Strukturelle Dynamik von Peptidyl Carrier Domänen in nicht-ribosomalen Peptid-Synthetasen

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich 14
Biochemie, Chemie und Pharmazie
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Alexander Koglin
aus Potsdam

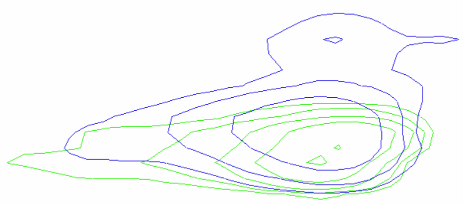
Frankfurt 2006
D30

vom Fachbereich 14 ‚Biochemie, Chemie und Pharmazie' der
Johann Wolfgang Goethe – Universität Frankfurt am Main als Dissertation angenommen.

Dekan: Prof. Dr. Harald Schwalbe

1. Gutachter: Prof. Dr. Volker Dötsch
2. Gutachter: Prof. Dr. Mohamed A. Marahiel (Phillips-Universität, Marburg)

Datum der Disputation:



1 Zusammenfassung

Eine große Zahl natürlicher, so genannter sekundärer Metabolite sind kleine und strukturell sehr verschiedene Polypeptide und Polyketide. Diese bioaktiven Substanzen haben im allgemeinen ein breit aufgestelltes therapeutisches Potential und werden von verschiedenen bakteriellen Stämmen und Pilzen biosynthetisiert. Sie sind sowohl biologisch, als auch therapeutisch wichtig als Cytostatika, Immunsuppressiva und Antibiotika mit einem sehr großen antibakteriellen und antiviralen Potential. Die trotz ihrer kleinen Größe, komplexen Polypeptide und Polyketide werden von modular aufgebauten Megaenzymen in mehrstufigen Mechanismen synthetisiert. Für die Synthese dieser Peptide sind sehr große Proteincluster verantwortlich, die meistens aus einer begrenzten Anzahl sehr großer, Multidomänen umfassenden, Superenzyme aufgebaut werden. Diese Proteincluster mit einem Molekulargewicht bis in den Bereich von MegaDalton werden als nicht-ribosomale Peptidsynthetasen (NRPS) und Polyketidsynthetasen (PKS) bezeichnet. Die NRPS Systeme zeichnen sich dadurch aus, daß für die biosynthetisierten Polypeptide keine Information in Form von Nukleinsäuren wie DNA oder RNA kodiert (Walsh, C.T., 2004; Sieber & Marahiel, 2005).

Für die Synthese der Polypeptide ist eine Aktivierung der einzelnen Bausteine, der Aminosäuren, durch Amino-acyl-adenylierung notwendig. Im Anschluß an die Aktivierung, wird die aktivierte Aminosäure über einen Thioester gebunden weitertransportiert. Die Thioesterbildung erfolgt an Cysteaminthiolgruppen intrinsischer 4'-Phosphopantethein-kofaktoren. Eine Modul einer NRPS stellt eine geschlossene Einheit zum Einbau einer Aminosäure mit einer hohen Spezifität für das Substrat und die biosynthetische Reaktion dar. Diese Module sind aus Domänen aufgebaut, die definierte Funktionen haben und mittels flexibler Linker miteinander verbunden sind. Die Domänen werden nach ihrer Funktion unterschieden. Die Acyl-adenylierung oder Aktivierung eines Substrates, beispielsweise einer Aminosäure, erfolgt durch die A-Domänen. Die Peptidyl- oder Acyltransportfunktion der aktivierten Substrate wird durch Thioester-domänen (T-

Domäne), auch PCP (peptidyl carrier domain) genannt, bewältigt. Die Biosynthese der Kopplungsreaktion, beispielsweise die Ausbildung der Peptidbindung in NRPS Systemen, erfolgt an den Kondensations-Domänen (C-Domäne). Für die Substratspezifität eines Synthesemoduls sind die A-Domänen verantwortlich, welche die Aktivierung eines Substrat durch ATP-Hydrolyse ermöglichen. In NRPS Systemen sind auch Zyklisierungsreaktionen, durchgeführt von Cyclase-Domänen (Cy-Domänen), L/D-Epimerase-funktionen (E-Domänen) und N-Methylierungen (M-Domänen) beschrieben. So wird in Tyrocidin A an zwei Positionen spezifisch Phenylalanin in die D-Form epimerisiert und anschließend in der Peptidbiosynthese verwendet. Die Interaktion und Erkennung zwischen den multi-modularen Superenzymen, zum korrekten Aufbau der kompletten Synthetase, wurden in letzter Zeit Kommunikations-Domänen (COM-Domänen) beschrieben. Wie die aufgebaute Synthetase die korrekte Sequenz der biosynthetischen Reaktionsschritte sicherstellt ist nicht bekannt. Die enorme Diversität biosynthetischer Reaktionen in NRPS Systemen und die hohe Substratvielfalt in den verschiedensten Synthetasen unterschiedlicher Stämme eröffnet ein weites Feld für mögliche Neukombinationen von Modulen und Modifikationen von Produkten, um neue bioaktive Polypeptide mit antibiotischen Eigenschaften durch die Gestaltung neuer biosynthetischer Reaktionswege zu erhalten.

Die Biosyntheseprodukte der NRPS und PKS Systeme lassen sich Gruppen kategorisieren wie Peptidantibiotika, beispielsweise β-Lactame und makrozyklischer Polypeptide. Weitere Gruppen sind die makrozyklischen Lactone, beispielsweise Polyene und Makrolide, aromatische Verbindungen, wie Chloramphenicol, und Chinone (Tetracyclin). Die näher diskutierten Beispiele sind die antibakteriellen Polypeptide Surfactin und Tyrocidin A. Surfactin ist ein antibakteriell wirkendes makrozyklisches Lipoheptapeptid, welches von *Bacillus subtilis* synthetisiert wird und ein enormes antivirales Potential besitzt. Tyrocidin A ist ein antibakteriell wirkendes makrozyklisches Decapeptid und wird von *Bacillus brevis* und *Brevisbacillus parabrevis* synthetisiert.

Zusätzlich werden viele bakterielle Toxine ebenfalls durch solche Systeme multi-modularer Synthetasen erzeugt. Ein Beispiel ist das Polyketid Vibriobactin, das Toxin des humanpathogenen Bakterium *Vibrio cholerae*.

Ein großes Problem der wachsenden Weltbevölkerung moderner Gesellschaften und in den Entwicklungsländern ist die wachsende Zahl multiresistenter Bakterienstämme. Die starke Progression in der Entwicklung von Resistenzen gegen Antibiotika ist auch Gegenstand des aktuellen WHO-Reports. Alarmierend ist die Resistenzentwicklung gegen die sogenannten Reserveantibiotika Vancomycin und Ceftazidim. Ein umfangreicheres Verständnis der Interaktion zwischen Domänen in einem Modul und zwischen Modulen eines NRPS Systems ist Grundlage für die Neukombination unterschiedlicher Module zur erfolgreichen Gestaltung neuer Biosynthesen. Da die meisten dieser Biosynthesen oder die Synthese alternativer Substanzen nicht in der Organischen Chemie zu realisieren sind oder die Produkte zu teuer wären, um diese in großen Mengen zu erzeugen, muß das Ziel sein die NRPS und PKS Systeme in ihrem modularen Aufbau und ihre Interaktion zu verstehen, um alternative Antibiotika biosynthetisch herzustellen.

Peptidyl Carrier Proteine (PCPs) sind kleine zentrale Transport-Domänen, integriert in den Modulen nicht-ribosomaler Peptidsynthetasen (NRPSs). PCPs tragen kovalent über eine Phosphoesterbindung einen aus dem Protein herausragenden 4′-phosphopantetheinyl (4′-PP) Kofaktor. Der 4′-PP Kofaktor ist an der Seitenkette eines hochkonservierten Serins gebunden, welche ein zentraler Bestandteil der Phosphopantethein-Erkennungs-Sequenz ist. Die Erkennungssequenz ist homolog in vielen Proteinen mit ähnlicher Funktion, inklusive Acyl Carrier Proteinen (ACPs) der Fettsäuresynthetasen (FAS) und der Polyketidsynthetasen (PKS). Die Thiolgruppe des 4′-PP Kofaktors dient zum aktiven Transport der Substrate und der Intermediate der NRPS Systeme. Die generelle Organisation und die Kontrolle der exakt aufeinander folgenden Reaktionsschritte in der Peptidsynthetase, ist die entscheidende Frage für die Funktion des Proteinclusters (*assembly line mechanism*). In Modulen der NRPS Systeme folgen die PCP-Domänen C-terminal auf die Adenylierungs-domänen (A-

Domäne). Die Aufgabe der A-Domänen ist die Selektion and die Aktivierung einer spezifischen Aminosäure für die „assembly line". Die eigentliche Bildung der Peptidbindung erfolgt an der Kondensations-Domäne (C-Domäne). Der Transfer der Peptidintermediate und der aktivierten Aminosäuren zwischen A-Domänen und C-Domänen ist Aufgabe der PCPs. Um diese Funktion erfüllen zu können, ist eine große Bewegung in PCPs, bzw. des 4'-PP Kofaktors notwendig, welche als „swinging arm model" (Weber et al., 2001) beschrieben wurde. Die PCPs koordinieren damit die Peptidbiosynthese während sie mit diversen Domänen der Synthetasen spezifisch wechselwirken müssen. Die molekularen Mechanismen des Transportes wurden bisher allerdings nicht untersucht. Eine Dynamik der Transport-Domänen wurde bereits postuliert (Kim & Prestegard, 1989; Andrec et al., 1995), konnte bisher aber nicht gezeigt werden (Weber et al., 2001). Interessanterweise zeigt sowohl apo-PCP (ohne den kovalent gebundenen 4'-PP Kofaktor) also auch holo-PCP langsamen chemischen Austausch, der als jeweils zwei stabile Konformationen beschrieben werden konnte. Diese jeweils zwei stabilen Zustände, welche sich im Austausch befinden, wurden als A und A*, für apo-PCP, und entsprechend H und H* für holo-PCP bezeichnet. Während der A- und der H-Zustand sich sowohl voneinander als auch von den entsprechenden A* und H*-Zuständen unterscheiden und spezifisch für die apo- und die holo-Form von PCP sind, ist die kalkulierte Struktur vom A*-Zustand größten Teils identisch mit der des H*-Zustandes. Die erhaltenen NMR-Strukturen des A-Zustandes, des H-Zustandes und des gemeinsamen A/H-Zustandes beschreiben in ihrer Gesamtheit ein neues Modell für ein allosterie-kontrolliertes System dualer konformationeller Zwei-Zustands-Dynamik. Zu dem beobachteten konformationellen Austausch der PCP-Domäne, konnte die Bewegung des 4'-PP Kofaktors koordiniert werden. Die Bewegung des 4'-PP Kofaktors in Verbindung mit dem konformationellen Austausch der PCP-Domäne charakterisiert die Interaktion mit katalytischen Domänen eines NRPS Moduls. Des weiteren konnte mit Hilfe des Modells die Wechselwirkung mit externen Interaktionspartnern, wie der Thioesterase II und der 4'-PP Transferase, untersucht werden.

Die externe Thioesterase II der Surfactin-Synthetase (SrfTEII) von *Bacillus subtilis* ist ein separat expremiertes 28 KDa Protein. Sie gehört zur Familie der α/β-Hydrolasen und ist verantwortlich für die Regenerierung falsch beladener 4'-PP Kofaktoren der Peptidyl Carrier Domänen. Die SrfTEII wurde mittels Lösungs-NMR untersucht, die Resonanzen wurden zugeordnet, erste strukturelle Modelle konnte berechnet werden und das Interaktionsverhalten mit verschiedenen modifizierten Kofaktoren und PCPs wurde analysiert. Die Spezifität der Substraterkennung durch die SrfTEII kann beschrieben werden. Interessanterweise zeigt auch die SrfTEII Doppelpeaks für einzelne Aminosäuren, diese können als Indikator für eine spezifische Substraterkennung durch das Enzym verwendet werden und helfen den funktionellen Unterschied zwischen der SrfTEI-Domäne und SrfTEII zu verstehen.

2 Introduction

2.1 Role of native antibiotic synthesis

A vast number of small and structurally diverse bioactive peptides and polyketides with broad therapeutical potential are produced by the modular megaenzyme super family of non-ribosomal peptide synthetases (NRPSs) and polyketide synthetases (PKS) (Walsh, C.T., 2004; Sieber & Marahiel, 2004). These complex peptides are synthesized by multiple-step mechanisms that require amino acyl adenylation of the building blocks followed by thioesterification on distinctive cysteamine thiol groups of intrinsic 4'-phosphopantetheine (4'-PP) cofactors. These secondary metabolites of different bacterial strains and fungus are biologically important as human pathogenic toxins, cytostatics, immune depressives and antibiotics with enormous antibacterial and antiviral potential.

These biosynthetic protein clusters are mostly assembled from a limited number of gene products with a molecular weight up to the MDa range to form non-ribosomal peptide synthetases (NRPS) and polyketide synthetases (PKS), offering an enormous diversity of biosynthetic reactions and products with a wide field of potential modifications. Many of these gene products are constructed of arranged modules with a well defined specificity for a substrate and for a specific biosynthetic reaction of each module. The modules are constructed of domains with defined functions and are connected to each other by flexible linkers. These are domain-functions like acyl-adenylation of a substrate like an amino acid residue (A-domain), a peptidyl or acyl carrier-function specific for the activated substrate (T-domain) and a condensation domain, which actually specifically biosynthesize the coupling reaction (C-domain), which actually forms the peptide bond in the NRPS systems. The substrate selectivity is determined by the A-domain, which activates the substrate by adenylation, although cyclase- (Cy-domains), epimerase- (E-domains) and N-methylation (M-domains) functions are described. Recently, the interacting domains for the correct assembly of the complete synthetase consisting of isolated gene products, could be described as communication domains (COM-domain). How the assembled synthetase ensures the correct sequence of biosynthetically reaction steps is not know yet.

The products of this NRPS- and PKS-systems are categorized in groups like peptide antibiotics (e.g. β-lactams, macro cyclic polypeptides), macrocyclic lactones (e.g. polyenes and macrolides), aromatics (chloramphenicol) and chinones (tetracycline). More examples for antibacterial polypeptides are Surfactin, a cyclic lipoheptapeptide biosynthesized by *Bacillus subtilis* with an enormous antiviral potential or Tyrocidine an antibacterial cyclic decapeptide from *Bacillus brevis* and *Brevisbacillus parabrevis*.

Additionally, many bacterial toxins are synthesized by these multi-modular synthetase systems, like vibriobactin, the polyketide toxin from *Vibrio cholerae*. Most of the biosynthetic reactions can not be realized by organic chemistry to produce these polypeptides or polyketides synthetically in large amounts or the products would be too expensive for a general application, especially in developing countries. Modules of NRPS or PKS systems can act stereospecifically on substrates, e.g. epimerized D-amino acid residues (e.g. D-Phenylalanine in Tyrocidine A), specifically attach modified fatty acids (e.g. β-hydroxy-myristoylic acid in Surfactin) or the product contains heterocyclic compounds (threonine in Acinetobactin).

One major problem of modern societies and developing countries is the progressive development of antibiotic resistance of bacteria and viruses. The latest WHO report on antibiotic resistance describes an increasing development of resistance mechanisms even on last line of defense antibiotics like vancomycin and ceftazidime. These facts make research for alternative antibiotics and new strategies in antibacterial treatment necessary and it is the incentive for working on non-ribosomal peptide synthetases and polyketide synthetases. The modular organization of NRPS and PKS suggests that the biosynthesis of new antibiotics by recombination of different modules with different specificities by designing a new biosynthetic pathway should be possible. This approach still needs enhanced efforts before new antibiotics will be available. Particularly, the limited knowledge about domain-domain interaction of subunits of NRPS and PKS systems is still a challenge.

2.2 The Tyrocidine A synthetase

Tyrocidine A is a macrocyclic decapeptide biosynthesized by an assembled protein cluster consisting of three gene products to form the multifunctional modular non-ribosomal peptide synthetase (NRPS). The antibiotic potential of the lipophilic molecule is due to the intercalation into the cell membrane and its function as an ionophore. The uncontrolled permeability for ions of the plasma membrane causes cell death.

The assembly line of Tyrocidine is organized in three transcripts TycA (123 kDa) for activation and epimerization of D-Phe, TycB (405 kDa) to finally synthesize a tetrapeptide and TycC a 742 kDa protein to synthesize the native decapeptide and for cyclization of the matured peptide.

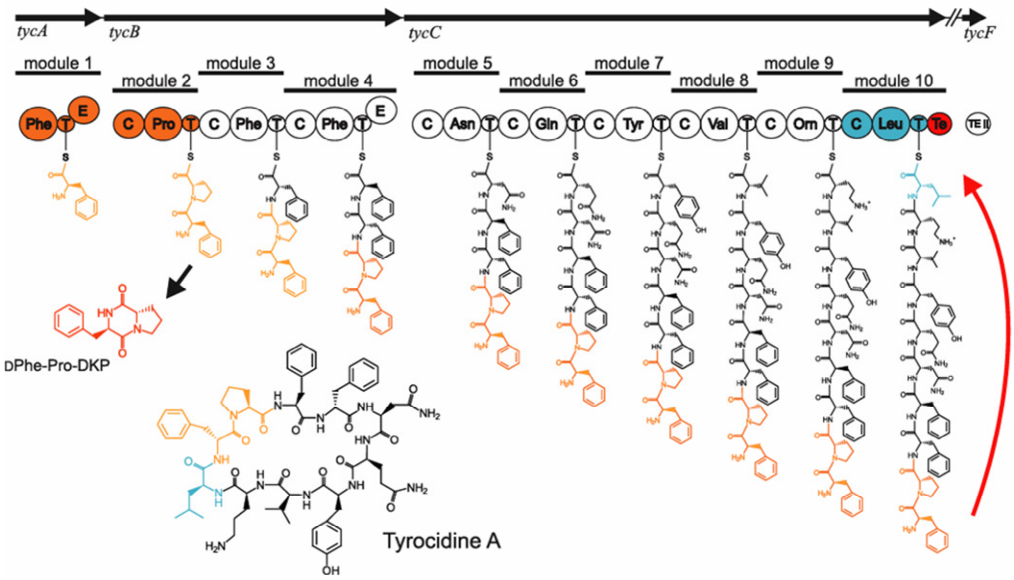


Fig. 1: Functional organization of the Tyrocidine A synthetase. The Tyrocidine A synthetase is organized out of three multimodular transcripts: tycA, tycB and tycC. Each module of the nonribosomal peptide synthetase is specified for the incorporation of one amino acid residue. The cyclization between the first D-Phe (orange) and the last L-Leu (domains shown in blue) is synthesized by the C-terminal thioesterase-domain (Mootz et al., 2000).

Tyrocidine A is biosynthesized by the assembled synthetase in the following order: D-Phe ·L-Pro ·L-Phe ·D-Phe ·L-Asn ·L-Gln ·L-Tyr ·L-Val ·L-Orn ·L-Leu and subsequently cyclisized by a C-terminal thioesterase domain, as shown in Fig. 1. Figure 1 points out the modular organization of the tyrocidine A synthetase and the biosynthetic pathway of Tyrocidine A.

The biosynthesis and subsequent cyclization are symbolized. The growing peptide chain is shown bound to Peptidyl carrier proteins (PCP) via their posttranslational covalently bound 4'-Phosphopantheteinyl-cofactor (*multiple thiotemplate-mechanism*; Stein et al., 1996). The PCPs are labeled 'T', while the C-domains are labeled 'C' and the substrate specific A-domains are labeled according to their substrate amino acid residue.

2.3 Function of Peptidyl carrier proteins

The general organization and control of the subsequent reaction steps in polypeptide synthesis by a NRPS-system is the crucial question for a new combination of modules. The thioester-domains (T-domains), also called Peptidyl carrier proteins (PCPs), are either centrally integrated or isolated shuttle elements of modular non-ribosomal peptide synthetases (NRPSs). These domains or isolated proteins carry a protruding 4'-phosphopantetheinyl (4'-PP) cofactor and are sequentially and structurally conserved (Fig. 2).

```
TYCC_BREVP     17 sKLAEIWERVLGVS------GIGILDNFFQ-IGGHSLKAMAVAAQVHREYQVELPLKVLF--AQPTIKALAQYV 81
SRFAA_BACSU   978 KAIAAIWQDVLNVE------KAGIFDNFFE-TGGHSLKAMTLLTKIHKETGIEIPQQFLF--EHPTITALAEEA 1042
SRFAA_BACSU  2015 QKVADIWAQVLQAE------QVGAYDHFFD-IGGHSLAGMKMPALVHQELGVELSLKDLF--QSPTVEGLAQVI 2079
SRFAB_BACSU   972 KKLAEIWEGILGVK-------AGVTDNFFM-IGGHSLKAMMMTAKIQEHFHKEVPIKVLF--EKPTIQELALYL 1035
SRFAB_BACSU  2017 ESLCRIWQKTLGIE------AIGIDDNFFD LGGHSLKGMMLIANIQAELEKSVPLKALF--EQPTVRQLAAYM 2081
PPS1_BACSU    968 MKLSQLWEDVLKNG------PVGIHDNFFD-RGGHSLKATALVSRIAKEFDVQVPLKDVF--AHPTVEGLATVI 1032
GRSB_BACBR    978 EKLAKIWEEVLGIS------QIGIQDNFFS_LGGHSLKAITLISRMNKECNVDIPLRLLF--EAPTIQEISNYI 1042
GRSB_BACBR   2014 MKLAEIWHNVLGVN------KIGVLDNFFE-LGGHSLRAMTMISQVHKEFDVELPLKVLF--ETPTISALAQYI 2078
GRSB_BACBR   4098 AQLVLIWQEVLGIE------LIGITDNFFE-LGGHSLKATLLVAKIYEYMQIEMPLNVVF--KHSTIMKIAEYI 4162
LYS2_YEAST    850 REVRDLWLSILPTKPA----SVSPDDSFFD-LGGHSILATKMIFTLKKKLQVDLPLGTIF--KYPTIKAFAAEI 916
LYS2_SCHPO    887 RDIRDIWLRIIPHAT-----DVNKKASFFD-IGGHSILATRLIFELRKKFAVNVPLGLVF--SEPTIEGLAKEI 952
ACVS_CEPAC    797 SDLAAIWGNILSVPAQ----DIGSESNFFR-LGGHSIACIQLIARVRQQLGQGITLEEVF--QTKTLRAMAALL 863
ACVS_EMENI    852 IALGKIWADVLGAHHL----SISRKDNFFR-LGGHSITCIQLIARIRQQLGVIISIEDVF--SSRTLERMAELL 918
ACVS_NOCLA    790 EQLRAIWSEVLGVPQN----RIGERDDFFR-LGGHSISCILLIARVRQRLSLSLGVEDVF--ALRTLDALAGHL 856
```

Fig. 2: Sequential alignment of Peptidyl carrier domains. The sequential alignment of PCPs with an averaged size of 80 amino acid residues from different organism and NRPS systems like the Surfactin synthetase from *Bacillus subtilis*, the Gramicidin synthetase from *Bacillus brevis* and the ACV synthetase from different fungus to the TycC3-PCP from Tyrocidine A synthetase from *Brevisbacillus parabrevis* show a number of highly conserved amino acid residues. Conserved amino acid residues are a Trp (Trp23 for TycC3-PCP) at the end of helix I, the function is scaffolding the protein fold, the domain-recognition sequence (G/S-X-X-D-N-F-F) right before the 4'-PP-recognition sequence (G-G-H-S-L) with the essential Ser residue and a C-terminal Thr residue in front of the last helix.

A primary sequence analysis of several carrier domains from different strains stresses the highly conserved 4'-phosphopantethein recognition sequence with the essential serine residue and the averaged fold of a three or four-helix bundle with an overall size of 70 to 90 amino acid residues (Fig. 2).

The 4'-PP cofactors are attached to the side chain of the conserved serine residues of the PCPs via a phosphoester-bond. The highly conserved and essential serine residue is embedded in the so called phosphopantheteine recognition sequence, which is colored red in the Fig. 2. The structural organization and the recognition sequence are conserved in a wide range of proteins with functions similar to PCPs, acyl carrier proteins in fatty acids synthetases (FAS) and polyketide synthetases (PKS). PCPs constitute central shuttle units of the individual NRPS modules where they act as carriers for all intermediates of the peptide synthesis during such an assembly line mechanism. Within NRPS modules the PCP domains are fused at their N-terminal end to the A-domains that are responsible for the selection and activation of the specific amino acid constituents of the final peptide products. Peptide-bond forming condensing domains (C-domains) are located C-terminal to the PCPs. The transfer of peptide intermediates between A- and C-domains by PCPs demands therefore large movements of the 4'-PP cofactor that have been described by a "swinging arm" model (model in Fig. 3).

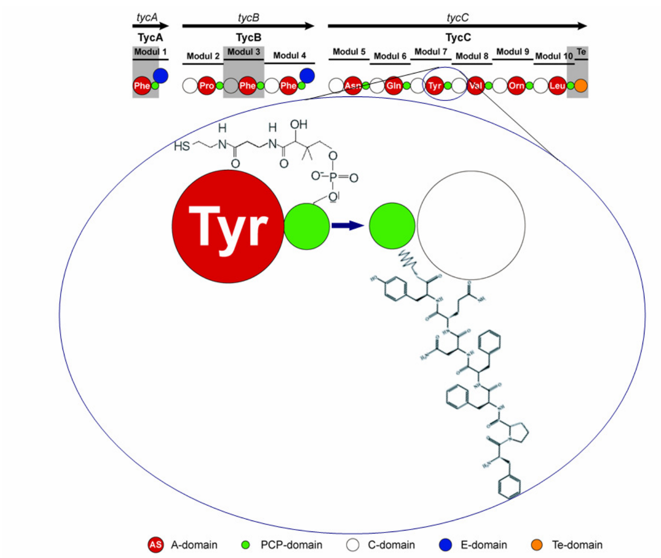


Fig. 3: Functional model for the carrier domain TycC3-PCP. The unloaded 4'-PP cofactor will be modified by the tyrosine specific A-domain (left) and the growing peptide will be carried to C-terminal C-domain (right). In the upper row the domain organization of the tyrocidine synthetase is shown.

However, underlying molecular details that induce this proposed mobility of the 4′-PP cofactor have not been analyzed so far. The logical conclusion is that PCPs coordinate peptide synthesis by interaction with a diverse set of enzymes.

The posttranslational modification of carrier domains with the attached 4′-PP cofactor is carried out by enzymes of the 4′-phophopantheteine transferase (PPTase) family. PPTases specifically catalyze the phosphoester formation of the conserved serine residue with Coenzyme A, where the 4′-PP part is transferred to the serine and the remaining 3′-5′-ADP is released. The PPtase Sfp from the Surfactin NRPS system shows a reduced selectivity and can be used for a wide range of PCPs including the TycC3-PCP, the third PCP domain of the tyrocidine C-module.

The intention to understand the functional organization of NRPS and PKS systems resulted in great endeavors to reveal structural details of isolated domains and larger constructs and finally culminated in the complete x-ray crystal structure of a fatty acid synthetase (Maier, T. et al., 2006). Structural studies where performed on isolated domains as well A-domains, for instance the phenylalanine activating domain of the gramicidine S-synthetase I [PheA] (Conti et al., 1997) or C-domains, like the VibH condensation domain from the vibriobactin-synthetase from *Vibrio cholerea* (Keating, T.A. et al., 2002) and carrier domains (others, Weber et al., 2001). Based on structural comparison of crystal structures from the free domain and the substrate bound state (PheA-domain) or rather by the observation of missing electron density in specific areas of modules, revealed the idea of dynamic systems (Kim & Prestegard, 1989; Maier, T. et al., 2006; Jenni, S. et al., 2006). A dynamic behavior for the carrier domains was postulated earlier (Kim & Prestegard, 1989; Andrec et al., 1995), but could not be described with structural data (Weber et al., 2001). The published crystal structure of the PPtase Sfp (Reuter et al., 1999), the crucial factor for the posttranslational modification of the central carrier units with the 4′-PP cofactor, is of interest for understanding the general domain interaction within the multi-domain complexes, like the interaction of Peptidyl carrier proteins with Sfp. Sfp is showing a new type of protein fold with an intra-molecular pseudo-symmetry. The Sfp-substrate Co A is bound in the inner core of the protein, so that the pyrophosphate of Co A is coordinated by an essential Mg^{2+}-ion. Unfortunately in the Sfp/Co A co-crystal just the ribose-phosphate residue of the adenyl-residue was visible. The fact that the thiol-group of the Co A can be modified as a thioester and will be still used as the substrate for loading the carrier proteins is remarkable (Quadri et al., 1998; Belshaw et al., 1999). The interaction site of Sfp with PCP was described by mutation analysis (Mofid et al., 2002).

2.4 NMR Spectroscopy

Nuclear magnetic resonance (NMR) is a well understood and established technique for the determination of protein structures with atomic resolution and is a perfect tool to study the structural dynamic and interaction of macromolecules on different time scales, as well as the interactions of proteins and nucleic acids with ligands and enzymatic activities. Due to the observation of hydrogen-bonds and other scaffolding interactions by NMR spectroscopy, the native folding of biomacromolecules can be studied (Smith, L.J. et al., 1996; Klein-Seetharaman et al., 2002).

For example a two-state dynamic between an inactive and the enzymatic active state in a ms to µs time regime was expected for the Gramicidin S PheA-adenylation-domain (Conti et al., 1997) and could be demonstrated for many different systems. An exemplary example for a two-state dynamic of a protein was recently published by D. Kern and co-workers, where they demonstrated the linkage between dynamics and catalysis of the adenylate-kinase (Adk) from *E. coli* and *Aquifex aeolicus* (Wolf-Watz et al., 2004). The Adk converts two molecules of ADP to ATP and AMP and this reaction is correlated with a huge domain movement, comparable to the closing of a hand on a set of finger tips. In the open conformation, two molecules of ADP will be bound as the substrates. In the closed conformation, these two ADPs are brought in tight proximity for converting to ATP and AMP. These structural changes could be demonstrated by NMR-spectroscopy.

A detailed functional research and structural investigation of biomacromolecules by high resolution NMR spectroscopy is limited due the size of the observed macromolecules. A severe line broadening with increasing molecular weight is caused by the inverse relation between the molecule's tumbling rate and its transverse relaxation. The tumbling rate, which is usually in the range of 10^{-9} seconds (van de Ven, 1995), decreases with increasing molecular weight. Fast relaxing signals of large macromolecules result in significantly broadened lines and additionally, the increasing number of spins in a naturally limited frequency range makes resonance-specific assignments more complicated and requires additional techniques (Trbovic *et al.*, 2005) or makes a resonance specific assignment even impossible. The T_2 relaxation mechanism of two ½ spins, J and I and a scalar coupling constant J_{JI}, is dominated by the dipolar coupling (DD) of these two spins and by the chemical shift anisotropy (CSA) of each spin individually.

Interference of these mechanisms give rise to different relaxation rates of the individual multiplet components of the coupled spin system. These different rates are averaged by a 180° pulse on spin I during t_1 evolution of spin S, resulting in elimination of DC/CSA interference. In contrast the transverse relaxation optimised spectroscopy (TROSY) reduces T_2 relaxation exploiting destructive interference between the two relaxation mechanisms (Pervushin, Riek et al. 1997). It is designed in order to select for the longest-lived coherences and isolate these from the ones that relax much faster.

The TROSY technique is well established for ^{1}H-^{15}N and ^{1}H-^{13}C (AX) spin systems resulting in spectra with a higher signal-to-noise ratio and a 60% (^{1}H), respectively 40% (^{15}N) reduced line width (Pervushin et al., 1997). It could be shown that the TROSY effect can also be applied to methyl group AX_3 spin systems (Tugarinov, Hwang et al. 2003) and additionally that the external contributions to the relaxation rates can be minimized by measurement of highly deuterated samples (Tugarinov and Kay 2004).

Finally the precision of NMR-derived structures is not at the same level like x-ray structures, but NMR structures contain information about the native dynamic of macromolecules. The limited spectral range of NMR-spectroscopy and the resulting overlap can be reduced by heteronuclear spectroscopy like double- and triple resonance experiments in three- and four-dimensional spectra. The labeling of macromolecules with NMR-active and stable isotopes like ^{15}N, ^{13}C or as already mentioned perdeuteration are established methods. Additional higher magnetic fields and higher sensitivity, e.g. usage of cryogenic probes, improves the spectral resolution and the signal/noise level. The development of cell-free expression systems and its adaptation to membrane bound proteins (Klammt et al., 2004) enables the development of techniques for amino acid type specific labeling and the combination of several amino acid types, also independent as ^{15}N and ^{13}C labeled amino acid residues (Trbovic et al., 2005).

3 Materials and Methods

3.1 Materials

NMR-Spectrometer Avance700	Bruker BioSpin, Germany
Cryo-probe 1H(13C, 15N), xyz-gradient	Bruker BioSpin, Germany
NMR-Spectrometer Avance900	Bruker BioSpin, Germany
Cryo-probe 1H(13C, 15N), z-gradient	Bruker BioSpin, Germany
Incubator Shaker	Infors AGSwitzerland
Evolution centrifuge (rotors: SS34, GS3)	Sorvall, Kendro Laboratory Products
Centrifuge 5810 R	Eppendorf AG, Germany
Centrifuge 5417C (rotor:F45-30-11)	Eppendorf AG , Germany
Cary-300 UV/Vis spectrophotometer	Cary Instruments, USA
Sonicator Labsonic® (probes: 5 T, 40 T)	B. Braun Biotech International
Waters™ System	Millipore
Ni-NTA™ Fast Flow	Amersham Biosciences GmbH
Q Sepharose™ Fast Flow	Amersham Biosciences GmbH
Amicon Stirred Ultrafiltration Cell	Millipore Corporation
Ultrafiltration Membrane	Millipore Corporation
Biometra Tpersonal Thermocycler	Westburg BV

All chemicals were purchased from Carl Roth AG, Germany or are specifically declared separately.

Media. LB Medium is prepared by dissolving 10 g Tryptone, 5 g Yeast extract and 10 g sodium chloride in 1000ml distilled water. For sterilization the medium is autoclaved for 20min at 121°C.

Minimal Medium consists of 6.35 g Na_2HPO_4, 3 g KH_2PO_4 and 0.5 g NaCl per liter H_2O_{dest}. After autoclaving for 20 min at 121°C the following salts and feedings have to be added: 1 ml 100mM $CaCl_2$, 1 ml 1M $MgSO_4$, 1 g NH_4Cl, 2 g Glucose. In case of auxotrophic bacteria the medium has to be supplemented with amino acids that the strain is not capable to synthesize

To increase the transformation efficiency it is recommended to incubate cells with S.O.C. Medium prior to plating. The latter is prepared with 20 g Tryptone, 5 g Yeast extract and 0.5 g NaCl per litre H_2O_{dest}. After having autoclaved the medium, 20 ml of 1 M Glucose are added. Ampicillin is provided as a 1:1000 stock solutions which correspond to 100 mg/ml. The concentration of the stock solution for IPTG is 1 M.

Bacterial strains. For cloning of DNA the *E.coli* DH5α strain [lacZ∆M15 recA1 endA1 hsdR17 (r_K^- m_K^+) supE44] is used. The transformation efficiency of this strain varies between $1*10^6$ – $3*10^6$ cfu/ng depending on how well the preparation of competent cells works. For a transformation about 100 ng DNA are required.

The *E.coli* BL21 strain [*E. coli*B F^- dcm ompT hsdS(r_B- m_B-) gal lon] is adapted for high-level expression of genes cloned into vectors containing the T7 promoter (e.g. pET). The gene encoding bacteriophage T7 RNA polymerase is carried on the bacteriophage λDE3, which is integrated into the chromosome of BL21. The transformation efficiency of BL21 usually is about ten times less than that of the DH5α strain.

Constructs. The expression constructs are based on the high-copy expression vector pQE70-Hexa-Histidine-tag (Qiagen, Germany) for histidine-tagged protein purification. For functional co-expression of proteins the low-copy expression vector pREP4 was used.

3.2 Methods

DNA Purification. Plasmid DNA is purified using the QIAprep® Spin Miniprep Kit with the according protocol (Qiagen May 2004). In brief, 2ml of an overnight culture are centrifuged and resuspended in RNase containing buffer. Upon lysis, the cell debris is removed and the DNA is bound to a QIAprep spin column, washed and in the end eluted with H_2O_{dest}.

DNA techniques and mutagenesis. The QuickChange kit (Stratagene) was used for DNA mutagenesis according to the manufacturer's recommendations. For the construction of the mutants TycC3-PCP_{S45A} and $SrfTEII_{S86A}$ two mutagenic primers containing the desired mutation were used in a PCR reaction with the pQE70-TycC3-PCP and pQE70-SrfTEII (Qiagen) as templates and *Pfu*Turbo DNA polymerase (Stratagene). For construction of plasmid pREP4-Sfp, the endogenous *SmaI* restriction site of pREP4 was first replaced by *EcoRI* by the QuickChange technique. The resulting plasmid pREP4(*EcoRI*) was then linearized at the single *EcoRI* site. *The Sfp* gene was amplified by PCR from pQE60-*sfp* (Mofid et al., 1999) and cloned into the linearized pREP4 (*EcoRI*) to give pREP4-*sfp*.

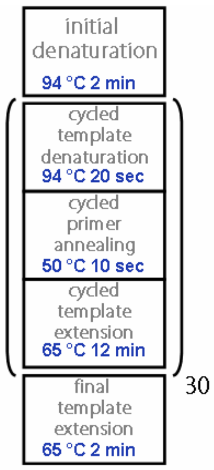


Fig. 4: PCR reaction Stratagene QuikChange® mutagenesis for the S45A-mutation.

In order to remove the template DNA in the PCR reaction it is digested with 1 µl DpnI and subsequent incubation at 37°C for 1 h. Having identified the remaining PCR product on a 0.9% agarose gel and after transformation of isolated plasmid DNA into *E.coli* DH5α cells the accumulated plasmid DNA is purified from an overnight culture. The subsequently isolated plasmid DNA is transformed into E.coli BL21 cells and protein test expressions and DNA sequencing give evidence for the successful mutagenesis.

Transformation of Plasmid-DNA. Competent cells are made using the $CaCl_2$ method according to D. Hanahan et al. (Sambrook and Russell 2001) and then stored at -80°C. For transformation an aliquot is thawed on ice for 2 minutes and 1 µl DNA (~10 ng) is added. After incubation on ice for additional 30 min, cells are heat shocked for 10 sec at 42°C and immediately stored on ice. After adding 150 µl S.O.C. medium the sample is incubated at 37°C for 30 min prior to plating. In order to have single colonies on the LB/selective antibiotic-Agar Petri dishes it proved useful to plate two different concentrations of cells 100 µl resp. 300 µl of the BL21 sample and 50 µl and 150 µl for DH5α.

Stringent control of pET expression systems. The pET System (Novagen) is a very powerful method for expression of small soluble proteins. The target genes are cloned into a pET vector under the control of a strong bacteriophage T7 promoter. This can only be recognized by the bacteriophage T7 RNA polymerase that must be encoded by the host cell, e.g. *E.coli* BL21 and DL39. Upon addition of IPTG, protein expression is induced. In the first step the sugar derivative binds the *lac* repressor so that the *lac* promoter can then be recognized by the bacterial RNA polymerase that transcribes the T7 gene. T7 RNA polymerase is a small, very efficient protein with a higher transcription rate than the bacterial counterpart.

Protein expression and labeling*.* Apo-TycC3-PCP and apo-TycC3-PCP_{S45A} were produced in *E. coli* and purified by Ni^{2+}-affinity chromatography followed by gel filtration as described (Weber et al., 2000). For the production of holo-TycC3-PCP containing a ^{15}N-labelled 4'-PP cofactor, the *E. coli* strain BL21 doubly transfected with pREP4-Sfp and pQE70-TycC3-PCP plasmids was used. The cells were grown and purified according to the apo-TycC3-PCP protocol. The amount of ^{15}N-holo-PCP protein produced was determined by an HPLC method essentially as described previously (Mootz et al., 2002). A final content of holo-TycC3-PCP of nearly 100% was obtained*,* indicating that TycC3-PCP was efficiently phosphopantetheinylated by co-expressing Sfp in *E. coli.*

The expression and purification of the non-labelled Sfp and SrfTEII proteins was described previously (Mofid et al., 1999; Schwarzer et al., 2002). The

concentrations of the purified proteins were determined by UV-spectroscopy based on the calculated extinction coefficient at 280 nm.
Uniform labeling is achieved by replacing 2 g/L ^{12}C-glucose with 2 g/L ^{13}C–glucose for uniform ^{13}C labeling whereas for a ^{15}N labeled sample 1 g/L $^{15}NH_4Cl$ is added instead of $^{14}NH_4Cl$.

In-vitro* modification of PCP (priming assay).* The *in-vitro* modifications were carried out in 5 ml reaction mixtures in 50 mM phosphate buffer, pH 8.0 containing 20 mM $MgCl_2$. 0.25 mM ^{15}N-apo-TycC3-PCP, 1 mM Co A or acetyl-Co A and 25 µM Sfp were incubated at 30°C for 0.5h. The reaction mixture was applied to a Superdex75 16/60 gel filtration column. Those fractions containing the ^{15}N-holo-TycC3-PCP or ^{15}N-acetyl-holo-PCP were pooled and concentrated using Vivaspin devices (VivaScience) or Millipore devices and stored at -80°C.

Synthesis of Tyrosine-CoA. N-Boc- and side-chain tBu-protected Tyrosine was purchased from Novabiochem. Coupling reagents were purchased from IRIS biotech, coenzyme A (trilithium salt) from MP biomedicals, and all other compounds from Sigma-Aldrich.

