

Dissecting the Mechanism of Protein Unfolding by the Unfoldase PAN using Structural Biology

Von der Naturwissenschaftlichen Fakultät der
Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation
von
Georg Ernst Heinrich Krüger, M. Sc.

2021

Referentin: Prof. Dr. ric. Teresa Carlomagno

Korreferent: Prof. Dr. rer. nat. Dietmar J. Manstein

Tag der Promotion: 17.06.2021

Abstract

The proteasome as the major cellular protease plays a central role in regulated protein degradation in all kingdoms of life. Besides the proteolytic core particle, the proteasome comprises an ATP-dependent unfoldase subcomplex from the family of AAA+ ATPases. Numerous structural, biochemical and biophysical studies have provided insights into the mechanism of ATP-dependent protein unfolding by hexameric AAA+ unfoldase from bacteria, archaea and eukaryotes, resulting in the suggestion of a common mechanism for how all AAA+ unfoldases couple ATP-hydrolysis to protein unfolding. However, conflicting biochemical, biophysical and structural data on unfoldases from various organisms challenge the universal validity of this mechanism, calling for further studies on unfoldases from all kingdoms of life. The proteasomal ATPase PAN from archaea is considered to be an evolutionary precursor to its eukaryotic homolog, the 19S regulatory particle, and its much simpler architecture makes it an excellent system for structural and mechanistic studies in solution. In this PhD thesis I employed solution-state NMR spectroscopy to study the conformational states of PAN and gain insights into the unfolding reaction of the model substrate GFP-ssrA by PAN.

After assigning NMR resonances of all subdomains and of the full complex, I could confirm a functional asymmetry in hexameric PAN particles. Asymmetry has also been found in most of the recent cryo-EM models from various AAA+ ATPases; however, it has been located mostly to the hexameric ring formed by the nucleotide-binding domain of the ATPases. By contrast, my studies show conformational heterogeneity and asymmetry in the so-called OB-ring which is located above the nucleotide-binding domain and is thought to allosterically link the nucleotide state of PAN to substrate recruitment. Further, studies on the binding of PAN to an ssrA-tagged model substrate demonstrate the absence of strong, stable interactions between the PAN N-terminal domains and its substrates, in disagreement with what has been previously proposed.

Further, I used time-resolved NMR experiments to study the dynamic process of unfolding of the model substrate GFP-ssrA. Detecting selectively the NMR-signals of isotopically labeled GFP-ssrA, I was able to monitor the states experienced by the substrate during the unfolding and degradation reaction by PAN at amino acid resolution. The disappearance of natively folded GFP signals and the concomitant build-up of unstructured proteolysis-products in the NMR experiments with similar time-constants confirm a tight coupling of the two processes and are in good agreement with findings of a collaborative study using time-resolved small angle neutron scattering.¹

These studies demonstrate the power of NMR spectroscopy in helping to understand dynamic processes like protein unfolding and emphasize its complementarity to other solution-based techniques like small-angle scattering which probe biological samples in a more native-like environment and thus allow dynamic processes to take place and be monitored.

Keywords: protein unfolding, proteasome, AAA+ ATPases, solution NMR spectroscopy, time-resolved NMR, structural biology

