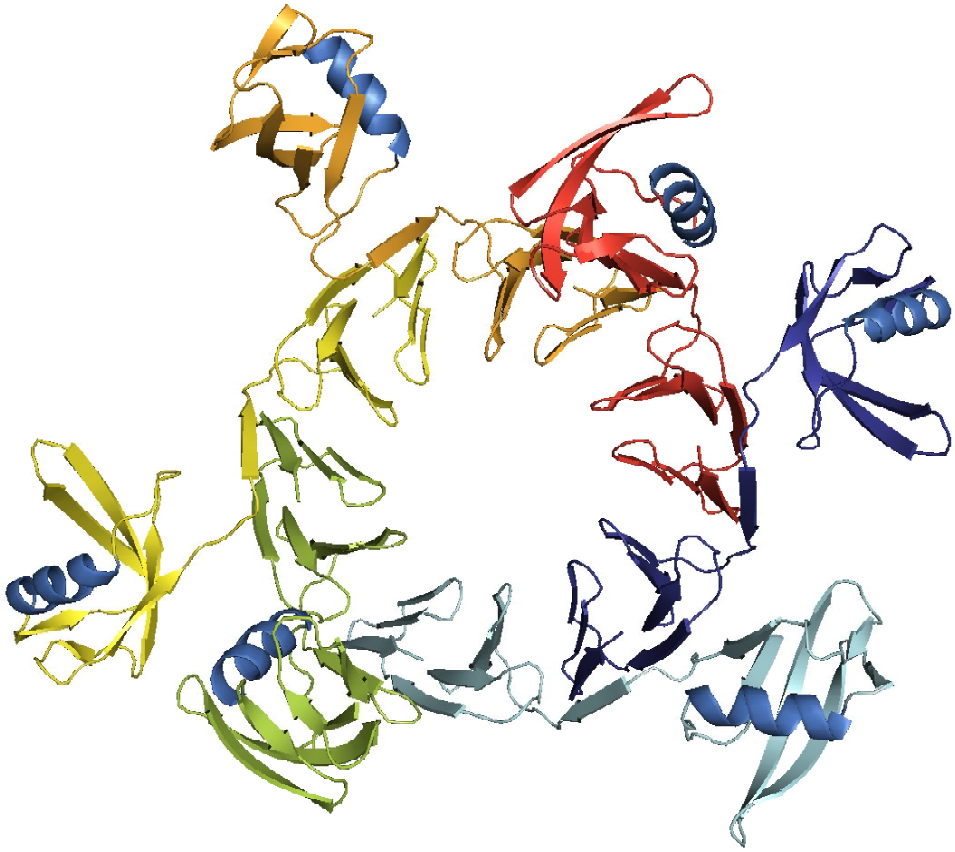
Investigations on the 100 kDa Homohexameric Protein Ph1500 by NMR Spectroscopy



Dissertation

Ilka Varnay

2008

TECHNISCHE UNIVERSITÄT MÜNCHEN
Institut für Organische Chemie und Biochemie

Investigations on the 100 kDa Homohexameric Protein Ph1500 by NMR Spectroscopy

Ilka Varnay

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. St. J. Glaser

Prüfer der Dissertation:

1. Univ.-Prof. Dr. H. Kessler
2. Hon.-Prof. Dr. A. Lupas, Eberhard-Karls-Universität Tübingen

Die Dissertation wurde am 04.11.2008 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 15.01.2009 angenommen.

Ich halte dafür, dass das Glück der Völker nicht in der Menge ihrer Eisenbahnen liegt. Auch nicht die Zukunft Bayerns und Tirols. Man soll mir die idyllische Einsamkeit und die romantische Natur, deren malerische Schönheit im Winter noch ungleich größer ist als im Sommer, nicht durch Eisenbahnen und Fabriken stören. Auch für zahllose andere Menschen, als ich einer bin, wird eine Zeit kommen, in der sie sich nach einem Land sehnen und zu einem Fleck Erde flüchten, wo die moderne Kultur, Technik, Habgier und Hetze noch eine friedliche Stätte weit vom Lärm, Gewühl, Rauch und Staub der Städte übriggelassen hat.

Ludwig II. über den technischen Fortschritt (1878)

Die vorliegende Arbeit wurde am Institut für Organische Chemie und Biochemie der Technischen Universität München in der Zeit von Mai 2004 bis Oktober 2008 unter der Leitung von Prof. Dr. Horst Kessler angefertigt.

Acknowledgements

First of all, I would like to express my gratitude to *Prof. Dr. Horst Kessler* for the excellent working conditions, in particular concerning the spectrometers, for giving me the opportunity to visit international conferences, for the scientific freedom he assured and for supporting me in setting up the cooperation with *Prof. Dr. Andrei Lupas*.

Notably, I sincere appreciate *Prof. Dr. Andrei Lupas* and *Dr. Murray Coles* for agreeing in the cooperation and for offering me a project, which allowed to combine NMR spectroscopy and protein evolution studies. Moreover I wish to thank *Dr. Murray Coles* for helpful discussions.

My further thanks goes to:

Dr. Vincent Truffault for his support especially in the initial stage of this work, for his introduction to the spectrometers, for his encouragement, and for the telephone-hotline when a problem occured at late hours on the spectrometer.

Dr. Sergej Djuranovic, *Astrid Ursinus* and *Dr. Jörg Martin* for providing me with all the samples of Ph1500, Ph1500-N and Ph1500-C, thus building the stable fundament of this work.

Dr. Beate Roeckel for the fabrication of the EM pictures of Ph1500.

All my colleagues and the people from Steffen Glaser's group, Michael Sattler's group, and all other related groups for the nice atmosphere and for interesting discussions.

Dr. Rainer Haeßner, who carries a large responsibility for all spectrometers and computers and is deeply engaged to keep them running.

Again *Rainer* and *Gustav* for their permission to measure at 80°C.

Andreas Enthart for helping me with various computer related problems.

Evelyn Bruckmaier and *Martha Fill* for their engagement in contract administration over all the years.

Dr. Tammo Diercks for supplying me with his pulse programmes.

Johannes Beck and *Murray Coles* for proof-reading parts of this manuscript.

Dominik Heckmann, Helge Menz, Wilbert Snijders, Andreas Zander and *Afra Torge* for their contribution to 'Emma is confused'.

My father for giving me the opportunity to study and for his trust and support.

Contents

List of Abbreviations

Å	angstroem (10^{-10} m)
γ	gyromagnetic ratio ($s^{-1} \cdot T^{-1}$)
η	viscosity ($10^3 kg \cdot m^{-1} \cdot s^{-1}$)
τ_c	correlation time
δ	chemical shift
2D	two dimensional
3D	three dimensional
4D	four dimensional
AAA	ATPases associated with diverse cellular activities
AMA	proteins occuring in *Archaeoglobus fulgidus* and methanogenic archaea
AQUA	suite of programs for Analyzing the QUAlity of biomolecular structures that were determined via NMR spectroscopy
ARIA	Ambiguous Restraints for Iterative Assignment
ATP	adenosine-5'-triphosphate
BLAST	Basic Local Alignment Search Tool
BMRB	BioMagnetic Research Bank
C41 (DE3)	mutant strain of Escherichia coli BL21(DE3) originally described by Miroux and Walker
Cdc48/p97	AAA-family chaperone whose cellular functions are facilitated by its interaction with ubiquitin binding cofactors
CheckShift	program for re-referencing of chemical shifts
COSY	correlation spectroscopy
CRINEPT	CRIPT + INEPT
CRIPT	cross relaxation-induced polarization transfer
CSA	chemical shift anisotropy
CSI	chemical shift index
CYANA	Combined assignment and dYnamics Algorithm for *NMR* Applications
Da	Dalton (g/mol)
DD	dipole-dipole coupling
DNA	desoxyribonucleic acid
E. Coli	Escherichia coli
EM	electron microscopy
FID	free induction decay
FT	fourier transform
GD box	$\beta\alpha\beta$-element with a highly conserved GD sequence
G-T	Guanine - thymin
Het-NOE	Heteronuclear-NOESY experiment
HSQC	heteronuclear single quantum coherence
HMQC	heteronuclear multiple quantum coherence

I	nuclear spin quantum number
INEPT	insensitive nuclei enhancing polarisation transfer
IPTG	isopropyl-beta-D-thiogalactopyranoside
kDa	kilo Dalton (10^3 g/mol)
MEXICO	measurement of exchange rates in isotopically labelled compounds
MPI	Max Planck Institute
MQ	multiple quantum
m_s	azimuthal quantum number
Ni-NTA	Nickel-nitrilotriacetat
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
NS	number of scans
OB fold	oligonucleotide/ oligosaccharide binding fold
PASTA	protein assignment by threshold accepting
PDB	protein data bank
PDB ID	Protein data bank identification number
Ph	*Pyrococcus horikoshii*
Ph1500-N	N-terminal domain of Ph1500
Ph1500-C	C-terminal domain of Ph1500
ppm	parts per million (10^{-6})
PRE	paramagnetic relaxation enhancement
QUEENS	QUantitative Evaluation of Experimental NMR restraints
RDC	Residual Dipolar Coupling
RMSD	root mean square deviation
SimShift	program for predicting dihedral angles from chemical shifts and homology
SHIFTOR	program for predicting dihedral angles from chemical shifts and homology
T_1	longitudinal / spin-lattice relaxation time
T_2	transversal / spin-spin relaxation time
TALOS	Torsion Angle Likelihood Obtained from Shift and sequence similarity
TOCSY	total correlation spectroscopy
tr	denotes spectra using the TROSY unit
TROSY	transverse relaxation optimised spectroscopy
TXI	triple resonance inverse detection
UV	ultraviolet
VAT	Valosine-containing-protein-like ATPase of *Thermoplasma acidophilum*
VATN-n	N terminal part of the N terminal domain of VAT
VATN-c	C terminal part of the N terminal domain of VAT
wt	wild type

Chapter

1

Introduction and Aim of the Work

Since the first weak radio frequency responses from atomic nuclei were observed independently by two groups around F. Bloch and E. M. Purcell in 1946, nuclear magnetic resonance (NMR) spectroscopy has developed to a powerful technique, alternative to X-ray crystallography for providing structures of proteins at atomic resolution. The interest in protein structures follows from the fact that proteins are involved in almost all physiological processes, including catalysis, signal transduction, recognition and regulation, and thus play also important roles in many diseases. From time immemorial it has been the dream of humans to cure the diseases he finds himself tormented with. While examining the effects of drugs, which were initially mostly provided by plants, has always been possible, the design of selective drugs became first possible when the understanding of the effects was brought to the molecular level. With the progress of organic synthesis and the definition of structure-activity-relationships (SAR), it became not only feasible to identify and synthesize natural drugs, but also artificial ones. *Rational Design* together with computational modelling techniques have allowed to optimise drugs for selectivity by fitting specific groups into the binding site of proteins. Providing highly resolved protein structures with accurately defined local conformations is still a time consuming process. Quite reliable predictions of the global fold can be derived much faster using homology modelling techniques, if the structure of a homologous protein is known.

With the increasing number of known structures it appears also possible to gain an understanding of the evolution on a molecular level. Beside the improvement of homology modelling techniques and protein structure classification schemes, the usefulness of protein evolution studies might be less obvious. Since understanding evolution and protein evolution touches the basics of life, investigations in these fields might contribute to solving the conflict between a scientific and a religious or believing conviction or the combination of both. Though, an interpretation of scientific results always needs to be done carefully; the theory of evolution has been often misused as a proof or counterevidence for the respective convictions by a superficial interpretation. Likewise, it has been often misinterpreted in ideological world-views. Apart from the application in structural biology the technique of NMR spectroscopy is also applied in various other important fields, e.g. in organic synthesis and magnetic resonance imaging (MRI).

The strength of X-ray cristallography compared to NMR lies in the possibility to study very large proteins and protein complexes. NMR spectroscopy is more limited in size of the molecules to be analyzed. Its advantage over X-ray cristallography is that experiments are performed in aqueous solution close to physiological pH and salt concentration as opposed to a crystal lattice. Thus, it allows the study of the internal dynamics of proteins and the investigation of interactions of proteins with other proteins, DNA, RNA and smaller ligands.

In the last decades many efforts have been directed to expand the technique of NMR spectroscopy towards the structure determination of larger proteins. For very large protein complexes, for example amyloid fibrils, cryo-electron microscopy becomes more and more meaningful. With the resolution of less than 5 Å already obtained today, the combination of both techniques can be a great benefit.

In the present work, the solution structure of the 100 kDa protein Ph1500 from the hyperthermophilic archae bacteria *Pyrococcus horikoshii*, has been investigated. The function of Ph1500 is still unclear, although its gene environment suggests a role in DNA repair. Electron microscopy with a resolution of about 20 Å has shown Ph1500 to form a hexameric ring around a central pore. The protein consists of two domains (Ph1500-N and Ph1500-C), which can be

expressed separately. The isolated N-terminal domain containing 76 residues is monomeric, while the isolated C-terminal domain contains 71 residues and remains hexameric.

The relevance of Ph1500 in protein evolution studies and the resulting interest in a structural investigation rose from the following context.
The N-terminal domain shows strong sequence similarity to the N-terminal domains of several AAA proteins (ATPases Associated with diverse cellular Activities), notably those of the Cdc48/p97 family. The fold of the relevant domain has been described as a β-clam. AAA proteins are typically involved in unfolding proteins or remodeling protein complexes, and their N-domain are involved in recognition of their substrates. It is therefore assumed that their prime role consists in protein binding. As Ph1500 has no ATPase domain it is clearly not functionally related to the AAA proteins. Ph1500-N is therefore potentially the first example of this type of β-clam outside of the AAA proteins. Moreover the β-clam fold proposed for Ph1500-N is distantly related to members of the cradle-loop barrel meta-fold, and indeed shares their hallmark structural and sequential motif, the GD box. However, the topological relationship between the β-clam and the other members of the meta-fold is complicated, and their evolutionary relationship is as yet unclear.
The C-domain shows no sequence similarity to proteins of known structure. Therefore, it is of interest from the point of view of structural biology. Indeed, homologues of Ph1500 have been defined as targets in structural genomics projects for several years. The fact that none of these projects progressed beyond the stage of purified protein coincides with the experiences from the MPI for Developmental Biology, Tübingen, Department of Protein Evolution, where attempts to crystallize the protein also failed.
The first sequence analysis of Ph1500C also suggested a GD box, and secondary structure prediction suggested an OB barrel fold. The OB barrel fold is another example of a fold that contains a GD box, but is topologically distinct from cradle-loop barrels. Thus both domains potentially represent links that could help to expand the cradle-loop barrel meta-fold.

Corresponding to the molecular weight of 8.5 and 49 kDa, the structural investigations of both domains pose different demands on the technique of NMR. Established standard NMR techniques and recent methodological developments for the investigation of large proteins above 25-30 kDa are described and applied in the present work.

As initially predicted the structure of the N-domain shows a β-clam fold. This fold is similar to a β-barrel fold with the barrel not being fully closed. The C-domain, initially expected to show a OB-fold, turned out to show the first discovered twelve-bladed β-propeller fold. Different relaxation rates of the two domains directed the formation of the full structure of Ph1500 with the N-domain being flexibly attached to the C-domain.

Chapter

2

Introduction to Protein Evolution

This short introduction to protein evolution is largely based on a book article written by A. N. Lupas and K. K. Koretke (2008) and a review written by N. V. Grishin (2001).

The world's proteome is approximated to comprise around a trillion protein-coding genes. This number is still small compared to the number of possible sequences, which is 20 to the power of hundred for a protein consisting of 100 amino acids. However, with the increasing number of known structures it became clear that the world's proteome is not a random sample of the polypeptid space, instead many proteins from different species share recognizable similarity in sequence and structure. From this follows the theory that proteins evolved by combination and recombination from a basic complement of autonomously folding units (domains). The antecedent domain segments are supposed to have emerged in the context of RNA-dependent replication and catalysis from small pre-optimized fragments with secundary structures, which as soon they became longer found out that they could exclude water between themselves ("hydrophobic collapse"). Apparently the ancestral set of folded peptides has not been greatly increased since the time of the last common ancestor.

Folding is essential for proteins to exert their activity. This process does not simply entail an approximate spatial arrangement referred to as molten globule,

but requires the polypeptide chain to assume a specific structure to within fractions of an Angstrom in a reproducible fashion. Even proteins that appeared to be active in a natively unfolded state turned out to also adopt a defined reproducible scaffold.
The process of folding is complicated and easy derailed, thus cells allocate substantial resources to systems ensuring the folding, quality control and turnover of proteins. Surprisingly it was found that most random polypeptide chains do not fold at all, thus natural proteins evidently represent a special case. Consequently a combination of several point mutations leads often to unfolded proteins and also protein design projects frequently do not yield more than molten globules or amyloid aggregates.

In an attempt to explain why proteins are so highly preserved, even after considerable divergence of their sequences, folded proteins can be imagined to be located on islands of stability scattered in a vast ocean of unfolded states. On their respective islands proteins can pace about by mutation, gradually diverging through adaptive changes and neutral drift, but they cannot leave their island without drowning. Occasionally, large and rare events allow proteins to cross to another island of stability, but mostly, proteins will be forced to maintain their folds over billions of years of evolution. As the sequences have to retain common features, they can be used for structure predictions for newly sequenced proteins. Domains in which fold changes took place are supposed to have closely spaced energy minima due to the interchangeability of their fold-determining interactions, thus they can be imagined as cases where islands of stability are close together. If a fold change takes place in evolution, it is still not assured if the new fold will establish, since folding does not imply function, but is rather a prerequisite for it. Furthermore there has to be a functional requirement and an advantage over established siblings.

In the previous years several examples were found for homologous proteins with globally distinct structures, thus giving evidence for a change of fold in evolution. According to the definition, a fold change takes place if the nature of one or more secondary structure elements or their topology is altered. Hence, if a fold change takes place in evolution, the so far commonly accepted presumption that a

similar sequences result in a similar structures is no longer valid. This brings new challenges to the homology modelling techniques and structure classification schemes. Dealing with homology is also problematic, as the question occurs of how it can be assured that two proteins are in fact homologous. However, if there is significant sequence conservation, local structural resemblance and functional similarity it strongly indicates an evolutionary relationship. Since the space of possible sequences exceeds the space of allowed structures by far, it is unlikely that nature can independently find similar sequences. Even more relaxed criteria can be applied to multidomain proteins. If the homologous major domains are more conserved, it is likely that the smaller domains with sequence and local structure similarity are also homologous. Typically, the strongest sequence similarity is solely observed in several conserved motifs, which can be either embedded in similar or quite different sequences and structures, thus it is important to differentiate between global and local homology.

The question how proteins can change their fold in evolution is up to now still largely unexplored. Mechanisms on the gene level that are able to generate new folds are point mutations, indels (insertions and deletions), circular permutations and gene duplications and recombinations. While point mutations and indels are most common, gene duplications and recombinations are less common. However, their power to generate new folds is in reverse order. Gene duplications are leading to tandem repeat sequences. Circular permutations are resulting in the ligation of the termini and the cleavage at another site. Proceeding point mutions can result in substitutions of secondary structures, e.g. the replacement of an α-helix by a loop or β-strand, strand invasions or withdrawals and β-hairpin flips or swaps, respectively.

Concerning the reconstruction of evolutionary events, there are various problems that need to be considered. Primarily, evolution happened just once, its path can therefore not be proven experimentally. The sequence of the common ancestor of a protein can only be indirectly supposed through comparison of the sequences of still existing proteins. As proteins do not fossilize, any direct observation of intermediate forms is impossible. Reconstructing the evolution of protein folds involves demonstrating homology between diverse folds by solving

the structures of putative intermediates in sequence space. This leads to the metafold concept of grouping folds by evolutionary origin. However, most often it is only possible to assume an evolutionary relationship between distantly related proteins with distinct topologies that cannot be inter-converted by a simple topological modification. The exact pathway is difficult to reconstruct, because of the large and unknowable body of missing data. That means that retracing the path of molecular evolution can only be guided by the principles of parsimony and maximum likelihood. Nevertheless, sequence comparison programs turned out to be very useful for structure predictions. There are many examples where the results showed to be reliable, e.g. the prediction that Ph1500-N has a β-clam fold (see also Chapter 5.4). Moreover knowledge of the evolutionary relationships appears to be the best way to classify and to understand protein structures.

Chapter 3

Biochemical and Evolutionary Context of Ph1500

Biochemical work on Ph1500 and homologous proteins and the examination of their functional and evolutionary context summerized in this chapter were done by Sergej Djuranovic at the MPI for Developmental Biology, Tübingen, Department of Protein Evolution, if not stated otherwise.

3.1 Homology between Ph1500-N and the AAA Protein Family

The N-domain of Ph1500 (isolated from *Pyrrococcus horikoshii)* shows strong sequence similarity to the N-domains of several AAA proteins, notably those of the Cdc48/p97 family. The fold of the relevant domain has been described as a β-clam or Cdc48 domain 2-like fold. The AAA protein family, first described by Erdmann et al. (1991) as „**A**TPases **A**ssociated with diverse cellular **A**ctivities", comprises proteins with very different biological functions, ranging from DNA repair and replication to organelle biogenesis, membrane trafficking, transcriptional regulation, protein quality control and protein folding or unfolding. Typical is a highly conserved domain responsible for ATP binding. This domain with a length of 200-250-amino acid residues contains Walker A and B motifs (Walker et al. 1982). Both are involved in binding of the triphosphate moiety of ATP and coordination of an Mg^{2+} ion, which is important for the subsequent hydrolysis.

The domain architecture of AAA proteins consists of a non-ATPase N-terminal domain, which is the putative substrate binding site, followed by one or two copies of AAA domains named D1 and D2 (Fig. 3.1.).

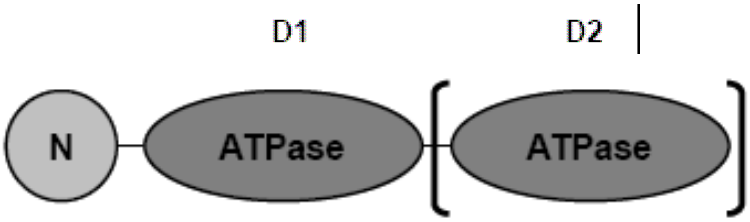


Figure 3.1. Schematic domain organization of AAA proteins.

In members that contain two copies of AAA domains, one of the domains may be degenerate and may be primarily involved in structural stability of the complex (Singh et al. 1999). All AAA proteins whose oligomeric structure has been investigated up to now form hexameric or dodecameric complexes or at least hexamerize in a substrate-dependent manner. Figure 3.2 shows P97 (Zhang et al. 2000) as an example, with the β-clam fold recognizable in the C-terminal part of the N-terminal domain.

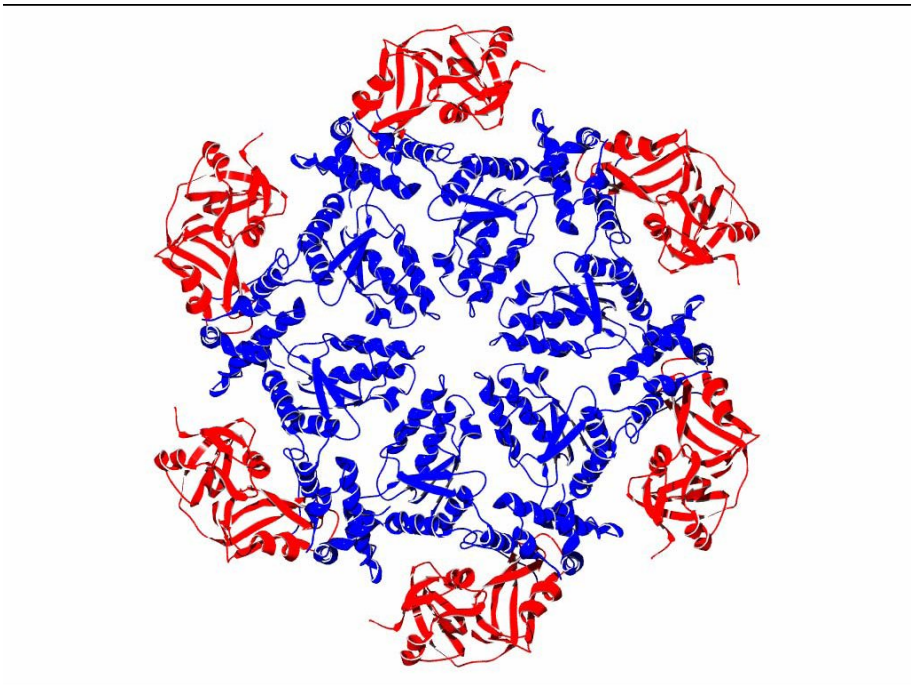


Figure 3.2. X-ray crystal structure of the hexameric protein p97 (figure taken from Zhang et al. 2000) with N-terminal (red) and D1 (blue) domains. The β-clam fold is recognizable in the C-terminal part of the N-terminal domain.

The wide variety in biological functions of AAA proteins originates primarily from divergence in their N-terminal domains. These domains are most often responsible for the interaction with substrates, either directly or through adaptor molecules (Dougan et al. 2002).
Besides the β-clam domain (residues 1-76), Ph1500 contains an additional second non-ATPase domain (Ph1500-C; residues 77-147) that mediates hexamerisation of the full protein. Since Ph1500 lacks the ATPase domain, it is clearly not functionally related to the family of AAA proteins. Thus Ph1500-N represents potentially the first example of this type of β-clam outside of the AAA proteins.

3.1.1 Homology between Ph1500-N and VatN-C

The sequence of Ph1500 was found in a PSI-BLAST homology search with VAT-Nc as query sequence. VAT (**V**alosine-containing-protein-like **A**TPase of *Thermoplasma acidophilum*) is an archaeal member of the CDC48/p97 group of AAA proteins and with its tripartite domain structure N-D1-D2 characteristic for this group. Electron microscopy revealed that the members of this family, including VAT, form hexameric rings with the kidney-shaped N-terminal domain positioned at the edge of the ring structure (Figure 3.3) (Rockel et al. 1999). These findings were further confirmed by the crystal structures of the p97 N-D1 complex (Figure 3.2) and of the complete protein (DeLaBarre and Brunger 2005), as well as by the NMR structure of the N-terminal domain of VAT (Coles et al. 1999).

VAT is able to refold or unfold heterologous protein substrates *in vitro* in dependence of the Mg^{2+} concentration (Golbik et al. 1999). The finding that the change in activity was accompanied with differences in thermal stability suggested the existence of at least two different conformational states in the presence of different concentrations of Mg^{2+} ions. Recently, it was shown that VAT has Mg^{2+} dependent ATP hydrolysis and *in vitro* unfoldase activity against an ssrA-tagged GFP protein (Gerega et al. 2005). Surprisingly, deletion of the N-terminal domain (VATΔN) increased ATP hydrolysis approximately 24 times,

and led to an even more drastic increase (250-fold) in unfoldase activity. These data indicate the role of the kidney-shaped N-terminal domain in the regulation of activity of the full protein, but do not generally exclude the idea that this domain might have a second role in substrate or adaptor molecule binding.

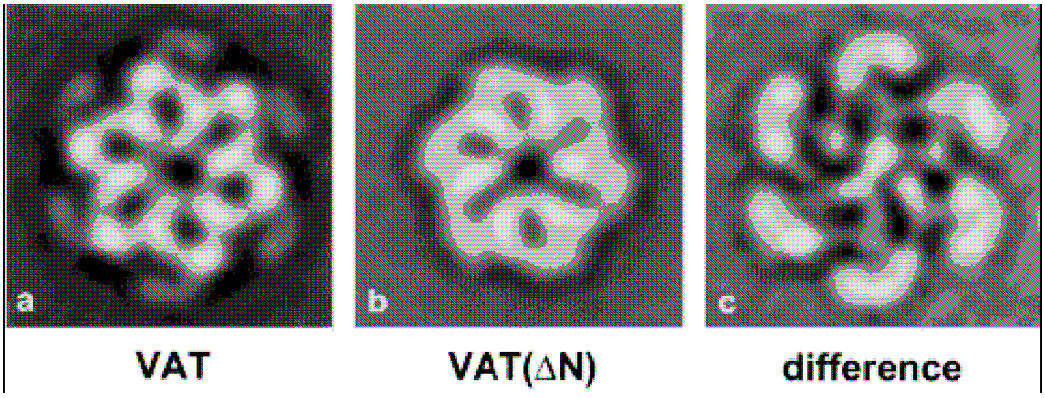


Figure 3.3. 2D average images of VAT (a) and VATΔN (b). Subtractions of image b from image a (c) clearly discloses the position and shape of the N-terminal domain (Rockel et al. 1999).

The NMR structure of the N-terminal domain of VAT (VAT-N) containing two subdomains was solved in 1999 by Coles et al. The first 92 amino acid residues (named VAT-Nn) adopt a double-psi barrel fold, and the C-terminal subdomain (VAT-Nc, 93-185) folds into the relevant β-clam.

The double-psi barrel fold of the N-terminal subdomain (VAT-Nn) is one of the most topologically complex folds in nature. The canonical fold consists of a six-stranded β-barrel capped from both sides by small α-helices. Figure 3.4 shows schematically the topology of the fold so that the pseudo-twofold rotational symmetry of the ββαβ-element becomes apparent. The double-psi structure is formed by two interlocked motifs, each of which comprises a loop and a strand that together resemble the Greek letter psi (Castillo et al. 1999).

Pseudotwofold symmetry of the double-psi barrel fold – also recognizable on the sequence level – suggests the evolution from a homodimer by duplication (Castillo et al. 1999). An evolutionary path for the double-psi barrel fold has been proposed, based on the conservation of a sequence motif with conspicuous Gly-Asp residues – i.e. hxhxxGDxh (h = hydrophobic residue, x = any residue) –

referred to as GD box (Coles et al. 1999). This motif forms an orthogonal turn within two β-strands that flank an α-helix (βαβ), (Figure 3.5). This path was found to connect the double-psi barrel fold with swapped hairpin barrels into a metafold named cradle-loop barrels that seem to originate from a proposed ancestral homodimeric RIFT barrel (Figure 3.6; Alva et al. 2008). Up to now it is unclear if the double-psi barrel fold originated from a RIFT barrel fold by a strand swap or via an alternative pathway in which the strand swapping preceded the duplication and fusion. A homo-dimeric double-psi barrel (Figure 3.6 bottom left; constructed at the MPI Tuebingen) also proved to be stable.

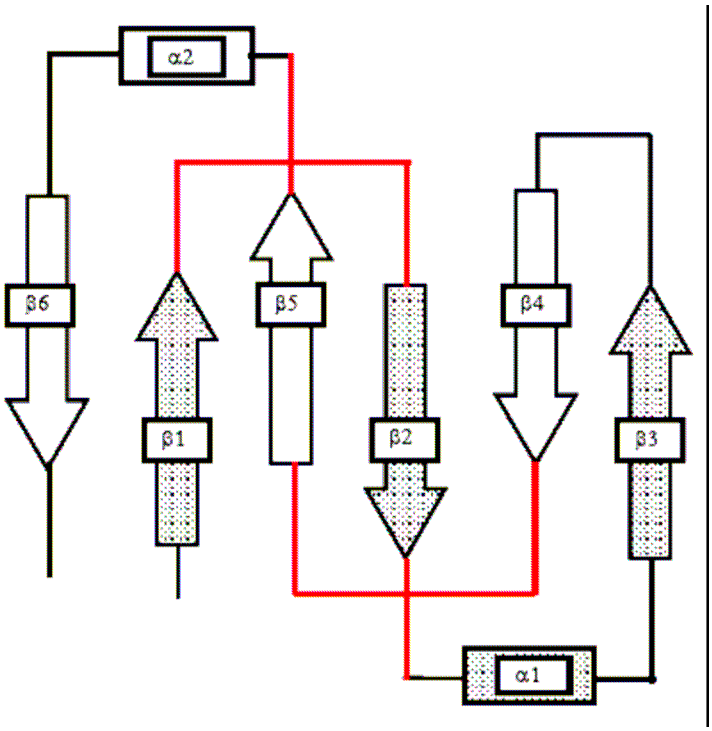


Figure 3.4. Topology of the double-psi β-barrel fold. The loop and β-strand that together resemble the Greek letter psi are colored in red.

A distinctive sequence feature of double-psi barrels is the lack of the proline at the start of the helix, that is highly conserved in the other folds. This is true for both halves, thus the proline was probably missing before duplication. This might indicate a dimeric double-psi ancestor, however an example of this fold has not been discovered so far.

The β-clam fold proposed for Ph1500-N is also distantly related to members of the cradle-loop barrel meta-fold, since it shares their hallmark, the GD box.

However, the topological relationship between the β-clam and the other members of the meta-fold is complicated, and a true evolutionary relationship is yet unclear.

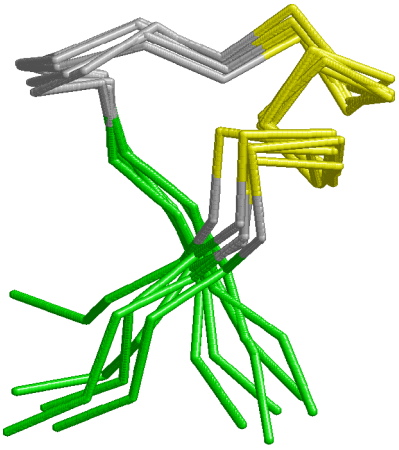


Figure 3.5: GD-box motif, consisting of a βαβ-element with conserved Gly-Asp residues, that make an orthogonal turn within this motif.

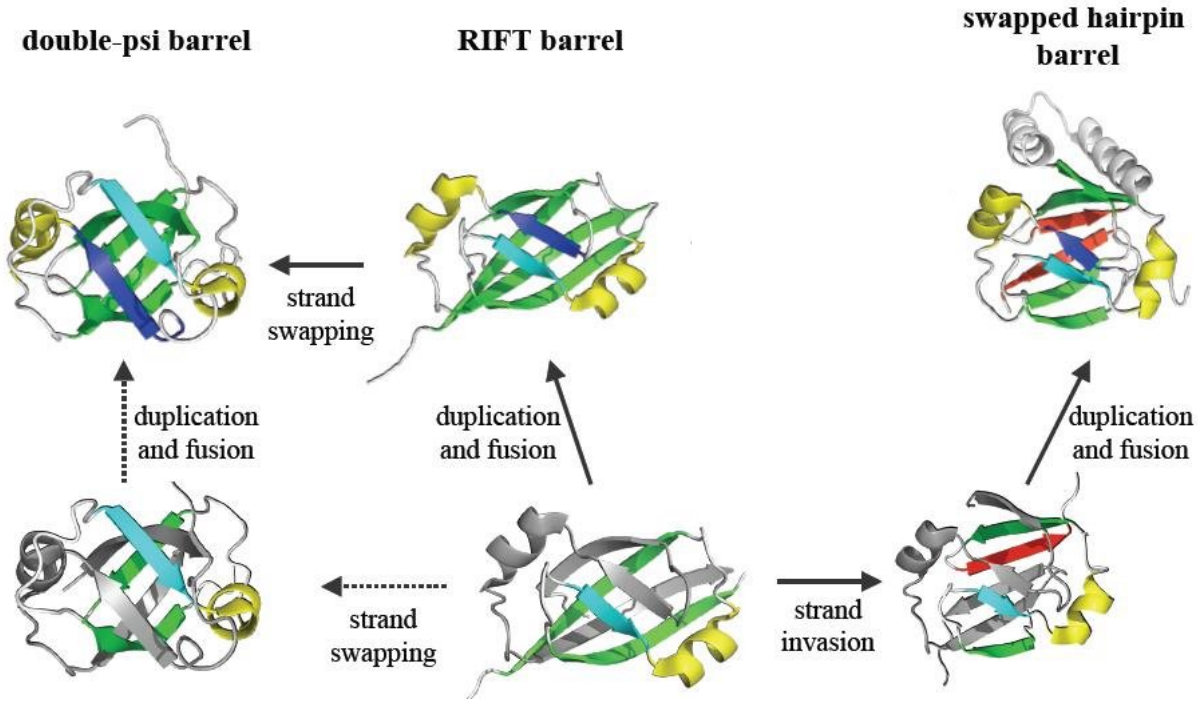


Figure 3.6. Assumed evolutionary relationships within the cradle-loop barrel metafold. Common feature is a βαβ-element with conserved Gly-Asp residues referred to as GD-box, and the peculiar shape of their ligand binding site (cradle loop). This fold family is supposed to originate from a modeled ancestral RIFT barrel (down middle).