The synthesis of the aminoacyl-CoA was performed according to the previously published protocol (Belshaw et al., 1999). A mixture containing the molar equivalents of protected Tyrosine (1.0), coenzyme A (1.5), PyBOP (1.5), and of potassium carbonate (4.0), dissolved in THF/water (1:1), was agitated for 2 h at room temperature. After removing the solvent, cleavage of the protecting groups at the acid-labile N-terminal and side-chain was performed by treatment of a mixture of trifluoroacetic acid (TFA), triisopropylsilane (TIPS), and water (95:2.5:2.5) for 3 h at room temperature. The deprotected aminoacyl-CoA was precipitated by addition of ice-cold diethyl ether and separated by centrifugation (4000 rpm, 4 °C) for 5 min. The aminoacyl-CoA pellet can be stored at –20 °C. For purification, the CoA-derivative was dissolved in 5 % acetonitrile/water (v/v) and applied to an Agilent preparative high-performance liquid chromatography (HPLC) system with a reversed-phase VP 250/21 Nucleodur 100-5 C18 ec column (Macherey and Nagel) using a gradient at a flow rate of 20 mL min^{-1} [buffer A, 0.1

% TFA/water (v/v); buffer B 0.1 % TFA/acetonitrile (v/v)]. The sample loading was performed at 5 % buffer B and kept constant for 5 min, afterwards a linear gradient up to 40 % buffer B within a 20 min time line, followed by a linear gradient to 95 % buffer B within 2 min, and hold 95 % buffer B for 3 min. The eluting compound was detected by UV absorption at 215 nm and automatically collected. Product containing fractions were pooled and lyophilized and the identity verified by MALDI-TOF mass spectrometry (calculated mass, 930.71 g mol^{-1}; observed mass, 930.81 g mol^{-1}). The thioester bind amino acid in the product was comfirmed by incubation with 0.1 M KOH to liberate the amino acid and analyzed by HPLC-MS versus amino acid standards.

NMR spectroscopy. Backbone and side chain assignments were based on ^{15}N-edited NOESY and TOCSY experiments collected using Bruker Avance NMR-spectrometer operating at proton frequencies of 700 and 800 MHz equipped with cryogenic 5mm z-axis gradient triple resonance probes at 298 K. DSS (4, 4-dimethyl-4-silapentane-1-sulfonate) was used as internal chemical shift reference. The TROSY-based ^{15}N-edited NOESY spectra for apo- and holo-TycC3-PCP were recorded with a mixing time of 80 ms and processed after zero-filling using squared sine-bell window function. The spectral width of both proton dimensions of these spectra was 10.8 ppm and 35 ppm for the ^{15}N-dimension. In addition homonuclear 2D-^{1}H·^{1}H-TOCSY and 2D-^{1}H·^{1}H-NOESY spectra after exchanging all labile protons with D_2O buffer were used in order to assign aromatic protons and to obtain distance constraints defining the hydrophobic core. Mixing times were 40 ms and 120 ms for the NOESY spectra and 70 ms for the TOCSY spectra. In each case a total of 2048 points in the direct and 1024 points in the indirect dimension were collected with spectra widths set to 12.4 ppm. The specificity of observable double peaks for TycC3-holo- and apo-PCP was determined by superimposing high resolution ^{1}H· ^{15}N-Transverse relaxation-optimized spectroscopy (Trosy)-hetero single quantum correlation (HSQC) spectra of all states, including a TycC3-PCP point mutation where Ser 45 was replaced by Ala (S45A-PCP) and TycC3-holo-PCP with acetyl-thioester on the covalently bound 4′-Phosphopantetheine cofactor

(TycC3-acetyl-holo-PCP). The superimposition of TycC3-holo- and apo-PCP displayed holo-PCP specific (H-state) and apo-PCP specific (A-state) double peaks and different areas of the supposed two-state-dynamic. The A-state and H-state specific signals were tested by superimposition with S45A-PCP and TycC3-acetyl-holo-PCP. The proposed structural dynamic should consequently result in the presence of two different NOE data sets of the corresponding residues, and the analysis of the NOE patterns indeed could confirm that assumption. So A-state and H-state specific signals could be also assigned separately in ^{15}N-TOCSY-HSQC and ^{15}N-NOESY-HSQC spectra.

The mentioned NMR experiments for assignment of resonances are already described and the assignment was performed according to standard structure determination procedures. All data were collected using Bruker NMR-spectrometer (Avance800 MHz, Avance700MHz or DMX600 MHz) equipped with a cryogenic 5mm z-axis gradient triple resonance ^{1}H [^{13}C,^{15}N]- probe at 298 K. DSS was used as an internal chemical shift reference. The TROSY-based ^{15}N-edited NOESY spectra for TycC3-holo- and TycC3-apo-PCP were recorded with a mixing time 70 ms and processed after zero-filling using squared sine-bell window function. The spectral widths for acquisition of these spectra were kept constant to 6613.74 Hz for ^{1}H-dimensions ($\omega2$ $\omega3$) and 1277.14 Hz for ^{15}N-dimension ($\omega1$). A 2D-^{1}H·^{1}H-TOCSY and two 2D-^{1}H·^{1}H-NOESY spectra of D_2O exchanged samples of TycC3-apo-PCP and TycC3-holo-PCP was used in order to assign aromatic Protons and to derive distance constraints defining the hydrophobic core. For each sample a NOESY spectra with mixing time of 40 ms and 120 ms was recorded to separate intraresiduale from interresiduale NOEs.

D_2O exchange measurements. The H_2O/D_2O buffer exchange experiments of uniformly ^{15}N-labeled apo-TycC3-PCP, holo-TycC3-PCP and apo-TycC3-PCP$_{S45A}$ were performed using 0.5 mM samples of each protein at 298 K. All data were collected using a Bruker Avance 800 MHz NMR-spectrometer. For detecting the solvent accessibility and exchange of backbone amides, a reference spectrum of each sample was recorded first in 50 mM sodium phosphate buffer (pH 6.8) supplied with 5 % D_2O and with 0.15 mM DSS as the internal chemical shift standard. After reference data acquisition the sample buffer was exchanged into D_2O buffer (50 mM sodium phosphate (previously D_2O exchanged and dried), pD 7.2) by concentrating it to approximately 100 µl by ultrafiltration (Millipore concentrator, 1 kDa MWCO) followed by loading the sample on a protein desalting spin column (PIERCE biotech) pre-equilibrated in D_2O buffer. The collected sample was subsequently diluted to a final volume of 500 µl. Each [^{15}N, ^{1}H]-TROSY-HSQC spectrum was measured within 20 min. with a spectral width in the ^{1}H and ^{15}N dimensions set to 10.8 ppm (512 points) and 35 ppm (256 points), respectively. Each experimental series lasted for 10 hours with an additional experiment after 16 h and 24 h. The processing of all spectra and the peak integration was performed using MestreC 4.3.8 (Cobas & Martin-Pastor, 2004; Cobas et al., 2004; Cobas & Sardina, 2003) [MESTRELAB RESEARCH - Rúa Xosé Pasín 6 - 5C, Santiago de Compostela, A CORUÑA; SPAIN - CP: 15706]. For the analysis of the exchange rates both the peak integrals and the peak heights were used providing virtually identical results.

Titration experiments. Interaction between the different PCP forms and their interaction partners were investigated by NMR titration experiments. For this purpose 0.5 mM solutions of the individual PCP samples were stepwise mixed with unlabeled protein samples (Sfp, SrfTEII and SrfTEII_{S86A}, for details see results section). Chemical shift changes were monitored by ^{15}N-TROSY-HSQC experiments, measured with spectral widths of 35 ppm (512 points) and 10.8 ppm (1024 points) in the ^{15}N and ^{1}H dimensions, respectively. Between 8 and 10 individual titration steps were performed starting from a concentration of 1:20, after recording the basis experiment of the pure PCP-sample, and ending at a concentration of 2:1 or 5:1 for strongly and weakly interacting complexes, respectively.

Analysis of residual dipolar couplings (RDCs). The alignment of uniformly ^{15}N labeled holo-TycC3-PCP was performed using a diluted solution of Pf1 phage

(Profos, Regensburg, Germany) (Hansen et al., 1998; Zweckstetter & Bax, 2001) added to the 0.75 mM protein sample in 50 mM sodium phosphate, pH 6.8, 10% D_2O, and 200 µM complete-EDTA solution (Complete protease inhibitor, Roche). The concentration of Pf1 phage for alignment was calibrated by 1-D ^2H spectrum on the basis of ^2H quadrupole splitting of DOH (Hansen et al., 1998; Zweckstetter & Bax, 2001). All measurements were performed at 298K. The splitting was determined to 10.2 Hz at a Bruker Avance800 MHz spectrometer with 18.8 *T* of magnetic field strength. The ^1H-^{15}N dipolar coupling values were measured using Mestrec 4.3.8 (Cobas & Sardina, 2003) and analyzed by ReDCAT V7 for Windows (Valafar, 2003). All experimental values were compared to predicted values based on the calculated structures (Tjandra, 1999).

The increasing interest in residual dipolar couplings (RDC) for solution structural determination of macromolecules is caused by the unique combination of angle and r^{-3} distance dependence information. Residual dipolar couplings derive from small degree of alignment of molecules in the static magnetic field B_0 which is cancelled to zero in isotropic solutions. These describe the orientating of an intermolecular vector dependent from the alignment tensor in a laboratory coordinate frame.

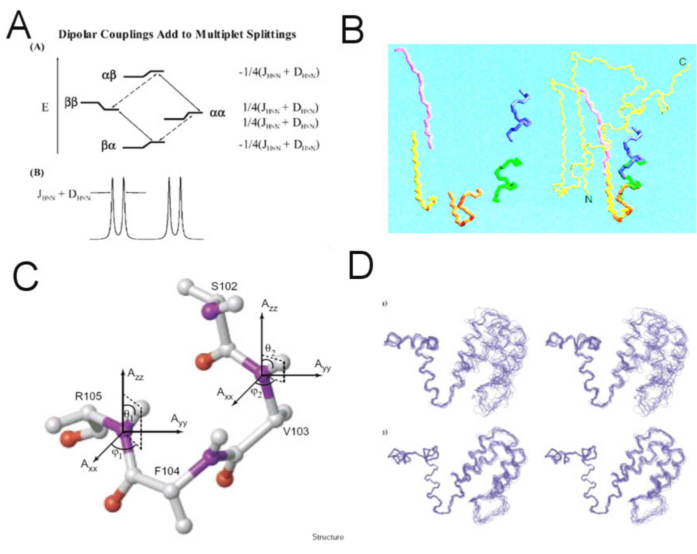


Fig. 5: Residual dipolar coupling. The effect of weakly aligned proteins on dipolar couplings to multiplet splitting (A) is used for fragment replacement to define global folds (B) or to refine the structure of macromolecules (D). In C the effect of weakly aligned proteins and the tensor composition is symbolized by two NH-vectors.

Figure 5 demonstrates the effect of protein alignment on dipolar couplings (Prestegard, 2004) and symbolizes the usage for fragment replacement (Bax, 2003) or structural refinement (Tjandra, 1999) by relative orientation of the dipolar coupling vector.

This coupled information for relative orientation of bond vectors and long-range structural information improves structures determination by simulated annealing protocols or even enables determining protein backbone structures. The alignment of the protein sample and the corresponding alignment tensor is effected by paramagnetic moieties of the sample or by diluted liquid crystal media. Nowadays some media are available for aligning samples in a liquid crystal state: bicelle-based liquid crystals made by dimyristol-phosphotidylcholine (DMPC) and dihexanoyl-phosphotidylcholine (DHPC) or media like cetylpyridinium, polyacrylamide gels, purple membranes, Polydimethylsiloxan (PDMS), PEG/Hexanol, Polystyrol, Polyglutamates, 4-n-Pentyl-4'-cyanbiphenyl (PCBP), orientating filamentous phages (PF1, fd) and rod-shaped viruses (TMV), also Gummibärchen (Haribo, Bonn) were successfully used as a protein alignment media (Kessler, 2005). This experiment was used for structural validation of AH- and H-state of TycC3-holo-PCP and to proof the structural changes within the two-state dynamic of this domain. The alignment of uniformly ^{15}N labeled TycC3-holo-PCP in the magnetic field B_0 was performed using a diluted solution of Pf1 phage (Profos, Regensburg, Germany) added to the 0.75 mM protein sample in 50 mM NaPi, 10% D_2O, 200 µM complete-EDTA protease inhibitor at pH 6.8. The concentration of Pf1 phage for alignment was calibrated by 1-D ^{2}H spectrum on the basis of ^{2}H quadrupol splitting of DOH. All measurements were performed at 298K. The splitting was determined to 10.2 Hz at a Bruker AV800 MHz spectrometer with 18.8 *Te*. The recorded 2D TROSY-type spectra with no ^{1}H decoupling of ^{15}N·^{1}H TycC3-holo-PCP in isotropic solution and in anisotropic aligned media to determine nonzero ^{1}H-^{15}N dipolar couplings were collected. Distance and angle information for Pf1 aligned macromolecules are derived from dipolar couplings between the nuclei *i* and *j* calculated by

$$DC_{ij}(\theta,\phi) = -\frac{\mu_0}{4\pi} y_i y_j \hbar S \frac{\left[A_a (3\cos^2\theta - 1) + \frac{3}{2} A_r \sin^2\theta \cos 2\phi \right]}{4\pi^2 r_{ij}^3}$$

Where μ_0 is the magnetic permittivity in vacuum, γ_i and γ_j are the gyromagnetic ratios of nuclei I and J, $\hbar$ as the Planck constant, S is a degree of disorder for internal motions of the inter-nuclear vector between i and j with the distance r_{ij} and θ and ϕ are the angle of the inter-nuclear vector r_{ij} and its projection to the *xy*-plane. With A_a as the unit less axially symmetric component and A_r the rhombic component of the molecular alignment vector. A_a and A_r are calculated with the Eigenvalues of the alignment tensor (A_T)

$$A_T = \begin{pmatrix} A_{zz} & A_{zx} & A_{zy} \\ A_{xz} & A_{xx} & A_{xy} \\ A_{yz} & A_{yx} & A_{yy} \end{pmatrix}$$

This equation can be simplified by replacing all experiment specific constant parameters by the introduced terms D^a and D^r.

$$DC_{ij} = D^a * \left[\left(3\cos^2\theta - 1\right) + \frac{3}{2} D^r \sin^2\theta \cos 2\phi \right]$$

Where D^a is the magnitude of the dipolar coupling tensor, which is normalized to the dipolar interaction between the nuclei i and j, and D^r is the rhombicity.

$$D_a = \frac{3}{4} \Omega A_{zz}$$

$$D_r = \frac{2}{3} \frac{\left(A_{xx} - A_{yy}\right)}{A_{zz}}$$

Where Ω is the maximum of dipolar coupling and is only modulated by the angle θ of the inter-spin vector r_{ij} to the static magnetic field B_0 and the length of r_{ij} as the dynamically averaged fixed distance between the coupling nuclei *i* and *j*.

$$\Omega = \frac{-\mu_0 \left(\frac{\hbar}{2\pi} \right) \gamma_i \gamma_j}{\left(4\pi^2 r_{ij}{}^3 \right)}$$

$$\Omega = -\frac{\mu_0 \hbar \gamma_i \gamma_j}{\left(8\pi^3 r_{ij}{}^3 \right)}$$

Based on the powder pattern from an experimental set of dipolar couplings, the magnitude D_a^{ij} and the rhombicity *R* where estimated. The largest absolute value of all bond vectors $2D_a^{ij}$ provides

$$S_{zz} = \frac{2D_{a^{ij}}}{\Omega}$$

and *R*

$$\frac{3}{2} R = \frac{\left(S_{xx} - S_{yy} \right)}{S_{zz}} = -\left[1 + \left(\frac{D_{opp}}{D_{a^{ij}}} \right) \right]$$

D_{opp} symbolize the other extreme value from measured dipolar couplings.

For protein alignments using Pf1 phage magnetic alignment and the more dominant electrostatic alignment caused by negative surface charge of phage particles (~ 0.5 *e*/nm^2) must be considered. So prediction of magnitude and orientation of the alignment vector based on molecular properties only will not work as long as electrostatic forces are not enclosed to calculations. Residual dipolar couplings are measured by determine the difference in splitting between the aligned, as the sum of scalar *J*-coupling and dipole-dipole coupling, and the isotropic sample which display *J*-coupling only.

$$DC_{ij} = \left(J_{ij} + DC_{ij}\right) - J_{ij}$$

Caused by the negative gyromagnetic ratio of ^{15}N, the measured coupling within one spectrum is negative.

Spectra analysis. NMR spectra were processed with Bruker XWinNMR 2.6 and analysed using Sparky Version 3.110 (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco). Molecular graphics images were produced using the UCSF Chimera (Pettersen, E.F. et al., 2004; Sanner, M.F. et al., 1996) beta version 1 build 2065 package from the Computer Graphics Laboratory, University of California, San Francisco (supported by NIH P41 RR-01081).

Assignment of the double peaks to individual forms of the PCP were based on comparison of the chemical shifts with either TycC3-PCP$_{S45A}$ showing exclusively resonances of the A-state or acety-holo-TycC3-PCP showing exclusively resonances of the H-state.

Structural calculation. For structural calculations, NOE-based distance restraints from 3D ^{15}N-NOESY-HSQC and D_2O-exchanged 2D-homonuclear-NOESY spectra were extracted using the CANDID module of CYANA 1.2 (Herrmann et al., 2002; Guntert, 2004). For each structural calculation by a simulated annealing protocol a set of approx. 950 unambiguous, non-redundant proton-proton upper distance restraints was used. These upper distance limits were derived from approx. 430 intraresidual, 480 sequential, 560 mid-range and 590 long-range NOEs with 120 mid-range and 190 long-range NOEs with exactly one possible assignment (Goddard & Kneller, 2004). Approx. 30 hydrogen bonds according to the three stable helices deduced from the D_2O exchange experiment were included as distance restraints. The final calculated structure ensemble of 20 Dyana generated conformers (Guntert et al., 1997) was structurally compared to the Aria 2.0 derived and energy minimized average fold (Nilges & O'Donoghue, 1998; Linge et al., 2003). As a proof for the structural differences of the individual

TycC3-PCP subforms, each structure was back-calculated using Aria 2.0 (Linge et al., 2001; Linge et al., 2003) based on CNS 1.1-simulated annealing protocol (Brünger et al., 1998). In Aria 2.0 each structural calculation and water refinement started with 100 random structures for each of 8 cycles and resulted in a set of 10 water refined structures with a precision of root mean square deviation (RMSD) < 0.9 Å for backbone resonances and without distance violations larger than 0.25 Å. Structures calculated with both protocols were virtually identical, showing that the data are interpreted identically by both automatic NOE assignment routines. Structural statistics for all three structures are shown in Fig. 26 in the result section. Analysis of the backbone dihedral angles revealed that in the A-state 93.2% are located in the most favoured and additionally allowed regions with no dihedral angles in the disallowed region. For the A/H-state the same analysis showed that 93.2% of dihedral angles are located in the most favoured and additionally allowed, 5.3% in the generously allowed and 1.5% in the disallowed regions. In the H-state structure 87.9% are located in the most favoured and additionally allowed regions and 3.4% are located in the disallowed region. Analysis of the dihedral angles was performed with the program package PROCHECK (Laskowski & MacArthur, 1993).

Complex modeling. The model of the complex between apo-TycC3-PCP and Sfp was calculated using the program HADDOCK1.3 (Dominguez et al., 2003) within the CNS1.1 software package (Brünger et al., 1998) based on the mutational analysis of Sfp (Reuter et al., 1999; Mofid et al., 2004) and the titration experiments of TycC3-holo- and apo-PCP as mentioned above. From the originally created 1000 structures, 200 were selected for the final iteration step followed by water refinement of 50 structures The RMSD of the finally saved 50 structures of the complex is 2 Å. This value is small for a non-covalent protein-complex and supports the quality of the calculation. Coenzyme A as the substrate for the PPTase Sfp was energy minimized in the protein fold of Sfp using AutoDock XXX based on the observed electron density of the adenyl phosphate of Co A in the published crystal structure of Sfp (Reuter et al., 1999). The substrate relaxed in a stretched and stable conformation inside Sfp with the pyrophosphate residue of Co A coordinated by the essential Mg^{2+} ion of Sfp (see results section) . The position of the thiol-group of the pantheteine part of Co A is remarkable due to its exposed position and totally free outside on the protein interface.

Investigation of hydrogen bonds. For the identification of hydrogen bonds a long range HNCO experiment (Cordier & Grzesiek, 1999) with a transfer time of 133 ms was used. Unfortunately, the low quality of the resulting spectra allowed to obtain information only about the A-state of the protein. Identification of hydrogen bonds in the H-state were based on the analysis of the structure, the results of the D_2O exchange experiments as well as on a 2ns molecular dynamics simulation (Gromacs 3.2 (Berendsen et al., 1995; Lindahl et al. 2001), using the Gromos96 (van Gunsteren & Berendsen, 1990; van Gunsteren et al., 1995) force field in an OPLS water box.

Molecular imaging. Molecular graphics images were produced using the UCSF Chimera beta version 1 build 2065 package from the Computer Graphics Laboratory, University of California, San Francisco (supported by NIH P41 RR-01081).

4 Results

Conformational diversity of apo- and holo-TycC3-PCP. Detailed analysis of TROSY-HSQC spectra of ^{15}N-labeled apo-TycC3-PCP and of holo-TycC3-PCP containing the 4'-PP cofactor indicated that several backbone resonances show two distinct chemical shifts. The superimposition of the complete spectra from TycC3-apo-PCP and TycC3-holo-PCP reveals this differences between the two samples (Fig. 6).

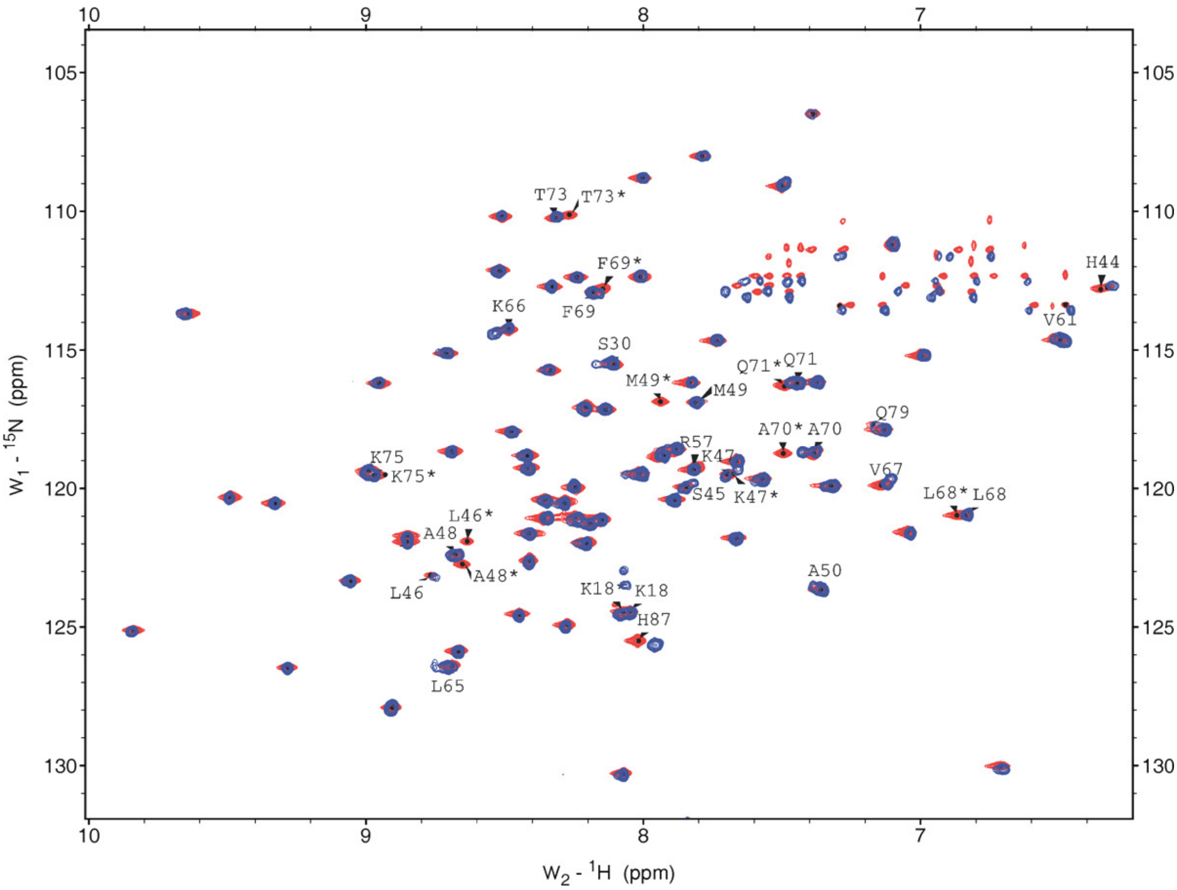


Fig. 6: Trosy-HSQC-spectra of TycC3-apo-PCP and TycC3-holo-PCP. The superimposition of the ^{15}N-Trosy-HSQC spectra of holo-PCP, shown in blue, and apo-PCP (red) demonstrates the specific double peaks of each form and imagine the specific and different dynamics of apo-PCP and holo-PCP. The amino acid residues that are involved in this slow chemical exchange are indicated in the spectra.

In a more detailed view onto this peak doubling clarifies that this slow chemical exchange rates are most prominent for apo-TycC3-PCP involving a total of 20 amino acid residues of which at least 8 show equal intensities (L46, K47, A48, M49, L68, F69, Q71, T73) (Fig. 7) and for holo-PCP separately.

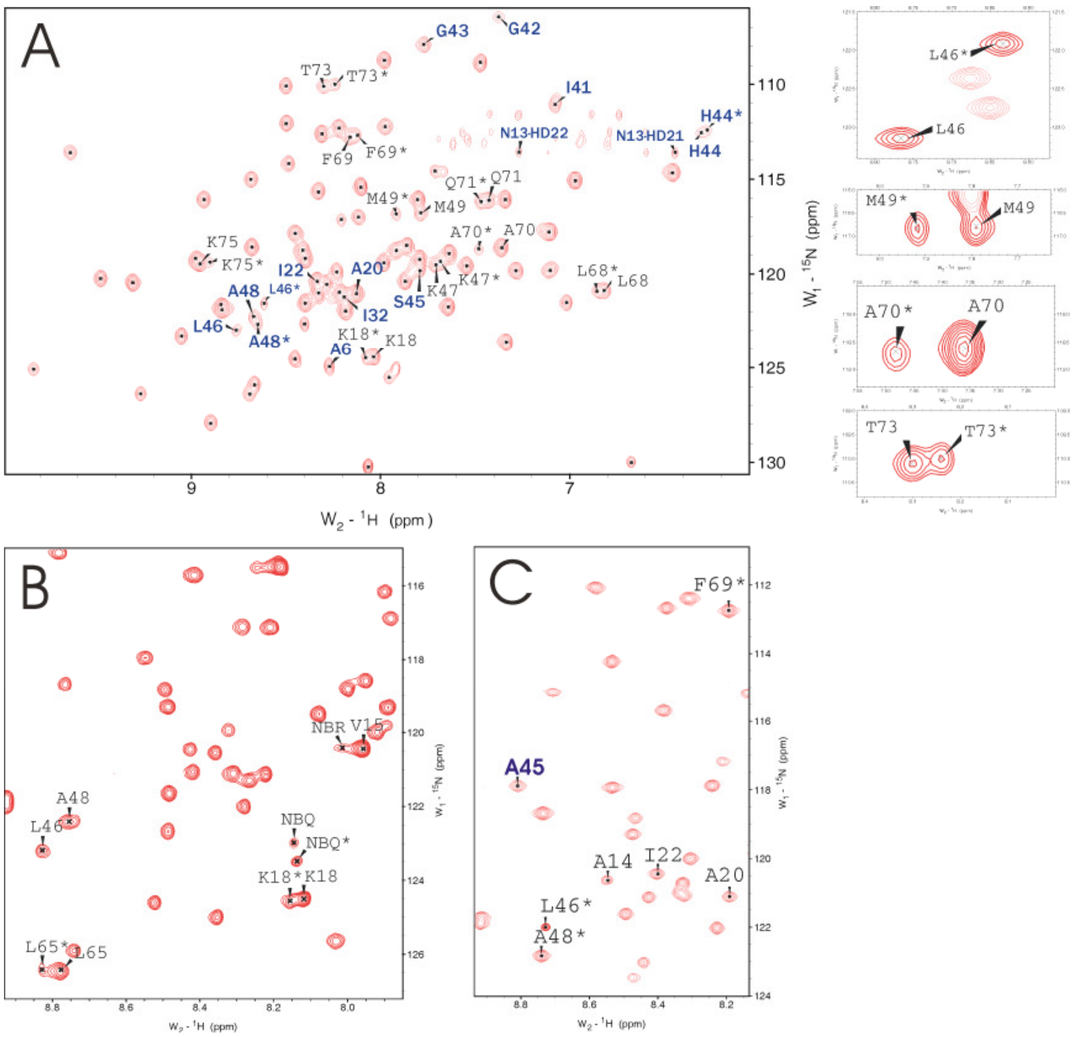


Fig. 7: . Refined assignment and conformational exchange of apo-TycC3-PCP. TROSY-HSQC spectra of different forms of TycC3-PCP. The spectra were recorded on a Bruker Avance 800 MHz or DMX600 MHz NMR-spectrometer equipped with a cryogenic probe at 298 K. A, apo-TycC3-PCP with the refined backbone amide proton assignment according to Weber (PDB accession code: 1dny). Newly defined and refined assignments are given in bold. Residues showing conformational exchanges observable by the formation of double peaks are indicated by asterisks which mark the signals specific for the A-state of apo-TycC3-PCP. The enlarged areas on the right show four examples. B, holo-TycC3-PCP. Amide proton signals of the 4'-PP cofactor and of selected PCP residues are indicated. Signals specific for the H-state of holo-TycC3-PCP are marked. C, apo-TycC3-PCP$_{S45A}$. The A45 signal and residues showing double peaks in the apo-TycC3-PCP spectrum (A) are indicated. Only signals specific for the A-state (marked by asterisks) are visible. (Koglin *et al. Science,* 2006, modified)

The residues in apo-TycC3-PCP showing double peaks are mainly clustered in two regions (Fig 8). The first region is located at the beginning of helix αII comprising residues H44 to A53 and including the active site residue S45 that represents the site of modification with the 4'-PP-cofactor. The second region spans residue R57 to T73 and includes most residues between helices αII and αIV. In the three-dimensional structure of TycC3-PCP, these two regions are directly adjacent to each other.

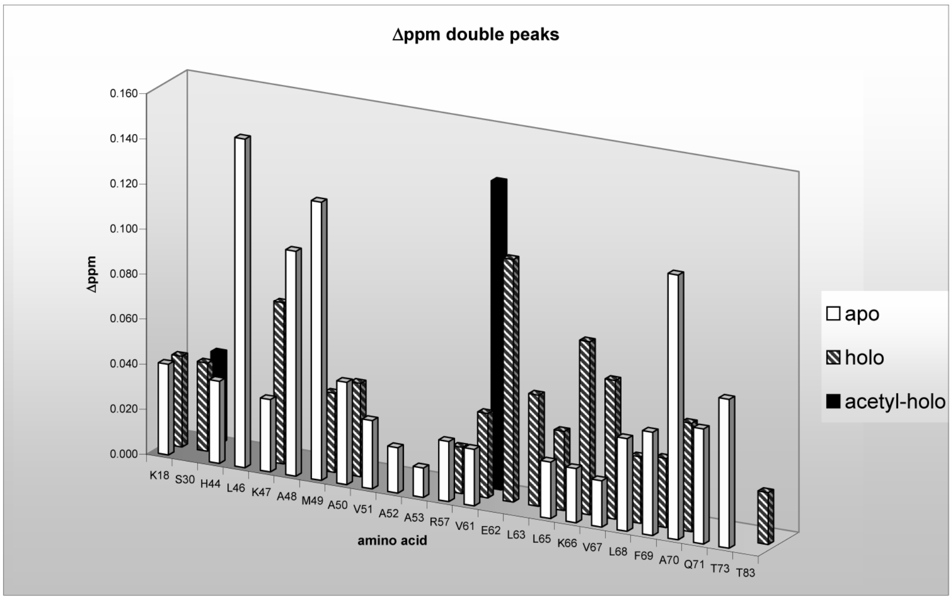


Fig. 8: Differences in chemical shift between double peaks observed in the individual PCP forms. The combined chemical shift differences between the observed two signals of the individual residues are given in Δppm. The differences were calculated from TROSY-HSQC spectra of the proteins taken at 800 MHz and 298 K. White bars: apo-TycC3-PCP; striped bars: holo-TycC3-PCP; dark bars: acetyl-holo-TycC3-PCP. The single double peak in the spectra of acetyl-holo-PCP seem to be effected by rotational freedom of the charge of the acetyl-group, which effects the Val 61 residue only, due to the position of the 4'-P-cofactor. (Koglin *et al. Science,* 2006, modified)

Upon conversion to the holo form by modification with the 4'-PP-cofactor, the pattern of peak doubling of the TycC3-PCP domain clearly changes (Fig. 7, 8). Most strikingly is the complete disappearance of many double peaks of the N-terminal region surrounding residue S45 including H44, L46, A48, V51, A52 and A53. The chemical shift differences of the other double peaks are considerably reduced with the exception of residue K47 that shows an increased chemical shift difference (Fig. 8). In contrast, many residues of the second region showing double peaks display increased chemical shift differences and the entire region is shifted towards the N-terminus, including the C-terminal part of helix αII.

Structural investigation of the individual TycC3-PCP forms. The described differences observed in the ^{15}N-TROSY-HSQC spectra of the individual TycC3-PCP forms most likely indicate the presence of an equilibrium of two distinct conformations that we designate A-state/A*-state for apo-TycC3-PCP and H-state/H*-state for holo-TycC3-PCP. In order to further structurally characterize these different conformations, we reassigned holo-TycC3-PCP and newly assigned apo-TycC3-PCP by multi-dimensional heteronuclear NMR spectroscopy. The availability of high field NMR spectrometers equipped with cryogenic probes allowed us to complete the previously published assignment of holo-TycC3-PCP (Weber et al., 2000) and to newly assign the active site residue Ser 45 in addition to the residues Gly 42, His 44, Leu 46 and Ala 48 that together form a substantial part of the 4'-PP cofactor binding site (Fig. 7).

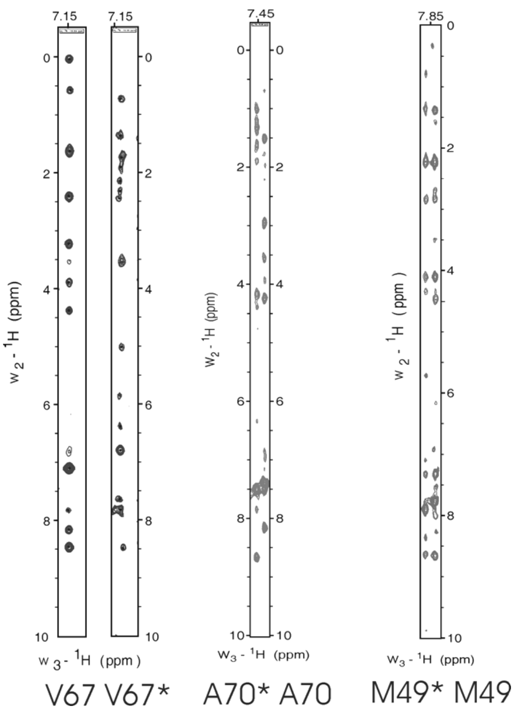


Fig. 9: Strips taken from the 3-D ^{15}N-edited NOESY-HSQC spectra along the indirect proton dimension at the amide proton frequency of the indicated residues. The strips corresponding to V67 and A70 are taken from the spectrum of holo-TycC3-PCP and the M49 strip from the apo-TycC3-PCP spectrum. Asterisks indicate H-state and A-state resonances. (Koglin *et al. Science,* 2006, modified)

Structure of the apo-TycC3-PCP specific A-state. The existences of different conformations for apo- and holo-PCP were further confirmed by the presence of distinct NOE patterns for the individual subforms (Fig. 9).

Structure determination followed standard protocols and was based on 932 and 851 unambiguously assigned non-redundant upper distance constraints for the A-state and the A*-state of the apo form, respectively. The structures of the H-state and the H*-state of the holo-form are based on 774 and 831 distance constraints, respectively.

Detailed comparison of the calculated structures indicated characteristic structural differences for the distinct subforms of apo-TycC3-PCP and holo-TycC3-PCP (Fig. 10).

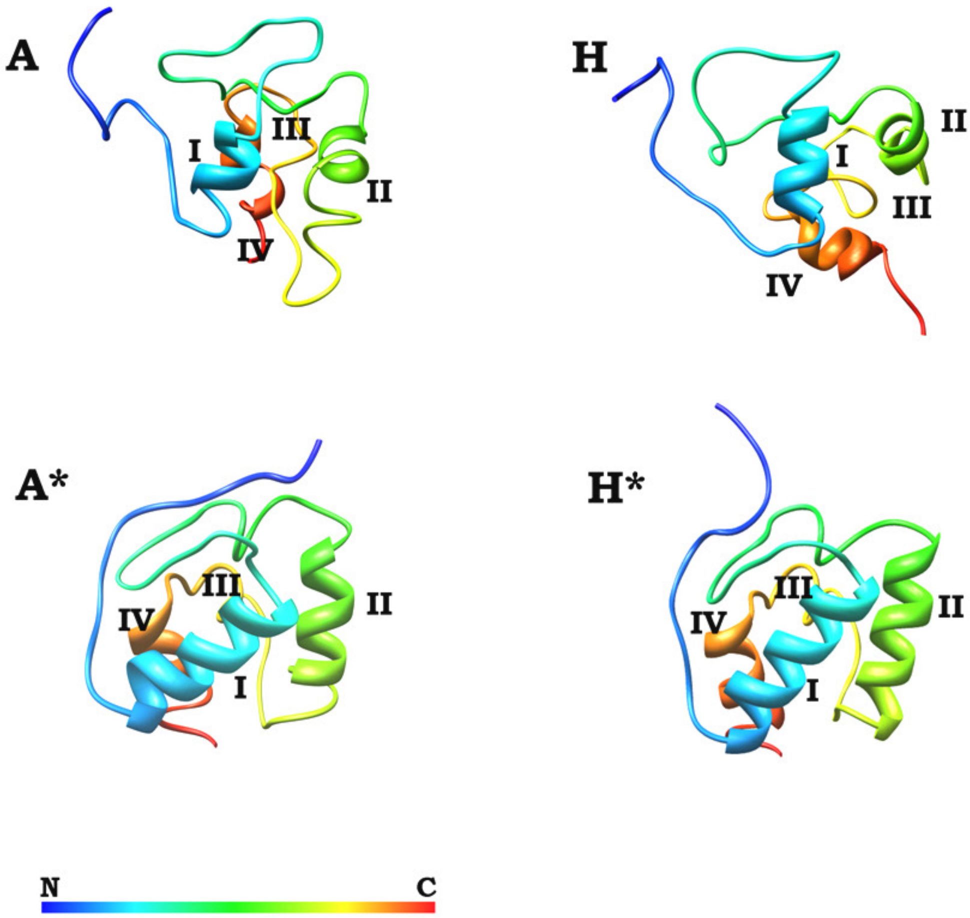


Fig. 10: Structural comparison of PCPs. The comparison of the calculated structures of the two isoforms of TycC3-apo-PCP (A and A*) and TycC3-holo-PCP (H and H*) reveals the apo-specific fold (A) based on apo-specific NOEs and the holo-specific fold (H) of the holo-specific NOEs respectively. The structures based on the NOEs of the signals give an identical fold (A* and H*) and is called the A/H-fold. The calculated structures are shown as ribbon diagrams of the protein backbone and colored from blue (N-terminus) to red (C-terminus) as shown in the symbolic diagram below.

A-state PCP generally shows the most flexible and extended three-dimensional structure having very large loop regions (Fig. 11). The N-terminus is highly flexible and mainly fixed by only one H-bond between A10 and E16 to the loop area between helix αI and αII (as determined by long range HNCO measurements (Cordier & Grzesiek, 1999)). Helix αIII basically does not exist and instead forms a large loop (loop III). Helix αI (V15 to A20) is shorter as compared to the A*-state and reduced to only 1.5 turns. The helices αII (M49 to A53) and αIV (A76 to Y80) are reduced to only two α-helical turns and helix αII shows a kink at position Q54. Loop III thus becomes stretched and embedded in the core of the protein between helices αII and αIV. The position of this stretched loop III is stabilized by H-bonds between G43 and L68 and between H44 and F69.

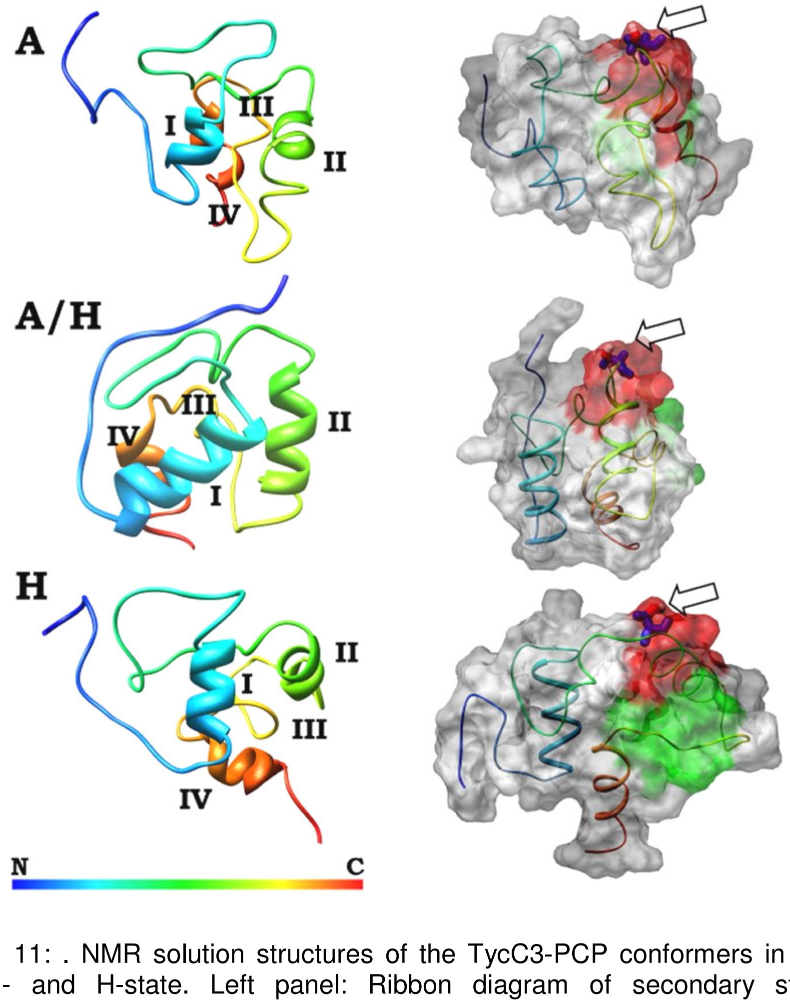


Fig. 11: . NMR solution structures of the TycC3-PCP conformers in the A-, A/H- and H-state. Left panel: Ribbon diagram of secondary structure elements. Right panel: Space filling models. The molecules are rotated by 90 degrees compared to the molecules in the left panel. The side chain of the active site residue S45 is shown. The highly conserved region surrounding the residue S45 is coloured in red. Surface exposed residues of helix/loop region III are shown in green. (Koglin *et al. Science,* 2006, modified)

Interestingly, a 2ns molecular dynamics simulations using the Gromos96 force field (van Gunsteren & Berendsen, 1990; van Gunsteren et al., 1995) suggests that this fixation of loop III allows the formation of an H-bond between the hydroxyl group of S45 and R47-backbone. The equivalent simulation of the A*-state PCP showed that these H-bonds are eliminated in favour of a new one that is located between S45-OH and the side chain of H44 (Nδ1). The active site residue S45, therefore, seems to play a crucial role for the observed conformational diversity of apo-TycC3-PCP. Accordingly, mutating the active site serine to alanine is sufficient to abolish any structural heterogeneity in apo-TycC3-PCP and the protein represents a completely "frozen" A-state (Fig. 7).