Outline

Abstract	iv
List of Abbreviations.....	vii
1 Introduction.....	1
1.1 The Ubiquitin proteasome system.....	1
1.1.1 The eukaryotic 26S proteasome	2
1.1.2 The 20S proteolytic core particle.....	4
1.2 Archaeal proteasome systems	5
1.2.1 PAN - an archaeal regulatory particle homologue	5
1.2.2 The archaeal 20S proteasome	7
1.2.3 Targeted protein degradation in archaea	7
1.3 Structural properties of AAA+ ATPases.....	8
1.4 The diverse architecture of protease associated AAA+ ATPases	11
1.5 Structure and function of proteasomal ATPases	13
1.6 A conserved mechanism of translocating ATPases	25
2 Methodological Background.....	27
2.1 NMR spectroscopy	27
2.1.1 Basic principles of NMR spectroscopy	27
2.1.2 NMR of large protein complexes	32
2.1.3 NMR assignment strategies for proteins	34
2.1.4 Protein dynamics by NMR spectroscopy.....	37
2.1.5 SOFAST-NMR for Real-time NMR studies.....	42
2.2 SEC-MALS.....	44
3 Aim of this Thesis Work.....	46
4 Materials and Methods	47
4.1 Materials	47
4.1.1 Instruments.....	47
4.1.2 Software	48
4.1.3 Chemicals, Enzymes and Consumables	48
4.1.4 Media, Supplements and Buffers	50
4.1.5 Plasmids and Oligonucleotides	51
4.1.6 Bacterial Strains and competent cells	54
4.2 Cloning and Mutagenesis.....	55
4.2.1 Cloning	55
4.2.2 Site-directed mutagenesis.....	56
4.3 Protein expression and purification	57
4.3.1 Recombinant protein expression in LB medium	57
4.3.2 Recombinant expression of isotopically labelled protein and deuteration	57
4.3.3 Protein purification	58
4.4 Biochemical assays.....	64
4.4.1 Fluorescence based GFP-unfolding assays.....	64
4.5 Bioanalytical methods.....	64
4.5.1 Analytical SEC and SEC-MALS	64

4.5.2	ESI-protein mass spectrometry	65
4.6	NMR spectroscopy	65
4.6.1	Resonance assignments.....	65
4.6.2	Backbone dynamics experiments on PAN's coiled coil domain	68
4.6.3	NMR-titrations with the ssrA-peptide.....	68
4.6.4	PAN/PAN-GFP CPMG relaxation dispersion experiments	69
4.6.5	Time-resolved NMR-experiments on the unfolding and degradation of GFP-ssrA	69
4.6.6	Data analysis of NMR-time series	70
5	Results	73
5.1	NMR studies on the archaeal PAN-20S proteasome system	73
5.1.1	PAN's coiled coil domain.....	74
5.1.2	The OB-ring is a hexamer with up to 6 conformations.....	79
5.1.3	ATPase domain	82
5.1.4	Studies on functionally working full length PAN.....	84
5.1.5	Characterization of PAN-substrate interactions by NMR-spectroscopy	97
5.2	Real-time NMR studies on the processing of a model substrate by the archaeal PAN-20S proteasome.....	102
5.2.1	Initial studies on the processing of GFP-ssrA substrates by the PAN-20S proteasome system	102
5.2.2	GFPuv-ssrA is not a suitable fluorescent protein for NMR-studies on isotopically enriched samples	104
5.2.3	A folding reporter GFP variant as a substrate for the PAN-20S proteasome	106
5.2.4	Real-time NMR studies on the unfolding and degradation of GFP by the PAN-20S proteasome system.....	107
5.2.5	Resonance-assignments of different GFP variants	117
6	Discussion, Conclusions and Outlook.....	119
6.1	NMR studies on the archaeal proteasome activator PAN	119
6.1.1	PAN-substrate interactions	123
6.2	Time-resolved NMR-studies on the processing of a model substrate by the PAN-20S proteasome.....	125
7	Bibliography.....	130
Appendix.....		vii
Additional figures		vii
Amino acid sequences.....		viii
Resonance assignments.....		ix
Acknowledgements.....		xxxiii
Curriculum Vitae Georg Krüger.....		xxxiv
Contributions to scientific publications		xxxvi