3.1.2 Homology between Ph1500-N and AMA proteins

Weak but significant sequence similarity was also found between the N-terminal domains of AMA proteins, occuring in *Archaeoglobus fulgidus* and Methanogenic Archaea (Djuranovic et al. 2006), and the β-clam part of N-terminal domains of the CDC48/p97 group, as well as between AMA and Ph1500-N. AMA proteins differ from the CDC48 group in the absence of the N-terminal double-psi barrel domain. Thus the clam-like domain is the sole N-terminal domain. A chaperone activity of AMA proteins was found to reside entirely within this β-clam like N-domain.

Another important difference between Cdc48 family and AMA proteins is that AMA proteins have a single ATPase domain. In contrast to all other AAA proteins, this ATPase domain is not responsible for hexamerisation, rather the N-domain itself was found to form a hexamer. The alignment of these domains to their homolog from *T. acidophilum* (VAT-Nc) discloses a relatively longer loop region between the first β-strand and α-helix with a conserved GYPL motif within this loop, which is characteristic of the AMA-N sequences. It was shown that this GYPL motif mediates the hexamerization, since single or double mutations within this motif resulted in a loss of the oligomeric ring structure. Modelling of the AfAMA-N domain based on EM data and its similarity to VAT-Nc was done by Dr. Beate Rockel from the MPI for Biochemistry and is shown in Figure 3.7. Loops containing the motif were placed in the central pore of the hexamer, thus favouring the proposed role of GYPL in oligomerization and/or binding of protein substrates.

An intact GYPL motif is also important for the function of the AfAMA-N domain and the full AfAMA protein, since single or double mutants totally abolish chaperone activity. Insertion of the loop containing GYPL motif into the homologous sequence of the VAT-Nnc beta-clam induced oligomerization of this domain (monomer to trimer). Furthermore, trimeric Vat-Nnc chimera gained chaperone activity in heat aggregation assays with citrate synthase as a substrate. Similar examinations are intended with Ph1500-N chimeras.

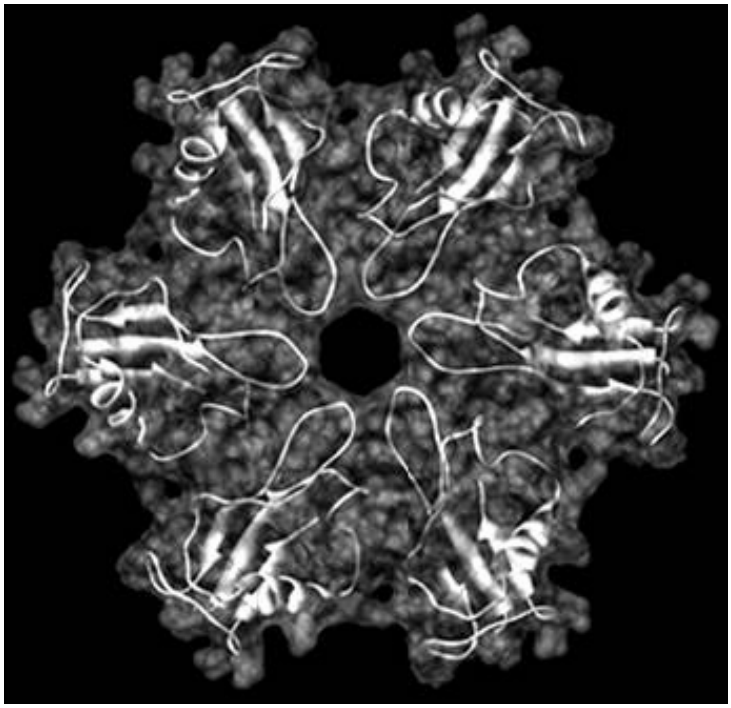


Figure 3.7. Model of the hexameric AMA-N β-clam domain based on EM data and its similarity to VAT-Nc and with the loops containing the motif placed in the central pore.

3.2 Functional Implications for Ph1500

A BLAST search (Altschul 1997) using the sequence of Ph1500 protein as a query found 3 homologous sequences, two from *Pyrococci* (*P. abyssi, P. furiosus*) and one from *Thermococcus kodakarensis*. Analysis of gene loci conservation with "The SEED" server (Overbeek 2005) indicated gene coupling of the Ph1500-like proteins with a gene that is annotated as an endonuclease III (Figure 3.8), an enzyme which makes excision repair of mismatched G-T pairs from damaged DNA molecules (Thayer et al. 1995).

In a pure hypothetical picture based on the gene environment of Ph1500, the protein was initially assumed to recruit or dock endonuclease III to sites of G-T mismatches induced by UV irradiation. According to the proposed mechanism, the hexameric C-domain was assumed to slide down the DNA until it encounters a G-T mismatch, and the β-clam recruits endonuclease III to fulfil its activity, which is excision of the mismatched pair.

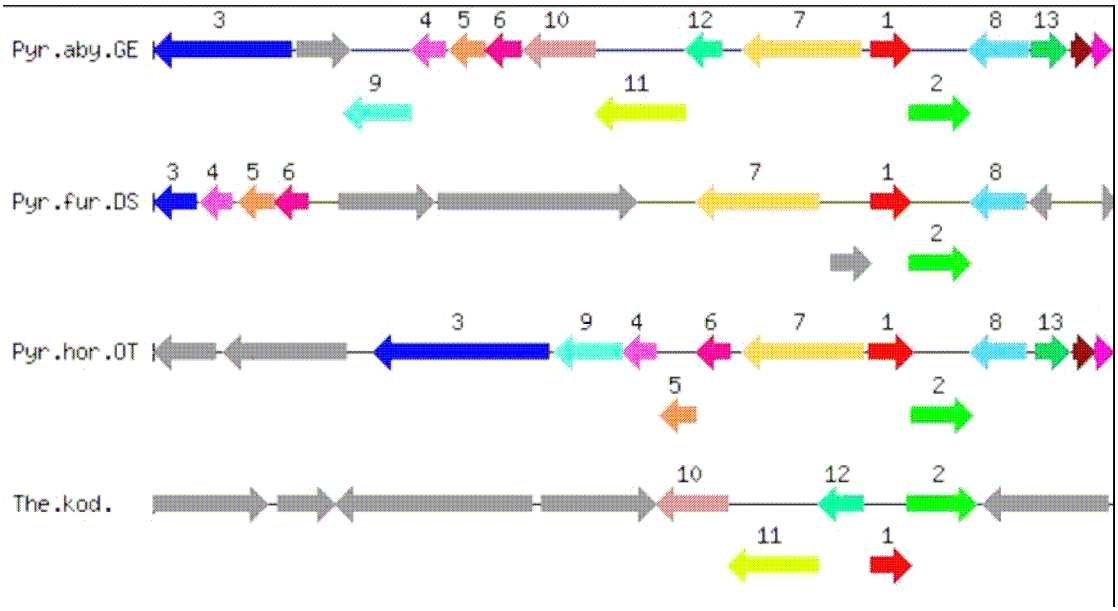


Figure 3.8. Analysis of the Ph1500 gene locus with "The SEED". Ph1500-like genes are labelled in red and numbered 1; endonuclease III genes are labelled in green and numbered 2. The genes were found only in the genomes of *P. abyssi, P. furiosus*, *P. horikoshii* and *T. kodakarensis*.

Constructs of Ph1500 were analyzed for possible interaction with unspecific DNA fragments, but no binding was observed. The endonuclease III from *P. horikoshii* was also cloned and expressed, and purified protein mixed with Ph1500 or its N-and C-domains. The proteins were tested for co-migration on a calibrated gel-sizing column, indicating complex formation. First results showed an interaction of endonuclease III with either N-domain (Ph1500-N) or full Ph1500 protein. However, these results could not to be confirmed by NMR titration experiments. Consequently, it appears more likely that Ph1500 binds to DNA that has been nicked by Endonuclease III. It is also planned to assay Ph1500 for interaction with single-stranded DNA, as well as with mismatched double-stranded DNA, mimicking DNA damage, and possibly to examine the interaction with electron microscopy. Apart from the assumed role in DNA repair, it can be supposed, that Ph1500-N has a role in protein binding, analogous to the substrate recognition role of homologous AAA protein domains.

3.3 Electron Microscopy Studies on Ph1500

Preliminary electron microscopy studies on Ph1500 (Figure 3.9A) were done at low resolution at the MPI Tübingen by Dr. Heinz Schwarz. Data collection of Ph1500 protein particles negatively stained using uranyl acetate, as well as selection and averaging of the obtained images, was done by Dr. Beate Rockel at the MPI for Biochemistry, Martinsried. From the images, 10003 particles were selected and subjected to translatory alignment in order to center the particles properly. An eigenvector-analysis of the aligned data set produced class averages with clear six-fold symmetry (Figure 3.9B). Rotational alignment using a hexameric class average revealed particles of two different ring sizes (Figure 3.9C and D). A schematic representation of Ph1500 is shown in Figure 3.10, based on the finding that separate expression of both domains yields a monomeric N-terminal domain and a hexameric C-terminal domain.

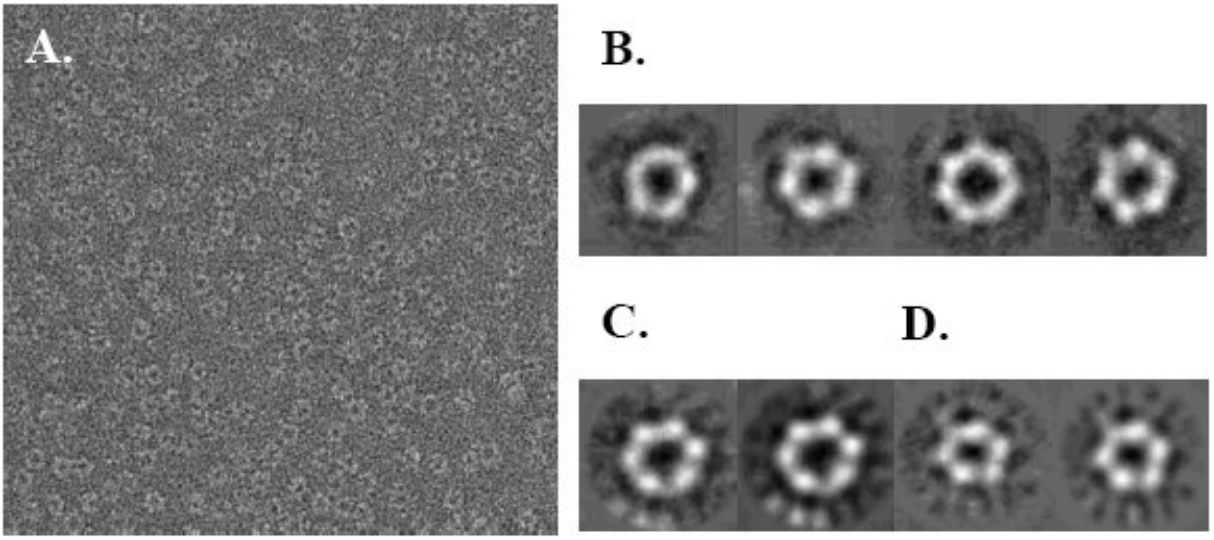


Figure 3.9 Low resolution EM of Ph1500 (A); Ph1500 particles negatively stained with uranyl acetate and class averaged after eigenvector-analysis (B); class averages with different ring sizes after rotational alignment (C and D).

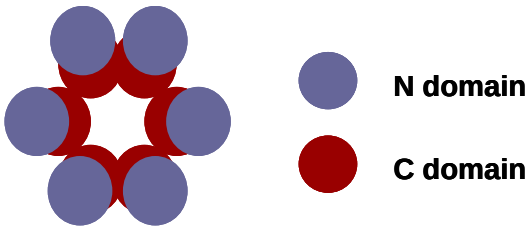


Figure 3.10. Schematic representation of Ph1500. The monomeric 76 residue N-domain is shown in blue and the hexameric 71 residue C-domain is shown in red.

3.4 The Rise of Interest in a Structural Investigation on Ph1500-C

The first sequence analysis of Ph1500-C suggested a GD box likewise Ph1500-N, and secondary structure prediction suggested an OB barrel fold. Beside the β-clam fold, the OB barrel fold is another example of a fold that contains a GD box, but is topologically distinct from cradle-loop barrels (Chapter 3.1.1). Thus both domains could possibly represent links that may help expand the cradle-loop barrel meta-fold.

Furthermore the interest in the structure of the C-domain of Ph1500 has originated by the finding that this domain lacks perceivable sequence similarity to proteins of known structure. Thus it is of interest from the point of view of structural biology. Indeed, homologous of Ph1500 have been defined as a target in structural genomics projects for several years. The fact that none of these projects progressed beyond the stage of purified protein reflects the experience in the MPI Tübingen, where attempts to crystallize the full protein also failed.

Accordingly, it was planned to solve the structure of this protein by NMR spectroscopy. The size of the Ph1500 hexamer (100 kD) is beyond the limits of NMR structure determinations, however the fact that the two domains of the protein can be expressed separately and exhibit high stability allowed the structure determination of the single domains. With a molecular weight of 49 kDa the C-domain was also beyond the limits of high resolution structure determinations, thus it was initially planned to identify the residues involved in the polymerisation and subsequent mutation of these residues to receive stable monomers in the ideal case.

Chapter 4

Solution Structure Determination of Proteins by NMR Spectroscopy

4.1 Characteristics of the Technique of NMR Spectroscopy

In general, spectroscopy characterizes the interaction of electromagnetic radiation with material under given resonance conditions and is normally applied to measure the differences of energy levels occupied by the material. In Nuclear Magnetic Resonance Spectroscopy the interaction takes place between a nucleus and a radiofrequency wave (continuous wave spectrometers) or between multiple nuclei and a high power radiofrequency pulse.

A prerequisite for this interaction is that the nucleus exhibits different energy levels. This prerequisite is only given in a magnetic field where a Zeeman-splitting is observed due to the interaction of the spin angular momentum of the nucleus (which is an intrinsic property of the nucleus) with the external magnetic field. As the size of the splitting, and with it the population difference of states corresponding to different energy levels, depends on the size of the magnetic field, it is advantageous that the magnetic field be as strong as possible. The homogeneous superconducting magnets common today induce resonance conditions at radiofrequencies in the range of 500 to 900 MHz.

A further prerequisite is that the size of the spin angular momentum of a nucleus, which is the sum of the single spin angular momentums of all neutrons

and protons within that nucleus, is different from zero. Nuclei typically used in NMR spectroscopy have a spin angular momentum of I=½ to result in two different energy levels expressed with the azimuthal quantum number $m_s=\{-½;+½\}$. With a spin angular momentum of I=≥1 larger number of energy levels exist, leading to very complex spectra.

If all spin ½ particles of one type, for example all ^1H spins present in a protein, possessed the same resonance frequency, it would be of little interest to measure. Due to the fact that the movement of the surrounding electrons produces tiny magnetic fields, each spin present in the protein experiences a little different total magnetic field. Dependent on this additional local magnetic field, the spins show tiny differences in resonance frequency. To measure the resonance frequencies of the single nuclei (which is the same as the frequencies of precessional motion), initially, all nuclei present in a sample have to be brought into coherence.

In NMR spectroscopy the tiny differences in resonance frequencies can not only be measured, rather in the development of this technique, it was discovered that it is possible to transfer magnetisation from one nucleus to other nuclei in a controlled way. The magnetisation transfer can be mediated via scalar couplings through bonds or via dipolar couplings through space. The former transfer via bonds provides information about the connectivity of atoms. Since the size of the scalar coupling constants is dependent on the conformation of the molecule, it is moreover possible to obtain information about torsion angles by measuring ^3J couplings. The phenomenon of the distance dependent magnetisation transfer via dipolar couplings is exploited in NOESY-type experiments to measure distances between atoms. If a high enough number of distance restraints can be obtained, a structure calculation can be performed.

A protein consists mainly of ^1H, ^{12}C, ^{14}N and ^{16}O isotopes. Spin ½ nuclei present in a protein are ^1H, ^{13}C and ^{15}N with a natural abundance of ~100%, 1.1% and 0.37%, respectively. To increase the sensitivity of carbon and nitrogen it is necessary to label the protein with ^{13}C and ^{15}N isotopes (McIntosh and Dahlquist 1990; Schreiber and Verdine 1991).

4.2 Landmarks towards the Structure Determination of Proteins

Since the early days of solution NMR spectroscopy the size of the molecules under investigation has been steadily increased. Molecules studied by solution NMR spectroscopy reach from small molecules to peptides and proteins and recently even large protein complexes. One of the major impacts on this development came from Richard Ernst and Weston Anderson. In 1966 they used a high-power radio frequency pulse to irradiate the whole spectral bandwidth at once and applied a Fourier Transformation (FT) on the data obtained (Ernst and Anderson 1966), thus they could dramatically reduce the expenditure of time compared to the continuous wave spectrometers common at that time. Moreover the signal to noise ratio could be significantly improved by summing the acquired free induction decays (FID) in the computer. For his contributions to the development of the methodology of high resolution NMR spectroscopy Richard Ernst was awarded the Nobel Prize in 1991.

In 1971, the first two-dimensional correlation spectrum for protons, today known as COSY, was reported by Jean Jeener at the Ampere Summer School in Yugoslavia (Jeener 1971 and 1994). For providing the fundament for all multi-dimensional NMR experiments introduced so far, he was dignified with the Russel Varian Prize in 2002.

The application of two-dimensional Fourier transform spectroscopy to proteins was first reported by the group of Kurt Wüthrich (Wüthrich et al. 1982; Wüthrich 2001), who was awarded the shared Nobel prize in 2002. By developing new pulse sequences, especially by using the nuclear Overhauser effect as a convenient way for measuring distances within proteins (Jeener et al. 1997; Kumar et al. 1980), they paved the way of NMR spectroscopy as a tool for the structure determination of proteins.

4.3 Standard Methods for Protein Structure Determination

4.3.1 Protein Structure Determination based on Dihedral Angle and Rotamer Restraints

The tertiary structure of a protein is described well by its internal torsion angles; the backbone dihedrals Phi (Φ), Psi (ψ) and Omega (ω) and the sidechain rotamers Chi1 (X_1) and Chi2 (X_2) (Figure 4.1). If these could be measured accurately enough, it would be possible to determine the structure via torsion angles alone. However, it is not possible to measure the exact backbone dihedral angles, and thus it is necessary to define the structure based on additional distance restraints. Backbone Φ and Ψ dihedral angles can be measured indirectly by measuring the ^{3}J-H^N(i)H$^\alpha$ couplings in a HNHA experiment and ^{3}J-N(i+1)H$^\alpha$ couplings in a HNHB experiment, respectively. Alternatively, they can be indirectly derived from secondary chemical shifts (deviation from random coil values; Wishart et al. 1992), which were shown to be sensitive to the backbone conformation (Wishart and Sykes 1994). Typical ^{3}J (HN-Hα)-coupling constants in secondary structure elements of proteins are given in Table 4.2 (Bystrov 1976; Wüthrich 1986).

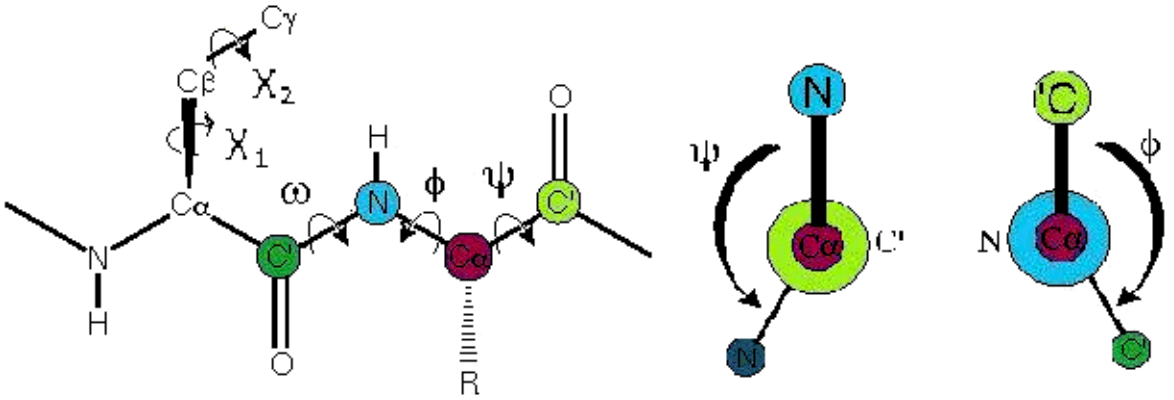


Figure 4.1. Phi (Φ), Psi (ψ), Omega (ω), Chi1 (X_1) and Chi2 (X_2) torsion angles in proteins.

Since the secondary structure elements (β-sheets and α-helices) possess characteristic Φ and Ψ angle combinations and accordingly secondary chemical shifts, their presence can be predicted from backbone and Cβ shifts. Typically, Cα and C' resonate downfield when located in an α-helix and upfield when located in a β-strand; Cβ and Hα behave contrarily (Table 4.1). The use of secondary chemical shifts for secondary structure prediction requires careful referencing, as resonances are also sensitive to pH, buffer conditions and temperature.

Secondary structure element	**^{3}J (HN–Hα) [Hz]**	**Secondary chemical shifts**			
		Cα	**C'**	**Cβ**	**Hα**
α-helix	3.9	> 0	> 0	< 0	< 0
antiparallel β–sheet	8.9	< 0	< 0	> 0	> 0
parallel β–sheet	9.7				
random coil	6-8				

Table 4.1. Typical ^{3}J (HN–Hα)-coupling constants in secondary structure elements of proteins (Bystrov 1976; Wüthrich 1986) and sign of the secondary chemical shifts of Cα, C', Cβ and Hα in α-helices and β-sheets with respect to random coil. According to the definition the random coil secondary chemical shifts are zero.

Different methods have been established to predict the secondary structure of a protein, either based only on secondary chemical shifts (CSI method; Wishart and Sykes 1992 and 1994) or empirically based on chemical shift and residue type homology (TALOS (Cornilescu et al. 1999), SHIFTOR (Neal et al. 2006) and SimShiftDB (Ginzinger and Fischer 2006)).

The widely used computer program TALOS searches a database for the 10 best matches to the chemical shifts and to the sequence of a given residue triplet in the protein of interest. If these 10 matches indicate consistent values for the central residue's phi and psi angles, then their averages and standard deviations are used as a prediction. However, if the 10 best matches have mutually inconsistent values of phi and psi, the matches are declared ambiguous, and no prediction is made for the central residue. The database contains ^{13}Cα, ^{13}Cβ, ^{13}C', ^{1}Hα and ^{15}N chemical shifts for 186 proteins for which a high resolution X-ray structure is available.

The program SHIFTOR is similar to TALOS, it predicts Φ, Ψ, Ω and X_1 torsion angles using only ^{1}H, ^{13}C and ^{15}N chemical shifts. The time needed for the database search is noticeable reduced.
The program SimShift also runs much faster, the chemical shifts of all proteins contained in the database as well as of the protein of interest are re-referenced with an extra program CheckShift and the database was greatly expanded by including shifts back-calculated from a large number of crystal structures. Moreover similar regions are identified instead of similar triplets with the attempt to direct the process towards a prediction of not only the secondary structure, but also the tertiary structure.

Backbone Ω angles are normally always 180° with a tolerance of about 10° due to the partial double bond character of the peptide N^{H}-C'-bond with the trans-conformation being less sterically hindered. The only exception is proline which lacks the amide proton and thus the cis- and trans-conformation have similar energies. The chemical shift of Cγ and the intensity of the H^{α}-H^{α} (i-1) cross peak must be used to determine the isomeric form. Cis-prolines show Cγ chemical shifts around 24-25 ppm and a strong H^{α}-H^{α} (i-1) crosspeak, whereas trans-prolines show Cγ chemical shifts around 27-28 ppm and a weaker H^{α}-H^{α}(i-1) crosspeak.

Protein sidechains interchange between three preferred rotameric states corresponding to torsion angles of 60°, 180° and –60° between N and Cγ for Chi1 (X_1) and Cα and Cδ for Chi2 (X_2) etc. Within these rotameric states there is moreover a more or less strong preference of one or two of the three possible conformations. This is due to sterical hindrance within the respective residue, dependent on the bulkiness of the sidechain and Φ and Ψ torsion angles, but also due to sterical hindrance between the respective residue and the surrounding residues. Thus, residues located on the surface of the protein often have more flexible sidechains, while residues located in the hydrophobic core of the protein normally show a stronger preference of one conformation. A strong preference of one conformation allows application of a dihedral restraint for the sidechain rotamer with a tolerance of usually 30°. Otherwise the sidechain is regarded as flexible.

Sidechain rotameric states, and thus stereospecific assignment of H^{β} atoms and valine methyl groups, can be derived from the information contained in the HNH-, CNH-, HCH- and CCH-NOESY spectra by examining the intensities of intraresidual H^{N}-$H^{\beta1}$, H^{N}-$H^{\beta2}$, H^{α}-$H^{\beta1}$ and H^{α}-$H^{\beta2}$ crosspeaks. Also sequential and across strand NOE crosspeaks can be consulted as well as ^{3}J-N^{H}-$H^{\beta1}$ and -$H^{\beta2}$ couplings derived from the HNHB experiment.

4.3.2 Protein Structure Determinations based on Distance Restraints

The nuclear Overhauser enhancement (NOE) terms the magnetisation transfer through cross-relaxation between a nucleus and the surrounding nuclei nearby and is exploited in NOESY-type experiments. The intensity of the NOE signals depends on the distance (r) of the nuclei and is with common magnetic field strengths and in the range of typical correlation times of proteins proportional to $1/r^{6}$. Thus, measurement of the NOE allows a semi-quantitative estimation of the distance between two nuclei and can be translated into a distance restraint (Kumar et al. 1980). The magnetisation can be passed on to a third nucleus by indirect magnetisation transfer over a longer time period. Consequently, it is advisable to measure a NOE build-up curve to decide for the optimal mixing time, which is a compromise between preferably far-reaching NOEs and the minimization of spin diffusion leading to misinterpretation of NOEs. Another problem is a rather poor spectral distribution and severe signal overlap in a 2D experiment, which can be partly overcome by combining a HSQC-type experiment with the NOESY experiment to a 3D NOESY-HSQC experiment (HNH-NOESY (Griesinger et al. 1987) and HCH-NOESY (Fesik and Zuiderweg 1988)). Additional information can be derived by attaching a further HSQC sequence and leaving out the proton incrementation time, thus resulting in a 3D HSQC-NOESY-HSQC experiment like NNH- (Zhang and Forman-Kay 1997), CNH- and CCH-NOESY (Diercks et al. 1999). The magnetisation transfer pathway together with an example of structural information derived from various common 3D NOESY experiments are shown in Figure 4.2.

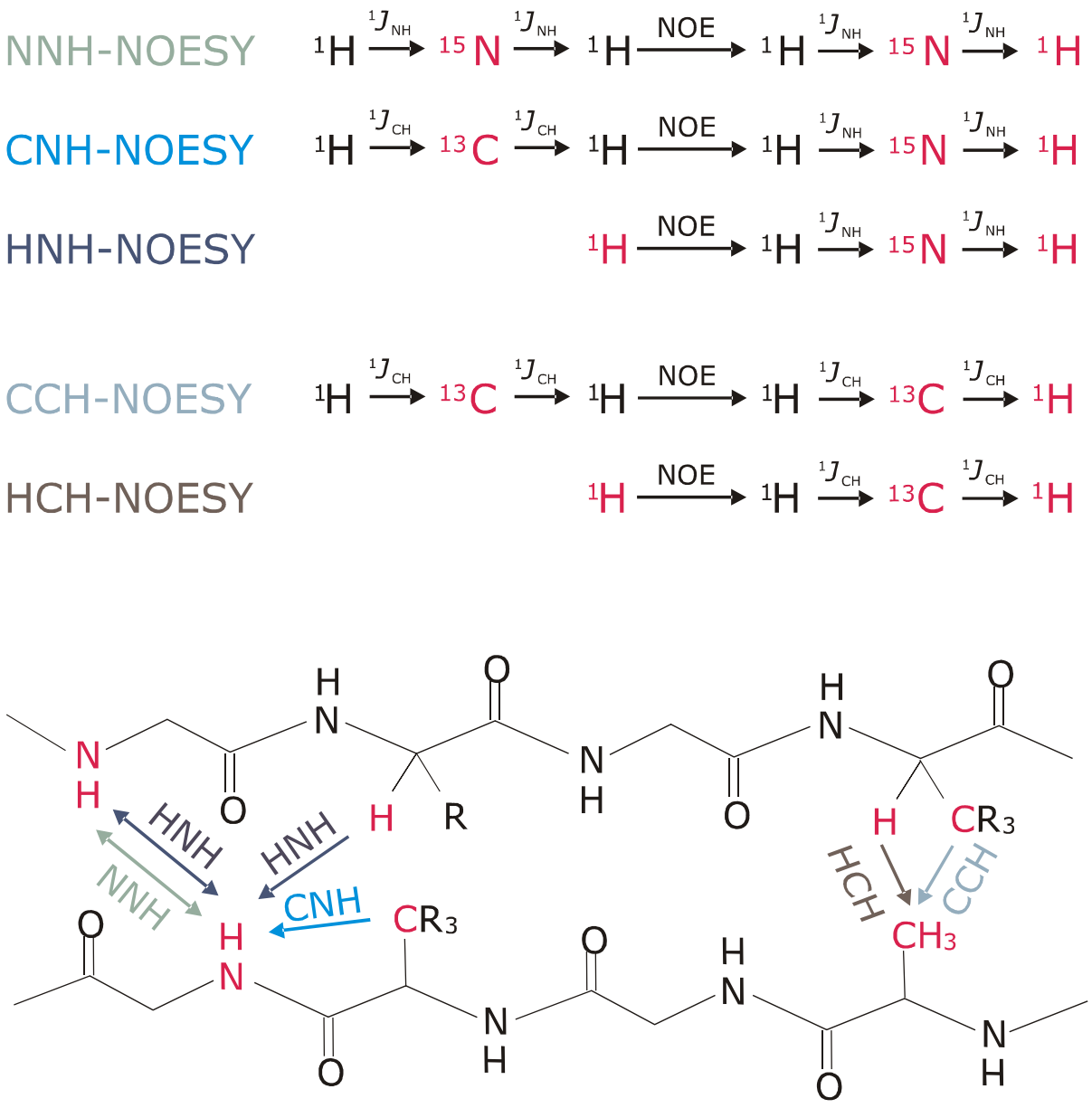


Figure 4.2. Coherence transfer in commonly used 3D NOESY type experiments. Nuclei on which time was incremented or which were directly detected are shown in red. Examples for possible structural information derived from the respective experiments are shown below in an excerpt from an antiparallel β-sheet.

4.3.3 Water Exchange Measurement with a MEXICO Experiment

Hyrogen exchange rates between labile protons such as backbone and sidechain amide protons and the solvent (typically water) can be measured with a MEXICO experiment (**M**easurement of **EX**change rates in **I**sotropically labelled **CO**mpounds) (Gemmecker et al. 1993). The speed of exchange directly reflects the solvent accessibility. As solvent accessibility also depends on protein interactions, MEXICO experiments can further be used for protein interaction studies.

The MEXICO experiment starts with excitation of all protons. Subsequently magnetisation of protons bound to ^{13}C or ^{15}N is filtered out (hence requiring isotropically labelled compounds). The only magnetisation that is retained, is that of water protons that have exchanged with labile protons during the mixing time. Afterwards a standard ^{15}N-HSQC pulse sequence follows. In a resulting MEXICO spectrum only amide groups with solvent accessibility are visible, with the intensity depending on the mixing time and exchange rates. The exchange rates can be calculated after measurement of the intensities obtained at different mixing times and by subtracting them from the standard 2D ^{15}N-HSQC.

4.3.4 Extent of Motion Measurement with a Heteronuclear-NOESY Experiment

The extent of motions in the pico- to nanosecond timescale can be derived from a $^{15}N\{^{1}H\}$-heteronuclear NOE experiment (Het-NOE) (Farrow et al. 1994), which allows an estimate of backbone flexibility, thus identifying more or less ordered regions.

The Heteronuclear-NOE experiment is similar to a ^{15}N-HSQC experiment, merely the magnetization is transferred from proton to nitrogen via space, rather than being transferred via bonds. The occurrence of a Het-NOE requires both, nitrogen and proton magnetization to be aligned along Z. During a presaturation of amide protons dipolar interactions occur between the saturated amid protons and their bound nitrogen. The intensity of the Het-NOE and therefore the NH signal in the Het-NOE spectrum is directly correlated to the flexibility of the backbone.

4.4 Selective Proton Flipback Techniques for Fast Pulsing

Since measuring time in NMR spectroscopy is normally limited and also very costly, it is desirable to design experiments that allow to receive the same information in less time. With this objective the selective proton flipback technique for fast pulsing was developed, focusing on the most time consuming part of the pulse sequence, the recovery of magnetisation.

Usually, only polarisation of a limited sub-set of all protons is converted into observable coherence. The selected proton polarisation is coupled via extensive dipolar interactions to the pool of unused polarisations. Consequently, recovery of unused polarisation can be used indirectly and unspecifically to cool the proton lattice (Pervushin et al. 2002) and thus, accelerate re-equilibration for the selected proton subset. This principle was transferred to HSQC-based multi-dimensional out-and-back experiments that exploit only polarisation of ^{15}N-bound protons and furthermore the flip-back technique was extended from water to all non-^{15}N-bound protons (Diercks et al. 2005). The underlying separation of H^N polarisation from unused proton polarisation can be either achieved through positive or negative selection by J-coupling, or by using band-selective pulses.

Fast pulsing achieved by selective proton flipback techniques was applied in various triple resonance experiments in the present work.

4.5 Recent Methodological Approaches for Proteins above 25-30 kDa or Protein Complexes

There are two major problems of NMR spectroscopy of proteins above 25-30 kDa, one is poor signal to noise ratios and the other is severe spectral overlap. Both are due to slower tumbling and therewith faster transverse relaxation resulting in broad lines. Spectral overlap is additionally intensified due to the increased number of atoms present in the protein and usually becomes the limiting factor. Thus, the number of NMR structures published in the PDB Databank for proteins with a molecular weight above 30 kDa is quite limited (Figure 4.3). In recent years there were a number of publications which introduced new methodological approaches aimed to overcome these problems in

addition to technical advances and advances in molecular biology (also perceivable in Figure 4.3).