Structure of the intermediate A/H-state of TycC3-PCP. Besides the presence of the 4'-PP cofactor, we did not observe differences in the structures of A*-state and H*-state and we therefore finally designated this intermediate form of TycC3-PCP as the A/H-state (Fig. 11). This state is most similar to the previously published structure of TycC3-PCP (Weber et al., 2000). The N-terminal end of the protein is fixed by an H-bond from A6 to G31 (identified by a long-range HNCO experiment). Additional hydrogen bonds involving the N-terminus including E7 to I32 as well as A10 to E16 are suggested based on the structure calculation and are supported by molecular dynamics simulations (Berendsen etal., 1995; Lindal et al., 2001). The orientation of the helices are very similar to that of apo-TycC3-PCP. However, the helices are longer and more stable. Helix αI is extended from A14 to V26, helix αII from K47 to Q54 and helix αIV from I74 to Y80. The core of the A/H-state is built by the helices αI and αIV and is stabilized by H-bonds between E16 and K75 and from I34 to T73 (determined by molecular dynamic simulations and supported by D_2O exchange experiments). The loop between helices αI and αII is fixed to helix αIV by H-bonds and contains the active site residue Ser 45 oriented towards the N-terminal end. The loop III now forms a very short and non-stable one-turn helix III (L65 to A70) that is located together with the loop between helix αII and αIV (V55 to T73) outside of the globular protein core. This relocation of the loop III area is one of the most striking differences to the A-state.

D_2O-exchange experiments. These structural differences between the A and the A/H forms are further supported by amide proton exchange experiments (Fig. 12).

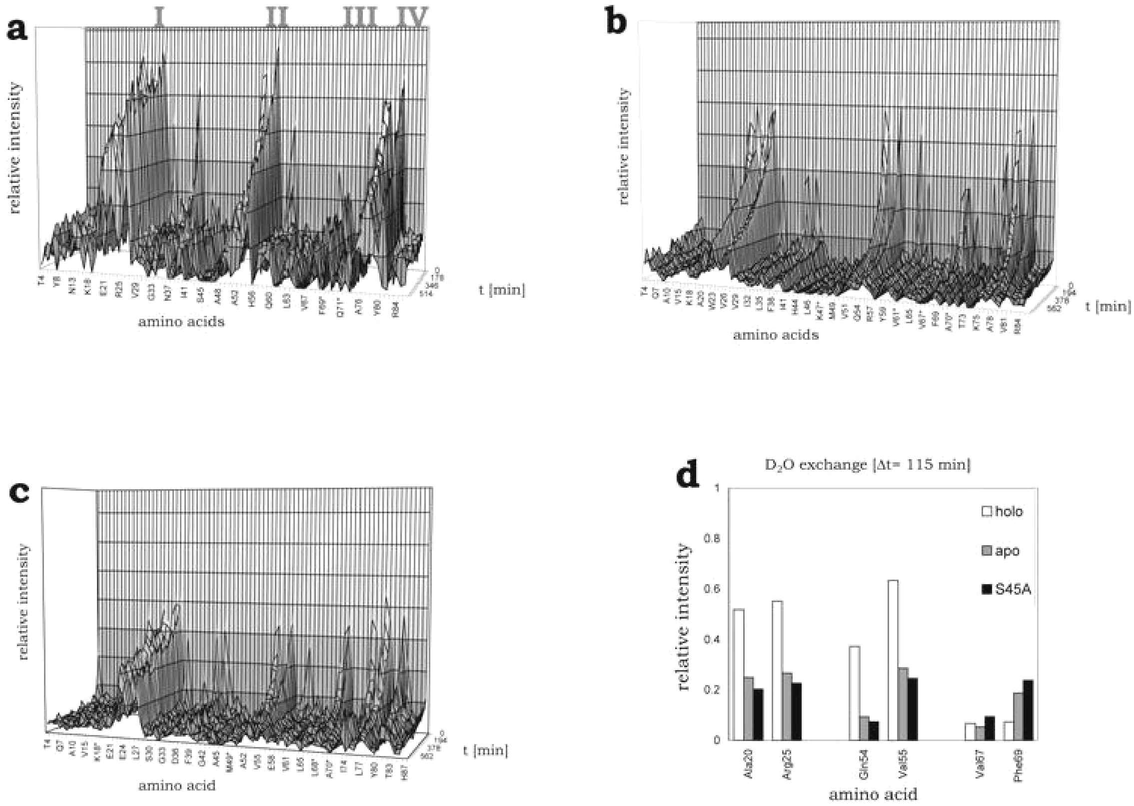


Fig. 12: D_2O exchange measurements of different forms of TycC3-PCP. The D_2O buffer exchange curves for each individual backbone amide are shown. Processing and integration parameters were kept constant for all [^{15}N, ^{1}H]-TROSY-HSQC spectra. For clarity, the positions of the helices αI, αII, αIII and αIV are indicated. a, holo-TycC3-PCP; b, apo-TycC3-PCP; c, apo-TycC3-PCP_{S45A}; d, D_2O exchange rates of selected residues of different TycC3-PCP forms. The relative amide proton peak intensity of selected residues measured after 115 min. is shown. A20/R25 and Q54/V55 are residues from the helices αI and αII, respectively, and F69, in contrast to V67, is located in the loop III stretch that is buried in the core of TycC3-PCP in the A-state. (Koglin *et al. Science,* 2006)

In contrast to the relatively stable helices αI, αII and αIV, the amide protons of helix III are almost completely invisible even in the earliest spectra, suggesting that the backbone amides are highly solvent accessible. Furthermore, the D_20 exchange of the mutant protein apo-TycC3-PCP_{S45A} supports our observation that turn III in the A-state is buried in the core of the protein and thus protected from the solvent by the helices αII and αIV. Apo-TycC3-PCP$_{S45A}$ does not show any conformational exchange and represents the A-state. As a consequence, the region from K66 to F69 forming the core of the loop III area is exchanging slower than the neighbouring amides as compared to the exchange rates of apo-TycC3-PCP, where still an equilibrium between A-state and A/H-state exists.

Structure of the holo-TycC3-PCP specific H-state. The major difference between the holo-TycC3-PCP specific H-state and the intermediate A/H-state is the orientation of helix αIV that is moved parallel to helix αI by approx. 3 Å (Fig. 11). As a consequence the single turn helix III unravels and becomes extended. This transition induces a dislocation of helix αII and the region connecting helices αI and αII containing the active site residue Ser 45 with the attached 4'-PP cofactor. In D_2O exchange experiments, the slowly exchanging hydrogen bonds within the three helices αI, αII and αIV of holo-TycC3-PCP are easily detectable (Fig. 12) while the solvent exposed loop III region is not visible. In addition, the hydrogen bonds between A6 and G31 as well as E7 and I32, which stabilize the N-terminus near helix αI reduce the solvent exchange rate of the corresponding amide protons and make them observable in the exchange experiments.

Selective interaction of TycC3-PCP conformers with Sfp and SrfTEII. The complex formation of TycC3-PCP with other proteins was analysed by NMR titration assays. ^{15}N-labeled TycC3-PCP was stepwise titrated with unlabeled potential interaction partners up to a final ratio of 1:2 (Fig. 13). Molecular interactions become visible by chemical shift differences of specific amino acid residues that are involved or affected by the complex formation (Zuiderweg, 2002). The 4'-PP transferase Sfp does not interact with holo-TycC3-PCP (data not shown). A titration with apo-TycC3-PCP in presence of CoA/Mg^{2+} immediately resulted in the conversion into holo-TycC3-PCP and thus no interaction could be monitored.

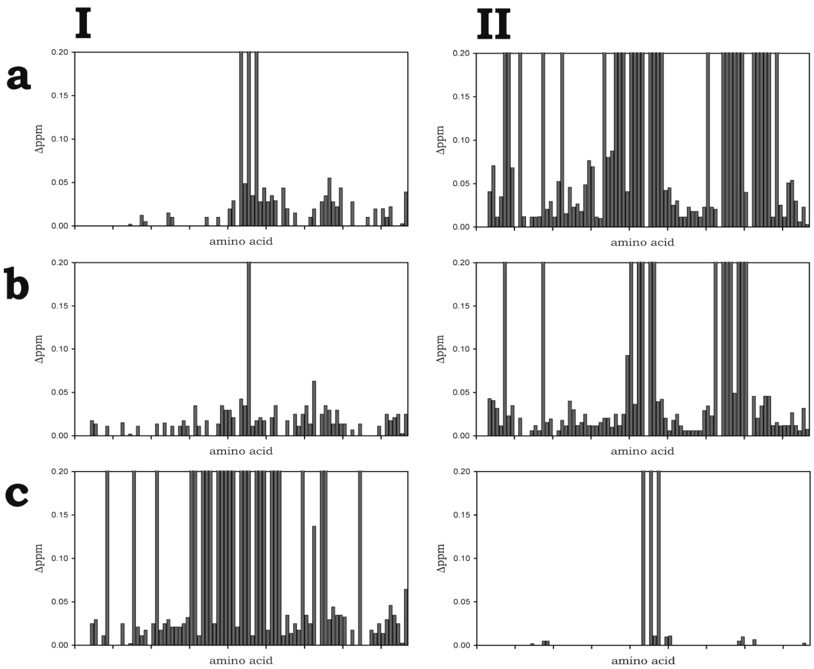


Fig. 13: Specific interaction of TycC3-PCP conformers with binding partners. Histogram of the chemical shift differences in NMR-Titration assays. I, Interaction with Sfp. A, ^{15}N-apo-PCP; B, ^{15}N-apo-PCP $_{S45A}$; ^{15}N-apo-PCP$_{S45A}$ with CoA and Mg^{2+}. II, Interaction with SrfTEII. A, ^{15}N-acetyl-holo-PCP with SrfTEII$_{S86A}$.; B, ^{15}N-holo-PCP; C, ^{15}N-apo-PCP. (Koglin *et al. Science,* 2006)

The titration of apo-TycC3-PCP in the absence of CoA/Mg^{2+} revealed a weak interaction primarily with a few residues that surround the active site residue Ser 45 (Fig. 13Ia). In order to prevent the rapid conversion to holo-PCP by Sfp in presence of CoA, we analyzed the interaction with the active site mutant TycC3-PCP$_{S45A}$. Again, only a weak interaction was monitored in absence of CoA (Fig. 13Ib). In contrast, a strong interaction with Sfp is visible in the presence of CoA/Mg^{2+} including the same residues previously identified in the titration experiments with wild type TycC3-PCP in the absence of CoA/Mg^{2+}. These include the amino acid residues 30 to 55 of helix αII and the 4'-PP recognition site within the loop between helix αI and αII (Fig. 13Ic). This result indicates that Sfp specifically recognizes the A-state TycC3-PCP as this is the only conformer detectable with the mutant TycC3-PCP$_{S45A}$ and that CoA is important for the complex formation.

The thioesterase SrfTEII is responsible for the regeneration of misprimed holo-PCP domains (Schwarzer et al., 2002). As expected, only a weak interaction is observable with apo-TycC3-PCP (Fig. 13IIc). Based on the number and magnitude of the observed chemical shifts the interaction is considerably enhanced with holo-TycC3-PCP and involves mainly the active site motif and the complete helix/loop III (Fig. 13IIb). Even more enhanced chemical shift differences are observable with the misprimed acetyl-holo-TycC3-PCP derivative, indicating that the acetyl moiety plays a major role in the recognition mechanism (Fig. 13IIa). For this titration, we used the active site mutant SrfTEII$_{S86A}$ containing a defective catalytic triad (Linne et al., 2004) in order to prevent the rapid hydrolysis of the acetyl residue. The interaction of SrfTEII$_{S86A}$ with acetyl-holo-TycC3-PCP was identical to that of wild-type SrfTEII with holo-TycC3-PCP suggesting that the point mutation did not affect the complex formation (see 'Additional' section about the external Thioesterase from Surfactin synthetase).

The titration experiments are shown as color indicated structures, from no interaction (white), very weak interaction (yellow) to strong interactions (red) to easier describe the region specific interaction and the formation of the interaction site by the structural diversity of TycC3-holo-PCP (Fig. 14).

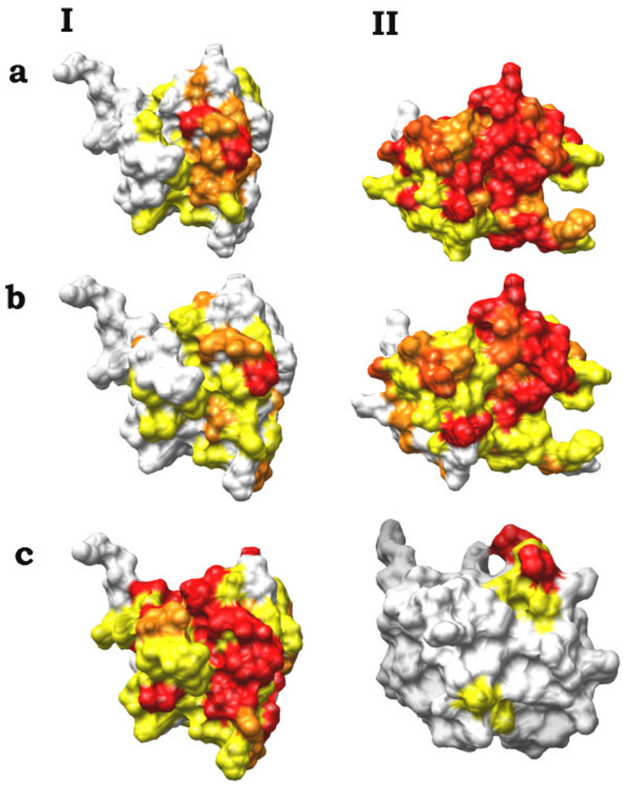


Fig. 14: Result of the titration experiment (Fig. 13) as a graphical presentation of holo- and apo-PCP with Sfp (I) and SrfTEII (II). (Koglin *et al. Science,* 2006)

The graphical presentation of the TycC3-PCP titration with Sfp and SrfTEII characterize the formation of a closed interaction site on the surface of PCP. In Fig. 14I the titration of apo-TycC3-PCP with Sfp (a), of S45A-PCP with Sfp (b) and of S45A-PCP with Sfp loaded with CoA and Mg^{2+} (c). In all three structures the A-state of TycC3-PCP depicts the formation of a closed interaction site, after Sfp-activation by loading CoA and Mg^{2+}. In Fig. 14II the NMR-titration experiment of acetyl-holo-TycC3-PCP with $SrfTEII_{S86A}$ (a), holo-TycC3-PCP with the wildtype thioesterase SrfTEII (b) and of apo-TycC3-PCP in the A/H-state with SrfTEII (c) is shown.

An H-state specific selection by SrfTEII is evident from (I) almost no interaction with apo-TycC3-PCP that also forms the A/H-state, (II) gradually reduced A/H-state specific signals in TROSY-HSQC spectra with increased SrfTEII concentrations (data not shown) and by line shape analysis (Mittag et al., 2003; Mittag et al., 2004) of double peak signals from selected residues that clearly demonstrate the shift of the preformed conformer equilibrium of holo-TycC3-PCP in favour towards the H-state upon titration with SrfTEII (Fig. 15).

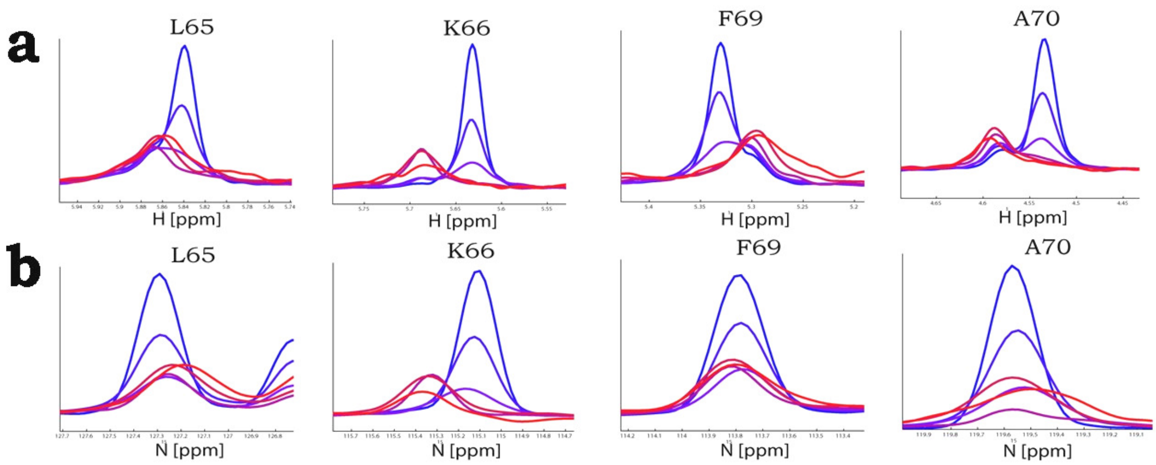


Fig. 15: Line shape analysis of selected residues of the ^{15}N-holo-TycC3-PCP / SrfTEII interaction. One-dimensional view of N-H signals of selected holo-TycC3-PCP residues showing double peaks in TROSY-HSQC spectra. The larger peak represents the A/H-state specific signal whereas the smaller shoulder in the proton dimension corresponds to the H-state specific signal. The titration proceeds from free ^{15}N-holo-TycC3-PCP (blue) to a final ^{15}N-holo-TycC3-PCP / SrfTEII ratio of 1:1 (red). a, proton dimension; b, nitrogen dimension. The changes in the equilibrium between the A/H-state and the H-state is in dynamic exchange dependent on interaction with the appropriate interaction partner and can be moved back. (Koglin *et al. Science,* 2006; Software: *NMRlab* Gunther *et al.,* 2000)

Key recognition elements in H-state TycC3-PCP for SrfTEII are (I) the terminal acetyl moiety, (II) helix/loop αIII, (III) the active site sequence motif, and (VI) potentially the 4'-PP cofactor. As a result of the molecular dynamic simulation it can be suggested that in the H-state of holo-TycC3-PCP the N-terminus is removed from its close proximity to the active site residues Ser 45, thus eliminating a putative steric hindrance of the SrfTEII interaction with the A/H state. A change in the surface charge in the area between helices III and αIV caused by the relocation of the strong negatively charged helix αIV could also be important for a specific recognition process. This results in transforming a cluster of negative charges into a positively charged area that forms a large positive belt along the whole surface of H-state holo-TycC3-PCP (Fig. 16).

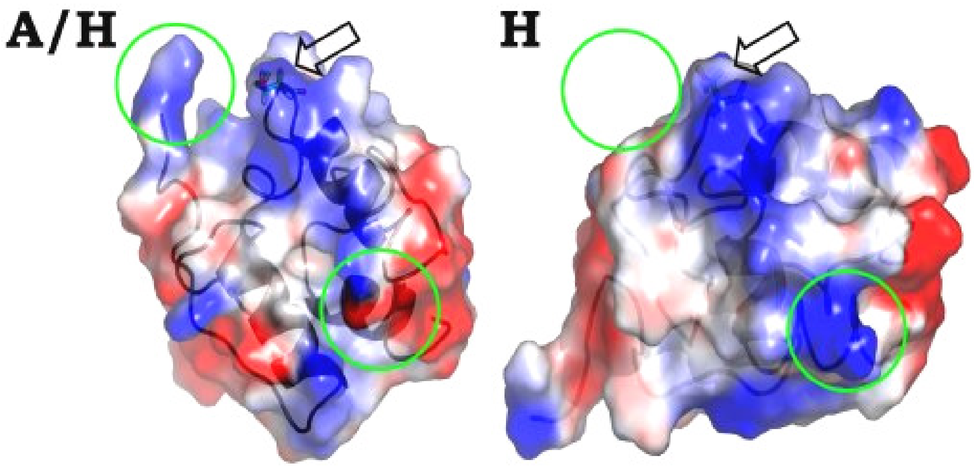


Fig. 16: Surface charge and morphology of A/H-state and H-state holo-TycC3-PCP. Space-filling model of holo-TycC3-PCP in the A/H-state (a) and H-state (b). Circles enclose the position of the N-terminal end and the area showing a change from negative to positive charges caused by a relative repositioning of helix αII compared to the scaffolding helix αI in the H-state. The arrow indicates the position of the active site residue Ser 45. (Koglin *et al. Science,* 2006)

Sfp/apo-TycC3-PCP complex modelling. The availability of a crystal structure of Sfp (Reuter et al., 1999) allowed modelling of its complex with the A-state conformer of TycC3-PCP. The complex formation of activated Sfp and A-state TycC3-PCP is primarily based on NMR-titration experiments of ^{15}N-labeled apo-TycC3-PCP and unlabeled Sfp (Zuiderweg E.R.P., 2002) and the mutation analysis of Sfp (Mofid et al., 2004). A-state PCP fits perfectly on top of CoA, which is located in a large groove in the surface of Sfp where the essential Mg^{2+} ion coordinates the pyrophosphate (Fig. 17a, b). It should be noted that the reactive terminal thiol group of CoA is easily accessible in the complex and is available for a variety of substrates. The interface of the complex is composed of helix αII including the loop to helix αI of A-state apo-TycC3-PCP and parts of helices αH2 and αH6, the β-strands B4 and B6 and the loop region between αH6 and αH7 of Sfp (Fig. 17c). The reactive hydroxyl group of Ser 45 in A-state PCP is brought into very close contact with the pyrophosphate group of CoA, putatively enabling an efficient hydrolysis and 4′-PP transfer as described as the direct Mg^{2+} coordinated phosphate hydrolysis (Fig. 17d).

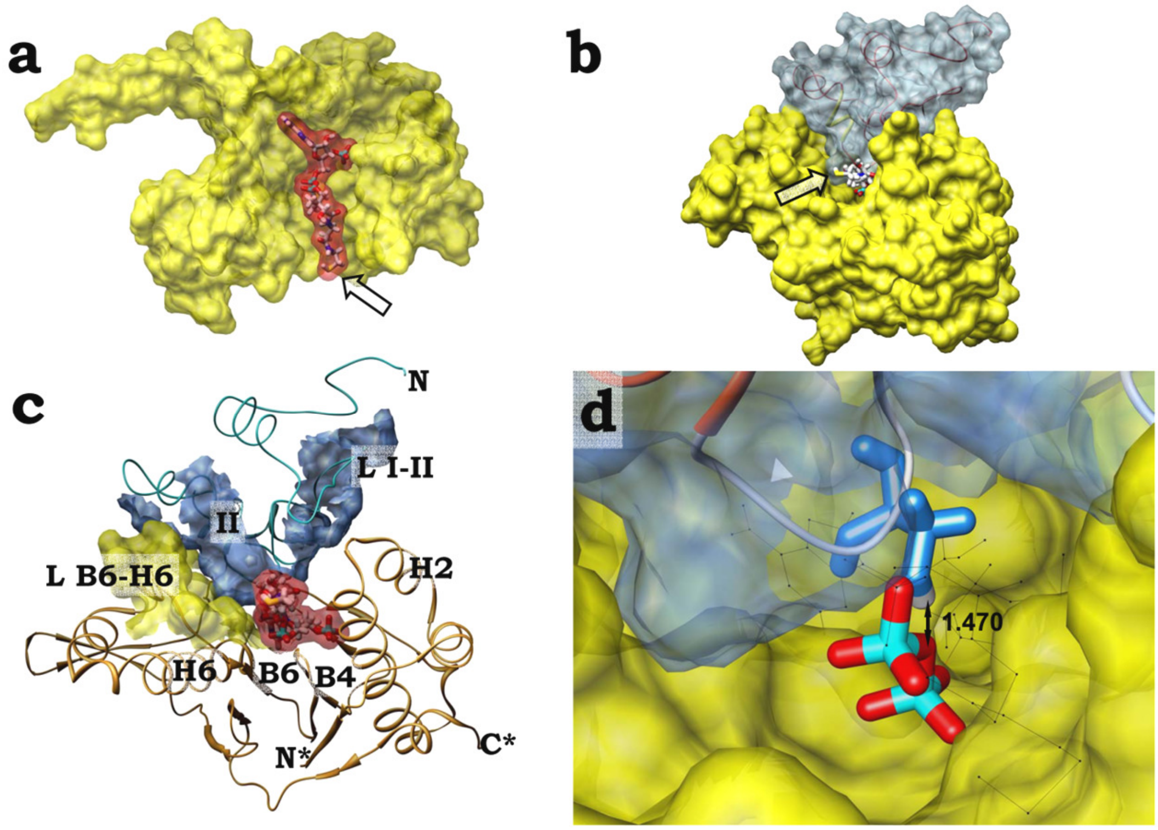


Fig. 17: . Model of the A-state TycC3-PCP/Sfp complex. The Sfp protein is shown in yellow, TycC3-PCP in blue and CoA in red. The arrow indicates the position of the reactive thiol group of CoA. (a) Proposed position of CoA in Sfp based on energy minimization routines from AutoDock 3. (b) Space filling model of the Sfp/H-state TycC3-PCP complex derived from complex modeling using Haddock 1.3 together with cns 1.1. (c) Secondary structural elements of Sfp and TycC3-PCP involved in the complex interface. Helices of TycC3-PCP are indicated with roman numbers while latin numbers indicate involved elements of the Sfp protein. (d) Insight the active complex of Sfp (yellow), apo-TycC3-PCP (blue) with the essential Ser 45 residue shown and colored blue and the pyrophosphate group of Coenzyme A (colored by the elements oxygen (red) and phosphorus (cyan)). The coordinating Mg^{2+} is not shown. (Koglin *et al. Science,* 2006, modified)

The structural heterogeneity of holo-TycC3-PCP causes defined movements of the 4'-PP cofactor. Structure calculations suggested that the transitions between the H- and A/H-states of holo-TycC3-PCP results in a major movement of the 4'-PP cofactor over a large part of the protein surface. In order to define these putative and different states of the cofactor the identification of the chemical shifts of its two amide protons and nitrogens, NBR and NBQ would help (Fig. 18).

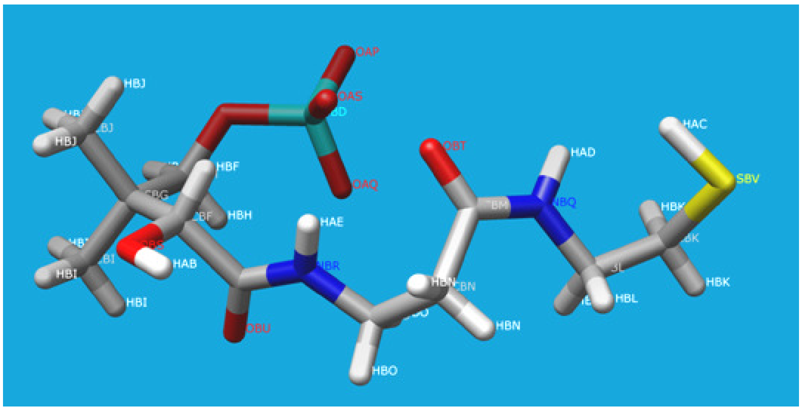


Fig. 18: Graphical presentation of the 4'-phosphopantheteine cofactor of PCPs. The cofactor is attached to the hydroxyl-group of Ser 45 via a phosphoester and the thiol-group is the active site of the mature holo-PCP. All atoms are labeled corresponding to the IUPAC nomenclature.

This identification was based on comparing a TROSY-HSQC spectrum recorded on a sample of 3mM CoA at the natural abundance of ^{15}N with a TROSY-HSQC spectrum of fully ^{15}N-labeled holo-TycC3-PCP, containing a ^{15}N-labeled cofactor (Fig. 19).

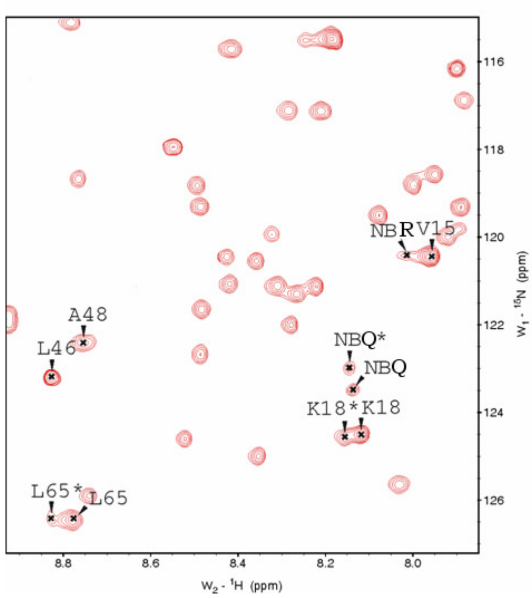


Fig. 19: Identification of the two amides of the 4'-PP cofactor. The thiol-group oriented NBQ is showing a clear double peak, while the phosphoester oriented amide NBR is showing broad lines of a less defined single peak. (Koglin *et al.* *Science,* 2006, modified)

The NBQ amide proximal to the thiol-group was detected as a well resolved double peak, whereas the distal amide NBR showed a single very broad signal. The observed double peak formation indicates the localization of the 4'-PP cofactor in two clearly distinct positions in holo-TycC3-PCP. In order to further identify these two locations, a highly resolved ^{15}N-NOESY-HSQC spectrum of completely ^{15}N labelled holo-TycC3-PCP was recorded with an exceptional long mixing time. The two signals belonging to the 4'-PP amide NBQ exhibited distinct NOE pattern, either showing NOEs to the amino acids L68, F69 and probably also L65, or to V26 and probably E24 (Fig. 20).

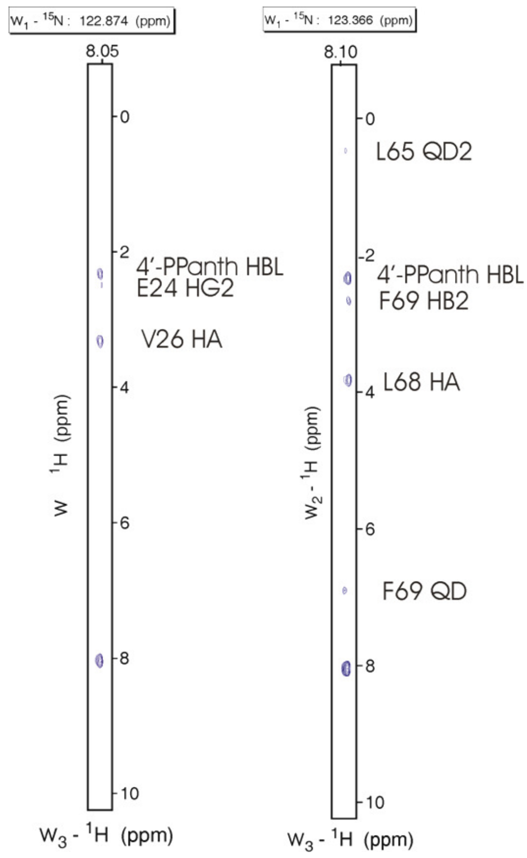


Fig. 20: Observed NOE-pattern of the thiol-group oriented amide double peak of the 4'-PP-cofactor NBQ and NBQ*, which describe distinct positions of the 4'-PP cofactor on surface of TycC3-holo-PCP. The assigned NOEs are mentioned beside the NOE-stripes. (Koglin *et al. Science,* 2006)

Modelling of holo-TycC3-PCP combined with the analysis of these NOE pattern revealed that in the A/H-state the cofactor is located close to the N-terminus, while in the H-state it is located in the vicinity to the C-terminal end (Fig. 21). This result indicates a large rotation of the 4'-PP cofactor by approx. 100 degree resulting in a 16 Å relocation of the thiol group induced by this two-state conformational equilibrium of holo-TycC3-PCP.

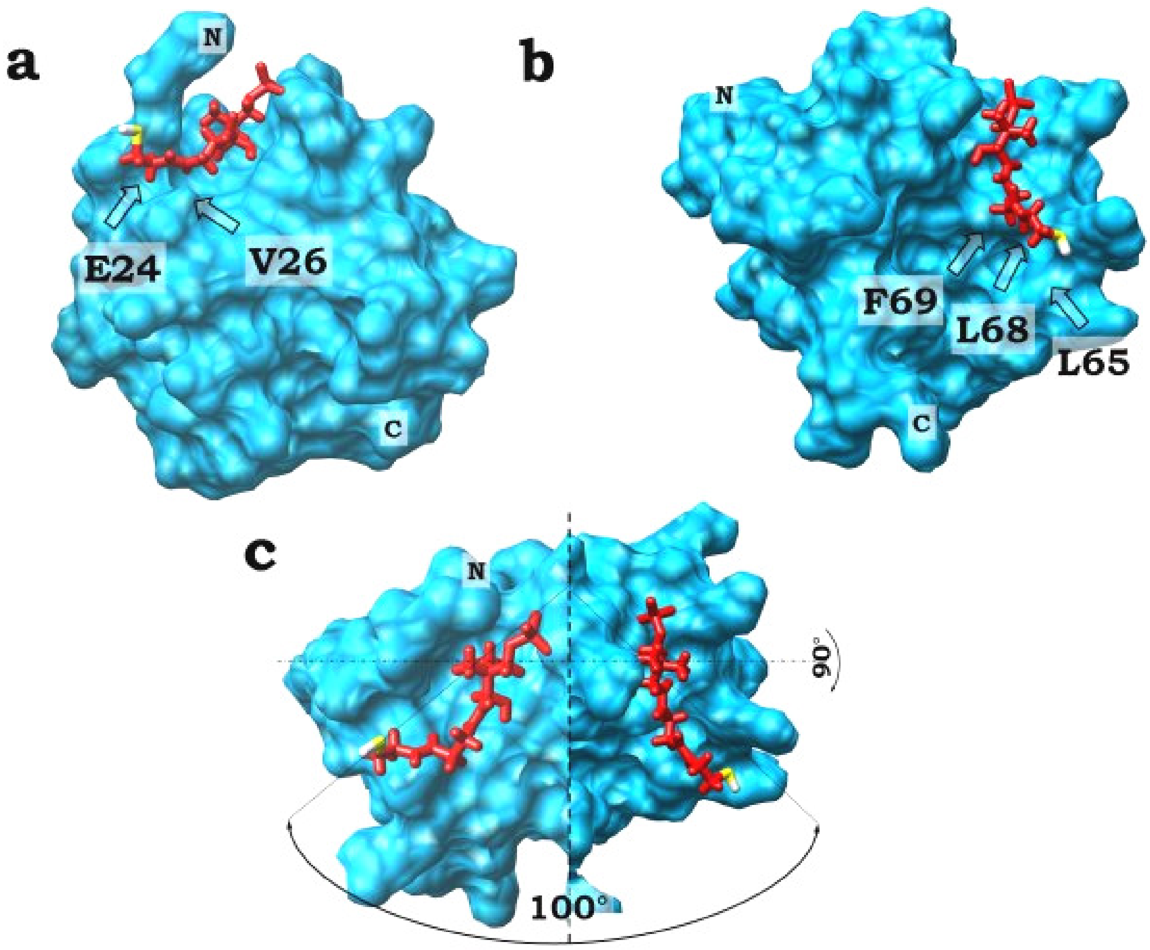


Fig. 21: Directed repositioning of the 4'-PP cofactor on the surface of holo-TycC3-PCP. a, A/H-state of holo-TycC3-PCP with the 4'-PP cofactor located close to residue V26 in the vicinity of the N-terminal end; b, H-state of holo-TycC3-PCP showing the 4'-PP cofactor located in the vicinity of residue F69 at the other site of the protein. The position of the protein is rotated by 180° as compared to a; c, rotation of the protein by 90° as compared with a and b and comparison of the two different locations of the 4'-PP cofactor. The degree of movement of the 4'-PP arm is indicated. N, N-terminus; C, C-terminus. Identified residues with contacts to the 4'-PP cofactor are indicated. (Koglin *et al. Science,* 2006, modified)

Residual Dipolar Coupling (RDC). The investigation of RDCs was used for supporting the fold difference between the A/H-state and the H-state of TycC3-holo-PCP. The distorsion of helix aII is large enough to estimate the relative position of this helix to the general fold of TycC3-holo-PCP in A/H-state and H-state by comparing of spectrally determined RDCs to the calculated RDCs based on the estimated alignment and the calculated structures. The alignment of uniformly ^{15}N labeled holo-TycC3-PCP was performed using a diluted solution of Pf1 phage (Profos, Regensburg, Germany) (Hansen et al., 1998; Zweckstetter & Bax, 2001) added to the 0.75 mM protein sample in 50 mM sodium phosphate, pH 6.8, 10% D_2O, and 200 µM complete-EDTA (protease inhibitor, Roche). The concentration of Pf1 phage for alignment was calibrated by 1-D ^2H spectrum on the basis of ^2H quadrupole splitting of DOH (Hansen et al., 1998; Zweckstetter & Bax, 2001). All measurements were performed at 298K. The splitting was determined to 10.2 Hz at a Bruker Avance800 MHz spectrometer with 18.8 *T* of magnetic field strength. The measured RDCs of both states of holo-TycC3-PCP (Fig. 22a) and the resulting pake-pattern (Fig. 22b) were used to estimate the alignment tensor for comparison to the structure-based RDCs.

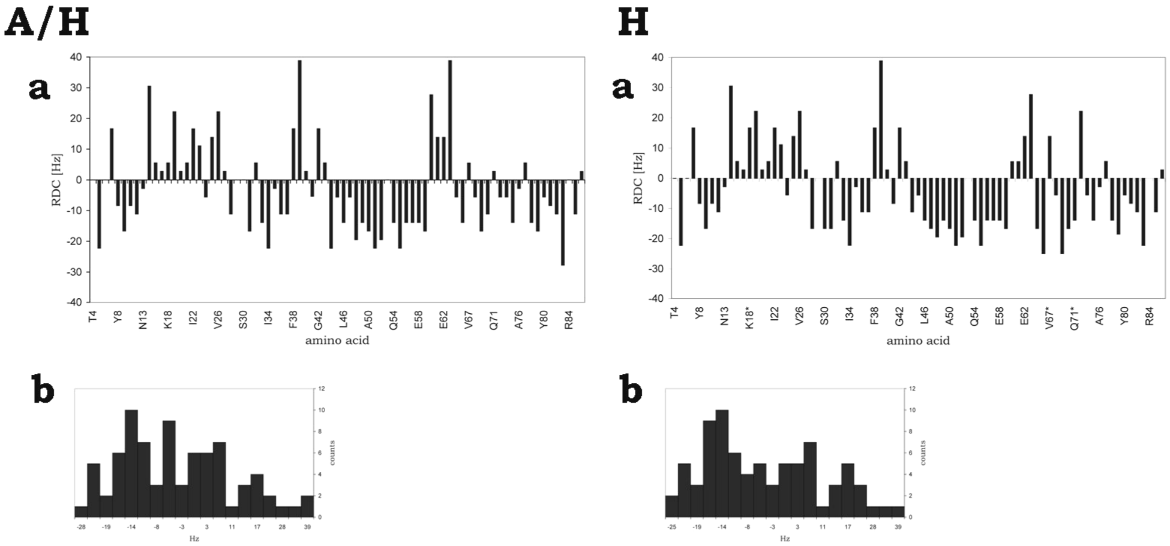


Fig. 22: RDC analysis of holo-TycC3-PCP. (a) Residual dipolar coupling analysis of an anisotropic sample of the 15N-labeled protein aligned by Pf1-phage shown for A/H-state and H-state separately. The differences of coupling rates between both states indicate local structural changes by comparable global orientation. (b) The corresponding pake pattern are support the local reorientation of a part of the protein.

Covalent modification of the 4'-PP cofactor. The covalent modification of the 4'-PP cofactor of holo-TycC3-PCP should support the positioning of the cofactor on the protein surface. For the modification the aromatic amino acid residue tyrosine was used, to detect the effect of the aromatic ring current on the protein by recording ^{15}N-Trosy-HSQC spectra. After loading the unlabeled modified cofactor to ^{15}N-labeled apo-TycC3-PCP and recording a Trosy-HSQC spectra, the H-state of TycC3-PCP was detectable only. In Fig. 23 the superimposition of the ^{15}N-Trosy-HSQC spectra of holo-TycC3-PCP with the typical double peaks and Trosy-HSQC of TYR-holo-TycC3-PCP is shown with the H-state specific double peaks only.