List of Abbreviations

Aa	amino acid
a. u.	Arbitrary units
ATP	Adenosintriphosphate
ADP	Adenosindiphosphate
BSA	Bovine serum albumin
β -ME	β -mercaptoethanol
CC	coiled coil
CP	core particle
Cryo-EM	Cryogenic-electron microscopy
CV	Column volume
kDa	kilodalton
dH ₂ O	distilled water
ddH ₂ O	Millipore filtered distilled water
DF	Degree of freedom
DTT	Dithiothreitol
ESI	Electro spray ionisation
frGFP-ssrA	folding reporter GFP C-terminally ssrA-tagged
fw	forward
<i>g</i>	<i>g</i> -force
GFP	Green fluorescent protein
HDX	Hydrogen-Deuterium Exchange
HMQC	Heteronuclear multiple quantum coherence
HSQC	Heteronuclear single quantum coherence
IM-labelled	Isoleucin δ 1 and methionine ϵ –methyl groups ¹ H- and ¹³ C-labelled in an otherwise perdeuterated background
IEX	Ion exchange chromatography
IPTG	
FPLC	Fast Protein Liquid Chromatography
LB	Lysogeny broth
MALS	Multi Angle Light Scattering
MS	Mass spectrometry
MES	2-(<i>N</i> -morpholino) ethanesulfonic acid
MWCO	Molecular weight cut-off
NMR	Nuclear magnet resonance
OD ₆₀₀	Optical density of cell cultures measured as absorption at 600 nm
PAN	Proteasome activating nucleotidase
PCR	Polymerase chain reaction
RP	Regulatory particle
rpm	Revolutions per minute
rt	Retention time
rv	reverse
S	Svedberg (unit)
SANS	Small-angle neutron scattering
SEC	Size exclusion chromatography
SDS-PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
SS	Sum of squares
TRIS	Tris(hydroxymethyl)aminomethane
TROSY	Transverse relaxation optimized spectroscopy
UBS	Ubiquitin-proteasome system
Ub	Ubiquitin

1 Introduction

1.1 The Ubiquitin proteasome system

The composition of the proteome as the universal ensemble of all proteins contained in all living cells is delicately regulated on two major levels to match the requirements of the cell under a broad range of conditions²: (1) Gene regulation of protein-encoding genes at transcriptional and translational level controls the rate at which certain proteins are synthesized by the cell's central protein factory, the ribosome.^{3,4} (2) After ribosomal protein biosynthesis two major pathways in eukaryotes remove superfluous proteins, when they are no longer needed, are misfolded or damaged by different kinds of cellular stress. Next to the autophagal-lysosomal system, which involves hydrolytic degradation of proteins in the compartmentalized lysosome,⁵ the Ubiquitin-proteasome system (UPS) is the major pathway for cellular protein removal in eukaryotes.⁶ Targeted protein degradation is crucial for cellular health since the accumulation of needless or damaged proteins can lead to highly cytotoxic aggregates with disastrous consequences not only for the cell but also the entire organism.^{7,8} Furthermore, the removal of regulatory proteins after their tasks has been fulfilled is critical to properly maintain a multitude of regulated processes like cell cycle progression, cell survival, proliferation and apoptosis, which could otherwise lead to unregulated cell growth and consequently cancer.^{9,10} These key roles make the entire UPS a valuable drug target, as testified by the development of a multitude of small-molecule inhibitors targeting several distinct components of this pathway.¹¹

The UPS activity pathway can be divided into a regulatory step, which marks proteins for degradation, and the actual proteolytic cleavage by the proteasome (Figure 1.1 A). The UPS degrades only proteins which are targeted for removal, to avoid unselective and harmful proteolysis. The first mark of proteins that need to be degraded is their covalent ligation to the 76-amino-acid protein ubiquitin (Ub),¹² is recognised by the proteasome. Ubiquitin is attached to lysine residues of the target protein via its C-terminal glycine in an ATP dependent reaction catalyzed by an enzyme-cascade E1, E2 and E3 (Figure 1.1A).¹³ The first step of this conjugation is the ATP-dependent activation of Ub through the formation of a thioester linkage to the enzyme E1. Next, the activated Ub is transferred to the ubiquitin carrier protein E2, again forming a thioester linkage. The third step is catalyzed by the ubiquitin ligase E3, which attaches the Ub C-terminus to a lysine ϵ -amino group of the substrate via an isopeptide linkage, resulting in monoubiquitylation. Generally, substrates can be linked to Ub via multiple lysine residues resulting in multimonoubiquitylation. Ub itself in turn also contains 7 lysine residues (Figure 1.1 B) which together with the N-terminal methionine are potential sites for ubiquitylation providing the basis for a potentially complex ubiquitylation pattern also referred to as the ubiquitin code.¹² The 500-1000 different E3 ligases in humans¹⁴ are mostly responsible for the substrate specificity of ubiquitylation by bringing the Ub-E2 conjugate together with specific substrate proteins. Since ubiquitylation is not exclusive to the UPS pathway, the specific pattern of ubiquitylation determines whether ubiquitylated proteins are targeted for proteasomal degradation or directed to other regulatory pathways. The 26S proteasome as the major protease in eukaryotic cells, mainly accepts substrates decorated with Lys48 or Lys11 linked poly-UB chains.¹⁵ Other variants may also be accepted. However,