It should be admitted that there is not only a problem of feasibility, as the time needed to solve a structure with NMR spectroscopy is not linearly increasing with the size of the protein, but rather exponentially due to more severe ambiguity in resonance and NOE assignments.

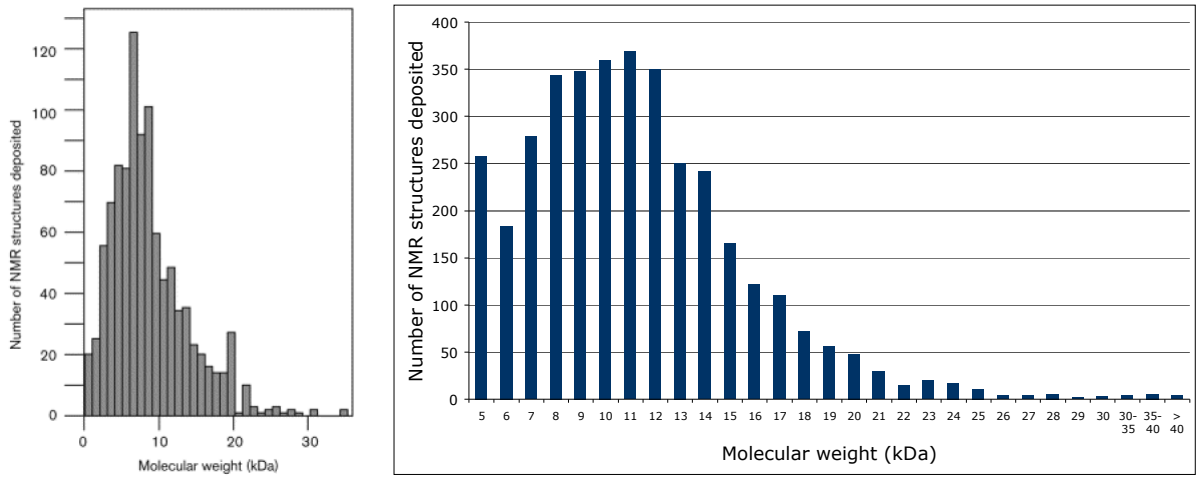


Figure 4.3. Molecular weight distribution of NMR structures deposited in the PDB Databank till 1997 (left) (figure taken from Güntert 1998) and till 2008 (right).

Apparently, the size of a protein domain (an autonomously folding unit) is limited to about 300 residues. If a protein consists of two or more domains and provided that the single domains can be expressed separately and that they are stable on their own, it is common to analyse the single domains and assemble them later on. In the present work this was done with the two domains of Ph1500. The various existing techniques that allow mapping of binding or complexation surfaces are less limited in size of the protein or protein complex.

Very large proteins are often homooligomers. In preferable cases NMR investigations can be also performed on the monomers, provided that they are stable. If the oligomerisation surface is not too large, there might be an opportunity to prevent the oligomerisation and receive stable monomers, for example by changing the salt concentration or by identifying and mutating or deleting the residues involved in the interface (this was initially attempted in the present work, see also chapter 6.7).

Overview of recent developments:

- NMR Hardware

Stronger magnets lead to reduced signal overlap and to increased signal to noise ratio. Cryogenic probes improve the signal to noise ratio up to a factor of 4 by cooling the probe and thus reducing the resistance and thermal noise. The new generation of consoles shows also reduced electronic noise and allows a faster switching between single pulses.

- Deuteration and TROSY

By deuteration of a protein (LeMaster and Richards 1988) the relaxation sources are essentially reduced ($\gamma^H=6.5\cdot\gamma^D$). On parallel the obtainable information from a fully deuterated sample is restricted. Thus the deuteration level has to be carefully chosen in dependence of the different experiments required. It is most often a compromise between reduced relaxation leading to line narrowing and a reduced proton concentration leading to reduced signal to noise in experiments with non-exchangeable protons involved in the magnetisation transfer pathway.
The gain in signal to noise is especially significant if **T**ransverse **R**elaxation-**O**ptimized **S**pectroscop**Y** (TROSY) (Pervushin et al. 1997) is applied on a deuterated sample. In TROSY type experiments only one (the narrowest) of the four components in a non-decoupled ^{15}N-HSQC (Figure 4.4) is selected. Each of the four components has different relaxation rates and thus linewidth due to constructive or destructive interference of the two main relaxation sources for ^{1}H and ^{15}N, namely dipole-dipole (DD) coupling and chemical shift anisotropy (CSA). At high magnetic fields CSA relaxation, which is proportional to B_0^2, becomes comparable to field independent DD relaxation and a constructive cancelling of transverse relaxation can be observed. The optimal field strength for amide NH is 1 GHz. The pulse sequence of TROSY differs from ^{15}N-HSQC in missing ^{1}H decoupling during ^{15}N evolution and ^{15}N decoupling during ^{1}H acquisition.
Besides of the field dependence, the TROSY effect is also dependent on the correlation time of the protein. In contrast to small proteins, the loss in signal intensity due to rejection of some coherence pathways is compensated for large proteins (>20 kDa) at high magnetic fields. The larger a protein (<200 kDa), the more pronounced is the gain in sensitivity and resolution by TROSY.

For the 49 kDa C-domain of Ph1500, sufficient information for a sequential assignment was provided by TROSY type backbone experiments recorded at an increased temperature (318 K) on a 85% deuterated sample.

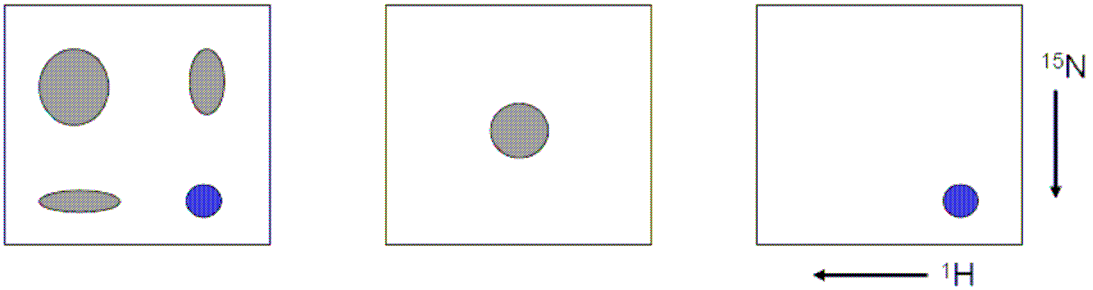


Figure 4.4. Non-decoupled HSQC (left), decoupled HSQC (middle) and TROSY-HSQC experiment (right). The narrowest peak (blue) is due to the constructive canceling of transverse relaxation caused by chemical shift anisotropy (CSA) and by dipole-diploe (DD) coupling at high magnetic field.

- CRIPT and CRINEPT

Due to very fast transverse relaxation (T2) for very large proteins (>200 kDa) the TROSY experiment with a through-bond scalar coupling via INEPT transfer becomes inefficient. Instead Cross Relaxation-Induced Polarization Transfer (CRIPT) between dipole-dipole coupling and CSA relaxation can be used to transfer in-phase ^{1}H coherence to ^{15}N coherence in ^{15}N-^{1}H moieties or a cross relaxation-enhanced polarisation transfer (CRINEPT=CRIPT+INEPT) (Riek et al. 1999).

A CRINEPT experiment was also recorded on the 100 kDa full protein Ph1500, though the size of the protein is on the lower limit where the CRINEPT gains over the TROSY experiment thus both spectra yielded a comparable information content.

- Selective isotope labelling

Selective isotope labelling is either aimed to reduce spectral overlap by reducing the number of signals and relaxation sources, or to provide further information for example about the type of residue by a selective labelling of single types of residues (see also chapter 6.3). A special case of selective isotope labelling is segmental labelling (Xu et al. 1999) that allows cancellation of all signals arising from the chosen segment.

- Selective isotope labelling and Methyl-TROSY

Selective isotope labelling can be used to minimize the relaxation sources by a deuteration and to reintegrate single protons. Normally those of methyl groups from isoleucine, leucine and valine are reintegrated which can be achieved by using for example 2-keto-3-d_2-1,2,3,4-^{13}C-butyrate and 2-keto-3-methyl-d_3-3-d_1-1,2,3,4-^{13}C-butyrate as precursors (Figure 4.5; Rosen et al. 1996; Goto et al. 1999) in minimal media. With ^{13}C in the sidechain it is possible to connect the methyl group to the backbone, the disadvantage is that the one-bond ^{1J}C-C coupling needs to be refocused. With one protonated and ^{13}C labelled methyl group of Isoleucine, leucine or valine and one ^{12}C and deuterated methyl group a better resolution can be achieved, since a major contribution to relaxation is eliminated. The removal of intraresidue NOE contacts further improves the sensitivity and simplifies the spectra. The spin system is also 'linearized' for an efficient coherence transfer in the MQ-CCH-TOCSY experiment.

Methyl-TROSY based experiments (Yang et al. 2004; Tugarinov and Kay 2004; Tugarinov et al. 2004; Tugarinov and Kay 2005), for example MQ-CCH-TOCSY and HMQC-HCH-NOESY, exploit the property that in the macromolecular limit destructive interference between the multiple ^{1}H-^{1}H and ^{1}H-^{13}C dipolar interactions that relax a $^{13}C^{1}H_3$ spin system cause some density matrix elements to relax much more slowly than others. The HMQC experiment was found to be already optimal with respect to the methyl-TROSY effect. In contrast, in HSQC experiments rapidly and slowly relaxing elements are mixed several times. It is possible to obtain a medium resolution structure based on methyl-methyl and methyl-NH contacts. The average content of isoleucine, leucine and valine residues is 21%.

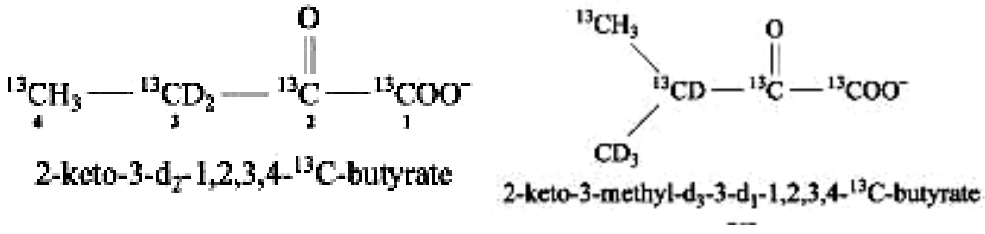


Figure 4.5. Precursors for a selective re-integration of protons in methyl groups of isoleucine (left), leucine and valine (right; Rosen et al. 1996 ; Goto et al. 1999).

- 4D Experiments

By introducing a further dimension 4D experiments (e.g. 4D HNNH-NOESY; Grzesiek et al. 1995) allow to essentially reduce the problem of spectral overlap, though by introducing a further incrementation time in the pulse sequence, the problem of faster transverse relaxation and thus a faster decay of signal is more severe.

- Residual Dipolar Couplings (RDC's)

Long-range orientational information can be obtained by measuring residual dipolar couplings (RDC's) (Tolman et al. 1995). Therefore measurement of RDC's can also be used to derive information about the orientation of domains towards each other.

Usually spatially anisotropic dipolar couplings are averaged to zero due to the isotropic motion of a protein in solution. Partial alignment of the protein leads to an incomplete averaging. An alignment can be achieved for example by adding a paramagnetic tag (if the molecule is not intrinsically paramagnetic) or in liquid crystalline media, lipid bicelles or polymer gels. The occurrence of residual dipolar couplings introduces an extra splitting in the ^{15}N dimension in a non-^{1}H decoupled HSQC spectrum. The size of the splitting depends on the angle towards the alignment tensor and the external magnetic field, respectively. The indirect measurement of this angle thus provides long-range information about the orientation of the ^{15}N-^{1}H-bond vectors towards each other.

- Paramagnetic Relaxation Enhancement (PRE)

By covalently attaching a paramagnetic group to the protein long range (~10-35 Å) distance information can be obtained by Paramagnetic Relaxation Enhancement (PRE) measurement (Gillespie and Shortle 1997), since the relaxation enhancement on single nuclei is dependent on their distance ($1/r^{-6}$) to the unpaired electron of the paramagnetic group. Typically pseudo-contact shifts are measured at the same time.

- ^{13}C-detection

For large proteins the sensitivity gain of an INEPT transfer to ^{13}C is strongly reduced due to the loss of coherence through relaxation, thus ^{13}C-detection (Bertini et al. 2004; Bermel et al. 2006; Hu et al. 2006) might be an alternative to ^{1}H-detection. Recently it was shown on a very large protein (480 kDa) that a ^{13}C detected ^{13}C-^{13}C NOESY experiment allows for the detection of one and two-bond carbon correlations (Matzapetakis et al. 2007). With increased molecular weight, NOE intensities are gaining from longer longitudinal relaxation times and increased cross relaxation.
^{13}C detection requires specially designed pulse programs and probe heads with the coil for ^{13}C detection as close as possible to the sample.

- Low viscosity solvents

A fundamental line narrowing can be achieved by encapsulating the protein in reverse micelles (Wand et al. 1998) with water in the interior space and low viscosity solvents in the exterior space (e.g. short chain alkanes or supercritical carbon dioxide), provided that its native tertiary structure is maintained. To keep these solvents liquefied specially designed high pressure tubes are required. The advantage of measuring the protonated protein in low viscosity solvents is that the effective correlation time becomes nearly independent of the size of the protein and critical data are maintained that can be lost by deuteration and selective reprotonation. It must be noted that this is an emerging technique that is not yet well established.

4.6 Automated versus Manual Structure Determination

Protein assignment and structure determination are laborious and time consuming processes, thus it is desirable to replace human work by the computer in as many steps of this process as possible. To date, there exist already a number of automated programs, e.g. ARIA (Rieping et al. 2007) and CYANA (Günthert 2004), for assignment and structure determinations, though several problems are still inherent in these programs and it is difficult to state how big

the gain of time actually is. Different factors responsible for usually less precise defined local conformations are for example a loose scaling of distance restraints, the exclusive use of upper-distance restraints, over-reliance on NOE data over other forms of data, more frequent misassignments and missed consideration of overall weak traces in NOESY spectra. Besides, it might be problematic to verify the results of an automated process without any experience in the manual process.

Thus for the structure determination of Ph1500-N a manual procedure was preferred. However, for Ph1500-C a semi-automated process was used that allowed to considerably reduce the expenditure of time without incorporating any of the problems mentioned above. This could be achieved by using a script (written by Dr. M. Coles) to automatically transfer the assigned resonances of the respective and the preceding residues to the respective strips in the HNH-, CNH-, HCH- and CCH-NOESY spectra. Though the assignment had to be revised, the process was thus considerably sped up and the proportionally smaller number of NOE's bearing nearly all the information content about the structure were distinguishable at once.

4.7 The Relative Information Content of Distance Restraints

The information content of NOE's resulting in intraresidual and sequential restraints is largely exhausted after being used for the sidechain rotamer assignment, if the latter was possible. Thus, with the information being contained in sidechain torsion angles, Φ and ψ torsion angles and medium and long range restraints, those restraints are predominantly redundant. Φ and ψ torsion angles are usually derived from secondary structure predictions and verified while defining the topology. Intraresidual and sequential upper and lower distance restraints, whose information is not obviously contained in torsion angles and medium and long range restraints, should be applied if the respective crosspeaks are significantly weak or strong.

It has been previously shown that the relative information content of a single restraint or a set of restraints can be calculated as the difference in uncertainty

of the system before and after adding the experimental restraint(s) (Nabuurs et al. 2003). Figure 4.6 shows the relative number of restraints within the four classes of intraresidual, sequential, medium range and long range restraints and their relative information content received for the immunoglobulin binding domain of protein G (GB1) using the program QUEEN. The relative information content of restraints is strongly dependent on the information that was previously added (context dependency); e.g. the information contained in medium range restraints is strongly reduced if long-range restraints are already incorporated in the calculation.

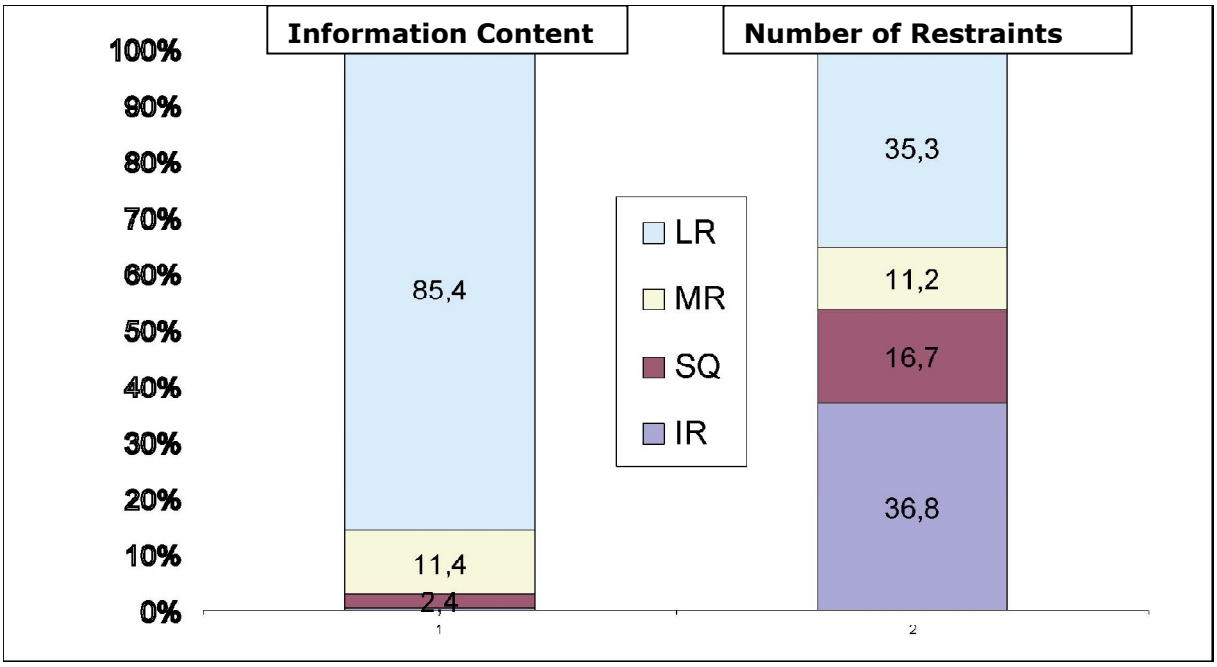


Figure 4.6. Relative number of restraints within the four classes of intraresidual, sequential, medium range and long range restraints and their relative information content received for the immunoglobulin binding domain of protein G (GB1) using the program QUEENS (figure taken from Nabuurs et al. 2003).

4.8 Validation of NMR Structures

With the increasing number of structures published in the RCSB Protein Databank (PDB), the number of inaccurate reported structures increases as well, and it is difficult to tell if there is a considerable proportional decrease. Apart from the

possibility that the validation parameters might be manipulated in single cases due to the pressure of publication, it could happen that a structure was made in all conscience and is still erroneous. This is due to the fact that for proteins the number of unambiguous NOEs is quite limited. "Unambiguousness" normally relies on several previously made assumptions, for example the assumption that the topology model and the resonance assignment are correct, and that previously assigned NOEs have been assigned correctly. Another problem arises if the NOE assignment is based on the assumption of various possible arrangements of structural elements with a further possibility that one might not have taken into consideration. The C-domain of Ph1500 serves as a good example; the possibility of an intermolecular character of the sheet consisting of strands β1 and β8 (Figure 4.7) might have not been conceived. Thus the question arises how a preferably extensive and meaningful structure validation can be achieved.

Typically natively folded proteins exhibit only a slightly different energy content compared to their unfolded state. Native folds are highly optimised structural assemblies and their stability is dependent on this energy content with respect to that of their unfolded state. Thus most validation methods are based on analysing the energy content of the protein, with the conformations and structural assembly exhibiting the lower energy being more likely. Those which substantially increase the energy content of the protein are unlikely present in the native fold. However, direct measurement of this energy is difficult, thus, these validation tools are almost all based on empirical data by comparison to a database of structures that are assumed to be correct.

One of these methods is the investigation of Φ and Ψ angle combinations. These were grouped into most favoured, additional allowed, generously allowed and disallowed regions by Ramachandran et al. (1963) based on computer simulations with the objective of finding stable conformations. Φ and Ψ angle combinations which cause atoms to collide correspond to sterically disallowed conformations.

In the same way sidechain rotameric states and the combination of X_1/X_2 angles, peptide bond planarity, bond lengths, bond angles, side chain planarity, unsatisfied hydrogen bond donors and acceptors, hydrogen bond geometry and

atom clashes can be analysed. There are various programs that perform this analysis, e.g. PROCHECK (Laskowski 1993), MOLPROBITY (Davis et al. 2007) and WHATIF (Vriend 1990). Moreover, the existence of an interior surface is unlikely, therefore the packing quality should be examined.
Further indication of the quality of a structure gives the NOE RMSD (averaged NOE violation) and the maximal restraint violation. Both should be possibly close to zero. By analysing individual restraints, it is important to note that consistently violated restraints are not necessarily the incorrect restraints. Often they are induced by nearby non-violated incorrect restraints. Thus if a consistently violated restraint is found, one should revise all restraints in that region of the structure.

The superimposition RMSD of the structural ensemble indicates the freedom of conformational motion limited through all experimentally derived restraints. A high RMSD reflects an imprecisely defined structure due to inadequate restraints or the presence of flexible regions. A low RMSD reflects a precisely but not necessarily correctly defined structure, since a structure based on erroneously assigned NOEs might also exhibit low RMSD values. Another problem is that the RMSD depends on the calculation protocols used, and these vary greatly.

The number of restraints defined per residue gives an idea of the precision of the structure. Though, it is less meaningful as long as double and redundant information and erroneous assigned NOEs are included. Redundant intraresidual restraints can be detected using the AQUA program (Laskowski et al. 1996). Moreover, as previously shown by the QUEEN method (Nabuurs et al. 2003), the relative information content of the respective classes of restraints, as well as of single restraints within these classes, is entirely different. Thus the number of restraints defined per residue is a less precise indicator for the quality of a structure.
A more precise indication can be obtained by performing a Completeness Check on the Aqua Server (Laskowski et al. 1996). This program subtracts all given restraints from all expected restraints for the given structure. Subsequently all missing expected restraints have to be revised.

Possibly, the strongest evidence for the precision and correctness of a structure might be given with a back-calculation of experimental spectra being in good agreement with the experimental ones. This approach is similar to the completeness method, but is more convenient and precise in its application, as increased or decreased intensities of all signals belonging to one strip can be easier detected. Back-calculation of the complete spectrum as well as of single strips, either to revise the full protein or only local regions, can be also used in a hypothesis-test approach. Thus for example rotamers and beta-turn types can be not only verified, but also defined.
Finally it has to be noted that it is less advantageous to perform the validation only at the end of the structure determination. Rather the structure validation should be integrated in an iterative structure determination process.

Returning to the example of the C-domain of Ph1500, one can imagine that the possibility of an intermolecular character of the sheet consisting of strands β1 and β8 (Figure 4.X right) would have not been taken into consideration. By reconsidering all stated validation tools, one can pose the question, which validation criteria would have been capable in detecting the error. Less likely are the RMSD for the structural ensemble, the number of restraints defined per residue, empirical statistics and validations considering the energy content unless the alternative structure resulted in steric clashes in the crossed loops connecting the sheets (Figure 4.7 left). More likely are the NOE RMSD, as strong restraint violations in the specific region would be expected, and a completeness check on the Aqua Server. The latter should have indicated that NOEs between atoms of the crossovers should be present, though not being present in the experimental spectra. A comparison of back-calculated and experimental spectra with a similar information content to the two previous stated validation criteria should also have been capable of detecting the error. A further indication is given by the solvent accessibility of amide protons located within the loops and by the packing quality.

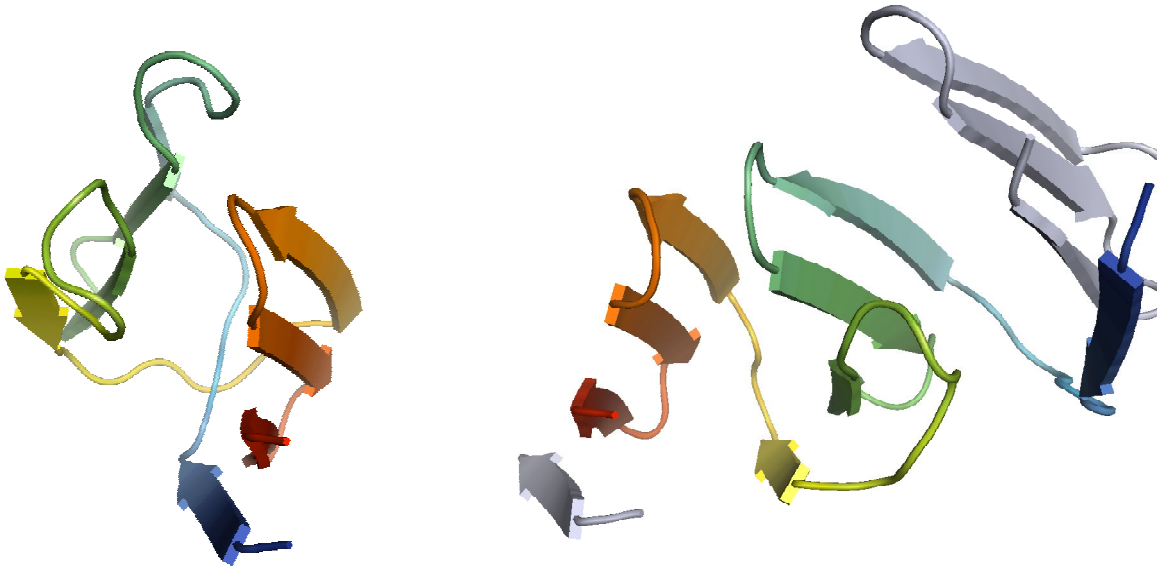


Figure 4.7. Structure of Ph1500-C with an intermolecular contact between strands β1 and β8 (right) and a model of Ph1500-C with an intramolecular contact between strands β1 and β8 (left) leading to a crossing of the two loops connecting the sheets.

Chapter 5

Structure Determination of the Monomeric N-domain of Ph1500

5.1 Protein Expression, Isotope Labelling and Purification

Protein expression, isotope labelling and purification of Ph1500-N were done by Sergej Djuranovic at the MPI Tübingen for Developmental Biology, Department of Protein Evolution.

Ph1500-N was expressed in *E. coli* C41 (DE3) at 37°C after induction with 1 mM IPTG at OD600~0.6. Minimal media supplement contained ^{15}N-labelled NH_4Cl as sole nitrogen source, and ^{13}C-labelled glucose as the sole carbon source. After purification, using a combination of Ni-NTA affinity and gel filtration chromatography, the protein was dialyzed against a 20 mM sodium phosphate buffer (pH 7.4) containing 250 mM NaCl.

5.2 NMR Methods and Experiments

All experiments were recorded at 300 K on Bruker DMX600 and DMX900 spectrometers on a 1.0 mM uniformly ^{13}C/^{15}N-labelled sample of Ph1500-N in 20 mM sodium phosphate buffer (pH 7.4) containing 250 mM NaCl. Backbone

sequential assignments were completed using an array of standard triple-resonance experiments (HNCO, HN(CA)CO, HNCA, HNCACB and CBCA(CO)NH) plus HNHA and HNHB experiments. Aliphatic sidechain assignments were achieved using a combination of HC(C)H-TOCSY, H(C)NH-TOCSY, (H)CCH-COSY and (H)CC(CO)NH-TOCSY experiments, while assignments of the aromatic residues were made by linking aromatic spin systems to the respective $C^{\beta}H_2$ protons in a 2D-NOESY spectrum.

Distance data were derived manually from a set of five 3D-NOESY spectra, including the heteronuclear edited NNH- (Zhang and Forman-Kay 1997), CCH- and CNH-NOESY spectra (Diercks *et al.* 1999) in addition to conventional ^{15}N- and ^{13}C-HSQC-NOESY spectra and a 2D-NOESY spectrum with a destructive filter of amid proton coherence. The mixing time chosen for all NOESY spectra was 80 msec.
All triple resonance experiments, as well as the HNH- and NNH-NOESY experiments, were implemented using selective proton flipback techniques for fast pulsing (Diercks et al. 2005). Processing of all NMR spectra was done using the program XWINNMR.

5.3 Resonance Assignment

To simplify the sequential assignment the automatic assignment program PASTA (Leutner et al. 1998) was used. The input was composed of amide proton and nitrogen chemical shifts picked from a ^{15}N-HSQC spectrum and C', Cα and Cβ chemical shifts of the respective (i) and the preceding residues (i-1) derived from an array of triple-resonance experiments (HNCO, HN(CA)CO, HNCA, HNCACB and CBCA(CO)NH).
Assignment of 97% of all backbone resonances was received with exception of the first (E1) and second last amino acid (H75). The fully assigned ^{15}N-HSQC spectrum recorded at 300 K and 600 MHz is shown in Figure 5.1.

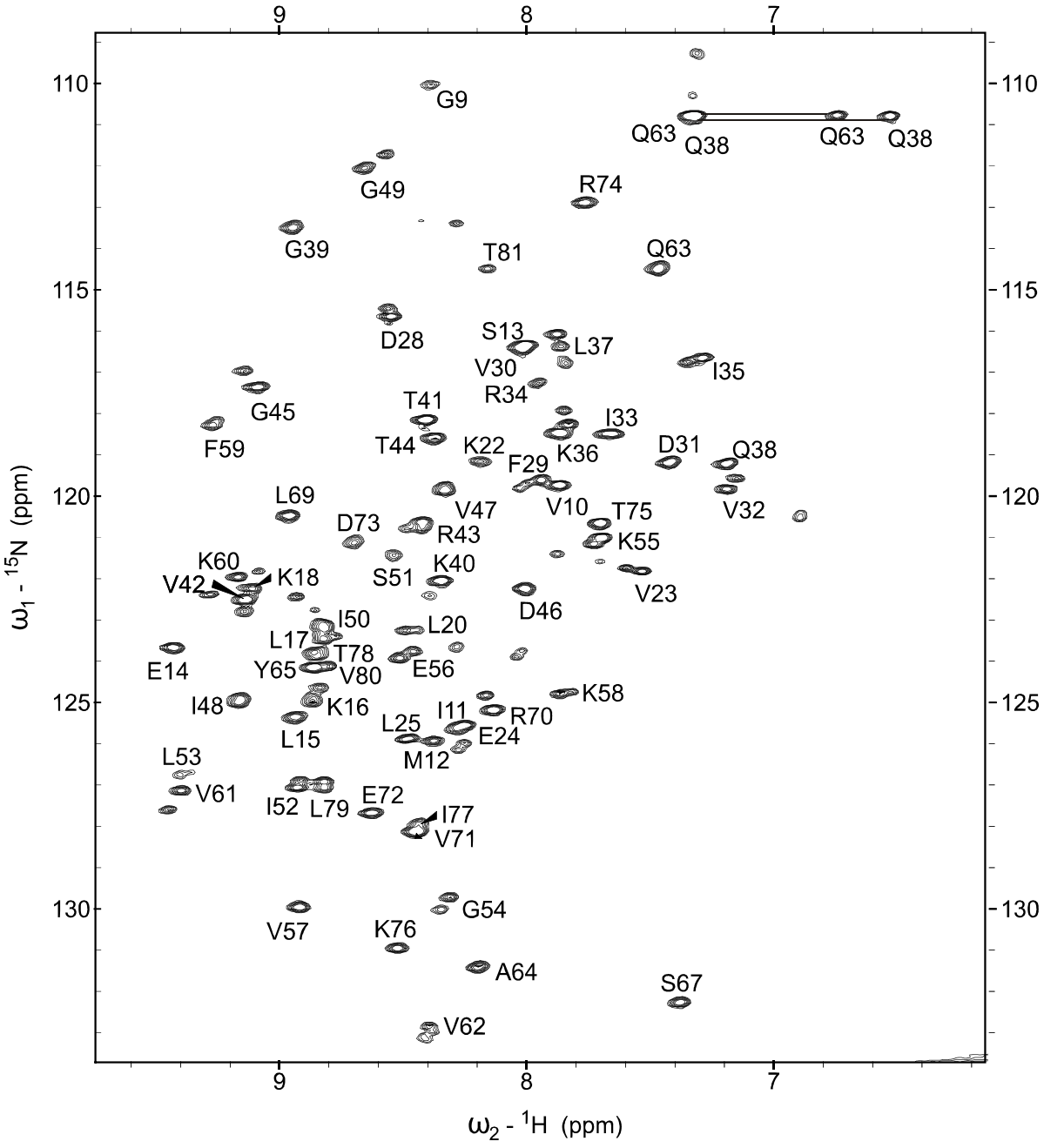


Figure 5.1. ^{15}N-HSQC of Ph1500-N recorded at 300 K and 600 MHz. Almost complete assignment of the backbone (97%) and sidechain resonances (97%) could be achieved. The only unassigned residues are the first (E1) and second last amino acid (H75). Unassigned resonances are belonging to a second conformation, to the His-tag or to impurities. Cross peaks connected by solid lines correspond to sidechain NH_2 groups of glutamine residues.

The assignment of sidechain resonances was completed to 97% (missing residues are E1 and H75) using a combination of H(C)CH-TOCSY, H(C)NH-TOCSY, (H)CCH-COSY and (H)CC(CO)NH-TOCSY experiments. The latter experiment contains theoretically all carbon resonances of the sidechain linked to the backbone amide proton and nitrogen resonances and thus strongly simplifies the assignment. Assignments of the aromatic residues were made by linking aromatic spin systems to the respective $C^{\beta}H_2$ protons in a 2D-NOESY spectrum.

5.4 Secondary Structure Prediction

Initially a homology search was made at the MPI for Developmental Biology in Tübingen and Ph1500-N was predicted to exhibit a β-clam fold. The high conformity of the secondary structure prediction with the experimental results, supplemental based on NOE patterns, is shown in Figure 5.2 together with the sequence of the homologous protein VatN-C and its experimentally found secondary structure (Coles et al. 1999). The secondary structure prediction included the prediction of the presence of a GD box motiv (see also Chapter 3.4). Moreover the assignment of the backbone and Cβ resonances allowed a secondary structure prediction with the program TALOS (Cornilescu et al. 1999) based on a fragment search using the HN, N, C', Cα and Cβ resonances (Figure 5.2) and with the CSI method (Wishart and Sykes 1992 and 1994), solely based on the chemical shifts (Figure 5.2 and 5.3). The different approaches to predict the secondary structure are described in Chapter 4.3.1.
The structure is mainly constituted of β-sheets with only one α-helix being present. For Ph1500-N the various secondary structure predictions all turned out to be reliable to a large extend.