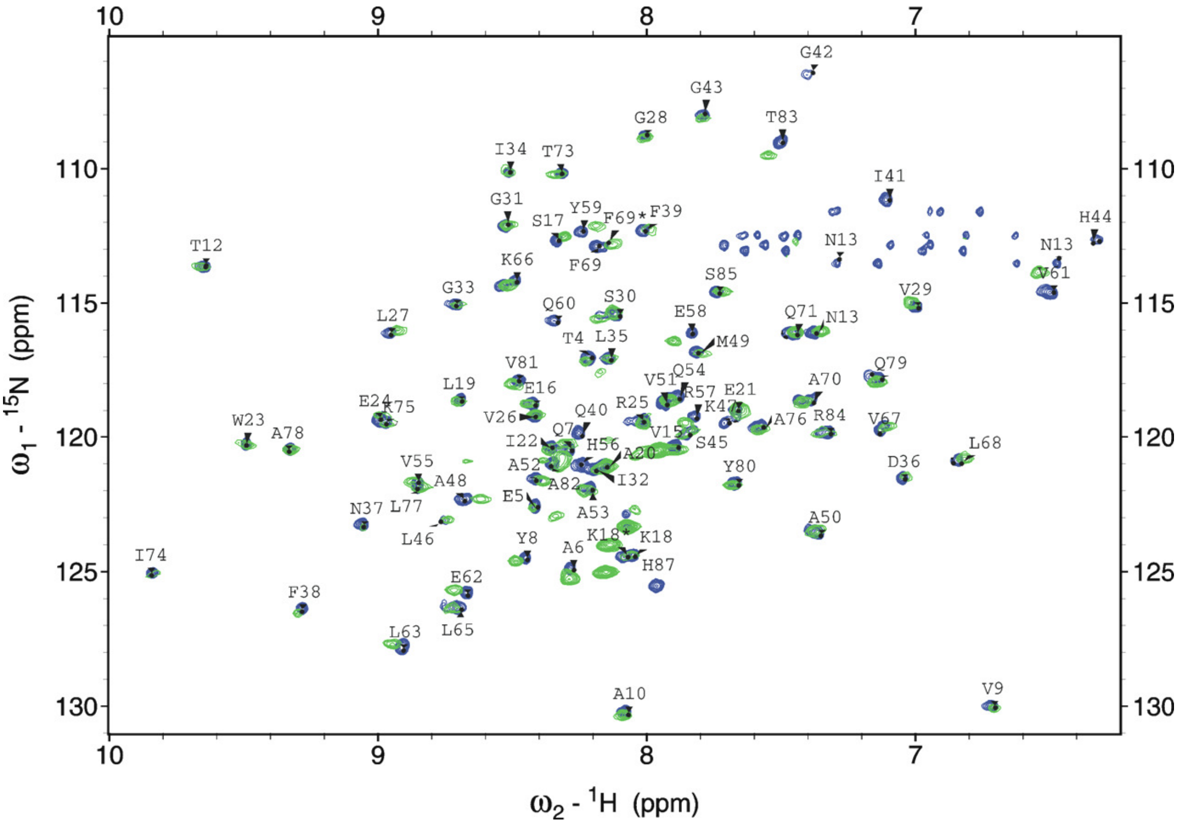


Fig. 23: Superimposition of the ^{15}N-Trosy-HSQC spectra of holo-TycC3-PCP with the typical double peaks (blue) and Trosy-HSQC of TYR-holo-TycC3-PCP with the H-state specific double peaks only (green) is shown. The assignment of TycC3-PCP is fully shown and indicate the additionally changed resonances in TYR-holo-TycC3-PCP.

Interestingly, TYR-holo-TycC3-PCP shows the H-state only, like acetyl-holo-PCP, a modification of the 4'-PP cofactor moves the dynamic system of holo-PCP into the 'frozen' state with the cofactor oriented to the C-terminus, similar to the putative interaction state with the C-terminal Condensation-domains of a NRPS module. Some amides in the Trosy-HSQC show a distinct change in chemical shifts compared to the holo-PCP. These amino acid residues are all structurally clustered near the C-terminus of the protein. The implementation of this information on the surface of the H-state of PCP reveals the following image (Fig. 24), where the chemical sift changes are colored on the surface and the 4'-PP cofactor is shown in its expected position for the H-state of TycC3-PCP and the chemical shift changes are effected by the ring current of the aromatic sidechain of tyrosine.

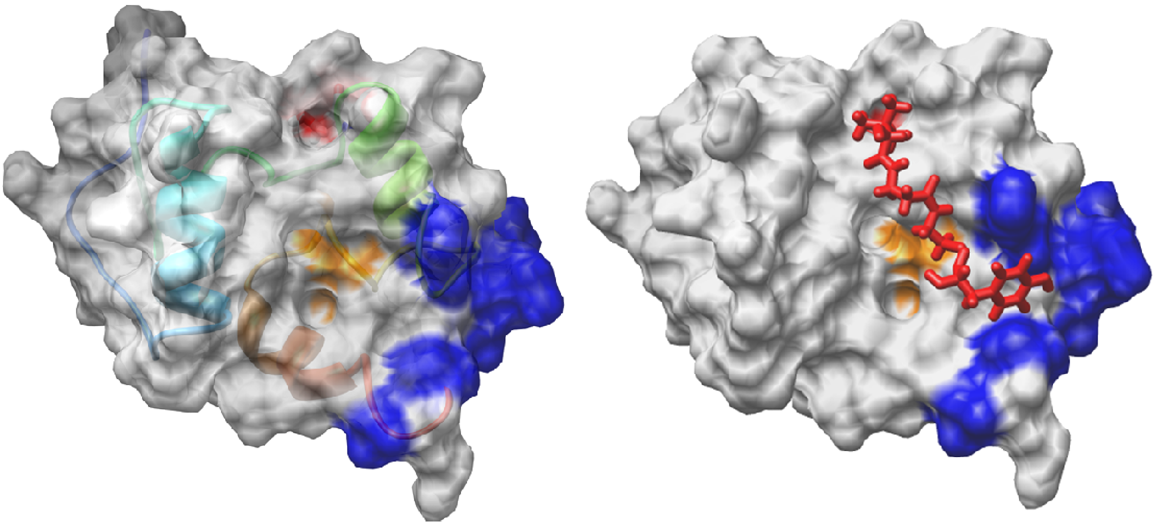


Fig. 24: Imaging of the frozen H-state of TYR-holo-TycC3-PCP. On the left side the structure of the H-state is shown with helix αI (light blue), helix αII (green), the hydroxyl-group of Ser 45 (red), the by NOEs detected contacts of the cofactor to Leu 68 and Phe 69 (orange) and dark-blue colored the amino acid residues with changed chemical shifts. On the right side is the tyrosine-modfied 4'-PP cofactor modeled on the protein surface. The observed changes in chemical shifts of the protein after the tyrosine was bound to the cofactor suggest the ring current of the aromatic sidechain as the effecter of this changes. This experiment supports the model of the coordinated movement of the 4'-PP cofactor dependent on the protein dynamic.

Finally, the refined assignment of holo-TycC3-PCP (Fig. 25), as the bases for the performed structural calculations and modeling, and the structural statistic of the final structures derived from the performed calculations (Fig. 26) are shown.

Group	Atom	Nuc	Shift	SDev	Assignments
T4	H	1H	8.177	0.005	12
T4	HA	1H	4.388	0.007	4
T4	HB	1H	4.223	0.005	2
T4	N	15N	117.034	0.130	10
T4	QG2	1H	1.154	0.004	2
E5	H	1H	8.385	0.004	22
E5	HA	1H	4.220	0.012	2
E5	HB2	1H	2.219	0.004	2
E5	HB3	1H	1.909	0.003	2
E5	N	15N	122.521	0.042	20
E5	QG	1H	2.045	0.007	3
A6	H	1H	8.247	0.005	13
A6	HA	1H	4.229	0.013	2
A6	N	15N	124.908	0.012	10
A6	QB	1H	1.288	0.004	4
Q7	H	1H	8.254	0.004	24
Q7	HA	1H	4.222	0.006	3
Q7	HB2	1H	1.894	0.010	3
Q7	HB3	1H	1.996	0.004	3
Q7	HE21	1H	6.776	0.002	21
Q7	HE22	1H	7.441	0.001	16
Q7	N	15N	120.572	0.055	21
Q7	NE2	15N	112.238	0.000	33
Q7	QG	1H	2.274	0.005	5
Y8	H	1H	8.420	0.005	19
Y8	HA	1H	4.266	0.050	2
Y8	HB2	1H	3.111	0.003	3
Y8	HB3	1H	2.700	0.000	3
Y8	N	15N	124.580	0.086	14
Y8	QD	1H	7.000	0.000	2
Y8	QE	1H	6.536	0.002	4
V9	H	1H	6.680	0.005	16
V9	HA	1H	3.677	0.009	3
V9	HB	1H	1.458	0.007	3
V9	N	15N	130.055	0.032	12
V9	QQG	1H	0.718	0.007	4
A10	H	1H	8.049	0.008	15
A10	HA	1H	4.097	0.010	4
A10	N	15N	130.249	0.032	11
A10	QB	1H	1.301	0.005	2
P11	HA	1H	4.242	0.008	2
P11	QB	1H	1.856	0.006	3
T12	H	1H	9.621	0.006	21
T12	HA	1H	4.404	0.005	3
T12	HB	1H	4.128	0.001	2
T12	N	15N	113.632	0.055	14
T12	QG2	1H	1.164	0.003	3
N13	H	1H	7.347	0.006	20
N13	HA	1H	4.713	0.006	9
N13	HB2	1H	3.039	0.003	6
N13	HB3	1H	2.887	0.004	6
N13	HD21	1H	7.258	0.004	22
N13	HD22	1H	6.480	0.006	23
N13	N	15N	116.085	0.036	15
N13	ND2	15N	113.876	0.000	39
A14	HA	1H	3.878	0.004	2
A14	QB	1H	1.512	0.000	1
V15	H	1H	7.861	0.004	24
V15	HA	1H	4.120	0.004	3
V15	HB	1H	1.947	0.004	3
V15	N	15N	120.385	0.055	17
V15	QG1	1H	0.679	0.001	2
V15	QG2	1H	0.802	0.006	2
E16	H	1H	8.392	0.006	31
E16	HA	1H	3.489	0.001	6
E16	HB2	1H	1.752	0.010	2
E16	HB3	1H	1.546	0.007	3
E16	N	15N	118.860	0.029	26
E16	QG	1H	2.304	0.002	2
S17	H	1H	8.308	0.003	36
S17	HA	1H	3.844	0.002	3
S17	HB2	1H	3.922	0.001	2
S17	HB3	1H	4.087	0.006	2
S17	N	15N	112.705	0.052	33
K18	H	1H	8.019	0.005	37
K18	HA	1H	4.218	0.004	3
K18	HB2	1H	1.948	0.003	2
K18	HB3	1H	2.364	0.009	4
K18	N	15N	124.389	0.031	32
K18	QD	1H	1.397	0.001	2
K18	QE	1H	3.197	0.000	1
K18	QG	1H	1.563	0.009	2
K18*	H	1H	8.105	0.018	3
K18*	N	15N	124.366	0.000	3
L19	H	1H	8.668	0.004	22
L19	HA	1H	3.971	0.003	5
L19	HB2	1H	1.910	0.004	2
L19	HB3	1H	1.814	0.004	3
L19	N	15N	118.624	0.045	17
L19	QD1	1H	0.661	0.005	3
L19	QD2	1H	0.800	0.003	3
A20	H	1H	8.127	0.005	36
A20	HA	1H	3.717	0.002	3
A20	N	15N	121.127	0.031	31
A20	QB	1H	1.345	0.003	2
E21	H	1H	7.637	0.004	39
E21	HA	1H	4.069	0.004	3
E21	HB2	1H	1.343	0.005	2
E21	HB3	1H	2.148	0.013	3
E21	HG2	1H	2.473	0.003	2
E21	HG3	1H	2.315	0.003	4
E21	N	15N	119.012	0.041	35
I22	H	1H	8.331	0.004	42
I22	HA	1H	3.714	0.004	2
I22	HB	1H	2.187	0.003	3
I22	HG12	1H	0.819	0.003	2
I22	HG13	1H	1.997	0.015	3
I22	N	15N	120.368	0.026	36
I22	QG2	1H	1.101	0.004	3
W23	H	1H	9.472	0.005	25
W23	HA	1H	4.128	0.003	2
W23	HB2	1H	3.713	0.006	2
W23	HB3	1H	3.262	0.004	4
W23	HH2	1H	6.860	0.005	4
W23	HZ2	1H	7.513	0.002	2
W23	HZ3	1H	7.255	0.002	2
W23	HE3	1H	7.623	0.000	1
W23	N	15N	120.296	0.039	17
E24	H	1H	8.979	0.004	28
E24	HA	1H	4.141	0.004	2
E24	HG2	1H	2.389	0.000	3
E24	HG3	1H	2.810	0.005	4
E24	N	15N	119.383	0.119	19
E24	QB	1H	2.172	0.001	2
R25	H	1H	7.992	0.004	38
R25	HA	1H	4.139	0.002	3
R25	HB2	1H	2.108	0.008	2
R25	HB3	1H	2.016	0.002	2
R25	HE	1H	7.279	0.002	2
R25	N	15N	119.426	0.046	35
R25	QD	1H	3.219	0.005	5
R25	QG	1H	1.762	0.003	3
V26	H	1H	8.385	0.003	32
V26	HA	1H	3.721	0.003	4
V26	HB	1H	1.938	0.008	2
V26	N	15N	119.255	0.052	28
V26	QQG	1H	0.984	0.008	2
L27	H	1H	8.931	0.005	28
L27	HA	1H	3.947	0.002	2
L27	HB2	1H	1.893	0.006	3
L27	HB3	1H	1.433	0.004	2
L27	HD11	1H	0.449	0.006	17
L27	HD12	1H	0.023	0.008	2
L27	HG	1H	1.248	0.005	2
L27	N	15N	116.170	0.089	24
L27	QD2	1H	0.986	0.002	2
G28	H	1H	7.983	0.004	13
G28	HA1	1H	4.132	0.012	2
G28	HA2	1H	3.869	0.002	3
G28	N	15N	108.702	0.046	10
V29	H	1H	6.967	0.005	20
V29	HA	1H	4.393	0.004	4
V29	HB	1H	1.898	0.005	2
V29	N	15N	115.128	0.116	14
V29	QG1	1H	0.804	0.003	2
V29	QG2	1H	0.781	0.010	3
S30	H	1H	8.083	0.005	16
S30	HA	1H	4.622	0.004	2
S30	HB2	1H	3.687	0.001	2
S30	HB3	1H	3.837	0.003	3
S30	N	15N	115.465	0.040	12
G31	H	1H	8.497	0.005	15
G31	HA1	1H	3.693	0.002	2
G31	HA2	1H	4.629	0.003	3
G31	N	15N	112.096	0.013	11
I32	H	1H	8.162	0.008	28
I32	HA	1H	3.840	0.004	3
I32	HB	1H	1.749	0.004	2
I32	HG12	1H	1.125	0.003	2
I32	HG13	1H	1.613	0.004	3
I32	N	15N	121.237	0.070	24
I32	QG2	1H	1.031	0.009	2
G33	H	1H	8.689	0.005	17
G33	HA1	1H	4.380	0.003	3
G33	HA2	1H	3.597	0.003	4
G33	N	15N	115.067	0.027	10
I34	H	1H	8.487	0.005	20
I34	HA	1H	3.956	0.002	3
I34	HB	1H	1.694	0.000	2
I34	N	15N	110.126	0.018	16
I34	QD1	1H	0.650	0.002	2
I34	QG1	1H	0.635	0.013	2
I34	QG2	1H	0.823	0.002	2
L35	H	1H	8.110	0.004	21
L35	HA	1H	4.644	0.002	3
L35	HG	1H	1.328	0.006	3
L35	N	15N	117.120	0.083	17
L35	QB	1H	1.680	0.007	3
L35	QQD	1H	0.842	0.005	2
D36	H	1H	7.015	0.004	18
D36	HA	1H	4.318	0.003	3
D36	HB2	1H	2.532	0.001	3
D36	HB3	1H	2.726	0.002	3
D36	N	15N	121.576	0.054	14
N37	H	1H	9.033	0.004	18
N37	HA	1H	4.886	0.003	2
N37	HD21	1H	7.974	0.002	18
N37	HD22	1H	7.298	0.003	23
N37	N	15N	123.377	0.065	10
N37	ND2	15N	114.688	0.000	37
N37	QB	1H	2.775	0.001	5
F38	H	1H	9.263	0.004	22
F38	HA	1H	4.887	0.005	3
F38	HB2	1H	2.715	0.002	2
F38	HB3	1H	3.351	0.004	6
F38	N	15N	126.503	0.067	15
F38	QD	1H	7.116	0.000	1
F38	QE	1H	7.359	0.000	1
F39	H	1H	7.994	0.004	20
F39	HA	1H	3.898	0.003	4
F39	HB2	1H	2.945	0.003	5
F39	HB3	1H	3.272	0.005	3
F39	N	15N	112.318	0.075	15
F39	QD	1H	7.337	0.000	2
F39	QE	1H	7.435	0.008	5
F39	HZ	1H	7.389	0.001	1
Q40	H	1H	8.221	0.005	25
Q40	HA	1H	4.342	0.007	4
Q40	HE21	1H	6.904	0.003	20
Q40	HE22	1H	7.517	0.003	18
Q40	N	15N	119.917	0.015	21
Q40	NE2	15N	112.143	0.000	34
Q40	QB	1H	2.135	0.007	4
Q40	QG	1H	2.396	0.002	4
I41	H	1H	7.079	0.004	22
I41	HA	1H	4.412	0.006	4
I41	HB	1H	2.104	0.008	2
I41	N	15N	111.211	0.073	18
I41	QD1	1H	1.103	0.007	2
I41	QG1	1H	1.225	0.005	3
I41	QG2	1H	0.898	0.006	5
G42	H	1H	7.369	0.005	20
G42	HA1	1H	4.279	0.004	4
G42	HA2	1H	3.447	0.005	3
G42	N	15N	106.543	0.034	15
G43	H	1H	7.765	0.005	21
G43	HA1	1H	1.840	0.004	2
G43	HA2	1H	2.617	0.003	10
G43	N	15N	107.979	0.074	15
H44	H	1H	6.299	0.005	18
H44	HA	1H	4.989	0.004	5
H44	HD2	1H	7.724	0.000	1
H44	HE1	1H	6.951	0.000	1
H44	N	15N	112.772	0.168	12
H44	QB	1H	3.477	0.002	2
S45	H	1H	7.808	0.007	14
S45	HA	1H	3.339	0.000	1
S45	N	15N	119.757	0.025	10
S45	QB	1H	1.945	0.002	2
L46	H	1H	8.741	0.006	17
L46	HA	1H	4.390	0.007	3
L46	HG	1H	1.743	0.001	2
L46	N	15N	123.116	0.070	14
L46	QB	1H	1.895	0.004	2
L46	QQD	1H	0.927	0.003	2
K47	H	1H	7.672	0.003	35
K47	HA	1H	4.144	0.001	3
K47	N	15N	119.521	0.031	31
K47	QB	1H	1.295	0.003	3
K47	QD	1H	0.988	0.004	2
K47	QG	1H	1.598	0.001	2
K47*	H	1H	7.670	0.002	26
K47*	HA	1H	4.142	0.000	1
K47*	N	15N	119.338	0.017	25
K47*	QB	1H	1.306	0.000	1
K47*	QD	1H	0.993	0.000	1
K47*	QG	1H	1.588	0.000	1
A48	H	1H	8.647	0.005	41
A48	HA	1H	3.518	0.003	7
A48	N	15N	122.404	0.062	35
A48	QB	1H	1.407	0.001	4
M49	H	1H	7.792	0.004	32
M49	HA	1H	4.116	0.007	3
M49	HG2	1H	2.696	0.004	4
M49	HG3	1H	2.845	0.006	15
M49	N	15N	116.804	0.029	29
M49	QB	1H	2.244	0.003	4
A50	H	1H	7.335	0.004	18
A50	HA	1H	4.332	0.002	3
A50	N	15N	123.639	0.030	13
A50	QB	1H	1.592	0.003	2
V51	H	1H	7.904	0.003	37
V51	HA	1H	3.375	0.003	4
V51	HB	1H	2.059	0.006	5
V51	N	15N	118.821	0.017	31
V51	QG1	1H	0.658	0.001	3
V51	QG2	1H	0.771	0.001	2
A52	H	1H	8.385	0.004	41
A52	HA	1H	3.921	0.003	3
A52	N	15N	121.647	0.044	35
A52	QB	1H	1.586	0.001	2
A53	H	1H	8.188	0.004	34
A53	HA	1H	4.290	0.004	4
A53	N	15N	122.052	0.034	29
A53	QB	1H	1.571	0.007	2
Q54	H	1H	7.859	0.004	37
Q54	HA	1H	4.095	0.004	3
Q54	HB2	1H	2.235	0.002	5
Q54	HB3	1H	2.102	0.005	3
Q54	HE21	1H	6.588	0.004	26
Q54	HE22	1H	7.452	0.007	26
Q54	HG2	1H	2.587	0.002	7
Q54	HG3	1H	2.477	0.006	6
Q54	N	15N	118.599	0.066	31
Q54	NE2	15N	111.890	0.004	44
V55	H	1H	8.830	0.003	33
V55	HA	1H	3.613	0.003	5
V55	HB	1H	2.294	0.005	3
V55	N	15N	121.634	0.027	27
V55	QG1	1H	1.068	0.004	7
V55	QG2	1H	1.021	0.001	2
H56	H	1H	8.233	0.004	31
H56	HA	1H	4.271	0.003	3
H56	HB2	1H	3.417	0.001	2
H56	HB3	1H	3.187	0.002	7
H56	HD2	1H	7.742	0.002	1
H56	HE1	1H	6.599	0.000	1
H56	N	15N	120.835	0.065	27
R57	H	1H	7.802	0.005	36
R57	HA	1H	3.895	0.002	2
R57	HD2	1H	3.165	0.001	3
R57	HD3	1H	3.392	0.008	3
R57	HE	1H	7.328	0.001	2
R57	HG2	1H	1.607	0.005	2
R57	HG3	1H	1.733	0.004	3
R57	N	15N	119.293	0.092	34
R57	QB	1H	1.938	0.003	4
R57	QH1	1H	7.203	0.002	2
E58	H	1H	7.825	0.004	27
E58	HA	1H	3.883	0.004	3
E58	HB2	1H	1.295	0.002	2
E58	HB3	1H	1.396	0.003	2
E58	HG2	1H	1.908	0.006	2
E58	HG3	1H	1.755	0.005	2
E58	N	15N	116.175	0.057	24
Y59	H	1H	8.202	0.007	20
Y59	HA	1H	4.619	0.005	4
Y59	HB2	1H	3.141	0.007	2
Y59	HB3	1H	2.519	0.001	5
Y59	N	15N	112.341	0.040	17
Y59	QD	1H	6.504	0.001	3
Y59	QE	1H	6.780	0.002	5
Q60	H	1H	8.288	0.007	13
Q60	HA	1H	3.439	0.002	4
Q60	HE21	1H	6.449	0.003	18
Q60	HE22	1H	7.111	0.002	20
Q60	N	15N	115.654	0.052	10
Q60	NE2	15N	112.807	0.034	34
Q60	QB	1H	2.207	0.008	4
Q60	QG	1H	2.124	0.007	5
Q60*	H	1H	8.266	0.006	5
Q60*	N	15N	115.670	0.000	5
V61	H	1H	6.469	0.005	24
V61	HA	1H	4.290	0.004	2
V61	HB	1H	1.648	0.006	3
V61	N	15N	114.561	0.059	22
V61	QG1	1H	0.841	0.008	3
V61	QG2	1H	0.799	0.003	2
V61*	H	1H	6.476	0.006	3
V61*	N	15N	114.364	0.063	3
E62	H	1H	8.642	0.005	28
E62	HA	1H	4.226	0.009	3
E62	HB2	1H	1.820	0.001	2
E62	HB3	1H	1.897	0.002	2
E62	N	15N	125.909	0.037	26
E62	QG	1H	2.114	0.000	2
L63	H	1H	8.892	0.004	23
L63	HA	1H	4.758	0.001	4
L63	HB2	1H	1.112	0.002	2
L63	HB3	1H	1.848	0.003	2

L63	HG	1H	1.387	0.006	2
L63	N	15N	127.916	0.023	16
L63	QD1	1H	0.686	0.005	3
L63	QD2	1H	0.559	0.008	3
P64	HA	1H	4.373	0.005	3
P64	QB	1H	2.401	0.007	5
P64	QG	1H	1.878	0.003	2
L65	H	1H	8.665	0.006	31
L65	HA	1H	3.661	0.003	5
L65	HG	1H	1.259	0.002	2
L65	N	15N	126.363	0.032	27
L65	QB	1H	1.609	0.007	3
L65	QD1	1H	0.868	0.004	4
L65	QD2	1H	0.756	0.001	3
K66	H	1H	8.455	0.007	22
K66	HA	1H	3.894	0.003	3
K66	HB2	1H	1.725	0.003	2
K66	HB3	1H	1.671	0.003	2
K66	HG2	1H	1.307	0.007	3
K66	HG3	1H	1.263	0.006	2
K66	N	15N	114.158	0.039	17
K66*	H	1H	8.519	0.012	11
K66*	N	15N	114.327	0.000	8
V67	H	1H	7.094	0.005	36
V67	HA	1H	3.532	0.002	4
V67	HB	1H	1.701	0.003	3
V67	N	15N	119.845	0.042	32
V67	QG1	1H	0.029	0.006	2
V67	QG2	1H	0.675	0.001	2
V67*	H	1H	7.096	0.005	11
V67*	HB	1H	1.698	0.000	1
V67*	N	15N	119.261	0.000	10
V67*	QG2	1H	0.674	0.000	1
L68	H	1H	6.801	0.005	40
L68	HA	1H	3.864	0.003	2
L68	HB2	1H	1.450	0.000	2
L68	HB3	1H	1.226	0.000	2
L68	HG	1H	1.705	0.005	2
L68	N	15N	121.008	0.068	35
L68	QD1	1H	0.033	0.004	4
L68	QD2	1H	0.156	0.002	6
L68*	H	1H	6.792	0.004	4
L68*	N	15N	120.648	0.000	3
F69	H	1H	8.154	0.006	38
F69	HA	1H	4.210	0.005	2
F69	HB2	1H	2.898	0.007	4
F69	HB3	1H	2.998	0.001	3
F69	N	15N	112.912	0.035	32
F69	QD	1H	6.208	0.001	4
F69	QE	1H	7.002	0.005	14
F69*	H	1H	8.144	0.004	2
F69*	N	15N	112.929	0.000	2
A70	H	1H	7.348	0.005	38
A70	HA	1H	4.219	0.001	3
A70	N	15N	118.597	0.016	36
A70	QB	1H	1.487	0.005	5
A70*	H	1H	7.363	0.002	8
A70*	N	15N	118.616	0.067	8
Q71	H	1H	7.415	0.005	33
Q71	HA	1H	4.764	0.000	3
Q71	HB2	1H	1.985	0.002	4
Q71	HB3	1H	1.881	0.001	3
Q71	HE21	1H	7.247	0.002	17
Q71	HE22	1H	6.712	0.006	18
Q71	N	15N	116.226	0.077	28
Q71	NE2	15N	110.850	0.000	31
Q71	QG	1H	2.193	0.005	5
P72	HA	1H	4.934	0.003	8
P72	QD	1H	4.131	0.006	3
P72	QG	1H	1.464	0.004	2
T73	H	1H	8.289	0.005	22
T73	HA	1H	5.129	0.006	7
T73	HB	1H	4.644	0.007	4
T73	N	15N	110.158	0.034	17
T73	QG2	1H	1.002	0.004	3
I74	H	1H	9.825	0.007	27
I74	HA	1H	4.669	0.001	3
I74	HB	1H	2.999	0.003	6
I74	HG12	1H	0.622	0.004	3
I74	HG13	1H	-0.051	0.006	10
I74	N	15N	125.127	0.030	19
I74	QD1	1H	-0.191	0.006	6
I74	QG2	1H	1.089	0.001	2
K75	H	1H	8.948	0.003	35
K75	HA	1H	3.728	0.002	3
K75	HD2	1H	1.466	0.005	2
K75	HD3	1H	1.114	0.004	2
K75	HG2	1H	0.635	0.007	2
K75	HG3	1H	0.792	0.001	2
K75	N	15N	119.473	0.054	31
K75	QB	1H	1.677	0.000	2
A76	H	1H	7.541	0.004	39
A76	HA	1H	4.151	0.008	5
A76	N	15N	119.671	0.064	33
A76	QB	1H	1.474	0.006	2
L77	H	1H	8.830	0.003	40
L77	HA	1H	4.033	0.003	4
L77	HB2	1H	1.860	0.002	2
L77	HB3	1H	2.034	0.005	3
L77	HG	1H	1.448	0.006	2
L77	N	15N	122.017	0.040	36
L77	QD1	1H	1.095	0.005	2
L77	QD2	1H	0.878	0.001	2
A78	H	1H	9.312	0.004	17
A78	HA	1H	3.689	0.008	4
A78	N	15N	120.524	0.020	11
A78	QB	1H	1.364	0.006	3
Q79	H	1H	7.110	0.008	23
Q79	HA	1H	4.005	0.004	5
Q79	HE21	1H	7.398	0.003	17
Q79	HE22	1H	6.773	0.009	19
Q79	HG2	1H	2.469	0.002	5
Q79	HG3	1H	2.295	0.003	4
Q79	N	15N	117.836	0.031	17
Q79	NE2	15N	111.688	0.000	30
Q79	QB	1H	2.125	0.005	5
Y80	H	1H	7.634	0.006	21
Y80	HA	1H	3.932	0.009	3
Y80	HB2	1H	3.019	0.004	5
Y80	HB3	1H	3.242	0.005	2
Y80	N	15N	121.746	0.036	16
Y80	QD	1H	6.789	0.003	4
Y80	QE	1H	6.931	0.000	1
V81	H	1H	8.460	0.005	19
V81	HA	1H	3.199	0.008	4
V81	HB	1H	1.745	0.002	3
V81	N	15N	117.941	0.048	15
V81	QG1	1H	0.145	0.002	3
V81	QG2	1H	0.843	0.002	2
A82	H	1H	8.337	0.004	36
A82	HA	1H	3.862	0.006	2
A82	N	15N	121.200	0.071	31
A82	QB	1H	1.373	0.002	5
T83	H	1H	7.474	0.004	18
T83	HA	1H	4.019	0.006	3
T83	HB	1H	4.112	0.004	2
T83	N	15N	109.038	0.096	13
T83	QG2	1H	1.170	0.006	4
R84	H	1H	7.294	0.005	21
R84	HA	1H	4.131	0.002	3
R84	HD2	1H	2.792	0.000	1
R84	HD3	1H	2.494	0.000	1
R84	HE	1H	6.835	0.002	2
R84	N	15N	119.962	0.058	17
R84	QB	1H	1.856	0.002	2
R84	QG	1H	1.321	0.001	2
S85	H	1H	7.706	0.004	16
S85	HA	1H	4.222	0.004	2
S85	HB2	1H	3.698	0.001	2
S85	HB3	1H	3.853	0.001	2
S85	N	15N	114.661	0.066	12

Fig. 25: Sequential resonance assignment of TycC3-holo-PCP. The resonance assignment of all observed protons and the amide nitrogen chemical shifts are shown. The amino acid residues that provide double peaks in holo-PCP are labeled by asterisks.

Structures calculated with both protocols were virtually identical, showing that the data are interpreted identically by both automatic NOE assignment routines.

Structural statistics for all three structures are shown in Fig. 26. Analysis of the backbone dihedral angles revealed that in the A-state 93.2% is located in the most favoured and additionally allowed regions with no dihedral angles in the disallowed region. For the A/H-state the same analysis showed that 93.2% of dihedral angles are located in the most favoured and additionally allowed, 5.3% in the generously allowed and 1.5% in the disallowed regions. In the H-state structure 87.9% are located in the most favoured and additionally allowed regions and 3.4% are located in the disallowed region. Analysis of the dihedral angles was performed with the program package PROCHECK (Laskowski & MacArthur, 1993).

	A-state	A/H-state	H-state
NMR distance & dihedral constraints			
Distance constraints	874	831	753
Total NOE	1462	1867	1298
Intra-residue	361	435	345
Inter-residue			
Sequential (|i-j| = 1)	412	480	299
Medium-range (|i-j| < 4)	518	560	440
Long-range (|i-j| > 5)	171	392	214
Intermolecular	-	-	-
Hydrogen bonds	8	25	19
Total dihedral angle restraints			
ϕ	-	-	-
ψ	-	-	-
Structural Statistics			
Violations (mean and s.d.)			
Distance constraints (Å)	0.19+/- 0.04	0.18/+/- 0.03	0.16+/- 0.04
Dihedral angle constraints (°)	-	-	-
Max. dihedral angle viol. (°)	-	-	-
Max. distance constraint viol. (Å)	0.23	0.22	0.19
Dyana target function (20 / 100)	1.19	1.40	1.37
Deviations from idealized geometry			
Bond lengths (Å)	0.02	0.02	0.02
Bond angles (°)	-	-	-
Impropers (°)	-	-	-
Averaged pairwise r.m.s.d. (Å)			
Heavy	0.79+/-0.13	0.81+/- 0.23	1.10+/-0.16
Backbone	0.56+/-0.15	0.67+/-0.19	0.89+/-0.20
Number of conformers (out of)	18 (100)	20 (100)	18 (100)

Fig 26: Structural statistic of the DYANA derived models of the A-state, the A/H-state and H-state of TycC3-PCP. (Koglin *et al. Science,* 2006)

5 Discussion

The importance of protein dynamics and conformational exchanges for the regulation of a protein's activity has become widely accepted and can be described by two different models. In the induced fit model specific conformational changes are induced by specific binding events of effectors. The pre-formed equilibrium model on the other hand is based on the assumption that the non effected protein exists in multiple conformations of which one is individually selected upon ligand binding (Kern & Zuiderweg, 2003; Goh et al., 2004).

This thesis describes a new dual two-state model of an integrated single domain protein. The PCP domain can exist either in an apo- or holo-form-specific preformed equilibrium that overlap in the A/H-state as a common conformer. This model basically matches with the pre-exisiting equilibrium hypothesis (Goh et al., 2004). Proteins like e.g. the regulator NtrC can exist in two different conformers that exhibit different activities (Volkman et al., 2001). In our case, the A- and H-states would represent the active conformations for the specific binding of the interacting partners Sfp and SrfTEII. A new feature however is the presence of two alternative equilibriums dependent on the modification status of the PCP domain. The binding of the 4'-PP cofactor transforms the A-state into the H-state thus creating a new conformer. 4'-PP binding alters the conformation especially of helix αIII at a remote site and enables its recognition by SrfTEII. This mechanism could therefore be interpreted as a kind of allosteric regulation (Kern & Zuiderweg, 2003) although the 4'-PP cofactor itself might additionally play a role in the recognition by SrfTEII, which is supported by additional NMR-experiments (see additional section). The TycC3-PCP conformational equilibrium therefore combines two classical models of protein dynamics and could represent an example of a new model for cofactor dependent allostery controlled equilibrium (Fig 27).

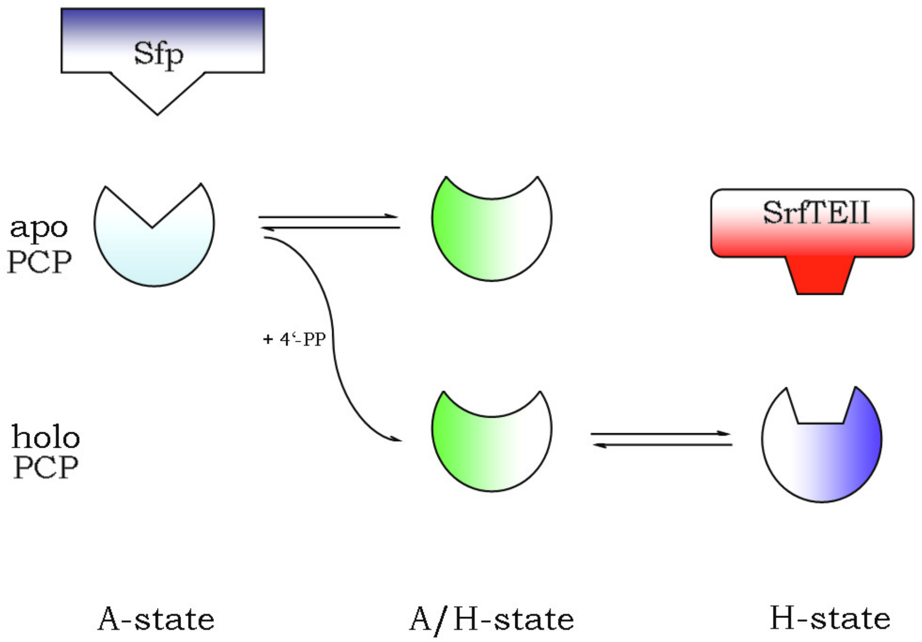


Fig. 27: Allosteric exchange model. The two-state dynamic of TycC3-PCP is controlled by the covalent modification with the 4'-PP cofactor and controls the specific interaction with Sfp (dark blue) in the A-state and SrfTEII (red) in the H-state. (Koglin *et al. Science,* 2006)

PCP domains share an extensive structural similarity to the functionally related ACPs of the fatty acid synthase (FAS) and polyketide synthase (PKS) systems, whereas homologues of the primary sequences are mostly confined to the area is surrounding the 4'-PP cofactor binding site. Common conserved elements of all analysed ACP and PCP structures so far are the three relatively large helices αI, αII, and αIV that are connected by loop regions of various lengths (Holak et al., 1988; Crump et al., 1997; Xu et al., 2001; Findlow et al., 2003; Wong et al., 2002; Reed et al., 2003; Weber et al., 2000). Those three helices are relatively parallel in the ACPs of *E.coli* (Holak et al., 1988), *B. subtilis* (Xu et al., 2001) and *M. tuberculosis* (Wong et al., 2002) while they are distorted in TycC3-PCP (Weber et al., 2000) and in the ACPs of *rattus spec.* (Reed et al., 2003) and *Streptomyces rimosus* (Findlow et al., 2003). A short fourth helix III composed of five to seven residues is located perpendicular to the three helix bundle in most ACP structures (Xu et al., 2001; Reed etal., 2003), whereas such a helix could not be clearly detected in the ACP solution structures of *E. coli* (Holak et al., 1988) and of *Streptomyces coelicolor* (Crump et al., 1997). Only one turn of this helix III can be defined in the A/H-state of TycC3-PCP, while it is absent in the A-and H-states. Superpositions of the ACP solution structures revealed the largest differences in the helix αIII/loop III area and in the long loop between helix αI and αII (Xu et al., 2001; Wong et al., 2002) containing the active site area with the conserved serine residue for the cofactor attachment. Holo-ACPs have to interact with the various enzymes of the FAS or PKS systems and it is therefore not surprising that some similarities to the exchange model of TycC3-PCP can be found. In the ACP of the oxytetracycline polyketide synthase of *S. rimosus* a rigid scaffold was found to be surrounded by flexible sites that constitute potential protein interaction sites (Findlow et al., 2003). Substantial motions in loop regions and a partial distortion of the helix αII have been suggested in the structure of *E. coli* ACP in earlier reports (Kim & Prestegard, 1989; Andrec et al., 1995). The recently published report about the crystal structure of the complete mammalian fatty acid synthetase (Maier et al., 2006) suggests a dynamic for the acyl carrier proteins. The attachment of the 4'-PP cofactor to *B. subtilis* -ACP induced distinct chemical shift changes in ^{15}N-HSQC spectra of few residues in close proximity to the active site serine residue at the N-terminal region of helix αII (Xu et al., 2001). However, so far no conformational heterogeneity of apo- and holo-ACPs has been proposed.

The dual two state conformational exchanges of TycC3-PCP have clear implications for recognition processes and describe two examples where strictly only one specific conformer is involved. In the A-state, the loop between helix αI and αII is extended up to position Ala 48. The highly conserved area surrounding the active site residue Ser 45 becomes therefore better exposed in an easily accessible loop as compared to the more compact A/H- and H-states and could thus serve as a recognition platform for 4'-PP-transferases like Sfp. The monomeric Sfp forms a topological pseudohomodimer since it folds into two very similar domains (Reuter et al., 1999; Mofid et al., 2004; Finking et al., 2004). Mutagenesis experiments revealed that helix αII of PCPs and ACPs seems to be involved in interactions with PPTases (Mofid et al., 2002; Finking et al., 2004).

This view is confirmed by the crystal structure of the *B. subtilis* ACP/AcpS complex (Parris et al., 2000). In addition, helix αII is also supposed to be the recognition helix for further FAS enzymes (Zhang et al., 2001). Accordingly, conformational exchange was detected in the long loop between helix αI and αII and in the helix αII in the homologous ACP of *S. rimosus* (Findlow et al., 2003). The exchange of the complete helix αII of PCP with an ACP-type helix αII was furthermore sufficient for a 4'-PP cofactor attachment by both, AcpS and Sfp-type enzymes (Mofid et al., 2002). Conserved residues proximal to the active site serine are reported to form hydrophobic contacts or salt bridges to the PP-transferase and residue K47 might play a role in scrutinizing between PCPs and ACPs by PP-transferases (Finking et al., 2004). Based on our complex model of Sfp/apo-TycC3-PCP, additional contacts especially to the loop area between helix αH1 and αH2 of Sfp can be proposed. A striking feature of the A-state is the embedding of the loop/helix III area in the core of apo-TycC3-PCP formed by helix αII and αIV. This region disappears completely from the surface of the A-state PCP and is therefore unlikely to be involved in recognition processes by Sfp. The proposed placement of the active site serine residue in close proximity to the cleavable pyrophosphate group coordinated by the Mg^{2+} ion would nicely enable the 4'-PP transfer according to conventional Mg^{2+} catalyzed phosphate hydrolysis.

Holo-PCPs have an array of diverse proteins they need to communicate with. Those include the various adjacent enzymatic domains of the NRPS systems that are involved in the peptide product formation like the A- or C-domains as well as thioesterase I and II, enzymes that are responsible for the final release of the peptide product or for the recycling of misprimed PCPs (Schwarzer et al., 2003; Brunner et al., 2002). In holo-TycC3-PCP, the conformational heterogeneity in the loop connecting helices αI and αII, that is necessary for the interaction of apo-TycC3-PCP with Sfp, is now almost completely absent. However, the conformational exchange of the helix/loop III region is increased and that area together with the surrounding loops is now located outside of the protein core and accessible for the interaction with other proteins. We could already identify these regions as recognition sites for the SrfTEII enzyme while the bound 4'-PP cofactor with a terminal acetyl residue additionally contribute to the complex formation. PCPs might furthermore become specifically recognized by adjacent A- and C-domains (Weinreb et al., 1998) and helix αII was already proposed not to be involved in those interactions (Mofid et al., 2002). In addition, the communication of PCP domains with C-terminal epimerization domains seemed to be affected by helix αII. This work was focused on interactions of the TycC3-PCP domains with individual enzymes known to be essential for functional NRPS systems *in trans*. Further analysis of intramolecular domain complexes *in cis* of a NRPS will certainly increase insights in the differential contact formation of PCP domains.

