A_{280}) and stored at 4°C for further use.
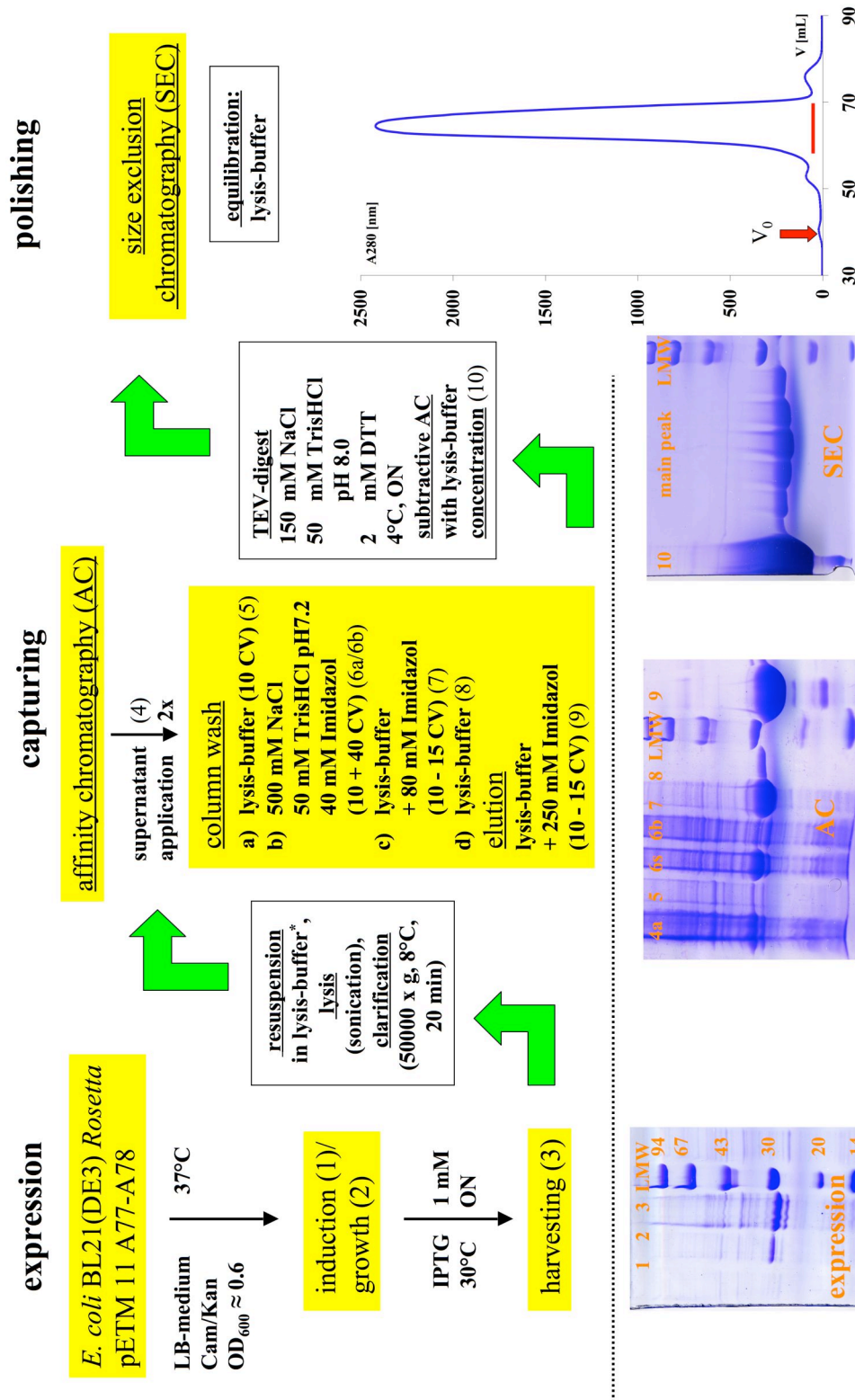
Figure 3.8 Scheme for the expression and purification of construct A77-A78.
The numbers in brackets correspond to the fractions on the SDS-gels.

3.3.2 Crystallization

Crystallization trials were performed using the vapor diffusion method at room temperature, where 1 μL of A77-A78 was mixed with 1 μL of reservoir solution to form a 2 μL drop. For initial screening, each droplet was equilibrated against 70 μL of reservoir solution in 96-well CrystalQuick protein crystallization plates (Greiner) using a sitting-drop setting. Commercial kits[1] yielded several initial crystallization conditions (Table 3.2).

Table 3.2 Hits in the initial screening of A77-A78.

Precipitant	Buffer	Additive
30 % PEG-4000	100 mM TRIS pH 8.5	200 mM Li_2SO_4 x H_2O
70 % MPD	100 mM HEPES pH 7.5	-
50 % PEG-200	100 mM Na/K-phosphate pH 6.2	200 mM NaCl
30 % 1,2-propanediol	100 mM HEPES pH 7.5	20 % PEG-400

Crystals grew as clusters of thin needles with different diameters. The conditions were reproduced using the hanging-drop vapor diffusion method with the same drop ratio and increased reservoir volume (500 μL) in VDX plates (Hampton Research). Conditions with PEG-4000, PEG-200 and propanediol as precipitant did not show significant improvement (independent of buffer, pH-range, PEG molecular weight, crystallization method, precipitant combination and protein concentration).

MPD as precipitant yielded crystals overnight, combined with a significant amount of amorphous precipitate. Subsequently the effect of precipitant and protein concentration were explored. Additionally, a possible temperature dependence of crystallization was examined at 4°C. Since MPD is a volatile precipitant, setups with MPD solely in the reservoir and drops with protein solution (increased volume, 1.5 μL) were produced and stored at different temperatures (4°C and 20°C). Only drops stored at 4°C remained transparent, from which clusters of needles with higher diameter grew after approximately two weeks. To improve the quality of crystals in these MPD evaporation setups, different additives (*e.g.* PEG-200, PEG-400, 1,6-hexanediol, glycerol) were tried. None of the additives is volatile. Thus, each had

---

[1] Sparse matrix kits CS-I & II from Hampton Research; Cryo I, Cryo II, Wizard I, Wizard II from Emerald Biostructures; Structure Screen 1, Structure Screen 2 from Molecular Dimensions

to be added to the drop (but not to the reservoir) from stock solutions (0.5 μL of a 20 % four-fold stock, final concentration 5%), which increased the drop size from 1.5 μL to 2.0 μL. Crystals (some >500 μm in diameter) suitable for X-ray diffraction experiments, were found with this volatile evaporation/additive approach (at 4 °C). Best crystals, grown from 40 % MPD, 5 % PEG-400 were directly frozen in mother liquor and used in data collection. For a schematic overview of the refinement of crystallization conditions of A77-A78 see Figure 3.9.
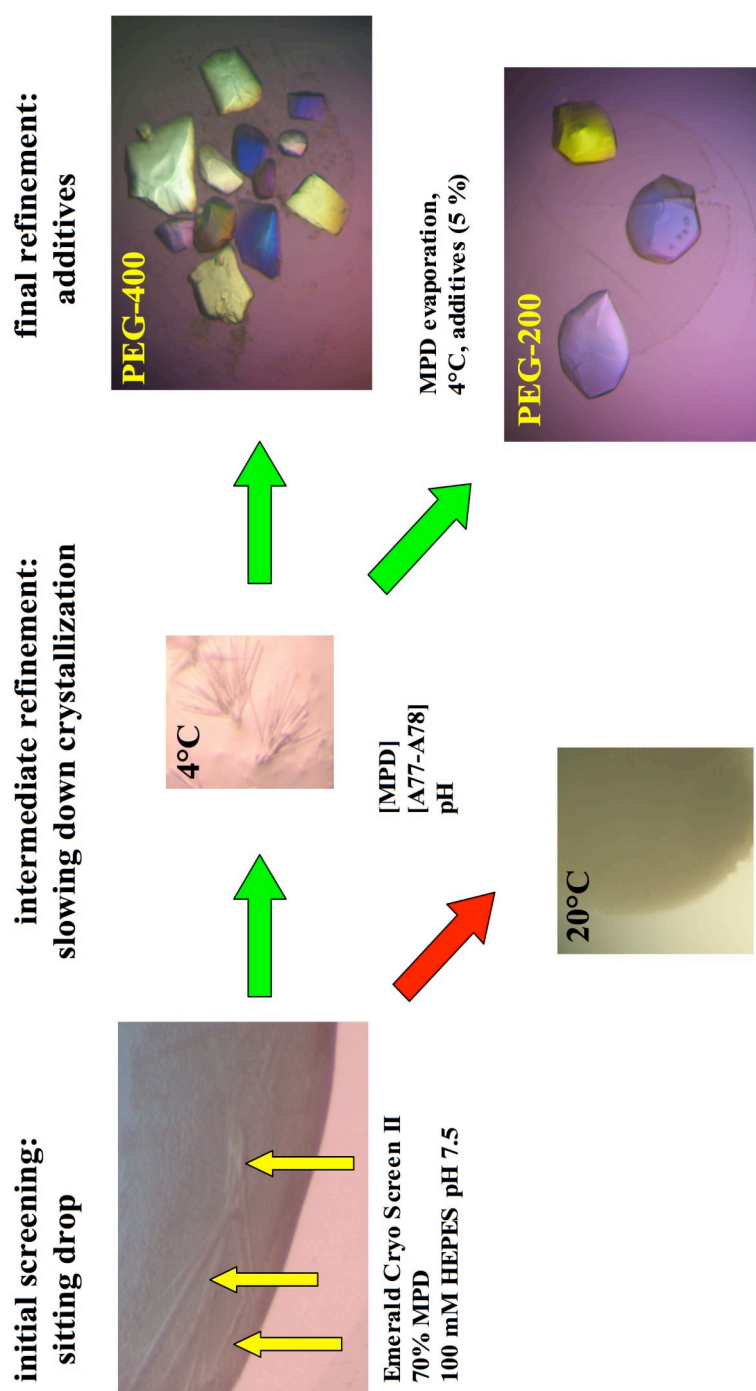
Figure 3.9 Refinement of crystallization conditions for the FNIII tandem A77-A78. Note: the images are not in scale. The needles in initial screening conditions (indicated by yellow arrows) and subsequent refinement trials (mid) were much smaller than crystals obtained in the final refinement (right). In intermediate refinement, setups stored at 20°C gave amorphous precipitate (red pathway), while setups stored at 4°C yielded needles with improved size (green pathway). In final refinement, crystals with PEG-400 as additive appeared first and were used for collection of native X-ray data.

3.3.3   Collection of a native data set

A native data set was collected from a best crystal of titin A77-A78 at 100 K on an Elliot GX20 CuK$_\alpha$ rotating-anode equipped with a mar345 detector. Data were collected as two series of non-overlapping oscillations (see Table 3.3 for data collection parameters).

Table 3.3 Data collection parameters for A77-A78.

Parameter	Value
Total rotation	103°
Oscillation range ($\Delta\rho$)	0.5°
Exposure time	240 s per frame
Detector to crystal distance	100 mm

The diffraction images (Figure 3.10) were processed up to 1.65 Å with the XDS/XSCALE package (Kabsch, 1988). For X-ray statistics see Table 3.4.
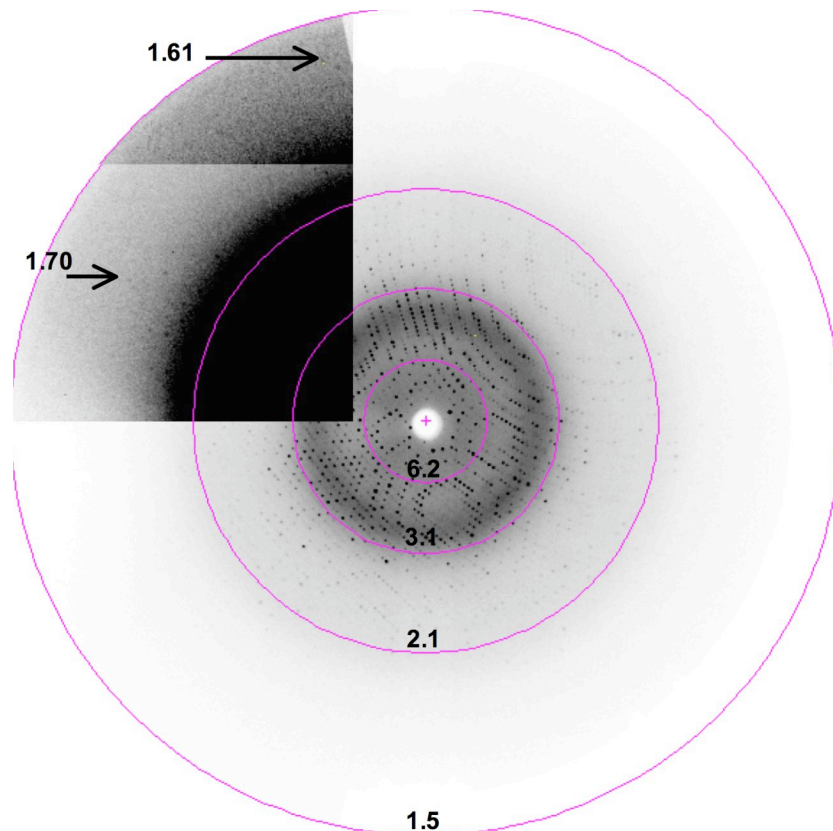
Figure 3.10 Diffraction images from crystals of A77-A78.
High resolution reflections are shown in the area with high contrast. Resolution rings [Å] are displayed.

A77-A78: insights into the A-band organization in human muscle

Table 3.4 X-ray statistics for A77-A78.

Parameter	Value
X-ray source	Elliot GX20
Detector	mar345
Wavelength	1.5405 Å
Spacegroup	C 2 2 2_1
Unit cell [Å]	115.90, 163.20, 65.20
Resolution	20 - 1.65 (1.68 - 1.65)[1]
Unique reflections	74213 (3775)[1]
R_{sym} (I) [%]	7.1 (36.8)[1]
I/σ (I)	11.9 (3.6)[1]
Multiplicity	3.90 (3.69)[1]
Completeness [%]	96.7 (99.5)[1]

[1]outer resolution shell

The R_{free} (Brünger, 1992) was used as a cross-validation indicator during refinement. For this, reflections were partitioned at an early state into a working and a free set using FREERFLAG (CCP4, 1994).

3.3.4 Detection of non-crystallographic symmetry (NCS)

The Matthews coefficient V_m (Matthews, 1968) was calculated in a way analogous to *Ss*TrpB2b (section 0). Considering that the molecular weight of A77-A78 is 21595 Da (as calculated from sequence data), the analysis of the Matthews coefficient (V_M; Matthews 1968) revealed that the a.u. must contain two to four copies of A77-A78 (Table 3.5) with three being the most probable.

Table 3.5 Matthews coefficients for crystals of A77-A78.

Molecules per a.u.	Matthews coefficient	Solvent content [%]
2	3.6	65.7
3	2.8	48.5
4	1.8	31.4

In the subsequent steps of structure elucidation, it became clear that there are two tandems of A77-A78 in the a.u. For the detection of the rotational component of the non-crystallographic symmetry (NCS) the programs AMORE (Navaza, 1994) (see Table 3.6 for solutions) and POLARRFN (CCP4, 1994) were used. The method correlates sets of self-vectors from two Patterson maps.

Table 3.6 Euler and Polar rotational NCS relations as determined using AMORE.

Alpha	Beta	Gamma	Peak height	Omega	Phi	Kappa
0.0	180.0	90.6	61.6 (%)	90.0	44.7	180.0

The self-rotation function was calculated between a resolution of 18.0 and 5 Å
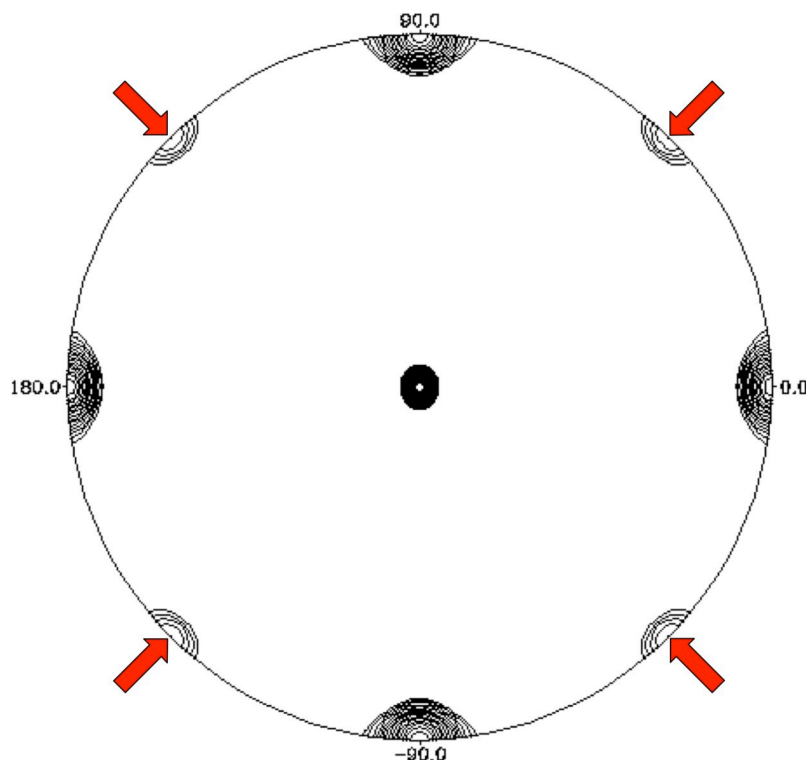
Figure 3.11 Graphical representation of the self-rotation function calculated using POLARFN. Stereographic projection of the section $\kappa = 180°$. The presence of NCS is shown by additional 2-fold axes indicated by red arrows.

This result indicated that a two-fold ($\kappa=180°$) NCS axis existed perpendicular to c (omega=90°) and roughly contained in the *ab* plane, bisecting it (phi=45°). The axis and the molecular copies that it relates is displayed on the three-dimensional crystal lattice shown in Figure 3.13 (section 3.4.1).

3.3.5 Phasing by Molecular Replacement (MR)

The structure of the FNIII domain A170 (determined previously in our group; Mrosek *et al.*, 2007) was kindly provided by Michael Mrosek to be used as search model in this study. A170 is positioned downstream of A77-A78 in the M-line region of titin. Two modified search models derived from A170 were produced that maximally approximated the sequence features of A77 and A78 separately. For this, corresponding sequence alignments of A170 to each of the modules of A77-A78 (38 % identity to A77 and 34 % to A78) were carried out. With the resulting models trimmed to the minimal common sequence, MR was pursued using PHASER (McCoy *et al.*, 2005) (Table 3.7). MR solutions were visually inspected for lattice packing and fitting into a calculated electron density map with the program O.

Table 3.7 Statistics for MR using PHASER.

Search model	Z_{rot}	Z_{trans}	log-likelihood gain[1]
A77 (module1)	3.9	5.5	22
A77 (module2)	4.6	7.9	73
A78 (module1)	3.4	12.4	162
A78 (module2)	3.3	16.1	289

[1]overall log-likelihood gain: 367

3.3.6 Refinement and model building

The initial MR model of A77-78 was subjected to ARP/wARP (Perrakis *et al.*, 1999) for automated model building (warpNtrace procedure). The warpNtrace procedure (overview in Figure 6.5 in the appendix) calculates an electron density from the pre-refined initial model and creates a free atom model in this electron density. The stereochemistry of the MR model is completely ignored and the free atom model expanded by adding atoms in regions of high density. Based on connectivity, the main-chain is traced first (normally this results in various fragments). Intermediate hybrid models are refined using REFMAC (Murshudov *et al.*, 1997). The docking of the sequence into the side-chain electron density produces the final ARP/wARP model. The intermediate structure so obtained was refined against native data between 1.65 and 19.80 Å resolution using REFMAC

(Murshudov *et al.*, 1997) without a translation/liberation/screw (TLS) tensor. Restrained maximum likelihood based refinement included sparse matrix minimization (without NCS restraints), overall individual isotropic B-factor refinement and bulk solvent correction. Solvent molecules were identified with ARP/wARP (solvent mode routine) and COOT. For manual interpretation of electron density and subsequent model building the software COOT (Emsley & Cowtan, 2004) was used.

The final model contains two copies of A77-78 present in the asymmetric unit (see Table 3.8 for model parameters). One copy shows all expected residues (Gly1-Asp197) where the first three residues (Gly1-Ala-Met3) are remnants of the TEV-cleavage site and have no biological function. In the second copy, those residues are disordered and not included in the model (the N-terminal residue of this chain is the physiological Asp4).

Table 3.8 Model parameters for A77-A78.

Variable	Value
R-factor/R_{free} [%][1]	21.6/23.4
Number of protein residues	392
Number of solvent molecules	482
Number of MPD molecules	2
rmsd (NCS copies)[2] [Å]	0.837
average B-factor (MC/SC)[3]	18.26/20.66 ‖ 19.09/21.24
rmsd bonds [Å]/angle [°]	0.012/1.361

[1]the R-free set comprised 978 reflections corresponding to 1.35% of the total data.
[2]for all atoms of the residues Asp4 - Asp197 calculated with LSQKAB
[3]MC = main-chain, SC = side-chain for copy 1 ‖ copy 2

The Ramachandran plot (Ramachandran *et al.*, 1963) shows that all residues of both NCS copies of A77-A78 are in allowed regions (Figure 3.12, with corresponding statistics in Table 3.9).

Table 3.9 Ramachandran diagram statistics[1]

	Number	Percentage
Residues in favored regions	305	90.8
Residues in additional allowed regions	31	9.2
Residues in generously allowed regions	0	0.0
Residues in disallowed regions	0	0.0

[1]for non-Gly and non-Pro residues (336) from analysis in PROCHECK
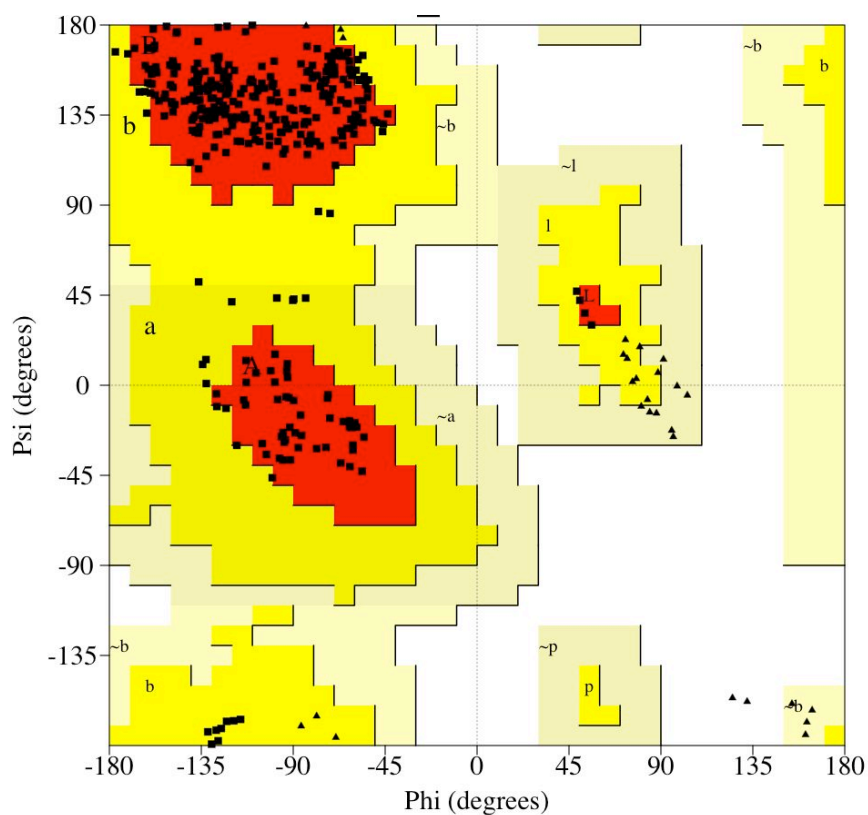
Figure 3.12 Ramachandran diagram of the two NCS-related copies of titin A77-A78.
Favored regions are indicated in red, additional allowed regions in yellow, generously allowed regions by light yellow and disallowed regions by white background, respectively. The plot was generated using PROCHECK (Laskowski *et al.*, 1993).

3.3.7   Sequence alignments and figures

The titin sequence (EMBL-data library accession code X90568) was fragmented to extract the FNIII domain sequence information. For each domain, N-terminal and C-terminal residues were defined as the first and last ones forming

interactions within each module in the structure of A77-A78. The resulting module sequences were aligned using ClustalW (Chenna *et al.,* 2003) and visually validated according to the integral structural features reported for FNIII modules (*e.g.* hydrophobic core packing). Sequence conservation was mapped onto the B-factor column using MODTRAFO (Schirmer, in-house software), which applies the BLOSOM matrix. Figures showing the structural features were prepared with PyMOL (DeLano, 2004).

3.3.8   Structure superposition

For comparison of A77-A78 to structures deposited in the PDB database (www.rcsb.org), a sequence-based search was performed. Structures involved in protein-protein complexes were excluded while multidomain arrangements were investigated further. These multidomain FNIII structures were superposed to A77-A78 using TOP (CCP4, 1994). Best superpositions were investigated further at the structural level (see section 3.4.4a).

For A77, best matches were found to the tandem of *Drosophila* neuroglican (Huber *et al.,* 1994, 1CFB) and module two (of four) of the FNIII tandem from human fibronectin (*Leahy et al.,* 1996, 1FNF). Best matches for A78 were module two of the heparin and integrin binding fragment of fibronectin (Sharma *et al*., 1999, 1FNH), module two of a tandem from chicken tenascin (Bisig et al., 1999, 1QR4) and two human FNIII modules of the cytoplasmic tail of integrin α6β4 (de Pereda *et al.,* 1999, 1QG3). The cell attachment modules of mouse fibronectin (Copie et al., 1998, 2MFN, *rmsd* 1.1 Å over 76 residues, identity 25 %) were discarded due to reported inter-modular flexibility.

Visual comparison and analysis of A77-A78 and investigated FNIII-tandems focused on the highly conserved N-terminal interface region (including loop conformation, composition and orientation) since conservation on the C-terminal side seems to be linked only to the core fold of FNIII modules (loops show conserved length and conformation in titin and other FNIII modules.

The superposition of A77 to A168-A170 (Mrosek *et al.,* 2007) using TOP yielded the template for homology model production

3.4 Results and discussion

3.4.1 The crystal lattice of A77-A78

The crystal structure of A77-78 has been elucidated at 1.65 Å resolution using the Molecular Replacement (MR) method. The asymmetric unit (a.u.) of A77-A78 crystal contains two NCS-related copies. The Matthews coefficient of these crystals (3.6 Å^3/Da) indicates an unusually high solvent content (66%). Protein molecules are not distributed evenly in the crystal, but form a quadratic lattice with large solvent-filled channels that co-align to "perforate" the crystal as micro-pores (Figure 3.13).
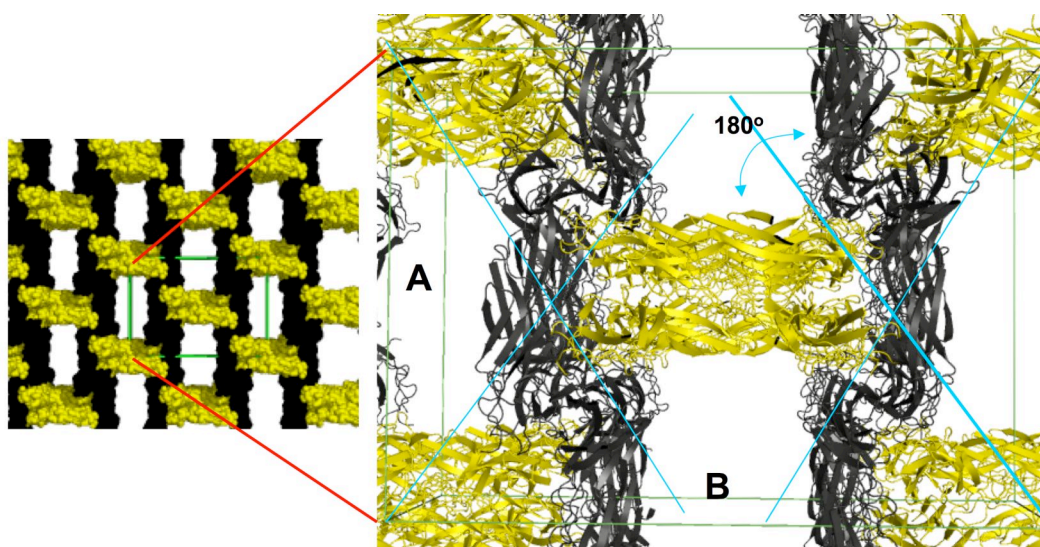
Figure 3.13 The crystal lattice of A77-A78 (view of the AB-plane).
In the crystal lattice A77-A78 is densely packed and forms solvent channels. One copy of A77-78 (and its symmetry equivalent molecules) is shown in black, the other in yellow. The unit cell is in green. The black tandems form layers parallel to the AC-plane, where A77-A78 molecules arrange anti-parallel. These layers are connected by parallel and antiparallel arranged A77-A78 molecules. The 2-fold relation between NCS copies in the a.u. is indicated in cyan.

The parallel arrangement of symmetry related molecules (Figure 3.14) is noteworthy. In this, two A77-A78 tandems associate over a broad surface (1430 Å^2 or 715 Å^2 per chain; analyzed with ProFace, Saha *et al.*, 2006), trapping multiple solvent atoms in between and where a predominantly hydrophobic interface is contributed by A78 domains.
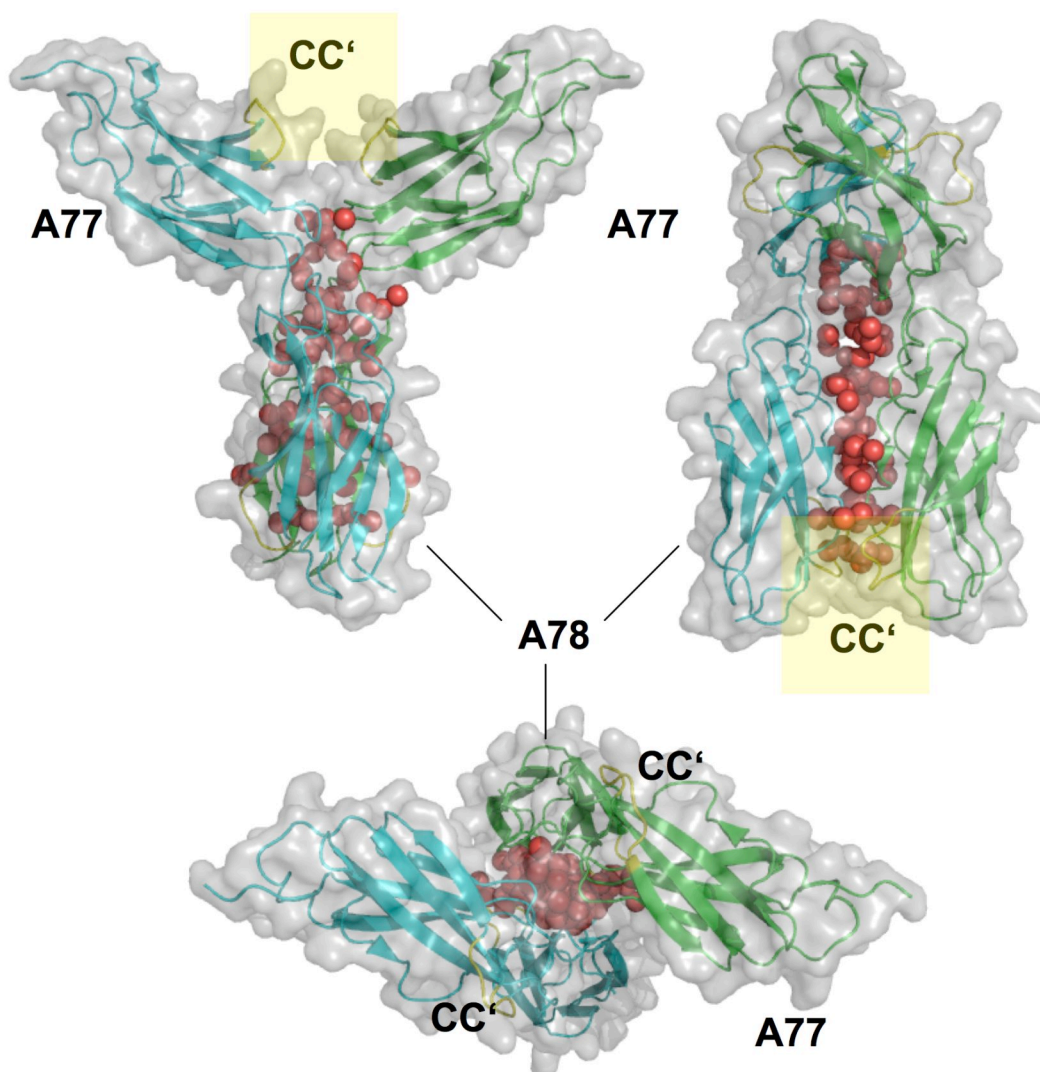
Figure 3.14 Y-shaped arrangement of NCS-related A77-A78 molecules in the crystal.
Orthogonal views of the two A77-A78 tandems (green and cyan with grey surface) associated over a broad surface with water molecules (red) in the interface. Highly conserved patches from both tandems cluster in the center of the dimer: the CC'-loop and the BC-loop.