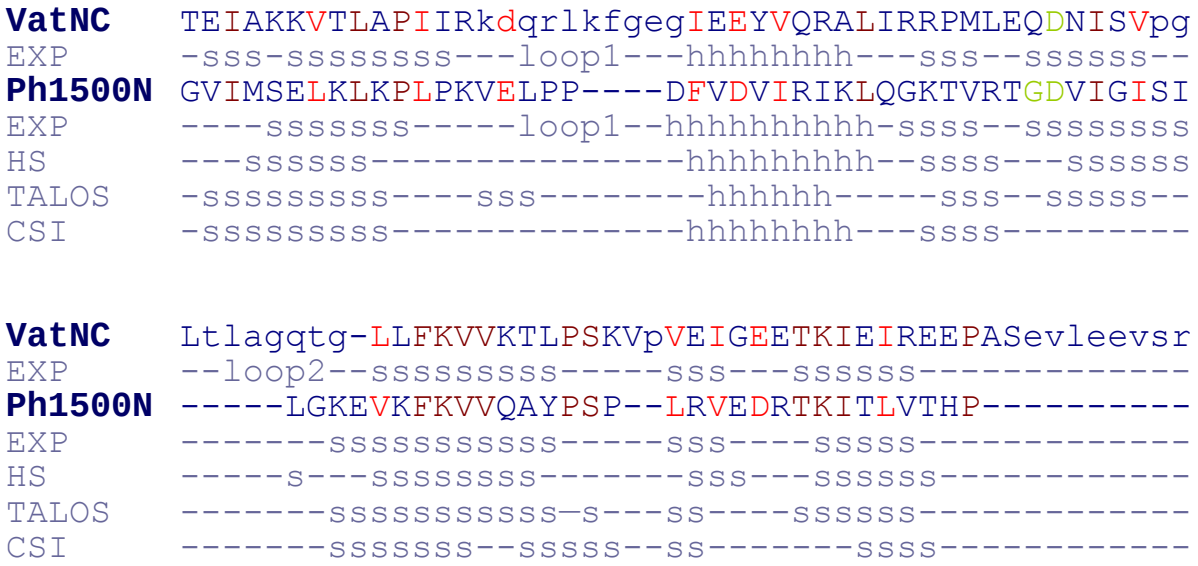


Figure 5.2. Comparison of the secondary structure prediction based on a homology search (HS), the prediction made with the program TALOS based on a fragment search using the HN, N, C', Cα and Cβ resonances, the CSI prediction based solely on the chemical shifts and the experimentally results (EXP) supplemental based on NOE patterns. The sequence of the homologous protein VatN-C is also shown together with its experimentally found secondary structure. Identical residues are marked in dark red and similar residues in red. The presence of a GD box motif (green) was assumed based on the homology search.

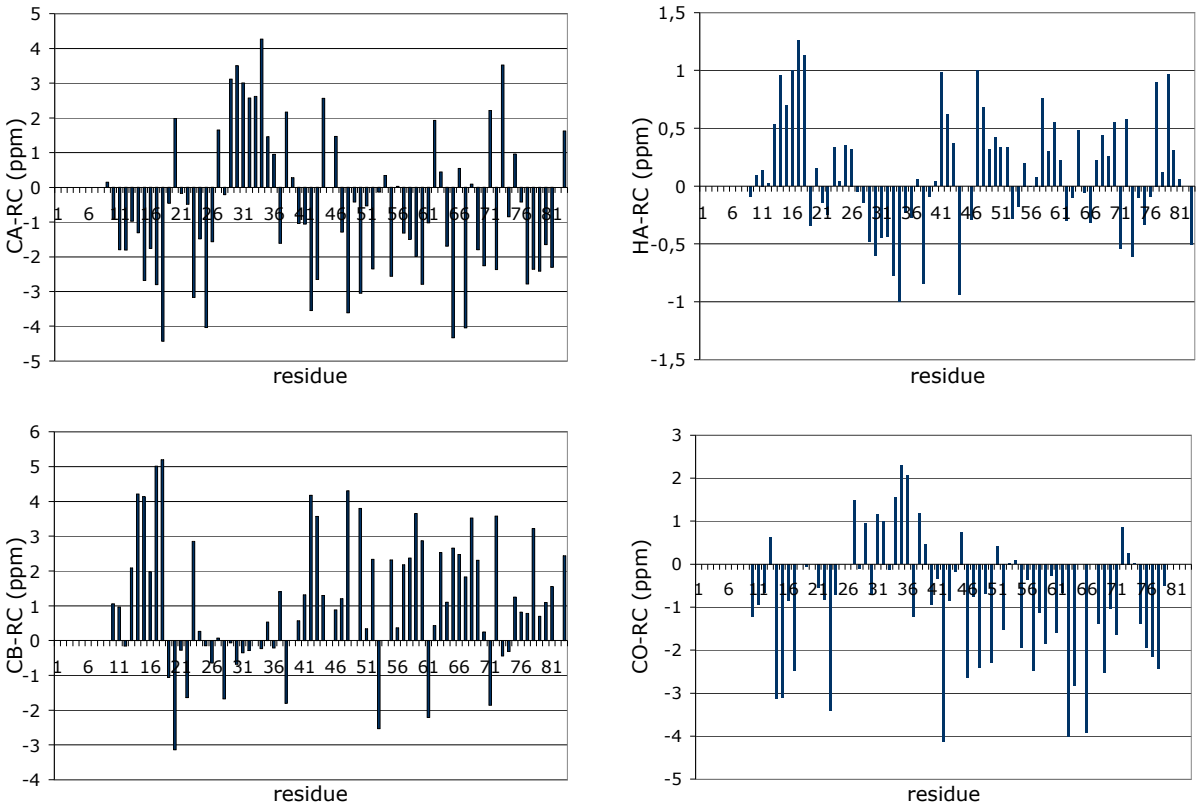


Figure 5.3. Plot of the secondary chemical shifts (deviations from the random coil values) for Cα, Hα, Cβ and C'. Cα and C' resonate downfield when located in an α-helix and upfield when located in a β-strand; Cβ and Hα behave contrarily.

5.5 Tertiary Structure

5.5.1 Topology Model

Advantageously, the first step in determining the tertiary structure is to define a topology model. Based on the secondary structure prediction, the H^N-H^N, H^N-H^α, H^N-C^α and H^α-H^α connectivity within β-sheets and α-helices, derived from HNH-, NNH-, CNH-, HCH- and CCH-NOESY spectra, was identified (Figure 5.4). Based on the distance restraints the topology can be additionally fixed by applying

distance and angle restraints for hydrogen bonds. Beside one α-helix (D28-L37) located between strand β1 and β2, the fold is constructed of two β-sheets, a short antiparallel β-sheet formed by strands β2 (G39-T41) and β5 (L69-V71) and a four stranded β-sheet, constituted of strands β6 (K76-L79), β1 (S13-K18), β4 (K55-A64) and β3 (G45-I52). Strand β1 and β6 are parallel, whereas β1, β4 and β3 are antiparallel. There is only one longer loop (P19-P27) between β1 and the α-helix.

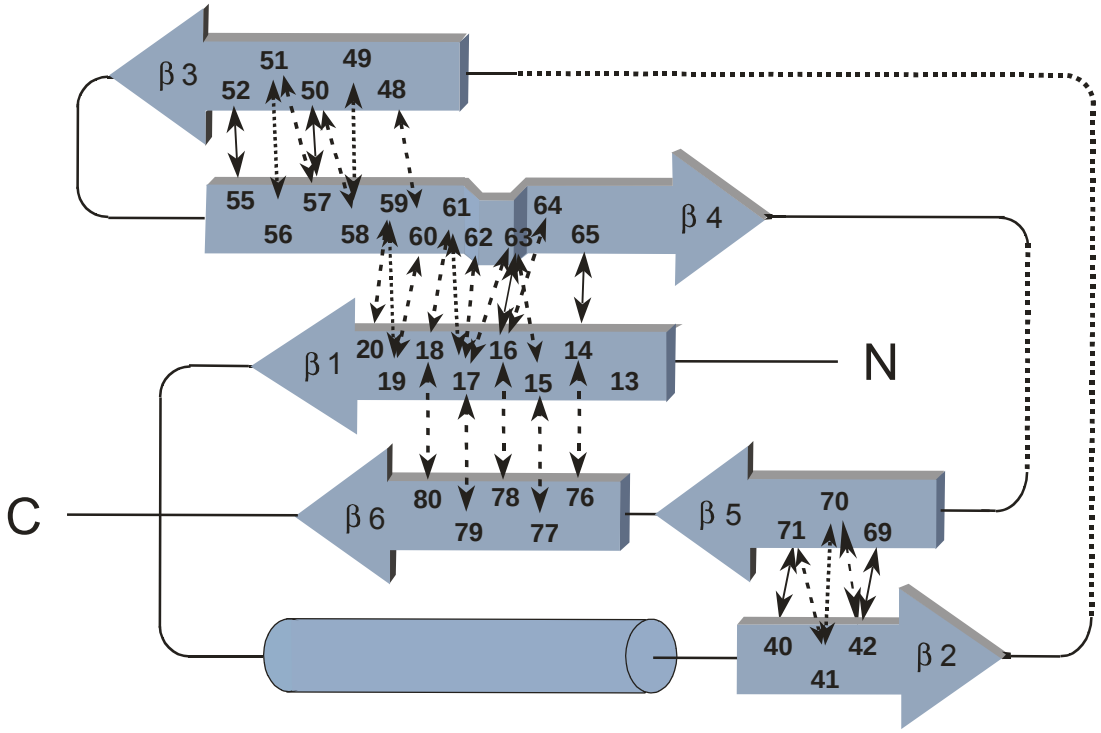


Figure 5.4. **Topology of Ph1500-N.** Arrows indicate the NOE H^N-H^N (drawn through), H^N-H^α (dashed) and H^α-H^α (dotted) connectivity in β-sheets, derived from HNH-, NNH-, HCH- and CCH-NOESY spectra, respectively. The fold is constructed of one α-helix and two β-sheets, a four stranded mixed parallel and antiparallel sheet and a short two stranded antiparallel sheet.

5.5.2 Water Exchange Measurement with a MEXICO Experiment

A MEXICO experiment (Chapter 4.3.3; Gemmecker et al. 1993) was measured on the $^{13}C/^{15}N$-labelled sample of Ph1500-N with five different mixing times (50, 100, 150, 200, 250 ms). The exchange rates of backbone amide protons with water were calculated by fitting the build-up of signal intensity with mixing time, and are plotted in Figure 5.5. Amide protons involved in hydrogen bonding are consistent with low exchange rates, thus reconfirming the topology model, whereas high exchange rates are reflecting solvent exposed amide protons.

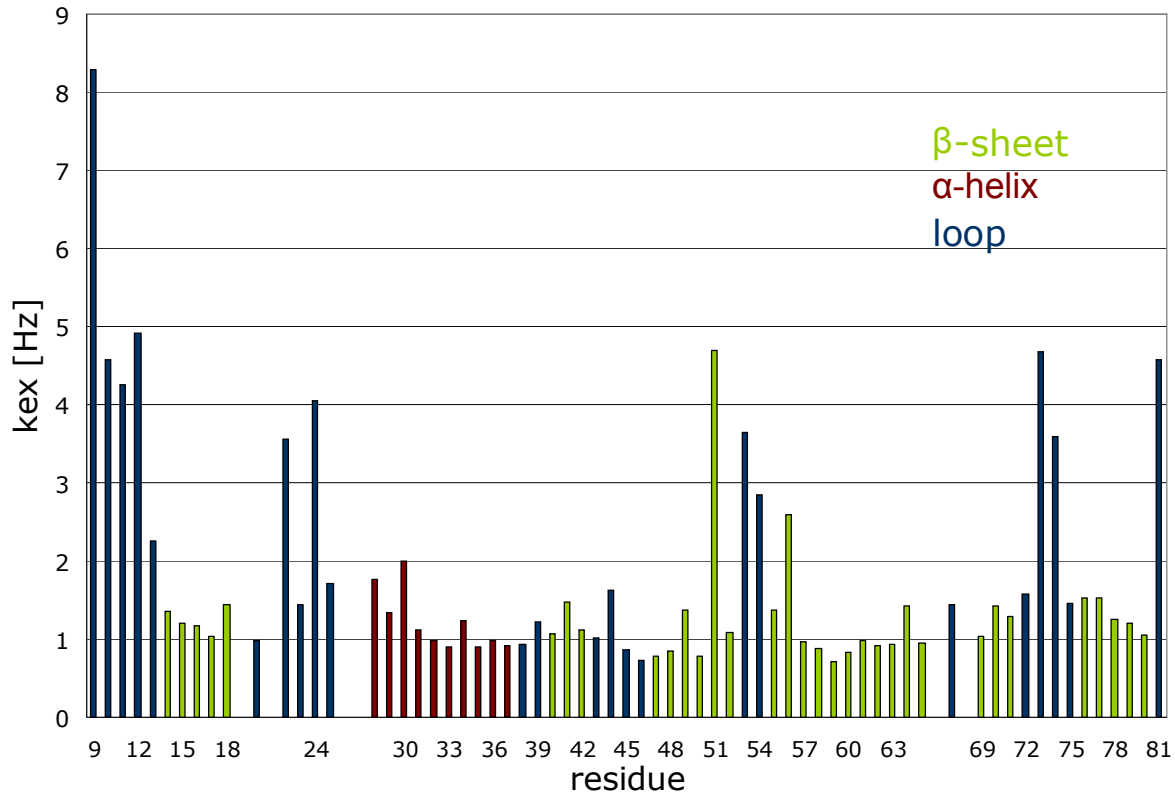


Figure 5.5. Exchange rates k_{ex} of backbone amide protons with water measured with a MEXICO experiment. Amide protons involved in hydrogen bonding are consistent with low exchange rates, whereas high exchange rates reflect solvent exposed amide protons. Residues located in a β-sheet, α-helix or loop are colored in green, red and blue, respectively.

5.5.3 Extent of Motion Measurement with a Heteronuclear-NH-NOESY Experiment

An estimation of backbone flexibility allowing identification of residues showing internal flexibility on fast (ps-ns) timescales was derived from a $^{15}N\{^{1}H\}$-heteronuclear NOE experiment (Chapter 4.3.4; Farrow et al. 1994) measured at 600 MHz with a proton presaturation delay of 3 s.

High Het-NOE values reflecting a low flexibility are observed for all amide groups located in secondary structure elements (Figure 5.6). Apart from the N-terminal residues, only one more flexible region located between residue 22 and 28 with slightly lower Het-NOE values can be identified. The results are in accordance with the results of the MEXICO experiment and the topology model.

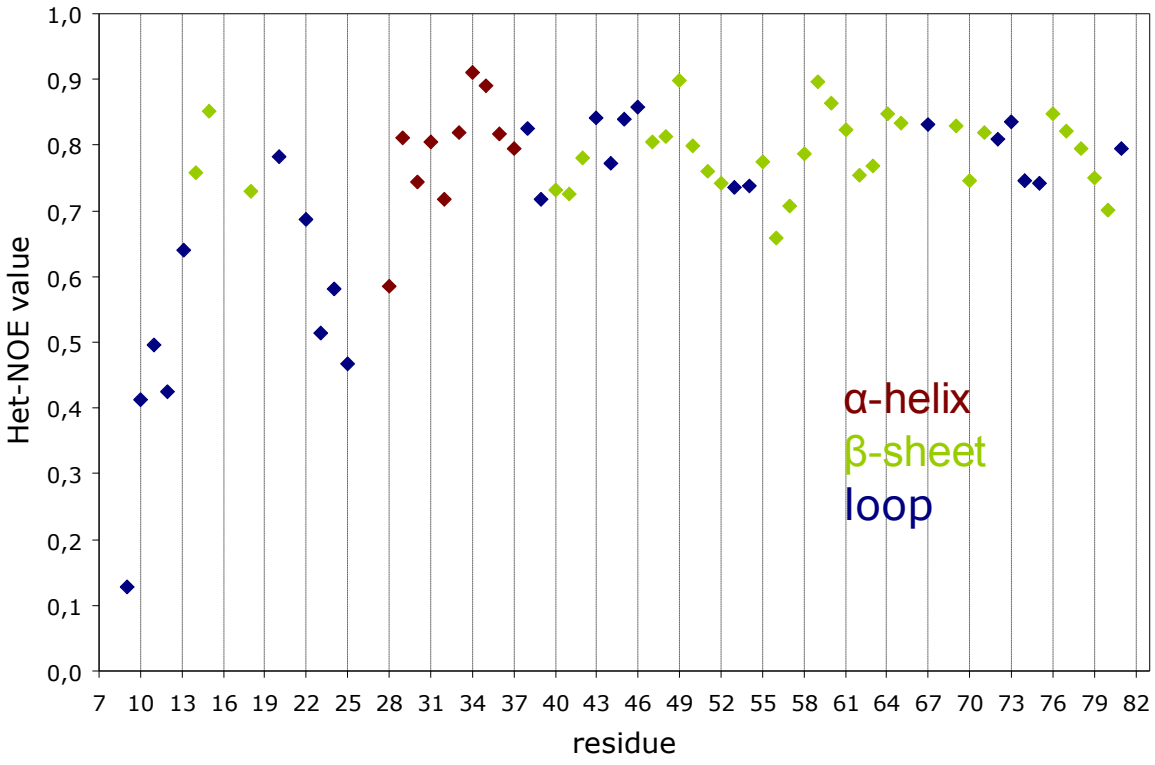


Figure 5.6. Heteronuclear-NOE values reflecting backbone flexibility on fast timescales. Values around 0.8 are observed for all residues located in secondary structure elements. One less ordered region was identified located between residue 22 and 28. Residues located in a β-sheet, α-helix or loop are colored in green, red and blue, respectively.

5.5.4 X_1 and X_2 Sidechain Rotamer Determination

Sidechain X_1 rotameric state determinations and stereospecific assignments were obtained for 26 of 45 prochiral $C^{\beta}H_2$ protons and for the $C^{\gamma}H_3$ groups of 8 of 11 valine residues. Determinations of X_1 rotamers were also available for all 7 isoleucine residues and all 5 threonine residues. Determinations of X_2 rotamers were made for 3 of 7 isoleucine and 4 of 8 leucine residues by consideration of patterns of intra-residue NOE connectivities derived from HNH- and HCH-NOESY spectra. Assignment of X_2 rotamers of leucine led to stereospecific assignment of the prochiral leucine $C^{\delta}H_3$ groups.

5.5.5 Structure Calculations

For the structure calculation performed with XPLOR (NIH version 2.9.3) the following input was used: 752 NOE distance restraints divided in 192 (25%) long range, 96 (13%) medium, 270 (36%) sequential and 194 (26%) intraresidue restraints and classified into four categories corresponding to upper interproton distance restraints of 2.7, 3.2, 4.0 and 5.0 Å, whereas lower distance restraints were included for very weak or absent intra-residue or sequential H^N-H^N or H^N-Hα crosspeaks using a minimum distance between 2.5 and 4.0 Å. In addition 43 Hydrogen bond restraints, derived from the topology model, Hetero-NH-NOESY and Mexico experiments, were applied via inclusion of pseudo-covalent bonds as described by Truffault et al. (2001). Moreover 51 Φ torsion angle restraints (derived from ^{3}J-HN(i)HA couplings in HNHA), 16 Ψ torsion angle restraints (derived from ^{3}J-N(i+1)HA couplings in HNHB), 46 Chi1 and 11 Chi2 rotamer restraints (received from HNH- and HCH-NOESY experiments) were included. Two prolines were defined as cis and five as trans isomers. Dihedral restraints for sidechain rotamers were applied with a tolerance of 30°, with the exception of proline residues where the X_1 rotamer was restrained to plus or minus 30° with a tolerance of 15°. Additionally, dihedral angle restraints were received for backbone Φ and Ψ angles based on C^{α}, C^{β}, C′ and H^{α} chemical shifts using the program TALOS (Cornilescu et al. 1999).

Structures calculated in an initial simulated annealing protocol were refined in two further slow cooling stages, the first including a conformational database potential and the second with the force constant on peptide bond planarity relaxed to 50 kcal/mol/rad^2. A final set of 22 structures was selected out of 50 calculated structures.

5.5.6 Structure Validation

A structure validation was made under the considerations described in chapter 4.5. The calculated structural ensemble shows a superimposition RMSD for the backbone and all residues of 0.276 Å (Figure 5.7) and for non-H atoms of 0.757 Å.

On the average 10 restraints were defined per residue with an average restraint violation (NOE RMSD) of 0.008 Å and no violations bigger than 0.11 Å. A completeness check of the NOE distance restraints was performed on the Aqua Server (Laskowski et al. 1996) showing an overall completeness of 63% for all amino acids and a shell of 2-3.5 Å and no disallowed restraints. The incompleteness represents crosspeaks expected from the structure that were not present in the restraint list. These missing restraints were revised, they were either due to assigned NOEs that were not used because they have no true information content, due to flexibility or due to overlap with water or other protein signals.

Backbone Φ and Ψ angle combinations were calculated with the program PROCHECK (Laskowski et al. 1993) for the regularised average structure and showed 94% of all residues to be in most favoured regions of the Ramachandran Plot (Ramachandran et al. 1963) and 6% in additionally allowed regions (Figure 5.8). Further structure validations were received using the programs Molprobity (Davis et al. 2007) and WHATIF (Vriend 1990) showing overall acceptable results.

A back-calculation of 3D HNH- and CNH-NOESY spectra was made. Comparison with the experimental ones showed that they were largely consistent with each other. If differences in intensities were significant, the distance restraints belonging to that residue were revised. If necessary also lower distance restraints were set.

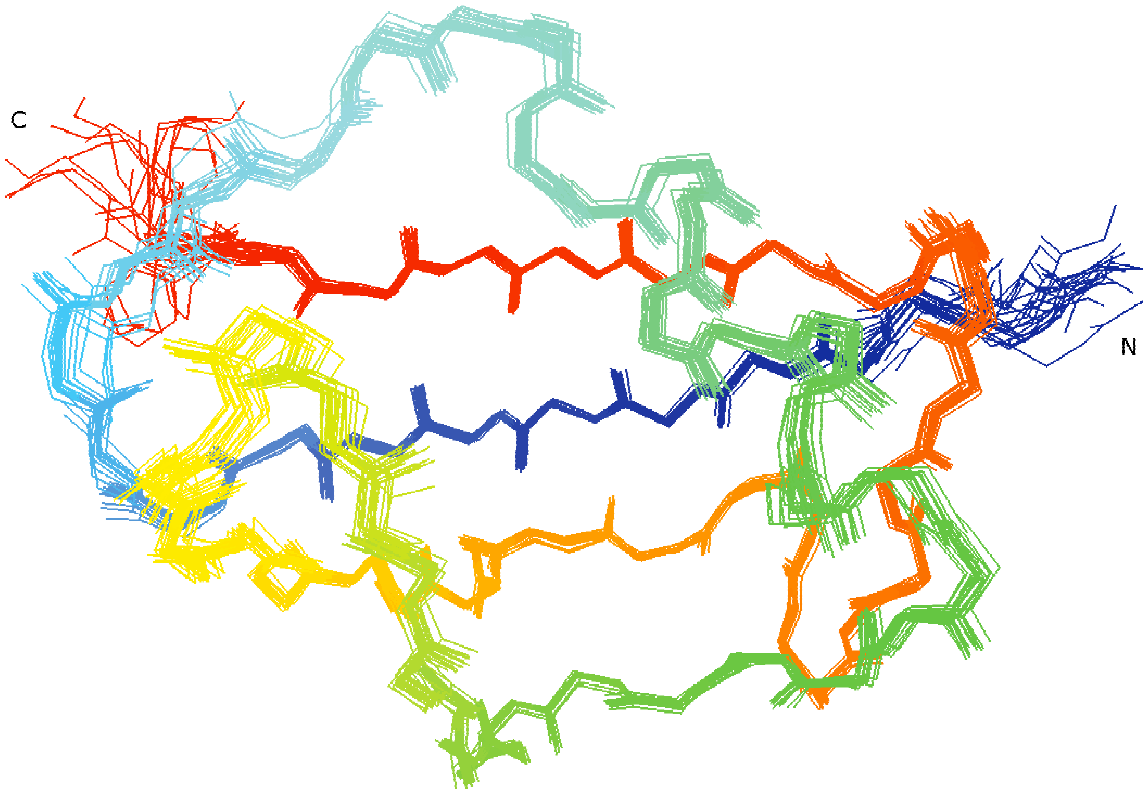
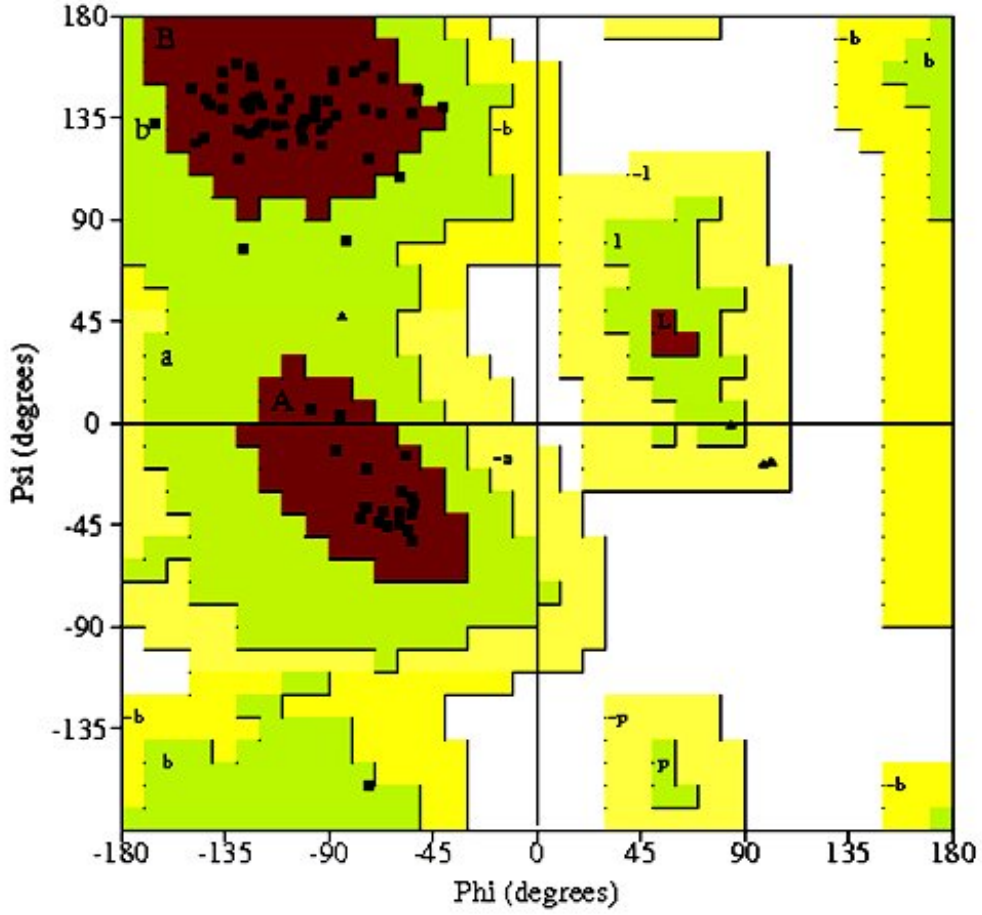


Figure 5.7. The superimposition RMSD obtained for the backbone is 0.276 Å. Less precise defined regions, due to a lack of experimental restraints predominantly caused by flexibility (see also Figure 5.6), are the N- and C-termini, the loop between strand β1 and the α-helix (light blue) and the β-turn (yellow) between strand β3 and β4.



Plot statistics

Most favoured regions (red)	59	93.7%
Additional allowed regions (green)	4	6.3%
Generously allowed regions (light yellow)	0	0%
Disallowed regions (white)	0	0%
N° of non-Gly and non-Pro residues	63	100%
N° of end-residues (excl. Gly and Pro)	1	
N° of Gly residues (shown as triangles)	5	
N° of Pro residues	7	
Total N° of residues	76	

Figure 5.8. Ramachandran Plot statistics for the regularised average structure of Ph1500-N calculated with PROCHECK (Laskowski et al. 1993). For glycine residues shown as triangles this classification is not representative. Region B corresponds to β-sheets, region A to right-handed α-helices and region L to left-handed α-helices or to a positive Phi region, respectively.

5.6 Description and Discussion of the Structure

The three-dimensional solution structure of Ph1500-N was determined manually as described in the previous section, based on distance, dihedral angle and rotamer restraints and can be described as a β-clam (Fig. 5.9).
The β-clam fold is similar to a β-barrel fold, with the main difference that a β-barrel fold would require hydrogen-bonded contacts between strands β2 and β3, while the present fold shows only sidechain contacts between these strands, thus opening the barrel. The hence exposed hydrophobic core is covered by an α-helix (D28-L37), lying diagonally across the four stranded sheet. This strongly twisted sheet is centred on the β1-strand (S13-K18), parallel to β6 (K76-L79) and antiparallel to β4 (K55-A64), with β4 being also antiparallel to β3 (G45-I52) and exhibiting a nearly orthogonal turn. Another small β-sheet formed by strands β2 (G39-T41) and β5 (L69-V71), consisting of three residues each, lies orthogonally upon the edge of the four stranded sheet. The fold contains only one longer less structured loop (P19-P27) between β1 and the α-helix.

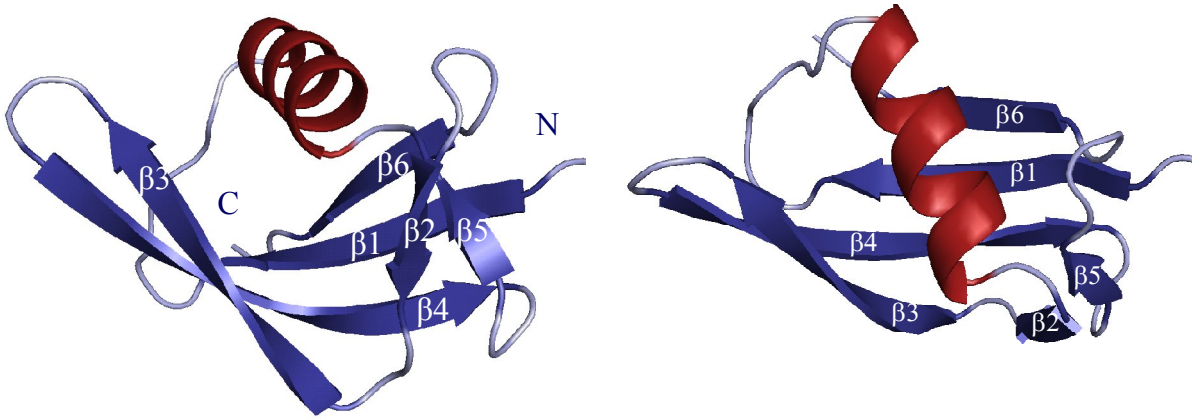


Figure 5.9. Solution structure of the monomeric N-domain of Ph1500. As initially predicted from sequence homology, Ph1500-N shows a β-clam fold.

As initially predicted, the C-domain of Vat-N and the homologous N-domain of Ph1500 show nearly identical folds, with two main differences. While VatN-C shows a parallel contact between the first third of β1 and β5, this contact is

absent in Ph1500-N. Moreover the strands β3 and β4 are longer in Ph1500-N and connected by a β-turn, in VatN-C there is a long loop connecting these strands. A comparison of the topology and structure of both protein domains is shown in Figure 5.10 and 5.11, respectively. A sequence comparison was previously shown in Chapter 5.4. Another homologous protein domain, the N-terminal domain of AfAMA is also shown in Firgure 5.11. The main difference is the relatively longer loop region between the first β-strand and α-helix containing the conserved GYPL motif, which mediates the hexamerisation in AfAMA and is characteristic for this group of proteins (see also chapter 3.5).

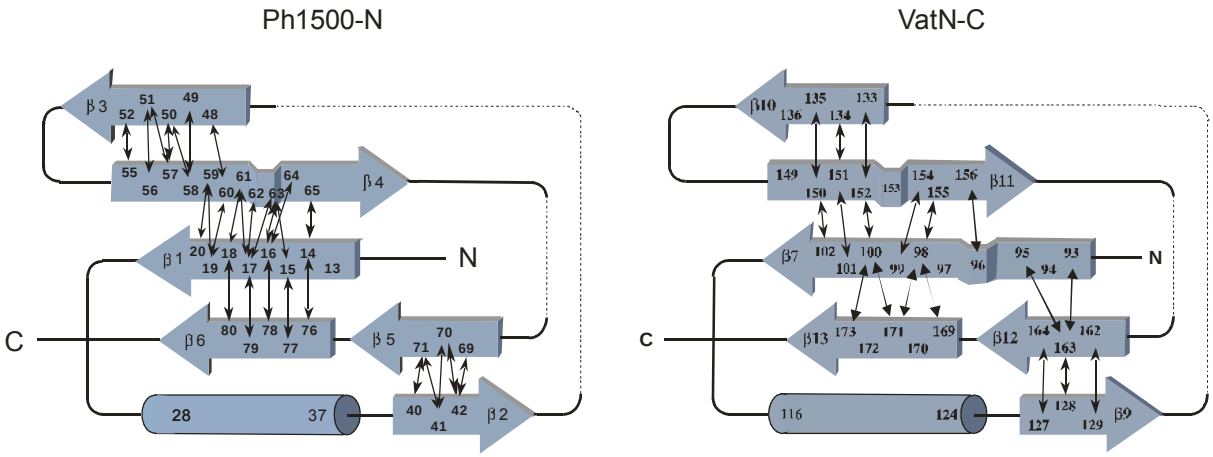


Figure 5.10. Comparison of the topology of Ph1500-N and VatN-C. While VatN-C shows a parallel contact between the first third of β1 and β5, this contact is absent in Ph1500-N. Both domains differ also in the length of the strands or loops, respectively.

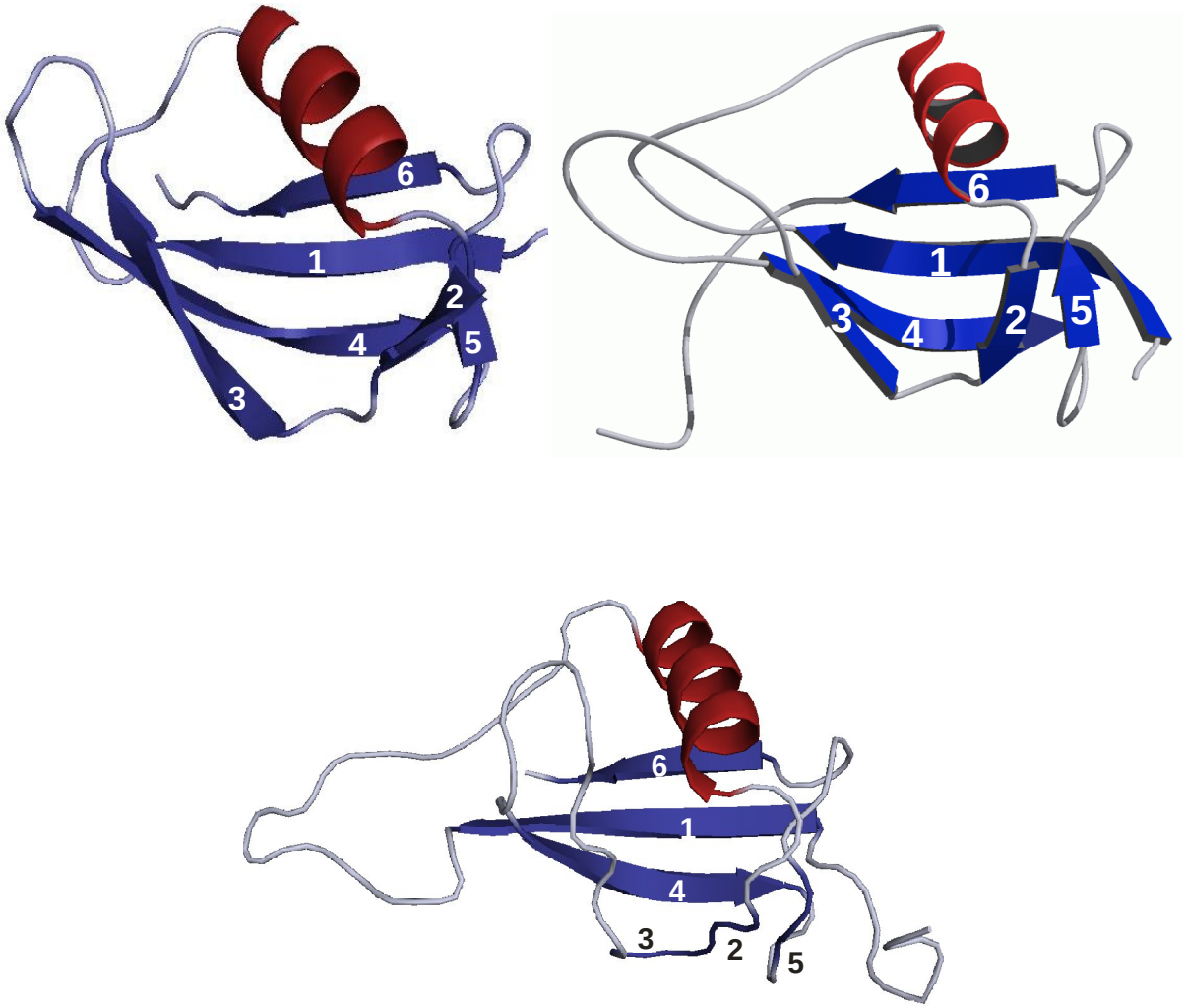


Figure 5.11. Comparison of the structure of the homologous protein domains Ph1500-N (top left), VatN-C (top right) and AfAMA-N (bottom). Visible are primarily the different extended loop regions and the backbone contacts between strand β1 and β5, that are missing in Ph1500-N.