Lys63 linked Ub-chains target substrates almost exclusively for autophagal-lysosomal hydrolysis and do not engage with the proteasome.¹⁶

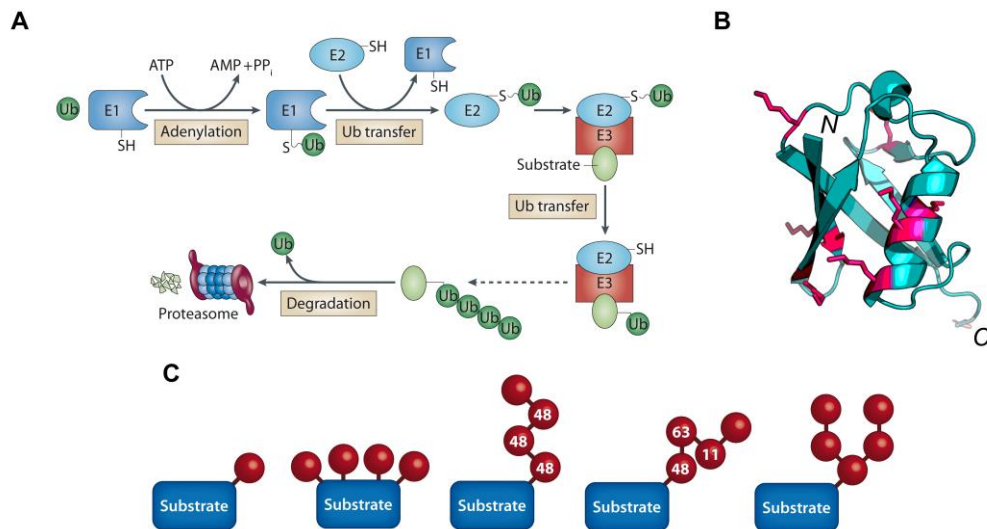


Figure 1.1 Overview of the Ubiquitin proteasome system A) The Ubiquitin proteasome pathway comprises several key steps. First, ubiquitin (Ub) is C-terminally linked to E1 in via an adenylated intermediate in an ATP dependent manner forming a thioester-linkage. Next, UB is transferred to E2, again forming a thioester-linkage. The Ub-E2 conjugate is recognized by substrate specific E3-ligases which catalyze the isopeptide linkage formation of the Ub-C-terminus to a substrate-lysine residue. Finally, ubiquitylated protein substrates are recognized by the 26S proteasome via their Ub-chains, ubiquitin is cleaved off and substrate proteins are degraded in an ATP-dependent manner. B) The ubiquitin structure (PDB: 1UBQ)¹⁷ exposes a flexible carboxy-terminal tail (C) for E1-E3 dependent ligation to substrates and seven lysine residues (red) for attachment of further ubiquitin molecules via isopeptide linkages. C) Different patterns of ubiquitylation are found (from left to right): Monoubiquitylation, multimonoubiquitylation, polyubiquitylation, mixed ubiquitin chains and branched ubiquitin chains. Part A is adapted from MAUPIN-FURLOW et al.¹⁸, while schematics in C are adapted from KOMANDER et al.¹²

1.1.1 The eukaryotic 26S proteasome

In eukaryotes the 26S proteasome is the main protease involved in regulated protein degradation. The enzyme removes harmful or no longer needed proteins as well as facilitates the recycling of basic amino acid building blocks and is thus essential for maintaining proteome homeostasis. As the final protease, degrading ubiquitylated proteins, the proteasome represents the terminal component of the ubiquitin proteasome system. The 2.5 MDa complex consists of two major subcomplexes comprising more than 30 subunits, with distinct functions in the processing of substrates (Figure 1.2): (1) The 19S regulatory particle (RP), which is responsible for recognition, deubiquitylation and ATP-dependent unfolding of ubiquitylated substrates and (2) a proteolytic core particle (CP) for proteolytic fragmentation. The strict compartmentalization of this destructive molecular machine is crucial, since especially the CP's proteolytic sites would be harmful to the cellular environment if not shielded from the cytosol. The structure and functions of the proteasome's individual subcomplexes will be briefly reviewed in the following sections.