3.4.2   Overall structure of A77-78

A77-A78 represent positions two and three in repeat four of the long super-repeat from the C-zone of titin, where the A77-A78 modules are flanked by IG domains. This is the first structure of an FNIII tandem from titin. Both copies of A77-A78 in the a.u. contain residues 4-197 (residues 1-3, remnants from protease digest, are disordered in one chain). The two NCS copies are essentially identical (*rmsd* 0.64 Å for C_{α}-atoms 4-196 calculated with LSQKAB, CCP4, 1994).

Each of the chains comprises two contiguous FNIII-modules (Figure 3.15) showing a classical topology, i.e. a two-layer β-sandwich consisting of a short three-stranded and a long four-stranded β-sheet packing against each other to form the hydrophobic core (Main *et al.,* 1992; Campbell & Spitzfaden, 1994).
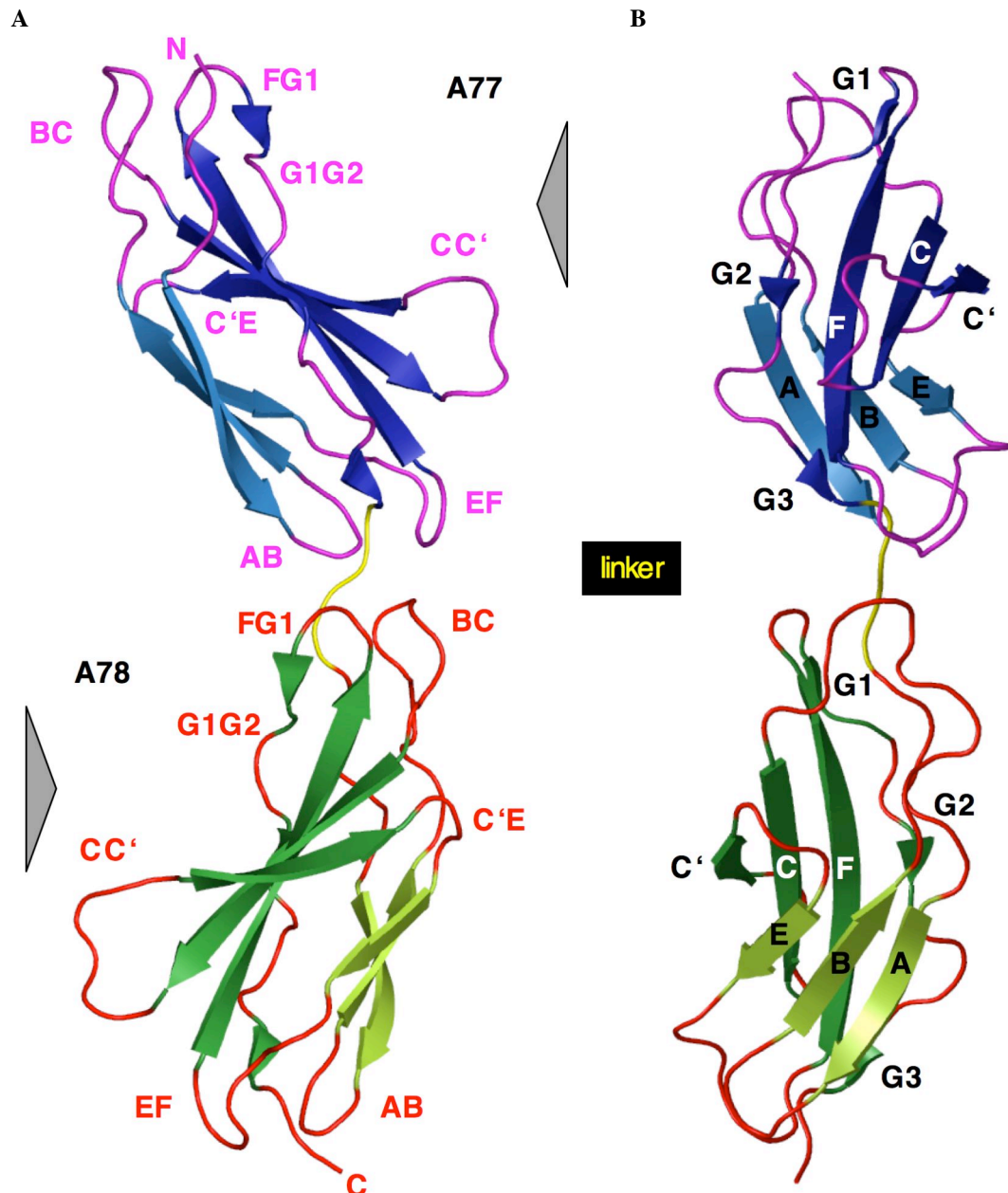
Figure 3.15 Overall structure of A77-A78.
The modules interact in a tail (A77) to head (A78) fashion involving their C- and N-terminal loop regions, respectively. The β-sheets of A77 are colored in blue with loops in magenta, β-sheets of A78 are colored in green with loops in red. Linker residues are colored in yellow. (**A**) Side-view with labeled loops. The front-sides are indicated with gray triangles. (**B**) View onto the concave plane of A77-A78 with labeled β-sheets (view onto the back-side of A77 and the front-side of A78).

The secondary structure elements of A77-A78 have been assigned with PROMOTIF (Hutchinson & Thornton, 1996). Analogous to the structure of A71 (Muhle-Goll *et al.,* 1998), the last β-strand (G-strand) in each module is fragmented. It comprises three segments with two amino acid residues each (referred to as G1, G2 and G3 in the following).

A77-A78 adopts a slightly bent conformation (opening angles of 157° and 158° for chain A and B, calculated with MOLEMAN2), while front and back sides roughly point in opposite directions (torsion angles of 176° for both chains, calculated with MOLEMAN2). The distance between the centers of gravity of A77 and A78 is approximately 41 Å, which is slightly larger than the observed periodicity in low-angle X-ray diffraction (40 Å; Squire *et al.,* 2004). The identical domain orientation of both copies in the a.u. suggests that the conformation of A77-A78 is well defined and probably affected by low modular variability.

3.4.3 Interfaces in A77-A78

The A77-A78 tandem contains an intra-domain and a domain-domain interface. In the intra-domain interface, residues between the layers of the β-sandwich form a tight hydrophobic packing with two centers, this is a "hydrophobic core" that is common to FNIII modules. Additionally, the extraordinarily long loops in the N-terminal loop cluster of titin's FNIII modules form interactions extending the hydrophobic core. The domain-domain interface shows a FNIII transition with an atypical interaction pattern (mainly polar contribution) involving an unusually high number of loops. A detailed description of both interfaces is given below.
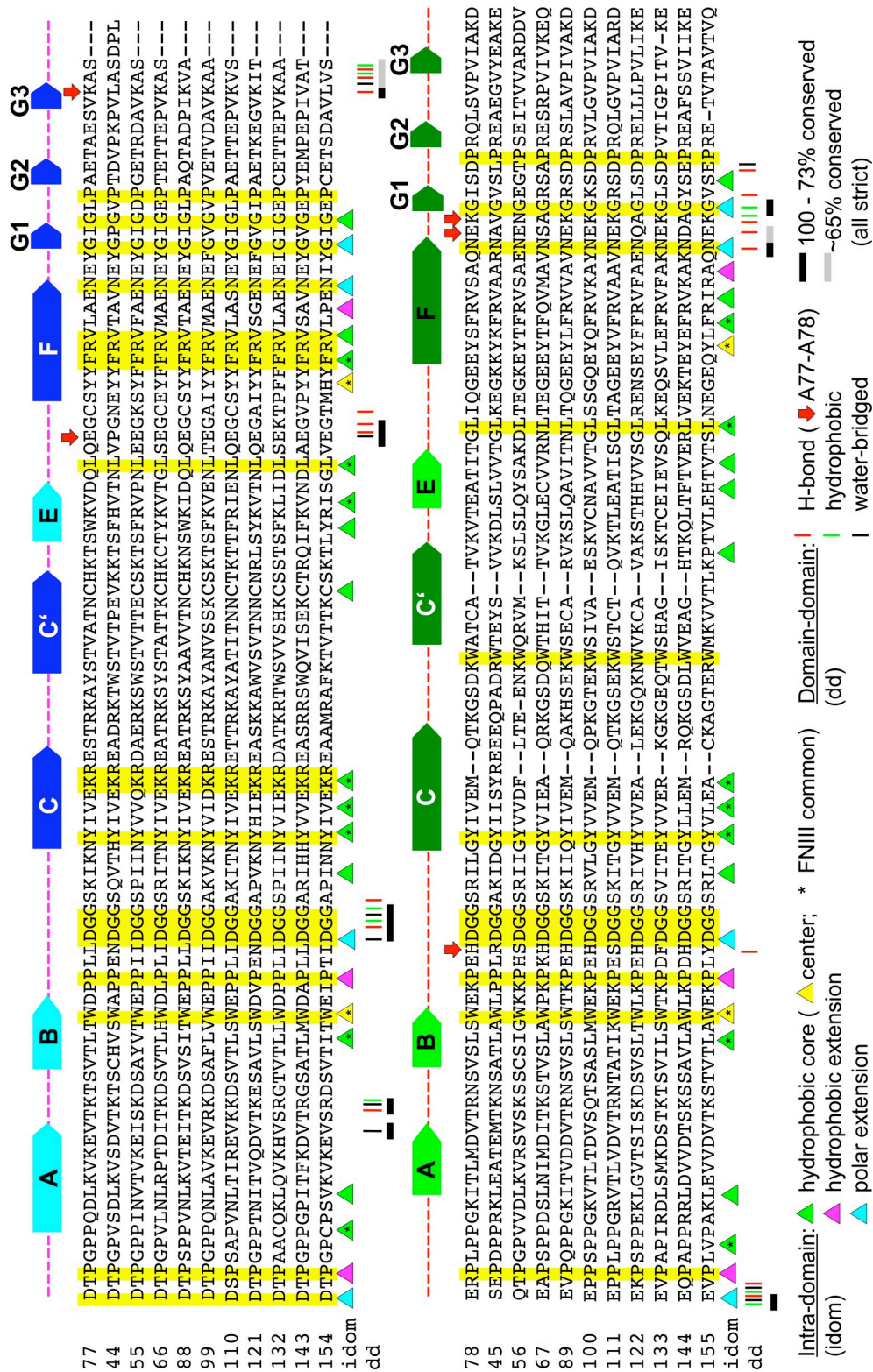
Figure 3.16 Interfaces in A77-A78 (and equivalent domains of the long super-repeat).
Secondary structure elements are shown in the color code of Figure 3.15. A legend explains signs indicating interactions indicated below the sequences.

3.4.3a The intra-domain interface: the hydrophobic core and its extension

The intra-domain interface in A77-A78 reveals hydrophobic interactions in the center of the domains (hydrophobic core) and a polar extension in the N-terminal loop cluster. The centers of the hydrophobic core common to FNIII modules are a strictly conserved Trp (Trp26/Trp123 in A77/A78) at the N-terminal side of the sandwich and a Tyr (Tyr75/Tyr171; involved in the Tyr-corner) at the C-terminal side. These highly conserved residues tightly pack together in titin's FNIIIs (as reported by Muhle-Goll *et al.,* 1998, Figure 3.16A). The OH-group of the Tyr-corner also coordinates a water molecule as reported by de Pereda *et al.* (1999) for the cytoplasmic tail of integrin.

On the N-terminal side, the common hydrophobic core is extended into the corresponding loop region by hydrophobic and polar interactions between conserved residues (Figure 3.16B). Central to the hydrophobic packing in the N-terminal loop region are four residues (Pro6-Pro9 in A77 and Pro103-Pro106 in A78) in a poly-Pro type II helix conformation (as reported by Huber *et al.,* 1994 for *Drosophila* neuroglian). These residues pack between strictly conserved residues in the BC-loop and the FG-turn. The central residue in this cluster is a strictly conserved Asn (Asn83/Asn179) part of a Asn-X-X-Gly conserved motif, which defines this β-turn and that forms a buried H-bonding network with highly conserved residues inter-connecting the linker, the BC-loop and the FG1-turn. Thus, the core packing of titin's FNIII modules comprises three centers, where the central Trp and the C-terminal Tyr form the hydrophobic core and the N-terminal loop cluster is stabilized by the polar interactions of an Asn. This may give the loop regions a dual role, since they are also involved in domain-domain interfaces.

Figure 3.17 The intra-domain interface of A78.
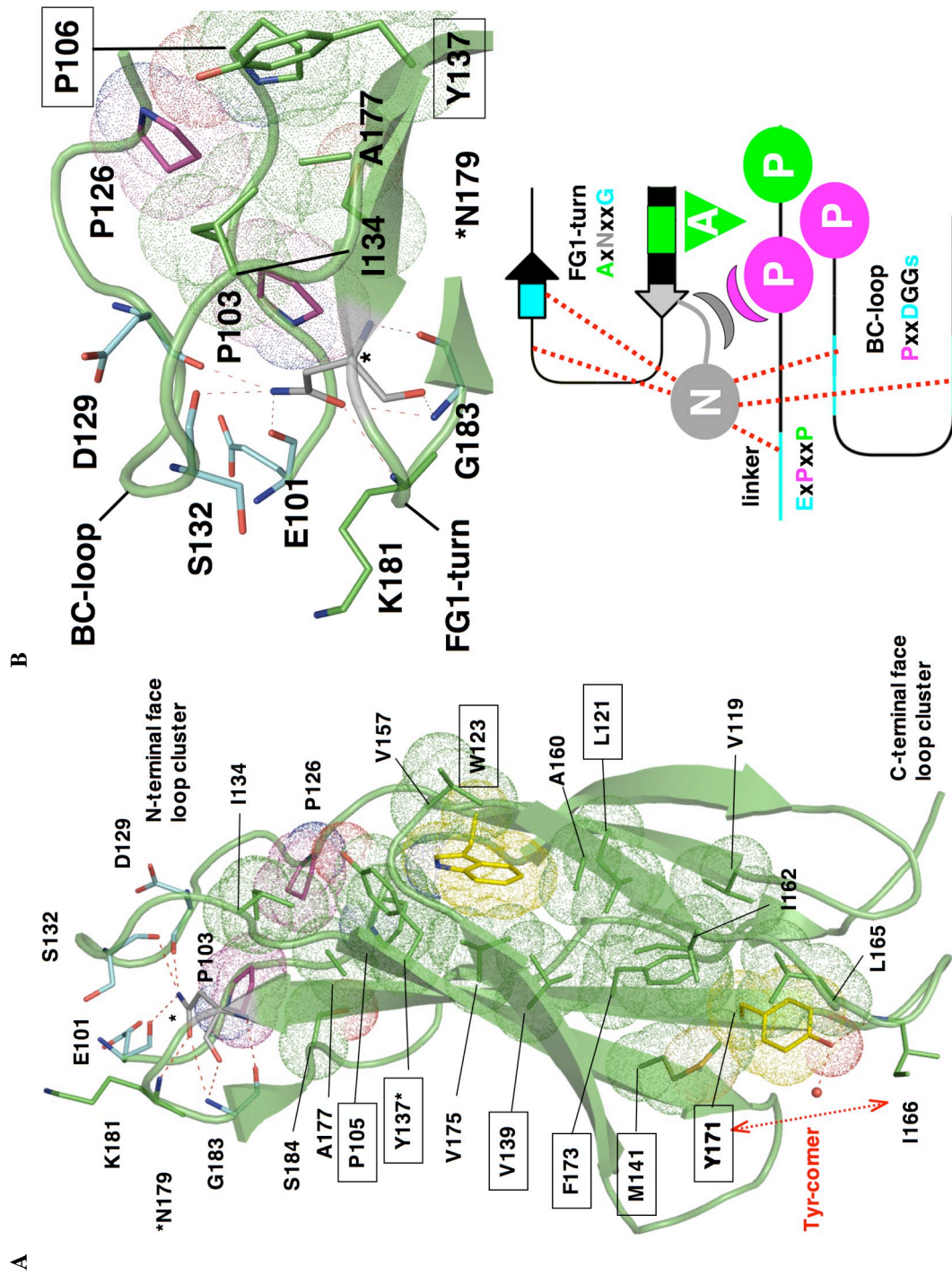
Figure 3.17 The intra-domain interface of A78.
(**A**) Residues involved in hydrophobic core formation are shown in green (sticks/dots) with the core residues in yellow. Residues forming additional hydrophobic interactions are shown in magenta (dots/sticks) and residues forming H-bonds (red dashed lines) to the central Asn179 (gray, sticks) in cyan (sticks). A red arrow indicated residues forming the Tyr-corner. Residues common to the hydrophobic core of FNIIIs are boxed. (**B**) Zoom into the N-terminal face loop cluster. Top: The H-bonding pattern of Asn179 (asterisk) is central to this loop cluster. Three highly conserved segments are inter-connected (see schematic representation in the bottom panel) and form a polar extension to the hydrophobic core. A similar tight organization of the N-terminal loop regions has also been observed in the so-called "N-conserved type" IG modules of titin (Marino *et al.*, 2005).

3.4.3b The interface between A77-A78

A77-A78 shows no free linker sequence, like *e.g.* Z1-Z2. The C-terminal border of A77 (Ser100) and the N-terminal border of A78 (Glu101) are in the center of the interface, so that the two domains are effectively connected by a zero-length linker. However, titin's FNIII transitions have variable linker lengths according to the position in the super-repeat (details in section 3.4.4b). Therefore, the "linker" will be defined here as the sequence downstream of the last secondary structure element of A77 to the first highly conserved Pro in A78: Ala99-Ser100-Glu101-Arg102 (for A77-A78). The central residues in this linker (Ser100 and Glu101) form H-bonds to all loop regions pointing into the interface (AB-turn/EF-loop in A77 and BC-loop/FG-turn in A78). There are also direct interactions between the modules. A tight H-bonding pattern (Figure 3.18; additional details in Table 6.1 in appendix) dominates the interaction pattern of the interface, although there is also a significantly low hydrophobic contribution (residues indicated in Table 3.10). The interface between A77 and A78 comprises a core and a rim region (determined with ProFace; Saha *et al.*, 2006) with three patches Table 3.10.

Table 3.10 Core and rim regions in the A77-A78 interface[1].

A77 (patch1)	A78 module border and linker (patch 2)	A78 (patch 3)
(1) **Val17-Thr-Lys19**	(2) **Val97-Lys98**	(3) **His128**-<u>Asp</u>-**Gly**-<u>Gly</u>-**Ser132**
(1) Glu71-Gly72	(2) **Ser100-Glu101**[2]	(3) Glu180-**Lys181**

[1]core regions are written in bold face and residues contributing exclusively by hydrophobic interactions are underlined
[2]Ala99 and Arg102, which have not been included into domain borders in the calculations, contribute to the interface by hydrophobic interactions and H-bonds. Ala99 and Arg102 have solvent accessibilities below 10% and 50%, respectively
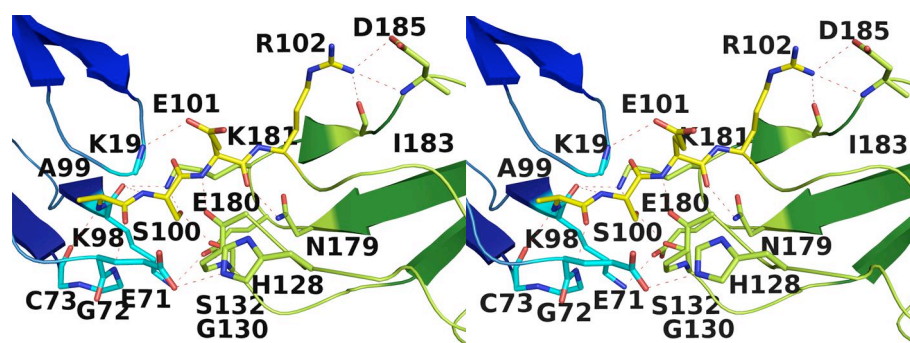
Figure 3.18 The A77-A78 interface.
A stereo view of the interface region between A77 (blue) and A78 (green) and the inter-connecting linker. The convex plane of the tandem is placed at the bottom. To give a better overview only involved parts of the residues forming H-bonds are given in stick representation (indicated by labeling, water molecules are not shown). Red dashed lines indicate H-bonds.

The interface resembles hetero-obligomer interfaces, which are depleted in hydrophobic interactions. This in contrast to other FNIII transitions (*e.g.* fibronectin, Leahy *et al.,* 1996) where predominantly hydrophobic interactions involving a low number of loops dominate the domain-domain interface. The tight H-bond network in the A77-A78 transition indicates that this should be a very stiff domain arrangement, likely to result in conserved, defined inter-domain orientations for positions two and three of the long super-repeat. This is supported by the fact that the interface of A77 and A78 reveals strictly conserved regions (Asp129-Gly-Gly130 in the BC-loop and Asn179/Gly182 in the FG1-turn of A78) besides others with lower conservation (AB-turn and EF-loop in A78 and the linker) (indicated in Figure 3.16). The bent, semi-extended conformation of A77-A78 results in distinct faces with different accessibilities and surface conservation, which may have a role in higher complexation, *e.g.* in myosin matrix assembly, which is likely to require similar tandem arrangements in domains with same functions. It could be speculated that the strong interaction pattern in the A77-A78 interface allows for variability in individual domains as long as the overall arrangement retains the function of the transition. The interaction density in A77-78 is likely to show higher tolerance against loss of single interactions than low interaction density transitions (*e.g.* FNIII transitions in the cell attachment module from mouse, Copie *et al.,* 1998).

3.4.4   Orientations of FNIII transitions in titin

The inter-domain orientation and dynamics of FNIII modules within tandems in titin must depend on the length of interface loops and linker sequences as well as the character of the interactions (polar *vs.* hydrophobic).

3.4.4a   A77-A78 and extracellular FNIII transitions

The superposition of A77-A78 to other tandems gave best results for three FNIII transitions (Table 3.11). Superposition of the tandems reveals a BC-loop with variable length (longest in A78), whereas the FG1-turn has similar length in all modules. In general, tandems with greater conservation in the N-terminal face and linker show higher similarity to the observed arrangement in A77-78.

Table 3.11 Superposition of other FNIII tandems to A78.

PDB code	*rmsd* [Å]	C_α-atoms	Description
1FNH	1.0	76	Heparin and integrin binding fragment of human fibronectin[1]
1QR4	1.1	78	Tandem from chicken tenascin
1QG3	1.3	77	Cytoplasmic tail of human integrin $\alpha6\beta4$

[1]modules two and three

The tandem arrangement suggests that the orientation of the linker and the neighboring module is determined by the structural features of the BC-loop, and to a minor extent of the EF-loop that is in close vicinity (Figure 3.19). Additionally, the conformation of the FG-turn shows dependence on the presence of the residue equivalent to the conserved Asn in the polar sequence motif of this loop - in modules lacking this residue, the FG1-turn orients away from the linker. This is combined with the lack of interaction between the FG1-turn and the linker, which results in further orientation of the linker an adjacent modules into the BC-loop direction.
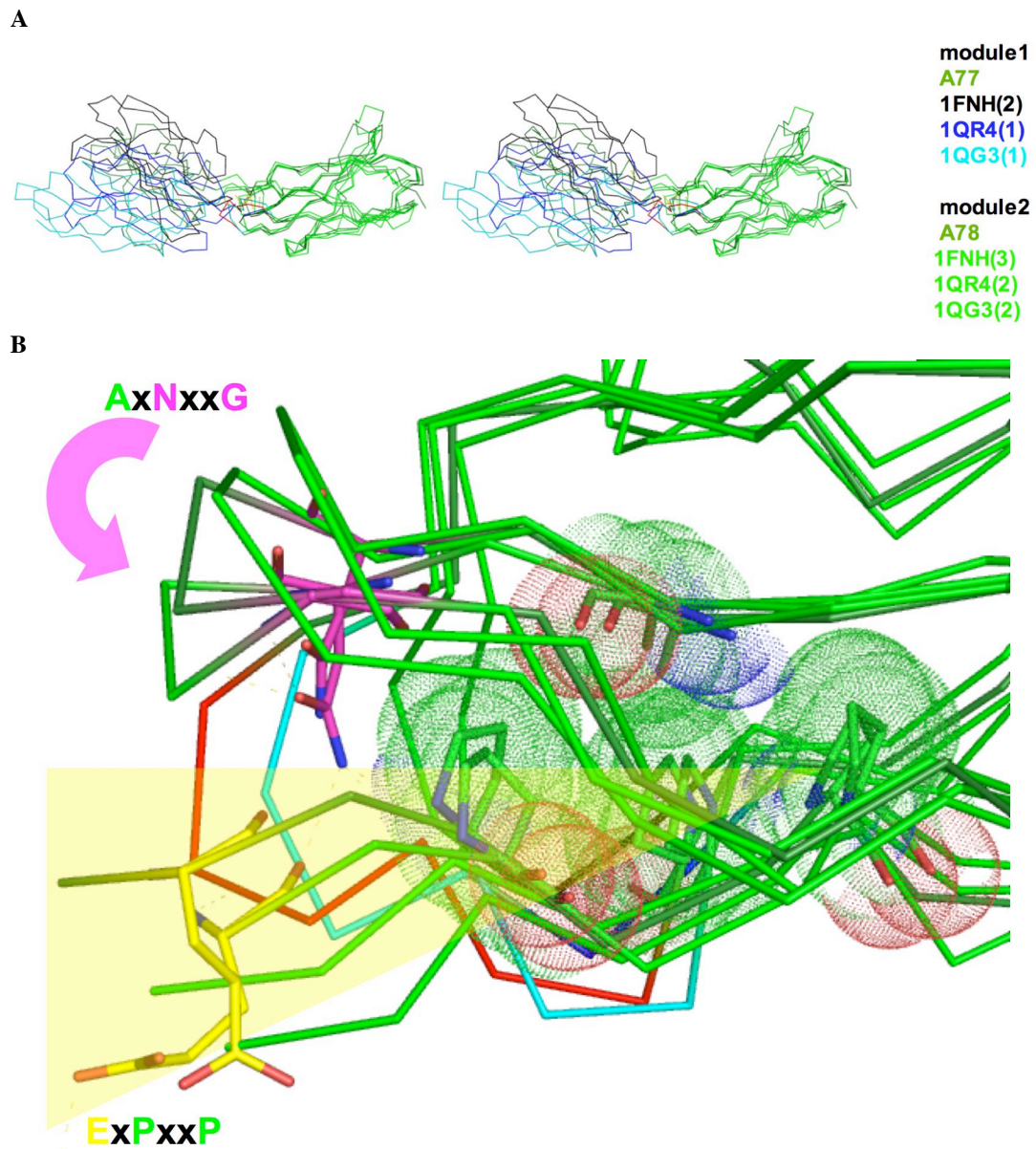
Figure 3.19 Comparison of A77-A78 to other tandems.
(**A**) Stereo view of A77-A78 superposed to other FNIII tandems. On the N-terminal side, A78 (dark green) shows the longest loops compared to the other modules (light green), while C-terminal sides show similar loop lengths. Especially the BC-loop (red in A78 and in the color of the first module in the others) is truncated. The first modules seem to "fall" into this direction suggesting the length of the BC-loop to be an effector of tandem organization. (**B**) Similar motifs on the N-terminal side result in similar organization of the corresponding elements. Structures without the two conserved motifs (as indicated in Figure 3.17 and shown as sticks here) seem to lack the stiff arrangement of the N-terminal side with subsequent loss of the interconnection of the corresponding elements.

Among the investigated FNIII tandems, the cytoplasmic tail of human integrin α6β4 (1QG3) reveals the highest similarity to A78. This includes a BC-loop of similar length (with low conservation) as well as the presence of conserved motifs in the linker and FG1-turn. The corresponding interactions result in an analogous

arrangement of the N-terminal face (extended hydrophobic core as well as inter-connection of the linker, BC- and FG1-loops). Even a similar interaction pattern across the surface is formed (numbering for A78): the side-chain of Asn179 (FG1-turn) interacts with the main-chain of Glu101 (linker), whose side-chain interacts with the main-chain of Lys19 (AB-turn). In general, the organization of structural elements, as observed in A78, seems to define the orientation of the module upstream.

3.4.4b The linkers between FNIII modules in titin

The linker regions in FNIII transitions of titin have not yet been investigated in detail. However, the domain positions reveal distinct linker lengths and conservation (Figure 3.20 and the corresponding sequence alignments in Figure 3.21). The studies by Muhle-Goll and coworkers focused on conservation of individual modules (front- and back-side). The study of the modular transitions, with their corresponding interaction patterns, makes it possible to draw conclusions on the relation between linker composition and transition rigidity.
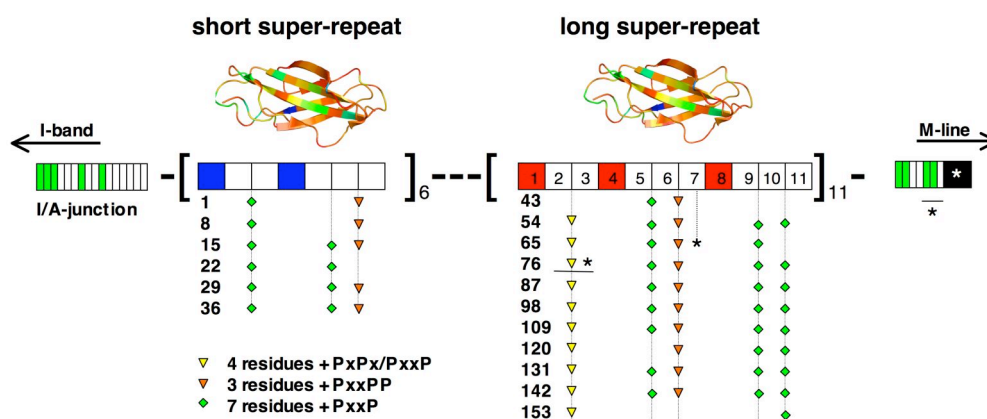
Figure 3.20 The A-band region of titin and conservation of FNIII modules in the super-repeats.

The sheme shows the I/A-junction, followed by the short super-repeat, the long super-repeat and the M-line region (left to right). FNIII modules are colored in white; IG modules belonging to the short and long super-repeat are colored blue and red, respectively. Other IG domains are colored in green. The titin kinase is colored in black. On top of the A-band scheme the conservation of the super-repeats is labeled on the structure of A77. Below the A-band scheme there is an overview of linker lengths between FNIII domains. An asterisk indicates modules for which experimental structures are available. Outliers concerning linker composition lack the triangles/squares in the lower panel.

The transitions between FNIII modules of the long super-repeat reveal distinct linker lengths and compositions (Figure 3.20):

(i)   Transition between modules two and three (with A77-78 as a representative): short linkers compared to other transitions and a Pro-x-Pro-x or Pro-x-x-Pro motif. Glx residue as the N-terminal domain border.

```
transition    C-terminal face      linker/Pro-rich    N-terminal face
A44-A45   NEYGPGVPTDVPKPVL  ASDPLSEPDPP  RKLEATEMTKNSATLAWLPPLRDGG  52
A55-A56   NEYGIGDPGETRDAVK  AS---QTPGPV  VDLKVRSVSKSSCSIGWKKPHSDGG  49
A66-A67   NEYGIGEPTETTEPVK  AS---EAPSPP  DSLNIMDITKSTVSLAWPKPKHDGG  49
A77-A78   NEYGIGLPAETAESVK  AS---ERPLPP  GKITLMDVTRNSVSLSWEKPEHDGG  49
A88-A89   NEYGIGLPAQTADPIK  VA---EVPQPP  GKITVDDVTRNSVSLSWTKPEHDGG  49
A99-A100  NEFGVGVPVETVDAVK  AA---EPPSPP  GKVTLTDVSQTSASLMWEKPEHDGG  49
A110-A111 NEYGIGLPAETTEPVK  VS---EPPLPP  GRVTLVDVTRNTATIKWEKPESDGG  49
A121-A122 NEFGVGIPAETKEGVK  IT---EKPSPP  EKLGVTSISKDSVSLTWLKPEHDGG  49
A132-A133 NEIGIGEPCETTEPVK  AA---EVPAPI  RDLSMKDSTKTSVILSWTKPDFDGG  49
A143-A144 NEYGVGEPYEMPEPIV  AT---EQPAPP  RRLDVVDTSKSSAVLAWLKPDHDGG  49
A154-A155 NIYGIGEPCETSDAVL  VS---EVPLVP  AKLEVVDVTKSTVTLAWEKPLYDGG  49
```

Figure 3.21a Linker region between modules two and three in the long super-repeat.
The alignment shows the region between the conserved Ala-x-Asn-x-x-Gly in the FG1-turn (first module) and the Pro-x-x-Asp-Gly-Gly motif in the BC-loop. The domain borders are indicated by gray and the Pro-rich motif by yellow background, respectively. Regions belonging to the modules (based on A77-A78 domain borders) are underlined.