Concerning the reconstruction of evolutionary events connecting the β-clam fold of Ph1500-N with other folds within the cradle-loop metafold family (Chapter 3.4), different scenarios are possible. However an evolutionary relationship that was previously proposed (Coles et al. 1999) still remains unclear. One scenario is that this fold evolved from fragments of an ancestral RIFT barrel by fusion and insertion of the β2-strand. This hypothesis is motivated

by sequence homology between Ph1500-N and a fold (PDB ID 1XE1) that is similar to a RIFT barrel, solely varying from the latter in the replacement of an α–helix by a β-strand. However, this hypothesis requires further verification.
Another β-clam (PDB ID 2JOV) which is clearly related to Ph1500-N, varies from the latter in that the first β-strand shows an antiparallel contact to the strand prior to the helix and a parallel contact to the strand subsequent to the helix. Both contacts are reversed in Ph1500-N. Moreover, this β-strand, located at the C-terminus in Ph1500-N, is circular permutated to the N-terminus. Thus it remains uncertain, if one fold arose from the other by circular permutation forcing the strand from antiparallel to parallel or vice versa, or if both folds have a common ancestor.

5.7 Data Deposition

The coordinates for a structure ensemble and a regularised average structure of Ph1500-N have been deposited in the Protein Data Bank (PDB ID: 2jv2).

Chapter 6

Structure Determination of the Hexameric C-domain of Ph1500

6.1 Protein Expression, Isotope Labelling and Purification

Protein expression, isotope labelling and purification of Ph1500-C were done by Sergej Djuranovic, Astrid Ursinus and Jörg Martin at the MPI Tübingen for Developmental Biology, Department of Protein Evolution.
Ph1500-C was expressed in *E. coli* C41 (DE3) at 37°C after induction with 1 mM IPTG at OD600~0.6. Minimal media supplement contained ^{15}N-labelled NH_4Cl as sole nitrogen source, and ^{13}C-labelled glucose as the sole carbon source, while ^{2}D/^{13}C-labelled glucose and D_2O were used as deuterium sources. After purification, using a combination of Ni-NTA affinity and gel filtration chromatography, the protein was dialyzed against a 20 mM sodium phosphate buffer (pH 7.4) containing 250 mM NaCl.

6.2 NMR Methods and Experiments

All experiments were recorded at temperatures between 318 and 353 K on a Bruker DMX600 and a Bruker DMX750 spectrometer with a TXI probehead equipped with triple-axis gradients. Due to technical limitations, transverse axis

gradients were not used at temperatures over 330 K. Several samples were prepared, all in 20 mM sodium phosphate buffer (pH 7.4) containing 250 mM NaCl. Firstly, a 0.6 mM uniformly $^{13}C/^{15}N$-labelled sample, a 0.6 mM uniformly $^{2}H,^{13}C/^{15}N$-labelled sample and a one with selective triple labelling of all amino acids except isoleucine, leucine and valine were prepared on a T101P mutant of Ph1500-C. This mutation was inadvertent and first traced during sequential assignment. Later, the concentration of the double-labelled mutant sample was doubled to 1.2 mM by centrifugal concentration after adding 50 mM arginine/glutamate (1:1) (Golovanov et al. 2004). Due to the fact that this concentrated sample was only stable for about 2 weeks at 80°C, and all experiments could not be completed during that time, a new sample of the wild type protein was prepared. This sample proved much more soluble and a concentration of 1.2 mM was reached without arginine/glutamate supplement.

The $^{2}H,^{13}C/^{15}N$-labelled mutant sample showed aggregation at temperatures around 323 K. In an attempt to increase the thermostability of the protein, different conditions of pH and salt concentration were tested. However, it was noticed that the $^{13}C/^{15}N$-labelled sample (both mutant and wild type) proved highly thermostable. Different thermostabilities of the double and triple labelled samples may eventually be caused by different purification levels rather than due to the effect of deuteration.

To determine the practically thermostability, a 40 µl volume of the protein solution proved sufficient. First, a 20 µl aliquot was heated slowly in a clear 200 µl Eppendorf-cap, with denaturation detected by visible aggregation (~95°C, 268 K). Another 20 µl aliquot was heated for one week at 85°C. This was then diluted and intact folding checked with a 1D NMR spectrum.

After recording experiments based on amide protons, water in the mutant probe was stepwise exchanged against D_2O to eliminate the signals arising from water and to reduce relaxation of $H\alpha$ protons. The effect was only slightly positive due to a noticeable loss of concentration during the exchange process.

Backbone sequential assignments were completed to 99% using an array of standard triple-resonance experiments (trHNCO, trHN(CA)CO, trHNCA and trHNCACB) recorded on the 85% deuterated mutant sample at 318 K. Aliphatic

sidechain assignments were completed to 97% using a combination of H(C)CH-TOCSY, (H)CCH-TOCSY and (H)CCH-COSY experiments measured at 353 K on the double labelled 1.2 mM sample of wild type Ph1500C. Assignments of the aromatic residues were made by linking the aromatic spin systems to the respective $C^{\beta}H_2$ protons in a 2D-NOESY spectrum, supported by a ^{13}C-HSQC of the aromatic region.

All distance restraints were manually assigned using a combination of the conventional HNH- and HCH-NOESY spectra, the heteronuclear edited trNNH-, CCH- and CNH-NOESY spectra (Diercks *et al.* 1999) as well as a 2D-NOESY spectrum with destructive filtering of amide protons, thus disclosing well resolved traces of the three aromatic residues. In addition a folded HCH-NOESY spectrum was recorded of the methyl region from 13 to 31 ppm in the indirect carbon dimension, which was essential to identify intra- and especially intermolecular methyl-methyl contacts. Reducing the spectral width to 29%, the resolution could be dramatically increased. All spectra were recorded at 353 K with exception of the trNNH-NOESY previously measured at 318 K, as well as the HNH-NOESY measured at 318 K as a TROSY version and at 353 K as a standard version. The optimal mixing time chosen for all NOESY spectra was 80 msec.
For Ph1500-C the structure determination was done manually with the exception of one script (written by Dr. M. Coles) used to transfer the assigned resonances of the respective and the preceding residues to the respective strips in the HNH-, CNH-, HCH- and CCH-NOESY spectra. Though the assignment had to be revised, the time needed for peak picking and labelling was considerably shortened.
The processing was either done with XWINNMR or TOPSPIN.

6.3 Resonance Assignment

Backbone and sidechain assignment of Ph1500-C was first performed on the T101P mutant. Later, the chemical shifts were reassigned on the wild type sample by comparing triple resonance and HNH-NOESY slices. A comparison of the amide proton and nitrogen chemical shifts is shown in the overlayed TROSY spectra of both samples (Figure 6.1).

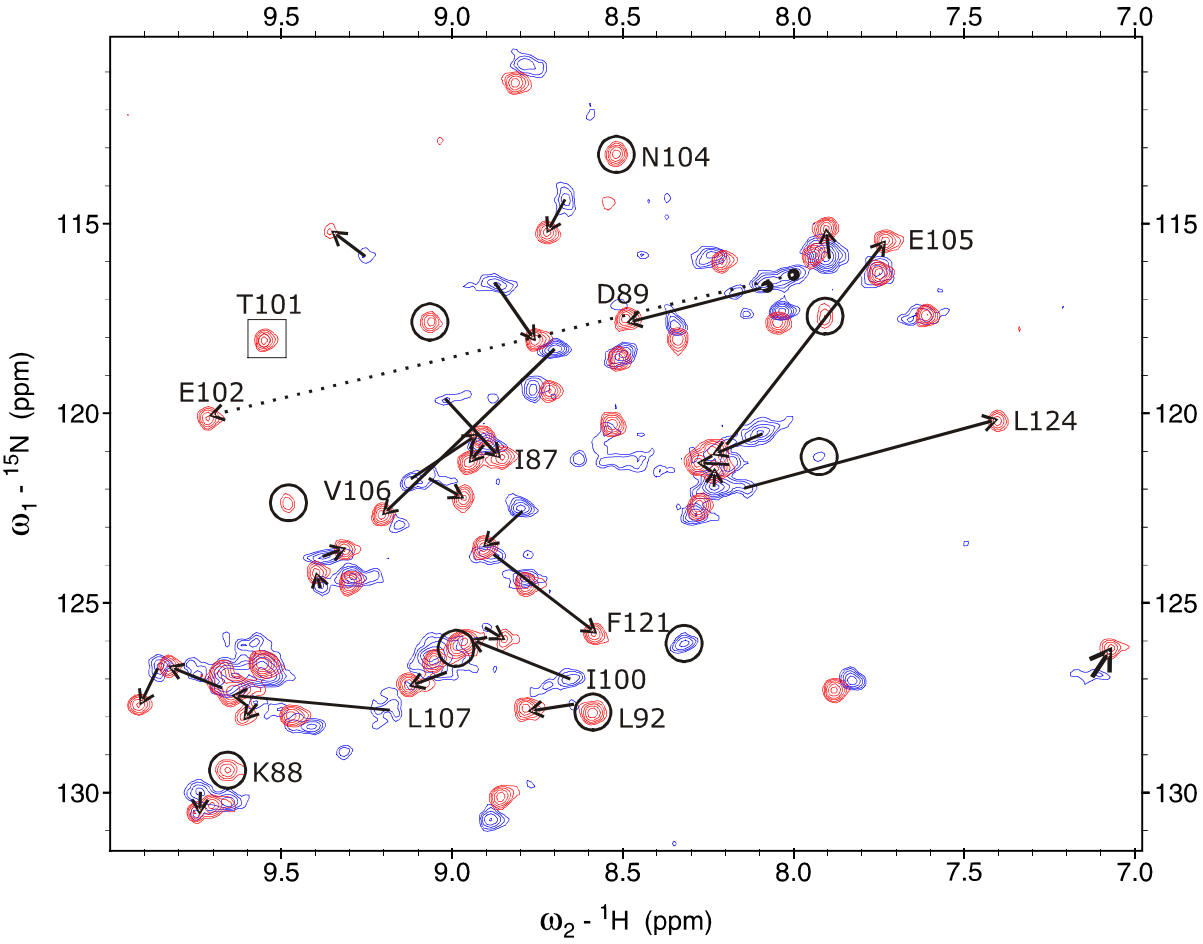


Figure 6.1. Overlayed TROSY spectra measured on the wild type sample and on the T101P mutant sample of Ph1500-C. The changes of chemical shifts are traced with an arrow and the mutated amino acid T101P is marked with a square. Residues with strong chemical shift changes are labelled. The strongest changes are observed in the neighbourhood of the mutation in sequence and space. Circles mark residues that either appeared or disappeared.

For Ph1500-C, the strategy has been to optimize the measurement temperatures with respect to the different experiments and to re-assign chemical shifts at the higher temperature later on. Sidechain assignment experiments and ^{13}C-HSQC-based NOESY experiments were carried out at the highest temperature tolerated by the hardware (353 K), assuming that the highest temperature would be required for sufficient resolution, as the protein has a high content (42%) of isoleucine (13%), leucine (16%) and valine (14%) residues, whose methyl–

methyl contacts were essential for the structure determination, especially for defining the intermolecular contact.

For backbone experiments with amide protons involved in the magnetization transfer pathway, temperature selection is less straightforward due to increased water exchange of amide protons at higher temperature. Accordingly, dependent on which effect is dominant, respective amide signals show either an increased or decreased intensity. In 3D backbone experiments measured at 318 K, three amide proton signals were missing or very weak and overlapped (K83, V135 and I136). Although V135 and I136 re-appeared at 353 K, seven additional signals (K85, N95, E103, K120, E122, N123 and N131) had disappeared and further signals showed strongly decreased intensities. Both residues that appeared at the higher temperature are located in a β-sheet, whereas all residues that disappeared are located in either loops or β-turns.

A comparison of TROSY and ^{15}N-HSQC spectra recorded at temperatures between 293 K and 353 K showed that the optimal temperature would have been about 328 K. However, since the triple labelled sample proved less thermostable, the temperature chosen for the backbone experiments measured on the triple labelled sample was 318 K, while that for the sidechain and NOESY experiments measured on the double labelled sample was 353 K.

Since backbone- and sidechain-experiments were recorded at different temperatures, the backbone resonances had to be re-assigned at the higher temperature. As the ^{13}C chemical shifts are less temperature dependent than amide protons, it was sufficient to use trHNCA, trHNCO and HN(CA)HA spectra recorded at the higher temperature, together with a series of TROSY spectra recorded at intermediate temperatures (323, 333 and 343 K) to trace chemical shift changes.

For the backbone assignment, an 85% deuterated sample and one with selective triple labelling of all amino acids, except isoleucine, leucine and valine, were prepared. The latter sample provides additional information, in that missing traces in the trHNCACB experiment can be assigned to isoleucine, leucine or valine residues and missing traces in the trHNCO experiment to amino acids with an isoleucine, leucine or valine in the preceding residue. Deuteration, together

with selective labelling of Ph1500-C, allowed an sequential assignment of about 20%. In an advanced state of the assignment process, after increasing the measurement temperature to 318K, it turned out that the latter sample would have been dispensable.

The sequential assignment was achieved using the automatic assignment program PASTA (Leutner et al. 1998). The input was composed of amide proton and nitrogen chemical shifts picked from a TROSY spectrum and C', Cα and Cβ chemical shifts of the respective (i) and the preceding residues (i-1) derived from an array of triple-resonance experiments (trHNCO, trHN(CA)CO, trHNCA and trHNCACB). Other experiments, such as the CBCA(CO)NH and HNHAHB, although potentially providing additional information, proved too insensitive for a protein of this size. Thus, only a trHNH-NOESY spectrum providing the information of possible preceding amide protons was additionally used.
Finally, increasing the temperature to 318 K allowed assignment of 96% of backbone resonances, with exception of K83, V135 and I136. The latter two residues were later assigned by increasing the temperature to 353 K, thus completing the backbone assignment to 99%. Figure 6.2 shows the fully assigned TROSY spectrum recorded at 353 K and 750 MHz.

Sidechain assignment was completed to 97% (missing residues are E53 and R61) using a combination of H(C)CH-TOCSY, (H)CCH-TOCSY and (H)CCH-COSY experiments measured at 353 K on the double labelled 1.2mM sample of wild type Ph1500C. Assignments of the aromatic residues were made by linking aromatic spin systems to the respective $C^{\beta}H_2$ protons in a 2D-NOESY spectrum and by additionally using a ^{13}C-HSQC of the aromatic region.

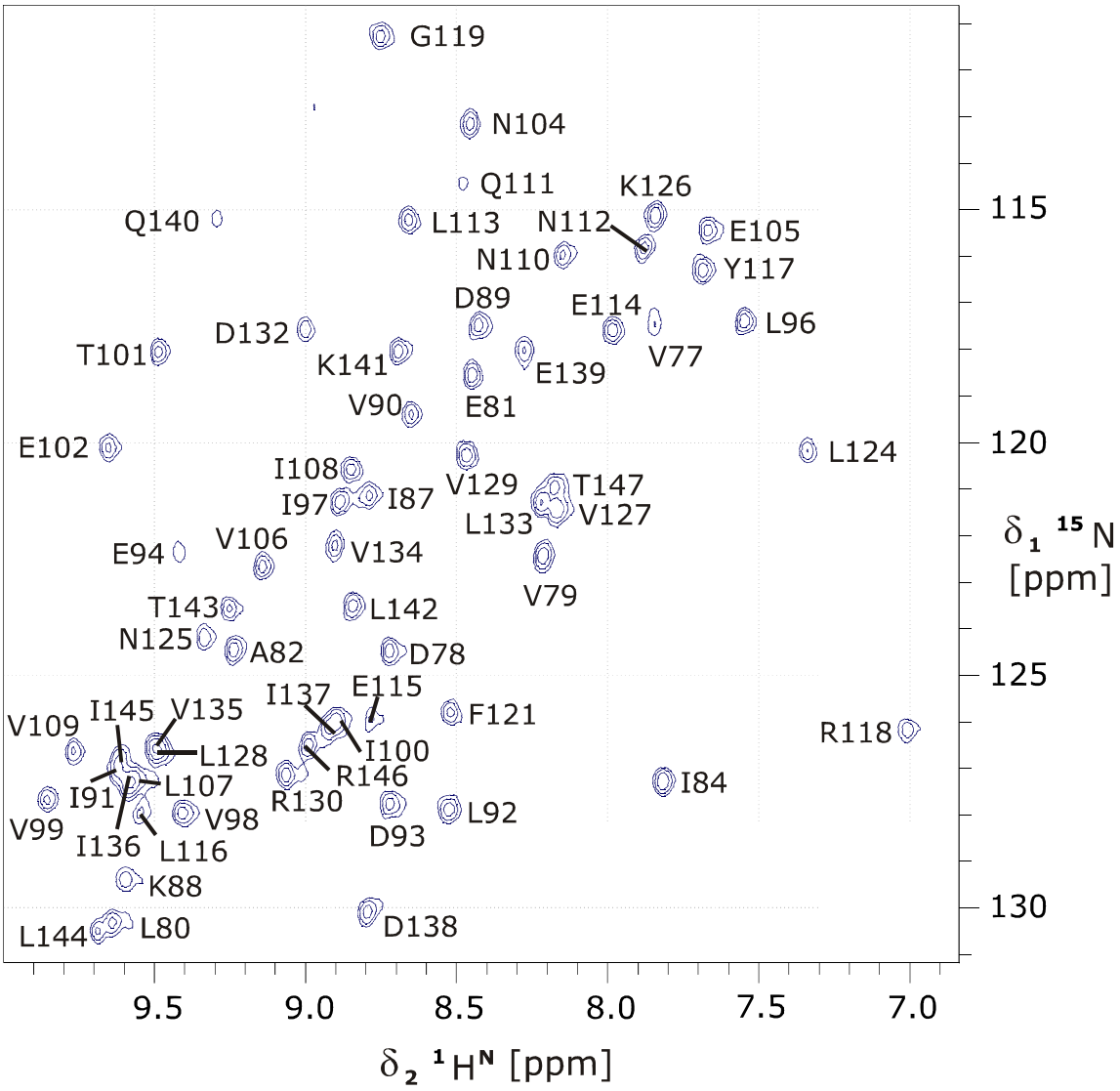


Figure 6.2. TROSY of Ph1500-C recorded at 353 K and 750 MHz. Almost complete assignment of backbone (99%) and sidechain resonances (97%) could be achieved. Various residues (K83, K85, N95, E103, K120, E122, N123 and N131) are missing in the TROSY spectrum recorded at 353 K, but were present in the TROSY spectrum recorded at 318 K, apart from K83.

6.4 Secondary Structure Prediction

Primarily a secondary structure prediction was made at the MPI for Developmental Biology in Tübingen using a powerful sequence search tool based on hidden Markov models, HHpred (Soding 2005) and only weak similarity was found, thus the results are not significant (Figure 6.3). With the assignment of the backbone and Cβ resonances an accurate prediction became possible.
For a careful referencing of the resonances, which are sensitive to pH, buffer conditions and also temperature (here 353 K) the program CheckShift (Ginzinger and Fischer 2006) was used. The results of the programs TALOS (Cornilescu et al. 1999) and SimSHIFT (Ginzinger and Fischer 2006) are shown in Figure 6.3 compared to the primarily made prediction and the experimentally results supplemental based on NOE patterns.

Figure 6.3. Comparison of the secondary structure prediction based on a homology search (HS), the prediction made with the programs TALOS (TS) and SimShift (SS) based on a fragment search using the HN, N, C', C α and Cβ resonances, and the experimentally results (EXP) supplemental based on NOE patterns. The unassigned residue is marked in red.

At this point it was already clear that the structure was purely constituted of β-strands, eight in number with similar length. The strand β5 was less clearly predicted, but later identified during the definition of the topology.

6.5 Tertiary and Quaternary Structure

6.5.1 Topology Model

A first anticipation of the global fold of Ph1500-C was obtained by defining the topology, based on the secondary structure prediction. Therefore the H^N-H^N, H^N-H^α, H^N-C^α and H^α-H^α connectivity in β-sheets, derived from HNH-, NNH-, CNH-, HCH- and CCH-NOESY spectra, was identified. All β-sheets showed to be antiparallel (Figure 6.4). Based on the distance restraints the topology can be additionally fixed by applying distance and angle restraints for hydrogen bonds.
The presence of two sheets was noticed, one sheet consisting of the anti-parallel strands β2 (D89-D93), β3 (L96-T101), β4 (E105-N110) and β5 (Y117-K120) and another sheet consisting of the anti-parallel strands β6 (K126-R130), β7 (L133-D138), β8 (K141-R146) and β1 (D78-K83).
The intermolecular character of the antiparallel connectivity between strand β8 and β1 was first detected during the further structure determination process. Consequently, the tertiary structure cannot be defined without defining the quaternary structure. A comparison of the two sheets is shown in Chapter 6.6, disclosing weak sequence similarity.

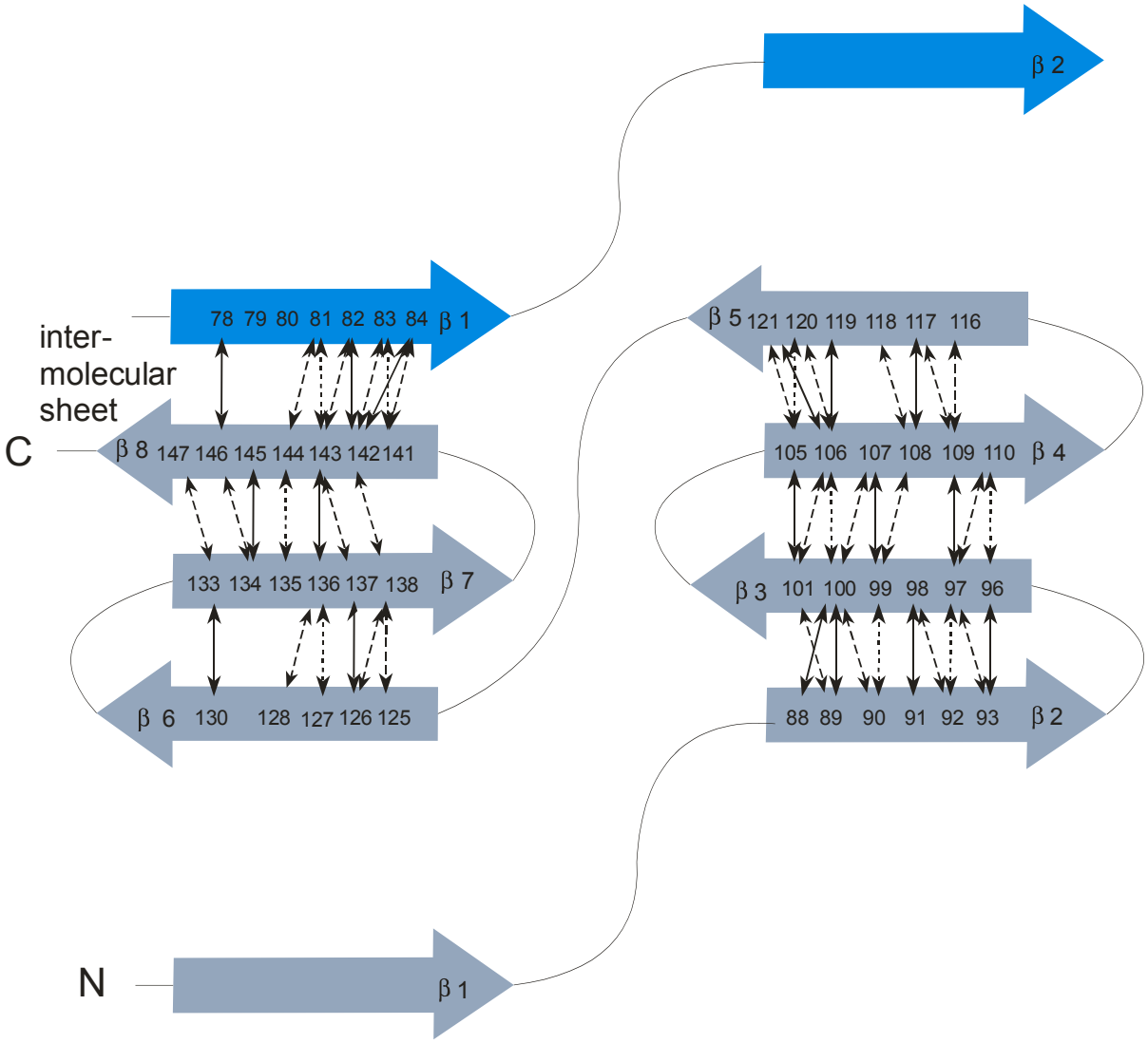


Figure 6.4. **Topology of Ph1500-C.** Arrows indicate the NOE H^N-H^N (solid line), H^N-H^α (dotted) and H^α-H^α (dashed) connectivities in the respective β-sheets, derived from HNH-, NNH-, HCH- and CCH-NOESY spectra. All β-sheets showed to be antiparallel. The intermolecular character of β1 and β8 was detected during the further structure determination process.

6.5.2 X_1 and X_2 Sidechain Rotamer Determination

Sidechain X_1 rotameric state determination and stereospecific assignment were made for 15 of 46 prochiral $C^\beta H_2$ protons. Determinations of X_1 rotamers were also available for 8 of 9 isoleucine residues, 2 of 3 threonine residues and 6 of 10

valine residues, leading to a stereospecific assignment of the $C^{\gamma}H_3$ groups. Determinations of X_2 rotamers were made for 7 of 9 isoleucine residues. All rotamer assignments were made by consideration of patterns of intra- and interresidue NOE connectivities derived from HNH- and HCH-NOESY spectra. Additionally, a hypothesis-test approach was applied using local back-calculation of expectation spectra and comparing them with the experimental ones.

6.5.3 Structure Calculations

The structure calculation with XPLOR (NIH version 2.9.3) was performed using the following input: 646 NOE distance restraints divided in 326 (50%) long range, 86 (13%) medium, 167 (26%) sequential and 67 (11%) intraresidue restraints and classified into four categories corresponding to upper interproton distance restraints of 2.7, 3.2, 4.0 and 5.0 Å. Lower distance restraints were included for weak or absent HN-HN and HN-Hα crosspeaks and by comparing experimental with backcalculated spectra using a minimum distance between 2.5 and 4.0 Å. Intraresidual and sequential upper and lower distance restraints were only applied for H^N-H^{α} and $H^N(i)$-$H^{\beta}(i-1)$ if these were significantly weak or strong, and for $H^N(i)$-$H^N(i-1)$ and $H^N(i)$-$H^{\alpha}(i-1)$. Furthermore 32 hydrogen bond restraints were applied *via* inclusion of pseudo-covalent bonds as described by Truffault et al. (2001). 31 Chi1 and 7 Chi2 rotamer restraints with a tolerance of ±30° were defined. Dihedral angle restraints were derived for backbone Φ and Ψ angles based on C^{α}, C^{β}, C' and H^{α} chemical shifts using the program TALOS (Cornilescu et al. 1999) and SIMSHIFT (Ginzinger and Fischer 2006).

50 structures were calculated in an initial simulated annealing protocol and refined in two further slow cooling stages. The first included a conformational database potential and the second included additional restraints to assure the symmetry of the homohexamer. The force constant on peptide bond planarity was relaxed to 50 kcal/mol/rad^2 in the second slow cooling stage. 17 structures were selected for the final set.

6.5.4 Structure Validation

The structure validation was made under the considerations described in Chapter 4.5. The cross section dimension of the hole in the middle of the C domain is approximately consistent with the electron microscopy data on Ph1500 with 20 Å resolution. The cross section dimension of the full protein could not be measured accurately due to a diffuse outline of the averaged EM pictures, likely caused by the flexibility of the N domain.

The calculated structure ensemble exhibits a superimposition RMSD for the backbone of 0.541 Å (Figure 6.5) and for non-H atoms of 0.984 Å. On the average 9 restraints were defined per residue with an average restraint violation (NOE RMSD) of 0.011 Å and no violations bigger than 0.10 Å. The number of restraints defined per residue (9) is below the average for structures deposited in the PDB Databank, though the percentage of long range restraints (50%) is far above the average due to the fact that redundant information was preferably avoided.

Ramachandran Plot statistics (Ramachandran et al. 1963) calculated with the program PROCHECK (Laskowski et al. 1993) for the regularised average structure showed 90% of all angle combinations to be in most favoured regions, 9% in additionally allowed regions, 1% in generously allowed regions and no residues in disallowed regions (Figure 6.6). The combination of Φ and ψ angles provides an indication of the quality of the structure. Disallowed combinations due to electrostatic repulsion of the carbonyl oxygens and steric hindrance, substantially increase the energy of the protein and thus are unlikely to be present in the native fold.
Further validation parameter were obtained from the program Molprobity (Davis et al. 2007) showing overall acceptable results.

A back-calculation of experimental spectra was used during the iterative process of the structure determination. It was also applied to identify or to verify the sidechain rotamer assignments and the β-turn types.

The consistency of back-calculated and experimental HNH-, CNH-, HCH-, CCH and 2D-NOESY spectra, especially for well-resolved aromatic sidechains and methyl groups provides further evidence for the accuracy of the structure. For signals found to be too strong in the back-calculated spectra a lower distance constraint was set, while for signals found to be to weak, the distance restraints for that residue were revised. A selection of back-calculated HNH-, CNH- amd HCH-NOESY traces in comparison with experimental traces is given in Figure 6.7 and 6.8. The location of the selected residues in the structure can be seen in the topology model (Figure 6.4). Slight differences in intensities that can still be noted were considered to be tolerable and might be partly caused by flexibility and spin-diffusion.

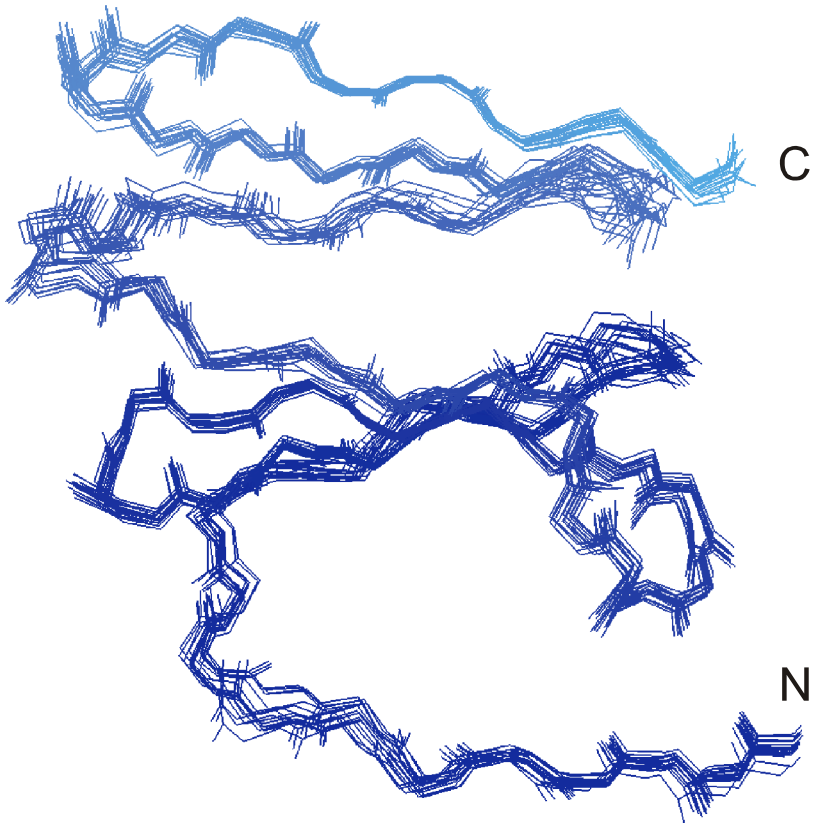
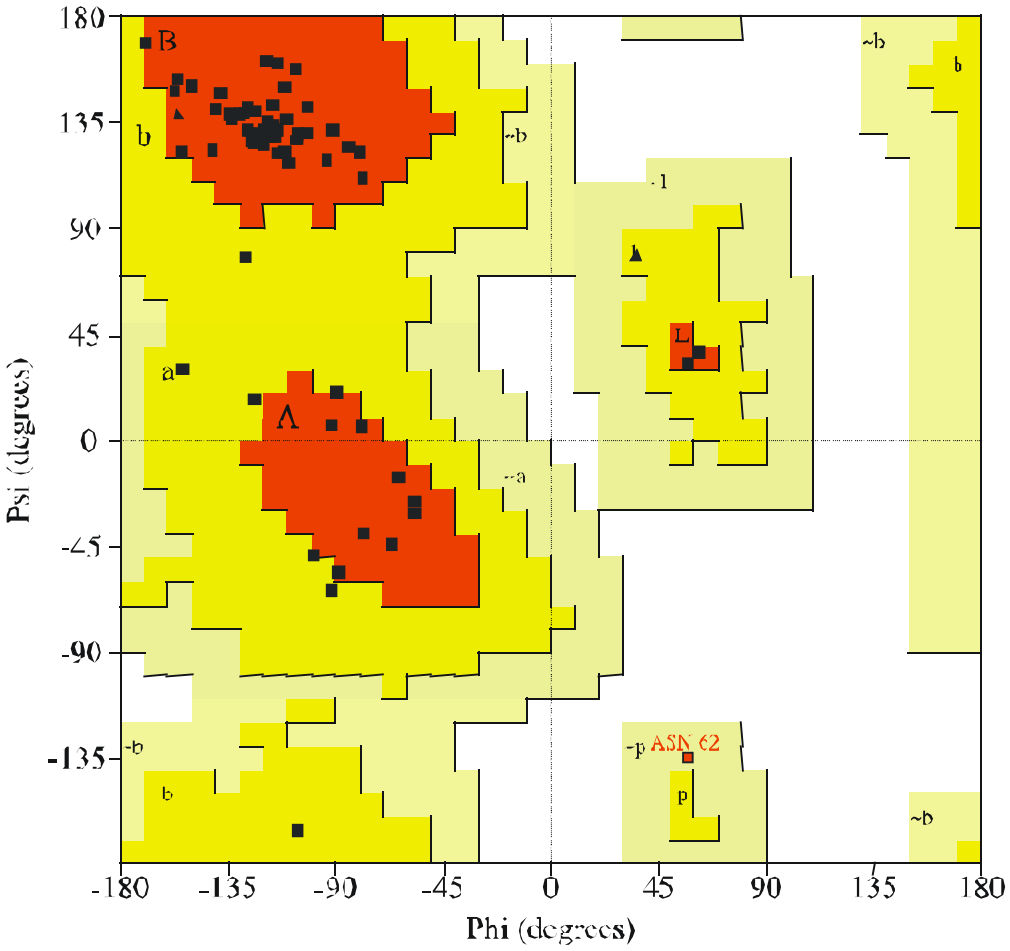


Figure 6.5. The superimposition RMSD obtained for the backbone is 0.541 Å. For the sake of clarity only one monomer unit is shown. Less precise defined regions are mainly both loops between the β-sheets and the β-turns between strand β2 and β3 and strand β6 and β7. Due to the accumulation of solvent exposed amide protons within these regions with an increased water exchange at 353 K, the information obtainable from HNH- and CNH-NOESY experiments was restricted.