The 4'-PP cofactor of TycC3-PCP was previously supposed to be flexible and extending away from the protein (Weber et al., 2000). This view would resemble the proposed situation in holo-ACPs. In the holo-ACP structures of *B. subtilis* and *M. tuberculosis*, no NOEs were observed between the 4'-PP cofactor and the protein itself observed from the protein side (Xu et al., 2001; Wong et al., 2002). This could indicate the absence of any preferential interactions with the ACP suggesting that the cofactor is either essentially solvent exposed or experiences a rapid exchange between multiple low-affinity ACP binding sites. However, small differences in the amide proton and nitrogen chemical shifts in a mixture of the apo- and holo-ACP of *M. tuberculosis* (Wong et al., 2002) gave evidence for a possible interaction of the 4'-PP cofactor with residues located in a proposed hydrophobic substrate binding pocket of ACPs proximal to the active site serine residue (Xu et al., 2001). Based on these observations and due to the required mobility of the 4'-PP cofactor upon transfer of acyl intermediates, the authors proposed a rapid exchange of the cofactor between a solvent exposed state and a second distinct state bound to the ACP. In this model, the 4'-PP cofactor of ACPs oscillates between a freely rotating state that is only fixed by the phosphoester at the active site serine and a more restricted protein bound state with the cofactor bound acyl ligand most likely being important for interaction (Wong et al., 2002).

Our results indicate two defined positions for the 4'-PP cofactor located in proximity to either the N-terminal end of holo-TycC3-PCP in the A/H-state or to the C-terminal end in the H-state. The relocation of the cofactor is initiated by a considerable extension of the loop/helix III region and dislocates helix αII causing the large rotation of approx. 100° of the cofactor. This conformational heterogeneity most likely allows the PCP to shuttle bound substrates from one side of the domain to the other. In addition, the N-terminal end in the A/H-state is tightly fixed by specific H-bonds that presumably results in a closer contact between the preceding A-domain and the PCP, thus placing the cofactor into an optimal position for its loading with an amino acid substrate by the previous A-domain. It is tempting to speculate that the associated structural rearrangements in the two holo-TycC3-PCP forms might additionally modulate specific interactions with the A- and C-domains of the NRPS system. In contrast to the integrated PCP domains type II ACPs in prokaryotic and plant cells are structurally independent small acidic proteins. Only type I ACPs found in animal tissues form integrated subunits of multifunctional FAS and PKS systems similar to the PCPs in NRPS systems. A non-ordered solvent exposed state of the cofactor in type II ACPs might be reasonable for the efficient loading with acyl intermediates by enzymes of the fatty acid synthetases complex that are present *in trans*. However, in NRPS systems the integrated PCP domains interact with enzymatic domains *in cis* and they have to use the 4'-PP cofactor as a flexible tether in order to act as peptide carrier between the adjacent regions. Therefore, an efficient shuttle function would rather benefit from rapid oscillations of the cofactor between defined positions in the proximity of either end of the PCP domain than from a non-ordered flexible orientation. These observations strongly support a directed mechanism of peptide synthesis enforced by a controlled 4'-PP cofactor mobility that is driven by the conformational equilibrium of holo-PCPs. The structural analysis of differentially loaded PCP domains will be an approach in order to further specify this mechanism

.

6 Additional

The external Thioesterase of the Surfactin synthetase

Enzymes that hydrolyze thioesters are omnipresent in prokaryotes and eukaryotes, since thioesters appear in many different metabolic and biosynthetic pathways. Thioesterases are found in protein metabolism to cleave fatty acids from cysteines of lipidated proteins or mainly associated with nonribosomal peptide synthetases (NRPS) and polyketide synthetases (PKS) for the 4'-phosphopantheteine (4'-PP) processing (Linne & Marahiel, 2000; Yeh et al., 2004). In NRPS- and PKS-systems are two types of thioesterases described: a C-terminal integrated Thioesterase I-domain, which is responsible for the product release from the biosynthetic machinery and the external thioesterase type II, which is expressed as a separated protein (Linne & Marahiel, 2000). The biological function of this type of separated thioesterases was recently studied and described as a regenerating system for misprimed 4'-PP cofactors of peptidyl- (PCP) and acyl (ACP) carrier domains from NRPS and PKS systems (Linne et al. 2004).

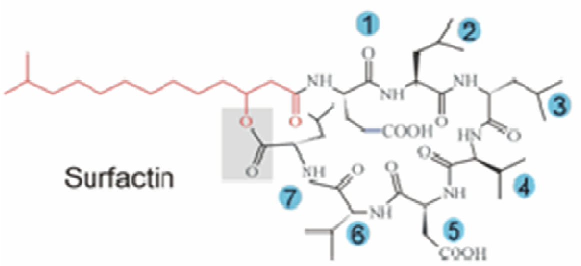


Fig. 28: Structure of Surfactin. The cyclic lipoheptapeptide Surfactin is biosynthesized in the indicated order and subsequently cyclisized by a carboxyester formation of L-Leu[7] and the β-hydroxy fatty acid.

The 28 kDa protein Surfactin Thioesterase II (SrfTEII), the external stand-alone Thioesterase, of *Bacillus subtilis* consist of 242 amino acid residues and is associated with the biosynthesis of the lipoheptapeptide surfactin (Fig. 28). Surfactin is one of the most potent biosurfactants known with an enormous antibacterial and antiviral potential. Surfactin is a macrocyclic heptapeptide bound to a β-hydroxy-myristoyl residue as the attached fatty acid, the peptide is biosynthesized in the order: L-Glu[1], L-Leu[2], D-Leu[3], L-Asp[4], L-Val[5], D-Leu[6] and L-

Leu[7] and subsequently cyclisized between amino acid residue L-LEU[7] and the β-hydroxy fatty acid by the internal Surfactin Thioesterase (SrfTEI).

Surfactin is produced by a multimodular nonribosomal peptide synthetase (NRPS), the Surfactin synthetase a three transcripts complex. Important examples for cyclic peptides biosynthesized by NRPS systems are vancomycin, cyclosporine, penicillin, tyrocidine and cephalosporin-precursors (Walsh, 2004; Sieber & Marahiel, 2005). All these peptides are potent antibiotic, immune suppressive or cytostatic substances. NRPS systems are multi-functional complexes of large multi-domain protein with up the 500 kD each transcript (Fig. 29).

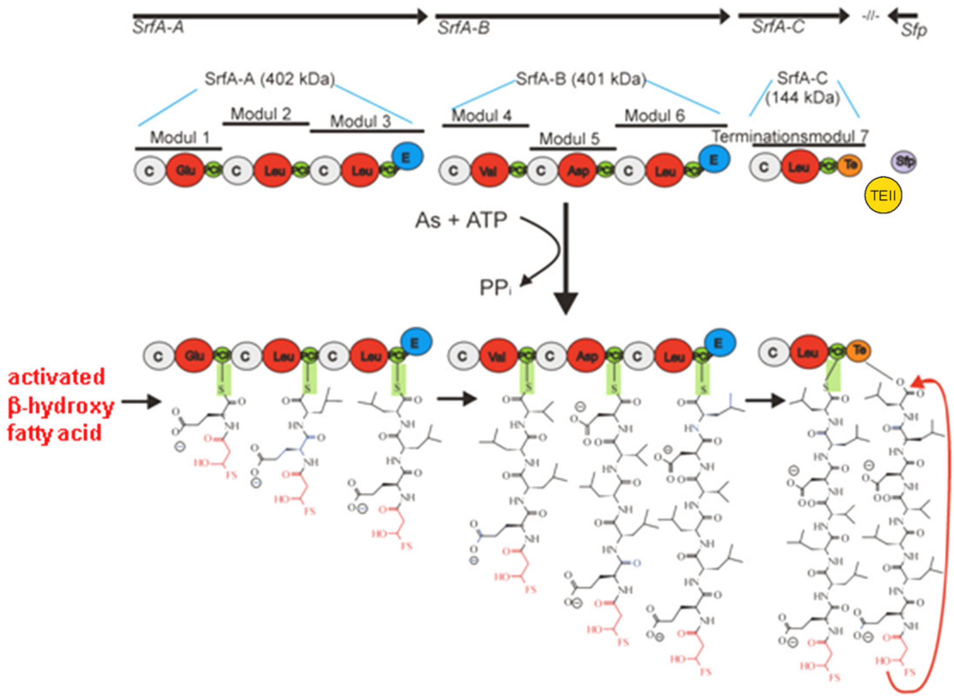


Fig. 29: Organisation of surfactin synthesis. The biosynthesis of the lipopeptide surfactin A is encoded by the srfA-operon which comprises four open reading frames codifying the protein component SrfA-A to A-D of the surfactin synthetase multienzyme system. The two large enzymes of the surfactin synthetase SrfA-A and SrfA-B, each comprising three amino acids activating modules, assemble amino acids as thioesters to surfactin in the following order: L-Glu, Leu, Leu (by SrfA-A) and L-Asp, L-Val and L-Leu (by SrfA-B). SrfA-A is responsible for initiating the surfactinA biosynthesis by containing an acyltransferase activity, which transfers a beta-hydroxy fatty acid to the L-Glu activating module. SrfC is a one-module enzyme which contributes the C-terminal L-leucin-residue. SrfD (SrfTEII) is an external thioesterase and carries a regeneration function for false modificated peptidyl carrier proteins (PCPs). Modification of the PCP domain by Sfp with acetyl-CoA instead of CoA inactivates PCP domains and blocks the NRPS machinery. SrfTEII affiliate to the a/b-hydrolase family and a catalytic-triade mechanism is presumed. (Bruner et al. (2002)

Each elongation module of the NRPS systems is responsible for the incorporation of one amino acid residue into the mature peptide (Walsh, 2004;

Sieber & Marahiel, 2005). The modules are organized at least in the three domains A-PCP-C, catalyzing the activation (A), thioester formation and transportation (peptidyl carrier protein, PCP) and condensation (C) of the specific amino acid substrates (Fig. 29). Epimerization domains (E) synthesize a stereospecific epimerization from L-from amino acid residues to the D-form. The covalent organization of the synthetase modules and the substrate specificity of the A-domains ensure the precise biosynthetic sequence of the peptides.

The genetic organization of the operon is symbolized in the upper row. The Condensation-domains for the initial loading of the β-hydroxy fatty acid and for the condensation of the growing peptide chain are shown in grey, labeled 'C'. Activation-domains and Peptidyl-carrier domains are shown in red with the specific amino acid residue labeled respectively in green, labeled 'PCP' with the growing peptide chain bound. The epimerization-domains of the SrfA-A and SrfA-B modules are shown in blue and labeled 'E'. The Thioesterase I domain, responsible for the release of the mature and heterocyclisized lipopeptide Surfactin is a component the SrfA-C module (orange). The external Thioesterase (SrfTEII) and the 4'-phosphorpantheteine-transferase (4'-PPtase) Sfp are expressed separately. The biosynthesis of the lipoheptapeptide Surfactin A is encoded by the *srfA*-operon which comprises four open reading frames codifying the protein component SrfA-A to A-D of the Surfactin synthetase multienzyme system. The two large enzymes of the Surfactin synthetase SrfA-A with a molecular weight of 402 kDa and SrfA-B (401 kDa), each comprise three amino acids activating modules, assemble amino acid residues to Surfactin by carrier-domain thioester formation in the following order: L-Glu, L-Leu, D-Leu (by SrfA-A) and L-Asp, L-Val and D-Leu (by SrfA-B). SrfA-A is responsible for initiating the Surfactin A biosynthesis by containing an acyltransferase activity, which transfers a β-hydroxy fatty acid to the L-Glu activating module and carries an epimerization-domain (E-domain) for the L- to D-form isomerization of the third leucine. The overall organization of SrfA-B is nearly similar, including the C-terminal E-domain for conversion of L-Leu^6 to D-Leu^6. SrfC is a one-module enzyme which contributes the C-terminal L-leucine-residue and the contains the Thioesterase I domain. SrfD (SrfTEII) is an external thioesterase and carries a regeneration function for false posttranslational modificated peptidyl carrier proteins (PCPs). Modification of the PCP domain by Sfp with acetyl-CoA instead of CoA inactivates PCP domains and blocks the NRPS machinery. This misleading of carrier domains is caused by the high concentration of acetylated or acylated Coenzyme A. In *E. coli* 80% of the intracellular Coenzyme A is acetylated and 14% are available as free Co A only.

SrfTEII affiliates to the α/β-hydrolase family and a catalytic-triad mechanism is presumed.(Bruner et al., 2002). The catalytic triad consists of a serine residue (Ser86) as the nucleophilic residue in a Gly-X-Ser-X-Gly core motif, a acidic residue aspartic acid (Asp190) and is completed by a histidine (His216) and was analyzed by mutational analysis (Linne et al. 2004). SrfTEII is essential for an efficient surfactin biosynthesis as it determines the correct loading of the 4'-phosphopantetheine (4'-PP) prosthetic groups attached to the carrier-domains (PCP), and it regenerates posttranslational mismodificated (misprimed) of 4'-PP-thiol groups (Fig. 30).

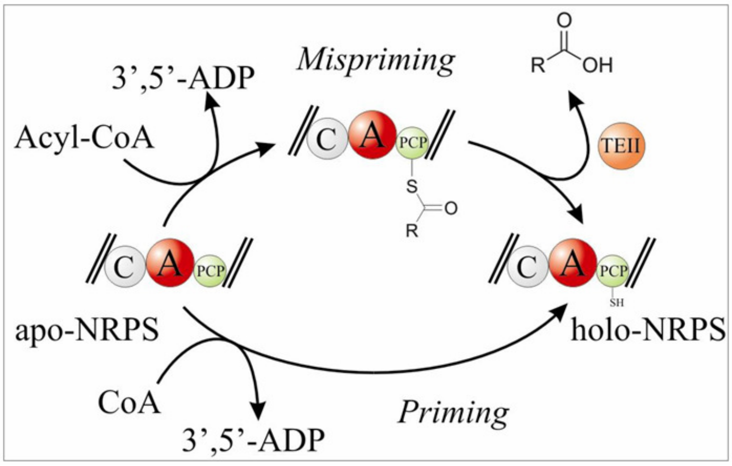


Fig. 30: Mechanism of SrfTEII. NRPS systems are often associated with external type II thioesterases of approx. 250 amino acids, that are essential for their functional regeneration. The SrfTEII enzyme is responsible for hydrolysing misloaded acetyl residues from otherwise blocked PCP domains. Mispriming occurs by phosphopantetheinyl transferases (PPTase) like Sfp that use acetyl-CoA instead of CoA as substrate and phosphopantetheinyl group donor for PCP domains (Quadri et al. (1998), Biochemistry 37:1585-95). The deletion of SrfTEII results in the reduction of the in vivo surfactin biosynthesis of more than 80 % (Linne et al. (2004).

NRPS systems are often associated with external type II thioesterases of approx. 250 amino acid residues, which are essential for their functional regeneration. The SrfTEII enzyme is responsible for hydrolyzing misloaded acetyl residues from otherwise blocked PCP domains. Mispriming occurs by phosphopantetheinyl transferases (4'-PPTase) like Sfp by loading acetyl-CoA instead of CoA as substrate as phosphopantetheinyl group donor for PCP domains (Quadri et al. 1998). The deletion of SrfTEII results in the reduction of the *in vivo* Surfactin biosynthesis of more than 80%. The catalytic triade of SrfTEII was clarified by mutational analysis. The S86A point mutation totally abolished the hydrolase activity of TEII (Linne et al., 2004).

Based on high resolution Nuclear Magnetic Resonance (NMR) techniques, the structure of the SrfTEII protein should verify the putative α/β-hydrolase fold. Structural analysis of the SrfTEII will provide an insight into the specificity of protein-protein interaction in large biochemical machineries like NRPS systems. The structure determination of proteins by liquid state high resolution NMR is based on sets of multidimensional heteronuclear NMR spectra that assign proton and heteronucleii of the protein backbone and of the sidechains. Based on inter-proton distance constraints derived from Nuclear Overhauser Enhancement Spectroscopy (NOESY) especially of not directly covalently bound amino acid residues and angle constraints, structure calculations can be performed.

Materials and Methods. The heterologous expression of SrfTEII was performed using the pQE30-protein expression vector (Qiagen) transformed into the *E. coli* DH5α strain. The SrfTEII-expression clone was selected on α-aminobenzylpenicillin (ampicillin) resistance.

The isotopic labeling is based on the standard protocol for M9-minimal media, where the ammonia and/or the carbons are/is replaced by NMR-actively labeled enriched sources. The ^{15}N-enriched Ammonia chloride (Cambridge Isotope Laboratory) as the only nitrogen source and uniformly 99% ^{13}C-enriched glycerol and glucose (Eurisotop) were used as the only ammonia- and carbon-sources at amounts of 1g ammonia chloride, 2g Glucose and 0.5g glycerol per liter M9-minimal media respectively. A 70% perdeuteration of the expressed SrfTEII protein was achieved by growing the cells in 99% D_2O-based M9-media with non-perdeuterated carbon sources and D_2O-exchanged salts. The prepared media can not be autoclaved, because of the aqueous atmosphere inside the autoclaved. The D_2O-media were sterilized by ultra-filtration and the D_2O can efficiently reused for fresh media after two passes in a rotary evaporator and a subsequently ultra-filtration.

The purification to homogeneity of the expressed SrfTEII was realized using a Ni-chelator resin (Amersham-Pharmacia) packed in a FPLC-column due to the C-terminal Hex-histidine-tag provided by the expression vector. The purified protein was eluted by a two step gradient with an increasing imidazole-concentration. The homogeneity of the protein sample was verified by SDS-polyacrylamide gel electrophoresis and using UV spectroscopy absorbance at 280 nm (absorption coefficient 16.860 $M^{-1}cm^{-1}$).

Using high resolution Nuclear Magnetic Resonance (NMR) techniques, the backbone- and the sidechain-atoms of the SrfTEII protein were manually assigned, using established standard-methods (van de Ven, 1995; Wider, G. & Wüthrich, K., 1999; Salzmann et al., 1999) and structure calculations were performed using the cyana 2.1 algorithm (Günthert, 2004). Sets of NMR-Spectra for backbone – and side chain assignment and structural calculation were recorded using doubly ($^{15}N/^{2}H$ and $^{15}N/^{13}C$) and triply ($^{15}N/^{13}C/^{2}H$) labeled samples of the purified SrfTEII protein. All spectra were recorded at 800 MHz and 900 MHz NMR-spectrometer (Bruker) equipped with triple resonance z-gradient cryogenic probes. For the backbone assignment following spectra were used: two dimensional-^{15}N-TROSY-HSQC, following three-dimensional spectra: HNCO, HNcaCO, HNCACB, HNcoCACB, HcacoNH and supported by backbone amide-amide contacts generated from a four-dimensional HNnoeNH-Noesy. For the assignment of aliphatic amino acid sidechains the following three dimensional NMR spectra were collected carbon- and proton-edited HCccoNH and TOCSY-HSQC. For the assignment of aromatic sidechains the three-dimensional TOCSY-HSQC, a D_2O-

exchanged two-dimensional HH-Noesy and a three-dimensional CH-Noesy-HSQC selective for aromatic sidechains was used. For performing structural calculations, distance constraints using the CANDID module of the simulated annealing protocol cyana2.1 were used. These distance constraints were based on mainly mid-range and long-range NOEs derived from a three-dimensional NH-Noesy-HSQC, a set of CH-Noesy-HSQC spectra selective for aromatic sidechains, for methyl-groups and for aliphatic sidechains. The four-dimensional HNnoeNH-Noesy was used in parallel to verify the global orientation of secondary structural elements. The secondary structure of SrfTEII was determined and applied as angle constraints for structural calculations based on the consensus of the results from calculated chemical shift index (csi) and the TALOS prediction.

The NMR-titration experiments were carried out by recording sets of ^{15}N-TROSY-HSQC experiments using a 1 mM sample of ^{15}N-labeled, perdeuterated SrfTEII in an aqueous 50 mM Na-phosphate buffer (pH 6.8) and adding stepwise unlabeled TycC3-PCP for each experiment. The titration experiments were performed at a 900 MHz NMR-spectrometer (Bruker) equipped with a triple-resonance, z-axis gradient probe and started with a ^{15}N-HSQC spectra of homogenous pure NMR-actively labeled SrfTEII protein sample. The final spectra was recorded with two times higher concentration of unlabeled TycC3-PCP compared to the SrfTEII-concentration. The number of scans were adjusted to the dilution of the SrfTEII protein. The holo-form of TycC3-PCP was generated by loading the 4'-PP cofactor *in vivo* due to coexpression of the Phosphopantheteine-transferase (PPTase) Sfp. The correct loading was checked by mass-spectrometry. The acetyl-holo-TycC3-PCP was generated by *in vitro* loading of acetyl-coenzyme A (purchased from Sigma-Aldrich) to apo-TycC3-PCP using the separately expressed PPTase Sfp. The complete and correct loading was verified by mass-spectrometry. For the titration experiment of SrfTEII wildtype protein with myristoyl-coenzyme A (purchased from Sigma-Aldrich) and of SrfTEII$_{S86A}$ mutant with myristoyl-coenzyme A, a 800 MHz-NMR-spectrometer (Bruker) equipped with a cryogenic triple-resonance probe was used.

Results. The SrfTEII protein is assigned to 95,5 % of all backbone resonances and 91% of all structurally relevant and detectable protons were assigned. The initial methionine could not be detected and the protein is assigned starting at position Ser2 and could be completed up to Gln241, the pre-last amino acid residue. The putative loop region between amino acid residue Asp67 and Pro76 could not be assigned. The signal intensity of amino acids sequentially close to this region are remarkable reduced in all NMR-spectra. The complete resonance assignment of the charged amino acid residues carrying nitrogen-containing sidechains like glutamine and asparagine was not possible, caused by the massive overlap of the sidechain resonances from 32 asparagines and glutamines. The nitrogen-containing sidechains of the 10 histidines and 10 arginines could be partially assigned. The two $N\varepsilon_1$-indoles of Trp200 and Trp203 could be identified, assigned and are labeled in Fig. 31. In Fig. 31 a ^{15}N-TROSY-HSQC spectra of perdeuterated ^{15}N-labeled SrfTEII and the assignments of all identified backbone resonances are shown and indicated by the one-letter-code for amino acid residues and the sequence-specific numeric counter. The sample conditions were 1 mM in protein concentration in a 50mM Na-phosphate buffer (pH 6.8) with 20 mM sodium chloride added. The spectra were recorded on an Avance 900 MHz NMR-spectrometer (Bruker) equipped with a triple resonance z-axis gradient, cryogenic probe at a temperature of 298K. Extremely helpful for the sequential resonance assignment of the protein backbone was the four dimensional HNnoeNH spectra. Visible within these experiment are only backbone amide resonances, so that sequential +1 and -1 amides could be directly identified by comparison with the assignment possibilities from the three dimensional spectra. Additional to the sequential amide NOEs, long range NOEs could be extracted from this four dimensional amide Noesy. The chemical shift resonance of the Gln179 backbone amide is remarkable at 133 ppm in nitrogen dimension and 12.2 ppm in proton dimension.

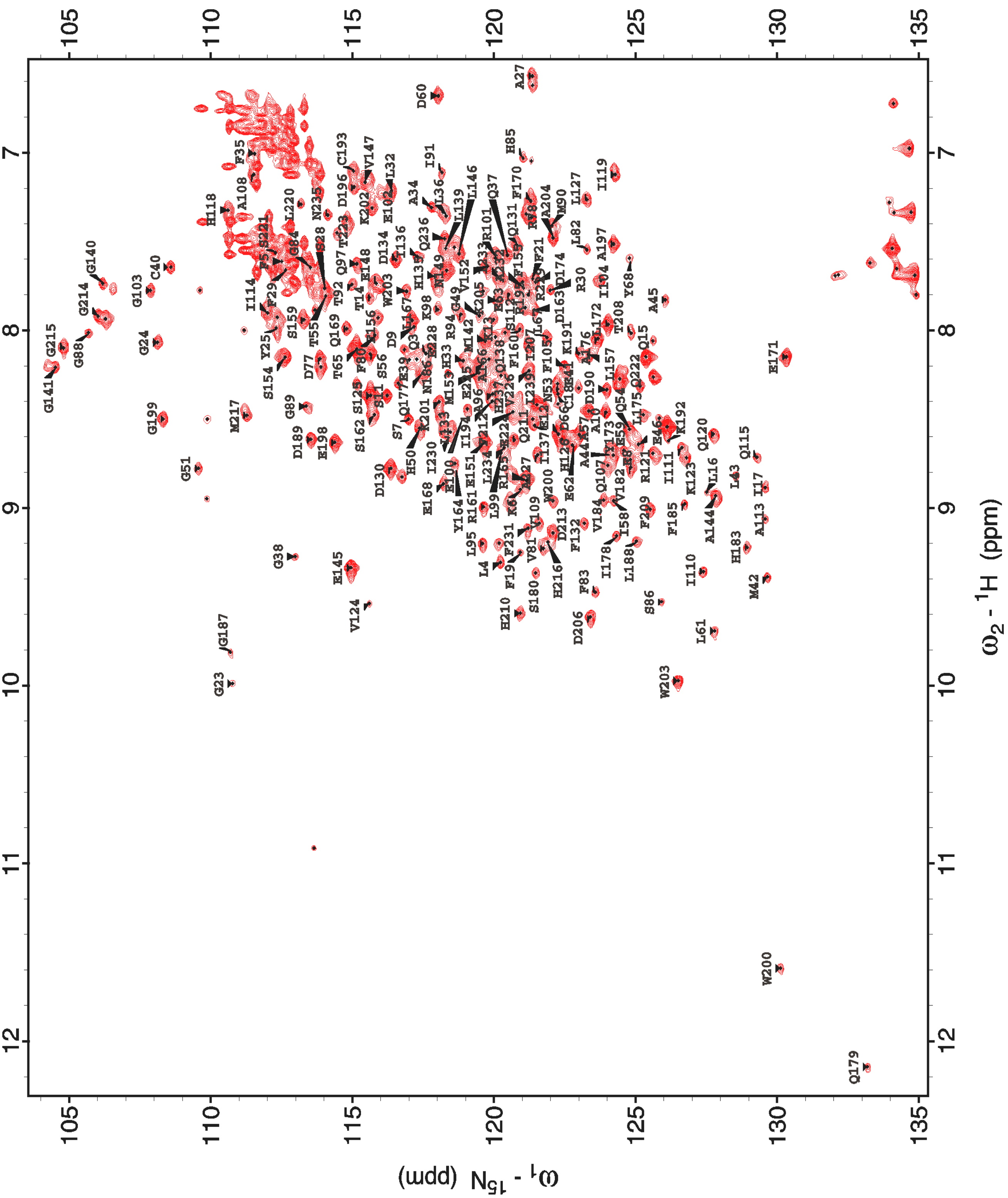


Fig. 31: [^{1}H-^{15}N]-HSQC NMR-spectra of [^{2}H-^{15}N] SrfTEII. The HSQC (heteronuclear single quantum correlation) is a 2 dimensional spectroscopy and is used to detect proton-heteronucleus correlations, here for ^{1}H-^{15}N correlation. This HSQC-spectra of SrfTEII represent a good dispersion of the amids. A good dispersion or rather resolution of the signals and a high intensity is essential for detection in higher dimensional experiments and reliable assignments. This HSQC was recorded using a 1 mM sample of ^{15}N labeled and 70% perdeuterated SrfTEII in 50 mM Na-phosphate pH 6.8 with 20 mM NaCl at a Bruker Avance 900 MHz spectrometer equipped with a cryoprobe at 298 K. The assignment is indicated

The organization of the secondary structural elements of SrfTEII is shown in a diagram plot (Fig. 32) according to the chemical shift index (csi). An index of +1 symbolizes a β-strand fold and an index of -1 predicts α-helical elements along the sequence of the amino acid residues, where every 10^{th} amino acid residue and position is indicated. The resulting secondary structure reflects the consensus of the chemical shift index, based on Cα, Hα, amide-proton and carbonyl resonances of a certain amino acid residue and the TALOS prediction, based on similarity searches in other protein structures. The consensus of both algorithms gives the average fold of the α/β-hydrolase family. Remarkable is the gap in the α-helix IV at position M153, which results from both algorithms and is supported as a kink in the helix by NMR-spectra.

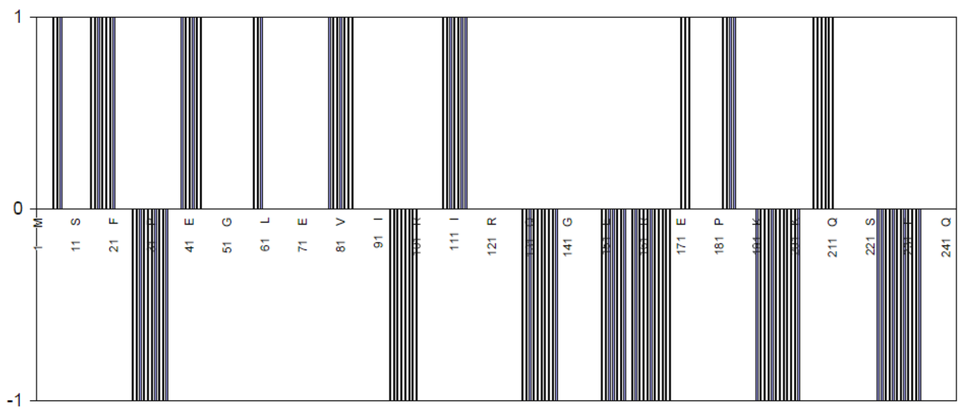


Fig. 32: Predicted secondary structure of SrfTEII. The prediction of this secondary structure reflects the consensus of Chemical Shift Index (csi) and TALOS prediction. It gives an average fold of the a/b-hydrolases family. An index of +1 symbolizes b-sheet fold and an index of -1 predicts an a-helical fold.

The general sequence organization of the secondary structural elements differs significantly from the known structures of related proteins like the Thioesterase I of the *B. subtilis* Surfactin synthetase. An alignment of the amino acid sequences with an additional comparison of the secondary structural elements provides an insight into the variability of the α/β-hydrolase family (Fig. 33). The alignment in Fig. 33 is shown for the sequence of SrfTEII in the upper row and is compared with SrfTEI in the lower row. The secondary structural elements are symbolized as tubes for α-helical stretches and arrows are used to show β-strands. The amino acid residues of the postulated catalytic triad of SrfTEII (Ser86, Asp190, His216) and of the SrfTEI (Ser80, Asp107 and His207) are highlighted in red letters. SrfTEII indicates a loop region from amino acid residue Ala45 up to Asp77, interrupted by 3 residue β-stretch. In comparison the SrfTEI shows within this region a β-strand, from amino acid 41 up to 49, and an α-helix (from amino acid 56 up to 67). For the global orientation of the secondary structural elements, this is a significant difference in the α/β-hydrolase family. The other main difference in comparison of the secondary structures is an a-helix in SrfTEII (Lys191 – Lys201), which is located between two amino acid residues of the catalytic triad (Bruner et al., 2002).

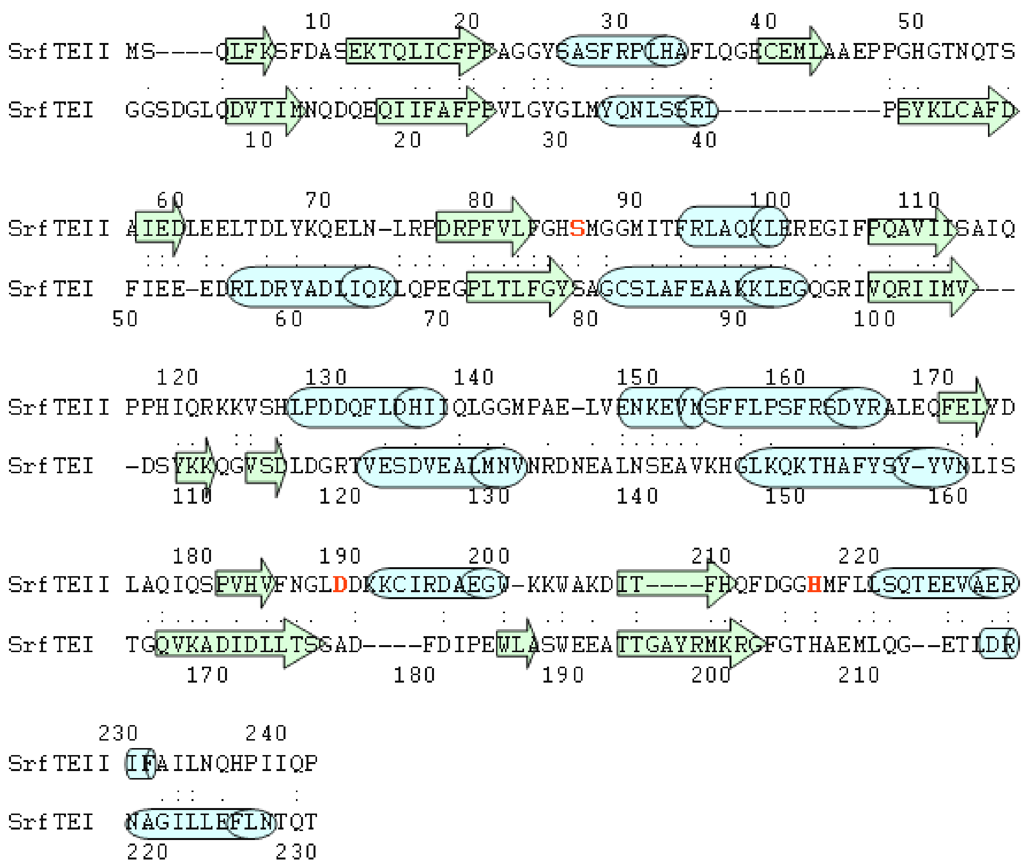


Fig. 33: Alignment of primary sequence and secondary structural elements of SrfTEII and SrfTEI. An alignment of the amino acid sequence of SrfTEII (upper row) and SrfTEI* (lower row) and its secondary structural elements (arrow: b-sheet and tube: a-helix) display significant differences in fold. The amino acid residues (aa) of the postulated catalytic triade of SrfTEII (Ser 86, Asp 190, His 216) and of SrfTEI (Ser 80, Asp 107, His 207) are highlighted in red. SrfTEII indicates a loop region from aa 45 to 78, interrupted by a 3 aa b-fold, instead of a b-fold (aa 41 - 49) and a a-helix (aa 56 - 67) in SrfTEI. The other main difference in secondary structure is an a-helix (aa 191 - 201) in SrfTEII, which is located between 2 residues of the active site (*Bruner et al., 2002).

After completing resonance assignments of the protein backbone and the side chains, analysis of the Noesy spectra enables insight into the orientation of the secondary structural elements by extracting and assigning long range distance information. With ^{15}N- and ^{13}C-edited Noesy-spectra of SrfTEII it becomes possible to describe the global orientation of the secondary structural elements of the protein (Fig. 34). The comparison of the global fold of SrfTEII with SrfTEI is showing the main differences of these two related proteins. The N- and C-termini of both compared proteins are labeled. The β-strands and the α-helices are indicated as blue arrows and orange tubes respectively. The active site residues of the putative catalytic triads are labeled by red stars. It is remarkable that the global fold of these two proteins is, beside the differences in the secondary structure, highly comparable. The obvious difference is the position of the active site residues. The active site of the SrfTEII protein, compared to the SrfTEI, seems to be moved to the C-terminal opening between the protein core, scaffold by the β-sheet, and the two covering α-helices.

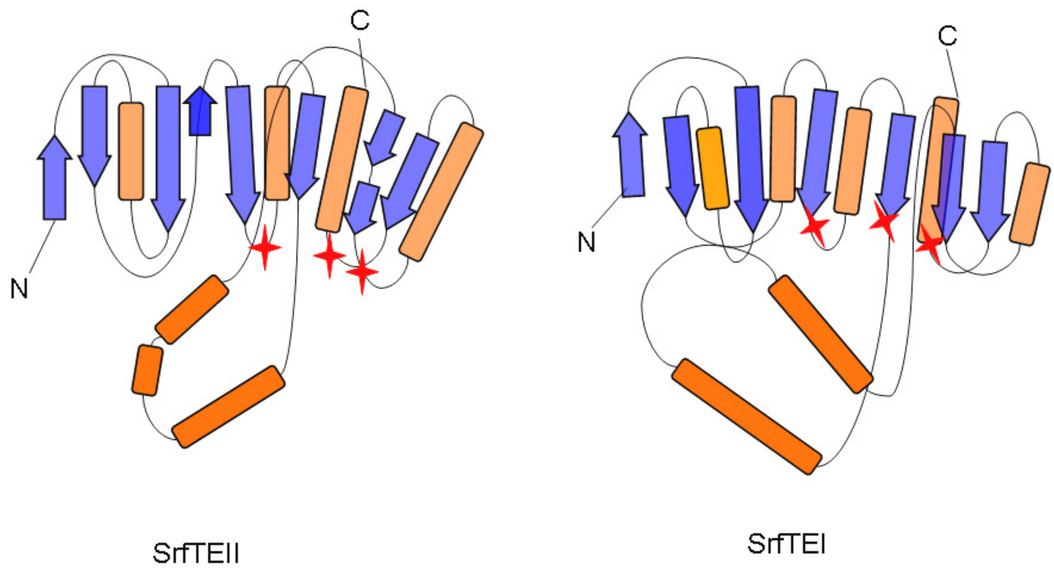


Fig. 34: Comparison organization of the secondary structural elements of SrfTEII (left) and SrfTEI (right side). The sequential organization of the secondary structural elements is different (Fig. 33), but remarkably the global orientation is similar and both proteins are showing the average fold of an α/β-hydrolase. The position of the active site residues to the global fold is slightly different (indicated by red asterisks). N- and C-termini are labeled, β-strands are indicated as blue arrows and α-helices as orange tubes.

For further structural investigation, long range and mid-range NOEs from all recorded Noesy spectra must be extracted and angle constraints need to be defined using the TALOS algorithm (Bonvin et al., 2001; Cornilescu et al., 1999). Due to the size of the SrfTEII protein 86% of all NOEs have more than one assignment possibility and a specific assignment of structurally important ambiguous NOEs is necessary. The four dimensional HNnoeNH spectra was outstandingly helpful to extract unambiguous distance constraints. Based on around 1430 manually assigned NOEs from the ^{15}N-Noesy-HSQC and HNnoeNH spectra and additional 700 NOEs from the ^{13}C-Noesy-HSQC selective for aliphatic sidechains and 420 manually assigned NOEs from the methyl-selective ^{13}C-Noesy-HSQC, 2100 meaningful distance constraints could be generated using the CANDID module of cyana-2.1. Additionally 190 angle constraints for the consensually predicted secondary structural elements and 90 manually defined distance constraints, based on Noesy data, to adjust the protein core were used for structural calculations. All calculations were performed using the automated structural calculation module of the cyana-2.1 algorithm. Each calculation was performed as a set of seven cycles with 50 calculated structures within each cycle. The final output of one run is a set of 15 structural models sorted by lowest energy. The shown preliminary structure of the SrfTEII protein is the averaged fold of one set of 15 models ranked by lowest energy (Fig. 35).

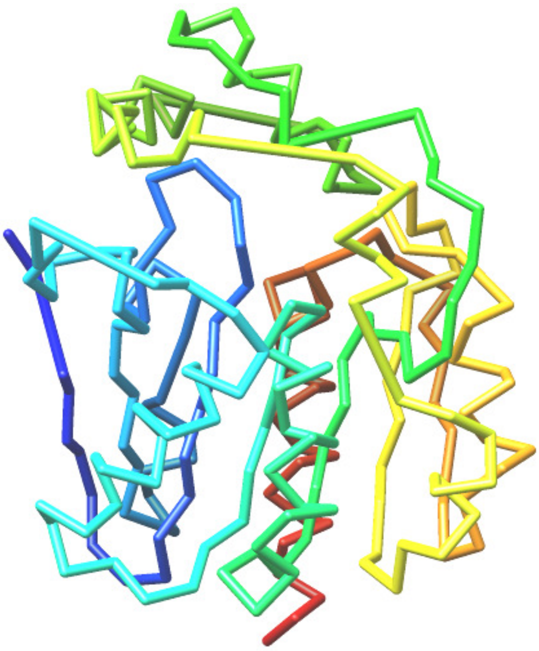


Fig. 35: Preliminary structural model of SrfTEII. The shown averaged model based on Ca traces is derived from 15 cyana structures sorted by lowest energy. Both structural features of α/β-hydrolases, the large β-sheet and the helix/helix-motif are shown. The structure is colored from blue (N-terminus) to red (C-terminus).

The C_α traces are shown only to reduce the structure to the symbolized protein backbone. The model is colored from N-terminus (blue) to the C-terminus in red. The protein core with the scaffold central β-sheet, can be identified and the position of the α-helices are already stable oriented around this β-sheet.

The helix/helix-motif is visible on top of the protein and is separately colored orange in Fig. 36. The stable definition of this structural motif is still a challenge for structural calculations. The position of this helix/helix-motif is not unambiguously defined by the structural relevant Noesy-data. Two positions appear to be possible.

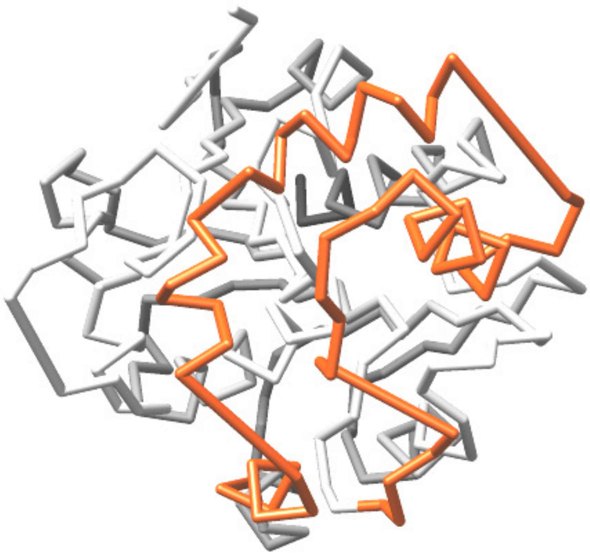


Fig. 36: Structure of SrfTEII. The helix/helix-motif is colored orange at its position on top of the protein core. This part of the protein is showing a high flexibility between two putative extreme states. The second extreme position of the motif is oriented to the left.

Interestingly, the putative two-state dynamic, based on the observation of double peaks in ^{15}N-TROSY-HSQC experiments shown by the SrfTEII (Fig. 37) seems to be even detectable in the averaged model. The region where the double peaks of SrfTEII appear (Fig. 37) are more flexible and are structurally less defined, compared to the rest of the protein and describe the structural dynamic in the regions colored in Fig. 36. Especially the two large helices that cover the protein core seem to be flexible between two defined extreme positions. Some of the double peaks in the ^{15}N-TROSY-HSQC of the SrfTEII are enlarged in Fig. 37.

Unfortunately, the two putative positions of this helix/helix-motif could not be spectrally separated. The non-active mutant SrfTEII$_{S86A}$ showed too many spectral differences, that a direct comparison within the titration experiments to the wild type protein by superimposition of both ^{15}N-TROSY-HSQC spectra is not possible.