(ii)   Transition between modules six and seven (with A71 as a structural representative for position seven): shorter linker length than in (i) and an extended Pro-rich motif (Pro-x-x-Pro-Pro). Preference for Asx as the domain border (equivalent to Glx in (i)).

```
transition    C-terminal face      linker/Pro-rich    N-terminal face
A48-A49   NKAGSSPPSKPTEYVT  AR---MPVDPP  GKPEVIDVTKSTVSLIWARPKHDGG  49
A59-A60   NLAGVGKPSLPSEPVV  AL---DPIDPP  GKPEVINITRNSVTLIWTEPKYDGG  49
A70-A71   NMAGKSKPSKPSEPML  AL---DPIDPP  GKPVPLNITRHTVTLKWAKPEYTGG  49
A81-A82   NIVGIGKPSKVSECYV  AR---DPCDPP  GRPEAIIVTRNSVTLQWKKPTYDGG  49
A92-A93   NIVGVGKASKNSECYV  AR---DPCDPP  GTPEPIMVKRNEITLQWTKPVYDGG  49
A103-A104 NIVGIGKPSKVSECFV  AR---DPCDPP  GRPEAIVITRNNVTLKWKKPAYDGG  49
A114-A115 NIAGIGKCSKSCEPVP  AR---DPCDPP  GQPEVTNITRKSVSLKWSKPHYDGG  49
A125-A126 NAAGVGKPSHPSEPVL  AI---DACEPP  RNVRITDISKNSVSLSWQQPAFDGG  49
A136-A137 NSAGLSSPSDPSKFTL  AV---SPVDPP  GTPDYIDVTRETITLKWNPPLRDGG  49
A148-A149 NAAGVSKASEASRPIM  AQ---NPVDAP  GRPEVTDVTRSTVSLIWSAPAYDGG  49
A158-A159 NARGSGKPSRPSKPIV  AMDPIAPPGKP  QNPRVTDTTRTSVSLAWSVPEDEGG  52
```

Figure 3.21b Linker region between modules six and seven in the long super-repeat.
The alignment region, the domain borders and the underlined regions are analogous to Figure 3.21a. The Pro-rich motif is indicated by red background.

(iii)   Remaining FNIII transitions, namely five to six, nine to ten and ten to eleven: long linkers and a Pro-x-x-Pro motif. The Pro residues are integral part of the FNIII modules and not of the linkers, so that they should not influence the conformational freedom of the inter-domain arrangement.

A

```
transition    C-terminal face    linker/Pro-rich    N-terminal face

A47-A48   NRFGISEPLT-SPKMV  AQFPFGV PSEP  KNARVTKVNKDCIFVAWDRPDSDGG  51
A58-A59   NKYGVGEGLK-SEPIV  ARHPFDV PDAP  PPPNIVDVRHDSVSLTWTDPKKTGG  51
A69-A70   NRYGVGPGIT-SAWIV  ANYPFKV PGPP  GTPQVTAVTKDSMTISWHEPLSDGG  51
A80-A81   NRYGKSTYLN-SEPTV  AQYPFKV PGPP  GTPVVTLSSRDSMEVQWNEPISDGG  51
A91-A92   NRYGQSFALE-SDPIV  AQYPYKE PGPP  GTPFATAISKDSMVIQWHEPVNNGG  51
A102-A103 NRYGKSAPLD-SKAVI  VQYPFKE PGPP  GTPFVTSISKDQMLVQWHEPVNDGG  51
A113-A114 NRYGKSSYSE-SSAVV  AEYPFSP PGPP  GTPKVVHATKSTMLVTWQVPVNDGG  51
A124-A125 NRFGISDHID-SACVT  VKLPYTT PGPP  STPWVTNVTRESITVGWHEPVSNGG  51
A135-A136 NRYGVSQPLV-SSIIV  AKHQFRI PGPP  GKPVIYNVTSDGMSLTWDAPVYDGG  51
A146-A147 NRFGVSKPLE-SAPII  AEHPFVP PSAP  TRPEVYHVSANAMSIRWEEPYHDGG  51
A157-A158 NRFGIGSYLQ-SEVIE  CRSSIRI PGPP  ETLQIFDVSRDGMTLTWYPPEDDGG  51
```

B

```
transition    C-terminal face    linker/Pro-rich    N-terminal face

A51-A52   NKYGVGDPVF-TEPAI  AKNPYDP PGRC  DPPVISNITKDHMTVSWKPPADDGG  51
A62-A63   NHYGKEPVQ-SEPVK   MVDRFGP PGPP  EKPEVSNVTKNTATVSWKRPVDDGG  51
A73-A74   NKYGVGEPLE-SEPVL  AVNPYGP PDPP  KNPEVTTITKDSMVVCWGHPDSDGG  51
A84-A85   NKYGVGEPLE-SEPVV  AKNPFVV PDAP  KAPEVTTVTKDSMIVVWERPASDGG  51
A95-A96   NKYGVGEPLE-SAPVL  MKNPFVL PGPP  KSLEVTNIAKDSMTVCWNRPDSDGG  51
A106-A107 NKYGIGEPLE-SGPVT  ACNPYKP PGPP  STPEVSAITKDSMVVTWARPVDDGG  51
A117-A118 NKYGVGEPLE-SVAIK  ALDPFTV PSPP  TSLEITSVTKESMTLCWSRPESDGG  51
A128-A129 NKYGIGEPLE-SDSVV  AKNAFVT PGPP  GIPEVTKITKNSMTVVWSRPIADGG  51
A139-A140 NKFGVGRPLD-SDPVV  AQIQYTV PDAP  GIPEPSNITGNSITLTWARPESDGG  51
A150-A151 NKYGPGVPVE-SEPIV  ARNSFTI PSPP  GIPEEVGTGKEHIIIQWTKPESDGG  51
A161-A162 NQFGISKPLKSEEPVT  PKTPLNP PEPP  SNPPEVLDVTKSSVSLSWSRPKDDGG 53
```

C

```
transition    C-terminal face    linker/Pro-rich    N-terminal face

A52-A53   NKAGPGKPSDASKAAY  ARDPQYP PAPP  AFPKVYDTTRSSVSLSWGKPAYDGG  52
A74-A75   NAAGISAPSPTSPFYK  ACDTVFK PGPP  GNPRVLDTSRSSISIAWNKPIYDGG  52
A63-A64   NRAGIGPPSEASDSVL  MKDAAYP PGPP  SNPHVTDTTKKSASLAWGKPHYDGG  52
A85-A86   NAAGLSEPSPPSAYQK  ACDPIYK PGPP  NNPKVIDITRSSVFLSWSKPIYDGG  52
A96-A97   NAAGVGEPSPATVYYK  ACDPVFK PGPP  TNAHIVDTTKNSITLAWGKPIYDGG  52
A107-A108 NAAGVGEPSEPSVFYR  ACDALYP PGPP  SNPKVTDTSRSSVSLAWSKPIYDGG  52
A118-A119 NAAGLSLPSETSPLIR  AEDPVFL PSPP  SKPKIVDSGKTTITIAWVKPLFDGG  52
A129-A130 NAAGQGPFSEPSEFYK  AADPIDP PGPP  AKIRIADSTKSSITLGWSKPVYDGG  52
A140-A141 NAAGVGPASGISRLIK  CREPVNP PGPP  TVVKVTDTSKTTVSLEWSKPVFDGG  52
A151-A152 NAAGNSEPSERSNFIS  CREPSYT PGPP  SAPRVVDTTKHSISLAWTKPMYDGG  52
A162-A163 NDVGLSETSPASEPVV  CKDPFDK PSQP  GELEILSISKDSVTLQWEKPECDGG  52
```

Figure 3.21c Linker regions of the long super-repeat with a long linker.
The alignment region, the domain borders and the underlined regions are analogous to Figure 3.21a. The Pro-rich motif is indicated by green background. The figure shows the transitions from the long super-repeat positions five to six (**A**), nine to ten (**B**) and ten to eleven (**C**).

Analogous to the linker regions of the long super repeat, the transitions of the short super-repeat can be aligned (Figure 3.22). Surprisingly, the linker lengths in the transition between modules two and three as well as six and seven reveal a long linker, which appears to be related to transitions of type (iii).

A

transition	C-terminal face	linker/Pro-rich	N-terminal face	
A2–A3	NKVGGGEYIELKNPVI	AQDPKQP**PDPP**	VDVEVHNPTAEAMTITWKPPLYDGG	52
A9–A10	NKFGCGPPVEIG-PIL	AVDPLGP**PTSP**	ERLTYTERQRSTITLDWKEPRSNGG	51
A16–A17	NRFGIGPPVETIQRTT	ARDPIYP**PDPP**	IKLKIGLITKNTVHLSWKPPKNDGG	52
A23–A24	NSIGMGPFVETSEALV	IREPITV**PERP**	EDLEVKEVTKNTVTLTWNPPKYDGG	52
A30–A31	NQYGRGPFVETPKPIK	ALDPLHP**PGPP**	KDLHHVDVDKTEVSLVWNKPDRDGG	52
A37–A38	NKVGVGPTIETKTPIL	AINPIDR**PGEP**	ENLHIADKGKTFVYLKWRRPDYDGG	52

B

transition	C-terminal face	linker/Pro-rich	N-terminal face	
A5–A6	NRFGPGPPCVSKPLV	AKDPFGP**PDAP**	DKPIVEDVTSNSMLVKWNEP-KDNG	50
A12–A13	NKYGISDECKSDKVV	IQDPYRL**PGPP**	GKPKVLARTKGSMLVSWTPPLDNGG	51
A19–A20	NKYGIGEPLDSEPET	ARNLFSV**PGAP**	DKPTVSSVTRNSMTVNWEEPEYDGG	51
A26–A27	NLYGISDPLVSDSMK	AKDRFRV**PDAP**	DQPIVTEVTKDSALVTWNKP-HDGG	50
A33–A34	NKIGVGPPLDSTPTV	AKHKFSP**PSPP**	GKPVVTDITENAATVSWTLPKSDGG	51
A40–A41	NKIGTGPPTESKPVI	AKTKYDK**PGRP**	DPPEVTKVSKEEMTVVWNPPEYDGG	51

C

transition	C-terminal face	linker/Pro-rich	N-terminal face	
A6–A7	NAAGPGKFSPPSDPKT	AHDPIS**PPGPP**	I-PRVTDTSSTTIELEWEPPAFNGG	51
A13–A14	NAAGIGPPSEPSDPEV	AGDPIF**PPGPP**	SCPEVKDKTKSSISLGWKPPAKDGG	52
A20–A21	NAAGVGPASLPSDPAT	ARDPIA**PPGPP**	-FPKVTDWTKSSADLEWSPPLKDGG	51
A27–A28	NEIGIGDPSPPSKPVF	AKDPIAKPSPP	VNPEAIDTTCNSVDLTWQPPRHDGG	52
A34–A35	NLAGLSKPSPSSDPIK	ACRPIK**PPGPP**	INPKLKDKSRETADLVWTKPLSDGG	52
A41–A42	NEIGIGEPSLPSRPVV	AKDPIE**PPGPP**	TNFRVVDTTKHSITLGWGKPVYDGG	52

Figure 3.22 FNIII transitions in the short super-repeat.
The alignment region, the domain borders and the underlined regions of short super-repeat modules are analogous to Figure 3.21a. The Pro-rich motif of the modules is indicated by green (compare Figure 3.21c) and red (compare Figure 3.21b) background. The figure shows the transitions from the short super-repeat positions two to three (**A**), five to six (**B**) and six to seven (**C**).

Modular connections dictate the conformational freedom of the titin chain, where the modules act as rigid building blocks. The current analysis suggests that the distinct conservation patterns of linker sequences and domain borders create defined interfaces at specific super-repeat positions. If the interaction pattern between the EF-loop of one module and the BC-loop of the next was conserved, elongation of the linker might either linearize the tandem conformation or weaken the modular interactions with a subsequent introduction of torsional freedom in the tandem, as observed in the IG-doublet Z1-Z2 (Marino et al., 2006).

The transition in positions two and three of the long-super repeat appears to be the most rigid. The positioning of a core residue in the linker (Glu101 in A77) may have the role of a selectivity filter in tandem pre-arrangement. Additional interactions between the domain ends and the linker could then support this initial interaction pattern and form super-repeat position dependent tandem arrangements.

The evolutionary relationship between the short and long super-repeat is, that the latter was created by gene duplication of the short super-repeat (Kenny *et al.,*

1999). The N-terminal Pro-rich motifs of equivalent transitions between the long and short super-repeat show similar sequence patterns, while the corresponding linkers have different lengths. The similar linker lengths of equivalent long super-repeat transitions suggest an evolutionary pathway from an eleven domain repeat prototype (as indicated in Figure 3.7) combined with linker truncations (Figure 3.23).
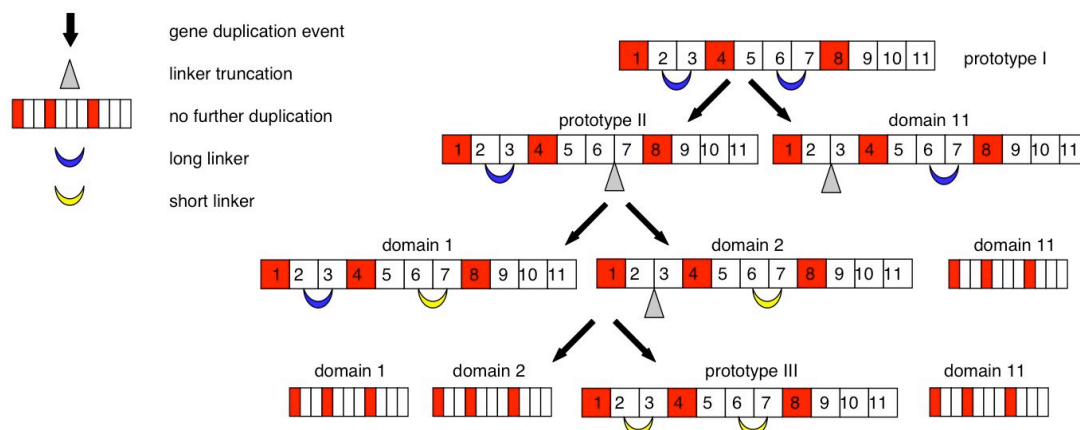
Figure 3.23 Model for linker truncation in the evolution of the long super-repeat.
The evolutionary pathway was adapted from Kenny *et al.*, 1999. The linker lengths suggest that prototype III has evolved from domain 2. According to Kenny and coworkers, the remainder of the long super-repeat has evolved from prototype III

3.4.5   The A76-A78 homology model

In the structure of A168-A170 (Mrosek *et al.*, 2007), an IG-IG-FNIII construct from the P-zone of human titin, the hydrophobic IG-FNIII interface involves a conserved motif (Asn-x-x-Gly in the FG-turn) in long-super repeat FNIIIs downstream to IG modules (illustrated in Figure 6.7 in the appendix). The putative interface residues in the tandem A76-A77 (IG-FNIII) share high conservation with A169-A170, which makes a similar interaction pattern likely and made it possible to produce a homology model for the tandem A76-A77, spanning the first three positions of domain four of the long super-repeat. For this purpose, the modules A77 and A170 were superposed using TOP (*rmsd* = 0.9 Å over 91 residues, identity 45 %). The resulting hypothetical transition A169-A77-A78 was used as a template and was subjected to homology modeling using SWISS MODEL (Guex & Peitsch, 1997; Schwede *et al.*, 2003). The model reveals a zigzag arrangement.

127

In A169, the FG-turn (the Asn corresponding to Asn179 in A77)) forms an H-bond to an Asp in the linker, which is equivalent to the interaction observed in A77-A78. The side-chain of the Asp in the linker forms a H-bond to the side-chain of an Arg in the A'B-loop of the IG-module. This pattern is similar to the interactions described in sections 3.4.3b & 3.4.4a, suggesting that conformational fixation of the main-chain in the FNIII N-terminal face creates a selectivity filter, which may be involved in pre-orientation of transitions. Additional residues in the interface may stabilize it with subsequent orientation For IG-FNIII transitions, the Asn-x1-x2-Gly motif in the FG-turn preferentially involves an acidic/basic residue (for x1) and an aromatic residue (for x2). An interaction pattern mediated by conserved residues suggests the observed zigzag arrangement to be applicable to analogous IG-FNIII transitions in the long super-repeat.
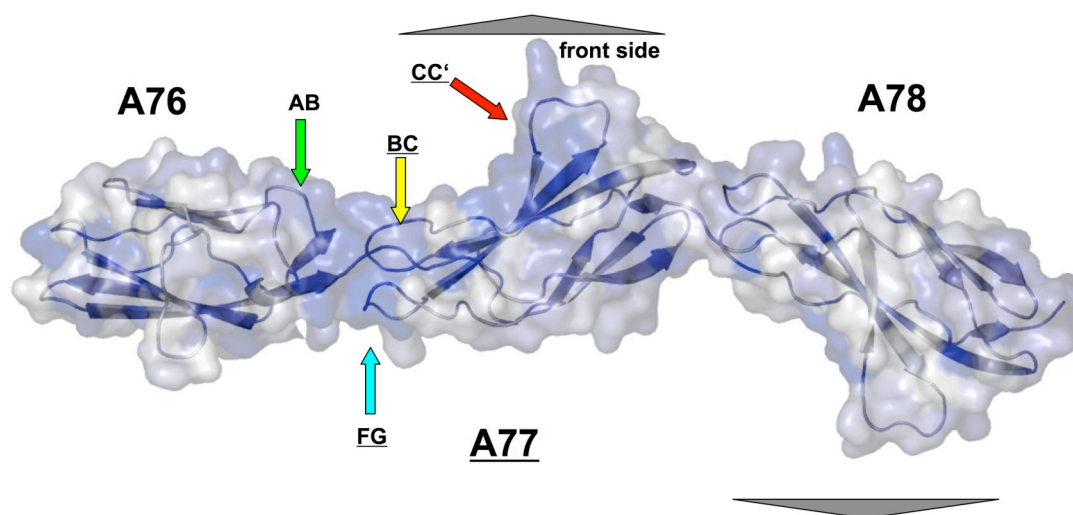
Figure 3.24 Homology model of A76-A78.
The modules are colored after their conservation in the first three positions of super-repeat domains from white (low conservation) to blue (high conservation). The AB-turn of the IG A76, which forms interactions to the linker is indicated by a green arrow. In A77, three highly conserved loop regions are indicated by yellow (BC-loop), red (CC'-loop) and cyan (FG1-turn) arrows. The gray triangles point away from the front sides of both FNIII modules.

The zigzag arrangement places the CC'-loop on the front-side (in the C-terminal moiety) of the module on the convex plane of each transition. This organization gives the CC'-loop high accessibility to other molecules and conserved patches in FNIII modules form an almost continuous stretch. This highly accessible stretch is likely to have a role in interface formation between titin and other proteins.
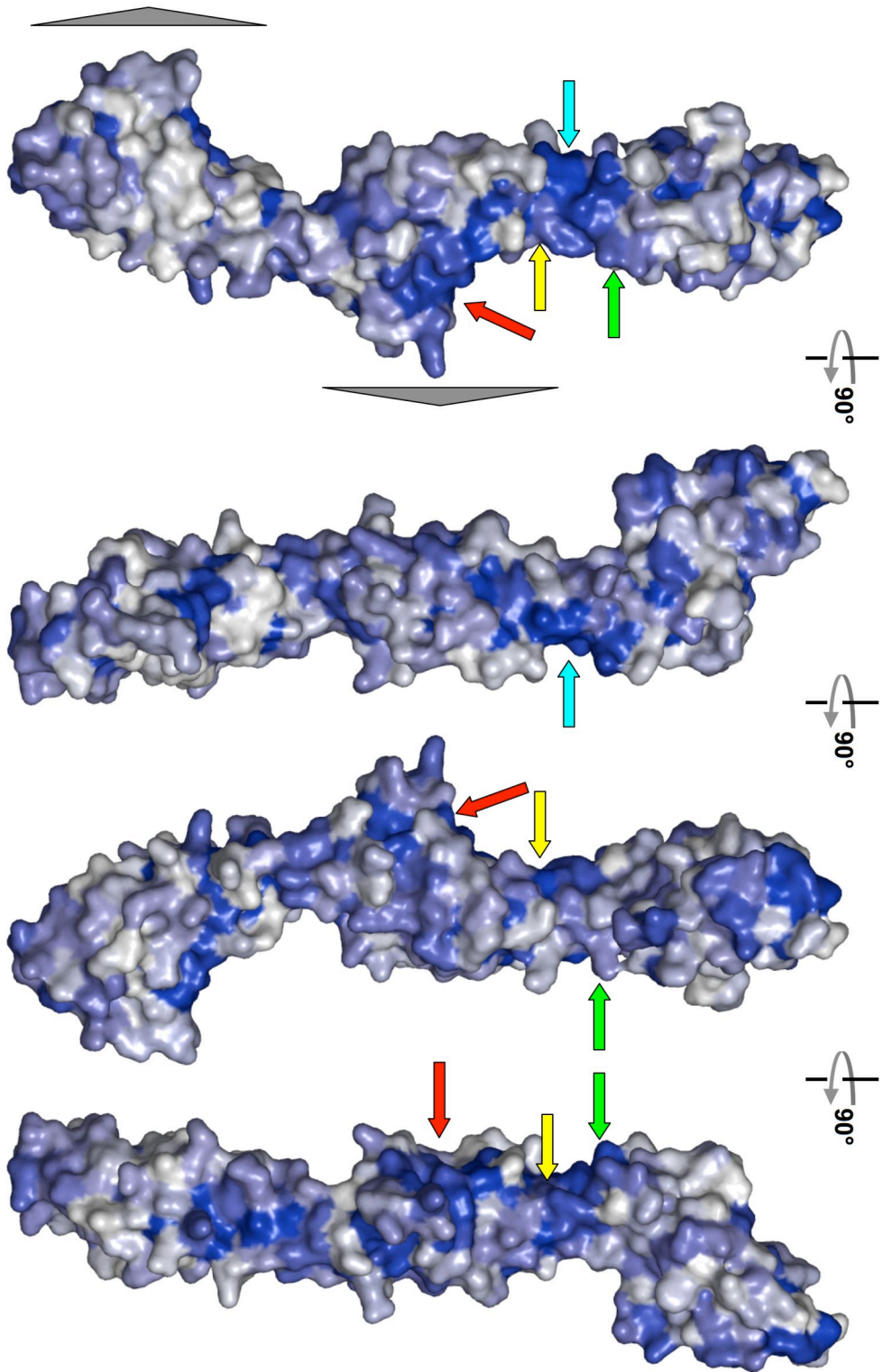
Figure 3.25 The conservation in long super-repeat modules one to three.
The coloring in this figure is analogous to the previous (without transparency to illustrate the surface conservation of the tandem). The most left view corresponds to the one shown in the previous figure.

3.5 Summary

The structure of the construct A77-A78 is the first structure of a FNIII tandem from human titin and the first atomic structure of an intracellular FNIII transition. Each of the modules of A77-A78 comprises a FNIII fold with a hydrophobic core characteristic of FNIII modules. This core is extended into the loop region of the N-terminal end of the domain. In this extension, a buried Asn residue interconnects the loops and the linker. These interactions are likely to introduce additional rigidity, making each FNIII module of titin a stiff building block.

In the unusual polar transition interface, a remarkable number of loops take place in inter-domain contacts and the linker by H-bond formation. The predominantly polar interface in combination with the high stiffness of the interface loops suggests that this conformation is highly defined and conserved across equivalent positions of the long super-repeat. Based on super-repeat position dependent conservation other transitions may comprise their own, discrete arrangements with distinct conformational freedom.

A hypothetical model for the A76-A78 tandem based on sequence conservation at the domain interface shows a zigzag, extended arrangement of domains. This conformation results in maximal surface exposure of conserved residues on the CC'-loop on the front-side of the modules.

The A77-A78 (and, in general, all transitions in positions two and three in the long super-repeat) is highly rigid and represents the tightest possible FNIII transition in titin. The crowded polar interface does not appear to allow for a detectable conformational freedom. This might directly reflect the putative role of this tandem in the creation of a defined template to modulate the assembly of the myosin-based thick filament (hetero-obligomer type). As in the case of Z1Z2 (Marino et al., 2006) and A168-A170 (Mrosek et al., 2007), a pre-existent defined conformation appears to help the recruitment of other molecules onto the surface of the titin filament, probably by means of steric effects. Due to the evolutionary relationship of the domains, the investigation of a distinct set of tandems should allow the production of a model for the whole A-band, which can help to understand the role of titin in the assembly of the thick filament.

After the completion of this study, SAXS data became available that confirm that the crystal structure of A77-A78 corresponds to the structure of the tandem in solution. Furthermore, an NMR study confirmed the interaction of A77-A78 with myosin heads and contact mapping experiments are currently under consideration.

4 Dystrophia Myotonica Protein Kinase

Dystrophia Myotonica Protein Kinase (DMPK) and related members of the family are complex multi-domain kinases that become activated by oligomerization. The self-assembly of DMPK (and related kinases) was thought to be mediated by a C-terminal coiled-coil fraction. However, contradictory biochemical data reported for different members of the family questioned the extent to which these kinases might share a general oligomerization principle.

At the time when this project was initiated, no structural data existed for any member of the family. A structural and biophysical study on DMPK was then undertaken in our lab that comprised full-length DMPK as well as truncated variants. Within this study, the work on a C-terminally truncated form lacking the coiled-coil domain, DMPKΔVR/CC, is my own. Crystallization trials of this form yielded no crystals suitable for X-ray data collection since protein samples did not show the long-term homogeneity critical to this process. A characterization of the oligomeric state of this construct is reported in:

Garcia P., Ucurum Z., Bucher R., Svergun D.I., Huber T., Lustig A., Konarev P.V., Marino M. and O. Mayans (2006). "Molecular insights into the self-assembly mechanism of dystrophia myotonica kinase." FASEB J. 20: 1142-1151

Molecular insights into the self-assembly mechanism of dystrophia myotonica kinase

Pilar Garcia,* Zöhre Ucurum,* Rainer Bucher,* Dmitri I. Svergun,[†] Thomas Huber,[‡]
Ariel Lustig,* Petr V. Konarev,[†] Marco Marino,* Olga Mayans*,[1]

*Division of Structural Biology, Biozentrum, University of Basel, Basel, Switzerland; [†]European
Molecular Biology Laboratory, Hamburg Outstation, c/o DESY, Hamburg, Germany and Institute of
Crystallography, Russian Academy of Sciences, Moscow, Russia; and [‡]Biochemisches Institut,
Universität Zürich, Zürich, Switzerland

ABSTRACT   Self-assembly via coiled-coil domains
(CC) is crucial for the regulation of the dystrophia
myotonica kinase (DMPK) -related family of kinases.
These CC domains are thought to form dimeric ar-
rangements and thus to mediate dimerization in these
enzymes. Using size exclusion chromatography com-
bined with multiangle static light scattering, we ana-
lyzed the oligomeric state of DMPK as well as that of a
truncated variant lacking the CC fraction. Remarkably,
both forms were found to assemble into robust dimers.
In contrast, the CC domain in isolation yielded trimeric
assemblies, indicating that the oligomerization proper-
ties of CC domains from this kinase family are more
diversified than anticipated. The crystal structure of
this CC has been elucidated to 1.6 Å resolution and its
properties in solution established using sedimentation
equilibrium and thermal denaturation. These data show
that, contrary to expectations, the self-assembly of
DMPK is not dictated by the association properties of
its CC domain. Instead, it appears to be driven by
sequence segments flanking both N and C termini of
the catalytic kinase fraction, as suggested by models of
head-to-head dimers based on small angle X-ray scatter-
ing data. Our findings support a shared pattern of
assembly across DMPK, ROCKs, and MRCK members
of this family.—Garcia, P., Ucurum, Z., Bucher, R.,
Svergun, D. I., Huber, T., Lustig, A., Konarev, P. V.,
Marino, M., Mayans, O. Molecular insights into the
self-assembly mechanism of dystrophia myotonica ki-
nase. *FASEB J.* 20, 1142–1151 (2006)

Key Words: *kinase · coiled-coil · protein assembly · protein
structure · X-ray crystallography · multiangle static light
scattering · small angle X-ray scattering*

SELF-ASSEMBLY IS A MAIN feature of the dystrophia
myotonica kinase (DMPK) -related family of kinases.
These complex multidomain kinases share a highly
conserved catalytic domain and a coiled-coil fraction
(CC) involved in the regulation of their activity. To this
group belong DMPK (1, 2), Rho-associated kinases I
and II (ROCKI/ II, also known as β and α, respectively)
(3, 4), *C. elegans* LET-502 (5), *Drosophila* Genghis Khan
kinase (Gek) (6), murine citron kinase (7), and rat

Cdc42 binding kinase (MRCK) (8), among others.
Most of these kinases act as effectors of small Rho
GTPases (Rho, Rac, and Cdc42 subclasses) known to
regulate cytoskeletal dynamics. Accordingly, several
members of the family are functionally linked to cy-
toskeletal reorganization events, as illustrated by the
involvement of Gek in actin polymerization (6) and the
role of citron kinase, LET-502, and ROCKII in regulat-
ing the cleavage furrow contraction during cytokinesis
(9–11).

DMPK is regarded as the defining member of this
kinase family. It is most closely related to MRCK and
ROCKs, with which it shares sequence similarity across
the N-terminal catalytic fraction but not across CC
moieties (**Fig. 1**). By analogy to other family members,
DMPK is suspected to regulate cytoskeletal events (12,
13) and may act as a downstream effector in the
regulatory pathways of RhoA (12) and Rac1 (14). Other
reports, however, defend that DMPK intervenes in the
modulation of the electrical excitability of the plasma
membrane (15, 16). Taken together, functional data
suggest that DMPK might be related to muscle stiffness
and hyperexcitability, heart conduction defects, and
cognitive deficit in a progressive neuromuscular disor-
der known as myotonic dystrophy (DM1), which is the
most common form of muscular dystrophy in adults
(17). DM1 patients suffer from DMPK insufficiency due
to the retention of mRNA transcripts in the nucleus
(18, 19). However, it is not currently understood how
low levels of this kinase affect cellular function, and
thus its involvement in disease remains polemic.

Oligomerization via CC segments is thought to be
central to the regulation of DMPK-related kinases. CC
motifs consist of two to five amphipathic α-helices that
wind around each other to form, typically, a left-
handed supercoil. They mediate self-assembly events in
proteins but can also be implicated in refolding pro-
cesses and molecular recognition (for a recent review,
see ref 20). CC sequences are characterized by a repeat
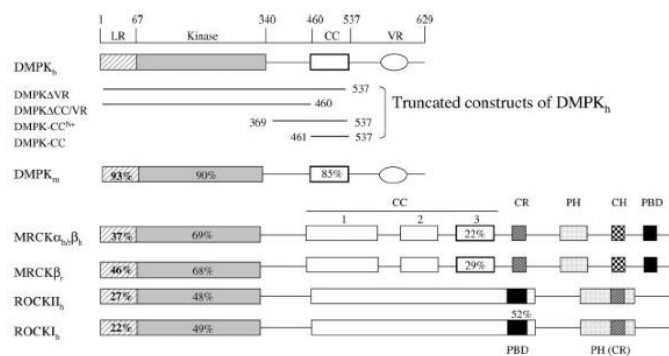
Figure 1. Domain organization in DMPK, MRCK, and ROCK. Domain composition of human DMPK (Q09013–9), mouse DMPK (P54265), human MRCKα (Q5VT26), rat MRCKα (AF021935), human MRCKβ (Q81WQ7), mouse ROCKII (V58513), and mouse ROCKI (V58512). LR: leucine-rich region; CC: coiled-coil; VR: region variable across DMPK spliceoforms. CR: cysteine-rich domain, PH: pleckstrin homology domain; CH: citron homology domain; PBD: p21 GTPase binding domain. Sequence identity to human DMPK is given as percentage where applicable. DMPK constructs used in the current study are shown.

pattern of seven amino acids denoted *a* to *g*. Positions *a* and *d* classically harbor hydrophobic residues, which constitute the structural core of the motif. Positions *e* and *g* often host charged groups that form networks of intra- and interhelical salt bridges crucial to fold stability and interchain recognition. CC domains from DMPK-related kinases are believed to form dimeric assemblies and thereby to mediate dimerization in these enzymes. To date, only two closely related CC fragments from this family have been characterized. These correspond to the C-terminal, Rho binding motifs of ROCKII-CC and ROCKI-CC that share 52% sequence identity (Fig. 1). For these, crystal structures have been reported in free form (21) and complexed to RhoA (22), respectively, confirming that both coils form parallel dimers. In a similar way, only a few analytical studies have reported on the oligomerization state of full-length kinases. Currently, data are available only for DMPK (23) and ROCKs (24, 25), which have been shown to adopt dimeric states. Despite the similarity of these kinases, different mechanisms have been proposed for their dimerization. While data on DMPK defended a self-assembly driven by its CC motif (23), the reports on ROCKs prove that CC moieties are not required for their oligomerization. The latter agrees with a biochemical study of MRCK, which also reported self-association in the absence of CC fractions (26). The associative interactions in ROCKs are the best understood so far due to the recent availability of the crystal structure from the dimeric N-terminal fraction (i.e., lacking CC segments) from ROCKI (25). This shows how head-to-head dimers interact via a small, N-terminal domain comprising the α-helical LR region (Fig. 1) as well as a tail, C-terminal to the kinase domain, which wraps around the catalytic core to reach the subunit interface. Given that dimerization segments show a certain homology across DMPK, ROCKs, and MRCK (Fig. 1), it would be reasonable to expect that related mechanism of assembly could be supported by these kinases, even if conflicting data have been reported (23).

Moreover, different biochemical roles have been attributed to the CC moieties of these kinases. The CC domains of ROCKs and citron kinase act as intrasteric inhibitors of their respective kinase domains and house at

their C termini binding motifs for small Rho GTPases, where binding of the latter releases the inhibition (7, 27). CC segments of MRCK (CC2/CC3) also act as intrasteric inhibitors but do not host a GTPase binding region (26), and physiological release factors are yet to be identified. On the contrary, the CC fraction of DMPK (DMPK-CC) has been proposed to act as an enhancer of its kinase activity (23), although it is unknown whether this involves transphosphorylation events as those observed in tyrosine kinases. In summary, both the functionality of CC domains from this family and the extent to which they govern molecular assembly remain unclear. Given that the CC from DMPK appears to differ from those of ROCKs and MRCK both in its regulatory properties and in its role during assembly (23), we have investigated the architectural details of DMPK to assess whether a common association principle exists in this family of kinases.