Plot statistics

Most favoured regions (red)	60	89.6%
Additional allowed regions (yellow)	6	9.0%
Generously allowed regions (light yellow)	1	1.4%
Disallowed regions (white)	0	0%
N° of non-Gly and non-Pro residues	67	100%
N° of end-residues (excl. Gly and Pro)	2	
N° of Gly residues (shown as triangles)	2	
N° of Pro residues	0	
Total N° of residues	71	

Figure 6.6. Ramachandran Plot statistics for the regularised average structure of Ph1500-C calculated with the program PROCHECK (Laskowski et al. 1993). For glycine residues shown as triangles this classification is not representative. Region B corresponds to β-sheets, region A to right-handed α-helices and region L to left-handed α-helices or to a positive Phi region, respectively.

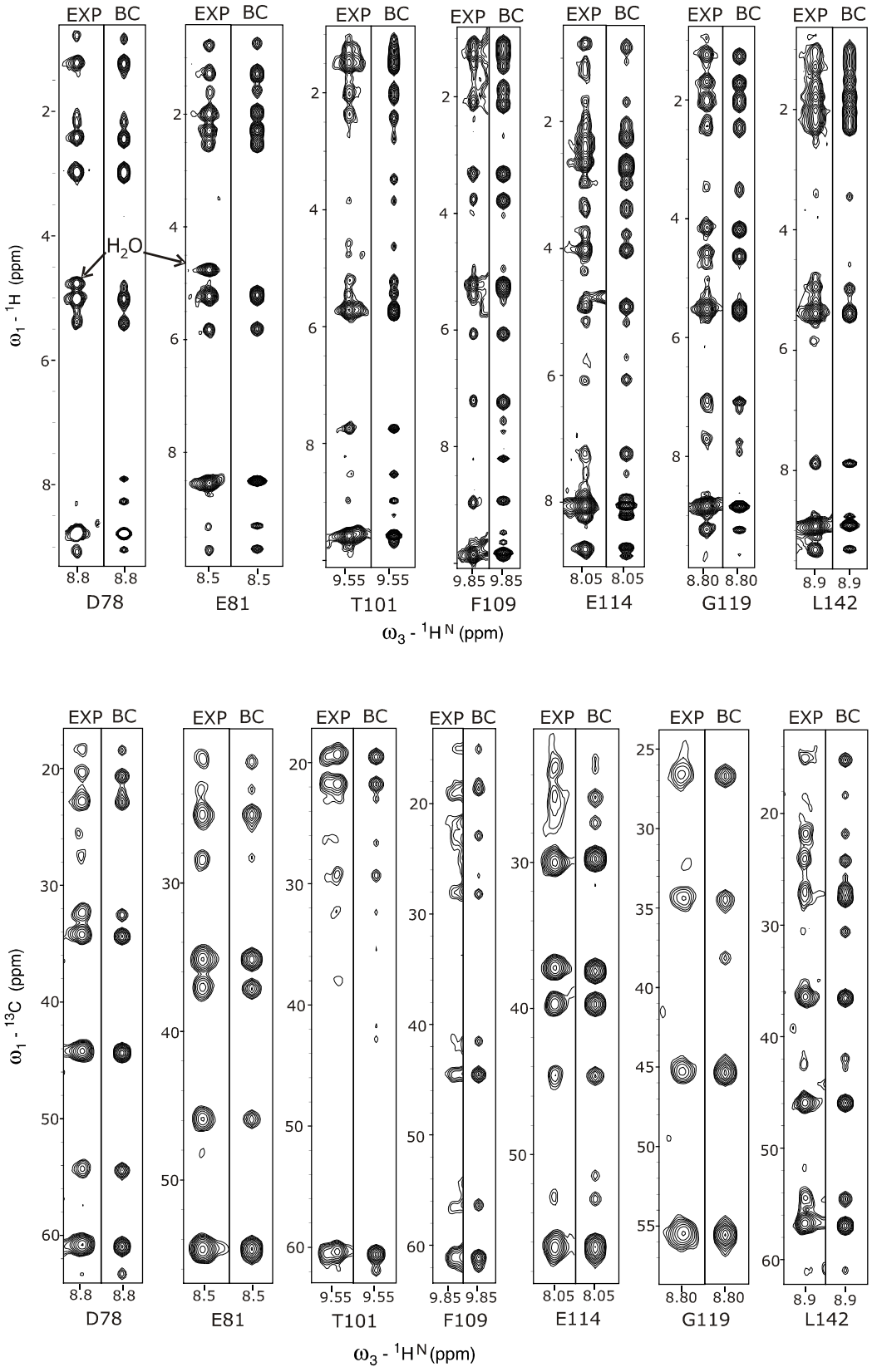


Figure 6.7. Selection of back-calculated (BC) HNH-NOESY (top) and CNH-NOESY traces (bottom) in comparison with experimental (EXP) traces of various well-resolved amide protons. The resonance frequency of water is 4.76.

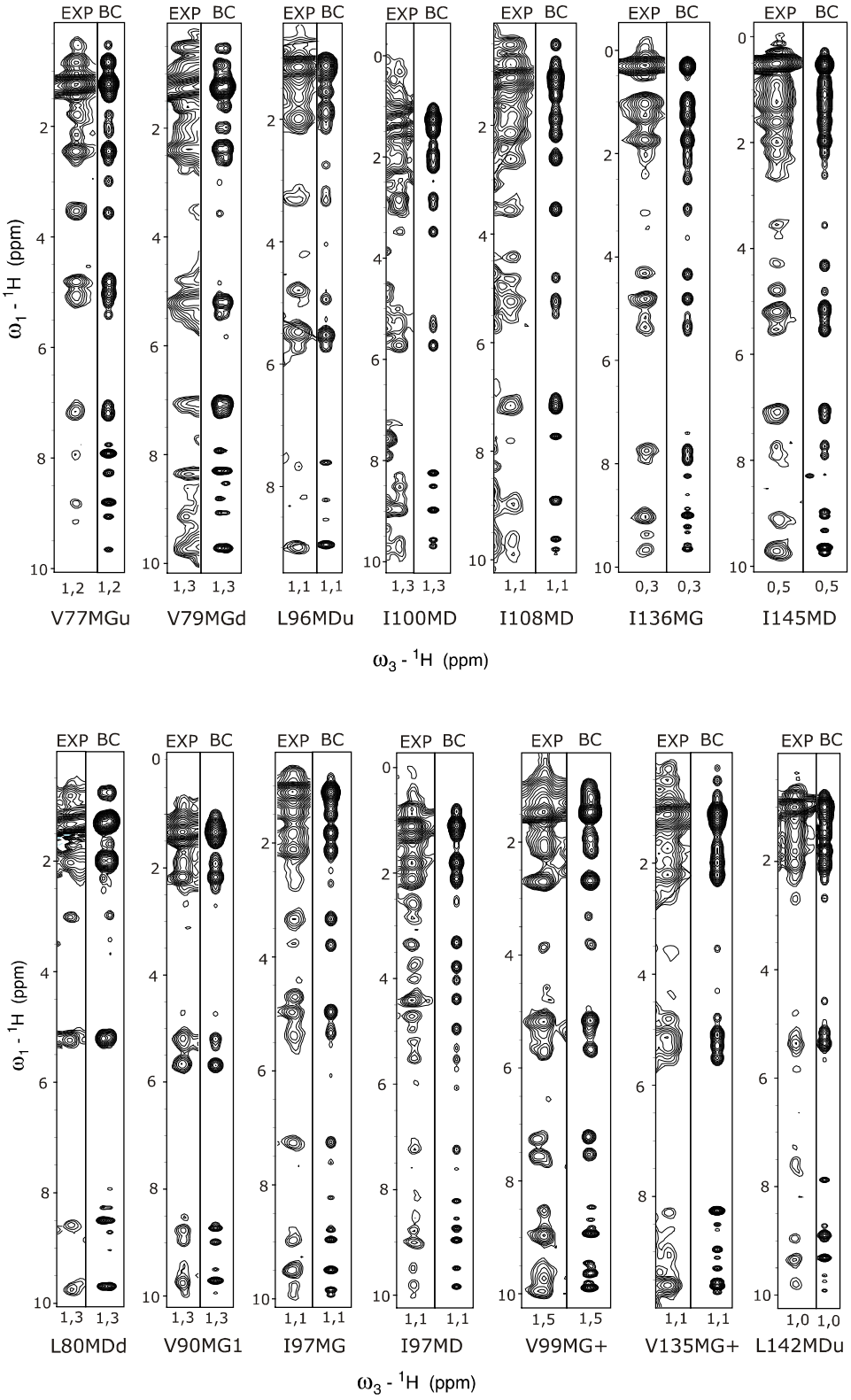


Figure 6.8. Selection of back-calculated (BC) HCH-NOESY traces in comparison with experimental (EXP) traces of various relatively well-resolved methyl groups. These are located either at the predominantly intramolecular (top) or intermolecular (bottom) interface of the four stranded β-sheets.

6.6 Description and Discussion of the Structure

Electron microscopy and NMR analysis made in the first stages of the structure investigation showed Ph1500 to form a hexameric ring. The C-domain shows no significant sequence similarity to proteins of known structure. A secondary structure prediction has initially supposed a OB-fold, indeed it emerged as a twelve bladed propeller fold.

The secondary structure consists of two similar anti-parallel β-sheets (termed blades), with four β-strands each (Figure 6.9). The length of the strands in the respective blades is nearly identical, while the internal twist of the blades is slightly different. Both blades have hydrophobic surfaces on both sides (Figure 6.10) which are covered by hexamerisation, thus resulting in a pseudo dodecamer.

In the initial investigation of the topology, three structure fragments were identified: firstly a β1-strand (D78-K83); secondly a sheet, denoted as blade 2, consisting of the anti-parallel strands β2 (K88-D93), β3 (L96-T101), β4 (E105-N110) and β5 (Y117-K120), with β2 and β3 connected by a β-turn (although identification of the turn type failed), β3 and β4 connected by an α-turn of type II α_{LU} and β4 and β5 connected by a short loop of 6 residues; thirdly another sheet consisting of the anti-parallel strands β6 (N125-R130), β7 (L133-D138) and β8 (K141-R146), with β6 and β7 being connected by a βII'-turn and β7 and β8 by a βI-turn. These three fragments are connected by two short loops (I84-I87) and (F121-L124) consisting of four residues each.

Showing antiparallel connectivity to strand β8 (K141-R146), strand β1 (D78-K83) forms part of the second sheet, denoted as blade 1, together with strands β8, β7 and β6. Contacts between the loop (I84-I87) and the sidechains of A82, V30 and T32 require this connectivity to be intermolecular (see also Chapter 4.8). Consequently the loops form very similar crossovers between the blades.

To cover their hydrophobic surfaces, the two blades have four possible intra- and intermolecular orientations, respectively, which can be described schematically as follows: 1. ‖ ‖; 2. ‖ ‖; 3. ‖ ‖ and 4. ‖ ‖, with ‖ representing one blade. In addition, the loops allow some freedom as to the rotation and tilt of the blades relative to each other. However, a number of unambiguously assigned contacts between the blades, namely between resolved methyl groups and aromatic

sidechains (Figure 6.11) excludes all but one orientation. Further, assigned distance restraints constrain the freedom of rotation and tilt. In the final structure, 52 and 28 distance restraints were assigned at each inter-blade interface, resulting in the structure shown in Figure 6.12. The higher number of restraints corresponds to the mostly intra-molecular interface, with the difference due to the larger number of well-resolved aromatic and aliphatic sidechains. The funnel shape of Ph1500-C is distinguishable in Figure 6.13. The termini are located at the smaller opening.

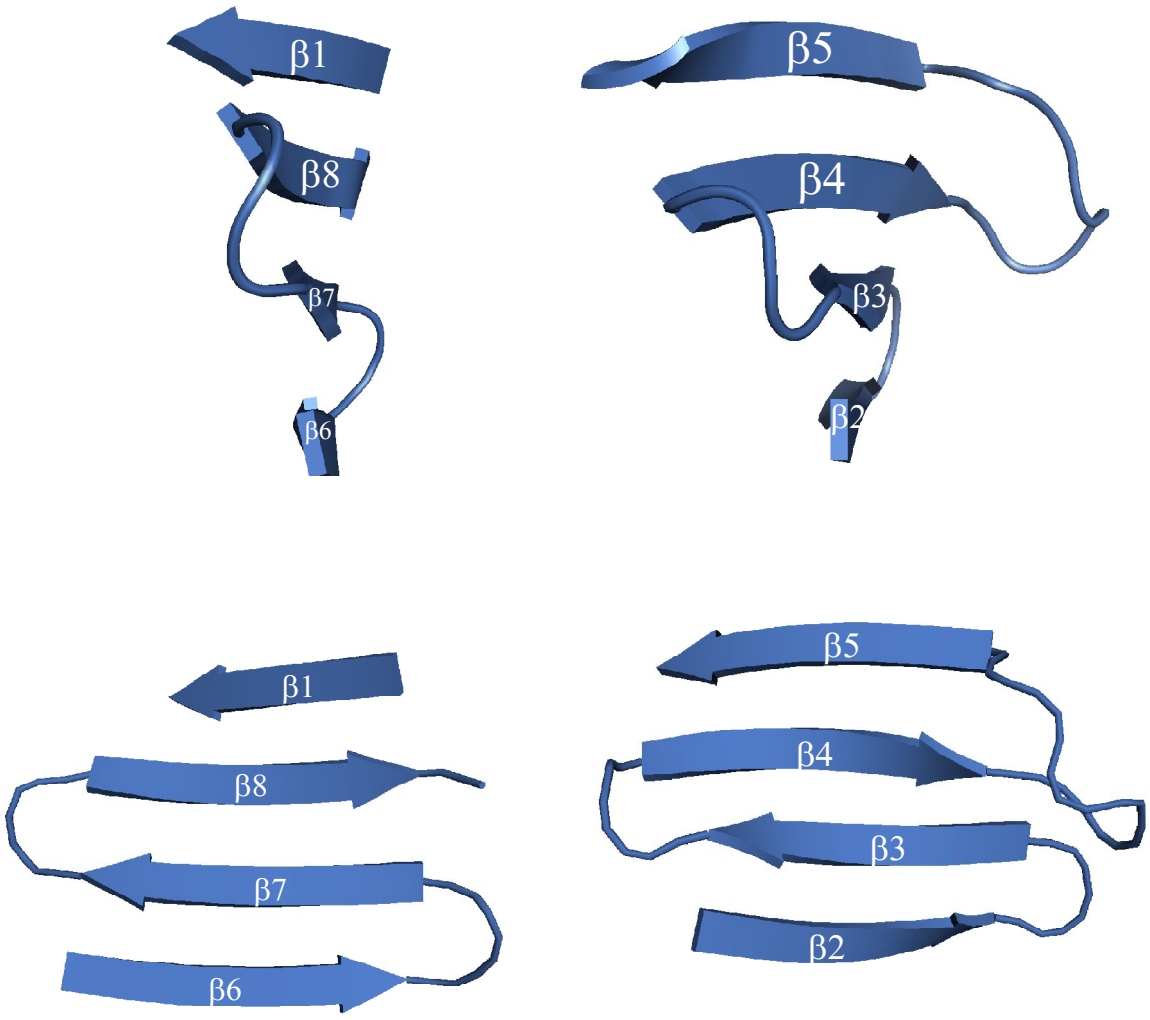


Figure 6.9. Comparison of the two blades of the pseudo dodecameric propeller fold of Ph1500-C. Blade 2 (right) shows a stronger internal twist than blade 1 (left). This is probably due to the β-bulge in blade 2 that is not present in blade 1.

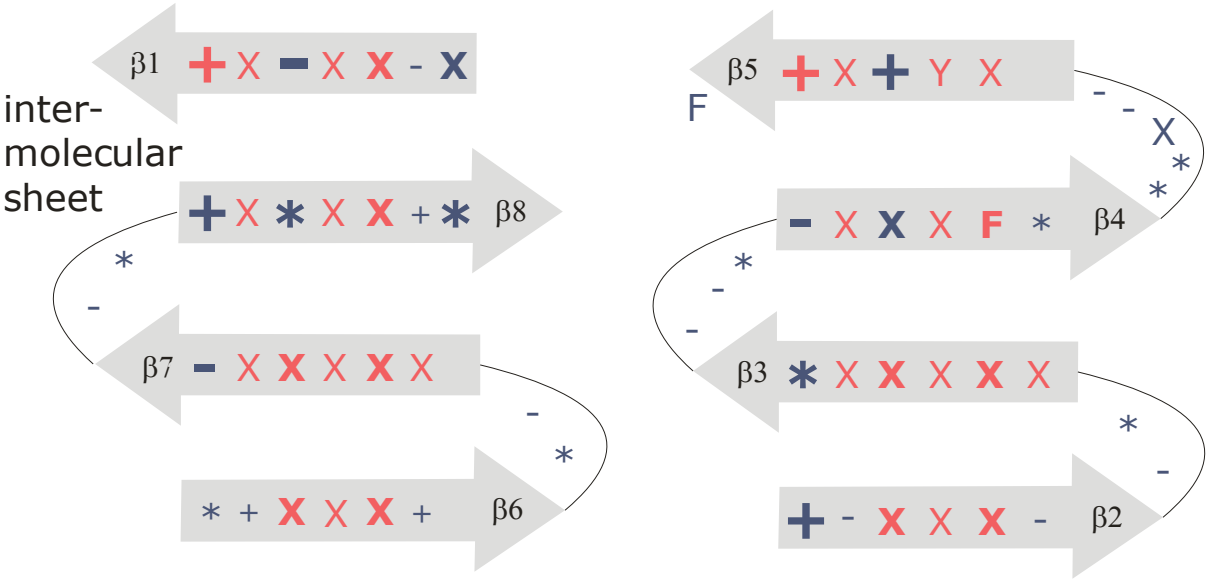


Figure 6.10. Residue types within the blades. Aromatic residues are labelled with their one letter code, hydrophobic residues (GAVLI) are labelled with X, polar residues (TNQ) with *, negatively charged residues (DE) with – and positively charged residues (KR) with +. Distinguishable is the highly hydrophobic surface of the blades and the symmetry between them. Identical or the same type of residues are labelled in red. The sidechains of bold residues are lying above the plane of the page, while the others are lying below.

Figure 6.11. The location of the aromatic residues is shown in a detail of Ph1500-C intra- and intermolecular interfaces. The aromatic residues together with a few well resolved methyl groups were essential for the structure determination.

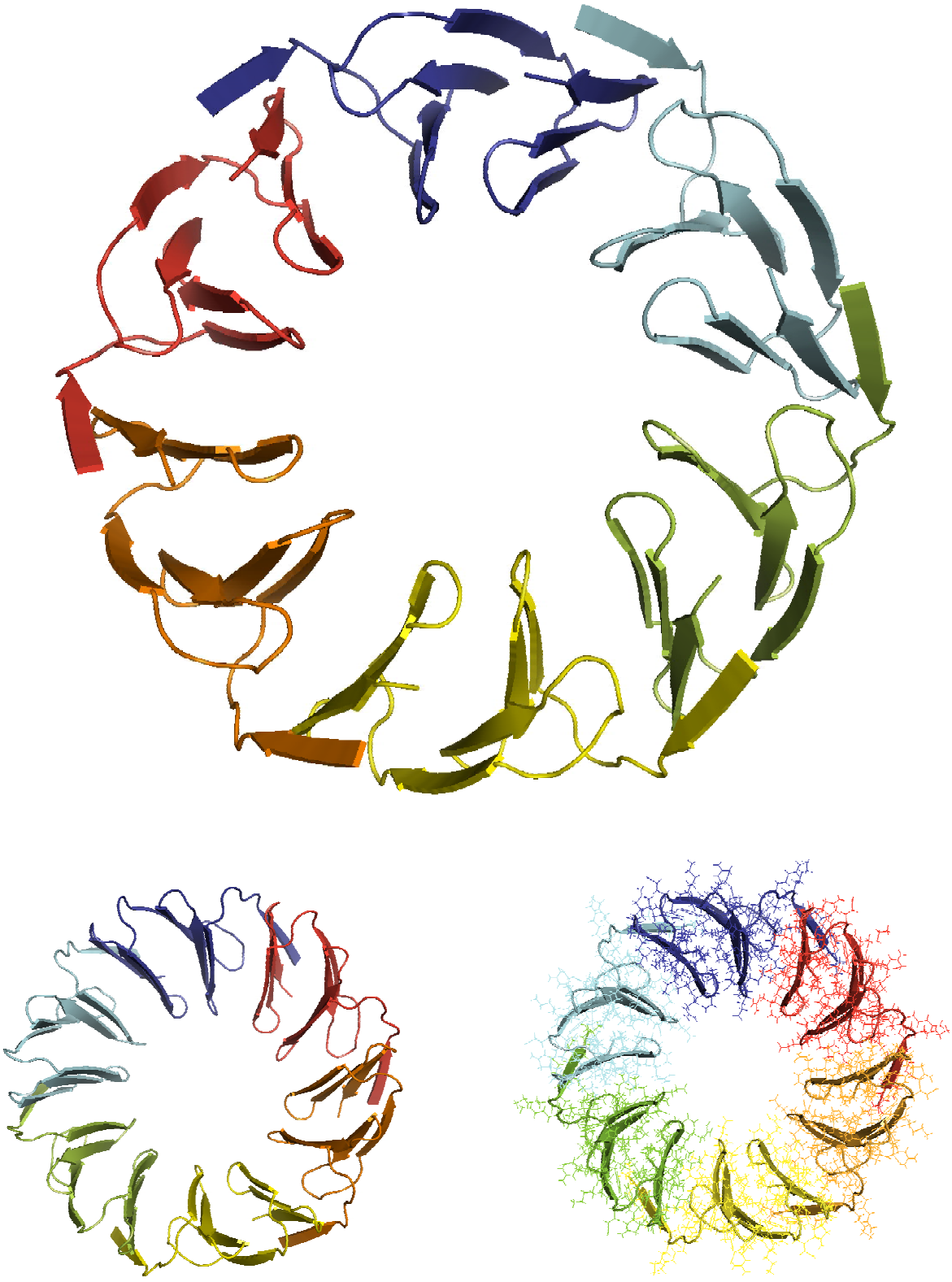


Figure 6.12. NMR solution structure of the hexameric C-domain of Ph1500. A cartoon representation is shown. The upper image shows the view on the smaller opening of the funnel, while the lower images show the view on the larger opening. Each colour represents one monomer unit. Sidechains are displayed in the image on the right.

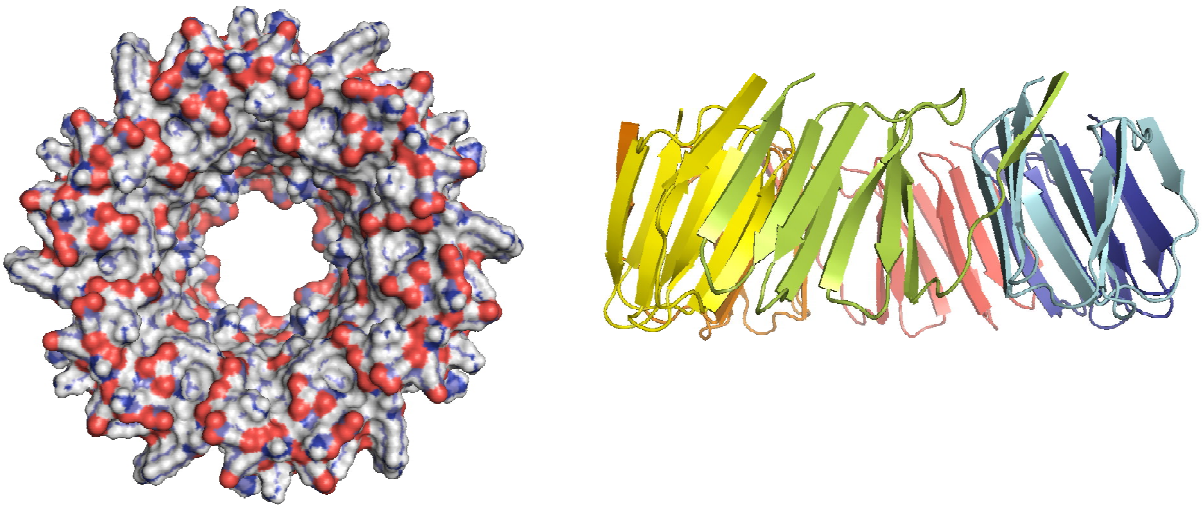


Figure 6.13. NMR solution structure of the hexameric C-domain of Ph1500. A surface representation is shown on the left. The right image shows a cartoon representation and a side view of the propeller fold. In both images the funnel shape of Ph1500-C is visible. The termini are located at the smaller opening of the funnel.

Ph1500-C is the first discovered homo-oligomeric β-propeller which is "velcroed" around the ring. A velcro closure terms the hydrogen bonded contact between both terminal strands as a consequence of a circular permutation of up to three strands of the last blade to the N-terminus. To date, only one other example of a multi-chain (oligomeric) β-propeller was found, which is a trimer of two-bladed monomers, where the N-terminal strand is domain swapped between the monomer blades, i.e. the monomer is a self-contained, two bladed entity. With twelve blades in one ring, Ph1500-C is moreover the largest single-ring β-propeller found in nature so far. In these respects the fold can be considered as unique, although, the topology is also found in single-chain β-propellers with up to nine blades in one subunit. Engineering projects yielded single chain propellers with up to ten blades in one ring. Various single-chain β-propellers with up to fourteen blades are known, which form double rings. Similarity on the level of sequence (Figure 6.14, especially concerning the hydrophobic pattern (Figure 6.10), and of structure between the two blades (Figure 6.9) suggests that the monomer unit arose through duplication. The RMSD of the overlayed single blades is 1.8 Å (Figure 6.15). The velcro closure is probably the consequence of a circular permutation that has increased the stability of the fold.

Generally, β-propellers can be divided into single chain and oligomeric propellers and into canonical propellers and less ordered propellers. Among the latter propellers there are for example propellers with the circle being opened, while others have extended or decorated blades. In single-chain propellers, some single blades show a diversity in sequence to a degree that homology cannot be detected. While canonical single-chain propellers were formed in evolution by repetition on the level of genes and the resulting oligomers, if established, were further stabilised during evolution, it can be supposed that for homooligomeric propellers subsequent point mutations resulted in the formation of different oligomers.

Sequence similarity between Ph1500-C and single-chain β-propeller folds has not yet been detected, despite of the high level of similarity in structure, that might suggest a common ancestor. The high diversity in sequence reflects the general findings about β-propellers. Based on the occurrence of specific sequence patterns, β-propellers are grouped into diverse families. The largest family, the WD40 propellers (Neer et al. 1994; Smith et al. 1999), with usually seven blades gained their name because of the common length of 40 residues of the repetitive unit and a prominent WD motif. Moreover the name imitates a popular brand of motor oil (WD-40), since a functional analogy was initially proposed. The characteristic WD motif is substituted by a FN motif in Ph1500-C. The central blade 2 (strand β2-5) of Ph1500-C has an RMSD of 1.5 Å to a blade from a canonical WD40 propeller (Figure 6.15).

Compared to most single-chain β-propellers observed to date Ph1500-C has a smaller repetitive unit of only 35 and 36 residues, with the crossover consisting of four residues each. Moreover the first and last strand of one blade are less twisted towards each other, especially in blade 1. This is probably due to the β-bulge in blade 2 that is absent in blade 1. However, the exact twist of the blades may not be defined precisely and may differ by several degrees in the native fold.

Since Ph1500-C is the biggest propeller discovered to date, the question arises which effects determine the stabilization of its size. Further investigations are required to answer this question. One would be to examine the influence of the

β-bulge contained in most canonical propellers in the first strand, thus being involved in the velcro closure. This β-bulge is absent in Ph1500-C. Its influence could be examined by introducing the β-bulge in Ph1500-C and deletion in the other constructs, respectively and subsequent analysis of the formed oligomers. Additionally it might be of interest to construct a single chain mutant of Ph1500-C to compare the timescales of the process of folding and to design a single chain mutant with seven and more repeats to find out to which bladenumbers it will fold, though the information available would probably not justify the efforts.

```
                  bSSSSS     SSSSSS    SSSSSS    SSbSSSS

blade 1'                                          VDV-LEAKIK--
blade 2       GIKDVILDE--NLIVVITEENEVLIFNQNLEELYRGKFE--
blade 1"      NLNKVLVRN--DLVVIIDE-QKLTLIRT

WD40          SVWGVAFSPDGQTIASASDDKTVKLWNRNGQLLQTLTGHSS
```

Figure 6.14. Comparison of the sequences of the two blades of Ph1500-C and with a blade from a canonical WD40-propeller. Consensus hydrophobic residues are marked in blue and identical residues are marked in red.

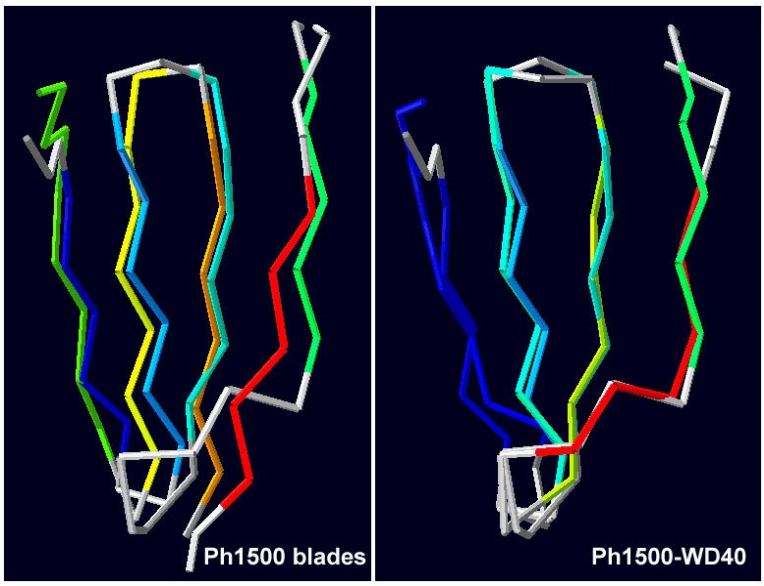


Figure 6.15. Comparison of the two blades of Ph1500-C (left) and between the central blade 2 of Ph1500-C and a blade of a canonical WD40 propeller (right).

6.7 Discussion of the Approach that gave Access to the High Resolution Structure of the 49 kDa Homohexamer

About ten years ago, structural studies of large proteins appeared limited to about 25 kDa, since the tumbling of the protein slows down about linearly with increased molecular weight. The tumbling rate affects the transverse relaxation and thus the linewidth of NMR signals, i.e. the signal to noise ratio and the resolution. The latter becomes the more important the higher the number of signals, thus structural investigations are most often limited by severe spectral overlap. In recent years, various methods have been developed that have increased this size limit (an overview is given in Chapter 4.5), however, structural studies of proteins above 35-40 kDa remain a challenge.

Presently, with a molecular weight of 49 kDa, the homohexameric C-domain of Ph1500 represents one of the largest protein structures undertaken *ab initio* by NMR spectroscopy. To solve the structure, the approach chosen initially was trying to identify the residues involved in the oligomerisation and subsequent mutation of these residues to receive stable monomers. For this purpose a model structure was constructed, based on assigned HN-HN contacts and a secondary structure prediction obtained with the program TALOS (Cornilescu et al. 1999). The constructed model revealed a β-sandwich fold with larger hydrophobic areas on the surfaces of both sheets. Various combinations of two or three hydrophobic residues located in these areas were chosen and mutated to arginine, though by mischance these mutants still formed hexamers. Only now, with the solved structure in hand, it is clear that this attempt had to fail, since it was supposed that the hexamerisation would be solely mediated via hydrophobic contacts. The structure shows that it is additionally mediated via backbone contacts.
Later on it was also planned to prepare samples with amino-acid specific labelling of isoleucine, leucine and valine methyl groups (Rosen et al. 1996 ; Goto et al. 1999) and to apply methyl-TROSY based experiments (Yang et al. 2004; Tugarinov and Kay 2004; Tugarinov et al. 2004; Tugarinov and Kay 2005).

However, the unfavourable relaxation properties could be overcome by exploiting the hyperthermophilic origin of the protein. Spectral overlap was a minor problem corresponding to the 71-residue monomer. Since the correlation time τ_c of a protein decreases with increasing the temperature, measurement of sidechain and NOESY experiments at 353 K made the 49 kDa protein appear like one with about a third of its size with respect to the linewidth. Thus these experiments could be completed on a uniformly $^{13}C/^{15}N$-labelled sample in substantially reduced measuring time and allowed complete sidechain assignment (97%) at full proton density.
Due to a faster water exchange of amide protons at elevated temperatures the sequential assignment was done at 318 K on a uniformly ~85% deuterated sample using TROSY-type experiments and later on most backbone resonances were re-assigned at 353 K. Hence, based on a nearly complete resonance assignment a high resolution structure determination has been possible.
The detailed process of the structure determination was described in this chapter. The positive feature of this approach was the simplicity and the low costs compared to standard methods applied for large proteins, such as selective labelling, which allow moreover only structures of medium resolution.

A comparison of the Hα region of ^{13}C-HSQC spectra at four temperatures (293, 313, 333 and 353 K) is shown in Figure 6.16. There is a continuous improvement of spectral quality with increasing temperature, since amide protons are not involved. Figure 6.17 shows the first processed FID's of HNCA spectra at four temperatures between 293 K and 353 K. A comparison of 3D trHN(CA)CO spectra measured at 298 K with 192 scan and a receiver gain of 4096 and at 313 K with 256 scan and a receiver gain of 16384 is shown in Figure 6.18. In both 3D experiments the effect of reduced transverse relaxation with increasing temperature is especially significant, due to the extra transfer step and further incremented time. In contrast, 3D NOESY experiments are less sensitive to transverse relaxation since the magnetization is along Z during the coherence transfer through space. While the influence of a faster water exchange of solvent exposed protons is not distinguishable in the shown 1D comparison, it can be easily detected in Figure 6.19, showing an area with NH- and NH_2-groups in

^{15}N-HSQC spectra of Ph1500-C measured at four temperatures between 293 and 353 K. At 353 K all signals arising from NH_2-groups have disappeared.