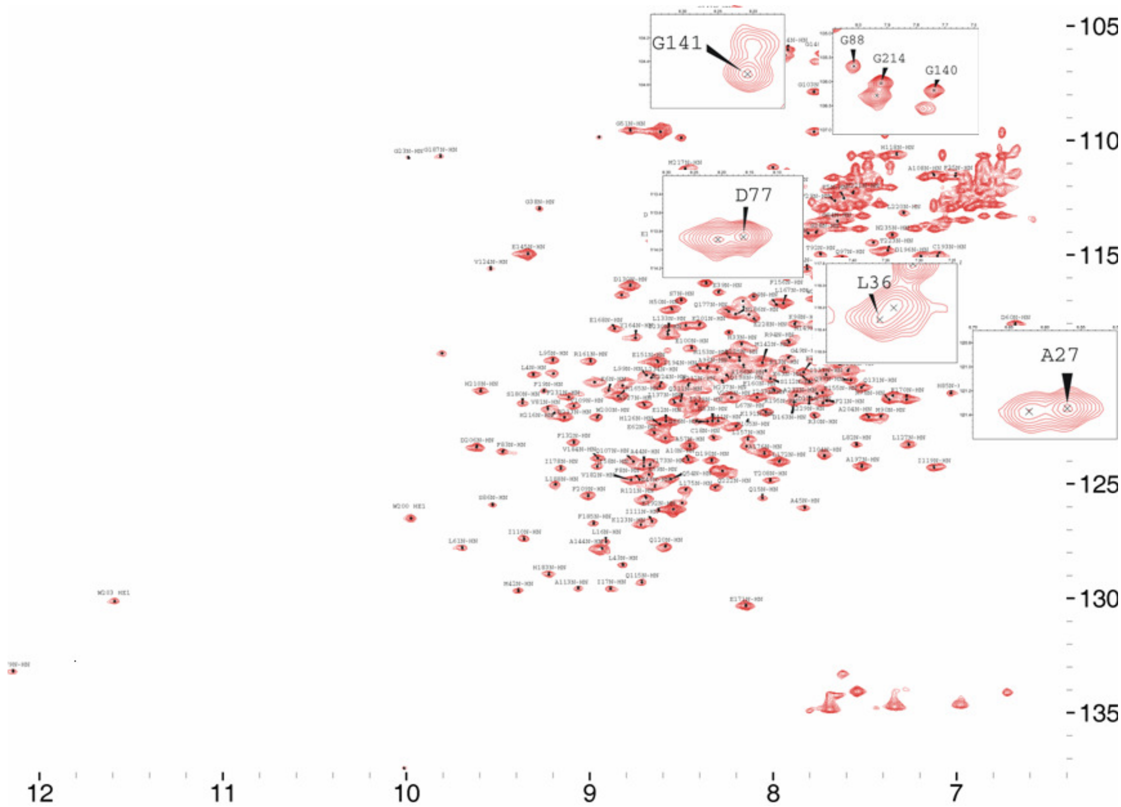


Fig. 37: [^{1}H ^{15}N] HSQC spectrum of perdeuterated ^{15}N- labeled SrfTEII. The SrfTEII is showing double peaks in HSQC spctra, suggesting slow chemical exchange and describe two putative dynamic states. Some of the double peaks are emphasized and labeled separately. This effect is detectable in defined regions of the protein only.

The regions where this chemical exchange happens can be identified by mapping the positions of the double peaks on the secondary structural model (Fig. 38). The positions of the amino acid residues that show slow chemical exchange rates and double peaks in the HSQC spectra are indicated by yellow suns. Based on the preliminary results of the structural calculations and the observation of the defined regions showing slow motion dynamic or slow chemical exchange rates, it can be speculated about the possibility of a well defined two-state movement of the helix/helix-motif on top of the protein core, which might modulate the protein activity and substrate selectivity. This putative two-state dynamic of proteins and domains involved multi-modular polypeptide- and polyketide synthetases seems to be a general feature to specifically modulate the selectivity and activity for the interaction with different reaction partners.

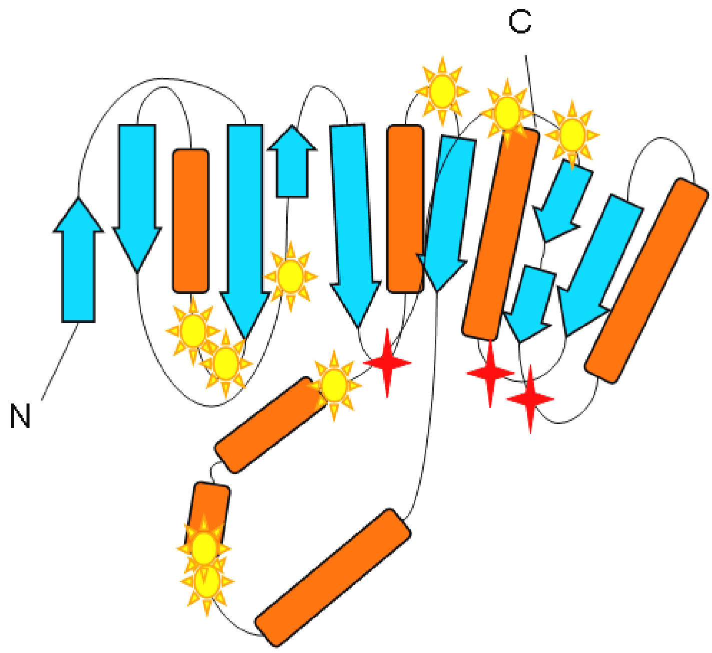


Fig. 38: Double peaks of the SrfTEII. The regions where the double peaks can be identified by mapping the positions of the double peaks on the secondary structural model. The positions of the amino acid residues, which show slow chemical exchange rates and double peaks in the HSQC spectra are indicated by yellow suns. Based on the preliminary results of the structural calculations and the observation of the defined regions showing slow motion dynamic, it can be speculated about the possibility of a well defined two-state movement of the helix/helix-motif on top of the protein core.

Furthermore, the specific interaction of SrfTEII with a separated PCP-domain has been analyzed by ^{15}N-HSQC based NMR-titration experiments. For this interaction-studies the already known and structurally described isolated TycC3-PCP, the third PCP-domain of the tyrocidine-synthetase module C, was chosen. For performing the titration experiments TycC3-PCP was used as the uniformly ^{15}N-labeled and unlabeled holo- or acetyl-holo-form and was titrated to SrfTEII unlabeled and perdeuterated/uniformly ^{15}N-labeled protein, respectively. The apo-form of TycC3-PCP does not show any hint of an interaction with SrfTEII. For the titration experiments of acetyl-holo-TycC3-PCP with the Surfactin thioesterase II, the enzymatic inactive point mutation $SrfTEII_{S86A}$ was used, to prevent the immediate hydrolysis of the acetyl-residue. The general interaction of wildtype SrfTEII with holo-TycC3-PCP is comparable to the interaction with the $SrfTEII_{S86A}$ mutant. Both titration experiments were performed to prove the interaction of the mutant with acetyl-holo-TycC3-PCP to be equal. In Fig. 39 the superimposition of the ^{15}N-HSQC-spectra of uniformly ^{15}N-labeled and perdeuterated $SrfTEII_{S86A}$ protein (red) with the 1:1 complex of [^{15}N ^{2}H]-$SrfTEII_{S86A}$ (blue) with unlabeled holo-TycC3-PCP is shown.

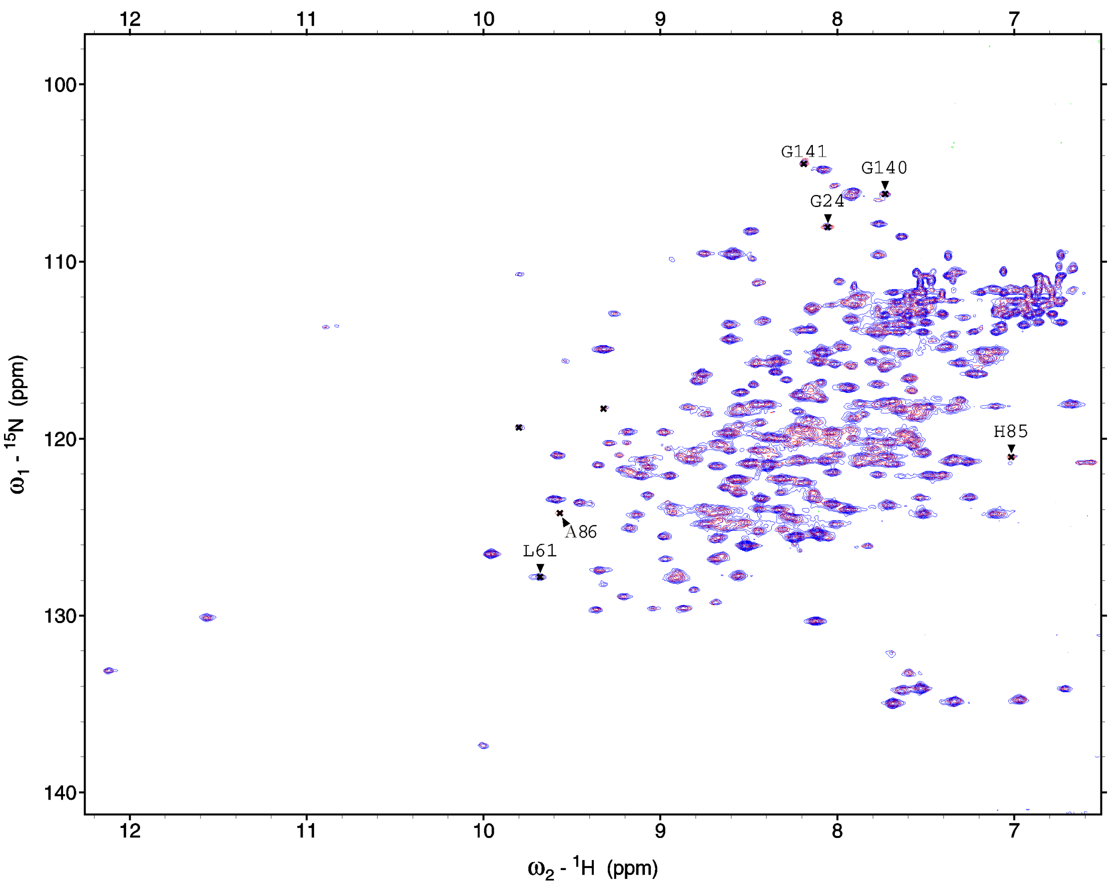


Fig. 39: ^{15}N-HSQC based titration experiment of ^{15}N-labled $SrfTEII_{S86A}$ with unlabeled holo-TycC3-PCP. The titration experiment is showing a weak interaction with holo-TycC3-PCP. Mainly detectable for just a few amino acid residues as indicated in the spectrum. The free ^{15}N-labled $SrfTEII_{S86A}$ is shown in red and ^{15}N-labled $SrfTEII_{S86A}$ in complex with holo-TycC3-PCP is colored blue.

This spectrum is used to have similar data for the titration with acetyl-holo-PCP (Fig. 13). Interestingly the interaction of SrfTEII with holo-TycC3-PCP is very limited and the effected amino acid residues of SrfTEII are indicated. The interaction of acetyl-holo-TycC3-PCP with SrfTEII$_{S86A}$ is shown in Fig. 40. SrfTEII in the 1:1 complex of both proteins is colored green and the involved amino acid residues of SrfTEII are indicated. Interestingly in the complex with acetyl-holo-PCP the SrfTEII is not showing double peaks or the putative two state dynamic.

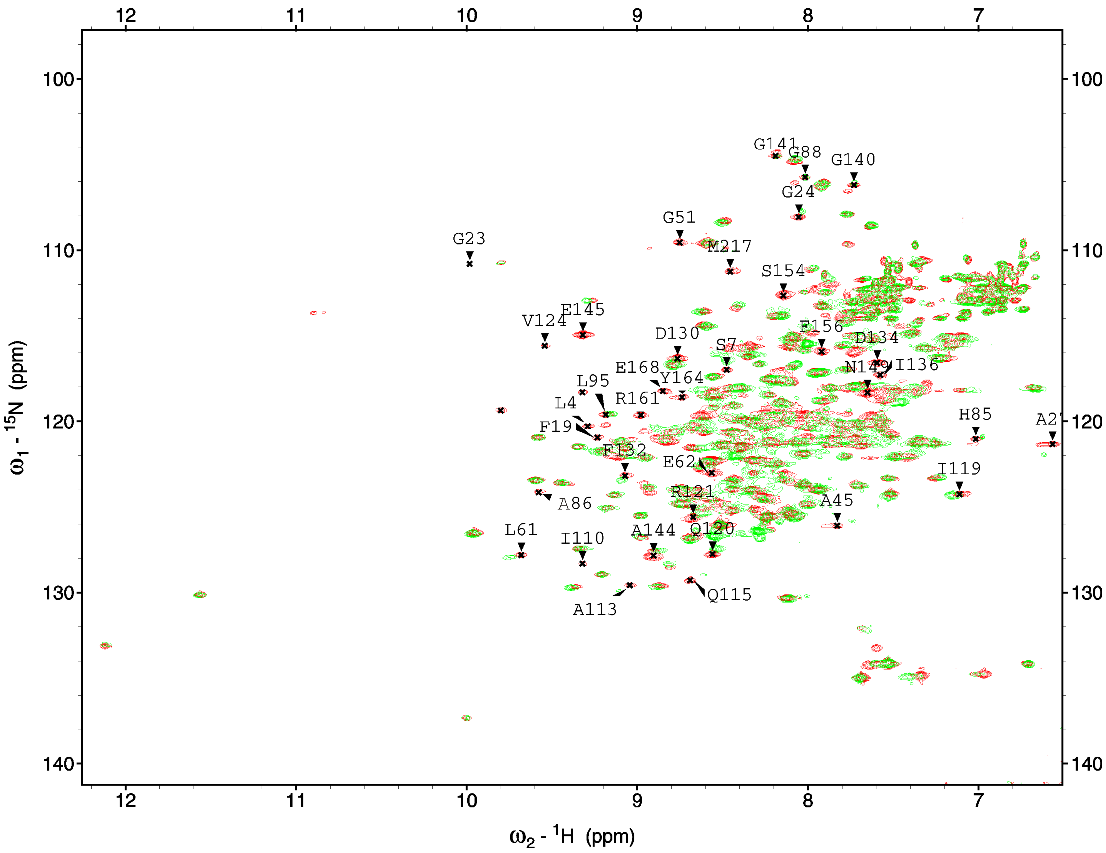


Fig. 40: ^{15}N-HSQC based titration experiment of ^{15}N-labled SrfTEII$_{S86A}$ with unlabeled acetyl-holo-TycC3-PCP. This titration experiment is showing a remarkable different behavior compared to the interaction with holo-TycC3-PCP. Changes of the chemical shifts are detectable for all indicated amino acid residues. The free ^{15}N-labled SrfTEII$_{S86A}$ is shown in red and ^{15}N-labled SrfTEII$_{S86A}$ in complex with acetyl-holo-TycC3-PCP is colored green. For the complex (green) double peaks can not be detect anymore.

Due to the observation that the modification of the 4'-PP cofactor seems to be recognized separately by SrfTEII within the interaction and the two-state dynamic of the thioesterase is frozen in one defined state, a titration experiment of SrfTEII with myristoyl-CoA (Fig. 41) was performed. The free SrfTEII is colored blue and in the myristoyl-CoA bound state is colored red. The isolated and modified 4'-PP-cofactor of PCPs is recognized by SrfTEII, but both putative state of

SrfTEII are showing a similar interaction. For the interaction of PCPs with SrfTEII both, the PCP as well as the modified 4’-PP cofactor, are crucial for the formation of the enzymatic active complex. The effected amino acid residues are labeled.

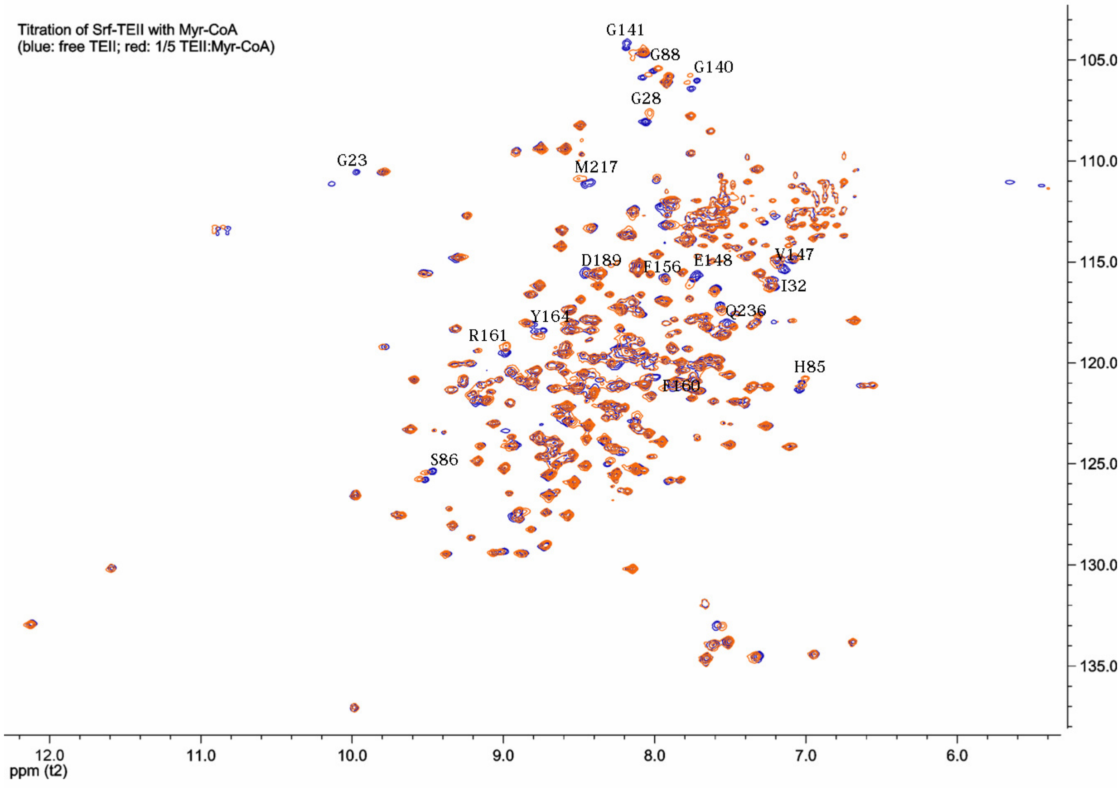


Fig. 41: Titration experiment of SrfTEII with myristoyl-CoA. The free SrfTEII is colored blue and in the myristoyl-CoA bound state is colored red. The isolated and modified 4’-PP-cofactor of PCPs is recognized by SrfTEII. All effected amino acid residues are indicated. In the complex SrfTEII is showing the double peaks and both are effected by the modified and isolated cofactor.

It could be demonstrated that not only the specific signaling amino acid residues of the frozen H-state of acetyl-holo-PCP-domain (see part about the structural analysis of the TycC3-PCP-domain) and the structural dynamic holo-TycC3-PCP by itself are recognized by SrfTEII, but also the 4’-PP-cofactor and its modification plays an important role in the recognition mechanism between these two proteins. Defined structural elements of the PCP domain obviously exhibit significant conformational switches upon modification with the 4’-PP cofactor and loading with the substrate, that presumably coordinate the interaction of the central PCP domains with other domains or enzymes of the NRPS system, like the external thioesterase SrfTEII.

Structural analysis of the *E. coli* telluride transporter TehA

Membrane proteins represent a major challenge in structural biology because they are hard to produce and difficult to become analyzed by high-resolution structural methods. The recent development of protocols for synthesis of membrane proteins in preparative scales using *in vitro* transcription/translation systems (Fig. 42) (Spirin et al., 1988; Klammt *et al.*, 2004) considerably facilitates the preparation of samples suitable for solution NMR spectroscopy.

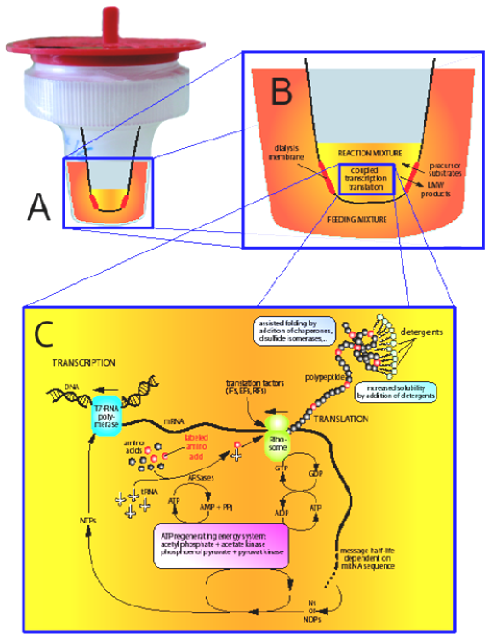


Fig. 42: Analytical set-up of cell-free protein expression. A: For optimization and screening reactions, commercially available microdialyzer units were used. B: The reaction compartment contains approx. 100 µl of reaction mixture separated by a semi permeable membrane from the feeding mixture. C: All high molecular weight compounds necessary for transcription and translation are present in the reaction compartment. Removal of waste products and supply of precursors is accomplished by the exchange of low molecular weight compounds with the feeding compartment. To increase the solubility of membrane proteins, detergents can be added directly to the reaction- and feeding mixture (Klammt *et al.*, 2005).

The bacterial protein TehA, which shows limited homology to the family of small multidrug resistance (sMDR) proteins, is an integral membrane transporter composed of ten putative transmembrane-segments (TMS). Its overexpression confers resistance to telluride compounds, while lipophilic dyes such as ethidium bromide, crystal violet and proflavin have also been reported as substrates (Turner et al., 1997). We have focused on a 28 kDa C-terminal deleted fragment of TehA (ΔTehA), basically retaining the resistance of the full-length protein, which contains seven out of the ten predicted transmembrane helices (Fig. 43).

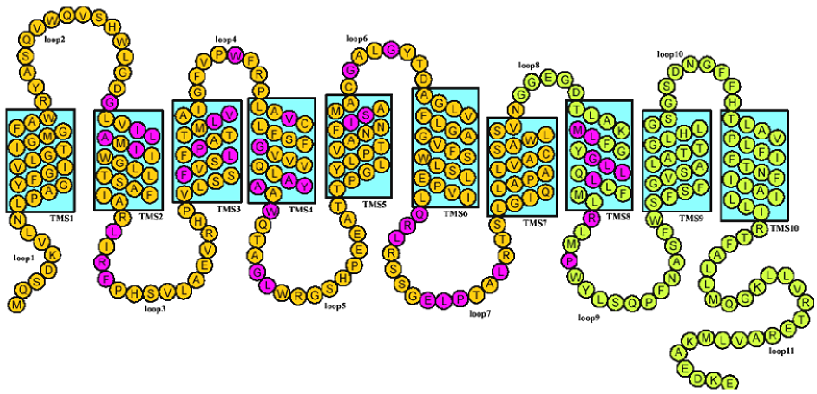


Fig. 43: Proposed secondary structure of TehA. The 10 transmembrane segments (TMS) are shown. The C-terminal truncated version ΔTehA is indicated in orange. Conserved residues related to the family of multi drug-transporters are indicated in purple.

As expected the ^{1}H-^{15}N correlation map of uniformly ^{15}N-labeled and perdeuterated, detergent-solubilized ΔTehA (Fig. 44) shows a poor dispersion for the majority of signals for a folded 28 kDa protein with presumably seven trans-membrane helices, even at the highest currently available field of 18.8 Tesla at the Avance 900 NMR spectrometer equipped with a cryogenic triple axis z-axis gradient probe. Detailed solution NMR studies, including the sequential assignment (shown in Fig. 44), require extensive use of selective isotopic labeling. Cell-free expression is ideally suited for this purpose as it allows convenient and cost-efficient production of such samples virtually in 24h.

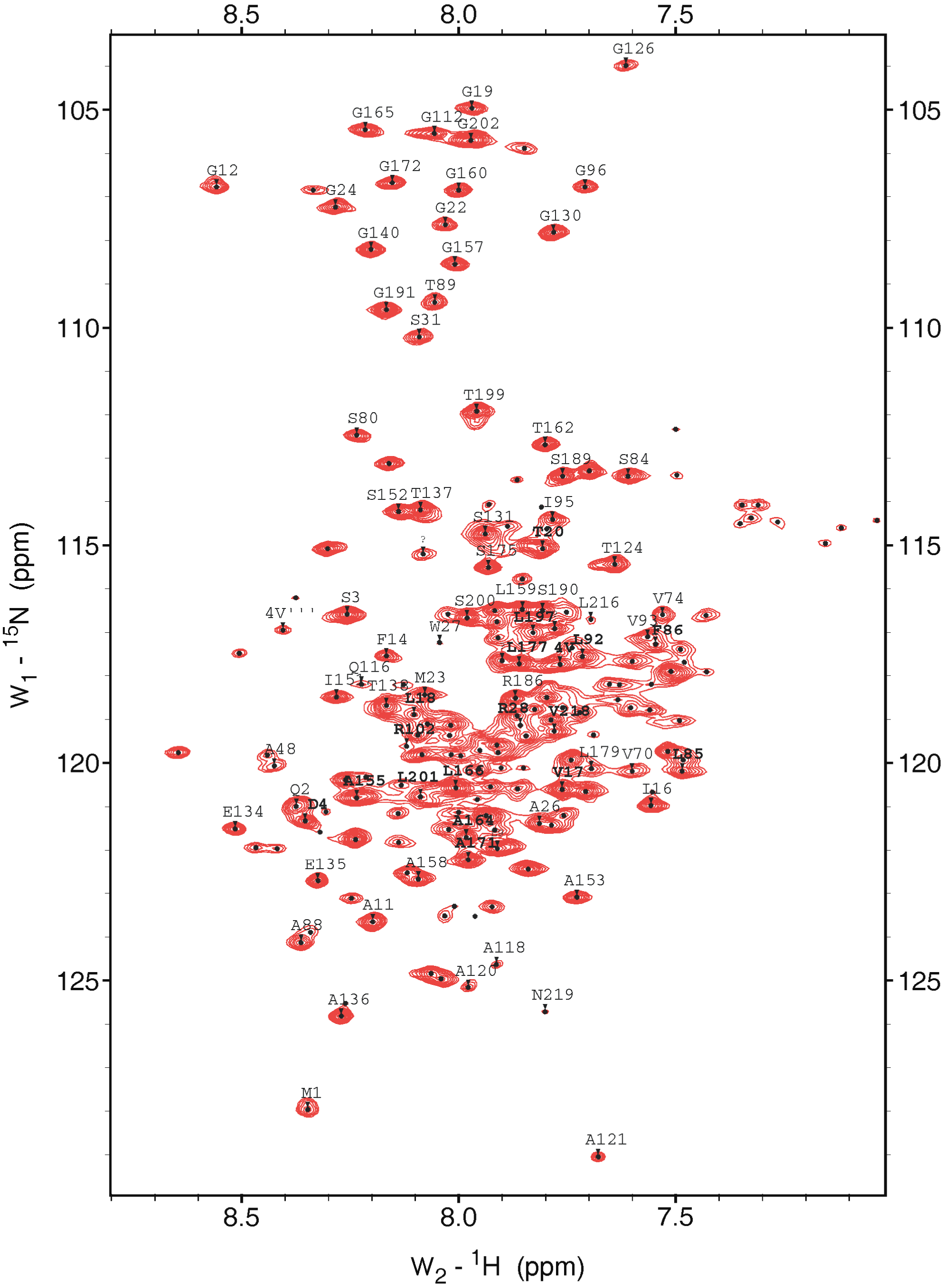


Fig. 44: [^{15}N,^{1}H]-TROSY of ΔTehA. A [*U*-^{2}H;^{15}N]-labelled protein sample containing 5% LMPG was employed. The spectrum was recorded at a 900 MHz NMR spectrometer equipped with cryogenic probe at a temperature of 40 °C. 80% of the protein backbone resonances could be assigned and are indicated by the amino acid residues one-letter code and the corresponding number of the amino acid position.

Furthermore, this method almost completely avoids problems due to isotopic scrambling and considerably improves the quality of the backbone resonance assignment using a combinatorial scheme for amino acid residue types selective labeling (Fig. 45) (Trbovic et al., 2005).

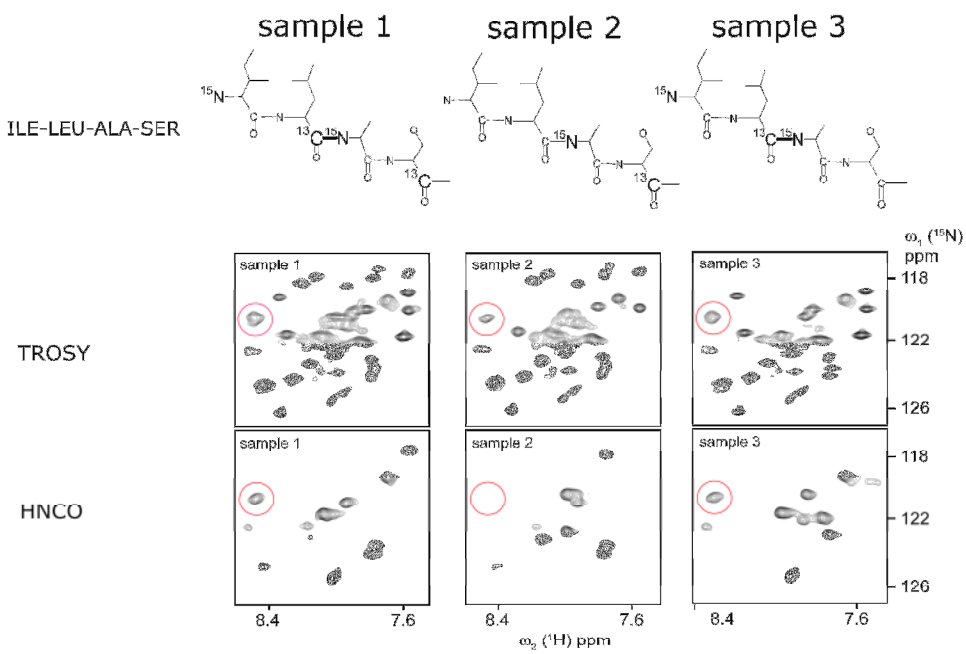


Fig 45: Combinatorial labeling of ΔTehA. Two-dimensional [^{15}N,^{1}H]-TROSY and [^{15}N,^{1}H]-TROSY-HNco spectra were recorded at 800 MHz. As an example, a cross peak at the circled position is present in each of the TROSY spectra, indicating that it belongs to an alanine residue. The preceding residue must be a leucine because a HNCO peak is observed for samples 1 and 3, but not for sample 2. Since the pair leucine-alanine is unique in TehA (47-48), the amide resonances of Ala48 were thus unambiguously identified (Trbovic *et al.*, 2005, modified).

Uniformly ^{15}N-, ^{13}C/^{15}N-, ^{2}H/^{15}N- and ^{2}H/^{13}C/^{15}N-labelled ΔTehA as well as a variety of selectively labeled protein samples containing up to three ^{15}N- (and in some cases $^{13}C_{1}$-) amino acid types, have been resolubilized in 1-myristoyl-2-hydroxy-sn-glycero-3-[phospho-rac-(1-glycerol (LMPG) micelles. Standard three-dimensional triple resonance experiments, i.e. HNCA, HNcoCA, HNCACB, HNcoCACB, HNCO, HNcaCO, and a four-dimensional $HNCO_{i-1}CA_{i}$ (all TROSY-type) allowed the assignment of 55% of the proteins backbone resonances in a relatively straightforward manner. Unfortunately, selective ^{15}N labeling of 10 different amino acid types (W, A, V, T, S, E, M, L, I, F) only provided additional 10% of unambiguous assignment. Therefore a simplified version of a combinatorial labeling scheme (Parker et al., 2004; Shi et al., 2004) was employed as well. It involves preparation of three samples with different combinations of ^{15}N- and $^{13}C_1$- labeled amino acids (as symbolized in Fig. 45). The amino acid types used here were carefully chosen to provide the highest possible information content for the previously unassigned stretches in ΔTehA. The presence or absence of cross peaks in TROSY and 2D HNco spectra (Fig. 45) for each of the samples identifies pairs of adjacent amino acid types. When such pairs are unique in the protein sequence, residues are unambiguously assigned and can serve as starting points for further sequential connectivities. For TehA, this strategy finally brought the total backbone assignment to 85% and enables the calculation of a preliminary model for TehA (Figs. 46-48) using the dyana 1.05 algorithm (Herrmann et al., 2002).

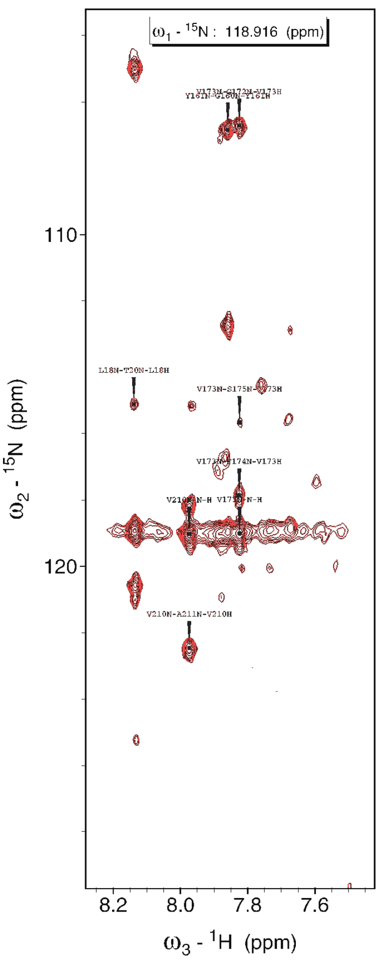


Fig. 46: Single plane from 3D-NhN-NOESY of TehA. Shown are backbone amide-amide contacts as NOEs between coupling amides. The spectrum was recorded at 900 MHz using a perdeuterated uniformly ^{15}N-labeled sample. According to the contacts from 3D-NH-NOESY-HSQC, the 3D-NhN-NOESY verifies the secondary structural elements and defines inter-helical contacts. Additional long-range distance constraints to define the global orientation of the helices could be derived from a 4D-NN-Noesy also recorded using a perdeuterated ^{15}N-labeled sample of TehA.

Since unambiguous long-range NOEs are limited NH based 3D and 4D NOESY spectra, paramagnetic relaxation enhancement (PRE) was employed to estimate inter-helical distances (Fig. 47).

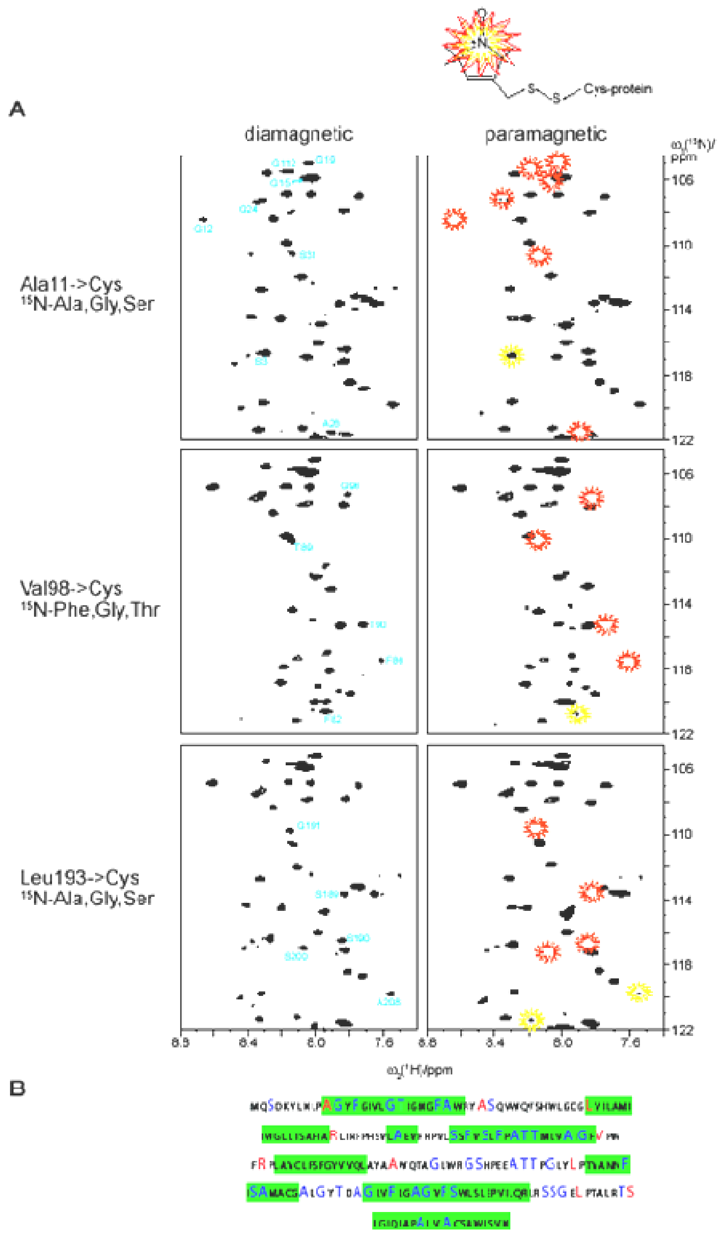


Fig 47: PRE studies on ΔTehA. A: Representative TROSY spectra (800 MHz) of three selectively ^{15}N-labelled mutants, containing a single cysteine residue with MTSL (left column: reduced state-diamagnetic; right column: oxidized state-paramagnetic) covalently attached are shown. Annotated peaks are either completely bleached out (colored red) or considerably broadened (colored yellow) in the spectra of the oxidized species due to their proximity to the spin-label. **B**: Amino acid sequence of TehA with helical segments as determined by secondary chemical shifts and NOEs marked in green. Red letters indicate positions of residues exchanged to cysteins in individual mutants. Distance constraints from PRE have been derived for residues printed in blue for each of the mutants. (Löhr, 2006)

For site directed spin-labelling with the nitroxide (1-oxyl-2,2,5,5-tetramethyl-3 pyrroline-3-methyl)methane- thiosulfonate (MTSL) mutants were generated where either one of the three native cysteine residues was retained or all three were replaced by alanine and another residue was mutated to a cysteine. Each NMR sample was selectively ^{15}N-labelled with three amino acid types to minimize overlap. TROSY spectra were recorded for both the oxidized (paramagnetic) and reduced (diamagnetic) states of MTSL and distance information was extracted from intensity differences of individual cross peaks in the two spectra. Using the PRE technique approx. 120 meaningful distance constraints as a pair of upper and lower limit for an individual Amide-Amide distance could be defined. Additionally, 205 distance constraints are based on unambiguous NOEs from 4D- NN- NOESY and 3D NhN-Noesy spectra (Fig. 46) to define inter-helical distance constraints. Based on the derived data a preliminary backbone model of ΔTehA could be obtained (fig. 48). The model indicates the existence of seven helices, where one helix (helix 4, colored green) seems to be flexible and not transmembrane oriented.

Fig. 48: Preliminary model of Δ-TehA as derived from dyana 1.05 simulated annealing protocol. A first model of a protein with seven-transmembrane helices could derived from 420 distance constraints using dyana 1.05 simulated annealing in vacuum. This calculated model symbolizes the proposed averaged topology of TehA in LMPG micelles. N-terminus (blue) and C-terminus (red) are oriented on opposite sites of a assumed conformational space of a membrane.

Imidazole-glycerol phosphate synthase

The imidazole glycerol phosphate (IGP) synthase is structural and functional part of the family of the glutamine-amidotransferases (GAT) (Carlomagno et al., 1988). Amidotransferases are enzymes that are involved in the biosynthetic pathways of several essential amino acids and biologically important cofactors and play an important roll of incorporating nitrogen into heterocyclic compounds like amino acid residues, purine- and pyrimidine nucleotides, amino sugars, coenzymes and antibiotics (table 1). Glutamine-amidotransferases carry a glutaminase function to catalyze reactions using the sidechain amide group derivative from glutamine. These enzymes are able to use free NH_3 in higher concentrations as well (Zalkin, 1993).

Table 1: Enzymatic pathways of amidotransferases

enzyme	biosynthetic pathway
Anthranilate-Synthase[1]	TRP
Carbamoyl-P-Synthase[1]	ARG, UTP, CTP
Aminodesoxychorismate-Synthase[1]	folic acid
CTP Synthetase[1]	CTP
FGAM-Synthetase[1]	AMP, GMP
GMP-Synthetase[1]	GMP
IGP-Synthase[1]	HIS
Glucosamine 6-P Synthase[2]	Glucosamine 6-P
Glutamate-Synthase[2]	GLU
Glutamine-PRPP-Amidotransferase[2]	AMP, GMP
Arylamine-Synthetase	chloramphenicole
GLU-tRNA^{Gln}-Amidotransferase	GLU-tRNA^{Gln}
NAD-Synthetase	NAD
Cobyrin acid a,c-diamid-Synthetase[1]	Coenzyme B_{12}

GATs are ordered by homology to the *purF* coded glutamine-PRPP-amidotransferase, for the biosynthesis of AMP and GMP (F-type; table index [1]), and the *trpG*-homologue (tryptophan biosynthesis) G-type (table index [2]). Especially the G-type GAT domains are evolutionary highly conserved over a wide range of different organisms like *Bacillus subtilis* (Kunst et al., 1997) and including human pathogens *Vibrio cholerae* (Heidelberg et al., 2000) and *Mycobacterium tuberculosis* (Cole et al., 1998).

The histidine biosynthesis of *E. coli* is genetically organized within one single operon. The genetic sequence of the cistronic genes within the operon is *hisL, hisD, hisC, hisB, hisH, hisA, hisF* and *hisIE* and the enzymatic sequence for synthesizing histidine from the substrates ATP and D-ribose-5-phosphate is shown in Fig. 49. The expression of the genes from the *his*-operon is regulated by three promoters. The main promoter and two distal weaker promoters are controlled by feedback-inhibition, by polycistronic mRNA (Carlomagno et al., 1988; Wagner, 2001), cell metabolism and attenuation of transcription to regulate the level of histidine biosynthesis (Kasai, 1974) highly precisely to prevent an intoxication by an accumulation of the cytotoxic precursor PRFAR (Murray & Hartmann, 1972; Winkler 1987).