MATERIALS AND METHODS

Cloning

Expression constructs were derived from human DMPK cDNA (L08835) spanning residue ranges as shown in Fig 1. DMPKΔVR was cloned into the *Eco*RI and *Sal*I sites of the insect cell expression vector pFastBac1 (Invitrogen, San Diego, CA, USA). A His$_6$-tag and a TEV (tobacco etch virus) protease cleavage site were inserted N-terminal to the target construct. DMPKΔCC/VR was cloned into pET-15b (Novagen, Madison, WI, USA) using primers 5′-GGAATTCCATA-TGTCAGCCGAGGTGCG-3′ and 5′-CGGGATCCTTACCCC-GTGACAGCTGTGG-3′ with *Nde*I and *Bam*HI restriction sites, respectively. This vector adds a His$_6$-tag and a thrombin cleavage site N-terminal to the target protein. Cloning of DMPK-CC has been previously reported (28). DMPK-CC^{N+} was cloned into the pETM-11 vector, a modified version of pET-24d (Novagen) that includes a His$_6$-tag plus a TEV protease cleavage site N-terminal to the expression construct. This used primers 5′-CGTTCGAAGACTACATGTTGGTG-GAGGACGGGCTC-3′ and 5′-GGGGTACCTTACCCCGTGA-CAGCTGTGG-3′ carrying BbsI (*Nco*I-compatible) and *Kpn*I restriction sites.

Protein production

Sf9 cells were transfected with the bacmic cDNA construct of DMPKΔVR according to the manufacturer's instructions

(Life Technologies Bac-to-Bac system). The transfected cells were incubated for 3 days at 27°C and the supernatant, containing recombinant virus for protein expression, was collected. Amplified virus (10 ml) was added to 1×10^6 Sf9 cells and incubated at 27°C for 3 days. Lysis was performed by incubation on ice for 30 min in the presence of DNase and protease inhibitors (Complete EDTA-free; Roche, Nutley, NJ, USA) in 50 mM Tris pH 8.0, 300 mM NaCl, 5 mM β-mercaptoethanol, 2% glycerol. A clarified supernatant was obtained by centrifugation at 20,000 rpm and 4°C for 1 h. The protein product was purified by affinity chromatography on a His-Trap HP column (GE Healthcare) equilibrated in lysis media. The bound protein was eluted by supplementing the medium with 200 mM imidazole. Subsequent purification used size exclusion chromatography on a Superdex 200 Tricorn 10/30 GL column (GE Healthcare) in 50 mM Tris pH 8.0, 150 mM NaCl, 5 mM β-mercaptoethanol, 2% glycerol.

DMPKΔCC/VR was expressed in *E. coli* Rosetta (DE3) (Novagen) grown in LB medium supplemented with 34 μg/ml of chloramphenicol and 100 μg/ml ampicillin. Induction used 1 mM IPTG and incubation was overnight at 25°C. Cells were harvested by centrifugation at 8000 *g* at 4°C for 15 min. Lysis was by French pressing in the presence of DNase and protease inhibitors in 50 mM Tris pH 8.0, 50 mM NaCl, 5 mM β-mercaptoethanol. The clarified supernatant, obtained by centrifugation at 20,000 rpm (4°C), was purified by affinity chromatography and subsequent gel filtration (as described for DMPKΔVR) in lysis media. Expression and purification of DMPK-CC were as reported (28). Production of DMPK-CC^{N+} used protocols as for DMPK-CC, except that all media were supplemented with 5 mM β-mercaptoethanol. All measurements used fresh preparations and sample identification was by peptide-digest mass spectrometry.

Structure determination

Crystallization, data collection, data statistics, and experimental phasing of DMPK-CC have been described (28). Briefly, crystals grew from 100 mM Tris-HCl pH 7.5, 2.0 M Na$^+$/K$^+$ tartrate and 5–10 mg/ml protein solutions. X-ray diffraction data collected at beamline ID14–2 (ESRF, Grenoble) revealed the crystals to belong to space group P2$_1$2$_1$2$_1$ with unit cell dimensions of a = 39.1 Å, b = 46.2 Å, c = 143.5 Å, and three polypeptide chains per asymmetric unit. Experimental phasing used 3λ MAD on an Se-Met derivative. Phase refinement was in SHARP (29) and automatic model improvement in ARP/warp v6.1 (30). Manual model building used the software O (31) and refinement was in REFMAC (32). For cross-validation, crystallographic data were divided into a working and a test set using FREERFLAG (33). Restraints for non-crystallographic symmetry were applied during the initial phase of refinement and released at later stages. Geometrical parameters of intermediate and final models were assessed using PROCHECK (34). The final model lacks several N- and C-terminal residues that could not be resolved in electron density maps, namely 461–465/530–537 in chain A, 461–463/528–537 in chain B and 461–465/527–537 in chain C. This amounts to ~20% of disorder in the structure and explains the slightly elevated R-factor values, which are somewhat above those characteristic of this resolution. Coordinates have been deposited at the Protein Data Bank with accession number 1WT6.

Analytical ultracentrifugation

Sedimentation velocity and equilibrium data were recorded using an Optima XL-A analytical ultracentrifuge (Beckman Instruments, Fullerton, CA, USA) equipped with 4 and 12-mm Epon double-sector cells in an An-60 Ti rotor and absorption optics. Runs were performed at 20°C in *1)* 50 mM Tris pH 7.5, 50 mM NaCl and *2)* 100 mM sodium phosphate pH 7.0 using detection wavelengths of 277 nm and 234 nm. Sedimentation velocity experiments were carried out at 54,000 rpm and 0.65 mg/ml sample concentration. Sedimentation equilibrium measurements were performed at 19,000, 21,000, 24,000, and 26,000 rpm, and protein concentrations of 0.16–3.3 mg/ml for DMPK-CC and 0.09–1.4 mg/ml for DMPK-CC^{N+}, where optical density values remained in a linear range. Average molecular masses were evaluated using SEGAL (G. Machaidze and A. Lustig, unpublished software), which adjusts the baseline absorbance to obtain the best linear fit of ln(absorbance) vs. radial distance square. A protein partial specific volume of 0.73 ml/g was used, while solution density was taken as 1.003 g/ml and viscosity as 1.02 centipoise.

Circular dichroism spectroscopy

Thermal denaturation of DMPK-CC was monitored at 222 nM on a 62A DS circular dichroism spectropolarimeter (AVIV) equipped with a temperature-controlled quartz cell of 2 mm optical path over a temperature range of 5–90°C. The scan speed was 0.5°C/min with a spectral bandwidth of 2.0 nm.

Binding assays

The binding of DMPK-CC to DMPKΔCC/VR was assayed in 50 mM Tris-HCl pH 7.5, 100 mM NaCl by *1)* affinity pull-down assays using His$_6$-tagged DMPKΔCC/VR immobilized in a His-Trap HP column (GE Healthcare) vs. a flow of untagged DMPK-CC; *2)* simultaneous coexpression of both constructs followed by attempts of copurification via metal affinity plus size exclusion chromatography (protocols as above).

Size exclusion chromatography combined with multiangle light scattering (SEC-MALS)

The oligomeric state of DMPKΔVR and DMPKΔCC/VR in solution was determined via SECMALS measurements performed on an ÅKTA explorer 10 system (GE Healthcare) connected to a triangle light scattering detector and a differential refractometer (miniDAWN Tristar and Optilab, respectively; Wyatt Technology, Santa Barbara, CA, USA). A Superdex 200 HR 10/30 column (GE Healthcare) was used in 50 mM Tris-HCl pH 8.0, 50 mM NaCl at a flow rate of 0.5 ml/min. Sample volumes of 50 μl (DMPKΔVR) and 100 μl (DMPKΔCC/VR) were injected at a concentration of 2 mg/ml. A specific refractive index increment (dn/dc) value of 0.185 ml/g was used. The data were recorded and processed using ASTRA software (Wyatt Technology). To determine the detector delay volumes and normalization coefficients for the MALS detector, a BSA sample (Sigma A-8531) was used as reference. Neither despiking nor band broadening correction was applied.

Small angle X-ray scattering (SAXS)

Synchrotron radiation SAXS data were collected on the X33 camera (35) of EMBL (DESY, Hamburg). Solutions of DMPKΔCC/VR were measured at protein concentrations of 1.4, 3.8, 5.7, 8.5, and 9.3 mg/ml in 50 mM TrisHCl pH 8.0, 50 mM NaCl, 2.5 mM β-mercaptoethanol at 4°C. Measurements were recorded using a MAR345 IP detector at a sample detector distance of 2.7 m and wavelength λ = 1.5 Å in the range of momentum transfer $0.011 < s < 0.45$ Å^{-1}. Exposure time for a single measurement was 5 min. To check for

135

TABLE 1. *X-ray data statistics and model parameters*

X-ray data	
Space group	P2₁2₁2₁
Resolution (Å)	20–1.6 (1.69–1.6)
Unique reflections	33876 (3951)
Completeness (%)	96.3 (79.2)
$R_{sym}(I)$	0.059 (0.330)
Multiplicity	4.4 (2.3)
$<I/\sigma(I)>$	15.0 (2.5)
Model refinement	
Resolution (Å)	20–1.6
Reflections working/free	32044/1712
R-factor/R-free (%)	22.7/27.0
Chains per asymmetric unit	3
No. protein/solvent atoms	1619/294
Average B (Å²)	26.05
Rmsd bond lengths (Å)/ angles (°)	0.019/1.717
Residues in Ramachandran core / disallowed (%)	99.5/0

radiation damage, two consecutive 3 min exposures at the highest protein concentration were compared and no changes were observed. The data were normalized, the scattering of the buffer subtracted, and the difference curves extrapolated to zero solute concentration using the program package PRIMUS (36). The maximum particle dimension D_{max}, forward scattering I (0) and the radius of gyration R_g were evaluated using GNOM (37). The molecular mass (MM) of the solute was calculated by comparing its forward scattering with that of a reference solution of BSA (MM=66 kDa). The volume of the hydrated particle was computed using the Porod invariant equation (38). The low resolution shape of the molecule was computed *ab initio* using the program DAMMIN (39), which uses an assembly of densely packed beads to represent the protein molecule. Simulated annealing was used to build a compact, interconnected configuration of beads inside a sphere with the diameter D_{max} that fits the experimental data $I_{exp}(s)$ to minimize the discrepancy:

$$\chi^2 = \frac{1}{N-1} \sum_j \left[\frac{I_{exp}(s_j) - cI_{calc}(s_j)}{\sigma(s_j)} \right]^2$$

where N is the number of experimental points, c is a scaling factor and $I_{calc}(s)$, and $\sigma(s_j)$ are the calculated intensity and the experimental error at the momentum transfer s_j, respectively.

The scattering from the crystallographic model of ROCKI (protein fraction of PDB entry 2ETR; 25) was calculated using CRYSOL (40). The model of dimeric DMPKΔCC/VR was constructed based on the structure of ROCKI using the program BUNCH (41). The dimerization domains consisting of the 70 N-terminal and 20 C-terminal residues of ROCKI remained fixed during the calculations, so that the subunit interface stayed unaltered. The kinase domain was allowed to move as a rigid body while the missing C-terminal tail in DMPKΔCC/VR was represented as an interconnected chain of dummy residues (for domain definition, see Fig 6a) (42). A simulated annealing protocol was used to position the kinase domain and find the configuration of the C-terminal fragment that could fit the experimental scattering while obeying the P2 symmetry of the dimer.

RESULTS

Crystal structure of DMPK-CC

The structure of DMPK-CC (residues 461–537; Fig. 1) has been elucidated by X-ray crystallography up to 1.6 Å resolution. X-ray data statistics and model refinement parameters are given in **Table 1**. Crystals contain three polypeptide chains in the asymmetric unit, which form a consecutive coil of 86.5 Å length, spanning residues 466–527 and comprising eight complete heptad repeats. They build a parallel, left-handed trimeric motif (**Fig 2a**).

The architecture of DMPK-CC is restrained at its terminal sections while a central region exhibits moderate structural fluctuations as indicated by crystallographic temperature factors and interchain conformational agreement (Fig 2a, b). The relative structural slackness at the middle point of the domain correlates with a widening of the coil at that position (Fig 2c), which is likely to be caused by the presence of a sequence motif, 493-IRTD-496, with low propensity for CC formation (estimated using MultiCoil; 43). Also noticeable is a residue located upstream, R483, which contributes a charged group in core position *a* with a reduced Crick-phase (−2.8° instead of the commonly

Figure 2. Crystal structure of DMPK-CC. *a)* Ribbon representation colored in a blue-to-red gradient according to crystallographic B-factors (8 to 55 Å²), where blue designates low values. Solvent atoms in the interior of the CC motif are displayed as red spheres; *b)* Superimposition of chains A, B, and C colored as in panel *a*. Cα atoms of residues influencing coil parameters are shown as black spheres; *c)* coil parameter values for pitch, radius and residue tendency to coil formation. Residues leading to coil alterations are boxed; *d)* sequence alignment of DMPK-CC and MRCK-CC3. Identity is shown in yellow and similarity in green.

Figure 3. Intra- and interchain interactions in DMPK-CC. Green represents hydrophobic groups, red is negatively charged, blue is positively charged, and black, all other residues. The central red sphere represents an internal solvent atom located at the 3-fold axis and coordinating the three side chains of N511. Interactions are shown as connecting lines, where salt bridges are in black.

encountered value of −40°, as calculated using Twister; 44). This leads to overtwisting of the subsequent coil portion (Fig 2c) and an associated suboptimal hydrophobic core packing in that region. Thus, central coil widening and pitch parameters appear not to simply reflect crystal lattice restraints but to be an intrinsic structural feature of this CC domain.

Residue arrangements around the 3-fold axis of DMPK-CC are mostly symmetrical, an exception are the side chains of F500, an aromatic residue in position *d*. The presence of aromatic groups in core positions is extremely rare in dimeric CCs, but it is observed in trimeric motifs (45). For example, the trimeric coil of lung surfactant protein D (46) contains F225 and Y228

in *a* and *d* positions, respectively, which exhibit similar asymmetrical arrangements of their side chains. This is likely to reflect the difficulty of packing bulky groups in the limited interior of a trimeric coil. Remarkably, DMPK-CC also houses several solvent molecules in its core. In particular, its C-terminal section comprises an internal cavity filled by solvent that stabilizes a polar residue, N511, in *a* core position. A similar arrangement has been found in other trimeric CCs, such as bacteriophage T4 fibritin (47) and a domain of a coat protein from MoMuLVirus (48). Such buried polar groups in CCs are thought to contribute to the specificity of the motif and to maintain chain registry (49). Other relevant interactions in DMPK-CC (**Fig. 3**) include 6 interchain and 2 intrachain salt bridges. Interchain bridges help defining the association state of CCs, while intrachain bridges are important for helix stability (50). The interchain bridges of DMPK-CC belong both to the classical types *eg*'RE and *eg*'ER and to unusual core contacts of type *ad*'RE (notation as in ref 40; where ' designate residues in a subsequent heptad repeat). Intrachain salt bridges in DMPK-CC are of type *bf*RE. It is striking that the sequence of DMPK-CC is rich in charged residues (16E, 9R, vs. 2D, 0K per chain), although less than half of these are involved in interactions (8E, 4R, 0D). Thus, availability of reactive groups in the surface of this domain is high (predominantly of a negatively charged character and toward the NH_2 terminus of the domain), so that it could support electrostatic binding to other molecular components.

Characterization of DMPK-CC in solution

In agreement with crystallographic data, the association state of DMPK-CC in solution is trimeric. One single oligomeric species could be detected using native-PAGE, size exclusion chromatography (**Fig 4a**), and a sedimentation velocity study, which produced a single migrating boundary corresponding to a S_{20W} of 2,1. Thus, the sample population in solution appeared to be highly homogeneous, with no dynamic states of assem-
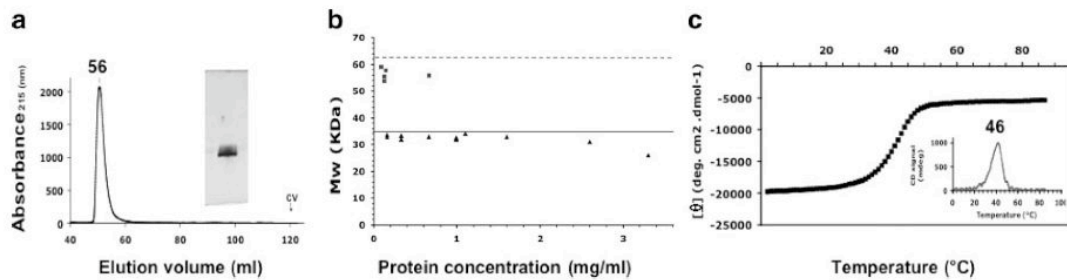
Figure 4. Biophysical characterization of DMPK-CC in solution. *a)* Size exclusion chromatogram in 50 mM Tris-HCl pH 7.5, 100 mM NaCl using a Superdex 75 Hi Load 16/60 column. A native-PAGE is shown in the inset; *b)* molecular mass calculated from sedimentation equilibrium measurements for DMPK-CC (triangles) and DMPK-CC^{N+} (squares). Buffer conditions as in panel *a*. Theoretical mass values for trimeric formations are indicated by horizontal bars, where a continuous line corresponds to DMPK-CC and a dashed line to DMPK CC^{N+}. *c)* Thermal denaturation curve monitored by circular dichroism. First derivative is shown.

137

bly being observed. Sedimentation equilibrium analysis of DMPK-CC yielded an average molecular mass (MM) of ~33 kDa over the entire protein concentration range assayed (0.16–3.3 mg/ml), consistent with the single presence of trimers in the solutions (theoretical MM for a trimer of this construct 34.9 kDa; Fig 4b). Since the sample remained fully assembled at the lowest concentration assayed, the Kd of the interaction must be lower than 4 µM.

These studies carried out at moderate ionic strength (50–100 mM NaCl) and limited protein concentration are in good agreement with crystallographic data obtained from high salt solutions (2.0 M Na$^+$/K$^+$ tartrate) or polyethylene glycol media and concentrated protein samples (5–10 mg/ml), both at pH 4.6 and 7.5 (28). This suggests that the trimeric formations of DMPK-CC are stable in a broad range of media conditions. To gain further insight into the stability of this domain, it was investigated by thermal denaturation monitored by circular dichroism. A monophasic melting curve was recorded characteristic of a cooperative helix-coil transition of the three chains as a single domain, with a $T_{1/2m}$ of 46°C (Fig 4c). This measurement was carried out in 100 mM sodium phosphate buffer pH 7.0, 100 mM NaCl at a protein concentration of 0.6 mg/ml. To account for specific buffer effects, measurements were repeated in 50 mM bis-Tris propane pH 7.5, 100 mM NaCl with equivalent results. This indicates a strong association of the component chains and confirms the robustness of the assembly. Therefore, it can be concluded that DMPK-CC has a strong tendency to self-associate into well-defined trimeric formations.

Oligomeric state of DMPK variants containing and lacking the CC domain

The assembly level of a DMPK construct comprising the LR sequence, the kinase domain, and the CC fraction, DMPKΔVR (Fig. 1), was investigated using multiangle light scattering coupled to size exclusion chromatography (SEC-MALS). This construct closely resembles an expressed DMPK spliceoform and is a suitable representative of this kinase in the cell. The study yielded MMs of 122′7 kDa and 124′5 kDa for protein concentrations estimated using UV-A$_{280}$ and refractive index, respectively. These values are in excellent agreement with the calculated MM of 120′2 kDa for a dimer of this sample (**Fig. 5**). This result is consistent with previous sedimentation equilibrium data on a comparable construct of this kinase (23).

To investigate the contribution of DMPK-CC to the dimerization of this kinase, a truncated variant lacking the CC fraction, DMPKΔCC/VR (Fig. 1), was examined. Data from size exclusion chromatography and native PAGE indicated the presence of a single species of DMPKΔCC/VR in solution, and SEC-MALS measurements confirmed that this corresponded to dimeric formations (Fig. 5). The experimental MM estimated for this sample was 101′4 kDa and 107′4 kDa using
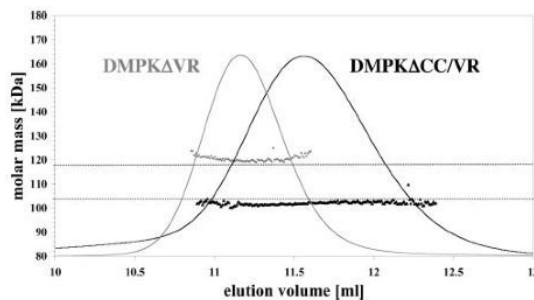
Figure 5. SEC-MALS characterization of DMPKΔVR and DMPKΔCC/VR. Molar mass (dots) and normalized UV signals are shown. The theoretical molar masses of the dimers of DMPKΔVR and DMPKΔVR/CC (120.0 kDa and 106.6 kDa, respectively) are indicated with horizontal lines.

UV-A$_{280}$ and refractive index to estimate the protein concentration, respectively. This is in close agreement with the theoretical MM of 106′6 kDa for a dimer of this construct. Furthermore, samples remained monodisperse and data were consistent through injected protein concentrations of 0.3–2 mg/ml, indicating that results are concentration independent (data not shown). The assembly state of DMPKΔVR coincides with that of DMPKΔCC/VR but not DMPK-CC, showing that the association level of this kinase is not dictated by the oligomerization properties of its CC domain and that segments associated with the kinase domain bear this function instead.

Analysis of DMPKΔCC/VR dimers by small angle X-ray scattering

The kinase domain of DMPK is flanked by an N-terminal region of ~65 residue length (LR) and a C-terminal tail containing ~120-residues that links it to the CC domain (Fig. 1). Recent crystallographic data on ROCKI (25) show that, in that kinase, sequences both N- and C-terminal to the kinase domain pack together to form a dimerization domain. It has been shown that the single presence of the α-helical LR region is not sufficient to secure assembly in this case, so that the removal of the kinase C-terminal tail results in ~80% monomeric species (24). Given that ROCKI and DMPK share homology in the mentioned regions (22% and 36% seq. id. in the LR domain and kinase tail, respectively; **Fig 6a**), we investigated whether DMPK might follow the same architectural principles of ROCKI by analyzing the overall conformation of DMPKΔCC/VR dimers using SAXS.

The experimental SAXS pattern of DMPKΔCC/VR is displayed in Fig 6b. The estimated MM, 95 ± 10 kDa, confirms that the protein is dimeric under the SAXS experimental conditions. This is further corroborated by the volume of the hydrated particle, 210 ± 10 × 10^{3}Å (3), noting that, for globular proteins, the hydrated volume in Å^3 should be approximately twice the MM in daltons. The radius of gyration R_g and the maximum particle size D_{max} were 39.9 ± 0.6 Å and

138
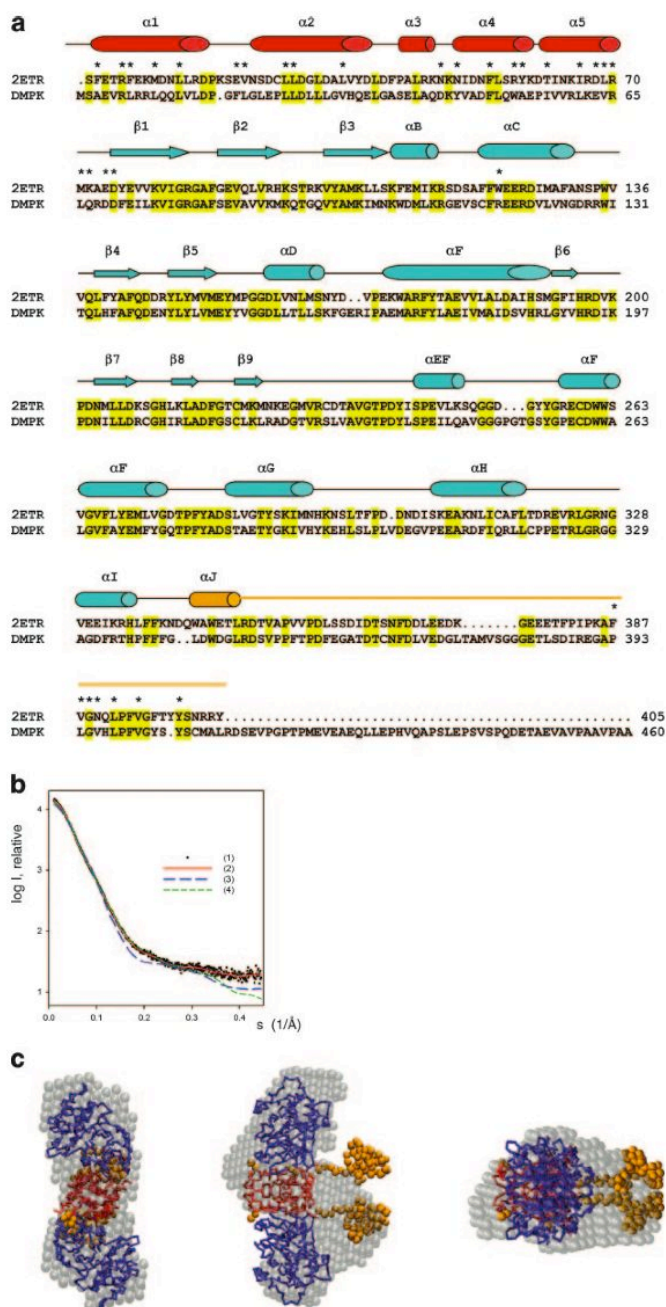
Figure 6. SAXS analysis of DMPKΔCC/VR dimers (*a*). Sequence similarity between DMPKΔCC/VR and the crystallographic model of ROCKI (PDB code 2ETR). Sequence identity is highlighted in yellow. Secondary structure elements are as in ref 25. Domain definition is as follows: red corresponds to the LR domain, blue to the kinase domain, and golden to the C-terminal tail. ROCKI residues involved in dimerization are marked with an asterisk. *b)* Experimental scattering of DMPKΔCC/VR (1) and calculated patterns from the *ab initio* low-resolution bead model (2), the ROCKI crystallographic model (3) and the reconstructed DMPKΔCC/VR model with an added C-terminal fraction (4); *c) ab initio* model of DMPKΔCC/VR (gray semitransparent beads) superimposed on the model constructed by adding C-terminal particles (orange beads) to the protomers of ROCKI (Cα-trace). Domain coloring as in *a*. Three orthogonal views are shown.

130 ± 10 Å, respectively, suggesting an elongated structure. This agrees with the low-resolution shape of DMPKΔCC/VR reconstructed *ab initio*, which fits neatly the experimental data with a discrepancy χ = 0.95 (Fig 6*b*). A comparison of this with the crystal structure of ROCKI indicates that DMPKΔCC/VR has a similar overall conformation, also appearing to form head-to-head dimers (Fig 6*c*). The low-resolution model of DMPKΔCC/VR, however, displayed extra electron density that was not satisfied by the ROCKI structure, as can be expected from the fact that each protomer of DMPKΔCC/VR has 65 additional residues at its C terminus (Fig 6*a*). The missing residues account for 14% of the total scattering mass of this sample. They

139

play a significant role in the scattering of this molecule, possibly due to their contribution to an anisometric molecular shape (see below), so that the patterns computed from ROCKI yielded a poor fit to the experimental data ($\chi = 6.6$; Fig 6b). To construct a model of DMPKΔCC/VR, the missing C-terminal portion was generated for each monomer using beads and maintaining a 2-fold symmetry as described in Materials and Methods. Although during the calculation the kinase domain was allowed to drift and rotate as a rigid body, its position was practically unchanged with respect to ROCKI. Models of DMPKΔCC/VR so generated showed that the C-terminal portions of each protomer formed rather compact formations, suggesting the existence of a protein domain at this position. A typical model of DMPKΔCC/VR in solution provides a good fit to the experimental data ($\chi = 1.75$; Fig 6b) and agrees well with the molecular shape calculated *ab initio* as revealed by a comparative superimposition (Fig 6c). A deviation between models, however, shows that the volumes occupied by the C-terminal fraction are represented somewhat differently by the two modeling methods. It should be noted that the *ab initio* model was obtained without imposing any prior knowledge or symmetry restrictions, whereas the reconstructed model applies a 2-fold symmetry that may not be entirely adhered to by a potentially flexible C-terminal tail. Such flexibility could underlie localized deviations between otherwise consistent models and scattering patterns.

CC domains can be found in proteins in combination with other motifs that influence their association. The most prominent example is that of the trimeric CC of fibritin, which is followed by a C-terminal extension of ~30 residue length that folds into a β-hairpin (51). This intertwines with equivalent β-hairpins from other subunits, being essential to achieve CC trimerization in that case. Given that SAXS models of DMPKΔCC/VR suggested the presence of a structural formation immediately prior to the CC fraction, we assayed whether this might act as an association motif that could modulate CC stoichiometry. For this, we constructed an N-terminally extended version of DMPK-CC, DMPK-CC^{N+} (Fig. 1), and analyzed its oligomeric state by sedimentation equilibrium (Fig 4b). Results indicated that DMPK-CC^{N+} remained trimeric (estimated MM 56.1 kDa compared to a theoretical MM of 63.5 kDa for a trimer of this construct), suggesting that possible structural elements of the kinase-CC linker sequence do not influence the oligomerization of this coil.

DMPK-CC is not involved in domain interactions

The CC fractions of ROCKs and MRCK have been shown to interact directly with their respective kinase domains, thereby acting as intrasteric inhibitors of kinase activity (26, 27). In ROCKs, the interaction involves the Rho binding site at the C terminus of its long CC domain, whereas in MRCK the binding has been mapped to a fragment spanning CC2/CC3, whose

deletion resulted in about a 3-fold increase in activity. Given that DMPK-CC and MRCK-CC3 share certain sequence similarity (up to 31% identity according to isoforms) (Fig 2d), we investigated the possible interaction of DMPK-CC with other domains in this kinase using pull-down assays as well as coexpression/copurification experiments on DMPK-CC and DMPKΔCC/VR. Binding between these species could not be detected by either procedure, which yielded independent original species. Hence, DMPK-CC appears to neither interact with the kinase domain nor form a hetero-coil against the α-helical LR sequence. This indicates that a CC association/dissociation regulatory model such as that proposed for MRCK and ROCKs is unlikely to be applicable to DMPK and that the kinase and CC fractions component of DMPK are likely to be arranged as architecturally distinct fractions, resulting in a globular portion and a filamentous shaft.

DISCUSSION

The CC domains of DMPK-related kinases are essential to the regulation of their catalytic activities. However, the molecular events mediated by these domains are yet to be accurately understood. CC sequences in this family exhibit large variability in length and composition, sharing little or no similarity. Only minimal structural data exist that aid to clarify the respective properties and functional roles of CC domains within this family. Here we report the crystal structure of DMPK-CC, which is unrelated to the ROCKs-CCs previously described (21, 22). Contrary to the dimeric character of the latter fragments, DMPK-CC forms a trimeric motif that is characterized by rigid termini and a central widening of its coil. It is currently unknown whether this feature might reflect functional properties of this domain. As proven by sedimentation equilibrium data, the assembly level of DMPK-CC in solution coincides with that in the crystal. No other oligomeric state has been detected in the current study under a broad range of protein concentration and ionic strength conditions. Furthermore, thermal denaturation studies have shown this assembly to be stable. Thus, our data prove that the oligomeric states of CC domains from this kinase family are more diversified than previously expected and that more coils must be investigated before general conclusions on their self-association properties can be drawn.

Experimental results on DMPK-CC contrast markedly with sequence-based predictions, which suggests this coil forms dimeric assemblies. For example, the software MultiCoil (43) estimates 82% probability of dimer vs. 10% trimer formation. The structural determinants of CC association remain poorly understood, and thus such predictions should be considered with caution. Only a few systematic analyses have addressed the basis of CC association experimentally, focusing on either the role of hydrophobic core residues (52, 53) or salt bridges (e.g., 54, 55). Consequently, cases are described where experimental data disagree with predictions (for

example, see ref 46). To understand the molecular basis of DMPK-CC trimerization, we analyzed its interchain contacts. The only factor that might be proposed at this time to favor formation of trimers over dimers is the presence of an aromatic residue, F500, at core position *a*, since this is rarely observed in dimeric coils (45). The contribution of the several salt bridges in DMPK-CC to its oligomerization cannot be evaluated based on current knowledge.

Given that CCs are established oligomerization motifs, these were initially thought to be indispensable for the molecular assembly of DMPK-related kinases. However, recent studies of ROCKs have shown that other domains mediate self-association in these kinases, namely, the N-terminal LR region and a tail C-terminal to the kinase domain (24, 25). Similarly, in the current study, a truncated DMPK variant lacking the CC moiety formed dimeric assemblies according to SEC-MALS data. A low-resolution structural analysis of these dimers using SAXS indicated that DMPK and ROCKI share similar dimerization principles involving head-to-head protomer interactions, even though a certain number of residues involved in the dimerization of ROCKI are not conserved in DMPK (Fig 6*a*). Our data on the truncated DMPK variant differ from those of a previous study (23), where a similar DMPK construct was argued to be monomeric according to sucrose density gradient sedimentation and pull-down assays. The dimeric form in our study (residues 1–460) contains 29 more residues at its C terminus than the construct previously reported. The truncation in that case might have disrupted a possible structural element C-terminal to the kinase domain, which has been insinuated by current SAXS data. More studies are required to identify the exact structural details of this segment of DMPK and to establish whether its integrity is essential for the assembly of this kinase. Contrary to that former study, we conclude that the CC domain of DMPK is not required for its self-association and that a common architectural design is likely to be shared by the closest members of this family: DMPK, MRCK, and ROCKs.