By choosing the optimal temperature it has to be considered that weak or missing amide protons will also show no or weak traces in ^{15}N-HSQC-based NOESY experiments, which need to be recorded at the same temperature as the sidechain experiments. Figure 6.20 shows 2D planes of ^{15}N-HSQC-NOESY spectra at four temperatures between 293 K and 353 K.

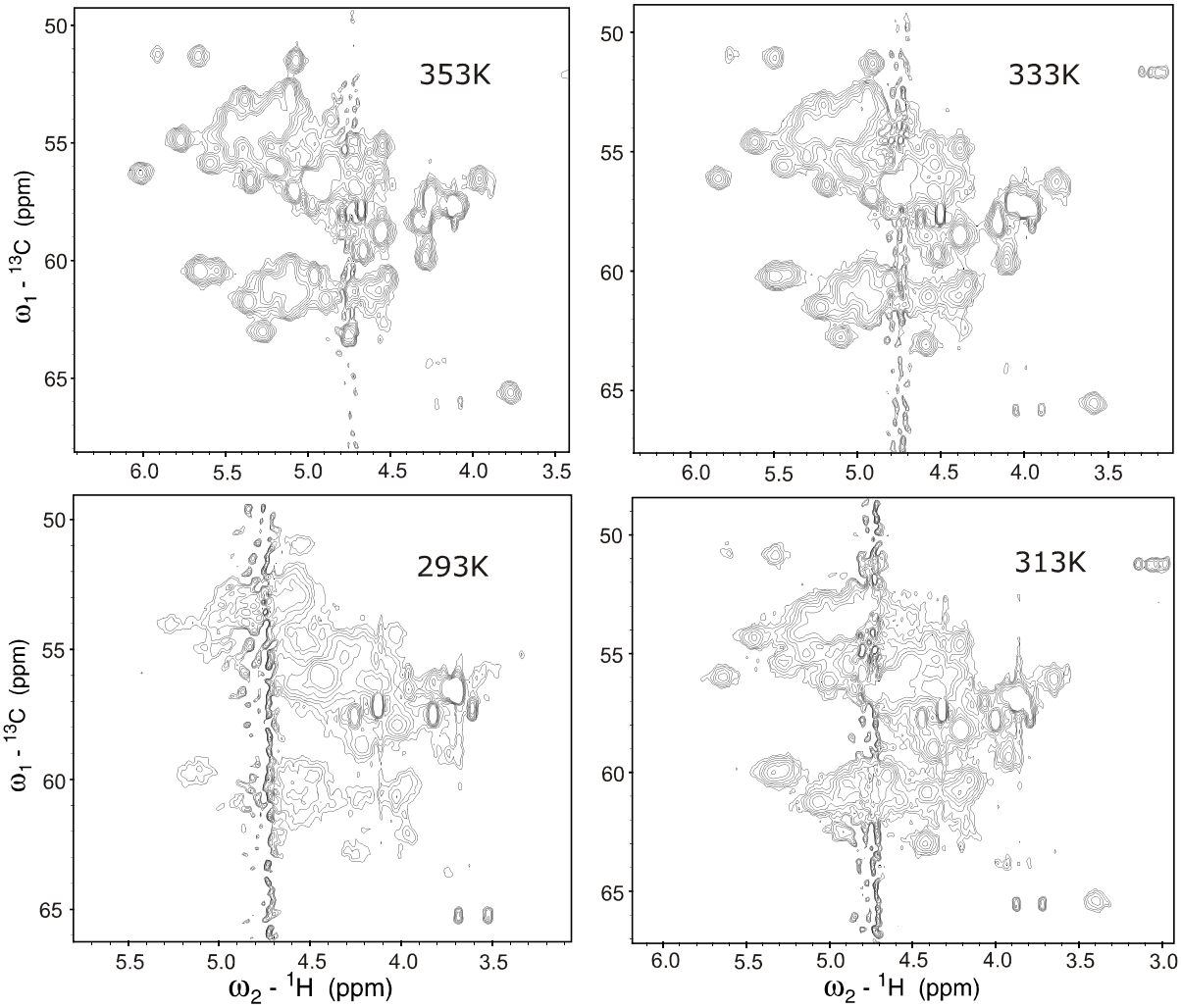


Figure 6.16. Comparison of the Hα region of ^{13}C-HSQC spectra at four temperatures (293, 313, 333 and 353K) measured at a 750 MHz spectrometer on the uniformly ^{13}C/^{15}N-labelled sample of Ph1500-C. Visible is a constant increase in signal to noise ratio and resolution with increasing temperature.

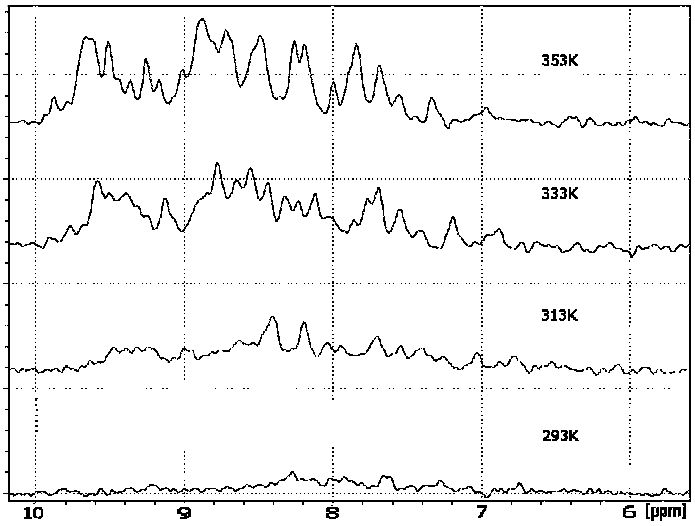


Figure 6.17. Comparison of the first processed FID's of the 3D HNCA experiment measured at various temperatures (293, 313, 333 and 353K) at a 750 MHz spectrometer on the uniformly $^{13}C/^{15}N$-labelled sample of Ph1500-C. Due to the extra dimension and thus a further transfer step the effect of line narrowing with increasing temperature is more fundamental than in 2D experiments. However, the influence of faster water exchange of solvent exposed protons is not detectable here.

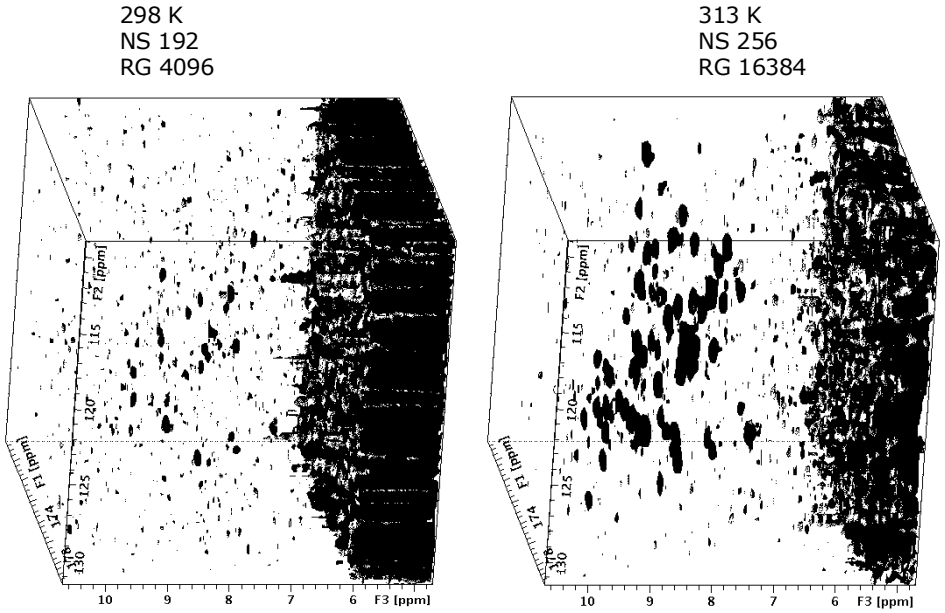


Figure 6.18. Comparison of 3D trHN(CA)CO spectra measured at 298 K with 192 scans and a receiver gain of 4096 (left) and at 313 K with 256 scans and a receiver gain of 16384 (right) at a 900 MHz spectrometer on the uniformly $^{2}H,^{13}C/^{15}N$-labelled mutant sample of Ph1500-C.

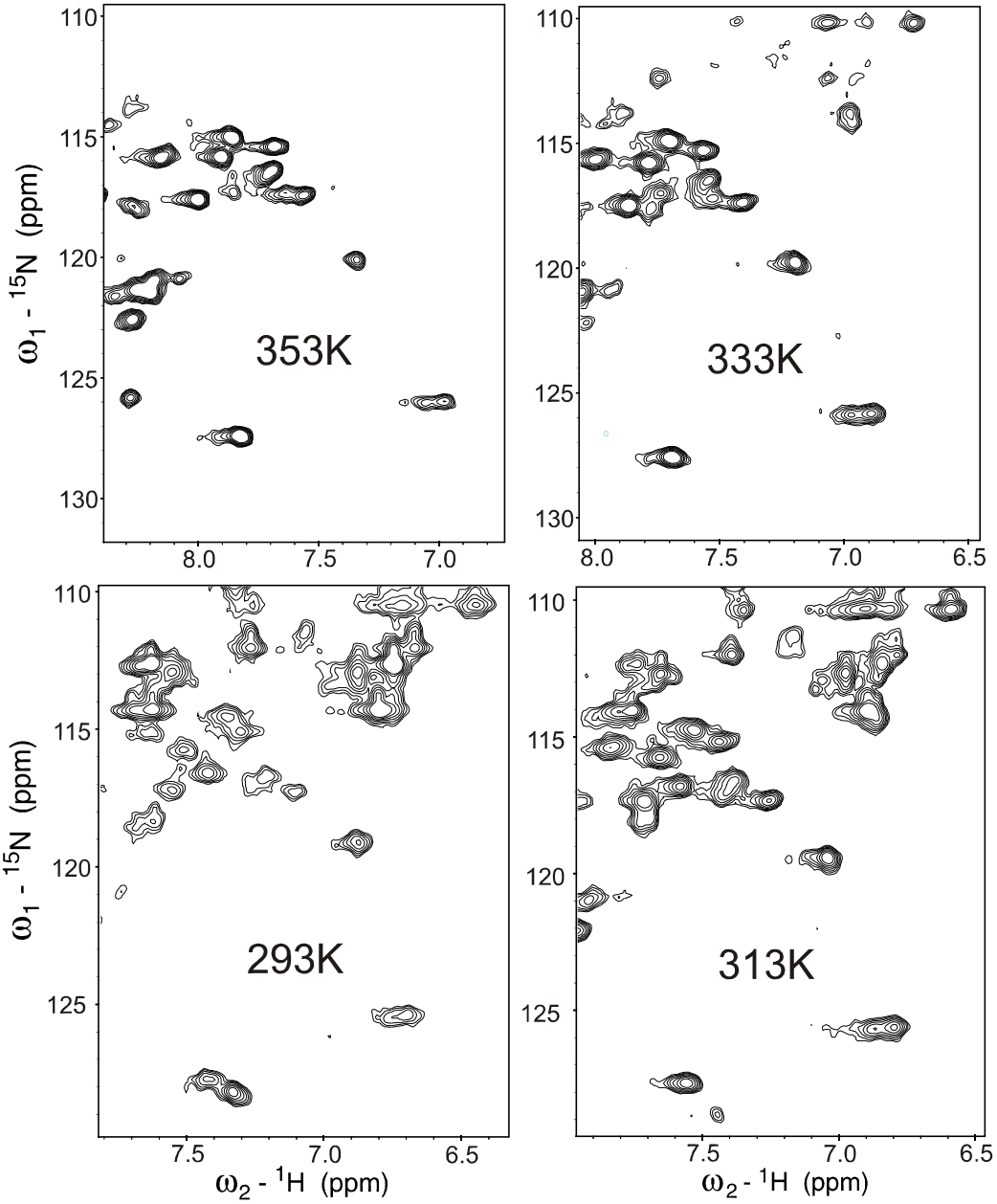


Figure 6.19. Comparison of an area of ^{15}N-HSQC spectra showing signals arising from NH and NH_2-groups. The experiments were recorded at a 750 MHz spectrometer on the uniformly ^{13}C/^{15}N-labelled sample of Ph1500-C. Measurement temperatures were 293, 313, 333 and 353K. At 353K all signals arising from NH_2-groups have disappeared due to the faster water exchange. Solvent exposed NH groups also show a descreased intensity at higher temperatures. Visible is also the line narrowing effect.

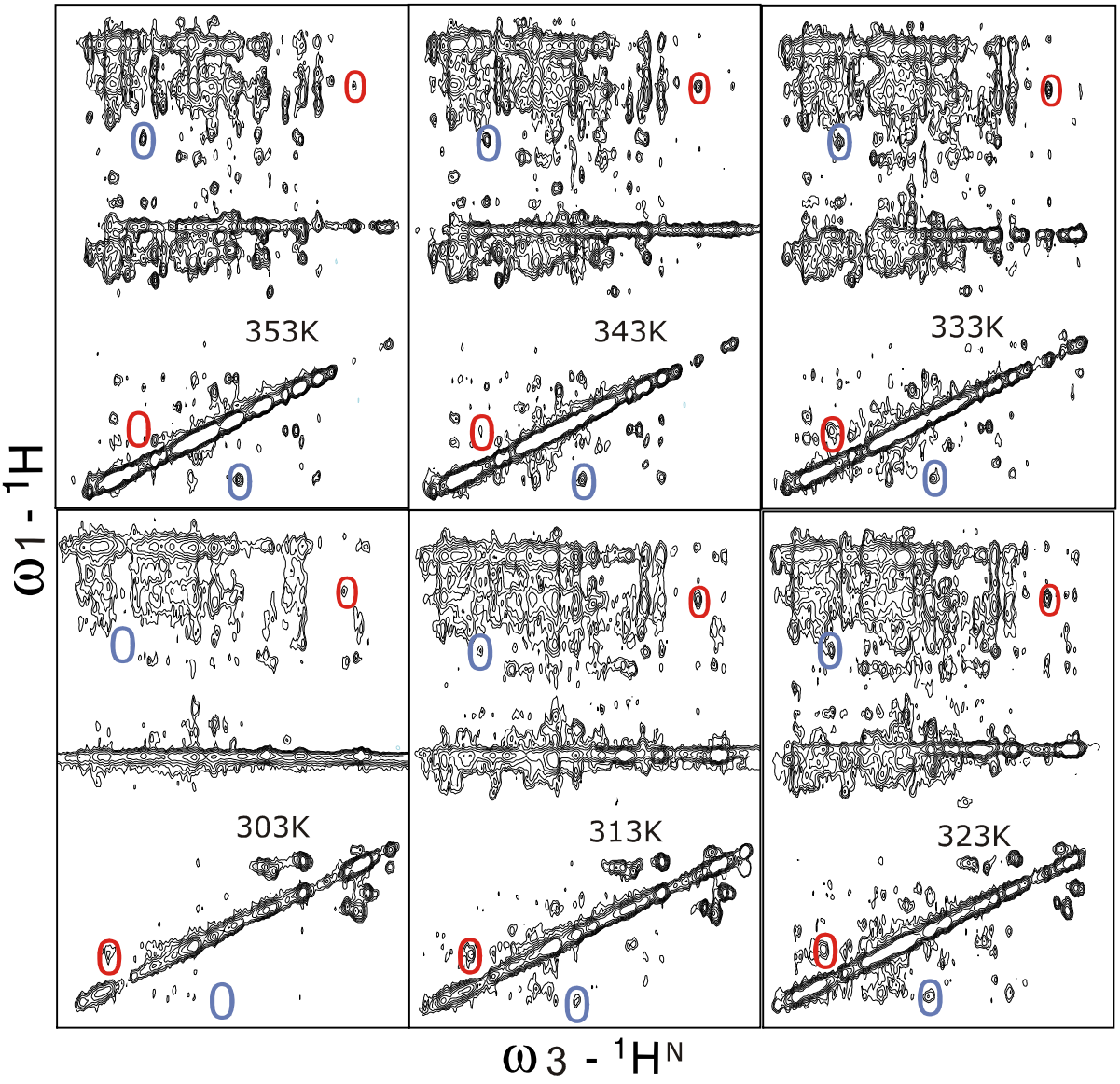


Figure 6.20. Comparison of 2D H^N-H planes of 3D HNH-NOESY experiments measured at six temperatures between 303 and 353K measured at a 750 MHz spectrometer on the uniformly $^{13}C/^{15}N$-labelled sample of Ph1500-C. Visible is the constant line narrowing effect with increasing temperature. In contrast the signal to noise ratio is either decreased or increased. Blue circles mark examples for solvent protected amide protons where the effect of faster water exchange is negligible. Red circles mark examples for solvent exposed amide protons where the effect of a faster tumbling is compensated by a faster water exchange of these protons at higher temperatures. For the sake of clarity chemical shift axis are omitted. Shown are the areas from ~6.8-10.2ppm in the direct dimension and from ~0.2-10.2ppm in the indirect dimension.

The global tumbling correlation time τ_c of an isotropically reorienting spherical molecule of volume V_h relates to the bulk solvent viscosity η through the well known Stokes-Einstein relation:

$$\tau_c = \frac{\eta V_h}{kT}$$

The effect of an increased absolute temperature T in the denominator is positive, but negligible, whereas the non-linear temperature dependence of the viscosity of water strongly effects the correlation time (Figure 6.21), particularly in the range just above room temperature. At 328 K, the correlation time is expected to be about half of that at room temperature and at 353 K about one third.
The linewidth of water becomes narrower at the same time, thus, in case of non-optimal water suppression with the water signal being the limiting factor for the receiver gain, the latter can be increased. The linewidth of water in dependence of the temperature are given in Figure 6.22. For the sake of clarity all comparisons were made using the same receiver gain.

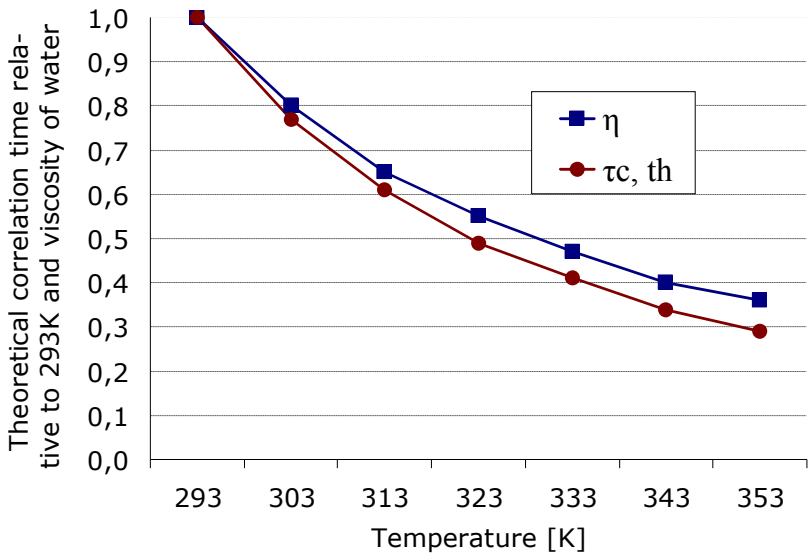


Figure 6.21. Temperature dependence of the viscosity of water (η) in [$\cdot 10^3 kg \cdot m^{-1} \cdot s^{-1}$] and of the theoretical correlation time ($\tau_{c,th}$) relative to 293K. The changes are slightly different, as the correlation time depends not only on the temperature dependent viscosity of water, but also on the absolute temperature.

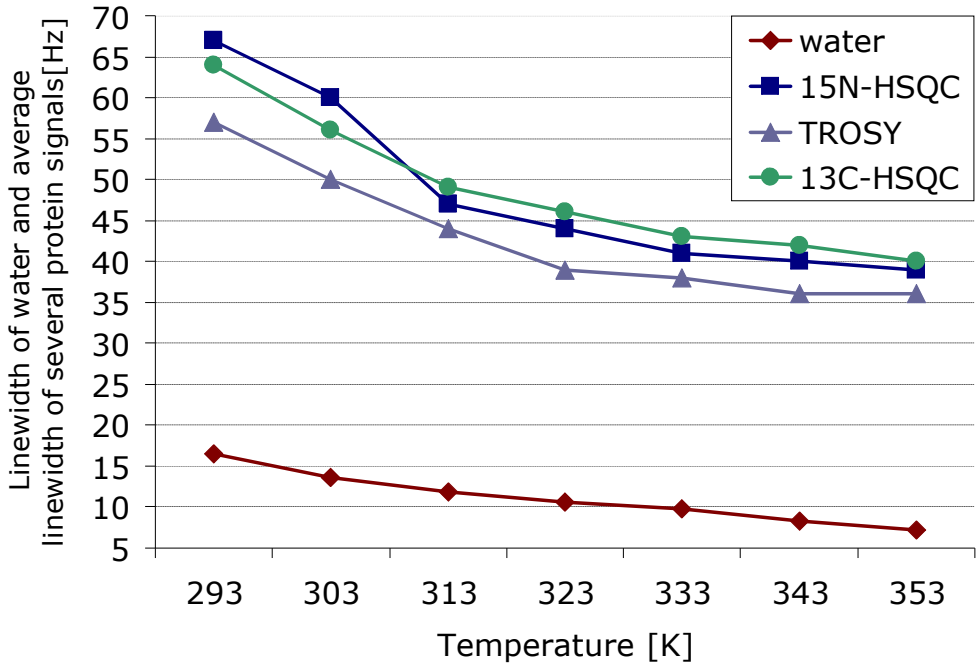


Figure 6.22. Temperature dependence of the linewidth of water and the average linewidth of various protein signals (T32, G50 and L55 in ^{15}N-HSQC and TROSY; I31γ1u, I67δ1, I68β and I76γ2 in ^{13}C-HSQC) in Hz, measured in the proton dimension at a 750 MHz spectrometer on the uniformly ^{13}C/^{15}N-labelled sample of Ph1500-C. All linewidth are given with an error of $\pm$ 0.4 Hz for water, $\pm$ 3.0 Hz for protein signals in ^{13}C-HSQC and 7.0 Hz for protein signals in ^{15}N-HSQC and TROSY.

A question that arises from the reported results of the present work, is the extent to which high temperature measurements can be applied for larger proteins in general for a structure determination at full or high proton density. At first glance, Ph1500-C appears to be an exception, perhaps representing the size limit for structure determination of uniformly labelled proteins due to the hexameric and hyperthermophilic nature. However, the hexamer exhibits a donut shape with a diameter around three times that of each subunit, resulting in a considerably longer effective correlation time than would be expected for spherical or rod-shaped proteins of similar molecular mass, or for multiple domains connected by flexible linkers. The linewidth of various selected signals are given in Figure 6.22. In addition, although Ph1500-C is extremely thermostable, it was advantageous, but not necessary, to measure at the highest possible temperature. Adequate NOESY spectra could have been obtained at

intermediate temperatures, where many proteins of thermophilic and even some of mesophilic origin would be stable. A prerequisite for interpreting NOESY information on uniformly labelled samples is a largely complete sidechain assignment, and this is a limiting factor. For Ph1500-C, the highest measurement temperatures were most advantageous for obtaining sidechain assignments with standard (H)CCH-TOCSY experiments. For proteins which are less thermostable, experiments less sensitive to transverse relaxation and/or deuteration level would be needed. In recent years several strategies involving direct ^{13}C-detection have been proposed, e.g. ^{13}C-detected HCC-TOCSY (Hu et al. 2005) and ^{13}C-^{13}C-NOESY spectra (Bertini et al. 2004, Bermel et al. 2006; Hu et al. 2006; Matzapetakis et al. 2007), which should allow full assignment at lower temperatures.

An optimization of the measurement temperature should be useful in general and the highest possible temperature for all experiments, except for backbone experiments, can be recommended. A plot of the linewidth against temperature (Figure 6.22) provides the information if there is a significant line broading or narrowing due to conformational exchange. For Ph1500-C, the plot has almost the same shape as that of the correlation time against temperature, thus it can be supposed that there is no relevant conformational exchange. Optimizing the temperature certainly takes some time, though this time can be gained later on.
In case that the optimal temperature for backbone assignment is below the optimal temperature for all other experiments, the question arises if nevertheless it can be completed to a sufficient degree at the higher temperature, or, if not, a lower temperature should be chosen. Admittedly this requires a later re-assignment at the higher temperature. The positive feature of the latter approach follows from the fact that a high deuteration level has a negative effect on NOESY and sidechain experiments by reducing the obtainable information, while it has a positive effect on backbone experiments. Thus, by using a highly deuterated sample for the backbone assignment, the longer correlation time at lower temperature can be compensated.

Assuming that the upper limit of a structure determination based on a highly-protonated, uniformly labelled sample would be 30 kDa at room temperature

(298 K), then the limit at 328 K would be 60 kDa with respect to the limited resolution caused by broad lines, neglecting the greater spectral complexity. Figure 6.23 shows a histogram of measurement temperatures for NMR structure determinations of proteins above 25 kDa (source: PDB website). The majority of these proteins were solved at temperatures between 298 to 303 K. Therefore an increase in measurement temperature of 10 to 20°C would place many proteins in a temperature range where structure determination at high, uniform proton density would be possible.

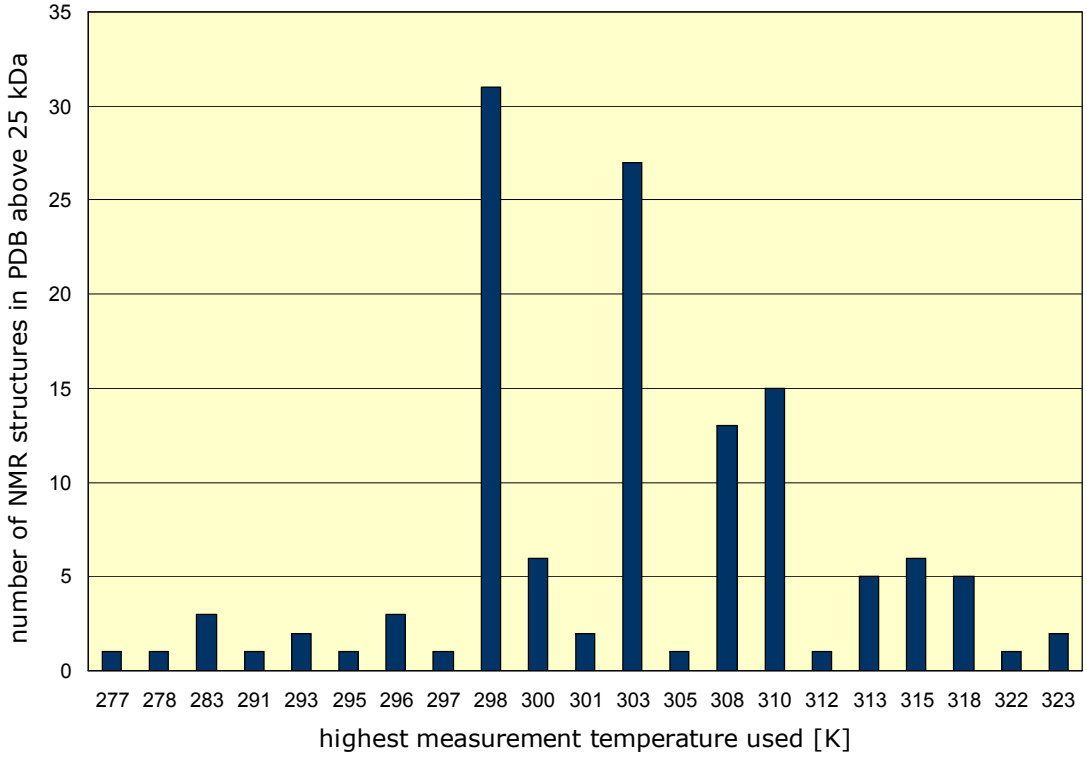


Figure 6.23 Reported measurement temperatures of PDB NMR structures above 25kDa (June 2008). To date, 323K is the highest reported measurement temperature used.

To address the problem of spectral complexity, 4D experiments (Grzesiek et al. 1995) and the technique of segmental labelling (Xu et al. 1999) have been introduced, though the application to much larger proteins is limited by transverse relaxation. Thus measurement at elevated temperatures could be also used as a complement to these techniques.

Despite the fact that manipulation of solution conditions and protein constructs, including the introduction of mutations, has been widely used to improve the solubility and general solution behaviour of NMR samples, it seems that similar efforts aiming at increasing the measurement temperature are less common. The results presented here suggest that optimising the measurement temperature is as important for larger proteins as optimising deuteration levels and labelling schemes.

In the last decades many efforts were put into the engineering of thermostability into mesophilic enzymes by stabilizing mutations with the aim to broaden their industrial use and to fasten reaction times (Eijsink et al. 2001, 2004 and 2005; Bloom et al. 2005). The most relevant here, is probably the semi-rational "consensus concept" (Lehmann and Wyss 2001; Lehmann, Pasamontes et al. 2000), which is based on the hypothesis that the consensus amino acids in a group of homologous proteins contribute more than average to the stability of the protein. Consequently, substitution of non-consensus by consensus amino acids based on a multiple sequence alignment should improve thermostability. It has been shown that the increase in thermostability is more considerable for substitutions of those residues that occur with a lower frequency at the respective position in the alignment (Wang et al. 1999; Wang et al. 2000), for a combination of stabilizing single mutations (Nikolova et al. 1998) and for applying the consensus concept to the entire sequence (Lehmann and Kostrewa et al. 2000). Applying the consensus concept, it appears likely that increases in measurement temperature of 10 to 20 K could be routinely achieved. Thus, further improvements in engineering proteins for thermostability might gain a strong impact on NMR measurements of large proteins, either by allowing larger proteins to be placed under the size limit of feasibility and to be solved at high, uniform protein density or simply to reduce measuring and analyzing time.
It is most definitely worth ascertaining if a homologous protein from a (hyper)-thermomophilic organism exists, before starting to solve the structure on a large and less thermostable protein.

Chapter 7

Structure and Function of the Full Protein Ph1500

7.1 Structure Determination of the Full Protein Ph1500

Protein expression, isotope labelling and purification of the full protein Ph1500 were done by Sergej Djuranovic at the MPI Tübingen for Developmental Biology under the same conditions used for the C-domain of Ph1500 (Chapter 6.1).
Two samples of the full protein were prepared in 20 mM sodium phosphate buffer (pH 7.4) containing 250 mM NaCl, one uniformly $^{2}H,^{13}C/^{15}N$-labelled sample and one sample containing a mixture (1:1) of uniformly $^{13}C/^{15}N$-labelled protein and uniformly $^{2}H,^{13}C/^{15}N$-labelled protein within one hexamer, obtained by mixing the two dissimilar labelled samples, unfolding the protein and refolding. The latter sample was initially prepared in the hope of gaining some information about which residues are located on the hexamerisation surface of the protein, which would be expected to show a faster relaxation than within the former sample.
On these samples a CRINEPT (Riek et al. 1999) with an optimized tranfer delay of 3.5 ms and a TROSY were recorded on a Bruker DMX900 spectrometer at 318 K. The number of signals that appeared in both spectra was substantially smaller than expected, presumably because the full protein exhibits a diameter around four to five times that of each subunit. This produces a considerably higher effective correlation time than for a spherical or rod-shaped molecule of the same molecular mass (100 kDa), leading to much faster transverse relaxation.

Overlaying the TROSY spectra with those of the single domains measured at the same temperature (Figure 7.1) shows that it contains only signals arising from the N-domain. Accordingly, the N-domain must have a slower transversal relaxation due to a higher flexibility compared to the C-domain, thus disallowing a fixed orientation of the two domains with respect to each other.

Attempts to determine the diameter of the full protein with EM failed due to a diffuse outline. In contrast, the cross section dimension of the central pore could be determined more accurately. A more precise determination of the uncertainty of the external diameter might give clues as to possible restrictions in the flexibility of the N-domain.

As a consequence of different transverse relaxation times, the structure of the full protein was calculated under the assumption that the N-domain is flexibly attached to the C-domain, hence symmetry restraints were applied only for the non-flexible parts.

The input for the structure calculation of the full protein with XPLOR was the sum of the inputs of both single domains. Figure 7.2 shows a cartoon representation of one calculated structure of the full protein with the N-domain exhibiting different orientations towards the C-domain. A line representation image of one calculated structure of Ph1500, which matches relatively well to the EM pictures of the class average of aligned particles, is shown in Figure 7.3.

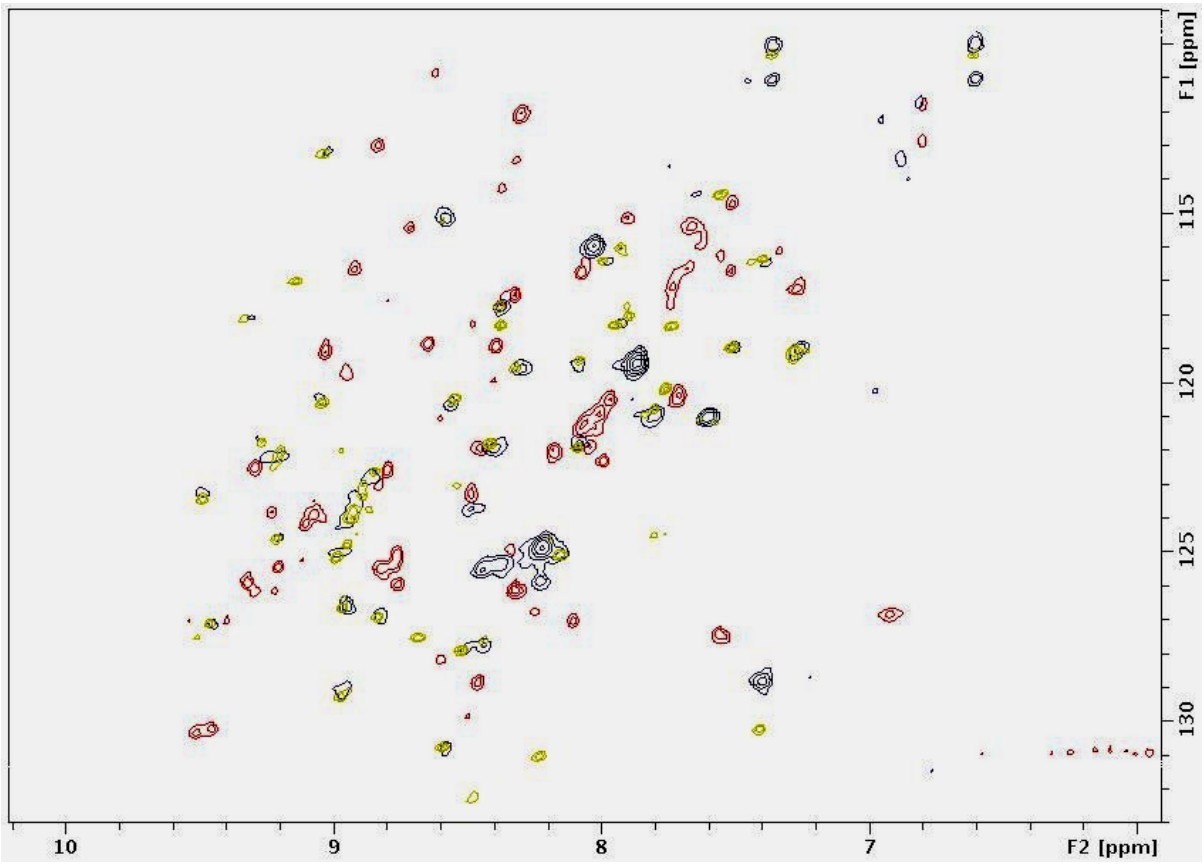


Figure 7.1. Overlayed TROSY spectrum of the full protein Ph1500 (in blue) with the ^{15}N-HSQC spectrum of the N-domain (in green) and the TROSY spectrum of the C-domain (in red). All shown experiments were measured at 318 K. The comparison shows that the TROSY spectrum of the full protein contains only signals arising from the N-domain.