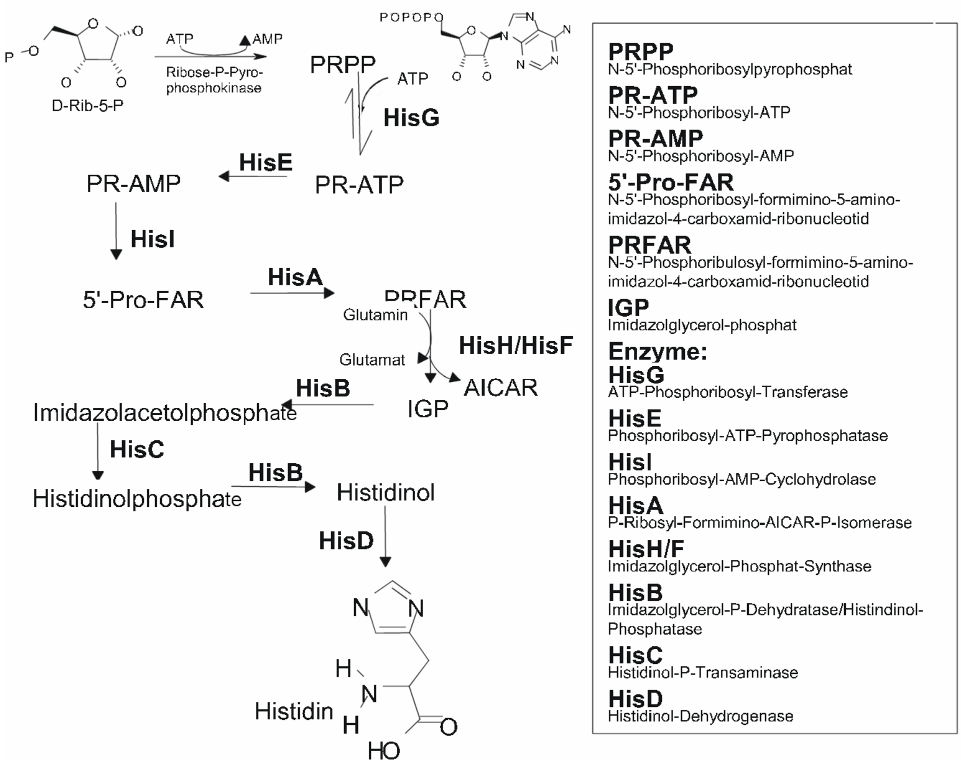


Fig. 49: Biosynthetic pathway for the essential amino acid histidine. The nine step reaction-sequence is coupled to a high energy consumption and starts with phosphor-ribosyl-pyrophosphate (PRPP) and ATP. All involved enzymes are labeled and intermediates and enzymes are described on the right side. (Carlomagno et al., 1988)

The imidazole glycerol-phosphate (IGP-) synthase from *E. coli* is a heterodimeric protein composed of the geneproducts from *hisH* and *hisF* of the *his*-operon and synthesizes the fifth step of the histidine biosynthesis (Fig. 49). The products of the heterodimeric enzyme are essential for the histidine biosynthesis and the second product is used in *de novo* purine biosynthesis. The IGP-synthase transfers an ammonium-group from glutamine to the cytotoxic nucleotide substrate N-5-phosphoribulosyl-formimino-5-aminoimidazol-4-carboxamid-ribonucleotide (PRFAR) and catalyzes the formation of 5-aminoimidazole-4-carboxamide-ribonucleotide (AICAR) and imidazole glycerol phosphate (IGP) the virtual product of these reaction to synthesize the essential amino acid histidine (Fig. 50) (Klem et al., 2001).

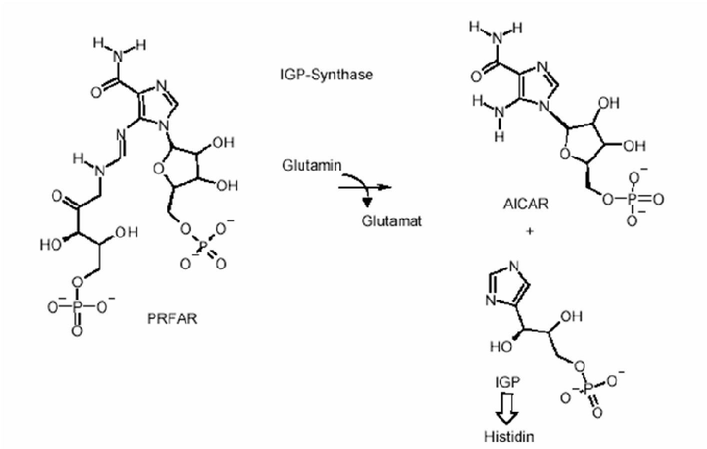


Fig. 50: Imidazole formation reaction of the Imidazole glycerol phosphate synthase (IGP-synthase). The substrate of the reaction N-5-phosphoribulosyl-formimino-5-aminoimidazol-4-carboxamid-ribonucleotide (PRFAR) is split by binding the NH_3 of Glutamine for the formation of the imidazole ring in IGP and 5-aminoimidazole-4-carboxamide-ribonucleotide (AICAR), which is used for *de novo* purine synthesis.

The IGP-Synthetase is a 454 amino acid residues protein with a molecular weight of 50.112 Da. The glutaminase activity of the assembled holo enzyme without nucleotide substrate binding is around 0.8% of the activity after PRFAR binding and glutamine is a 10^3 times more efficient substrate compared to pure ammonia. This dual substrate activation of the IGP synthesis is remarkable because the two substrate binding sites are separated on different proteins and the binding site of the free NH_3 is identical with the glutamine generated ammonia residue. Interestingly the activity of the IGP-synthase can be improved by a factor of 30-times when bound with the pre-substrate 5'-Pro-FAR or the product IGP compared to native concentrations and the reaction rate between synthesis and glutamine hydrolysis is kept to a ratio of 1:1. No intermediate states of this reaction could be described up to now, especially no transition state of the nucleotide substrate could be detected. The concerted reaction of the HisH-glutamine amidotransferase and of the hisF coded cyclase and its mechanism is not known.

X-ray crystallography of homologue glutamine amidotransferases from halothermophilic bacteria (Beismann-Driemeyer & Sterner, 2001) and kinetic studies of the carbamoyl-phosphate synthetase (Thoden et al., 1997) caused evidences for an up to 90 Angstrom water-free channel between the GAT-domain binding site and the active center of the cyclase domain (Bera et al., 2000; Douangamath et al., 2002). The published crystal-structure of the *T. maritima* imidazole glycerol phosphate synthetase (PDB code: 1gpw; Beismann-Driemeyer & Sterner, 2001) suggests the existence of this channel (Fig. 51). The HisH protein of the active complex is colored in blue and the HisF protein in red.

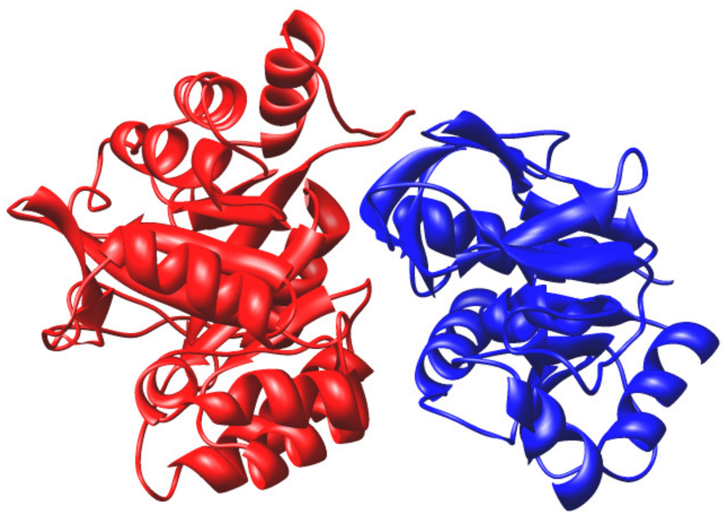


Fig. 51: Crystal structure of the heterodimeric *T. maritima* Imidazole glycerol-phosphate synthase. PDB code: 1gpw (Beismann-Driemeyer & Sterner, 2001). The HisF protein is colored red and the HisH is colored blue.

The *E. coli hisH* gene codes for a *trpG*-homologue glutamine amidotransferase, which hydrolyzes glutamine to glutamate and ammonia (Klem & Davisson, 1993). The released NH_3 is used by a separated synthase. The HisH protein of the IGP-synthase is a 196 amino acid residue protein with a molecular weight of 21.655 Da and expressed as a separated protein. For its enzymatic activity its hetero-oligomersation with HisF is essential (Klem & Davisson, 1993). The HisH monomer is described as an inactive protein with no isolated function. The glutamine hydrolysis and even the substrate binding is described to be dependent on the hetero-dimerisation with HisF. The binding of the intrinsic substrate PRFAR affects the glutaminase activity of HisH additional. The active site residues for dimerization (for HisH from *E. coli* Cys77, His97, Glu99) are highly conserved within all trpG-homologue GATs and the complete sequence of the GAT-domains of the IGP-synthase in a wide range of protozoa/monads, including a vast number of human pathogenic strains, is nearly identical. The structural model of HisH (Fig. 52) is colored from N-terminal (blue) to C-terminal (red) and the conserved amino acid residues of the active dimerization site are shown as stick models and the amino acid type is labeled.

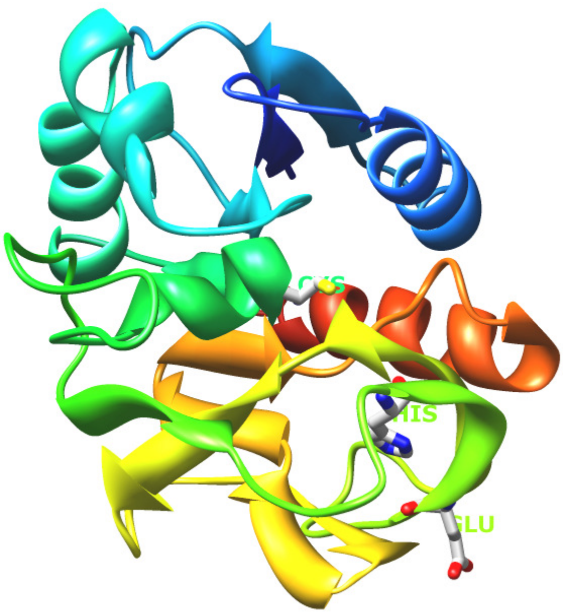


Fig. 52: The crystal structure of HisH is colored from N-terminal (blue) to C-terminal (red) and the conserved amino acid residues of the active dimerization site are shown as stick models and the amino acid type is labeled.

Interestingly the dimerization site is located on one β-strand and the conserved cysteine is separated and partially covered in the protein core.

The E. coli *hisF* gene codes for a 28.457 Da cyclase with 259 amino acid residues and is a member of the $(\alpha\beta)_8$-barrel family of proteins and synthases (Klem et al., 2001). The NH_3-dependent activity of HisF is specific for its nucleotide-substrate PRFAR. PRFAR is the product of the first four biosynthetic steps of the histidine synthesis and is derived from covalent modification of a D-ribose-5-phosphate modified to N-5-phospho-ribosyl-pyrpphosphate with ATP, where the purine residue of the ATP was opened. The enzymatic activity of the monomeric HisF with free NH_3 is decreased by the factor of $\frac{1}{10^3}$ compared to the glutamine dependent activity of the heterodimeric IGP-synthase. This suggests that the specific activity of the cyclase is modulated by hetero-dimerization with HisH. According to the missing reactivity of isolated HisH, mutational analysis demonstrated the importance of highly conserved amino acid residues from the HisF cyclase for the glutaminase activity (Klem et al., 2001). The cyclase HisF is highly conserved in a wide range of bacteria and the amino acid residues Arg5, Glu46, Gln123 and Val126 of the *E. coli* HisF can be found on nearly identical position in several IGP-synthases and play a crucial role for the glutaminase activity of HisH.

The simplified backbone model with enlarged α-helices and β-strands (Fig. 53) of the HisF model from the *Thermotoga maritima* IGP-synthase crystal structure demonstrates the $(\alpha\beta)_8$-barrel fold of HisF in the opened confirmation. The inner core made of 8 β-strands is surrounded by 8 α-helices and the model is colored from N-terminus (blue) to C-terminus (red). The essential amino acid residues for dimerization and control of the glutaminase function of HisH are shown as stick models and named by residue type. These amino acid residues are located within one region of the inner core.

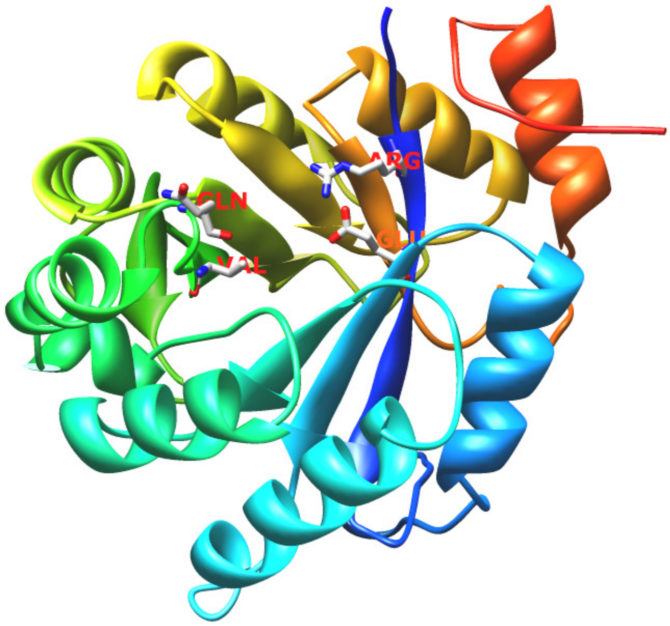


Fig. 53: The backbone model of the HisF protein from the *Thermotoga maritima* IGP-synthase crystal structure demonstrates the $(\alpha\beta)_8$-barrel fold of HisF in the opened confirmation. The inner core made of 8 β-strands is surrounded by 8 α-helices and the model is colored from N-terminus (blue) to C-terminus (red). The essential amino acid residues for dimerization and control of the glutaminase function of HisH are shown as stick models and named by residue type. These amino acid residues are located within one region of the inner core.

Based on the observations that the glutaminase activity of HisH is dependent on interaction with HisF and the increased substrate (PRFAR) binding probabilities of HisF effected by HisH interaction the following mechanism was described (Fig. 54) (Klem & Davisson, 1993; Beismann-Driemeyer & Sterner, 2001).

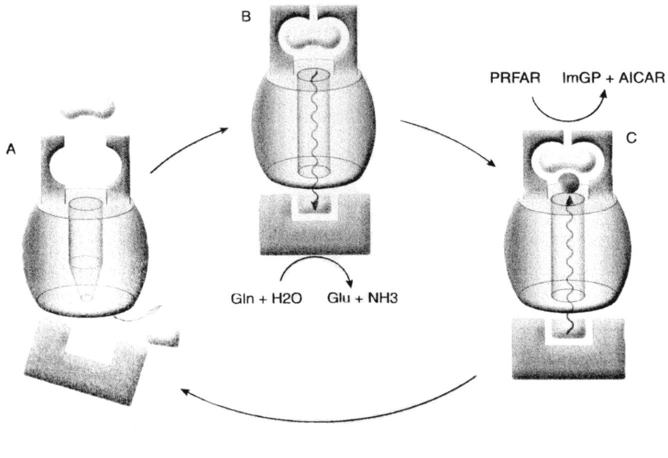


Fig. 54: Model for the reactivity of the IGP synthase. The substrate binding of HisF (barrel like) and HisH (presented as an u-iron like drawing) is dependent on dimerization (A). The glutaminase activity and glutamine binding are directly correlated to the HisF interaction (B) and enzymatic activity of HisF is controlled by the ammonia-transfer (C).

The substrate binding of HisF (barrel like) and HisH (presented as an u-iron like drawing) is dependent on dimerization (A). The glutaminase activity and glutamine binding are directly correlated to the HisF interaction (B) and enzymatic activity of HisF is controlled by the ammonia-transfer (C).

Methods. Genomic double frame shift mutations of a type *+1/-1 313* where performed in the Institute of Microbiology of Frankfurt University using genetic transduction. The double frame shift mutant PU102 *(HisH 313)* is a 9 amino acid residues frame shift within HisH between amino acid residue Val73 and Gln81, respectively PU107 (*HisH 313*) is five amino acid residues frame shift mutant between the amino acid residues Leu74 and Leu78. The double frame shift mutant in HisF and HisH PiWi01 (*HisH313,HisF313*) was constructed based on PU107.

The hisF gene from *E. coli* has been cloned into the mhr-vector pMS470∆8 and was transformed into *E. coli* strain XL10. For expression of HisH, the gene was cloned into the 6His·tag vector pQE30 (Qiagen) and co-transformed with the chaperone-expression system CL33 into strain XL10. Unlabeled proteins were expressed in bacteria grown in LB media. For expression of the labeled proteins ^{15}N, ^{15}N and ^{2}H or ^{15}N, ^{13}C and ^{2}H enriched M9 media (Sambrook et al., 2001) was used. Both protein domains were purified to homogeneity.

To analyze the structure and function of IGP synthase a combination of circular dichroism (CD), fluorescence spectroscopy and Nuclear Magnetic Resonance (NMR) techniques was used. The CD experiments were performed at the Institut für Biophysik of the Universitätsklinikum Frankfurt together with Dr. T. Link. Fluorescence and NMR spectroscopy were performed at the Institute for Biophysical Chemistry. Fluorescence experiments were carried out at 10°C on a SLM Aminco AB2 spectrometer. An excitation wavelength of 292 nm was used and the maximum emission was determined at 345 nm.

NMR experiments are used for sequential backbone assignment of HisF and dynamic studies of the HisH/HisF complex formation. These experiments are performed at 290 K on Bruker DMX-600 or DMX-800 spectrometers.

For NMR titration experiments of monomeric and dimerized HisF with the reaction product imidazole glycerol phosphate (IGP) and 5-aminoimidazole-4-carboxamide-ribonucleotide (AICAR purchased from Sigma-Aldrich), IGP was synthesized by stirring reaction of D-ribose-5-phosphate and formamidine at 60°C for 10h and purified using the chelator Dowex-10. The purity was verified by thin-layer chromatography and NMR-spectroscopy.

Results. The growing rates of the genomic double frame shift mutations in HisH PU102 and PU107 on fully supplemented MiA minimal media (DIFCO manual) with added histidinol as a direct histidine precursor, adenosine and nicotinamide was reduced in both cases by a factor of 1/7 compared to the wild type *E. coli* FP123. This significant effect could also being shown for the double frame shift mutation in HisH and HisF PiWi01 as well (Fig. 55).

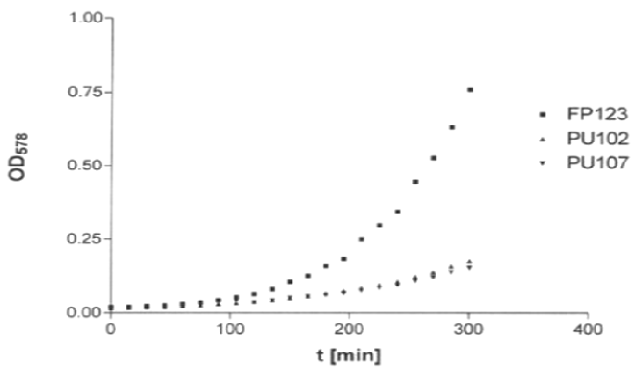


Fig. 55: Growing rates of the genomic double frame shift mutations in HisH PU102 and PU107 on fully supplemented minimal media with added histidinol was reduced in both cases by a factor of 1/7 compared to the wild type *E. coli* FP123.

Both enzymatic subunits were expressed in separate bacterial systems and could be purified to homogeneity. The association and *in vitro* dimerization of the two subunits into the intact enzyme IGP-synthase was followed by Blue-Native-PAGE to verify the stability of the glutamine free complex of HisH and HisF, by native PHAST-gel (Amersham-Pharmacia) (Fig. 56) and isoelectric focusing PHAST-gel electrophoresis (result shown in Fig. 57). The dimerization could be also demonstrated using fluorescence spectroscopy and by NMR spectroscopy for uniformly ^{15}N-labeled HisF in a titration experiment with unlabeled HisH without and with added glutamine in equal molar concentration to HisH as well (Fig. 62).

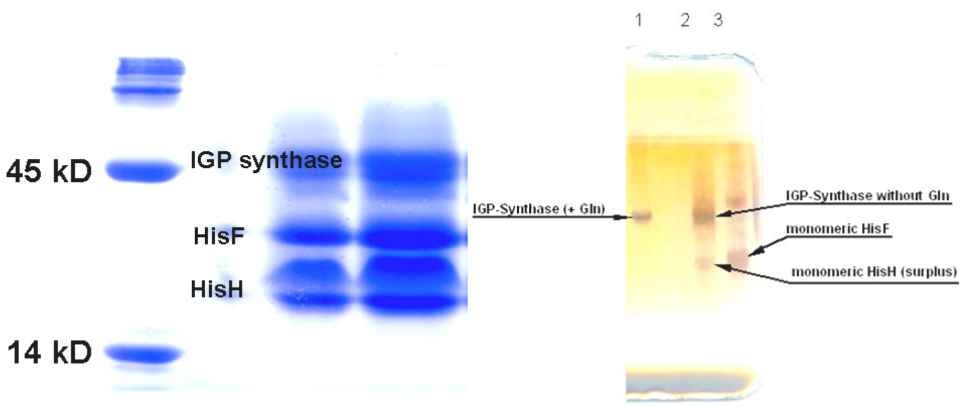


Fig. 56: Association and *in vitro* dimerization of the two subunits into the intact enzyme IGP-synthase shown by Blue-Native-PAGE (left side) to verify the stability of the glutamine free complex of HisH and HisF and by native PHAST-gel (Amersham-Pharmacia) on the right side. The protein marker to estimate the size and the lines are labeled.

Due to dimerization followed by isoelectric focusing the pI of the complex, shown in Fig. 57 line 1 (around pI 4.7) was shifted compared to the averaged pI of the isolated monomers HisF (lane 2) and HisH (lane 3) which were determined experimentally to be pI $_{(HisF)}$ 6.4 respectively pI $_{(HisH)}$ 6.0. This shift of p_i caused by dimerization provide a first hint of possible structural changes within the complex formation independent from substrate binding.

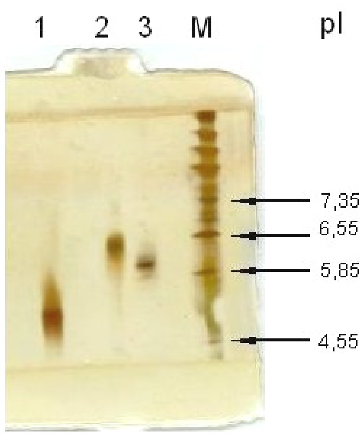


Fig. 57: Dimerization of HisH and HisF shown by isoelectric focusing the pI of the complex (line 1 (around pI 4.7)) was shifted compared to the averaged pI of the isolated monomers HisF (lane 2) and HisH (lane 3) which were determined experimentally to be pI $_{(HisF)}$ 6.4 respectively pI $_{(HisH)}$ 6.0.

The specific binding of glutamine to HisH resulted in considerable changes in secondary structure to the isolated HisH-domain, which could be shown using CD and fluorescence spectroscopy. A decrease in the α-helical portion of this domain was found using CD experiments compared to glutamine-free and glutamine-bounded forms (Fig. 58). The a-helical portion was reduced by 25 – 30% and was saturated at equimolar concentration. A further change in secondary structure could not be observed for a 10-times exceeded concentration of glutamine.

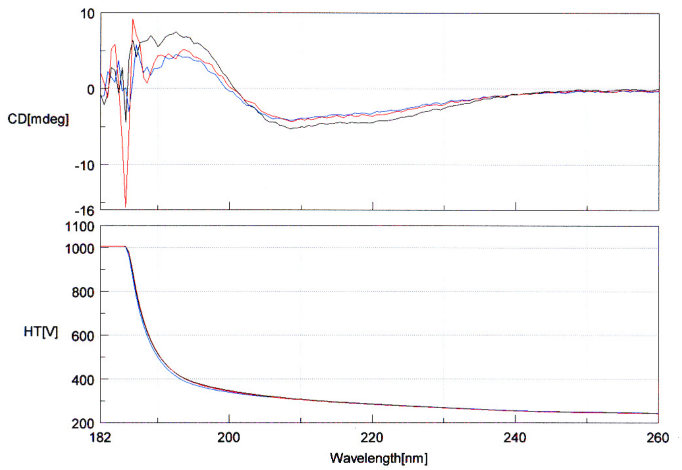


Fig. 58: CD-spectra of the specific binding of glutamine to HisH. This experiments resulted in considerable changes in secondary structure compared to the isolated HisH-domain. A decrease in the α-helical portion for this domain was found using CD experiments compared to glutamine-free (black lines) and glutamine-bounded forms (red and blue lines). The α-helical portion was reduced by 25 – 30% and was saturated at equimolar concentrations of glutamine.

Fluorescence experiments demonstrated a clear decrease in tryptophan fluorescence. This effect was detected at 20 µmolar Glutamine concentration in HisH solutions and is saturated at equimolar ratio. It is remarkable that the tryptophan quenching effect was not detected with any other amino acid residue tested (Fig. 59). The structural changes in the HisH-domain determined by substrate binding has no effect on dimerization compared to the substrate free HisH-domain. In both cases dimerization took place at the same rate.

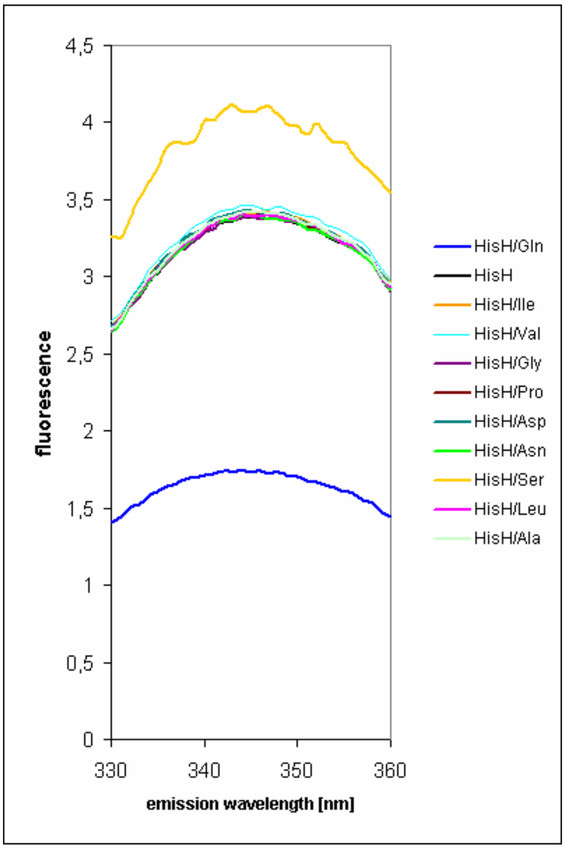


Fig. 59: Tryptophan fluorescence experiments of monomeric glutamine amido transferase HisH with different amino acids. The fluorescence experiments demonstrated a clear and specific quenching of the tryptophan fluorescence of HisH with glutamine only. This fluorescence queniching was detected at 20 µmolar Glutamine concentration in HisH solutions and is saturated at equimolar ratio. It is remarkable that the tryptophan quenching was not detected with any other amino acid tested. This results in the impression that the recognition and binding of the native substrate to the HisH protein is highly specific. All tested amino acids are shown in the legend on the right side of this figure.

The resonance assignments of the protein backbone of HisF is based on following spectra: ^{15}N-TROSY-HSQC, HNCO, HNcaCO, HNCACB, HNcoCACB, CBCACONH, HCCH-TOCSY and HNHA. The backbone amide resonance assignments are shown for monomeric HisF in a ^{15}N-TROSY-HSQC recorded at an Avance800 NMR-spectrometer at 290K (Fig. 60).

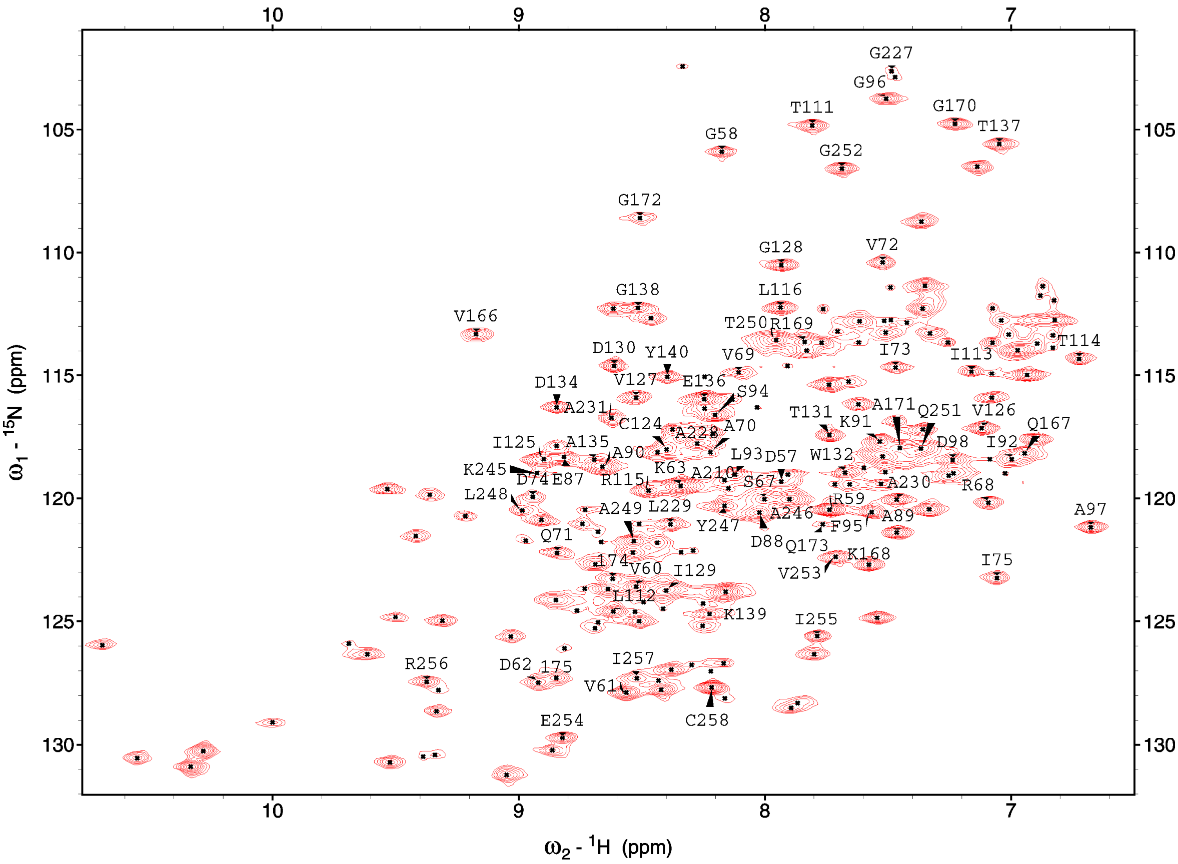


Fig. 60: ^{1}H^{15}N-TROSY-HSQC spectrum of ^{15}N labeled HisF. The assignments of the backbone amide resonances are indicated for the monomeric HisF by the one-letter code for the amino acid residues and the position. The spectra was recorded at a DRX800 NMR spectrometer at 290 K of a 1 mM protein sample in 50 mM NaPi, 100 mM NaCl, pH 7.2. Remarkable is the poor number of signals for this 258 amino acid residues protein. Interestingly the α–helices and the loop regions could be assigned, the internal β-strands could not be assigned.

The projection of the amide backbone resonances of *E. coli* HisF on the crystal structure of the *T. maritima* HisF identified the assigned and observable resonances located on the outer α-helices and loop regions. It is remarkable that the complete inner scaffolding core made out of 8 β-strands could not be assigned and was not observable in any kind of NMR spectra. In Fig. 61 the crystal structure of HisF from the published *T. maritima* IGP-synthase is shown with the NMR observable and assigned regions (red) and the not observable inner β-strands (khaki) from the top site or rather the PRFAR binding site of HisF.

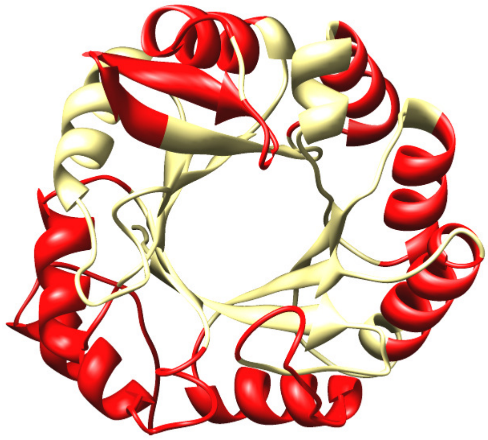


Fig. 61: Projection of the amide backbone resonances of *E. coli* HisF on the crystal structure of the *T. maritima* HisF identified the assigned and observable resonances located on the outer α-helices and loop regions.

D_2O exchange experiments from samples expressed in aqueous media and measured in a D_2O-based buffer and samples expressed in perdeuterated media and measured in aqueous solution did not result in additional assigned sequence stretches and demonstrated a limited number of resonances with no solvent exchange, probably inside the β-barrel. Combined R1/R2 NMR Carr Purcell Meiboom Gill (CPMG) -relaxation experiments (suggest the idea of an intermediate state exchange in the inner core of the HisF protein. This medium exchange rates on the small msec (approx. 1000 sec^{-1}) to large µsec (approx. 20.000 sec^{-1}) time scale, is causing the lack of information for the internal β-barrel. Also the dimerization does not effect additional resonances of the inner β-barrel.

^{1}H-^{15}N, ^{2}H-^{15}N and ^{2}H-^{15}N-^{13}C heteronuclear, multidimensional NMR-experiments of the monomeric and dimerized proteins revealed structural and dynamic information for this enzyme. ^{1}H- ^{15}N-HSQC-TROSY titration experiments of perdeuterated HisF revealed numerous chemical shift changes in the spectra due to conversion of monomeric HisF to the native protein-complex HisF·HisH (Fig. 62). This is an initial indication of structural changes in HisF by complex formation. Fig. 62 shows an overlay of the two spectra and some of the shifts marked by arrows. Monomeric ^{15}N-HisF is shown in blue and the IGP-synthase of ^{15}N-HisF and unlabeled HisH in red. Tested prediction about the regions that are involved in interaction and possible structural changes within the dimerization can be performed using the residue specific resonance assignment of HisF (Fig. 60). Interestingly not only the dimerization site of HisF is effected by this interaction, but also the PRFAR binding site, on the opposite side of the protein, is involved in dramatic changes of the chemical and obviously structural environment.

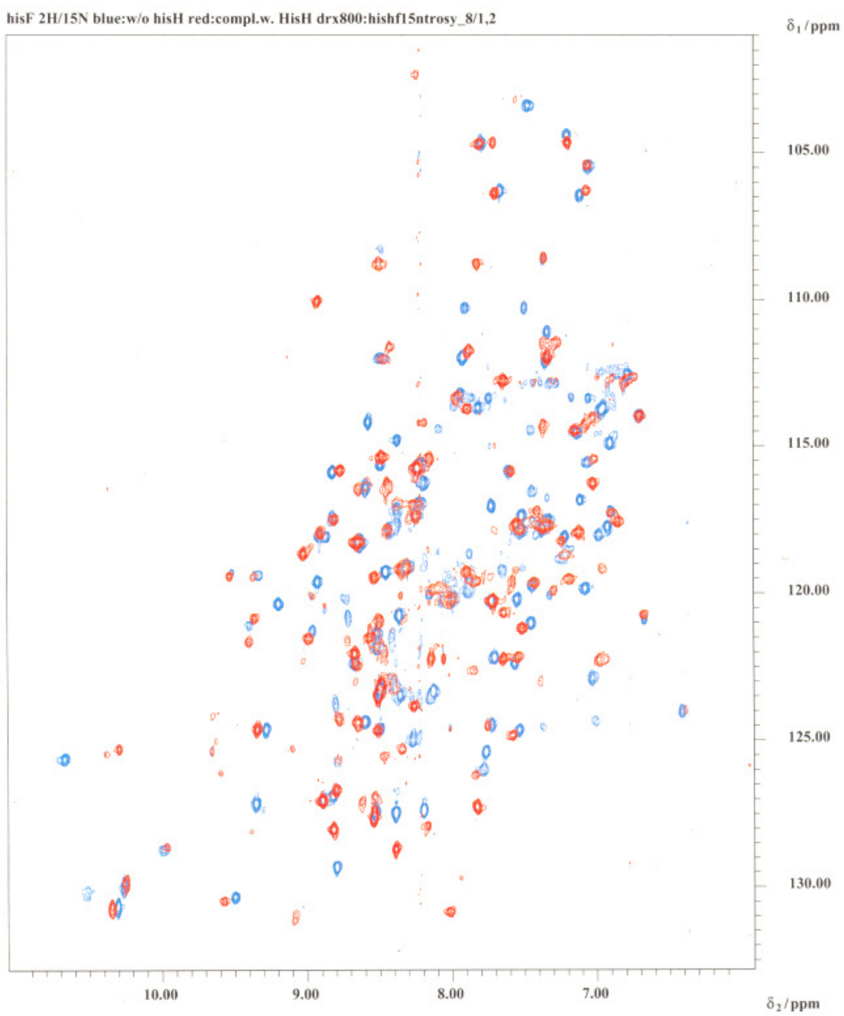


Fig. 62: ^{1}H- ^{15}N-HSQC-TROSY titration experiments of perdeuterated HisF revealed numerous chemical shift changes in the spectra due to conversion of monomeric HisF (blue) to the native protein-complex HisF·HisH (red). This is an additional indication of structural changes in HisF by complex formation.

The titration experiments of 15N-labeled HisF with the products of the IGP-synthase, IGP and AICAR, supplied small chemical shift changes around the PRFAR binding site only and do not cause major effects on the dimerization site.

Conclusions. The structural changes of the isolated HisH protein caused by glutamine binding and the fact that there is no effect on the formation of the HisH·HisF complex is a clear hint that the glutamine binding position and the dimerization region are not at the same position and both effects are structurally independent from the each other. The glutamine binding and structural changes could be shown by fluorescence spectroscopy. This prediction also supported by the calculation of the CD-spectra. These two separated functions of HisH, the glutamine binding and the independent dimerization site suggest a refined mechanism of the IGP-synthase (Fig. 63).

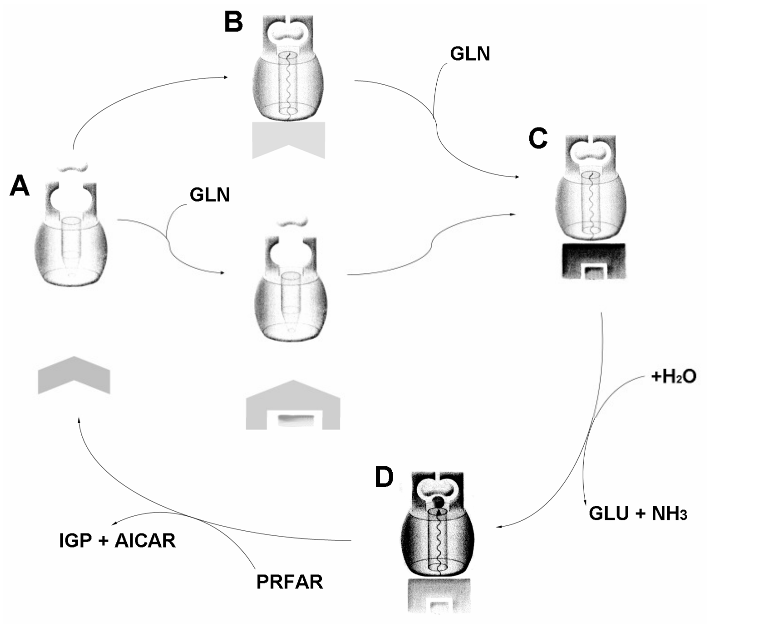


Fig. 63: Modified functionality of the IGP synthase. The dimerization of HisH and HisF is independent (B and C) from glutamine binding (A and C) . The glutaminase activity of HisH is dependent on PRFAR binding to HisF (C) and the products are formed and the complex is released after the NH_3 could be shuttled trough HisF.

The heteronuclear, multidimensional NMR-spectroscopic experiments and relaxations experiments suggest a intermediate exchange in the inner β-barrel-core of HisF. This postulated structural exchange of the β-strands to each other could be necessary for the specific channel function of HisF and could cause the water-free environment inside the channel. A detailed structural and functional study of the HisF protein by NMR, the effects and dynamic of the dimerization and enzymatic mechanism of ammonia-transfer can not be performed.

The inhibition IGP-synthase and HisF in particular has a strong antibacterial potential especially due to the fact that biosynthetic systems for the essential amino acid histidine is not present in vertebrates and would not effect vertebrate metabolism.

7 Literature

Almeida, F. C., G. C. Amorim, et al. (2001). "Selectively labeling the heterologous protein in Escherichia coli for NMR studies: a strategy to speed up NMR spectroscopy." J Magn Reson **148**(1): 142-6.

Andrec, M., Hill, R.B., and Prestegard, J.H. (1995) Amide exchange rates in Escherichia coli acyl carrier protein: correlation with protein structure and dynamics. *Protein Sci*. **4**, 983-993.

Bax, Ad (2003) Neurath Award lecture: Weak alignment offers new NMR opportunities to study protein structure and dynamics, *Protein Science*; **12**: 1-16

Beismann-Driemeyer, S., Sterner, R., (2001) Imidazole glycerol phosphate synthase from Thermotoga maritima. Quaternary structure, steady-state kinetics, and reaction mechanism of the bienzyme complex . *Journal of Biological Chemistry* 276 (23), 20387-20396

Bera, Smith, Zalkin (2000) Dual Role for the Glutamine Phosphoribosylpyro-phosphate Amidotransferase Ammonia Channel. INTERDOMAIN SIGNALING AND INTERMEDIATE CHANNELING. *J. Biol. Chem*., **275**, 7975-7979

Berendsen, H.J.C., van der Spoel, D. and van Drunen, R. (1995) GROMACS: A message-passing parallel molecular dynamics implementation, *Comp. Phys. Comm*.; **91**: 43-56

Bonvin, A.M.J.J., Houben, K., Guenneugues, M., Kaptein, R., Boelens, R. (2001) Rapid protein fold determination using secondary chemical shifts and cross-hydrogen bond 15N-13C′ scalar couplings (3hbJNC ′) *Journal of Biomolecular NMR* **21** (3), 221-233

Brünger 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., Read R.J., Rice L.M., Simonson T. and Warren G.L. (1998) Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta Cryst*. **D54**, 905-921

Brunner, S.D., Weber, T., Kohli, R.M., Schwarzer, D., Marahiel, M.A., Walsh, C.T., & Stubbs, M.T. (2002) Structural basis fort he cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. *Structure* **10**, 301-310.