The fact that an interaction between DMPK-CC and other domains in DMPK could not be identified in the current study suggests it is unlikely that the assembly properties of this CC are altered by a possible embedding within the kinase core. Neither is the influence of flanking structural elements expected, given that an N-terminally extended CC construct retained the same oligomerization concentration. Thus, we speculate that the stoichiometry of this coil within the context of full DMPK must be dominated by the prior association of the N-terminal catalytic fraction into dimers. Prediction data confirm that the DMPK-CC sequence is geometrically compatible with a dimeric association, and it is known that CC oligomerization can be dynamically altered by local factors (20).

Oligomerization is crucial for the activation of numerous kinases. Often self-assembly is directed to ensure efficient autophosphorylation in *trans* by increasing the local concentration of catalytic domains (56). Given that DMPK-CC does not determine the self-association of this kinase and that, contrary to MRCK and ROCKs, it does not act as an intrasteric inhibitor but as an activator, further investigations are needed to clarify whether its function might relate instead to the achievement of an active conformation or even substrate docking. Moreover, given the heterogeneity in CC sequences and their diversified interaction patterns across DMPK-related kinases, studies are required to reveal the exact role of these moieties and to establish the principles of their regulatory mechanisms. [FJ]

REFERENCES

1. Fu, Y. H., Pizzuti, A., Fenwick, R. G., Jr., King, J., Rajnarayan, S., Dunne, P. W., Dubel, J., Nasser, G. A., Ashizawa, T., de Jong, P., et al. (1992) An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* **255,** 1256–1258

2. Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H., Hunter, K., Stanton, V. P., Thirion, J. P., Hudson, T., et al. (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* **68,** 799–808

3. Nakagawa, O., Fujisawa, K., Ishizaki, T., Saito, Y., Nakao, K., and Narumiya, S. (1996) ROCK-I and ROCK-II, two isoforms of Rho-associated coiled-coil forming protein serine/threonine kinase in mice. *FEBS Lett.* **392,** 189–193

4. Ishizaki, T., Maekawa, M., Fujisawa, K., Okawa, K., Iwamatsu, A., Fujita, A., Watanabe, N., Saito, Y., Kakizuka, A., Morii, N., and Narumiya, S. (1996) The small GTP-binding protein Rho binds to and activates a 160 kDa Ser/Thr protein kinase homologous to myotonic dystrophy kinase. *EMBO J.* **15,** 1885–1893

5. Wissmann, A., Ingles, J., McGhee, J. D., and Mains, P. E. (1997) Caenorhabditis elegans LET-502 is related to Rho-binding kinases and human myotonic dystrophy kinase and interacts genetically with a homolog of the regulatory subunit of smooth muscle myosin phosphatase to affect cell shape. *Genes Dev.* **11,** 409–422

6. Luo, L., Lee, T., Tsai, L., Tang, G., Jan, L. Y., and Jan, Y. N. (1997) Genghis Khan (Gek) as a putative effector for Drosophila Cdc42 and regulator of actin polymerization. *Proc. Natl. Acad. Sci. USA* **94,** 12963–12968

7. Di Cunto, F., Calautti, E., Hsiao, J., Ong, L., Topley, G., Turco, E., and Dotto, G. P. (1998) Citron rho-interacting kinase, a novel tissue-specific ser/thr kinase encompassing the Rho-Rac-binding protein Citron. *J. Biol. Chem.* **273,** 29706–29711

8. Leung, T., Chen, X. Q., Tan, I., Manser, E., and Lim, L. (1998) Myotonic dystrophy kinase-related Cdc42-binding kinase acts as a Cdc42 effector in promoting cytoskeletal reorganization. *Mol. Cell. Biol.* **18,** 130–140

9. Madaule, P., Eda, M., Watanabe, N., Fujisawa, K., Matsuoka, T., Bito, H., Ishizaki, T., and Narumiya, S. (1998) Role of citron kinase as a target of the small GTPase Rho in cytokinesis. *Nature* **394,** 491–494

10. Piekny, A. J., and Mains, P. E. (2002) Rho-binding kinase (LET-502) and myosin phosphatase (MEL-11) regulate cytokinesis in the early *Caenorhabditis elegans* embryo. *J. Cell Sci.* **115,** 2271–2282

11. Kosako, H., Yoshida, T., Matsumura, F., Ishizaki, T., Narumiya, S., and Inagaki, M. (2000) Rho-kinase/ROCK is involved in cytokinesis through the phosphorylation of myosin light chain

141

and not ezrin/radixin/moesin proteins at the cleavage furrow. *Oncogene* **19**, 6059–6064

12. Jin, S., Shimizu, M., Balasubramanyam, A., and Epstein, H. F. (2000) Myotonic dystrophy protein kinase (DMPK) induces actin cytoskeletal reorganization and apoptotic-like blebbing in lens cells. *Cell Motil. Cytoskeleton* **45**, 133–148

13. Schulz, P. E., McIntosh, A. D., Kasten, M. R., Wieringa, B., and Epstein, H. F. (2003) A role for myotonic dystrophy protein kinase in synaptic plasticity. *J. Neurophysiol.* **89**, 1177–1186

14. Shimizu, M., Wang, W., Walch, E. T., Dunne, P. W., and Epstein, H. F. (2000) Rac-1 and Raf-1 kinases, components of distinct signaling pathways, activate myotonic dystrophy protein kinase. *FEBS Lett.* **475**, 273–277

15. Mounsey, J. P., Mistry, D. J., Ai, C. W., Reddy, S., and Moorman, J. R. (2000) Skeletal muscle sodium channel gating in mice deficient in myotonic dystrophy protein kinase. *Hum. Mol. Genet.* **9**, 2313–2320

16. Lee, H. C., Patel, M. K., Mistry, D. J., Wang, Q., Reddy, S., Moorman, J. R., and Mounsey, J. P. (2003) Abnormal Na channel gating in murine cardiac myocytes deficient in myotonic dystrophy protein kinase. *Physiol. Genomics* **12**, 147–157

17. Harper, P. S. (1989) *Myotonic Dsytrophy*, Saunders, London/Philadelphia

18. Davis, B. M., McCurrach, M. E., Taneja, K. L., Singer, R. H., and Housman, D. E. (1997) Expansion of a CUG trinucleotide repeat in the 3′ untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. *Proc. Natl. Acad. Sci. USA* **94**, 7388–7393

19. Jiang, H., Mankodi, A., Swanson, M. S., Moxley, R. T., and Thornton, C. A. (2004) Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons. *Hum. Mol. Genet.* **13**, 3079–3088

20. Lupas, A. N., and Gruber, M. (2005) The structure of alpha-helical coiled coils. *Adv. Protein Chem.* **70**, 37–78

21. Shimizu, T., Ihara, K., Maesaki, R., Amano, M., Kaibuchi, K., and Hakoshima, T. (2003) Parallel coiled-coil association of the RhoA-binding domain in Rho-kinase. *J. Biol. Chem.* **278**, 46046–46051

22. Dvorsky, R., Blumenstein, L., Vetter, I. R., and Ahmadian, M. R. (2004) Structural insights into the interaction of ROCKI with the switch regions of RhoA. *J. Biol. Chem.* **279**, 7098–7104

23. Zhang, R., and Epstein, H. F. (2003) Homodimerization through coiled-coil regions enhances activity of the myotonic dystrophy protein kinase. *FEBS Lett.* **546**, 281–287

24. Doran, J. D., Liu, X., Taslimi, P., Saadat, A., and Fox, T. (2004) New insights into the structure-function relationships of Rho-associated kinase: a thermodynamic and hydrodynamic study of the dimer-to-monomer transition and its kinetic implications. *Biochem. J.* **384**, 255–262

25. Jacobs, M. D., Hayakawa, K., Swenson, L., Bellon, S. F., Fleming, M., Taslimi, P., and Doran, J. (2005) The structure of dimeric rock 1 reveals the mechanism for ligand selectivity. *J. Biol. Chem.* Oct 24; [Epub ahead of print]

26. Tan, I., Seow, K. T., Lim, L., and Leung, T. (2001) Intermolecular and intramolecular interactions regulate catalytic activity of myotonic dystrophy kinase-related Cdc42-binding kinase alpha. *Mol. Cell. Biol.* **21**, 2767–2778

27. Amano, M., Chihara, K., Nakamura, N., Kaneko, T., Matsuura, Y., and Kaibuchi, K. (1999) The COOH terminus of Rho-kinase negatively regulates rho-kinase activity. *J. Biol. Chem.* **274**, 32418–32424

28. Garcia, P., Marino, M., and Mayans, O. (2004) Crystallization and preliminary X-ray analysis of the coiled-coil domain of dystrophia myotonica kinase. *Acta Crystallogr. D* **60**, 2336–2339

29. La Fortelle, E., and Bricogne, G. (1997) *Methods Enzymol.* **276**, 472–494

30. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**, 458–463

31. Jones, T.A, Zou, J.Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119

32. Murshudov, G. N., Vagin A.A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* **53**, 240–255

33. Collaborative Computational Project, N. 4 (1994) *Acta Crystallogr. D* **50**, 760–763

34. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291

35. Koch, M. H. J., and Bordas, J. (1983) X-ray diffraction and scattering on disordered systems using synchrotron radiation. *Nucl. Instrum. Methods* **208**, 461–469

36. Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M. H. J., and Svergun, D. I. (2003) PRIMUSā Windows-PC based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282

37. Svergun, D. I. (1992) Determination of the regularization parameter in indirect transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503

38. Porod, G. (1982) Small angle X-ray scattering. (Glatter, O., and Kratky, O., eds) Academic Press, London

39. Svergun, D. I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–2886

40. Svergun, D. I., Barberato, C., and Koch, M. H. J. (1995) CRYSOLā program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773

41. Petoukhov, M. V., and Svergun, D. I. (2005) Global rigid body modelling of macromolecular complexes against small-angle scattering data. *Biophys. J.* **89**, 1237–1250

42. Petoukhov, M. V., Eady, N.A., Brown, K. A., and Svergun, D. I. (2002) Addition of missing loops and domains to protein models by x-ray solution scattering. *Biophys. J.* **83**, 3113–3125

43. Wolf, E., Kim, P. S., and Berger, B. (1997) MultiCoil: a program for predicting two- and three-stranded coiled coils. *Protein Sci.* **6**, 1179–1189

44. Strelkov, S. V., and Burkhard, P. (2002) Analysis of alpha-helical coiled coils with the program TWISTER reveals a structural mechanism for stutter compensation. *J. Struct. Biol.* **137**, 54–64

45. Walshaw, J., and Woolfson, D. N. (2001) Socket: a program for identifying and analysing coiled-coil motifs within protein structures. *J. Mol. Biol.* **307**, 1427–1450

46. Hakansson, K., Lim, N. K., Hoppe, H. J., and Reid, K. B. (1999) Crystal structure of the trimeric alpha-helical coiled-coil and the three lectin domains of human lung surfactant protein D. *Structure Fold Des.* **7**, 255–264

47. Tao, Y., Strelkov, S. V., Mesyanzhinov, V. V., and Rossmann, M. G. (1997) Structure of bacteriophage T4 fibritin: a segmented coiled coil and the role of the C-terminal domain. *Structure* **5**, 789–798

48. Fass, D., Harrison, S. C., and Kim, P. S. (1996) Retrovirus envelope domain at 1.7 Å resolution. *Nat. Struct. Biol.* **3**, 465–469

49. Lumb, K. J., and Kim, P. S. (1998) A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil. *Biochemistry* **34**, 8642–8648

50. Burkhard, P., Ivaninskii, S., and Lustig, A. (2002) Improving coiled-coil stability by optimizing ionic interactions. *J. Mol. Biol.* **318**, 901–910

51. Tao Y, Strelkov, S.V., Mesyanzhinov, V.V., and Rossmann, M. G. (1997) Structure of bacteriophage T4 fibritin: a segmented coiled coil and the role of the C-terminal domain. *Structure* **5**, 789–798

52. Harbury, P. B., Zhang, T., Kim, P. S., and Alber, T. (1993) A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science* **262**, 1401–1407

53. Gonzalez, L., Jr., Woolfson, D. N., and Alber, T. (1996) Buried polar residues and structural specificity in the GCN4 leucine zipper. *Nat. Struct. Biol.* **3**, 1011–1018

54. Kammerer, R. A., Jaravine, V. A., Frank, S., Schulthess, T., Landwehr, R., Lustig, A., Garcia-Echeverria, C., Alexandrescu, A. T., Engel, J., and Steinmetz, M. O. (2001) An intrahelical salt bridge within the trigger site stabilizes the GCN4 leucine zipper. *J. Biol. Chem.* **276**, 13685–13688

55. Meier, M., Lustig, A., Aebi, U., and Burkhard, P. (2002) Removing an interhelical salt bridge abolishes coiled-coil formation in a de novo designed peptide. *J. Struct. Biol.* **137**, 65–72

56. Hubbard, S. R., and Till, J. H. (2000) Protein tyrosine kinase structure and function. *Annu. Rev. Biochem.* **69**, 373–398

142

Final remarks on protein-protein interfaces and their evolution

Proteins work in the cell within the context of multi-modular networks, where individual components are connected through a succession of interactions, as in the case of an information-processing signalling cascade, or by the product of one enzyme being the substrate of the next in a metabolic cascade. "Network connections" can be established between (i) dynamic functional protein units which do not interact but communicate "through space"; (ii) weakly and transiently interacting units; (iii) stable, obligate protein assemblies or (iv) fused multi-modular protein systems where different functional units are covalently linked. The assembly of protein components brings up a direct functional advantage, since the protein system gains additional properties that exceed the sum of their individual functions. For example, clustering of modules in signalling cascades or metabolic pathways ensures a rapid information exchange and eliminates delays that would occur as a result of cytosolic crowding and diffusion limitation. Complex formation can also provide a unique, local physicochemical environment that leads to an increase in specificity, affinity and, at times, enhanced activity through cooperativity. In addition, formation of higher-order assemblies provides an increase in mechanical and chemical stability, an advantage most clearly exhibited by the homo-oligomeric filamentous assemblies of the cytoskeleton. Thus, it can be concluded that molecular interactions stand substantial selection pressure. This is supported by the fact that proteins interacting with other proteins in strong, structural interactions (for example, histones, actin and tubulin) have changed little during evolution.

Although the mechanisms of protein evolution have been amply addressed (e.g. gene shuffling, duplication, …), the nature of the evolutionary pressure that drives formation of protein-protein interfaces – and, moreover, inter-subunit communication - is less well understood, since an advantage at the level of organismic function might only become deterministic once the evolution of a number of correlated factors has taken place. For example, it is currently not known how full, complex metabolic pathways have formed.

At a molecular level, the study of the evolution of protein-protein recognition is complicated. First, difficulties arise from the fact that many recognition intermediates appear not to have been retained in current organisms. But, more

importantly, a major limitation is the impossibility to correlate the multiple genome sequences available to date with the assembly properties of the protein the code for. Amino acid residues involved in the stabilization of protein interactions are able to act through all their functional groups (including main chain, aliphatic fraction and/or terminal groups). In fact, the global players in interface formation are often residues able to display a broad spectrum of molecular interactions of different nature (such as e.g. Arg). Thus, sequence conservation at interfaces is often rather permissive. Effectively, current studies that emphasize key functional groups in interface amino acids can be misleading. Since conclusions based on pure sequence comparison data are not sufficient to elucidate the evolutionary principles governing macromolecular recognition, the experimental investigation of protein structures, their interfaces and their assembly properties is inevitable.

5   Bibliography

(1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr **50**, 760-763.

Altschul, S. F., and Lipman, D. J. (1990). Protein database searches for multiple alignments. Proc Natl Acad Sci U S A **87**, 5509-5513.

Anderson, K. S., Miles, E. W., and Johnson, K. A. (1991). Serine modulates substrate channeling in tryptophan synthase. A novel intersubunit triggering mechanism. J Biol Chem **266**, 8020-8033.

Au, Y. (2004). The muscle ultrastructure: a structural perspective of the sarcomere. Cell Mol Life Sci **61**, 3016-3033.

Bahadur, R. P., Chakrabarti, P., Rodier, F., and Janin, J. (2003). Dissecting subunit interfaces in homodimeric proteins. Proteins **53**, 708-719.

Banik, U., Zhu, D. M., Chock, P. B., and Miles, E. W. (1995). The tryptophan synthase alpha 2 beta 2 complex: kinetic studies with a mutant enzyme (beta K87T) to provide evidence for allosteric activation by an aminoacrylate intermediate. Biochemistry **34**, 12704-12711.

Beltrao, P., and Serrano, L. (2007). Specificity and evolvability in eukaryotic protein interaction networks. PLoS Comput Biol **3**, e25.

Bernal, J. D. (1965). The structure of water and its biological implications. Symp Soc Exp Biol **19**, 17-32.

Bisig, D., Weber, P., Vaughan, L., Winterhalter, K. H., and Piontek, K. (1999). Purification, crystallization and preliminary crystallographic studies of a two fibronectin type-III domain segment from chicken tenascin encompassing the heparin- and contactin-binding regions. Acta Crystallogr D Biol Crystallogr **55**, 1069-1073.

Bogan, A. A., and Thorn, K. S. (1998). Anatomy of hot spots in protein interfaces. J Mol Biol **280**, 1-9.

Bork, P., Holm, L., and Sander, C. (1994). The immunoglobulin fold. Structural classification, sequence patterns and common core. J Mol Biol **242**, 309-320.

Brinda, K. V., and Vishveshwara, S. (2005). Oligomeric protein structure networks: insights into protein-protein interactions. BMC Bioinformatics **6**, 296.

Brunger, A. T. (1993). Assessment of phase accuracy by cross validation: the free R value. Methods and applications. Acta Crystallogr D Biol Crystallogr **49**, 24-36.

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., et al. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr **54**, 905-921.

Brzovic, P. S., Kayastha, A. M., Miles, E. W., and Dunn, M. F. (1992a). Substitution of glutamic acid 109 by aspartic acid alters the substrate specificity and catalytic activity of the

beta-subunit in the tryptophan synthase bienzyme complex from Salmonella typhimurium. Biochemistry **31**, 1180-1190.

Brzovic, P. S., Ngo, K., and Dunn, M. F. (1992b). Allosteric interactions coordinate catalytic activity between successive metabolic enzymes in the tryptophan synthase bienzyme complex. Biochemistry **31**, 3831-3839.

Caffrey, D. R., Somaroo, S., Hughes, J. D., Mintseris, J., and Huang, E. S. (2004). Are protein-protein interfaces more conserved in sequence than the rest of the protein surface? Protein Sci **13**, 190-202.

Cambillau, C., and Claverie, J. M. (2000). Structural and genomic correlates of hyperthermostability. J Biol Chem **275**, 32383-32386.

Campbell, I. D., and Spitzfaden, C. (1994). Building proteins with fibronectin type III modules. Structure **2**, 333-337.

Cantino, M. E., Brown, L. D., Chew, M., Luther, P. K., and Squire, J. M. (2000). A-band architecture in vertebrate skeletal muscle: polarity of the myosin head array. J Muscle Res Cell Motil **21**, 681-690.

Cantino, M. E., Chew, M. W., Luther, P. K., Morris, E., and Squire, J. M. (2002). Structure and nucleotide-dependent changes of thick filaments in relaxed and rigor plaice fin muscle. J Struct Biol **137**, 164-175.

Chakrabarti, P., and Janin, J. (2002). Dissecting protein-protein recognition sites. Proteins **47**, 334-343.

Chakravarty, S., and Varadarajan, R. (2000). Elucidation of determinants of protein stability through genome sequence analysis. FEBS Lett **470**, 65-69.

Chatzidimitriou-Dreismann, C. A., Vos, M., Kleiner, C., and Abdul-Redah, T. (2003). Comparison of electron and neutron compton scattering from entangled protons in a solid polymer. Phys Rev Lett **91**, 057403.

Christen, P., and Mehta, P. K. (2001). From cofactor to enzymes. The molecular evolution of pyridoxal-5'-phosphate-dependent enzymes. Chem Rec **1**, 436-447.

Clackson, T., and Wells, J. A. (1995). A hot spot of binding energy in a hormone-receptor interface. Science **267**, 383-386.

Clough, S., Beers, Y., Klein, G., and Rothman, L. (1973). Dipole moment of water from Stark measurements of H2O, HDO, and D2O. J chem Phys **59**, 2254-2259.

Cole, C., and Warwicker, J. (2002). Side-chain conformational entropy at protein-protein interfaces. Protein Sci **11**, 2860-2870.

Copie, V., Tomita, Y., Akiyama, S. K., Aota, S., Yamada, K. M., Venable, R. M., Pastor, R. W., Krueger, S., and Torchia, D. A. (1998). Solution structure and dynamics of linked cell attachment modules of mouse fibronectin containing the RGD and synergy regions: comparison with the human fibronectin crystal structure. J Mol Biol **277**, 663-682.

Crawford, I. P., and Ito, J. (1964). Serine Deamination by the B Protein of Escherichia Coli Tryptophan Synthetase. Proc Natl Acad Sci U S A **51**, 390-397.

de Pereda, J. M., Wiche, G., and Liddington, R. C. (1999). Crystal structure of a tandem pair of fibronectin type III domains from the cytoplasmic tail of integrin alpha6beta4. Embo J **18**, 4087-4095.

DeLano, W. L. (2002a). The PyMOL molecular Graphics System (San Carlos, CA, USA: DeLano Scietific).

DeLano, W. L. (2002b). Unraveling hot spots in binding interfaces: progress and challenges. Curr Opin Struct Biol **12**, 14-20.

DeLano, W. L., Ultsch, M. H., de Vos, A. M., and Wells, J. A. (2000). Convergent solutions to binding at a protein-protein interface. Science **287**, 1279-1283.

Dill, K. A. (1990). Dominant forces in protein folding. Biochemistry **29**, 7133-7155.

Dill, K. A., Truskett, T. M., Vlachy, V., and Hribar-Lee, B. (2005). Modeling water, the hydrophobic effect, and ion solvation. Annu Rev Biophys Biomol Struct **34**, 173-199.

Elcock, A. H. (1998). The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins. J Mol Biol **284**, 489-502.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr **60**, 2126-2132.

Evans, P. (2006). Scaling and assessment of data quality. Acta Crystallogr D Biol Crystallogr **62**, 72-82.

Ferrari, D., Niks, D., Yang, L. H., Miles, E. W., and Dunn, M. F. (2003). Allosteric communication in the tryptophan synthase bienzyme complex: roles of the beta-subunit aspartate 305-arginine 141 salt bridge. Biochemistry **42**, 7807-7818.

Finney, J. L. (2004). Water? What's so special about it? Philos Trans R Soc Lond B Biol Sci **359**, 1145-1163; discussion 1163-1145, 1323-1148.

Fitz-Gibbon, S. T., Ladner, H., Kim, U. J., Stetter, K. O., Simon, M. I., and Miller, J. H. (2002). Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum aerophilum. Proc Natl Acad Sci U S A **99**, 984-989.

Futterer, O., Angelov, A., Liesegang, H., Gottschalk, G., Schleper, C., Schepers, B., Dock, C., Antranikian, G., and Liebl, W. (2004). Genome sequence of Picrophilus torridus and its implications for life around pH 0. Proc Natl Acad Sci U S A **101**, 9091-9096.

Gerstein, M., and Levitt, M. (1997). A structural census of the current population of protein sequences. Proc Natl Acad Sci U S A **94**, 11911-11916.

Goldberg, M. E., and Baldwin, R. L. (1967). Interactions between the subunits of the tryptophan synthetase of Escherichia coli. Optical properties of an intermediate bound to the alpha-2 beta-2 complex. Biochemistry **6**, 2113-2119.

Granzier, H. L., and Labeit, S. (2005). Titin and its associated proteins: the third myofilament system of the sarcomere. Adv Protein Chem **71**, 89-119.

Guex, N., and Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis **18**, 2714-2723.

Bibliography

Guharoy, M., and Chakrabarti, P. (2005). Conservation and relative importance of residues across protein-protein interfaces. Proc Natl Acad Sci U S A **102**, 15447-15452.

Gutteridge, A., and Thornton, J. (2004). Conformational change in substrate binding, catalysis and product release: an open and shut case? FEBS Lett **567**, 67-73.

Gutteridge, A., and Thornton, J. M. (2005). Understanding nature's catalytic toolkit. Trends Biochem Sci **30**, 622-629.

Hamill, S. J., Cota, E., Chothia, C., and Clarke, J. (2000). Conservation of folding and stability within a protein family: the tyrosine corner as an evolutionary cul-de-sac. J Mol Biol **295**, 641-649.

Haney, P. J., Badger, J. H., Buldak, G. L., Reich, C. I., Woese, C. R., and Olsen, G. J. (1999). Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic Methanococcus species. Proc Natl Acad Sci U S A **96**, 3578-3583.

Hanson, J., O'Brien, E. J., and Bennett, P. M. (1971). Structure of the myosin-containing filament assembly (A-segment) separated from frog skeletal muscle. J Mol Biol **58**, 865-871.

Harpaz, Y., and Chothia, C. (1994). Many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to a new structural set which is close to that containing variable domains. J Mol Biol **238**, 528-539.

Hermansson, K. (1984) The electron distribution in the bound water molecule, University of Uppsala, Sweden, Uppsala.

Hettwer, S., and Sterner, R. (2002). A novel tryptophan synthase beta-subunit from the hyperthermophile Thermotoga maritima. Quaternary structure, steady-state kinetics, and putative physiological role. J Biol Chem **277**, 8194-8201.

Higgins, D. G., Labeit, S., Gautel, M., and Gibson, T. J. (1994). The evolution of titin and related giant muscle proteins. J Mol Evol **38**, 395-404.

Hioki, Y., Ogasahara, K., Lee, S. J., Ma, J., Ishida, M., Yamagata, Y., Matsuura, Y., Ota, M., Ikeguchi, M., Kuramitsu, S., and Yutani, K. (2004). The crystal structure of the tryptophan synthase beta subunit from the hyperthermophile Pyrococcus furiosus. Investigation of stabilization factors. Eur J Biochem **271**, 2624-2635.

Holm, L., and Sander, C. (1993). Protein structure comparison by alignment of distance matrices. J Mol Biol **233**, 123-138.

Huber, A. H., Wang, Y. M., Bieber, A. J., and Bjorkman, P. J. (1994). Crystal structure of tandem type III fibronectin domains from Drosophila neuroglian at 2.0 A. Neuron **12**, 717-731.

Hunter, C. A., Singh, J., and Thornton, J. M. (1991). Pi-pi interactions: the geometry and energetics of phenylalanine-phenylalanine interactions in proteins. J Mol Biol **218**, 837-846.

Hur, O., Niks, D., Casino, P., and Dunn, M. F. (2002). Proton transfers in the beta-reaction catalyzed by tryptophan synthase. Biochemistry **41**, 9991-10001.

Hutchinson, E. G., and Thornton, J. M. (1996). PROMOTIF--a program to identify and analyze structural motifs in proteins. Protein Sci **5**, 212-220.

Huxley, A. F., and Niedergerke, R. (1954). Structural changes in muscle during contraction; interference microscopy of living muscle fibres. Nature **173**, 971-973.

Huxley, H. E., and Brown, W. (1967). The low-angle x-ray diagram of vertebrate striated muscle and its behaviour during contraction and rigor. J Mol Biol **30**, 383-434.

Hyde, C. C., Ahmed, S. A., Padlan, E. A., Miles, E. W., and Davies, D. R. (1988). Three-dimensional structure of the tryptophan synthase alpha 2 beta 2 multienzyme complex from Salmonella typhimurium. J Biol Chem **263**, 17857-17871.

Improta, S., Politou, A. S., and Pastore, A. (1996). Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity. Structure **4**, 323-337.

Ispolatov, I., Yuryev, A., Mazo, I., and Maslov, S. (2005). Binding properties and evolution of homodimers in protein-protein interaction networks. Nucleic Acids Res **33**, 3629-3635.

Jaenicke, R., and Bohm, G. (1998). The stability of proteins in extreme environments. Curr Opin Struct Biol **8**, 738-748.

Jahn, T. R., and Radford, S. E. (2005). The Yin and Yang of protein folding. Febs J **272**, 5962-5970.

Jones, S., Marin, A., and Thornton, J. M. (2000). Protein domain interfaces: characterization and comparison with oligomeric protein interfaces. Protein Eng **13**, 77-82.

Jones, S., and Thornton, J. M. (1997). Analysis of protein-protein interaction sites using surface patches. J Mol Biol **272**, 121-132.

Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr A **47** (Pt 2), 110-119.

Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J Appl Cryst **26**, 795-800.

Kawarabayasi, Y., Hino, Y., Horikawa, H., Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Ankai, A., Kosugi, H., Hosoyama, A., et al. (2001). Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, Sulfolobus tokodaii strain7. DNA Res **8**, 123-140.

Kawasaki, H., Bauerle, R., Zon, G., Ahmed, S. A., and Miles, E. W. (1987). Site-specific mutagenesis of the alpha subunit of tryptophan synthase from Salmonella typhimurium. Changing arginine 179 to leucine alters the reciprocal transmission of substrate-induced conformational changes between the alpha and beta 2 subunits. J Biol Chem **262**, 10678-10683.

Kenny, P. A., Liston, E. M., and Higgins, D. G. (1999). Molecular evolution of immunoglobulin and fibronectin domains in titin and related muscle proteins. Gene **232**, 11-23.

Kirschner, K., Lane, A. N., and Strasser, A. W. (1991). Reciprocal communication between the lyase and synthase active sites of the tryptophan synthase bienzyme complex. Biochemistry **30**, 472-478.

Kirschner, M., and Gerhart, J. (1998). Evolvability. Proc Natl Acad Sci U S A **95**, 8420-8427.

Knochel, T., Ivens, A., Hester, G., Gonzalez, A., Bauerle, R., Wilmanns, M., Kirschner, K., and Jansonius, J. N. (1999). The crystal structure of anthranilate synthase from Sulfolobus solfataricus: functional implications. Proc Natl Acad Sci U S A **96**, 9479-9484.

Kryger, G., Silman, I., and Sussman, J. L. (1999). Structure of acetylcholinesterase complexed with E2020 (Aricept): implications for the design of new anti-Alzheimer drugs. Structure **7**, 297-307.

Kumar, S., and Nussinov, R. (2002). Close-range electrostatic interactions in proteins. Chembiochem **3**, 604-617.

Labeit, S., and Kolmerer, B. (1995). The complete primary structure of human nebulin and its correlation to muscle structure. J Mol Biol **248**, 308-315.

Lane, A. N., and Kirschner, K. (1983a). The catalytic mechanism of tryptophan synthase from Escherichia coli. Kinetics of the reaction of indole with the enzyme--L-serine complexes. Eur J Biochem **129**, 571-582.

Lane, A. N., and Kirschner, K. (1983b). The mechanism of binding of L-serine to tryptophan synthase from Escherichia coli. Eur J Biochem **129**, 561-570.

Laskowski, R. A., Moss, D. S., and Thornton, J. M. (1993). Main-chain bond lengths and bond angles in protein structures. J Mol Biol **231**, 1049-1067.

Lazaridis, T. (2001). Solvent size vs cohesive energy as the origin of hydrophobicity. Acc Chem Res **34**, 931-937.

Lazaridis, T., and Paulaitis, M. (1992). Entropy of hydrophobic hydration: a new statistical mechanical formulation. J Phys Chem **96**, 3847.

Leahy, D. J., Aukhil, I., and Erickson, H. P. (1996). 2.0 A crystal structure of a four-domain segment of human fibronectin encompassing the RGD loop and synergy region. Cell **84**, 155-164.

Leahy, D. J., Hendrickson, W. A., Aukhil, I., and Erickson, H. P. (1992). Structure of a fibronectin type III domain from tenascin phased by MAD analysis of the selenomethionyl protein. Science **258**, 987-991.

Lee, B. (1985). The physical origin of the low solubility of nonpolar solutes in water. Biopolymers **24**, 813-823.

Lee, C. E., Goodfellow, C., Javid-Majd, F., Baker, E. N., and Shaun Lott, J. (2006). The crystal structure of TrpD, a metabolic enzyme essential for lung colonization by Mycobacterium tuberculosis, in complex with its substrate phosphoribosylpyrophosphate. J Mol Biol **355**, 784-797.

Lee, S. J., Ogasahara, K., Ma, J., Nishio, K., Ishida, M., Yamagata, Y., Tsukihara, T., and Yutani, K. (2005). Conformational Changes in the tryptophan synthase from a hyperthermophile upon alpha2beta2 complex formation: crystal structure of the complex. Biochemistry **44**, 11417-11427.

Leja, C. A., Woehl, E. U., and Dunn, M. F. (1995). Allosteric linkages between beta-site covalent transformations and alpha-site activation and deactivation in the tryptophan synthase bienzyme complex. Biochemistry **34**, 6552-6561.

Leopoldseder (2005) Neuartige Tryptophan Synthase aus Hyperthermophilen: Charakterisierung der Enzyme aus Sulfolobus solfataricus.

Leopoldseder, S., Hettwer, S., and Sterner, R. (2006). Evolution of multi-enzyme complexes: the case of tryptophan synthase. Biochemistry **45**, 14111-14119.

Leslie, A. G. (2006). The integration of macromolecular diffraction data. Acta Crystallogr D Biol Crystallogr **62**, 48-57.

Levy, Y., and Onuchic, J. N. (2004). Water and proteins: a love-hate relationship. Proc Natl Acad Sci U S A **101**, 3325-3326.

Levy, Y., and Onuchic, J. N. (2006). Mechanisms of protein assembly: lessons from minimalist models. Acc Chem Res **39**, 135-142.