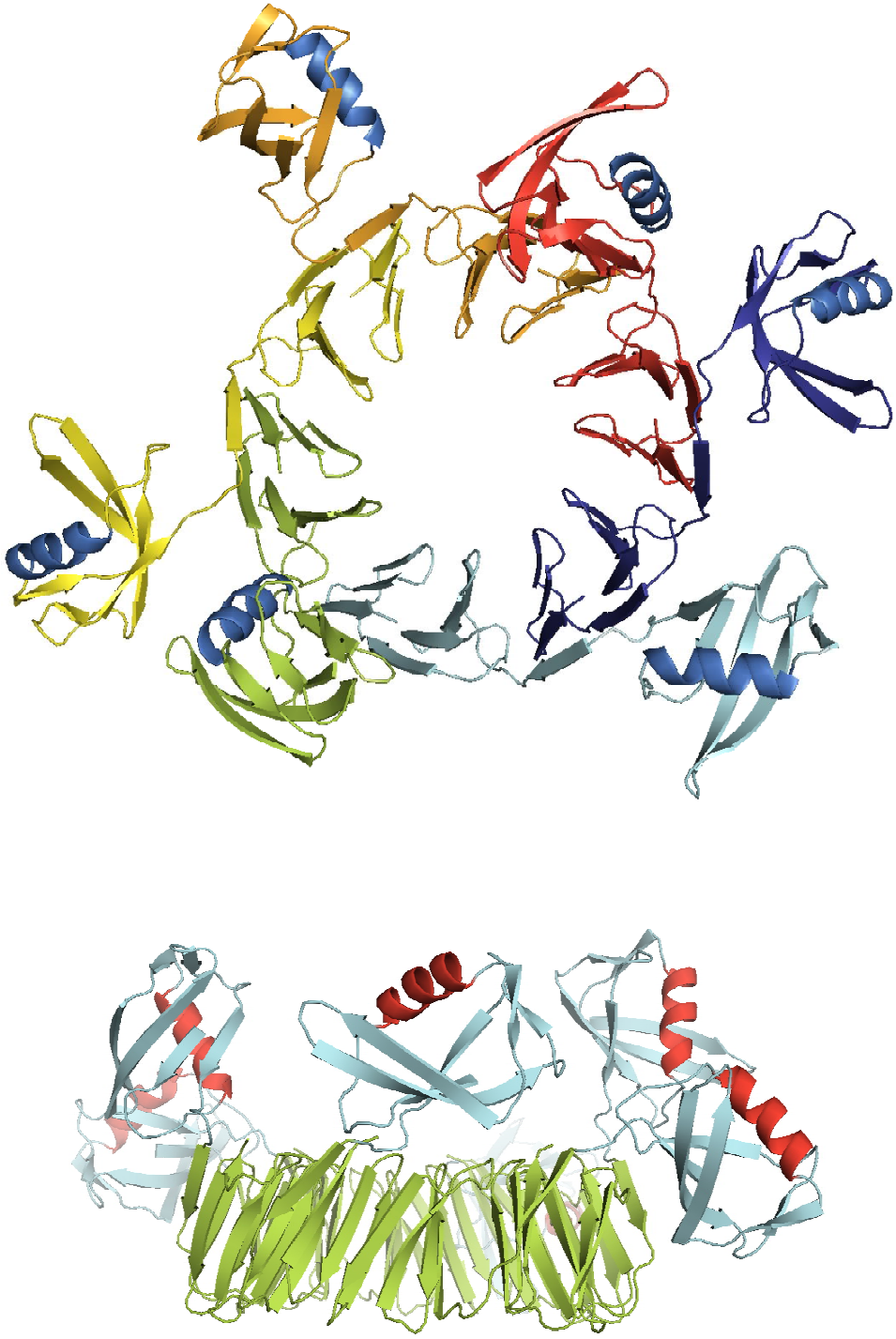


Figure 7.2. NMR solution structure of the full protein Ph1500. The structure was calculated under the assumption that the N-domain is flexibly attached to the C-domain, thus applying symmetry restraints only for the non-flexible parts. The input for the structure calculation was the sum of the inputs of both single domains. In the upper image one colour represents one monomer unit.

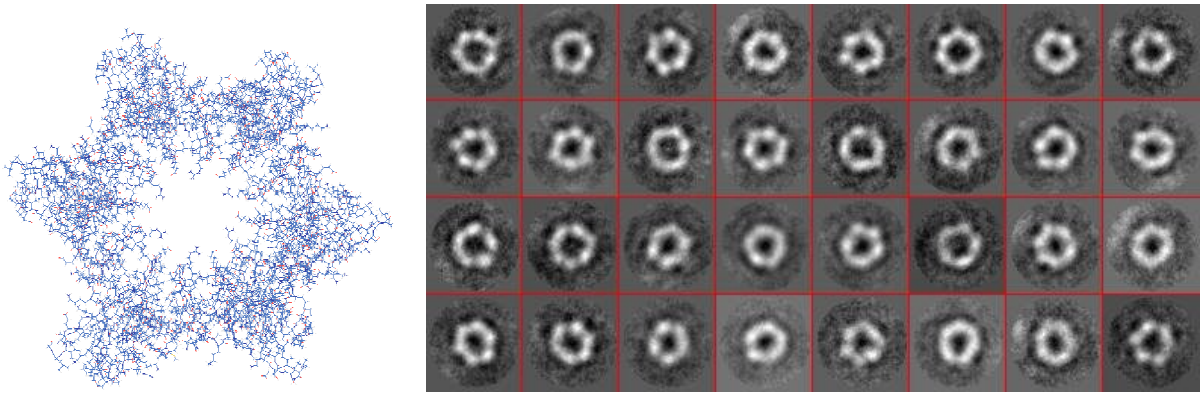


Figure 7.3. Line representation of one calculated structure of Ph1500, that matches relatively well to the EM pictures of the class average of aligned particles.

7.2 Supposed Function of Ph1500

Presently, the function of Ph1500 is still unclear. Based on its gene environment to Endonuclease III, primarily the hypothesis arose that the hexameric C-domain slides down DNA until it encounters a G-T mismatch, and the β-clam recruits or docks endonuclease III to sites of these mismatches to fulfill its activity, which is excision of the mismatched pair (see also Chapter 3.2). Though, a titration of Ph1500-N with endonulease III with ratios ranging from 1:0.5 to 1:5.7 and chemical shift mapping in ^{15}N-HSQC did not show any observable changes in chemical shifts (Figure 7.4). This gives rise to the hypothesis that Ph1500 binds to mismatched DNA that has been nicked by Endonuclease III to recruit the next factor. Constructs of Ph1500 analyzed for possible interaction with unspecific DNA fragments did not show any binding, thus the next step will be the examination of an interaction with single-stranded DNA as well as with mismatched double-stranded DNA mimicking DNA damage. On parallel it is intended to use Electron Microscopy to detect a possible complex formation with DNA. Additionally, it might be possible to map changes in chemical shifts in a CRIPT experiment recorded on the supposed complex of the full protein with DNA and endonuclease III. At a structural level, it can be noticed that the size of the central pore would allow the supposed binding to DNA.

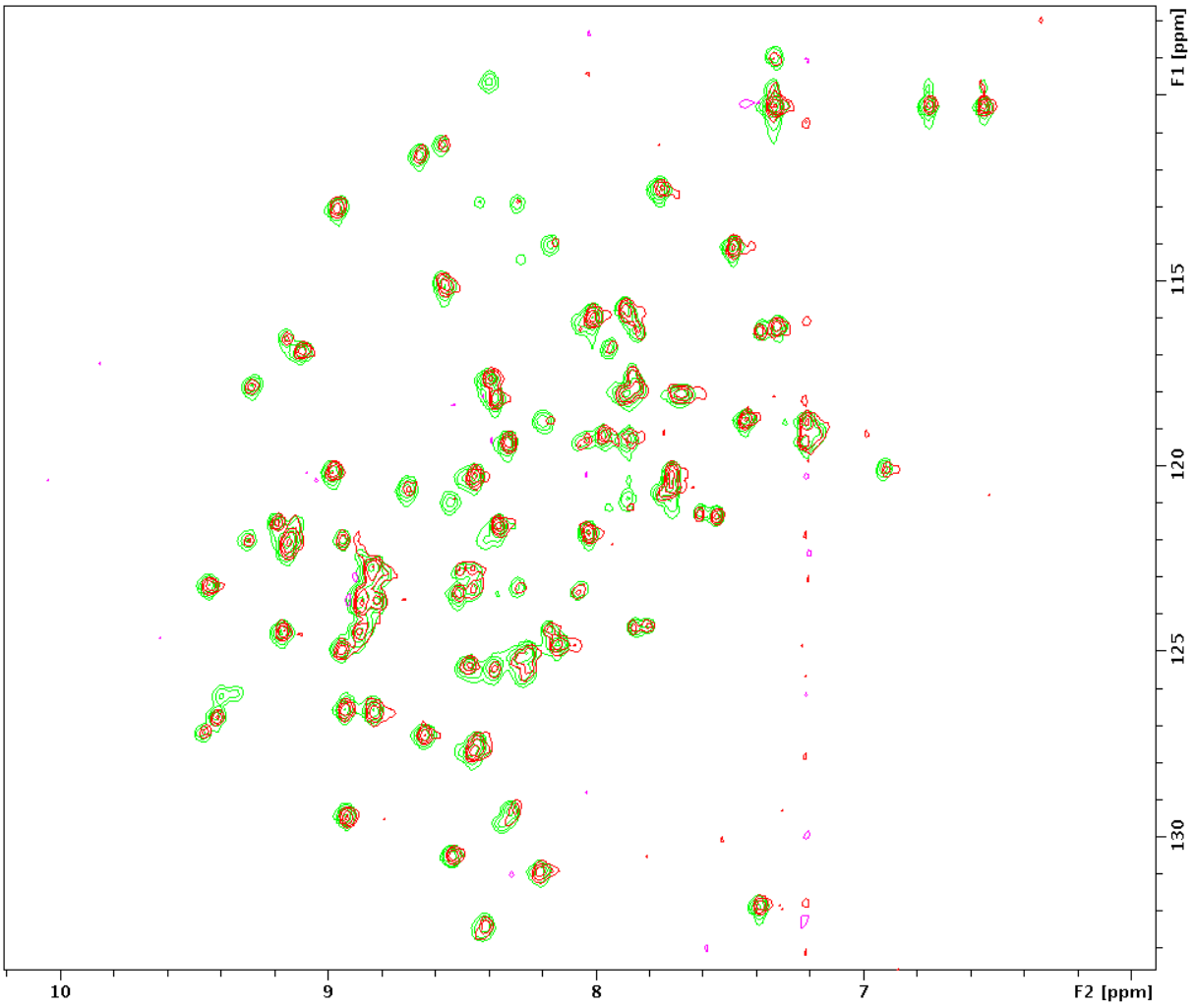


Figure 7.4. Chemical shift mapping in ^{15}N-HSQC for a titration of Ph1500-N with endonulease III in four steps with ratios 1:0.5, 1:1, 1:2 and 1:5.7. The reference spectrum is shown in red and the last titration point in blue. Since chemical shifts do not change, a binding to endonuclease can be excluded.

The evolutionary context of Ph1500 (Figure 7.5) is discussed in Chapter 3.1 and in the respective chapters of the single domains.

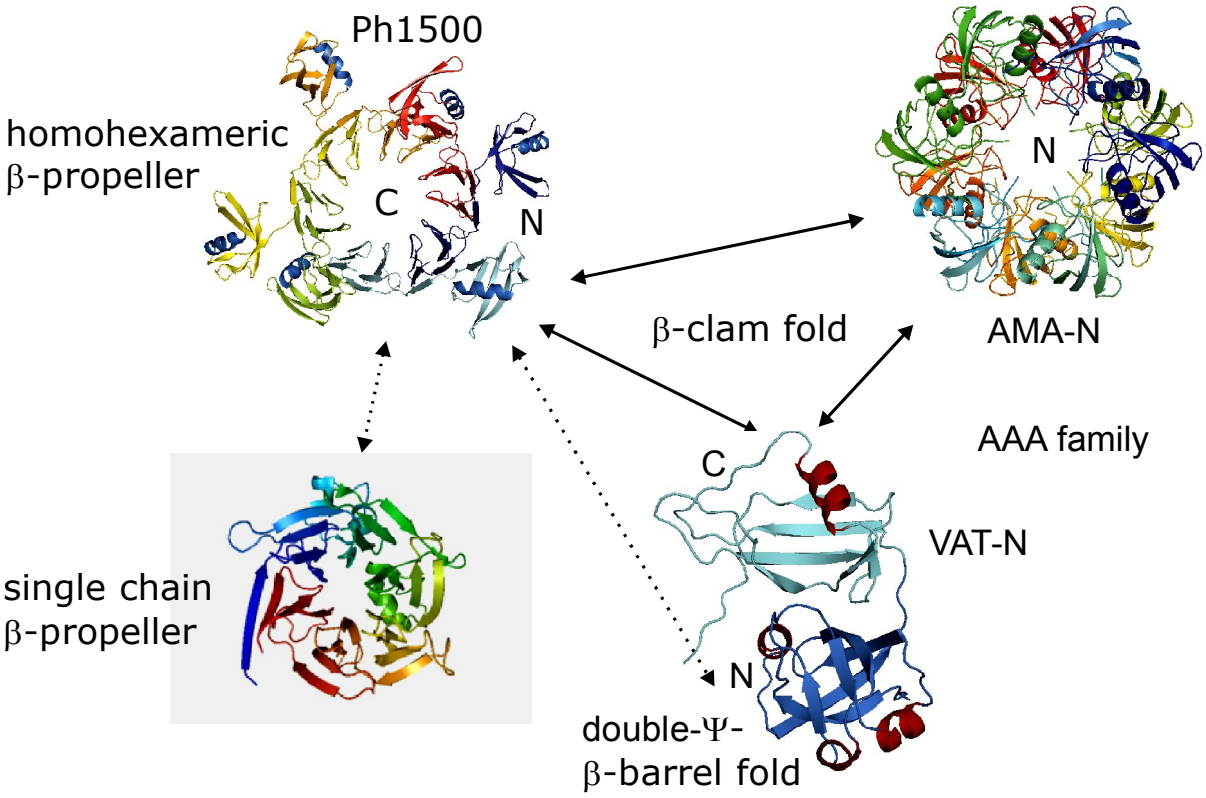


Figure 7.5. Evolutionary context of Ph1500. Drawn through arrows mark a distinct evolutionary relationship, while dotted arrows mark a supposed evolutionary relationship.

Summary

In this work, the solution structure of the 100 kDa homohexameric protein Ph1500 was investigated. Ph1500, from the hyperthermophilic archae bacteria *Pyrococcus horikoshii*, was one protein chosen in a study of the evolution of complex protein folds from simpler peptide fragments (Chapter 2). Ph1500 appears to be related to members of the cradle-loop β-barrel meta-fold, though the evolutionary transformations involved are still unclear. Ph1500 is assumed to have a role in DNA repair, based on its gene environment to Endonuclease III (Chapter 3). Electron microscopy has previously shown it to form a hexameric ring. The protein consists of two domains that have been expressed separately, with the 76 residue N-domain (8.5 kDa) being monomeric and the 71 residue C-domain forming a hexamer of 49 kDa. Thus they are posing quite different demands on the technique of NMR spectroscopy. Recent methodological approaches developed for the investigation of large proteins (above 25-30 kDa) or protein complexes are discussed in Chapter 4.

To solve the structure of Ph1500, first the structures of the individual domains were solved and then assembled them to form the complete structure (Chapter 7). The structure of the N-domain was derived manually using established standard techniques described in Chapter 4. As predicted from sequence homology, Ph1500-N shows a β-clam fold (Chapter 5), which was previously found in the substrate recognition domains (N-domains) of various AAA proteins (ATPases associated with diverse cellular activities). Nevertheless it contains no AAA-like ATPase domain and therefore represents the first incidence of this fold outside the AAA family. It is assumed that Ph1500-N has a role in protein binding, analogous to the substrate recognition role of homologous AAA protein domains, and that Ph1500 recruits a further factor in DNA repair.

According to secondary structure predictions, the C-terminal domain was predicted to have an OB-fold. It turned out to show a unique twelve bladed propeller fold (Chapter 6), thus being one of the first homo-oligomeric β-propellers and the largest single-ring β-propeller discovered to date. Various approaches along with the one leading up to the final successful resolution are

discussed in Chapter 6.7. By exploiting the hyperthermophilic nature of the protein, it has been possible to overcome unfavourable relaxation properties due to the high molecular mass. An optimization of the measuring temperature with respect to the experiments was performed and high temperature measurements were used as a tool to derive a high-resolution structure.

Investigations on the binding properties of Ph1500 and its relationship to the cradle-loop β-barrel meta-fold are ongoing.

Sequence of *Pyrococcus horikoshii* Ph1500

Ph1500-N

MHHHH HHEGV IMSEL KLKPL PKVEL PPDFV DVIRI KLQGK TVRTG DVIGI
SILGK EVKFK VVQAY PSPLR VEDRT KITLV THP

Ph1500-C

MHHHH HHVDV LEAKI KGIKD VILDE NLIVV ITEEN EVLIF NQNLE ELYRG
KFENL NKVLV RNDLV VIIDE QKLTL IRT

Ph1500

EGVIM SELKL KPLPK VELPP DFVDV IRIKL QGKTV RTGDV IGISI LGKEV
KFKVV QAYPS PLRVE DRTKI TLVTH PVDVL EAKIK GIKDV ILDEN LIVVI
TEENE VLIFN QNLEE LYRGK FENLN KVLVR NDLVV IIDEQ KLTLI RT

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25(17):3389-402

Alva V, Koretke KK, Coles M, Lupas AN (2008) Cradle-loop barrels and the concept of metafolds in protein classification by natural descent. Curr. Opin. Struct. Biol. 18:358-365

Bermel W, Bertini I, Felli IC, Kümmerle R, Pierattelli R (2006) Novel ^{13}C direct ditection experiments, including extension to the third dimension, to perform the complete assignment of proteins. J Magn Reson 178:56-64

Bertini I, Felli IC, Kümmerle R, Moskau D, Pierattelli R (2004) ^{13}C-^{13}C NOESY: An attractive alternative for studying large macromolecules. J. Am. Chem. Soc. 126: 464-465

Bloch F, Hansen WW, Packard M (1946) Nuclear Induction. Phys. Rev., 69:127

Bloom JD, Meyer MM, Meinhold P, Otey CR, MacMillan D, Arnold FH (2005) Evolving strategies for enzyme engineering. Curr. Opin. Struct. Biol. 15:447-452

Bystrov VF (1976) Spin–spin coupling and the conformational states of peptide systems. Prog. Nucl. Magn. Reson. Spectrosc. 10:41–81

Castillo RM, Mizuguchi K, Dhanaraj V, Albert A, Blundell TL, Murzin AG (1999) A six-stranded double-psi beta barrel is shared by several protein superfamilies. Structure 7(2):227-36

Coles M, Diercks T, Liermann J, Groger A, Rockel B, Baumeister W, Koretke KK, Lupas A, Peters J, Kessler H (1999) The solution structure of VAT-N reveals a

'missing link' in the evolution of complex enzymes from a simple βαββ element. Curr Biol 9(20):1158-68

Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J. Biomol. NMR 13:289-302

Davis et al. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Research 35:375-383

DeLaBarre B, Brunger AT (2005) Nucleotide dependent motion and mechanism of action of p97/VCP. J Mol Biol 347(2):437-52

Diercks T, Coles M, Kessler H (1999) An efficient strategy for assignment of cross-peaks in 3D heteronuclear NOESY experiments. J. Biomol NMR 15:177-180

Diercks T, Daniels M, Robert K (2005) Extended flip-back schemes for sensitivity enhancement in multidimensional HSQC-type out-and-back experiments. J. Biomol NMR 33:243-259

Djuranovic S, Rockel B, Lupas AN, Martin J (2006) Characterization of AMA, a new AAA protein from Archaeoglobus and methanogenic archaea. J Struct Biol. 156:130 138

Dougan DA, Mogk A, Zeth K, Turgay K, Bukau B (2002) AAA+ proteins and substrate recognition, it all depends on their partner in crime. FEBS Lett 529(1):6-10

Eijsink, VGH, Bjork A, Gaseidnes S, Sirevag R, Synstad B, van den Burg B, Vriend G (2004) Rational engineering of enzyme stability. J Biotechnol 113:105-120

Eijsink VGH, Gaseidnes S, Borchert TV, van den Burg B (2005) Directed evolution of enzyme stability. Biomol Eng 22:21-30

Eijsink VGH, Vriend G, Van Den Burg B (2001) Engineering a hyperstable enzyme by manipulation of early steps in the unfolding process. Biocatalysis Biotransform 19:443-458

Erdmann R, Wiebel FF, Flessau A, Rytka J, Beyer A, Frohlich KU, Kunau WH (1991) PAS1, a yeast gene required for peroxisome biogenesis, encodes a member of a novel family of putative ATPases. Cell 64(3):499-510

Ernst RR, Anderson WA (1966) Application of Fourier transform spectroscopy to magnetic resonance. Rev. Sci. Instr. 37:93-102

Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, Kay LE (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. Biochemistry 33:5984-6003

Fesik SW, Zuiderweg ERP (1988) Heteronuclear Three-Dimensional NMR Spectroscopy of the Inflammatory Protein C5a. J. Magn. Reson 78:588-93

Gemmecker G. Jahnke W, Kessler H (1993) Measurement of fast proton exchange rates in isotopically labelled compounds J. Am. Chem. Soc. 115:11620-21

Gerega A, Rockel B, Peters J, Tamura T, Baumeister W, Zwickl P (2005) VAT, the thermoplasma homolog of mammalian p97/VCP, is an N domain-regulated protein unfoldase. J. Biol. Chem. 280(52):42856-62

Gillespie JR, Shortle D (1997) Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels. J. Mol. Biol. 268:158-169

Gillespie JR, Shortle D (1997) Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. J. Mol. Biol. 268:170-184

Ginzinger SW, Fischer J (2006) SimShift: Identifying structural similarities from NMR chemical shifts. Bioinformatics 22(4):460-465

Golbik R, Lupas AN, Koretke KK, Baumeister W, Peters J (1999) The Janus face of the archaeal Cdc48/p97 homologue VAT: protein folding versus unfolding. Biol Chem 380(9):1049-62

Golovanov AP, Hautbergue GM, Wilson SA, Lian L-Y (2004) A Simple Method for Improving Protein Solubility and Long-Term Stability. *J. Am. Chem. Soc.* 126:8933-8939

Goto NK, Gardner KH, Mueller GA, Willis RC, Kay LE (1999) A robust and cost-effective method for the production of Val, Leu, Ile (δ1) methyl-protonated ^{15}N-, ^{13}C-, ^{2}H-labelled proteins. J. Biomol. NMR 13:369

Griesinger C, Sorensen OW, Ernst RR (1987) A practical approach to three-dimensional NMR spectroscopy. J. Magn. Reson 73:574-579

Griesinger C, Sorensen OW, Ernst RR (1987) Novel Three-Dimensional NMR Techniques for Studies of Peptides and Biological Macromolecules. J. Am. Chem. Soc 109:7227-28

Grishin NV (2001) Fold change in evolution of protein structures. J Struct Biol 134(2-3):167-185

Grzesiek S, Wingfield P, Stahl S, Kaufmann JD, Bax A (1995) Four-dimensional ^{15}N-separated NOESY of slowly tumbling perdeuterated ^{15}N-enriched proteins. J. Am. Chem. Soc. 117:9594-95

Güntert P (1998) Structure calculation of biological macromolecules from NMR data. Q Rev Biophys. 31(2):145-237

Günthert P (2004) Automated NMR Structure Calculation With CYANA. Meth. Mol. Biol. 278:353-378

Hu K, Vögeli B, Clore GM (2006) ^{13}C-detected HN(CA)C and HMCMC experiments using a single methyl-reprotonated sample for unambiguous methyl resonance assignment. J Biomol NMR 36:259-266

Hu K, Vögeli B, Pervushin K (2005) Side-chain H and C resonance assignment in protonated/partially deuterated proteins using an improved 3D ^{13}C-detected HCC-TOCSY. J. Magn. Res. 174(2):200-208

Jeener J (1971) lecture given at the Ampere Summer School in Basko Polje, Yugoslavia: Multidimensional Fourier NMR spectroscopy and imaging (today known as COSY), later published in Les editions de physique (1994) "NMR and More in Honour of Anatole Abragam", Eds. M. Goldman and M. Porneuf.

Jeener J, Meier BH, Bachmann P, Ernst RR (1979) Investigation of exchange process by two-dimensional NMR spectroscopy. J. Chem. Phys. 71:4546-53

Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Ono AM, Guentert P (2006) Optimal isotope labelling for NMR protein structure determinations. Nature 440:52-57

Kumar A, Ernst RR, Wüthrich K (1980) A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the eclucidation of complete proton–proton cross-relaxation networks in biological macromolecules. Biochem. Biophys. Res. Comm. 95 :1–6

Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst 26:283-291

Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thronton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8(4):477-486

Lehmann M, Kostrewa D, Wyss M, Brugger R, D'Arcy A, Pasamontes L, Van Loon APGM (2000) From DNA sequence to improved functionality: using protein sequence comparisons to rapidly design a thermostable consensus phytase. Protein Eng. 13:49-57

Lehmann M, Pasamontes L, Lassen SF, Wyss M (2000) The consensus concept for thermostability engineering of proteins. Biochim. Biophys. Acta 1543:408-415

Lehmann M, Wyss M (2001) Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. Curr. Opin. Biotechnol. 12:371-375

LeMaster DM, Richards FM (1988) NMR sequential assignment of Escherichia coli thioredoxin utilizing random fractional deuteration. Biochemistry 27(1):142-150

Leutner M, Gschwind RM, Liermann J, Schwarz C, Gemmecker G, Kessler H (1998) Automated backbone assignment of labelled proteins using the threshold accepting algorithm. J. Biomol. NMR 11:31-43

Lupas AN, Koretke KK, Evolution of Protein Folds, In *Computational Structural Biology*, edited by Peitsch M, Schwede T, *World Scientific Publishing Co*, 2008.

Matzapetakis M, Turano P, Theil EC, Bertini I, (2007) ^{13}C-^{13}C NOESY spectra of a 480 kDa protein: solution NMR of ferritin. J Biomol NMR 38:237-242

McIntosh LP, Dahlquist FW (1990) Biosynthetic incorporation of ^{15}N and ^{13}C for assignment and interpretation of nuclear magnetic resonance spectra of proteins. Q Rev Biophys. 23(1):1–38

Nabuurs SB, Spronk CA, Krieger E, Maassen H, Vriend G, Vuister GW (2003) Quantitative evaluation of experimental NMR restraints. J. Am. Chem. Soc. 125(39):12026-34

Neal S, Berjanskii M, Zhang H, Wishart DS (2006) Accurate prediction of protein torsion angles using chemical shifts and sequence homology. Magn Reson Chem. 44:58-67

Neer EJ, Schmidt CJ, Nambudripad R, Smith TF (1994) The ancient regulatory-protein family of WD-repeat proteins. Nature 371:297-300

Nikolova PV, Henckel J, Lane DP, Fersht AR (1998) Semirational design of active tumor suppressor p53 DNA binding domain with enhanced stability. Proc. Natl. Acad. Sci. USA 95:14675-14680

Overbeek et al. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res 33(17):5691-702

Pervushin K, Riek R, Wider G, Wüthrich K (1997) Attenuated T_2 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. USA 94:12366-12371

Pervushin K, Vögeli B, Eletsky A (2002) Longitudinal ^{1}H Relaxation Optimization in TROSY NMR Spectroscopy. J. Am. Chem. Soc. 124:12898-902

Purcell EM, Torrey HC, Pound RV (1946) Resonance absorption by nuclear magnetic moments in a solid. Phys. Rev. 69:37-38

Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. J. Mol. Biol. 7:95–99

Riek R, Wider G, Pervushin K, Wüthrich K (1999) Polarization transfer by cross-correlated relaxation in solution NMR with very large molecules. Proc. Natl. Acad. Sci. USA 96:4918–4923

Rieping W, Habeck M, Bardiaux B, Bernard A, Malliavin TE, Nilges M (2007) ARIA2: automated NOE assignment and data integration in NMR structure calculation. Bioinformatics 23:381-382

Rockel B, Walz J, Hegerl R, Peters J, Typke D, Baumeister W (1999) Structure of VAT, a CDC48/p97 ATPase homologue from the archaeon Thermoplasma acidophilum as studied by electron tomography. FEBS Lett 451(1):27-32

Rosen MK, Gardner KH, Willis RC, Parris WE, Pawson T, Kay LE (1996) Selective Methyl Group Protonation of Perdeuterated Proteins. J. Mol. Biol. 263:627-636

Schreiber SL, Verdine GL (1991) Protein Overproduction for Organic Chemists. Tetrahedron 47:2543-2562

Singh SK, Guo F, Maurizi MR (1999) ClpA and ClpP remain associated during multiple rounds of ATP-dependent protein degradation by ClpAP protease. Biochemistry 38(45):14906-15

Smith TF, Gaitatzes C, Saxena K, Neer EJ (1999) The WD repeat: a common architecture for diverse functions. Trends Biochem Sci 24:181-185

Soeding J (2005) Protein homology detection by HMM-HMM comparison. Bioinformatics 21(7):951-960

Thayer MM, Ahern H, Xing D, Cunningham RP, Tainer JA (1995) Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure. Embo J 14(16):4108-20

Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH (1995) Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. Proc. Natl. Acad. Sci. USA 92(20):9279-83

Truffault V, Coles M, Diercks T, Abelmann K, Eberhardt S, Lüttgen H, Bacher A, Kessler H (2001) The solution structure of the N-terminal domain of Riboflavin Synthase. J. Mol. Biol 309:949-960

Tugarinov V, Kay LE (2004) An isotope labelling strategy for methyl TROSY spectroscopy. J. Biomol. NMR 28:165-172

Tugarinov V, Sprangers R, Kay LE (2004) Line narrowing in methyl-TROSY using zero-quantum ^{1}H-^{13}C NMR spectroscopy. J. Am. Chem. Soc. 126:4921-4925

Tugarinov V, Kay LE (2005) Methyl groups as probes of structure and dynamics in NMR studies of high-molecular-weight proteins. ChemBioChem 6:1567-1577

Vriend G. (1990) WHAT IF: A molecular modeling and drug design program. J. Mol. Graph. 8:52-56

Walker JE, Saraste M, Runswick MJ, Gay NJ (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATPrequiring enzymes and a common nucleotide binding fold. Embo J 1(8):945-51

Wand AJ, Ehrhardt MR, Flynn PF (1998) High-resolution NMR of encapsulated proteins dissolved in low-viscosity fluids. Proc. Natl. Acad. Sci. USA 95(26):15299-15302

Wang Q, Buckle AM, Fersht AR (2000) Stabilization of GroEL minichaperones by core and surface mutations. J. Mol. Biol. 298:917-926

Wang Q, Buckle AM, Foster NW, Johnson CM, Fersht AR (1999) Design of highly stable functional GroEL minichaperones. Protein Science 8:2186-2193

Wishart DS, Sykes BD, Richards FM (1992) The Chemical-Shift-Index – A fast and simple method for the assignment of protein secondary structure through NMR-spectroscopy. Biochemistry 31:1647-51

Wishard DS, Sykes BD (1994) Chemical-shifts as a tool for structure determination. Methods Enzymol. 239:363-392

Wishard DS, Sykes BD (1994) The C-13 Chemical-Shift-Index – A simple method for the identification of protein secondary structure using C-13 chemical shift data. J. Biomol. NMR 4:171-180

Wüthrich K, Wider G, Wagner G, Braun W (1982) Sequential resonance assignments as a basis for determination of spatial protein structures by high resolution proton nuclear magnetic resonance. J. Mol. Biol. 155:311–319

Wüthrich K (1986) NMR of Proteins and Nucleic Acids. Wiley-Interscience, New York.

Wüthrich K (2001) The way to NMR structures of proteins. Nat. Struct. Biol. 8:923-925

Xu R, Ayers B, Cowburn D, Muir TW (1999) Chemical ligation of folded recombinant proteins: Segmental isotopic labelling of domains for NMR studies. Proc. Natl. Acad. Sci. USA 96(2):388-393

Yang D, Zheng Y, Liu D, Wyss DF (2004) Sequence-specific assignments of methyl groups in high-molecular weight proteins. J. Am. Chem. Soc. 126:3710-3711

Zhang O, Forman-Kay JD (1997) NMR studies of unfolded states of an SH3 domain in aqueous solution and denaturing conditions. Biochemistry 36:3959-70

Zhang X, Shaw A, Bates P, Newman R, Gowen B, Orlova E, Gorman M, Kondo H, Dokurno P, Lally J (2000) Structure of the AAA ATPase p97. Molecular Cell 6:1473-1484

Contributions to International Conferences and Meetings

Posterpresentation: *Structure determination of Ph1500*, I. Varnay, S. Djuranovic, V. Truffault, A. Lupas, M. Coles, H. Kessler, 22nd International Conference on Magnetic Resonance in Biological Systems, Goettingen, Germany, August 20-25, 2006.

Posterpresentation: *Structure determination of a 100 kDa homohexamer*, I. Varnay, S. Djuranovic, V. Truffault, A. Lupas, M. Coles, H. Kessler, 48th Experimental NMR Conference, Daytona Beach, Forida, USA, April 22-27, 2007.

Posterpresentation: *Structure determination of a 100 kDa homohexamer*, I. Varnay, S. Djuranovic, V. Truffault, A. Lupas, M. Coles, H. Kessler, EUROMAR 2007, Magnetic Resonance Conference, Tarragona, Spain, July 01-05, 2007.

Posterpresentation: *Structure determination of a 100 kDa homohexamer*, I. Varnay, S. Djuranovic, V. Truffault, A. Lupas, M. Coles, H. Kessler, 29th Annual Discussion Meeting, GDCh, Magnetic Resonance in Biophysical Chemistry, Goettingen, Germany, September 26-29, 2007.

Posterpresentation: *The solution structure of the 100kDa homohexameric protein Ph1500 reveals a β–clam attached to a twelve-bladed β-propeller*, I. Varnay, S. Djuranovic, A. Ursinus, J. Martin, V. Truffault, B. Roeckel, M. Coles, A. Lupas, H. Kessler, NMR-Life, Advances in NMR of protein and nucleic acid molecular recognition, Murnau, October 16-18, 2008.