Carlomagno, Chiariotti, Alifano, Nappo, Bruni (1988) Structure and Function of the *Salmonella typh.* and *E. coli* K12 Histidine Operons. *J. Mol. Biol.*, **203**, 585-606

Cobas, J.C. and Martin-Pastor, M. (2004) A homodecoupled diffusion experiment for the Analysis of Complex Mixtures by NMR. *J. Magn. Reson.*, **171**, 20-24.

Cobas, J.C., Tahoces, P.G., Martin-Pastor, M., Penedo, M. and Sardina, F.J. (2004) Wavelet-based ultra-high compression of multidimensional NMR data sets. *J. Magn. Reson.*, **168**, 288-295.

Cobas, J.C. and Sardina, J.F. (2003) Nuclear magnetic resonance data processing. MestRe-C: A software package for desktop computers. *Concept. Magn. Reson.*, **19A**, 80-96.

Cordier, F. and Grzesiek, S. (1999) Direct observation of hydrogen bonds in proteins by interresidue ${}^{3h}J_{NC'}$ scalar couplings, *J. Am. Chem. Soc.*; **121**: 1601–1602

Cole, Brosch, Parkhill, Garnier, Churcher, Harris, Gordon, Eiglmeier, Gas, Barry III, Tekaia, Badcock, Basham, Brown, Chillingworth, Connor, Davies, Devlin, Feltwell, Gentles, Hamlin, Holroyd, Hornsby, Jagels, Krogh, McLean, Moule, Murphy, Oliver, Osborne, Quail, Rajandream, Rogers, Rutter, Seeger, Skelton, Squares, Sqares, Sulston, Taylor, Whitehead, Barrell (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, **393**, 537-544

Cornilescu, G., Delaglio, F., Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology *Journal of Biomolecular NMR* **13** (3), 289-302

Crump, M.P., et al., and Simpson, T.J. (1997) Solution structure of the actinorhodin polyketide synthase acyl carrier protein from *Streptomyces coelicolor* A3(2). *Biochemistry* **36**, 6000-6008.

Dominguez, C., Boelens, R. and Bonvin, A. M.J.J. (2003) HADDOCK: a protein-protein docking approach based on biochemical and/or biophysical information. *J. Am. Chem. Soc*. **125**, 1731-1737.

Douangamath, A., Walker, M., Beismann-Driemeyer, S., Vega-Fernandez, M.C., Sterner, R., Wilmanns, M., (2002) Structural evidence for ammonia tunneling across the (βα)8 barrel of the imidazole glycerol phosphate synthase bienzyme complex. *Structure* **10** (2), 185-193

Findlow, S.C., Winsor, C., Simpson, T.J., Crosby, J., and Crump, M.P. (2003) Solution structure and dynamics of oxytetracycline polyketide synthase acyl carrier protein from *Streptomyces rimosus*, *Biochemistry* **42**, 8423-8433.

Finking, R., Mofid, M.R., and Marahiel, M.A. (2004). Mutational analysis of peptidyl carrier protein and acyl carrier protein synthase unveils residues involved in protein-protein recognition, *Biochemistry* **43**, 8946-8956.

Goh, C.S., Milburn, D., and Gerstein, M. (2004) Conformational changes associated with protein-protein interactions. *Curr. Opin. Struct. Biol*. **14**, 104-109.

Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol*. **273**, 283-298.

Guntert P. (2004) Automated NMR structure calculation with CYANA. *Methods Mol Biol*.; **278**:353-78

Hansen, M.R., Mueller, L. and Pardi, A. (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat. Struct. Biol*.; **5**(12):1065-1071

Heidelberg Eisen, Nelson, Clayton, Gwinn, Dodson, Haft, Hickey, Peterson, Umayam, Gill, Nelson, Read, Tettelin, Richardson, Ermolaeva, Vamathevan, Bass, Qin, Dragoi, Sellers, McDonald, Utterback, Fleishmann, Nierman, White, Salzberg, Smith, Colwell, Mekalanos, Venter, Fraser (2000) DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae. Nature,* **406**, 477-483

Herrmann T, Guntert P. and Wuthrich K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol*.; **319**(1): 209-27

Holak, T.A., Kearsley, S.K., Kim, Y., and Prestegard, J.H. (1988) Three-dimensional structure of acyl carrier protein determined by NMR pseudoenergy and distance geometry calculations, *Biochemistry* **27**, 6135-6142.

Kasai (1974) Regulation of the Expression of the histidine operon in *Salmonella typhimurium. Nature (London)*, **249**, 523-527

Kay, L. E. (2001). "Nuclear magnetic resonance methods for high molecular weight proteins: a study involving a complex of maltose binding protein and beta-cyclodextrin." Methods Enzymol **339**: 174-203.

Kern, D., & Zuiderweg, E.R.P. (2003) The role of dynamics in allosteric regulation. *Curr. Opin. Struct. Biol.* **13**, 748-757.

Kim, Y., and Prestegard, J.H. (1989) A dynamic model for the structure of acyl carrier protein in solution. *Biochemistry* **28**, 8792-8797.

Klammt, C., Löhr, F., Schäfer, B., Haase, W., Dötsch, V., Rüterjans, H., Glaubitz, C., Bernhard, F., (2004) High level cell-free expression and specific labeling of integral membrane proteins , *European Journal of Biochemistry* **271** (3), 568-580

Klammt, C., Schwarz, D., Fendler, K., Haase, W., Dötsch, V., Bernhard, F., (2005) Evaluation of detergents for the soluble expression of α-helical and β-barrel-type integral membrane proteins by a preparative scale individual cell-free expression system , *FEBS Journal* **272** (23),6024-6038

Klem, Davisson (1993) IGP-Synthase: The Glutamine Amidotransferase in Histidine Biosynthesis. *Biochemistry,* **32**, S.5177-5186

Klem, Chen, Davisson (2001) Subunit interactions and glutamine utilization by *Escherichia coli* Imidazole glycerol Phosphate Synthase. *J. Bacteriol.*, **182**, 989-996

Kunst, Ogasawara, Moszer, Albertini, Alloni, Azevedo, Bertero, Bessieres, Bolotin, Borchert, Boriss, Boursier, Brans, Braun, Brignell, Bron, Brouillet, Bruschi, Caldwell, Capuano, Carter, Choi, Codani, Connerton, Cummings, Daniel, Denizot, Devine, Duesterhoeft, Ehrlich, Emmerson, Entian, Errington, Fabret, Ferrari, Foulger, Fritz, Fujita, Fujita, Fuma, Galizzi, Galleron, Ghim, Glaser, Goffeau, Golightly, Grandi, Guiseppi, Guy, Haga, Haiech, Harwood, Henaut, Hilbert, Holsappel, Hosono, Hullo, Itaya, Jones, Joris, Karamata, Kasahara, Klaerr-Blanchard, Klein, Kobayashi, Koetter, Koningstein, Krogh, Kumano, Kurita, Lapidus, Lardinois, Lauber, Lazarevic, Lee, Levine, Liu, Masuda, Maueel, Medigue, Medina, Mellado, Mizuno, Moestl, Nakai, Noback, Noone, O'Reilly, Ogawa, Ogiwara, Oudega, Park, Parro, Pohl, Portetelle, Porwolik, Prescott, Presecan, Pujic, Purnelle, Rapoport, Rey, Reynolds, Rieger, Rivolta, Rocha, Roche, Rose, Sadaie, Sato, Scanlon, Schleich, Schroeter, Scoffone, Sekiguchi, Sekowska, Seror, Serror, Shin, Soldo, Sorokin, Tacconi, Takagi, Takahashi, Takemaru, Takeuchi, Tamakoshi, Tanaka, Terpstra, Tognoni, Tosato, Uchiyama, Vandenbol, Vannier, Vassarotti, Viari, Wambutt, Wedler, Wedler, Weitzenegger, Winters, Wipat, Yamamoto, Yamane, Yasumoto, Yata, Yoshida, Yoshikawa, Zumstein, Yoshikawa, (1997) The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis. Nature,* **390**, 249-256

Lambalot, R.H., Gehring, A.M., Flugel, R.S., Zuber, P., LaCelle, M., Marahiel, M.A., Reid, R., Khosla, C., and Walsh, C.T. (1996). A new enzyme superfamily - the phosphopantetheinyl transferases. *Chem. Biol.* **3**, 923-936.

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

Lee, K. M., E. J. Androphy, et al. (1995). "A novel method for selective isotope labeling of bacterially expressed proteins." J Biomol NMR **5**(1): 93-6.

Lindahl, E., Hess, B. and van der Spoel, D. (2001) GROMACS 3.0: A package for molecular simulation and trajectory analysis, *J. Mol. Mod.*; **7**, 306-317

Linge, J.P., O'Donoghue, S.I. and Nilges, M. (2001) Assigning Ambiguous NOEs with ARIA. *Methods in Enzymology*; **339**, 71-90

Linge, J.P., Habeck, M., Rieping, W. and Nilges, M. (2003) ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics;* **19**, 315-316.

Linne, U., Marahiel, M.A. (2000) Control of directionality in nonribosomal peptide synthesis: Role of the condensation domain in preventing misinitiation and timing of epimerization . *Biochemistry;* **39** (34), 10439-10447

Linne, U., Schwarzer, D., Schroeder, G., and Marahiel, M.A. (2004) Mutational analysis of a type II thioesterase associated with nonribosomal peptide synthesis. *Eur. J. Biochem.*; **271**, 1536-1545.

Maier, T., Jenni, S., and Ban, N. (2006) Architecture of Mammalian Fatty Acid Synthase at 4.5 A Resolution. *Science;* **311**, 1258 - 1262

Mittag T, Schaffhausen B, Gunther UL, (2004) Tracing kinetic intermediates during ligand binding., *J Am Chem Soc.*;**126**(29):9017-23

Mittag T, Schaffhausen B, Gunther UL., (2003) Direct observation of protein-ligand interaction kinetics., *Biochemistry*.;**42**(38):11128-36

Mofid, M.R., Marahiel, M.A., Ficner, R., and Reuter, K. (1999) Crystallization and preliminary crystallographic studies of Sfp: a phosphopantetheinyl transferase of modular peptide synthetases, *Acta Crystallogr. D: Biol. Crystallogr.* **55**, 1098-1100.

Mofid, M.R., Finking, R., and Marahiel, M.A. (2002). Recognition of hybrid peptidyl carrier proteins/acyl carrier proteins in nonribosomal peptide synthetase modules by the 4'-phosphopantetheinyl transferases AcpS and Sfp. *J Biol Chem*, **277**, 17023-17031.

Mofid, M.R., Finking, R., Essen, L.O., and Marahiel, M.A. (2004) Structure-based mutational analysis of the 4'-phosphopantetheinyl transferases Sfp from *Bacillus subtilis*: carrier protein recognition and reaction mechanism. *Biochemistry* **43**, 4128-4136.

Mootz, H.D., Schwarzer, D., Marahiel, M.A. (2000) Construction of hybrid peptide synthetases by module and domain fusions *Proc. Natl. Acad. Sci. U.S.A.* **97** (11), 5848-5853

Mootz H.D., Schwarzer D., Marahiel M.A., (2002) Ways of assembling complex natural products on modular nonribosomal peptide synthetases., *Chembiochem*.;**3**(6):490-504

Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K. and Olson, A.J. (1998) *J. Comput. Chemistry* **19**, 1639-1662

Murray & Hartmann (1972) *Canad. J. Microbiol.*, **18**, 671-681

Nilges, M. and O'Donoghue, S.I. (1998) Ambiguous NOEs and automated NOE assignment. *Prog. NMR Spect.* **32**, 107-139

Ollerenshaw, J. E., V. Tugarinov, et al. (2003). "Methyl TROSY: explanation and experimental verification." Magn. Reson. Chem. **41**: 843-82.

Parker, M.J., Aulton-Jones, M., Hounslow, A.M., Craven, C.J., (2004) A Combinatorial Selective Labeling Method for the Assignment of Backbone Amide NMR Resonances, *Journal of the American Chemical Society*, **126** (16), 5020-5021.

Parris, K.D., Lin, L., Tam, A., Mathew, R., Hixon, J., Stahl, M., Fritz, C.C., Seehra, J., and Somers, W.S. (2000) Crystal structures of substrate binding to Bacillus subtilis holo-(acyl carrier protein) synthase reveal a novel trimeric arrangement of molecules resulting in three active sites, *Struct. Fold. Des*. **8**, 883-895.

Patel, H.M., Tao, J., Walsh, C.T.(2003) Epimerization of an L-cysteinyl to a D-cysteinyl residue during thiazoline ring formation in siderophore chain elongation by pyochelin synthetase from Pseudomonas aeruginosa *Biochemistry* **42** (35), 10514-10527

Pervushin, K., R. Riek, et al. (1997). Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U S A* **94**(23): 12366-71.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004) UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **25**, 1605-1612.

Prestegard, J. H., Bougault, C. M., and Kishore, A. I., (2004) Residual Dipolar Couplings in Structure Determination of Biomolecules. *Chem. Rev.*; **104**(8): 3519 - 3540

Quadri, L.E.N., Weinreb, P.H., Lei, M., Nakano, M.M., Zuber, P., Walsh, C.T., (1998) Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carder protein domains in peptide synthetases. *Biochemistry* **37** (6), 1585-1595

Qiagen (May 2004). QIAprep Miniprep Handbook.

Rajagopalan, S., C. Chow, et al. (2004). "NMR spectroscopic filtration of polypeptides and proteins in complex mixtures." J Biomol NMR **29**(4): 505-16.

Reed, M.A.C., Schweizer, M., Szafranska, A.E., Arthur, C., Nicholson, T.P., Cox, R.J., Crosby, J., Crump, M.P., and Simpson, T.J. (2003) The type I rat fatty acid synthase ACP shows structural homology and analogous biochemical properties to type II ACPs. *Org. Biomol. Chem*. **1**, 463-471.

Reuter, K., Mofid, M.R., Marahiel, M.A., and Ficner, R. (1999) Crystal structure of the surfactin synthetase-activating enzyme sfp: a prototype of the 4'-phosphopantetheinyl transferase superfamily, *EMBO J* **18**, 6823-6831.

Salzmann, M., Wider, G., Pervushin, K., Senn, H., Wüthrich, K. (1999) TROSY-type triple-resonance experiments for sequential NMR assignments of large proteins *Journal of the American Chemical Society* **121** (4), 844-848

Sambrook, J. and D. W. Russell (2001) Molecular cloning: a laboratory manual (3rd ed.). Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press.

Sanner, M.F., Olson, A.J., and Spehner, J.C. (1996) Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* **38**(3), 305-320.

Schagger, H. and G. Von Jagow (1987). "Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa." Anal. Biochem. **166**: 368-379.

Schwarzer, D., Mootz, H.D., Linne, U., and Marahiel, M.A. (2002) Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proc Natl Acad Sci U S A* **99**, 14083-14088.

Schwarzer, D., Finking, R., and Marahiel, M.A. (2003) Nonribosomal eptides: from genes to products, *Nat. Prod. Rep*. **20**, 275-287.

Shi, J., Pelton, J.G., Cho, H.S., Wemmer, D.E., (2004) Protein signal assignments using specific labeling and cell-free synthesis**.** *Journal of Biomolecular NMR*, **28** (3), 235-247

Sieber SA, Marahiel MA., (2005) Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem Rev.*;**105**(2):715-738.

Spirin, A.S., Baranov, V.I., Ryabova, L.A., Ovodov, S.Y., Alakhov, Y.B., (1988), A continuous cell-free translation system capable of producing polypeptides in high yield, *Science*, **242** (4882), 1162-1164.

Stachelhaus, T., Hüser, A., and Marahiel, M.A. (1996) Biochemical characterization of peptidyl carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases, Ch*em. Biol.***3**, 913-921.

Stein, D.B., Linne, U., Marahiel, M.A.(2005) Utility of epimerization domains for the redesign of nonribosomal peptide synthetases *FEBS Journal* **272** (17), 4506-4520

Stein, T., (1996) The multiple carrier model of nonribosomal peptide biosynthesis at modular multienzymatic enzymes. *J. Biol. Chem*. **271**, 15428-15435.

Tjandra, N. (1999) Establishing a degree of order: obtaining high-resolution NMR structures from molecular alignment. *Structure*; **7**: R205-R211

Thoden, Holden, Wesenberg, Raushel, Rayment (1997) Structure of Carbamoyl Phosphate Synthetase: A Journey of 96 Å from Substrate to Product. *Biochemistry*, **36**, 6305-6316

Trbovic, N., Klammt, C., Koglin, A., Löhr, F., Bernhard, F., Dötsch, V., (2005) Efficient strategy for the rapid backbone assignment of membrane proteins, *J. Am. Chem. Soc.* **127** (39), 13504-13505

Tugarinov, V., P. M. Hwang, et al. (2003). "Cross-correlated relaxation enhanced 1H[bond]13C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes." *J. Am. Chem. Soc.* **125**(34): 10420-10428.

Tugarinov, V. and L. E. Kay (2004). An isotope labeling strategy for methyl TROSY spectroscopy. *J Biomol NMR* **28**(2): 165-72.

Turner, R.J., Taylor, D.E., Weiner, J.H., (1997), Expression of *Escherichia coli* TehA gives resistance to antiseptics and disinfectants similar to that conferred by multidrug resistance efflux pumps**.** *Antimicrobial Agents and Chemotherapy*, **41** (2), 440-444.

Valafar, H., (2003) REsidual Dipolar Coupling Analysis Tool, *Bioinformatics* **19** (12): 1549-1555

van de Ven, F.J.M., (1995) Multidimensional NMR in Liquids Basic Principles and

Experimental Methods *Verlag Chemie*, Weinheim, Germany

van Gunsteren, W.F. and Berendsen, H.J.C. (1990) Computer Simulation of Molecular Dynamics: Methodology, Applications and Perspectives in Chemistry, *Angew. Chem. Int. Ed. Engl.*; **29**: 992-1023

van Gunsteren, W.F., Hünenberger, P.H., Mark, A.E., Smith, P.E. and Tironi, I.G. (1995) Computer simulation of protein motion, *Computer Phys. Communications*; **91**: 305-319

Volkman, B.F., Lipson, D., Wemmer, D.E., and Kern, D. (2001) Two-state allosteric behaviour in a single-domain signalling protein. *Science* **291**, 2429-2433.

Wagner (2001) attenuation of transcription, *Transcription regulation in Prokaryotes*, 128-136; Oxford University press, Oxford

Walsh C.T., (2004) Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science* **303**(5665):1805-1810.

Weber, T., Baumgartner, R., Renner, C., Marahiel, M.A., and Holak, T.A. (2000). Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. *Struct. Fold. Des.* **8**, 407-418.

Weinreb, P.H., Quadri, L.E.N., Walsh, C.T., and Zuber, P. (1998) Stoichiometry and specificity of in vitro phosphopantetheinylation and aminoacylation of the valine-activating module of surfactin synthetase. *Biochemistry* **37**, 1575-1584.

Wider, G., Wüthrich, K. (1999) NMR spectroscopy of large molecules and multi-molecular assemblies in solution. *Current Opinion in Structural Biology* **9** (5), 594-601

Winkler (1987) *E.coli* and *S. typh*, Cellular and molecular biology : Biosynthesis of Histidine. *Am. Society for Microbiology*, **1987**, 395-411

Wolf-Watz, M., Thai, V., Henzler-Wildman, K., Hadjipavlou, G., Eisenmesser, E.Z., Kern, D. (2004) Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nature Structural and Molecular Biology* **11** (10), 945-949

Wong, H.C., Liu, G., Zhang, Y.-M., Rock, C.O., and Zheng, J. (2002). The solution structure of acyl carrier protein from *Mycobacterium tubercilosis*. *J.Biol. Chem*. **277**, 15874-15880.

Xu, G.Y., Tam. A., Lin, L., Hixon, J., Fritz, C.C., and Powers, R. (2001) Solution structure of B. subtilis acyl carrier protein. *Structure (Cambridge)* **9**, 277-287.

Yeh, E., Kohli, R.M., Bruner, S.D., Walsh, C.T. (2004) Type II thioesterase restores activity of a NRPS module stalled with an aminoacyl-S-enzyme that cannot be elongated *ChemBioChem* **5** (9), 1290-1293

Zalkin (1993) The Amidotransferases; A Review: *Adv.Enzymol.Relat.Areas Mol.Biol*., **66**, 203-309

Zhang, Y.M., Rao, M.S., Heath, R.J., Price, A.C., Olson, A.J., Rock, C.O., and White, S.W. (2001) Identification and analysis of the acyl carrier protein (ACP) docking site on beta keto-acyl-ACP synthase III. *J. Biol. Chem*. **276**, 8231-8238.

Zuiderweg, E.R.P. (2002) Mapping Protein-Protein Interactions in Solution by NMR Spectroscopy. *Biochemistry*; **41**(1):1 – 7

Zweckstetter, M. and Bax, Ad (2001) Characterization of molecular alignment in aqueous suspensions of Pf1 bacteriophage. *J. Biomol NMR*.; **20**: 365-377

Own publications

Trbovic, N., Klammt, C., Koglin, A., Löhr, F., Bernhard, F., Dötsch, V., (2005) Efficient strategy for the rapid backbone assignment of membrane proteins. *Journal of the American Chemical Society* **127** (39),13504-13505

Koglin, A., Klammt, C., Trbovic, N., Schwarz, D., Schneider, B., Schäfer, B., Löhr, F., Bernhard, F., Dötsch, V., (2006) Combination of cell-free expression and NMR spectroscopy as a new approach for structural inverstigation of membrane proteins. *Magnetic resonance in chemistry : MRC* , accepted

Koglin, A. Mofid, M.R., Löhr, F., Schäfer, B., Rogov, V.V., Blum, M.-M., Mittag, T., Marahiel, M.A., Bernhard, F., Dötsch, V., (2006) Conformational Switches Modulate Protein Interactions in Peptide Antibiotic Synthetases. *Science* **312**, 273-276

Schwarzer, D., Klammt, C., Koglin, A., Löhr, F., Schneider, B., Dötsch, V., Bernhard, F., (2006) Preparative scale cell-free expression systems: New tools fort he large scale preparation of integral membrane proteins for functional and structural studies. accepted

Rogov, V.V., Löhr, F., Rogova, N., Klammt, C., Koglin, A., Bernhard, F., Dötsch, V., (2006) A new doublet organization of phosphoreceiver domains in a hybrid sensor kinase: resonance assignment and solution structure of *Escherichia coli* C-RcsC. *submitted*

Rashid, U.J., Chandra, T., Koglin, A., Chen, J. C.-H., (2006) Structure of *Aquifex aeolicus* Argonaute highlights conformational flexibility of the PAZ domain as a potential regulator of RISC function. *accepted*

DANKSAGUNG

Besonderer Dank gilt Herrn Prof. Volker Dötsch für ausgezeichnete und inspirierende Gespräche, die eine kritisch hinterfragende und effektive Arbeit ermöglicht haben. Vielen Dank für die Unterstützung bei den Projekten, das offene Ohr bei Problemen, die Momente neben der Arbeit und die sehr angenehme Atmosphäre im Arbeitskreis.

Mein herzlicher Dank gilt allen Mitgliedern des Arbeitskreises ‚Biophysikalische Chemie', denn erst dieser hat in seiner Gesamtheit die freundliche Atmosphäre, die über die Arbeit hinaus anhält, und den Raum für eine effektive Zusammenarbeit geschaffen.

Vielen Dank an Dr. Frank Bernhard und Prof. Mohamed Marahiel, die mir erst den Zugang zu den NRPS Systemen ermöglicht haben und mich, durch Ihre Unterstützung und Betreuung, für diese Systeme begeistert haben. Vielen Dank an Prof. Marahiel für die Hilfe und Unterstützung während der Arbeit und bei der weiteren Gestaltung meiner nahen Zukunft. Des weiteren möchte ich dem ganzen Arbeitskreis von Mohamed Marahiel, für all die bereitwillige Hilfe danken, insbesondere Mohammmed Mofid und Daniel Stein.

Ein ganz besonders herzliches Dankeschön gilt Birgit Schäfer, die mit Ihrer Riesenhilfe und Ihrer geduldigen und ausgleichenden Art mich oft genug vor dem Chaos bewart hat und mir mit Ihrer großartigen Unterstützung oft sehr geholfen hat.

Einen Riesendank für die Aufnahme und Hilfe bei NMR-Experimenten und die Geduld beim Erklären diverser Probleme an Dr. Frank Löhr und Sigi Oguzer-Fachinger möchte ich für Ihre hilfreiche und wunderbare Organisation des ganzen Instituts danken.

Ein schmetternder Dank an den Herren Hpt. Marc-Michael Blum! Vielen Dank für die vielen intensiven und oft zeitlich ausufernden Gespräche, die nie Gefahr liefen, langweilig zu werden. Vielen Dank für besonnene Ruhe in kritischen Momenten, die Freundschaft und nicht zuletzt für die tolle Zusammenarbeit beim Errichten und Erhalten des Computernetzwerkes und darüber hinaus. Wer geht nach dem Kino schon noch mal arbeiten?

Dear Dr. Vladimir V. Rogov thank you for your help and support, your time to explain all the big and small problems and the fascinating feeling and knowledge about structures. Florian Durst und Dr. Wesley McGinn-Straub möchte ich für Ihre Hilfe in vielen Lagen, die freundschaftliche Atmosphäre und die bereichernde Zusammenarbeit danken. Dieser Dank gilt auch Birgit Schneider, Daniel Schwarz und Christian Klammt.

Christina Fischer und Dr. Tanja Mittag möchte ich für ihre Freundschaft und Ihre ehrliche, direkte Art danken, auch wenn sie rechtzeitig vor mir geflohen sind, um schlimmeres zu verhindern. Ein Dank auch an ihren ehemaligen Chef, der mit seiner äußerst diplomatischen, umsichtigen und aufrichtigen Art, nicht nur für mich eine nicht einzuschätzende Unterstützung war.

Dr. Michael Humbert möchte ich für seine stete Hilfe und Unterstützung in allen Lebenslagen, insbesondere in den chaotischen danken.

Bei meinen Eltern und meinem Bruder bedanke ich mich für eure immerwährende Unterstützung und Hilfe während der gesamten Ausbildung und darüber hinaus bedanken. Eure Zuneigung hat mir oft geholfen kontinuierlich weiter zumachen, grade in schwierigen Situationen.

In memoriam: Joachim Thum, Ruth und Arno Kessel.

LEBENSLAUF

Name: Koglin

Vorname: Alexander

Adresse: Buchenweg 15
74861 Neudenau
Deutschland

eMail: alexk@bpc.uni-frankfurt.de ALEXANDER_KOGLIN@hms.harvard.edu

Nationalität: Deutsch

Geburtsdatum: 08. September 1975

Geburtsort: Potsdam

Ausbildung

1982 - 1989 Polytechnische Oberschule, Magdeburg
1989 - 1990 Bettina von Arnim-Gymnasium, Magdeburg
1990 - 1995 Theodor-Heuss-Gymnasium, Heilbronn
1996 - 2001 Johann Wolfgang Goethe-Universität, Frankfurt/Main, Deutschland
Diplomarbeit bei Prof. Heinrich Rüterjans, Institut für Biophysikalische Chemie, Johann- W. Goethe Universität Frankfurt/Main, Deutschland
2001 Abschluss: Diplom Biochemiker (Note: mit Auszeichnung)
2001 - 2006 Promotion: „ Strukturelle Dynamik von Peptidyl Carrier Domänen in nicht-ribosomalen Peptid-Synthetasen " (Prof. Volker Dötsch, Institut für Biophysikalische Chemie,
Johann W. Goethe Universität Frankfurt/Main, Deutschland)

Stipendien

2002-2003 DFG-Graduiertenkolleg
2004-2006 DFG-Einzelantrag BE-19/11 (Dr. Frank Bernhard)

Publikationen

2005 Trbovic, N., Klammt, C., Koglin, A., Löhr, F., Bernhard, F., Dötsch, V., (2005) Efficient strategy for the rapid backbone assignment of membrane proteins. *Journal of the American Chemical Society* **127** (39),13504-13505

2006 Koglin, A. Mofid, M.R., Löhr, F., Schäfer, B., Rogov, V.V., Blum, M.-M., Mittag, T., Marahiel, M.A., Bernhard, F., Dötsch, V., (2006) Conformational Switches Modulate Protein Interactions in Peptide Antibiotic Synthetases. *Science* **312**,273-276

Koglin, A., Klammt, C., Trbovic, N., Schwarz, D., Schneider, B., Schäfer, B., Löhr, F., Bernhard, F., Dötsch, V., (2006) Combination of cell-free expression and NMR spectroscopy as a new approach for structural inverstigation of membrane proteins. *Magnetic resonance in chemistry : MRC* , accepted

Schwarzer, D., Klammt, C., Koglin, A., Löhr, F., Schneider, B., Dötsch, V., Bernhard, F., (2006) Preparative scale cell-free expression systems: New tools for the large scale preparation of integral membrane proteins for functional and structural studies. accepted

Rashid, U.J., Chandra, T., Koglin, A., Chen, J. C.-H., (2006) Structure of *Aquifex aeolicus* Argonaute highlights conformational flexibility of the PAZ domain as a potential regulator of RISC function. *accepted*

Rogov, V.V., Löhr, F., Rogova, N., Klammt, C., Koglin, A., Bernhard, F., Dötsch, V., (2006) A new doublet organization of phosphoreceiver domains in a hybrid sensor kinase: resonance assignment and solution structure of *Escherichia coli* C-RcsC. *Submitted and revised*

Präsentationen und Kurse

2002	102nd General Meeting ASM Conference, Salt-Lake City, Utah, USA [Poster]
2003	4th European Biophysics Congress, Alicante, Spanien [Poster]
	EuroLab Course Advanced Computing in NMR Spectroscopy, Florenz, Italien
2004	8th IUBMB Conference and ASBMB Annual Meeting, Boston, Massachusetts, USA [Poster]
2005	7th Course: Structure and Biophysics, NATO Adv. Study Inst; Erice; Italien [Poster, Talk]
	27th Discussion Meeting GDCh Magnetic Resonance Division, Mainz [Poster]

Beteiligung an der Lehre

2001-2004	Fortgeschrittenen-Praktikum der biophysikalischen Chemie (BPCII): NMR-Zuordnung
2004	Grundlagen-Praktikum der biophysikalischen Chemie (BPCI): Fluoreszenz-Spektroskopie
2005	Betreuung der Diplomarbeiten von Jörg Rinnenthal und Nikola Trbovic

Akademische Lehrer

Prof. Dr. Dötsch	(Biophysikalische Chemie)
Prof. Dr. Dr. Fasold	(Biochemie)
Prof. Dr. Ludwig	(Biochemie, Genetik)
Prof. Dr. Rüterjans	(Biophysikalische Chemie)
PD Dr. Günther	(Biophysikalische Chemie)
Prof. Dr. Bamberg	(Biophysikalische Chemie)
Prof. Dr. Göbel	(Organische Chemie)
Prof. Dr. Griesinger	(Organische Chemie)
Prof. Dr. Steiger	(Virologie)
Prof. Dr. Entian	(Mikrobiologie)
Prof. Dr. Pons	(Mikrobiologie)
Prof. Dr. Kröger	(Mikrobiologie)
Prof. Dr. Kolbesen	(Anorganische Chemie)
Prof. Dr. Bereiter-Hahn	(Zellbiologie)
Prof. Dr. Zimmermann	(Zellbiologie)
Prof. Dr. Prinzinger	(Physiologie)
Prof. Dr. Hainer	(Mathematik)
Prof. Dr. Aßmus	(Physik)

Zivildienst

1995 - 1996	Johanniter Unfall-Hilfe, Kreisverband Franken, Heilbronn

Zusammenfassung

Eine große Zahl natürlicher, so genannter sekundärer Metabolite sind kleine und strukturell sehr verschiedene Polypeptide und Polyketide. Diese bioaktiven Substanzen haben im allgemeinen ein breit aufgestelltes therapeutisches Potential und werden von verschiedenen bakteriellen Stämmen und Pilzen biosynthetisiert. Sie sind sowohl biologisch, als auch therapeutisch wichtig als Cytostatika, Immunsuppressiva und Antibiotika mit einem sehr großem antibakteriellen und antiviralem Potential. Die trotz ihrer kleinen Größe, komplexen Polypeptide und Polyketide werden von modular aufgebauten Megaenzymen in mehrstufigen Mechanismen synthetisiert. Für die Synthese dieser Peptide sind sehr große Proteincluster verantwortlich, die meistens aus einer begrenzten Anzahl sehr großer, Multidomänen umfassenden, Superenzyme aufgebaut werden. Diese Proteincluster mit einem Molekulargewicht bis in den Bereich von MegaDalton werden als nicht-ribosomale Peptidsynthetasen (NRPS) und Polyketidsynthetasen (PKS) bezeichnet. Die NRPS Systeme zeichnen sich dadurch aus, daß für die biosynthetisierten Polypeptide keine Information in Form von Nukleinsäuren wie DNA oder RNA kodiert (Walsh, C.T., 2004; Sieber & Marahiel, 2005).

Peptidyl Carrier Proteine (PCPs) sind kleine zentrale Transport-Domänen, integriert in den Modulen nicht-ribosomaler Peptidsynthetasen (NRPSs). PCPs tragen kovalent über eine Phosphoesterbindung einen aus dem Protein herausragenden 4'-phosphopantetheinyl (4'-PP) Kofaktor. Der 4'-PP Kofaktor ist an der Seitenkette eines hochkonservierten Serins gebunden, welches ein zentraler Bestandteil der Phosphopantethein-Erkennungs-Sequenz ist. Die Erkennungssequenz ist homolog in vielen Proteinen mit ähnlicher Funktion, inklusive Acyl Carrier Proteinen (ACPs) der Fettsäuresynthetasen (FAS) und der Polyketidsynthetasen (PKS). Die Thiolgruppe des 4'-PP Kofaktors dient zum aktiven Transport der Substrate und der Intermediate der NRPS Systeme. Die generelle Organisation und die Kontrolle der exakt aufeinander folgenden Reaktionsschritte in der Peptidsynthetase, ist die entscheidende Frage für die Funktion des Proteinclusters (*assembly line mechanism*). In Modulen der NRPS Systeme folgen die PCP-Domänen C-terminal auf die Adenylierungs-domänen (A-Domäne). Die Aufgabe der A-Domänen ist die Selektion and die Aktivierung einer spezifischen Aminosäure für die „assembly line". Die eigentliche Bildung der Peptidbindung erfolgt an der Kondensations-Domäne (C-Domäne). Der Transfer der Peptidintermediate und der aktivierten Aminosäuren zwischen A-Domänen und C-Domänen ist Aufgabe der PCPs. Um diese Funktion erfüllen zu können, ist

eine große Bewegung in PCPs, bzw. des 4'-PP Kofaktors notwendig, welche als „swinging arm model" (Weber et al., 2001) beschrieben wurde. Die PCPs koordinieren damit die Peptidbiosynthese während sie mit diversen Domänen der Synthetasen spezifisch wechselwirken müssen. Die molekularen Mechanismen des Transportes wurden bisher allerdings nicht untersucht. Eine Dynamik der Transport-Domänen wurde bereits postuliert (Kim & Prestegard, 1989; Andrec et al., 1995), konnte bisher aber nicht gezeigt werden (Weber et al., 2001). Interessanterweise zeigt sowohl apo-PCP (ohne den kovalent gebundenen 4'-PP Kofaktor) also auch holo-PCP langsamen chemischen Austausch, der als jeweils zwei stabile Konformationen beschrieben werden konnte. Diese jeweils zwei stabilen Zustände, welche sich im Austausch befinden, wurden als A und A*, für apo-PCP, und entsprechend H und H* für holo-PCP bezeichnet. Während der A- und der H-Zustand sich sowohl voneinander als auch von den entsprechenden A* und H*-Zuständen unterscheiden und spezifisch für die apo- und die holo-Form von PCP sind, ist die kalkulierte Struktur vom A*-Zustand nahezu identisch mit der des H*-Zustandes. Die erhaltenen NMR-Strukturen des A-Zustandes, des H-Zustandes und des gemeinsamen A/H-Zustandes beschreiben in ihrer Gesamtheit ein neues Modell für ein allosterie-kontrolliertes System dualer konformationeller Zwei-Zustands-Dynamik. Zu dem beobachteten konformationellen Austausch der PCP-Domäne, konnte die Bewegung des 4'-PP Kofaktors koordiniert werden. Die Bewegung des 4'-PP Kofaktors in Verbindung mit dem konformationellen Austausch der PCP-Domäne charakterisiert die Interaktion mit katalytischen Domänen eines NRPS Moduls. Des weiteren konnte mit Hilfe des Modells die Wechselwirkung mit externen Interaktionspartnern, wie der Thioesterase II und der 4'-PP Transferase, untersucht werden.

Die externe Thioesterase II der Surfactin-Synthetase (SrfTEII) von *Bacillus subtilis* ist ein separat expremiertes 28 KDa Protein. Sie gehört zur Familie der α/β-Hydrolasen und ist verantwortlich für die Regenerierung falsch beladener 4'-PP Kofaktoren. Die SrfTEII wurde mittels Lösungs-NMR untersucht, die Resonanzen wurden zugeordnet, erste strukturelle Modelle konnte berechnet werden und das Interaktionsverhalten mit verschiedenen Kofaktoren und PCPs wurde analysiert.

Aufschlüsselung der Einzelexperimente und Eigenleistung

Eigenleistung am Projekt TycC3-PCP:

Expression aller markierter und nicht markierter Proteine und deren Aufreinigung

Zuordnung der Resonanzen des Proteins in der apo- und holo-Form mit Hilfe der Software Sparky 3.1 (T.D. Goddard und Kneller)

Strukturrechnung und Strukturverfeinerung der drei strukturell relevanten Zustände in den vier Subformen unter Zuhilfenahme der Software CYANA2.1 (P. Günthert)

Molekulardynamik-simulation des A-states, A/H-states und H-states des TycC3-PCPs mittels Gromacs3.0 (van Gunsteren)

Aufnahme aller NMR-Experimente zur Analyse der Protein-Protein-Interaktion

Modellierung der Komplexe der 4'-PP-Transferase (Sfp) mit Substrat Coenzym A mittels Autodock und des aktiven Komplexes von Sfp, Coenzym A, Mg^{2+} und A-state apo-PCP mit der Software Haddock1.3 (cns 1.1) (A. Bonvin)

Vorbereitung der Proben zum Messen dipolarer Restkopplungen (RDCs) und die Auswertung der RDCs mit der Software RedCAT7 (M. Prestegard) und Vergleich der errechneten Amidvektoren mit den ermittelten

Probenvorbereitung, Aufnahme der NMR-Experimente und Auswertung der D_2O-Austauschexperimente

Vorbereitung und Durchführung der Experimente zur Modifikation des 4'-PP-Kofaktors mit einer Aminosäure (TYR) und Aufnahme und Auswertung der NMR-Experimente

Fremdleistung am Projekt TycC3-PCP:

Klonierung der Konstrukte zur Expression von TycC3-PCP und Sfp (M. R. Mofid)

Aufnahme der zur Strukturanalyse notwendigen NMR-Experimente (F. Löhr)

Installation der Softwares Autodock und Gromacs (M.M. Blum)

Alle verwendeten Materialen wurden durch das Institut (Biophysikalische Chemie) zur Verfügung gestellt.

Eigenleistung am Projekt SrfTEII:

Expression aller markierter und nicht markierter Proteine und deren Aufreinigung

Zuordnung der Resonanzen des Proteins mit Hilfe der Software Sparky 3.1 (T.D. Goddard und Kneller)

Strukturrechnung unter Zuhilfenahme der Software CYANA2.1 (P. Günthert)

Aufnahme aller NMR-Experimente zur Analyse von Protein-Protein-Interaktionen bzw. der Interaktion der SrfTEII mit isolierten Liganden bzw. Experimente die für die Zuordnung des Proteins relevanten sind

Probenvorbereitung, Aufnahme der NMR-Experimente und Auswertung der D_2O-Austauschexperimente

Fremdleistung am Projekt SrfTEII:

Klonierung der Konstrukte zur Expression von SrfTEII (F. Bernhard)

Eigenleistung am Projekt TehA:

Entwicklung von Zuordnungsstrategien und Zuordnung der backbone-Resonanzen der TehA (mit N. Trbovic)

Entwicklung von Strategien zum Strukturrechnen von Membranproteinen basierend auf Abstandbeschränkungen des Protein-backbone, generiert mittels Lösungs-NMR unter Zuhilfenahme der Software CYANA2.1

Entwicklung einer ersten Kalibrierung der Abstandsparameter generiert mit paramagnetischen tags (PRE) zur Verwendung in Strukturrechnungen

Berechnung des ersten Models der TehA, eines Proteins mit sieben Transmembran-helices, ausschließlich mit Lösungs-NMR ermittelten Parametern

Fremdleistung am Projekt TehA:

Herstellung der Proteinproben (F. Bernhard, C. Klammt)

Aufnahme aller NMR-Spektren (F. Löhr)

Mutationen für PREs (N. Trbovic, C. Klammt)

Klonierung (F. Bernhard)

Eigenleistung am Projekt IGP-S:

Sämtliche Experimente, inklusive genomischer Mutationen, Klonierungen, Entwicklung von Aufreinigungsstrategien, Interaktions-studien, Zuordnung der NMR-Spektren sind Eigenleistung.

ERKLÄRUNG

Ich erkläre hiermit, dass ich mich bisher keiner Doktorprüfung unterzogen habe.

Frankfurt am Main, 08.06.2006

Alexander Koglin

EIDESSTATTLICHE VERSICHERUNG

Ich erkläre hiermit an Eides Statt, dass ich die vorgelegte Dissertation über „Strukturelle Dynamik von Peptidyl Carrier Domänen in nicht-ribosomalen Peptid-Synthetasen" selbständig angefertigt und mich anderer Hilfsmittel als der in ihr angegebenen nicht bedient habe, insbesondere, dass aus Schriften Entlehnungen, soweit sie in der Dissertation nicht ausdrücklich als solche mit Angabe der betreffenden Schrift bezeichnet sind, nicht stattgefunden haben.

Frankfurt am Main, 08.06.2006

Alexander Koglin