Liversage, A. D., Holmes, D., Knight, P. J., Tskhovrebova, L., and Trinick, J. (2001). Titin and the sarcomere symmetry paradox. J Mol Biol **305**, 401-409.

Lo Conte, L., Chothia, C., and Janin, J. (1999). The atomic structure of protein-protein recognition sites. J Mol Biol **285**, 2177-2198.

Luther, P. K., Munro, P. M., and Squire, J. M. (1981). Three-dimensional structure of the vertebrate muscle A-band. III. M-region structure and myosin filament symmetry. J Mol Biol **151**, 703-730.

Lymn, R. W., and Taylor, E. W. (1971). Mechanism of adenosine triphosphate hydrolysis by actomyosin. Biochemistry **10**, 4617-4624.

Main, A. L., Harvey, T. S., Baron, M., Boyd, J., and Campbell, I. D. (1992). The three-dimensional structure of the tenth type III module of fibronectin: an insight into RGD-mediated interactions. Cell **71**, 671-678.

Makhatadze, G. I., and Privalov, P. L. (1995). Energetics of protein structure. Adv Protein Chem **47**, 307-425.

Marino, M., Svergun, D. I., Kreplak, L., Konarev, P. V., Maco, B., Labeit, D., and Mayans, O. (2005). Poly-Ig tandems from I-band titin share extended domain arrangements irrespective of the distinct features of their modular constituents. J Muscle Res Cell Motil **26**, 355-365.

Marino, M., Zou, P., Svergun, D., Garcia, P., Edlich, C., Simon, B., Wilmanns, M., Muhle-Goll, C., and Mayans, O. (2006). The Ig doublet Z1Z2: a model system for the hybrid analysis of conformational dynamics in Ig tandems from titin. Structure **14**, 1437-1447.

Marsden, R. L., Ranea, J. A., Sillero, A., Redfern, O., Yeats, C., Maibaum, M., Lee, D., Addou, S., Reeves, G. A., Dallman, T. J., and Orengo, C. A. (2006). Exploiting protein

structure data to explore the evolution of protein function and biological complexity. Philos Trans R Soc Lond B Biol Sci **361**, 425-440.

Maruyama, K., Kimura, S., Ishi, T., Kuroda, M., and Ohashi, K. (1977). beta-actinin, a regulatory protein of muscle. Purification, characterization and function. J Biochem (Tokyo) **81**, 215-232.

Matthews, B. W. (1968). Solvent content of protein crystals. J Mol Biol 33, 491-497.

Mayans, O., Ivens, A., Nissen, L. J., Kirschner, K., and Wilmanns, M. (2002). Structural analysis of two enzymes catalysing reverse metabolic reactions implies common ancestry. Embo J **21**, 3245-3254.

Mayans, O., van der Ven, P. F., Wilm, M., Mues, A., Young, P., Furst, D. O., Wilmanns, M., and Gautel, M. (1998). Structural basis for activation of the titin kinase domain during myofibrillogenesis. Nature **395**, 863-869.

Mayans, O., Wuerges, J., Canela, S., Gautel, M., and Wilmanns, M. (2001). Structural evidence for a possible role of reversible disulphide bridge formation in the elasticity of the muscle protein titin. Structure **9**, 331-340.

McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C., and Read, R. J. (2005). Likelihood-enhanced fast translation functions. Acta Crystallogr D Biol Crystallogr **61**, 458-464.

Meyer, E. A., Castellano, R. K., and Diederich, F. (2003). Interactions with aromatic rings in chemical and biological recognition. Angew Chem Int Ed Engl **42**, 1210-1250.

Miles, E. W. (1979). Tryptophan synthase: structure, function, and subunit interaction. Adv Enzymol Relat Areas Mol Biol **49**, 127-186.

Miles, E. W. (1995). Tryptophan synthase. Structure, function, and protein engineering. Subcell Biochem **24**, 207-254.

Miles, E. W. (2001). Tryptophan synthase: a multienzyme complex with an intramolecular tunnel. Chem Rec **1**, 140-151.

Miles, E. W., Houck, D. R., and Floss, H. G. (1982). Stereochemistry of sodium borohydride reduction of tryptophan synthase of Escherichia coli and its amino acid Schiff's bases. J Biol Chem **257**, 14203-14210.

Miles, E. W., Kawasaki, H., Ahmed, S. A., Morita, H., and Nagata, S. (1989). The beta subunit of tryptophan synthase. Clarification of the roles of histidine 86, lysine 87, arginine 148, cysteine 170, and cysteine 230. J Biol Chem **264**, 6280-6287.

Mitchell, J. B., Nandi, C. L., McDonald, I. K., Thornton, J. M., and Price, S. L. (1994). Amino/aromatic interactions in proteins: is the evidence stacked against hydrogen bonding? J Mol Biol **239**, 315-331.

Moolman-Smook, J., Flashman, E., de Lange, W., Li, Z., Corfield, V., Redwood, C., and Watkins, H. (2002). Identification of novel interactions between domains of Myosin binding protein-C that are modulated by hypertrophic cardiomyopathy missense mutations. Circ Res **91**, 704-711.

153

Morokuma, K., and Pedersen, L. (1968). Molecular-Orbital Studies of Hydrogen Bonds. An Ab Initio Calculation for Dimeric H2O. J Chem Phys **48**, 3275.

Morollo, A. A., and Bauerle, R. (1993). Characterization of composite aminodeoxyisochorismate synthase and aminodeoxyisochorismate lyase activities of anthranilate synthase. Proc Natl Acad Sci U S A **90**, 9983-9987.

Mozzarelli, A., Peracchi, A., Rossi, G. L., Ahmed, S. A., and Miles, E. W. (1989). Microspectrophotometric studies on single crystals of the tryptophan synthase alpha 2 beta 2 complex demonstrate formation of enzyme-substrate intermediates. J Biol Chem **264**, 15774-15780.

Mozzarelli, A., Peracchi, A., Rovegno, B., Dale, G., Rossi, G. L., and Dunn, M. F. (2000). Effect of pH and monovalent cations on the formation of quinonoid intermediates of the tryptophan synthase alpha(2)beta(2) complex in solution and in the crystal. J Biol Chem **275**, 6956-6962.

Mrosek, M., Labeit, D., Witt, S., Heerklotz, H., von Castelmur, E., Labeit, S., and Mayans, O. (2007). Molecular determinants for the recruitment of the ubiquitin-ligase MuRF-1 onto M-line titin. Faseb J **21**, 1383-1392.

Muhle-Goll, C., Habeck, M., Cazorla, O., Nilges, M., Labeit, S., and Granzier, H. (2001). Structural and functional studies of titin's fn3 modules reveal conserved surface patterns and binding to myosin S1--a possible role in the Frank-Starling mechanism of the heart. J Mol Biol **313**, 431-447.

Muhle-Goll, C., Pastore, A., and Nilges, M. (1998). The three-dimensional structure of a type I module from titin: a prototype of intracellular fibronectin type III domains. Structure **6**, 1291-1302.

Nagata, S., Hyde, C. C., and Miles, E. W. (1989). The alpha subunit of tryptophan synthase. Evidence that aspartic acid 60 is a catalytic residue and that the double alteration of residues 175 and 211 in a second-site revertant restores the proper geometry of the substrate binding site. J Biol Chem **264**, 6288-6296.

Navaza (1994). AMoRe: an Automated Package for Molecular Replacement. Acta Crystallogr D Biol Crystallogr **50**, 157-163.

Nicholls, A., Sharp, K. A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins **11**, 281-296.

Ofran, Y., and Rost, B. (2003). Analysing six types of protein-protein interfaces. J Mol Biol **325**, 377-387.

Onuchic, J. N., and Wolynes, P. G. (2004). Theory of protein folding. Curr Opin Struct Biol **14**, 70-75.

Oshima, K., Takezawa, Y., Sugimoto, Y., Kiyotoshi, M., and Wakabayashi, K. (2003). Modeling analysis of myosin-based meridional X-ray reflections from frog skeletal muscles in relaxed and contracting states. Adv Exp Med Biol **538**, 243-249.

Paci, E., and Karplus, M. (2000). Unfolding proteins by external forces and temperature: the importance of topology and energetics. Proc Natl Acad Sci U S A **97**, 6521-6526.

Pagni, M., Ioannidis, V., Cerutti, L., Zahn-Zabal, M., Jongeneel, C. V., and Falquet, L. (2004). MyHits: a new interactive resource for protein annotation and domain identification. Nucleic Acids Res **32**, W332-335.

Pearl, F., Todd, A., Sillitoe, I., Dibley, M., Redfern, O., Lewis, T., Bennett, C., Marsden, R., Grant, A., Lee, D., et al. (2005). The CATH Domain Structure Database and related resources Gene3D and DHS provide comprehensive domain family information for genome analysis. Nucleic Acids Res **33**, D247-251.

Peracchi, A., Bettati, S., Mozzarelli, A., Rossi, G. L., Miles, E. W., and Dunn, M. F. (1996). Allosteric regulation of tryptophan synthase: effects of pH, temperature, and alpha-subunit ligands on the equilibrium distribution of pyridoxal 5'-phosphate-L-serine intermediates. Biochemistry **35**, 1872-1880.

Pereira-Leal, J. B., Levy, E. D., Kamp, C., and Teichmann, S. A. (2007). Evolution of protein complexes by duplication of homomeric interactions. Genome Biol **8**, R51.

Pereira-Leal, J. B., Levy, E. D., and Teichmann, S. A. (2006). The origins and evolution of functional modules: lessons from protein complexes. Philos Trans R Soc Lond B Biol Sci **361**, 507-517.

Perrakis, A., Morris, R., and Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. Nat Struct Biol **6**, 458-463.

Pfuhl, M., and Pastore, A. (1995). Tertiary structure of an immunoglobulin-like domain from the giant muscle protein titin: a new member of the I set. Structure **3**, 391-401.

Pickett, S. D., and Sternberg, M. J. (1993). Empirical scale of side-chain conformational entropy in protein folding. J Mol Biol **231**, 825-839.

Prasad, G. S. (2001). Glycine rich P-loop motif in deoxyuridine pyrophosphatase. Curr Protein Pept Sci **2**, 301-311.

Raboni, S., Bettati, S., and Mozzarelli, A. (2005). Identification of the geometric requirements for allosteric communication between the alpha- and beta-subunits of tryptophan synthase. J Biol Chem **280**, 13450-13456.

Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. J Mol Biol **7**, 95-99.

Raushel, F. M., Thoden, J. B., and Holden, H. M. (2003). Enzymes with molecular tunnels. Acc Chem Res **36**, 539-548.

Rhee, S., Miles, E. W., and Davies, D. R. (1998a). Cryo-crystallography of a true substrate, indole-3-glycerol phosphate, bound to a mutant (alphaD60N) tryptophan synthase alpha2beta2 complex reveals the correct orientation of active site alphaGlu49. J Biol Chem **273**, 8553-8555.

Rhee, S., Miles, E. W., Mozzarelli, A., and Davies, D. R. (1998b). Cryocrystallography and microspectrophotometry of a mutant (alpha D60N) tryptophan synthase alpha 2 beta 2 complex reveals allosteric roles of alpha Asp60. Biochemistry **37**, 10653-10659.

Rhee, S., Parris, K. D., Ahmed, S. A., Miles, E. W., and Davies, D. R. (1996). Exchange of K+ or Cs+ for Na+ induces local and long-range changes in the three-dimensional structure of the tryptophan synthase alpha2beta2 complex. Biochemistry **35**, 4211-4221.

Rhee, S., Parris, K. D., Hyde, C. C., Ahmed, S. A., Miles, E. W., and Davies, D. R. (1997). Crystal structures of a mutant (betaK87T) tryptophan synthase alpha2beta2 complex with ligands bound to the active sites of the alpha- and beta-subunits reveal ligand-induced conformational changes. Biochemistry **36**, 7664-7680.

Ro, H. S., and Miles, E. W. (1999). Structure and function of the tryptophan synthase alpha(2)beta(2) complex. Roles of beta subunit histidine 86. J Biol Chem **274**, 36439-36445.

Rodier, F., Bahadur, R. P., Chakrabarti, P., and Janin, J. (2005). Hydration of protein-protein interfaces. Proteins **60**, 36-45.

Rome, E., Offer, G., and Pepe, F. A. (1973). X-ray diffraction of muscle labelled with antibody to C-protein. Nat New Biol **244**, 152-154.

Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H., and Zehfus, M. H. (1985). Hydrophobicity of amino acid residues in globular proteins. Science **229**, 834-838.

Rossmann, M. G. (1990). The molecular replacement method. Acta Crystallogr A **46** (Pt 2), 73-82.

Rossmann, M. G. (2001). Molecular replacement--historical background. Acta Crystallogr D Biol Crystallogr **57**, 1360-1366.

Ryu, S. E., Kwong, P. D., Truneh, A., Porter, T. G., Arthos, J., Rosenberg, M., Dai, X. P., Xuong, N. H., Axel, R., Sweet, R. W., and et al. (1990). Crystal structure of an HIV-binding recombinant fragment of human CD4. Nature **348**, 419-426.

Sachpatzidis, A., Dealwis, C., Lubetsky, J. B., Liang, P. H., Anderson, K. S., and Lolis, E. (1999). Crystallographic studies of phosphonate-based alpha-reaction transition-state analogues complexed to tryptophan synthase. Biochemistry **38**, 12665-12674.

Saha, R. P., Bahadur, R. P., Pal, A., Mandal, S., and Chakrabarti, P. (2006). ProFace: a server for the analysis of the physicochemical features of protein-protein interfaces. BMC Struct Biol **6**, 11.

Saraste, M., Sibbald, P. R., and Wittinghofer, A. (1990). The P-loop--a common motif in ATP- and GTP-binding proteins. Trends Biochem Sci **15**, 430-434.

Sarker, K. D., and Hardman, J. K. (1995). Affinities of phosphorylated substrates for the E. coli tryptophan synthase alpha-subunit: roles of Ser-235 and helix-8' dipole. Proteins **21**, 130-139.

Schiaretti, F., Bettati, S., Viappiani, C., and Mozzarelli, A. (2004). pH dependence of tryptophan synthase catalytic mechanism: I. The first stage, the beta-elimination reaction. J Biol Chem 279, 29572-29582.

Schneider, T. R., Gerhardt, E., Lee, M., Liang, P. H., Anderson, K. S., and Schlichting, I. (1998). Loop closure and intersubunit communication in tryptophan synthase. Biochemistry **37**, 5394-5406.

Bibliography

Schultheiss, T., Lin, Z. X., Lu, M. H., Murray, J., Fischman, D. A., Weber, K., Masaki, T., Imamura, M., and Holtzer, H. (1990). Differential distribution of subsets of myofibrillar proteins in cardiac nonstriated and striated myofibrils. J Cell Biol **110**, 1159-1172.

Schuttelkopf, A. W., and van Aalten, D. M. (2004). PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr D Biol Crystallogr **60**, 1355-1363.

Sharma, A., Askari, J. A., Humphries, M. J., Jones, E. Y., and Stuart, D. I. (1999). Crystal structure of a heparin- and integrin-binding segment of human fibronectin. Embo J **18**, 1468-1479.

She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., Chan-Weiher, C. C., Clausen, I. G., Curtis, B. A., De Moors, A., et al. (2001). The complete genome of the crenarchaeon Sulfolobus solfataricus P2. Proc Natl Acad Sci U S A **98**, 7835-7840.

Shirvanee, L., Horn, V., and Yanofsky, C. (1990). Escherichia coli mutant trpA34 has an Asp----Asn change at active site residue 60 of the tryptophan synthetase alpha chain. J Biol Chem **265**, 6624-6625.

Soding, J., and Lupas, A. N. (2003). More than the sum of their parts: on the evolution of proteins from peptides. Bioessays **25**, 837-846.

Spitzfaden, C., Grant, R. P., Mardon, H. J., and Campbell, I. D. (1997). Module-module interactions in the cell binding region of fibronectin: stability, flexibility and specificity. J Mol Biol **265**, 565-579.

Squire, J. M., Al-Khayat, H. A., Knupp, C., and Luther, P. K. (2005). Molecular architecture in muscle contractile assemblies. Adv Protein Chem **71**, 17-87.

Squire, J. M., and Knupp, C. (2005). X-ray diffraction studies of muscle and the crossbridge cycle. Adv Protein Chem **71**, 195-255.

Squire, J. M., Roessle, M., and Knupp, C. (2004). New X-ray diffraction observations on vertebrate muscle: organisation of C-protein (MyBP-C) and troponin and evidence for unknown structures in the vertebrate A-band. J Mol Biol **343**, 1345-1363.

Sterner, R., and Hocker, B. (2005). Catalytic versatility, stability, and evolution of the (betaalpha)8-barrel enzyme fold. Chem Rev **105**, 4038-4055.

Sterner, R., and Liebl, W. (2001). Thermophilic adaptation of proteins. Crit Rev Biochem Mol Biol 36, 39-106.

Tanford, C. (1978). The hydrophobic effect and the organization of living matter. Science **200**, 1012-1018.

Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., Krylov, D. M., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., et al. (2003). The COG database: an updated version includes eukaryotes. BMC Bioinformatics **4**, 41.

Tesmer, J. J., Klem, T. J., Deras, M. L., Davisson, V. J., and Smith, J. L. (1996). The crystal structure of GMP synthetase reveals a novel catalytic triad and is a structural paradigm for two enzyme families. Nat Struct Biol **3**, 74-86.

Bibliography

Thorn, K. S., and Bogan, A. A. (2001). ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. Bioinformatics **17**, 284-285.

Vogel, C., Bashton, M., Kerrison, N. D., Chothia, C., and Teichmann, S. A. (2004a). Structure, function and evolution of multidomain proteins. Curr Opin Struct Biol **14**, 208-216.

Vogel, C., Berzuini, C., Bashton, M., Gough, J., and Teichmann, S. A. (2004b). Supra-domains: evolutionary units larger than single protein domains. J Mol Biol **336**, 809-823.

Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng **8**, 127-134.

Wang, K., McClure, J., and Tu, A. (1979). Titin: major myofibrillar components of striated muscle. Proc Natl Acad Sci U S A **76**, 3698-3702.

Waters, M. L. (2002). Aromatic interactions in model systems. Curr Opin Chem Biol **6**, 736-741.

Weber-Ban, E., Hur, O., Bagwell, C., Banik, U., Yang, L. H., Miles, E. W., and Dunn, M. F. (2001). Investigation of allosteric linkages in the regulation of tryptophan synthase: the roles of salt bridges and monovalent cations probed by site-directed mutation, optical spectroscopy, and kinetics. Biochemistry **40**, 3497-3511.

Weyand, M., and Schlichting, I. (1999). Crystal structure of wild-type tryptophan synthase complexed with the natural substrate indole-3-glycerol phosphate. Biochemistry **38**, 16469-16480.

Weyand, M., Schlichting, I., Marabotti, A., and Mozzarelli, A. (2002). Crystal structures of a new class of allosteric effectors complexed to tryptophan synthase. J Biol Chem **277**, 10647-10652.

Whiting, A., Wardale, J., and Trinick, J. (1989). Does titin regulate the length of muscle thick filaments? J Mol Biol **205**, 263-268.

Witt, C. C., Olivieri, N., Centner, T., Kolmerer, B., Millevoi, S., Morell, J., Labeit, D., Labeit, S., Jockusch, H., and Pastore, A. (1998). A survey of the primary structure and the interspecies conservation of I-band titin's elastic elements in vertebrates. J Struct Biol **122**, 206-215.

Woehl, E., and Dunn, M. F. (1999a). Mechanisms of monovalent cation action in enzyme catalysis: the first stage of the tryptophan synthase beta-reaction. Biochemistry **38**, 7118-7130.

Woehl, E., and Dunn, M. F. (1999b). Mechanisms of monovalent cation action in enzyme catalysis: the tryptophan synthase alpha-, beta-, and alpha beta-reactions. Biochemistry **38**, 7131-7141.

Woehl, E. U., and Dunn, M. F. (1995). Monovalent metal ions play an essential role in catalysis and intersubunit communication in the tryptophan synthase bienzyme complex. Biochemistry **34**, 9466-9476.

Xiao, L., and Honig, B. (1999). Electrostatic contributions to the stability of hyperthermophilic proteins. J Mol Biol **289**, 1435-1444.

Xie, G., Forst, C., Bonner, C., and Jensen, R. A. (2002). Significance of two distinct types of tryptophan synthase beta chain in Bacteria, Archaea and higher plants. Genome Biol **3**, RESEARCH0004.

Xie, G., Keyhani, N. O., Bonner, C. A., and Jensen, R. A. (2003). Ancient origin of the tryptophan operon and the dynamics of evolutionary change. Microbiol Mol Biol Rev **67**, 303-342, table of contents.

Xu, D., Tsai, C. J., and Nussinov, R. (1998). Mechanism and evolution of protein dimerization. Protein Sci **7**, 533-544.

Yamagata, Y., Ogasahara, K., Hioki, Y., Lee, S. J., Nakagawa, A., Nakamura, H., Ishida, M., Kuramitsu, S., and Yutani, K. (2001). Entropic stabilization of the tryptophan synthase alpha-subunit from a hyperthermophile, Pyrococcus furiosus. X-ray analysis and calorimetry. J Biol Chem **276**, 11062-11071.

Yanofsky, C. (2001). Advancing our knowledge in biochemistry, genetics, and microbiology through studies on tryptophan metabolism. Annu Rev Biochem **70**, 1-37.

Yutani, K., Ogasahara, K., Tsujita, T., Kanemoto, K., Matsumoto, M., Tanaka, S., Miyashita, T., Matsushiro, A., Sugino, Y., and Miles, E. W. (1987). Tryptophan synthase alpha subunit glutamic acid 49 is essential for activity. Studies with 19 mutants at position 49. J Biol Chem **262**, 13429-13433.

Zalkin, H. (1993). Overview of multienzyme systems in biosynthetic pathways. Biochem Soc Trans **21**, 203-207.

Zou, P., Pinotsis, N., Lange, S., Song, Y. H., Popov, A., Mavridis, I., Mayans, O. M., Gautel, M., and Wilmanns, M. (2006). Palindromic assembly of the giant muscle protein titin in the sarcomeric Z-disk. Nature **439**, 229-233.

6 Appendix

6.1 Materials

6.1.1 Organisms

Escherichia coli DH5α, XL-1 BLUE, BL21 (DE3) *Rosetta*

6.1.2 Plasmids for protein expression

Plasmids for protein expression were kindly provided from C. Muhle-Goll, EMBL Heidelberg.

Vector	Construct
pETM11	A77-A78

6.1.3 Proteins

Low Molecular Weight Marker (LMW-Marker, GE Healthcare)

Protein (from)	Molecular weight [kDa]
Phosphorylase b (rabbit muscle)	97
Albumin (bovine serum)	66
Ovalbumin (chicken egg white)	45
Carboanhydrase (bovine erythrocyte)	30
Trypsin inhibitor (soybean)	20
α-Lactalbumin (bovine milk)	14

Other proteins: HEW Lysozyme
                       DNase I

6.1.4 Chemicals

Antibiotica:

Chloramphenicol
Kanamycin

Protease inhibitors:

Complete Protease Inhibitor Cocktail (Roche AG)
PMSF (AppliChem)

Appendix

Other chemicals:

Chemical	Manufacturer
Acetic Acid, AcOH	Merck
Acrylamide, AA	BioRad
Ammonium peroxodisulfate, APS	Merck
Coomassie Brilliant Blue R-250	Fluka
Dioxane	Merck
Dithiothreitol, DTT	AppliChem
Ethanol, EtOH	Merck
Na_2Ethylendiamintetraacetic acid, EDTA	AppliChem
Ethylenglycol	Fluka
Glycerin	Fluka
Glycine, Gly	Fluka
Guanidinium hydrochloride, GndHCl	Sigma
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid sodium salt (HEPES)	Sigma
1,6-Hexanediol	Fluka
Hydrochloric acid, HCl	Merck
Imidazole	Fluka
iso-Propanol, *i*PrOH	Fluka
Isopropyl-β-D-thiogalactopyranoside, IPTG	AppliChem
Magnesium chloride hexahydrate, $MgCl_2$ x 6 H_2O	Fluka
Magnesium sulfate heptahydrate, $MgSO_4$ x 7 H_2O	Fluka
β-Mercaptoethanol, β-ME	Fluka
N,N'-Methylenbisacrylamide	Serva
Methylpenthandiol, MPD	Merck
PEG-200[1]	Fluka
PEG-400[1]	Fluka
PEG-1500[1]	Fluka
PEG-4000[1]	Fluka
PEG-8000[1]	Fluka
Sodium cacodylate trihydrate, Na_3AsO x 3 H_2O	Fluka
Sodium chloride, NaCl	Merck
Sodium Dodecylsulfate, SDS	BioRad
Sodium hydroxide, NaOH	Merck
N,N,N',N'-Tetramethylenethylendiamine, TEMED	Fluka
Tris(hydroxymethyl)-aminomethan, Tris	Merck
Tryptone	Beckton, Dickinson & Co.
Yeast extract	Beckton, Dickinson & Co.

[1]PEG-X: Polyethylenglycol-M_w

Crystallization kits:

Cryo I, Cryo II, Wizard I, Wizard II	Emerald Biostructures
Structure Screen 1, Structure Screen 2	Molecular Dimensions
Crystal Screen, Crystal Screen 2, Quick Screen (Phosphate)	Hampton Research

6.1.5 Buffers and media

All buffers, stock solutions and media, if not mentioned here, were prepared exactly like described in Sambrook & Russel (2001).

LB-medium:

Tryptone
Yeast extract
NaCl

IPTG stock solution:

IPTG stock solutions have been prepared by dissolving 238 mg/mL IPTG in H_2O. The solution has been filtered sterile, aliquoted and stored at -20°C.

Chloramphenicol stock solution (1000x):

Chloramphenicol stock solutions have been prepared by dissolving 34 mg/mL Chloramphenicol in EtOH (abs.). The stock solution was stored at -20°C.

Kanamycin stock solution (1000x):

Kanamycin stock solutions have been prepared by dissolving 50 mg/mL Kanamycin in ddH_2O and subsequent sterile filtration. The stock solution has been stored in aliquots at -20°C.

PEG stock solutions:

PEG stock-solutions from solid PEGs ($\geq$ 1000 M_w) were prepared in 50 % (w/v) stock solutions.

6.2   Additional material (projects)

6.2.1   Introduction
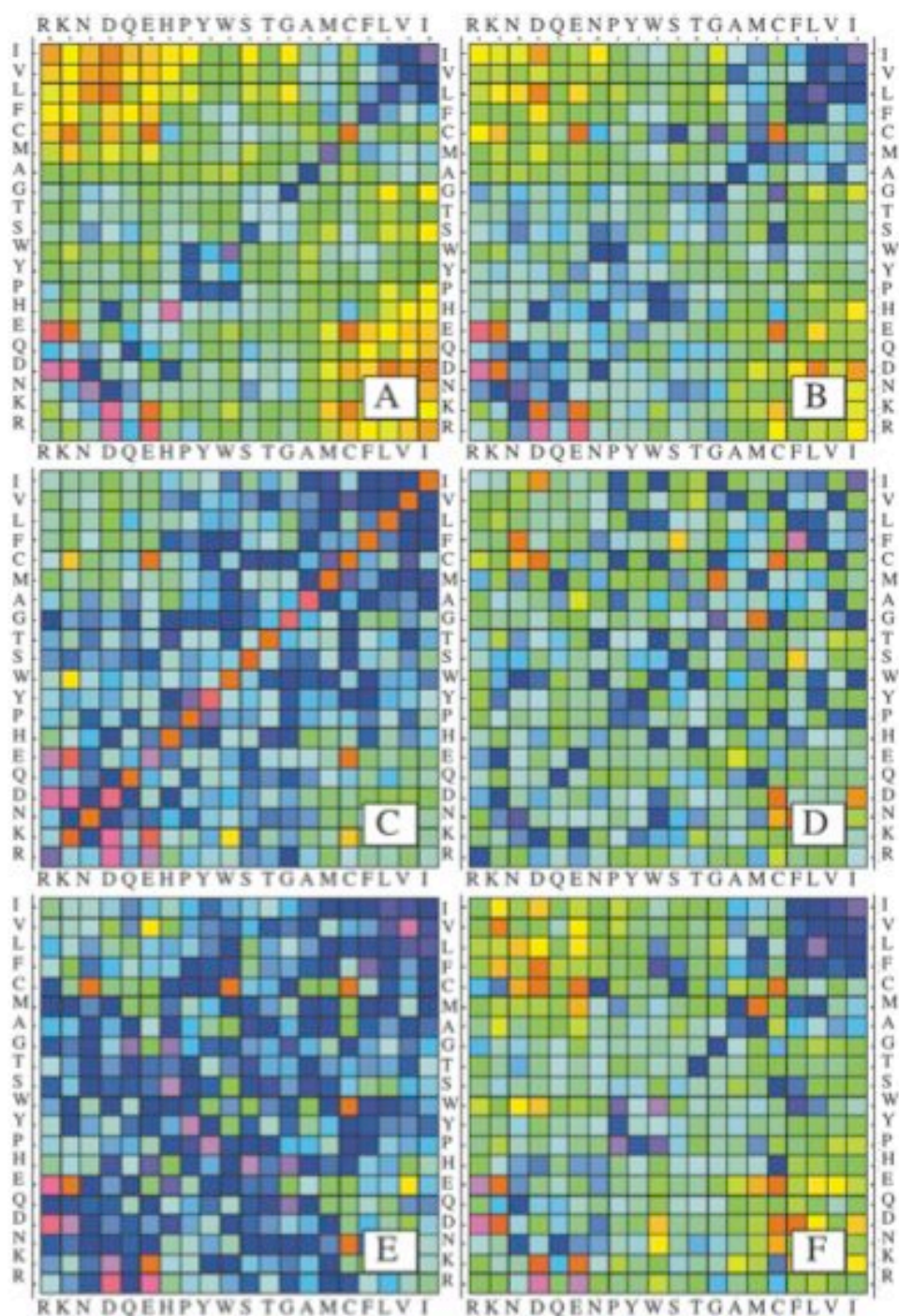
Figure 6.1 Residue-residue preferences.
(**A**) Intra-domain, (**B**) domain-domain, (**C**) obligatory homo-oligomers (homo-obligomers), (**D**) transient homo-oligomers (homo-complexes), (**E**) obligatory hetero-oligomers (hetero-obligomers), and (**F**) transient hetero-oligomers (hetero-complexes). A red square indicates that the interaction occurs more frequently than expected; a blue square indicates that it occurs less frequently than expected. The amino acid residues are ordered according to hydrophobicity,42 with Ile as the most hydrophobic and Arg as the least hydrophobic (adapted from Ofran & Rost, 2003).

6.2.2 *Ss*TrpB2b
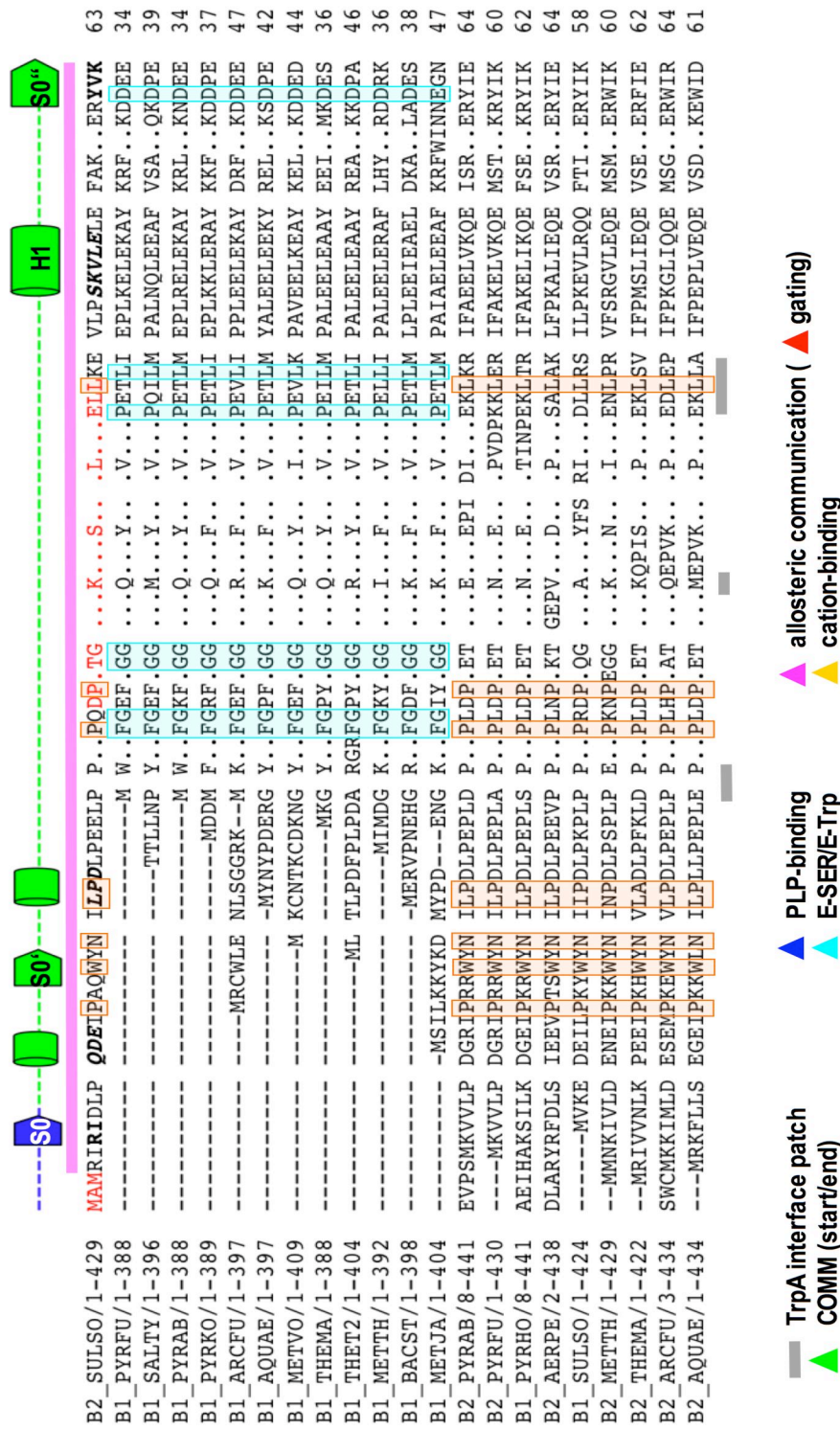
Figure 6.2a Structure-based sequence alignment between proteins of the TrpB1 and TrpB2 families.
This block corresponds to the N-terminal region in which the proteins adapt a different fold (indicated by the magenta bar) and subsequently there is only conservation within the TrpB1 (boxed in cyan) and TrpB2 (boxed in orange) families. Residues with important interactions in TrpB1 are given in a legend below the figure. Remaining elements are as in Figure 2.19.

Figure 6.2b Active site with the PLP/substrate/product binding segments.
Due to the active site this region shows high conservation in this region.

Legend:
- TrpA interface patch
- COMM (start/end)
- PLP-binding
- E-SER/E-Trp
- allosteric communication (▶ gating)
- cation-binding

Figure 6.2c Central COMM domain.

```
                 H4         S4         H5a  H5b    S5         H5'              H6
B2_SULSO/1-429  VALASALFRM KAHIFMVRTS YYAKPYRKYM MQMYGAEVHP SPSDITEFGR QLLAKDSNHP GSLGIAISDA VEYAH--KNG 220
B1_PYRFU/1-388  TAMAGALLGM KVDIYMGAED VERQKMNVFR MKLIGANVIP VNSGS.....            RTLKDAINEA LRDWVATFEY 178
B1_SALTY/1-396  SALASALLGL KCRIYMGAKD VERQSPNVFR MRLMGAEVIP VHSGS.....            ATIKDACNEA LRDWSGSYET 182
B1_PYRAB/1-388  TAMAGALLGM KVDIYMGAED VERQKMNVFR MKLIGANVIP VHTGS.....            KTIKDAINEA LRDWVATFEY 178
B1_PYRKO/1-389  TAMAGALLGM KVDVYMGAED VERQKMNVFR MGLIGARVIP VESGS.....            RTIKDAINEA LRDWVATFEY 181
B1_ARCFU/1-397  TAMAAALIGL EAEIYMGAED YERQKMNVFR MELIGAKVTA VESGS.....            RTIKDAINEA LRDWVESFEH 190
B1_AQUAE/1-397  TATASALFGL ECVVYMGEED AERQALNVFR MKLLGAKVEI VKSGS.....            RTIKDAINEA LRDWTNVES  186
B1_METVO/1-409  TSMAGALFGL ETEIFMGRVD TERQQPNVAR MKLLGAKVTP VDTGS.....            KVLKDAVNEA MRNWTATFEN 186
B1_THEMA/1-388  TATAAALFGM ECVIYMGEED TIRQKPNVER MKLLGAKVVP VKSGS.....            RTLKDAINEA LRDWITNLQT 179
B1_THET2/1-404  VATVAALFGL ECVVYMGEED VRRQALNVFR MKLLGAEVRP VAAGS.....            RTLKDAINEA IRDWITNVRT 190
B1_METTH/1-392  TAAAGALFGM DVDIYMGTED VERQKLNVFR MEVSGAEVIP VDSGS.....            RTLKDATNEA MRDWISNVDD 179
B1_BACST/1-398  AATVAAHFGM DCIVFMGEED IKRQELNVFR MKLIGAEVVP VSSGN.....            RTLKDATNEA IRYWVAHCDD 182
B1_METJA/1-404  TAAACAKLGL ECIIYMGAKD VERQKLNVFR MELMGAKVIP VFGGS.....            QTIKDAVNEA LRDWTTNVRT 190
B2_PYRAB/8-441  LSIAGALLGL KVRVYMARAS YQQKPYRKTL MRIYGAEVFP SPSENTEVGK RFLKEDPNHP GSLGIAISEA IEDVLK-DEK 222
B2_PYRFU/1-430  LSIAGALMGI KVRVYMARAS YEQKPYRKVL MRIYGAEVFP SPSENTEIGK RFLSENPNHP GSLGIAISEA IEDVLK-DEK 218
B1_PYRHO/8-441  LSIAGALIGL KVRVYMTRAS YYQKPYRKIL MEIYGAEVFP SPSENTEVGK RFLREDPNHP GSLGIAISEA IEDVLS-DEK 220
B2_AERPE/2-438  LSAAGAYFGV KVRVYMVRVS YLQKPYRRTL MELYGAEVYP SPSDKTEFGR KLLAENPNHP GSLGIAISEA IEDVINSGGN 224
B1_SULSO/1-424  VAIAASMYNM KSTIFMVKVS YEQKPMRRSI MQLYGANVYA SPTNLTEYGR KILETNPQHP GSLGIAMSEA IEYAL--KNE 215
B2_METTH/1-429  LSIACSLMDL QCKVYMVRVS FNQKPFRKTI MQLYGGEVVP SPSNHTEFGR RMLKEDPEHP GSLGIAISEA MEEALQ-EEN 218
B2_THEMA/1-422  LSYAGAKFGL EVKVFMVKVS YQQKPMRKYM MNLFGGKVTP SPSEETNFGR KILSEDPDNP GSLGIAISEA LEVAVS-DPN 219
B2_ARCFU/3-434  LCFATKLFEM ACTVYMVKVS FMQKPYRRVM METWGGEVIP SPSDRTEVGR KILAENPDTP GSLGIAISEA IEDAAK-NEN 221
B2_AQUAE/1-434  LSFATQFFDL QCRVYMVRVS YNQKPYRRIL METWKGEVIP SPSPYTNAGR KYYEENPEHP GSLGIAISEA IEEAAS-RED 218
```

Legend:
- ▬ TrpA interface patch
- ◤ COMM (start/end)
- ◀ PLP-binding
- ◀ E-SER/E-Trp
- ◀ allosteric communication
- ◀ cation-binding
- ◀ allosteric communication (◀ gating)

Helix H6 is the central component of allosteric communication in TrpB1 family members. However, none of these residues is conserved between the families. The insertion helix H5' and the truncation of the C-terminus of helix H6 are only present in TrpB2 family members.

Structural elements (left to right): **S6**, **H7a**, **H7b**, **S7**, **H8**, **S8**

Sequence alignment:

```
B2_SULSO/1-429  GKYVVGSV- ----VNSDIM FKTIAGMEAK KQMELIGE-D PDYIIGVVGG GSNYAALAYP FLGDELRS. ·GKVRRKYIA  290
B1_PYRFU/1-388  THYLIGSVVG PHPYPTIVRD FQSVIGREAK AQILEAEGQL PDVIVACVGG GSNAMGIFYP FVND----  ·KKVKLVG   249
B1_SALTY/1-396  AHYMLGTAAG PHPYPTIVRE FORMIGEETK AQILDKEGRL PDAVIACVGG GSNAIGMFAD FIND----  ·TSVGLIG   253
B1_PYRAB/1-388  SHYLIGSVVG PHPYPIIVRD FQSVIGREAR EQILEAEGDL PDVIVACVGG GSNAMGIFYP FVKD----  ·KSVRLIG   249
B1_PYRKO/1-389  SHYLIGSVVG PYPYPVIVRD FQSVIGREAR EQILEAEGTL PDVVACVGG  GSNAMGIFYP FVND----  --RVRLIG   251
B1_ARCFU/1-397  THYLIGSVVG PHPFPTIVRD FQAVIGKEAR RQIIEAEBGM PDAIIACVGG GSNAMGIFHP FLNDD---  --VRLIG    260
B1_AQUAE/1-397  THYVIGSVVG PHPFPMIVRD FQSVIGRETK EQILQKEGRL PDAIVACVGG GSNAMGIFYP FVED----  ·KGVQLIG   257
B1_METVO/1-409  THYLLGTVMG PHPFPTMVRD FQSVIGKEVK KQIMEQEERL PDYIVACIGG GSNAMGLFHP FLSNNISTGN DDAKNVKMIG 266
B1_THEMA/1-388  TYYVIGSVVG PHPYPIIVRN FQKVIGEETK KQILEKEGRL PDYIVACVGG GSNAAGIFYP FI-D-...  ·SGVKLIG   249
B1_THEMA/1-409  TFYILGSVVG PHPYPMVVRD FQSVIGEEVK RQSLELFGRL PDALIAAVGG GSNAIGLFAP FAYLPE...  ·GRPKLIG   263
B1_THET2/1-404  THYLIGSTMG PHPYPTMVKH FQSVIGREAR EQILEVEGEL PDTVIACVGG GSNAIGIFSA FMDDD...  ·VELIG     249
B1_METTH/1-392  HFYMIGSVVG PHPYPKMVRE FQRIIGDEAK EQFLACEGKI PDVIVACVGG GSNAIGMFYP FLQDD...  --VRLVG    260
B1_BACST/1-398  THYLIGSVVG PHPYPYMVRE FQRVIGKELK EQILEKEGRL PDVIVACVGG GSNAIGAFYE FLDDD...  ·--VELYA   260
B1_METJA/1-404  TYYLLGSALG PHPYPMMVRE FORVIGKELK EQILEKEGRL PDVIVACVGG GSNAIGAFYE FLDDD...  --VELYA    260
B1_PYRAB/8-441  ARYSLGSV-  --LNHVLM   HQTVIGLEAK EQMEEFEE-- PDVIVACVGG GSNAIGAFYP FVKDVLD...  ·GKSEYEFIA 290
B2_PYRAB/8-441  ARYSLGSV-  --LNHVLM   HQTVIGLEAK EQMEEFE--- ·PDVIGCVGG GSNFAGIAYP FVKDVLD...  GDNEYEFIA  286
B2_PYRFU/1-430  ARYSLGSV-  --LNHVLM   HQTVIGLEAK QQMEEFEE-- PDVIGCVGG  GSNFAGIAYP FVKEVLD...  GKAEYEFIA  288
B1_PYRHO/8-441  ARYSLGSV-  --LNHVLM   HQTVIGLEAK KQVKEFEE-- PDVIGCVGG  GSNFAGLSYP FIKDVLD...  GKAEYEFIA  288
B2_AERPE/2-438  AKYSLGSV-  --LNHVLL   HQTVIGLEAE KQFREAGV-Y PDIMIGAVGG GSNFAGFTYP FIRHRLK...  GSSSTRFIA  293
B2_SULSO/1-424  FRYLVGSV-  ---LDVVLL  HQSVIGQETI TQLDLLGE-D ADILIGCVGG GSNFGGFTYP FIGNK....  KGKRYIA    280
B1_SULSO/1-424  VYYSLGSV-  ----LDVVLL HQTVIGLETK KQLEIAGE-T PDIMIGCVGG GSNFGGAIFP FVKDKLD...  GKLDCEFIA  287
B2_METTH/1-429  TKYSLGSV-  ----LNHVLL HQTVIGLETK KQLELIGE-K PDILIGCHGG GSNFGGTIIP FVPDKLS...  GRDIRFVA   287
B2_THEMA/1-422  TKYSLGSV-  ----LNHVLL HQTVIGLETK AQLEKVDE-K PDVLIGCVGG GSNFAGLTYP FVNDAK....  NGDLEIIA   292
B2_ARCFU/3-434  TKYSLGSV-  ----LNHVLL HQTVIGLETK PDILIGCHGG GSNFAGLSFP FLADVLRGD. KRKEDLKVIA 287
B2_AQUAE/1-434  TKYSLGSV-  ----LNHVLL HQTVIGLEAK KQMEEAGY-Y PDVIGAVGG  GSNFAGLSFP FLADVLRGD. KRKEDLKVIA 287
```

Legend:
- ▬ TrpA interface patch
- ▶ (green) COMM (start/end)
- ▶ (blue) PLP-binding
- ▶ (cyan) E-SER/E-Trp
- ▶ (magenta) allosteric communication
- ▶ (orange) cation-binding
- ▶ (red) gating

Figure 6.2d End of the COMM subdomain and start of the C-terminal domain with the P-loop.
Downstream of the end of the COMM subdomain there is a truncation in TrpB2 family members. SsH1 packs into the resulting cavity. The P-loop between S7 and H8 is strictly conserved between the families. Residues of the P-loop are also involved in cation binding in TrpB1 proteins.
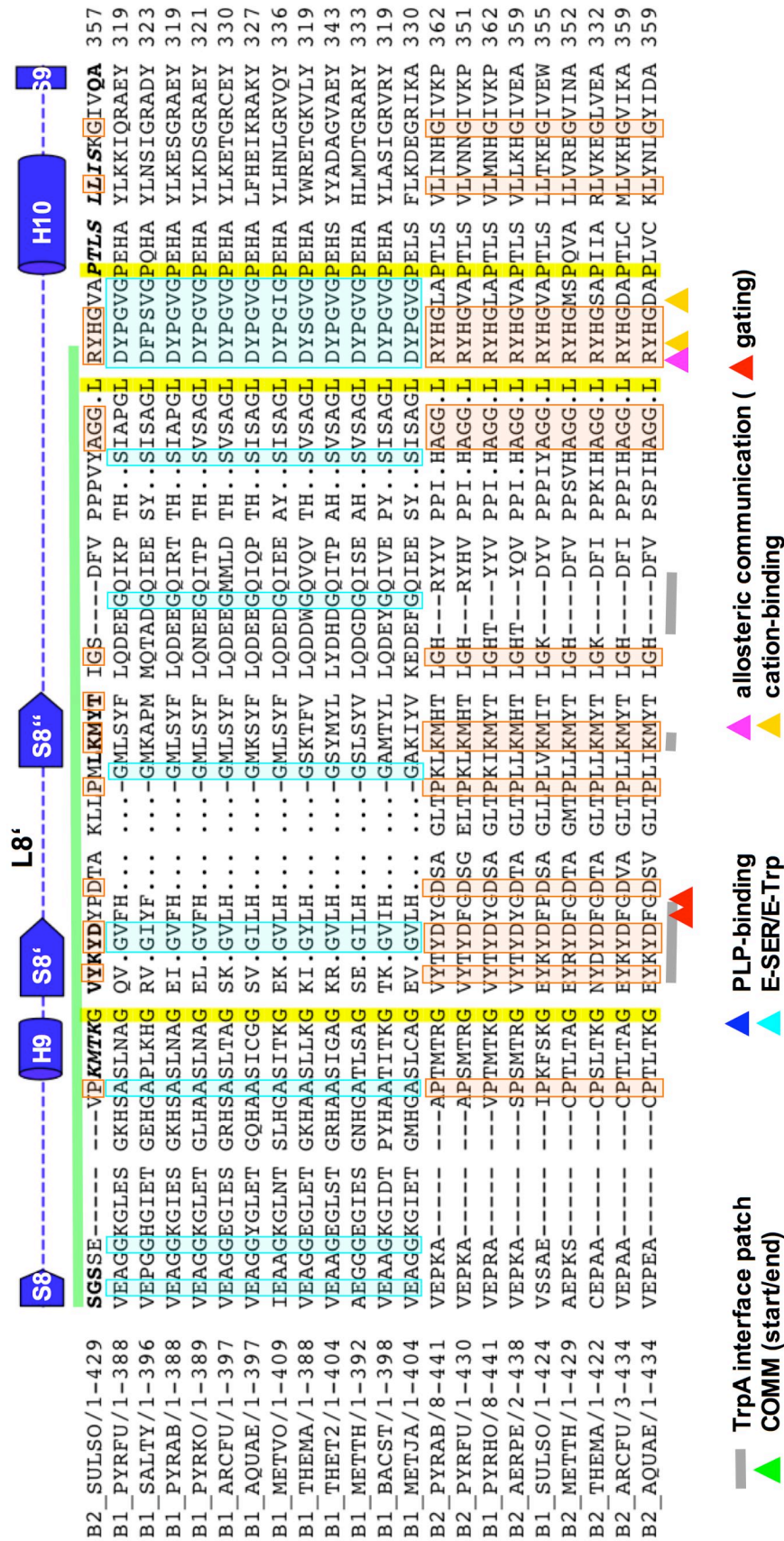
Figure 6.2e L8', the gating residues and the C-terminal portion of the cation binding-site
This region reveals a structural shift between TrpB1 and TrpB2 family members that is in context with the cation binding region carries an insertion in TrpB2 proteins. However, the segment is highly conserved in TrpB2 family members.

Figure 6.2f The C-terminal region.

Conserved segments are involved in interactions to the COMM domain, PLP-binding and structural organization of the C-terminal domain.

Legend:
- ▬ TrpA interface patch
- ▲ COMM (start/end)
- ▶ PLP-binding
- ▷ E-SER/E-Trp
- ▶ allosteric communication (▲ gating)
- ▶ cation-binding

Structural elements: S9 · H11 · H12 · S10 · H13

Sequence	Alignment
B2_SULSO/1-429	RDYSQEESFK WAKLFSELEG YIPAPETSHA LPIL-AEIAE EAKKSGERKT VLVSFSGHGL LDLGNYASVL FKE---- 429
B1_PYRFU/1-388	VTVTDEEALK AFHELSRTEG IIPALESAHA VAYAMKLAKE M..-SRDEI IIVNLSGRGD LDLDIVLKVS GNV---- 388
B1_SALTY/1-396	VSITDDEALE AFKTLCRHEG IIPALESSHA LAHALKMMRE Q..-PEKEQL LVVNLSGRGD KIFTVHDIL KARGEI 396
B1_PYRAB/1-388	VTVTDEEALR AFHELSRTEG IIPALESAHA VAYAIKLARE M..-SRDDV IIVNLSGRGD KDLDIVLKVS GNV---- 388
B1_PYRKO/1-389	VTVTDEEALR AFHELSRTEG ILPALESAHA VAYAMKIAPE M..-DKDEI IIVNLSGRGD KDLDIVRRVG NV----- 389
B1_ARCFU/1-397	VTVNDEEALR AFKTLSKIEG ILPALESAHA IAYAMKMAEE M..-QRDDV LVVNLSGRGD KDMDIVRRRI A------ 397
B1_AQUAE/1-397	TTATDEEALE GFKLLARTEG IIPALESAHA VIKAVEVARE L..-GKDGI VVINLSGRGD KDMAHVMKHL SLEG-- 397
B1_METVO/1-409	ASATDKQALK AFMELTRTEG IIPALESSHA IAYAIENAGN M..-DKDDI MVINLSGRGD HKLGC-- YVRERI 409
B1_THEMA/1-388	DAVTDEEALD AFIELSRLEG IIPALESSHA LAYL---KK I..-NIKGKV VVVNLSGRGD KDLESVLNHP YVRERI 388
B1_THET2/1-404	ASVTDEEALE GFKLLARLEG IIPALESAHA IAYAAKVVPE M..-DKDQV VVINLSGRGD KDVTEVMRLL GGEL-- 404
B1_METTH/1-392	EPVTDTHALR GFRLLSRYEG MPALESAHA IACLERYAEK P..-ENRGKT VIVNLSGRGD KDMFMVAGLL GVGV-- 392
B1_BACST/1-398	ESVTDEEAVA AFRLLAQTEG IIPAIESAHA VAKAVELAQS M..-SPDET VLICLSGRGD KDVQTMMRHL GAKEGE 398
B1_METJA/1-404	VCVTDDEALE AFQLLCRLEG ILPALESSHA LAYAVKLADK L..-DKDDI MVINLSGRGD GREI-- 404
B2_PYRAB/8-441	IAYHQTEVFE AAVLFAKAEG IVPAPESAHA VKAVIDKALE AR.REGKEMV ILFNLSGHGL LDLKGYDYL DGKLED 441
B2_PYRFU/1-430	IAYHQTEVFE AAALFAKLEG IVPAPESAHA IKATIDKAIE AK.REGKEIV ILFNLSGHGL LDLHGYEEYL EGRLQD 430
B1_PYRHO/8-441	MAYHQTEVFE AAVLFAKAEG IVPAPESAHA VRAVIDKALE AK.RKGEETV ILFNLSGHGL LDLKGYEDYL EGKLED 441
B2_AERPE/2-438	RAYHQREVFR AAHMFAKAEG VKAAIDEAIK AR.DEGRRVV IAFNLSGHGL LDLSNYESMM KRLNGN 438
B1_SULSO/1-424	REYNEREIFE AAKIFIENQG IVPAPESAHA IRAVVDEAIE AR.KNNERKV IVFNLSGHGL LDLSNYESMM KRLNGN 424
B2_METTH/1-429	RAVPQHTIFE SGVKFAKAEG VVPAPETCHA ISVAIDEARK CR.ETGEEKT IVISFSGHGL LDLKGYGDYL EGKI-- 429
B2_THEMA/1-422	QAFDQDETFE AAKIFAKLEG IIPAPESAHA IAGAIREAKK AK.EEGKERV IVFTLSGHGL LDLTAYV--- ------ 422
B2_ARCFU/3-434	RAVKQLPTFE AGLIFARTEG IIPAPETNHA VRAAIDEAIK AR.ENNEEKV IVFGFSGHGL LDLQAYDDYL AGRLAD 434
B2_AQUAE/1-434	VAYKQTEVFE AAVTFARTEG IVPAPESAHA IKAAIDEALK CK.ETGEEKV ILFNLSGHGY FDLSAYDKYL HGELTD 434

H0   S1   H1   S2

```
PYRFU/1-247   ---------- --MFKDGSLI PYLTAGDPDK QSTLNFLLAL DE-YAGAIEL
SALTY/1-267   MERYENLFAQ LNDRREGAFV PFVTLGDPGI EQSLKIIDTL IDAGADALEL
PYRAB/1-247   ---------- --MFRDGSLI PYLTAGDPSA KATLRFLLAI EE-YSGAIEL
PYRKO/1-249   ---------- --MFEKGSLI PYLTAGDPSV EKTLEFLLAV EE-FAGLIEL
PASMU/1-267   MRRFDTLFSQ LQQKQQGAFV PFVTLCDPNL AQSFEIIQTL VENGADALEL
METTH/20-266  ---------- -------AFV PFVVAGDPDM ETSLEIIRTL VDAGADALEV
METVO/26-267  ---------- --------LV SFLVSGDPNI EATLKFMNAL DE-YCGVIEL
METJA/9-272   -------FEE LKNKGEKAFV AFYVGGDPNL EISEKALEVI CK-HADIVEI
ARCFU/3-245   ---------- -----DRSLI VFFTACYPTA EKTVEFMLTA AESGADVIEL
THETN/1-258   MNRIDRRFDD LKKEGKKALI TFITAGDPDI DTTYDIVLAL EEAGADIIEL
BACST/11-256  ---------- PFIVAGDPSP EVTVDLALAL EEAGADLLEL
AQUAE/1-262   MGRISDKFTE LKEKREKALV SYLMVGYPDY ETSLKAFKEV LKNGTDILEL
THET2/7-258   -------FAK ARSEGRAALI PYLTAGFPSR EGFLQAVEEV LP-YADLLEI
THEMA/4-238   ---------- --------FI AYIPAGFPDL ETTRKILIAL NELGITGVEI
SULSO/7-237   ---------- --------LV VYMTLGYPNV QSFKDFIIGA VENGADILEL
```
△

L2

```
PYRFU/1-247   GIPFSDPIAD GKTIQESHYR ALKNGFKLRE AFWIVKEFRR HSS-TPIVLM
SALTY/1-267   GVPFSDPLAD GPTIQNANLR AFAAGVTPAQ CFEMLALIRE KHPTIPIGLL
PYRAB/1-247   GIPFSDPIAD GKTIQQSHFR ALKGGFKLEH AFNIVREFRK HSD-VPIVLM
PYRKO/1-249   GIPFSDPMAD GKTIQESHYR ALRNGFKLDD TFRILREFRR HSS-TPVILM
PASMU/1-267   GFPFSDPMLD GPVIQAANQR ALVAGCSTDA SFQLLTQVRS TYPTLPISLL
METTH/20-266  GFPFSDPIAD GTSVQGADLR ALRAGMTTEK CFQLIERVRE FTS-IPIGLL
METVO/26-267  GIPFSDPIAD GSTIQEANVR SLSNGYKIHQ SFDVLREFRK FSD-TPVVLM
METJA/9-272   GIPFSDPVAD GITIQKADVR ALNSGMNPLK AFELAKKLNE KAPNVPKVFL
ARCFU/3-245   GVPFSDPVAD GKTIQESYVR ALRN-FRVER VFEIAKAFRA ESD-KPLVLM
THETN/1-258   GIPYSDPLAD GPTIQASSQR ALAKGIKIKD IMDLVMRIRE KSD-IPIVYL
BACST/11-256  GVPYSDPLAD GPTIQRAAAR ALAGNMTLPK AIHLVAEMRK KGVTIPIILF
AQUAE/1-262   GFPFSDPVAD GPTIQVAHEV ALKNGIRFED VLELSETLRK EFPDIPFLLM
THET2/7-258   GLPYSDPLGD GPVIQRASEL ALRKGMSVQG ALELVREVRA LTE-KPLFLM
THEMA/4-238   GVPFSDPVAD GPVIQLAHSV ALRNGVTIKK ILEMLSEISV DYD---LYLM
SULSO/7-237   GIPPKYAKYD GPVIR..... .KSYDKVKGL DIWPLIEDIR KDVGVPIIAL
```
▲▲ △   ▬   ▬   ▬

H3   S4   H4

```
PYRFU/1-247   TYYNPI..YR AGVRNFLAEA KASGVDGILV VDLPVFHAKE FTEI...ARE
SALTY/1-267   MYANLV..FN NGIDAFYARC EQVGVDSVLV ADVPVEESAP FRQA...ALR
PYRAB/1-247   TYYNPV..FR VGLREFIGKA KDSGVDGMLI VDLPVMHASE FLEV...ARE
PYRKO/1-249   TYYNPV..FR TGVKKFLGEA KASGADGILV VDLPVSHAGE FLDA...AKE
PASMU/1-267   LCANLV..YA QGIERFYQRC AEVGVDAVLI ADLPLFAHRE FTVM...AEK
METTH/20-266  VYYNLI..YR MGVDEFYRRA AEAGVTGILA ADLPPEEASD ALRA...AEK
METVO/26-267  TYYNPI..YK RGIENFVIQA KEAGANGLII VDLPLDEAEQ YRAI...CKK
METJA/9-272   TYYNII..FK MGEEEFVKKC KEAGVSGIIV PDLPIEEADS LYNY...CKK
ARCFU/3-245   SYYNPI..YR RGVESFVEKA YSSGIDAMLV VDLPYDEAGD FVEV...CSR
THETN/1-258   VYYNSV..FK YGIEKFLEDS KNAGVDGLII PDLPLEERKE VLDV...ADK
BACST/11-256  TYYNPV..LQ LGEESFFALA RENGANGVLI PDLPFEESGP LREL...GER
AQUAE/1-262   TYYNPI..FR IGLEKFCRLS REKGIDGFIV PDLPPEEAEE LKAV...MKK
THET2/7-258   TYLNPV..LA WGPERFFGLF KQAGATGVIL PDLPPDEDPG LVRL...AQE
THEMA/4-238   SYLNPIVNYP EGKEKLLDEL KKLGVKGLII PDLPLREVKN VDIA...---
SULSO/7-237   TYLED-..WV DQLENFLNMI KDVKLDGILF PDLLIDYIDD LDKIDGIIKN
```

Figure 6.3a Structure based sequence alignment of TrpAs and the interactions to TrpB1.
The secondary structure elements are shown for *Pf*TrpA as assigned by Yamagata *et al.*, 2001. *Ss*TrpA shows a different L2 (catalytic loop) composition. Strictly conserved residues are highlighted. The first helix (H0), which is not present in *Pf*TrpA is shown transparent. The legend for allosteric/catalytic residues as well as residues interacting to TrpB1 is given in the second part.

```
        --▷  S5 ▷-------  ◖   H5  ◗----◖ S6 ▷--    L6  --------      ◖  ◗

PYRFU/1-247   EGIKTVFLAA PNTPDERLKV IDDMTTGFVY LVSLYGTTGA REEIPKTAYD
SALTY/1-267   HNIAPIFICP PNADDDLLRQ VASYGRGYTY LLSRSGVTGA... .ENRGAL
PYRAB/1-247   EGIKTVFLAA PNTPDERLKE IDKASTGFVY LISLYGTTGA RDKIPETAFN
PYRKO/1-249   EGLKTVFLAA PNTPDERLRE IDKASTGFVY LISLYGTTGA RDRLPETAFE
PASMU/1-267   YSIQQVFICP PNACEDTIKR VAQQTQGYTY LVSRSGVTGV ...ENQQSAD
METTH/20-266  YDIDQIFIVA PTTGSERLKR ISEVSSGFHY LVSVMGVTGA R....SRVED
METVO/26-267  HDMGTVFLVA PNTPDERLMY SDEASTLFLY VISTFGITGA R.........
METJA/9-272   YGVDLIFLVA PTTPDERLKK ILEKCSGFVY VVSVTGITGA R....EKVAE
ARCFU/3-245   TGMKNVFLAA PNTPEDRLRA MDELS-AFVY LVSTYGVTGE RDRISPLAFE
THETN/1-258   YEIYLIPLVA P-TSKERIKG ITENGKGFVY CVTLTGVTGA R.......EE
BACST/11-256  FDLPLISLVA P-TSKQRIER IASVAQGFLY CVSSLGVTGM R...ETLPES
AQUAE/1-262   YVLSFVPLGA PTSTRKRIKL ICEAADEMTY FVSVTGTTGA R......EKL
THET2/7-258   IGLETVFLLA PTSTDARIAT VVRHATGFVY AVSVTGVTGM R....ERLPE
THEMA/4-238   --YPIVPFVA PNTKDEEIDL INSVQAPFVY YISRYGVTGE R.....EDLP
SULSO/7-237   KGLKNVIFTS PSVPDLLIHK VSKISDLFLY YGVRP.TTGV ......PIPV
                            ▬▬                         ▬ △

        ◖  H6  ◗----------------◖ S7 ▷-----◖  H7  ◗---◖ S8 ▷

PYRFU/1-247   LLRRAKRIC- ...--RN... ....KVAVGF GVSKREHVVS LLKEGANGVV
SALTY/1-267   PLHHLIEKLK ...EYHAA.. ....PALQGF GISSPEQVSA AVRAGAAGAI
PYRAB/1-247   LLKRAKRIC- ...--KN... ....KVAVGF GVSKREHVEM LLNAGANGVV
PYRKO/1-249   FVRRARKIC- ...--NN... ....KLAVGF GVSRREQVEE LLKAGADGVV
PASMU/1-267   NLKTLVERLK ...AYQAP.. ....PILQGF GISQPQQVSD VLKIGVNGAI
METTH/20-266  ATIELIKRVK ...AEGSL.. ....PVMVGF GVSRPEHVRM LRDAGADGVI
METVO/26-267  ...GSFEKMT FEFIARAKNL CDKNKLYVGF GISNGEHAEK IIENGADGVI
METJA/9-272   ETKELIKRVK ...KFSKI.. ....PACVGF GISKREHVEE -ITEIADGAI
ARCFU/3-245   ALKRAKGIC- ...--RK... ....PVAVGF GVSKADHVRQ LISAGADGVV
THETN/1-258   ITTDLEEYMR LVSSYTSM.. ....PKALGF GISGPE-MAR KLKSLSDGIV
BACST/11-256  LGDFVSEVKR ...HSRV... ....PVAVGF GISTPEQV-A MLKEVCDGVV
AQUAE/1-262   PYERIKKVE ...EYRELC. ..DKPVVVGF GVSKKEHARE -IGSFADGVV
THET2/7-258   EVKDLVRRIK ...ARTAL.. ....PVAVGF GVSGK--ATA AQAAVADGVV
THEMA/4-238   FADHIKRVKE ...RIKL... ....PLFVGF GISRHEQVKK -VWEIADGVI
SULSO/7-237   SVKQLINRVR N..LVEN... ....KLIVGF GLSSESDLRD ALSAGADGIA

        ▷--◖   H8'  ◗--◖  H8 ◗---------

PYRFU/1-247   VGSALVKIIG EKGR-..... ...EATEFLK KKVEELLG-- ---
SALTY/1-267   SGSAIVKIIE KNLASP.... ..KQMLAELR SFVSAMKAAS R--
PYRAB/1-247   VGSALINIIA EHGE-..... ...NAEEKLR EKVRELAG-- ---
PYRKO/1-249   VGSALIELIS RSE--..... ...NPVEELR RKVAELSGYS R--
PASMU/1-267   SGSATVQIIE KYHHDP.... ..QACLAELA QFTQKMKQAT ---
METTH/20-266  VGSAIIDVIS GNLGDR.... ..ELMLQRIH EMAGTLKAA- ---
METVO/26-267  VGSAFVDIIK EYGDSN.... ...ETIYKLK ELAREL---- ---
METJA/9-272   VGSAIVKIVE KHLDENGQIK DEEKFLKELE EFVKNLKEGT K--
ARCFU/3-245   VGSAFVRLIN EKGE-..... ...RATEDIR ALTENLRS-- ---
THETN/1-258   VGSAIVERIA KGYNRC.... ...EMLKEVK EFVLSL---- ---
BACST/11-256  IGSALVQKVE QLGERL.... ...LAPEEKE AAIAEFAAYA R--
AQUAE/1-262   VGSALVKLAG QKKIEDL... ..GNLVKELK EGLRE----- ---
THET2/7-258   VGSALVRALE EGRSLA.... ...PLLQEIR QGLQRLEA-- ---
THEMA/4-238   VGSALVRIME ENPKD..... ...EIPRKVV EKVKELLG-- ---
SULSO/7-237   IGTVFIEEIE RNGVK..... ...SAINLVK KF-------- ---
```

△ **catalytic base**　　△ **allosteric residue**　　▬ **interaction to TrpB1**

Figure 6.3b Structure based sequence alignment (part two).

6.2.3 A77-A78

FNIII modules of the short super-repeat (A1-A42)

Figure 6.4a Sequence alignment of FNIII modules from the A-band of human titin.

Note: strictly conserved residues are indicated by yellow background. Titin's FNIII modules exhibit a position dependent conservation pattern (compare Figure 3.16 and Figure 3.24; indicated by orange/red boxed residues in loop segments). The CC'-loop forming the center of a highly conserved residue cluster on the front-side of the modules is boxed in blue.

FNIII modules of the long super-repeat (A43-A163)

| | A | B | C | C' | E | F | G1 | 2 | 3 |

```
A78                    A        B              C              C'             E        F              G1  2  3

A44  DTPGPV-SD--LKVSDVTKTSCHVSWAPFENDGGSQVTHIVIEKR------EADR---KTWSTVTPEVKRTSFHVT----NLVPG--NEYYFPRVTAVNEYG-PGVPTDVP---KPVLLASDPL----  100
A45  SEDPDR-K--LEATEMTKNSAILAWLFPIRDGGAKIDGYIISYR------EEEQPADRWTEYSVVKDLS-------LVVTGLKEG--KKYKFRVAARNAVG-VSLPRRAEG---VYEAKE-------  99
A47  DAPGPFQPP--FDISDIDADACSLSWHIPLEDGGSNIINNYIVEKC------DVSR---GDWVTALASVIKTSCRVG------KLLPG--QEVIFPRVRAENRFG-ISEPLTS----PKMVAQPPF----  100
A48  GVPSEFKNA--RVTKVNKDCIFVAWDRPDSDGGSPIIGYLIERK------ERNSIL---WVKAN-DT-LIVRSTEY-----PCAGLVEG--LEYSFRIYAIAINKAG-SSPPSKPTE---YVTAR-------  98
A49  MPVDPFGKP--EVIDVTKSTVSLIWARPKHDGGSKIIGYFVEAC------KLLPGDK---WVRCNTAPHQITPQEEY----TATGLEEK--AQYQFRALAFTAVN-ISPPSEPPSD---PVTILA------  101
A51  DKPGPFAS--VKINKYSDRAMLSWEFPLEDGGSEINYIVDKR------ETSR---PNWAQVSATVPITHSCSVE----KLIEG--HEYQFRICAENKYG-VGDPVFT----EPAIAKNPY----  99
A52  DPPGRCDPP--VISNITKDHMTVSWKFPADDGGSPIIGYLLEKR------ETQAVN---WTKVN-RK-PIIERTL----KATGLQEG--TEYEFRVTAINKAG-PGRPSDASK---AAYARDPQ----  101
A53  YPPAPFAPP--KVYDTTRSSVSLSWGKPAYDGGSPIIGYLVEVK------RADSDN---WVRCNLPQ--NLQKT----RFEVTGLMED--TQYQPRVTAVNKIG-YSDPSDVP---DKHYPKD----  100
A55  DTPGPF-IN--VTVKEISKDSAYVTWEPPIIDGGSPIINNVVQKR------DAER---KSWSTVTTECSKTSFRVP----NLEEG--KSYYFFRVFAENEYG-ICDPGETR----DAVRAS------  97
A56  QTPGPVVDL--KVRSVSKSSCIGWKKPHSDGGSRIIGYVVDFL------TE--ENK---WQRVMKSLS--LQYSA----KDLTEG--MEYTFRVAAENKYG-VGEGLKS----EPIVARHPF----  100
A58  GKPGIFTGP--IKFDEVTAEAMTLNWAPENDGGSEITNYILEKR------DSVN---NKWVTCASAVQKTFRVT----RLHEG--MEYTFRVAAENKYG-VGEGLKS----EPIVARHPF----  100
A59  DVPDAFPPP--NIVDVRHDSVSLTWTDPKRTGGSPITGYHLEFK------ERNSIL---WVKRAN-KT-PIRMRDF----KVTGLTEG--LEYEFRVMAINIAG-VGKPSLPSE---PVVAL-------  97
A60  DPIDPFGKP--EVINTTRNSVTLIWTEPKYDGGHKLTGYIVEKR------DLPSKS---WMKANHV--NVPECAF----TVTDLVEG--GRYEFRIRARNTAG-AISAPSESTE---TICKDE-----  101
A62  DVPGPF--GP--VEISNVSAEKATLWTEPLEDGGSPIKSYILEKR------ETSR---LLWTVVSEDIQSCRIVAC----KLIQG--NEYIFRVAAVNHYG-KGEPVQS----EPVKMVDRF----  99
A63  GPPGPFEKP--EVSNVTKNTATVSWKEPVDDGGSEITGYHVERR------EKKSLR---WVRAI-KT-FVSDLRC----KVTGLQEG--STYEFRVSAENRAG-IGPPSEASD---SVLMKDAA----  99
A64  YPPGPFSNP--HVTDGTTKSAGWKAPHYDGGLEITGYVVEHQ------KVGDEA---WIKDITGT--ALRIT----QFVVPDLQTK--EKYNFRISAIINDAG-VGEPAVIP---DVEIVE------  99
A66  DTPGPV-LN--ILRPTDITTKDSVTLIHWDIPLIDGGSRINTNYIVERK------EATR---KSYSTATCKCHCTYKVT----GLLSEG--CEYFFRVMAENEYG-IGEPTETT----EPVKAS------  97
A67  EAPSPEDSL--NIMDITKSTVSLAWPKFKHDGGSKITGYVIEAQ------RKGSDQ---WTHTTVK--G--LE----CVVRNITEG--EEYTFQVMAVNSAG-RSAPRRESRP---VIVKEQ------  97
A69  DIPGPFTGP--IKFDEVSDPVTFSWDFPENDGGVPISNYVVEMR------QTDS---TTWVELATTVIRRTYKAT----RLITG--LEYQPRVRAQNRYG-VGPGITS----AWIVANYPF----  100
A70  KVPGPFGTP--QVTAVTKDSMTISWHEPLSDGGSPILGYHVERK------ERNGIL---WQTVS-KA-LVPGNIF----KSSGLTDG--IAYEFRVTAENMAG-KSKPRKPSE---PMLAL-------  98
A71  DPIDPGKP--VPLNITRRTVTVLKWAKPEYTGGEKITSYIVEKR------DLPNGR---WLKANFS--NIIENEF----TVSGLTED--AAYEFRVIAKNAAG-AISPPSEPSD---AITCRDD----  101
A73  DRPGEPEGP--LAVTEVTESEKCVLSWFFPLDDGGAKIDHYIVQKR------ETSR---LAWTNVASEVQVTKLKVT----KLLKG--NEYIFRVMAVNKYG-VGEPLES----EPVLAVNPY----  100
A74  GPPDPFKNP--EVTIITKDSMVVCWGHPDSDGGSEIINYIVERR------DKAGQR---WIKCN-KK-TLTDLRY----KVSGLTEG--HEYEFRVMAENAAG-ISAPSPTSP---FYRACDTV----  101
A75  FKPGPEGNP--RVLDTSRSSISIAWNKPIYDGGSEITGYMVEIA------LPEEDE---WQIVTPPA--GLKAT----SYTITGLTEN--QEYKIRIYAMNSEG-LGEPALVP---GTPKAEDR----  101
A77  DTPSPF--QD--LKVKEVTKTSVTLTWDFPLLDGGSKIKNYIVEKR------ESTR---KAYSTVATNCHKTSWKVD----QLLOEG--CSYYFRVIAENEYG-IGLPAETA----ESVKAS------  97
A78  ERPLPEGKI--TLMDVTRNSVGLSWEKPEHDGGSRILGYIVEMQ------TKGSDK---WAATCATVK--V--TE----AITTGLIQG--EEVSFRVSAQNEKG-ISDPRQLSV---PVIAKD------  100
A80  DKPGPFTGP--VKMDEVTADSITLSWGFPYDGGSSINNIVEKR------DTIST---TTWOIVSATVARTTIHAC----RLKTG--CEYQFRIAAENRYG-KSTYLNS----EPTVAQPPF----  100
A81  KVPGPGTP--VVTLSSRDSMEVQWNNEPISDGGRVIGIVHLERK------ERNSIL---WVKLN-KT-PIPQTKF----KTTGLEEG--VEYEFRVSAENIVG-IGKPSKVSE---CYVAR-------  98
A82  DPCDPEGRP--EAIIVTRNSVTLQWKKPTYDGGSKITGYIVEKK------ELPEGR---WMKASFT--NIIDTHF----EVTGLVED--HRYEFRVIATNAAG-SFSPSESSTG---AITARD-----  100
A84  DRPGEEGCP--VVISGVTAEKCTLAWKFPLQDGGSDIINYIVERR------ETSR---LVWTVVDANVQTLSCKVT----KLLEG--NEYTFRIMAVNKYG-VGEPLES----EPVAKNPF----  100
A85  VVPDAFKAP--EVTTVTKDSMIVWWERPASDGGSEILGYVLEKR------DKEGIR---WTRCH-KR-LIGELRI----RVTGLIEN--HDYEFRVSAENAAG-LSEPSPPSA---CDPI-------  101
A86  YKPGPFNNP--KVIDITRSSVFLSWSKPIYDGGCEIQGYIVEKC------DVNVGE---WTMCTPPT--GINKT----NIEVEKLLEK--HEYNFRICAINKAG-VGEHADVP---GPIIVEKR----  101
A88  DTPSPF-VN--LKVTEITKDSVSIWWEPPLLDGGSKIKNYIVEKR------EATR---KSYAAVVINCHKNSWKID----QLOEG--CSYYFRVTAENEYG-IGLPAQTA----DPIKVA------  97
A89  EVPQPEGKI--TVDDVTRNSVSLSWTKPEHDGGSKIIQYIVEMQ------AKHSEK---WSECARVK--S--LQ----AVITNITOG--EEYLFRVVAVNEKG-RSDPRSLAV---DPIKYA------  100
A91  DKPDPEKGP--VKFDDVSAESITLSWNFPLYTGGCQITNYIVQKR------DTTT---TTWDDVVSATVARTTLKVT----KLKTG--TEYQFRIFAENRYG-QSFALES----DPIVAQPPY----  100
A92  KEPGPGTP--FATAISKDSMVIQWHEPVNNGGSPVIGYHLERK------ERNSIL---WTKVN-KT-IIHDTQF----KAQNLEEG--IEYEFRVVAENIVG-VGKASKNSE---CYVAR-------  98
A93  DPCDPEGTP--EPIWVKRNEITLQWTKPVYDGGSMITGYIVEKR------DLPDGR---WMKASFF--NVIITQF----TVSGLTED--QRYEFRVIAANAAG-NISRPSDSTG---PITAKDE----  100
A95  DRPGEEGCP--VQVTGVTSEKCSLTWSPPLQDGGSDISHVVEKR------ETSR---LAWTVVASEVVTNSKLKVT----KLIEG--NEYVFRIMAVNKYG-VGEPLES----APVLMKNPF----  101
A96  VLPGPFKSL--EVTNIAKDSMVVCWNRPDSDGGSEIIGYIVEKR------DRSGIR---WIKCN-KR-RITDLRI----RVTGLTEH--HEYEFRVSAENAAG-VGEPSPATV---YYKACDFV----  100
A97  FKPGPFNA--HIVDITKNSITLAWGKPIYDGGSBILGYVVEIC------KADEEE---WQIVTPQT--GLRVT----RFEISKLTEH--QEYKIRVCALNKVG-LGEATSVP----GTVKPEDK----  101
```
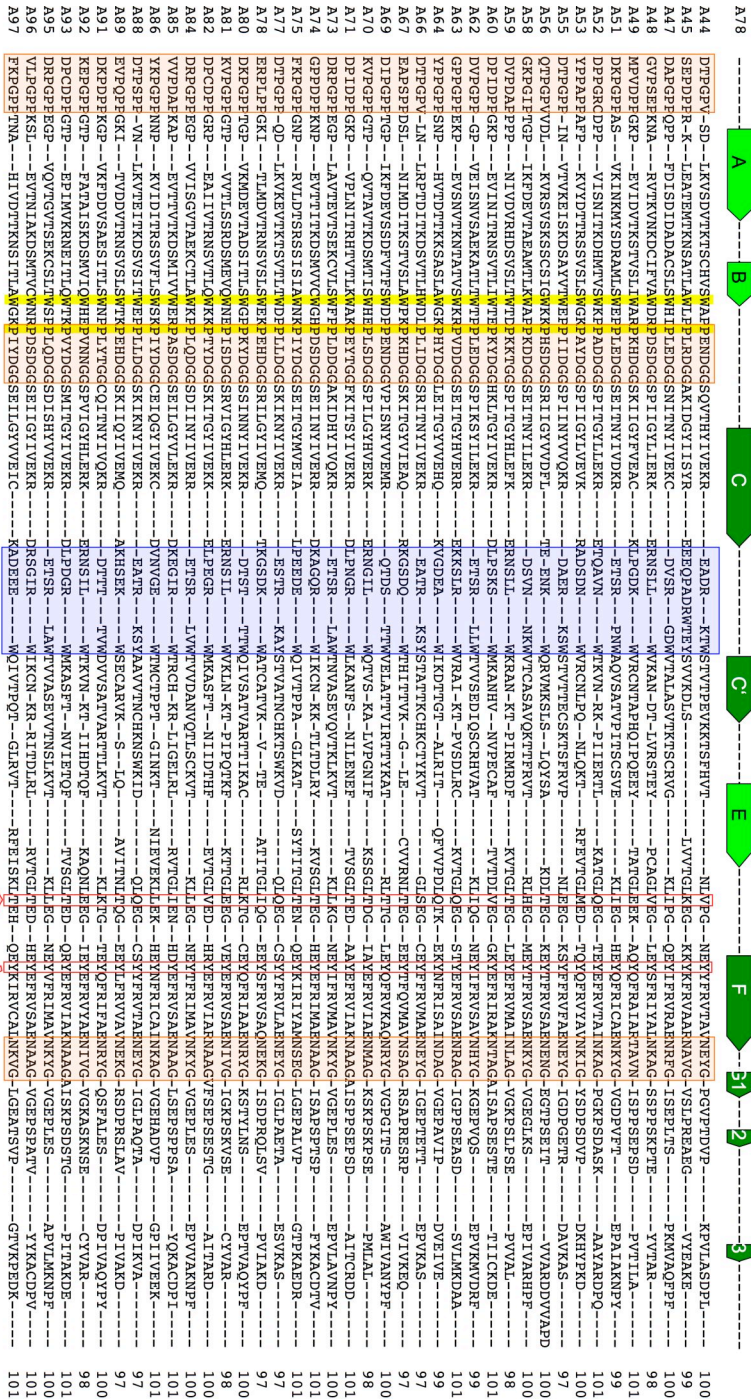
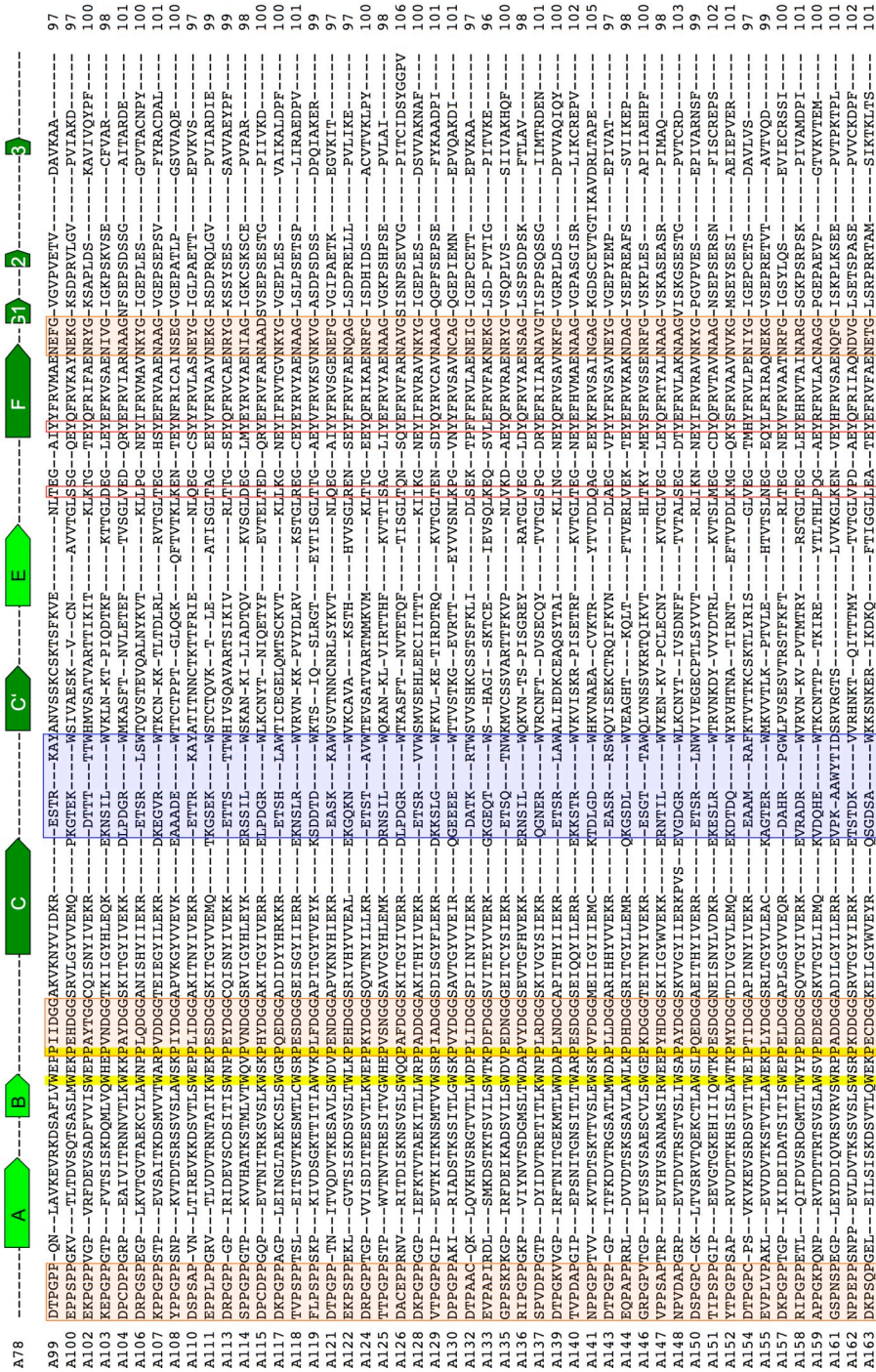Figure 6.4b FNIII sequence alignment (continued)

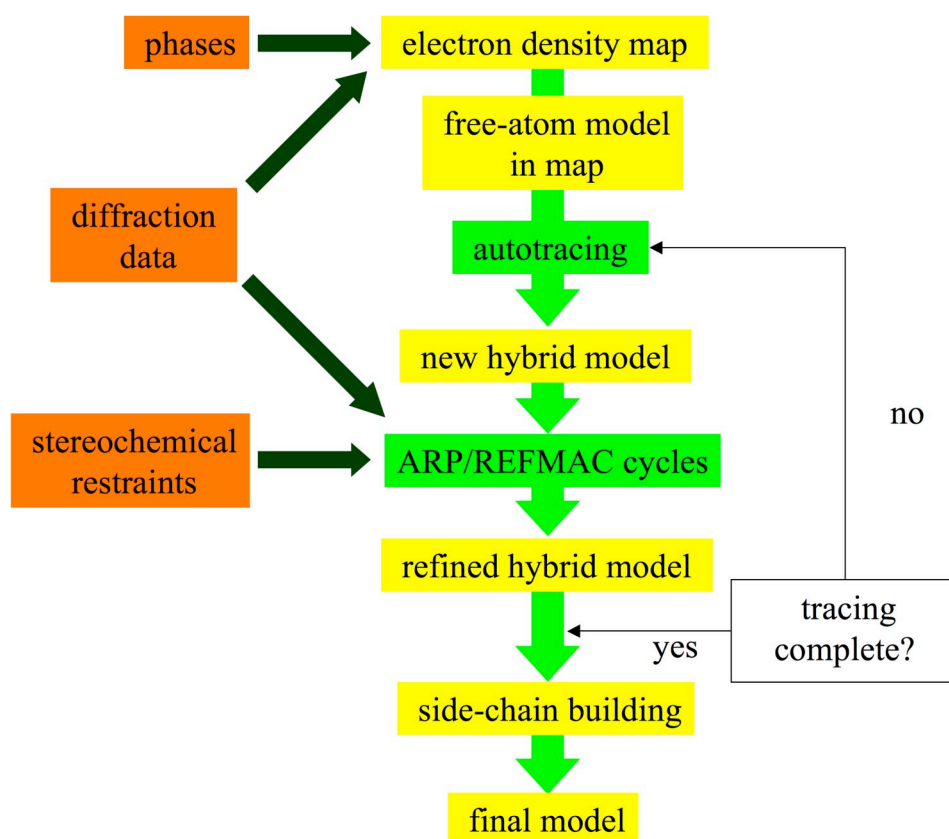**Figure 6.4b FNIII sequence alignment (continued)**

Figure 6.5 A flowchart of the warpNtrace procedure (adapted from Perrakis et al., 1999).
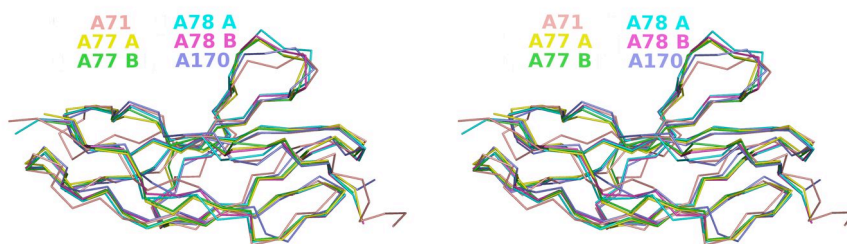

Figure 6.6 Stereo image of the superposition of titin's FNIII modules with known structure.
The superposition reveals the high similarity of these modules. Only the N- (left) and C-termini (right) deviate in in C_α-positions.

6.2.3a   H-bonding in the interface between A77 and A78

Table 6.1 H-bonding network in the A77-A78 interface.

Module-module interactions (direct and via water molecules)					
A77 residue	Atom	A78 residue	Atom	Distance [Å]	Notes[1]
Glu71[e]	OE2	His128[b]	NE2	2.9	direct
Lys98[e]	NZ	Glu180[f]	OE2	2.8	direct
Lys98[e]	O	Lys181[f]	Nz	2.9	direct
Lys19[a]	NZ	His128[b]	ND1	2.9/2.8	via water
Glu71[e]	OE2	Asp129[b]	OD1	3.2/2.7	via water
Lys98[e]	NZ	Gly131[b]	O	2.8/2.7	via water

Linker-to-module interactions (direct and via water molecules)[3]					
Linker residue	Atom	Module residue	Atom	Distance [Å]	Notes[1]
Ala99	N	Cys73[c]	O	2.9	direct
Ala99	O	Gly72[c]	N	2.7	direct
Ser100	N	Lys98[e]	O	3.1	direct
Ser100	OG	Ser132[b]	OG	2.7	direct
Ser100	OG	Lys181[d]	NZ	2.8	direct
Ser100	O	Val17[a]	O	3.1/3.0	via water
Glu101	N	Gly130[b]	O	2.8	direct
Glu101	OE1	Lys19[a]	N	3.1	direct
Glu101	OE2	Arg102	NH2	2.8/2.8	via water
Glu101	O	Asn179[d]	ND	3.0	direct
Arg102	NH2	Ile183[g]	O	3.1	direct
Arg102	NH2	Asp185[h]	N	3.1	direct
Arg102	NH2	Asp185[h]	OD1	3.3/2.6	via water
Arg102	NH2	Asp185[h]	OD2	2.9/3.4	via water
Arg102	NH2	Asp185[h]	OD1	3.3	direct

[1]listed are only waters with electron density observed in both chains in the asymmetric unit. interactions via water are given in: distance of water molecule first atom/second atom

positions of the residues on loops: [a]AB-loop A77, [e]EF-loop (A77), [e]G3-strand (A77), [g]G1-strand (A78), [b]BC-loop (A78), [d]F-strand (A78), [f]FG1-loop (A78), [h]G1G2-loop (A78)

[3]linker: residues between the last strand of A77 and the first residue of the PxxP-motif in A78
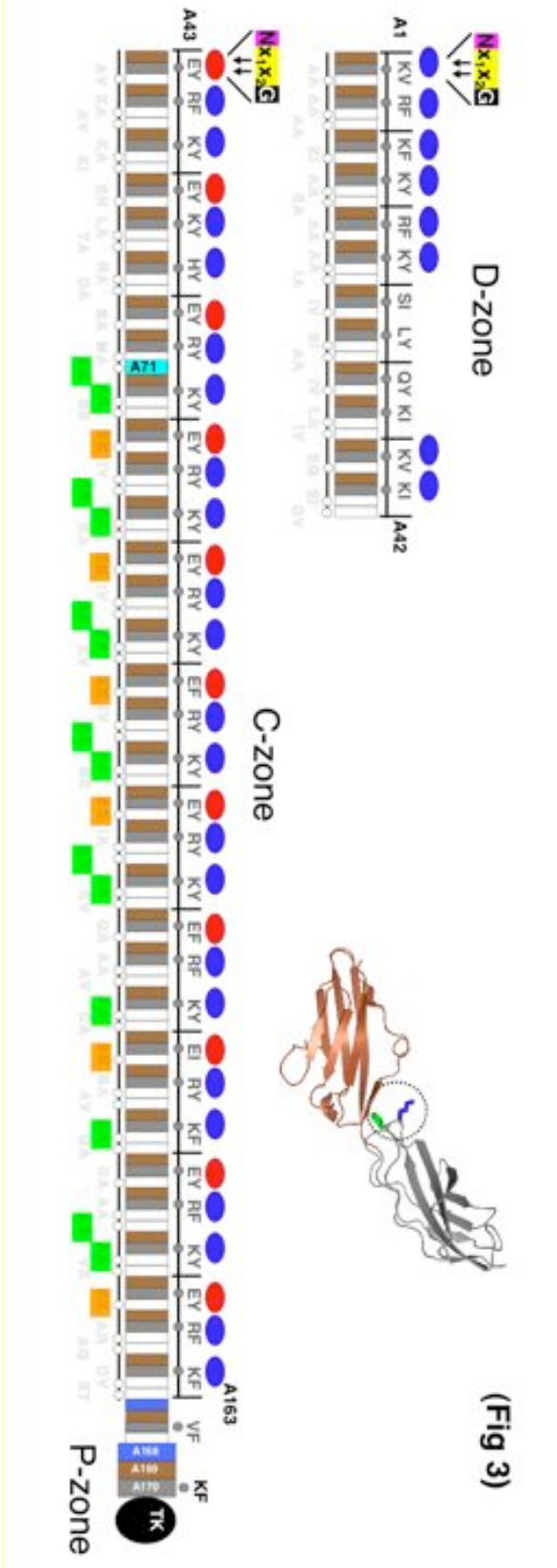
Figure 6.7 IG-FnIII domain interfaces in the titin A-band.
Ig and FnIII domains are shown as filled boxes, where Ig domains flanked by FnIII are in brown, FnIII following an N-terminal Ig in grey, other FnIII in white and Ig followed by another C-terminal Ig in blue. Fn^{A71}, the only other FnIII of titin with known structure (1BPV), is in cyan. The two residues in variable positions of the NxxG -turn of FnIII domains are displayed. Domain super-repeats are indicated. The pattern of alternating charges is displayed, where red and blue indicate negative and positive charges, respectively. Figure adapted from Mrosek *et al.*, 2007

Acknowledgement

Research described in this thesis was carried out at the Biozentrum of the University of Basel. I am particularly thankful to Prof. Olga Mayans for giving me the opportunity to work in her group. In this excellent scientific environment I was able to explore my research interests and ideas. Her constant guidance, advice and support as well as her personal attention to the progress of the work created a motivating and constructive atmosphere throughout my entire PhD, which will also be inspiring for me beyond this thesis.

I would also like to thank Claudia Muhle-Goll, our collaborator in the A77-A78 project as well as Siegfried Sterner, our collaborator in the *Ss*TrpB2b project. Without their help realization of these projects would not have been possible.

Now I would like to thank Prof. Dr. Michael N. Hall and PD Dr. Michael Hennig for their effort in reviewing my thesis and accepting to be my "Prüfungsvorsitzender" and co-examiner, respectively.

Many thanks to the people from the Mayans group (Biozentrum Basel), particularly Marco Marino and Michael Mrosek my true accompanists from Freiburg, who provided me with an inspiring environment and a lot of help through all phases of my work. And also Darko Skegro, Zöhre Ucurum, Eleonore von Castelmur, Sonja Leopoldseder and Pilar Garcia for their help, discussions and scientific interest beyond topics of my thesis.

Ohne meine Freunde, die schon seit Jahren respektvoll, treu und verständnisvoll an meiner Seite stehen wäre meine Leben sicherlich ärmer. Ich möchte all diesen Menschen danken, daß sie mir stets eine große Stütze waren, ohne die ich ins Wanken gekommen wäre. Dafür daß sie mir neue Horizonte eröffnet haben und mir die Kraft gegeben haben durch dunkle Täler auf fruchtbare Wiesen zu treten. Speziell möchte ich meinen besten Freunden Heike-Alice Markel und Frank Engesser wie auch meiner Schwester Christine Bucher-Flad und deren Partner Günter Flad danken, daß sie mich in meinem Wachstum unterstützt haben und dies teilweise initiiert haben. Ich möchte außerdem all den lieben Menschen in der „Brennessel" und meinen Freunden Sebastian Geissert, Stefanie Wienert, Sandro Müller, Kathi Bieck und Thilo Schäfer dafür danken, daß sie immer für mich da waren wenn ich sie gebraucht habe.

Ein stiller Dank geht an die Menschen, die hier nicht genannt werden, die aber nichts desto trotz wichtige Meilensteine auf meinem Weg bis zum jetzigen Zeitpunkt darstellen.

Am Ende möchte ich meinen Eltern danken, denen ich diese Arbeit widmen möchte. Dafür daß sie mich treu auf all meinen Wegen mit ihrer Liebe begleitet haben und mich, wo immer sie das konnten, bei der Verwirklichung meiner Träume mit allem was sie hatten unterstützt haben.

Curriculum vitae

Personal details

Name:	Rainer Michael Bucher
Date of birth:	15.05.1975 in Lörrach, Germany
Nationality:	German
Languages:	German (native), English (fluent), Spanish (B-level)
Marital status:	unmarried

Education

October 2002 - to date　　　　**PhD thesis in protein X-ray crystallography**

Biozentrum, University of Basel, Switzerland.
Group of Dr. Olga Mayans.

Topic: "Structural characterization of prokaryotic and eukaryotic proteins by X-ray crystallography"

Examiners: Prof. Olga Mayans, PD Dr. Michael Hennig
Prüfungsvorsitzender: Prof. Dr. Michael N. Hall

May 2001 - February 2002　　　**Diploma thesis in protein X-ray crystallography**

Department of Chemistry, University of Freiburg, Germany.
Group of Prof Georg Schulz.

Title:	Darstellung der Mutante D204C des Porins aus *Rhodobacter blasticus* and Modifikationen durch kovalente Anbindung geladener Moleküle zur Induktion von Leitfähigkeitsasymmetrie
Grade:	2.0 (good)

October 1998 - April 2001　　　**Diploma in Chemistry (major in Biochemistry)**

Department of Chemistry, University of Freiburg, Germany.

Graduation:	Diplom-Chemiker
Grade:	2.0 (good)

October 1996 - September 1998　　**Vordiplom in Chemistry**

Department of Chemistry, University of Freiburg, Germany.

Grade:	2.0 (good)

September 1995 - September 1996 Civil service

> Tüllinger Höhe in Lörrach, Germany.

September 1992 - August 1995   Wirtschaftsgymnasium, Lörrach

> Grade: 1.6 (very good - good)
> Graduation: Abitur (A-level)

September 1986 - August 1992   Realschule, Efringen-Kirchen

> Grade: 1.6 (very good - good)
> Graduation: Mittlere Reife

September 1982 - August 1986   Grundschule, Efringen-Kirchen

> Primary school

Oral presentations:

October 2003   **Amersham Biosciences seminar:** "Aekta User Club"
Title: Purification of a protein switch involved in nuclear migration and anchorage

August 2003   **Amersham Biosciences seminar:** "From Gene to Functional Protein"
Title: Expression and purification of a protein switch involved in nuclear migration and anchorage

Workshops:

August 2004   **PEPC4: Practical course at the EMBL outstation at the DESY, Hamburg, Germany.**
Topic: "Protein expression, purification and crystallization"
Oral presentation: "Expression and purification of the SUN domain of UNC-84"

Publications:

Garcia P., Ucurum Z., Bucher R., Svergun D.I., Huber T., Lustig A., Konarev P.V., Marino M. and O. Mayans (2006). "Molecular insights into the self-assembly mechanism of dystrophia myotonica kinase." FASEB J. 20: 1142-1151