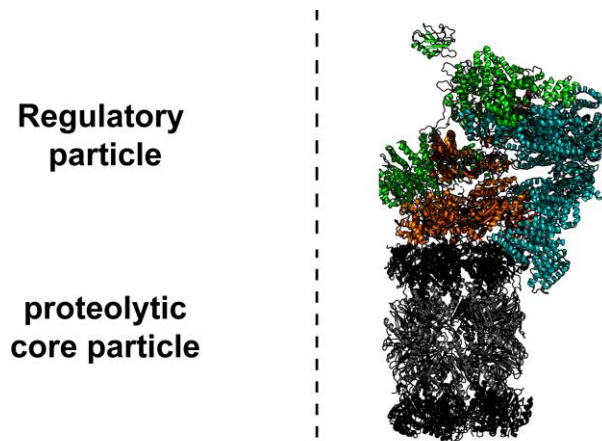


Figure 1.2 Architecture of the 26S proteasome. The full assembly of the eukaryotic 26S proteasome comprises the 19S regulatory particle (orange, green, cyan) and the 20S proteolytic core particle (greys). Shown is a singly-capped version of the yeast 26S proteasome (PDB: 6FVT).¹⁹

1.1.1.1 The 19S regulatory particle

As the name suggests the 19S regulatory particle regulates the access of substrates to the proteasome for proteolytic degradation. The RP forms a sophisticated entrance control to prevent unregulated proteolysis. The process of substrate admittance to the 20S CP can be divided into substrate-recognition and -deubiquitylation with subsequent unfolding and translocation. These events happen on two specialized subcomplexes of the RP: (1) the base complex resides directly on the CP and unfolds and translocates the substrate in an ATP depended manner and (2) the lid flanks one face of the RP and, acts as a structural scaffold and comprises the mayor deubiquitinase Rpn11 (Figure 1.3).

A heterohexameric AAA+ ATPase ring, formed by Rpt1-6 and located at the core of the RP-base, is the proteasome's motor converting the chemical energy of ATP into mechanical work. The ATPase ring's architecture can be described as a trimer of dimers, in which N-terminal helices of neighbouring subunits (Rpt1/2, Rpt6/3, Rpt4/5) form three coiled coil structures (CC)²⁰. The CC domain is followed by a ring assembly of OB-fold containing domains (OB-ring). A tight pore formed by the OB-ring poses the first constriction for folded substrates during their way through the proteasome. The mechanical work is generated by the six ATPase domains, which apply force to the substrates via their conserved pore-loops. Interaction to the proteolytic core particle is mediated by a conserved hydrophobic-tyrosine-X (HbYX) motif on Rpt2, Rpt3 and Rpt5 which induce gate opening in the CP's α -ring. Additionally, three non-ATPase proteins, named Rpn1, Rpn2 and Rpn13, accompany the AAA-ring in the base subcomplex. These proteins exhibit multiple binding sites for Ub or Ub-like proteins (UBL). An additional ubiquitin receptor Rpn10 is neither considered part of the base nor lid but rather bridges the two in the fully assembled RP. The lid acts as a scaffold keeping all components of the RP together and also keeping it attached to the core particle. It consists of 7 structural protein subunits (Rpn3, Rpn5, Rpn6, Rpn7, Rpn8, Rpn9, and Rpn12) and a Zn²⁺-dependent deubiquitinase (DUB) of the JAMM/MPN family of peptidases, called Rpn11. Rpn11 is mainly responsible for cleavage of ubiquitin chains attached to proteasomal substrates before they are translocated through the ATPase- ring's pore. The RP is further associated with two more DUBs (Ubp6 and Uch37); these have more specific Ub cleaving activities than Rpn11, which mostly cleaves the substrate bound Ub-chains.²¹ During the active cycle of proteasomal processing the entire RP undergoes substantial conformational

changes to facilitate the recognition of Ub-linked substrates, cotranslocational deubiquitylation, unfolding and translocation into the proteolytic core particle. The conformation adopted by the RP before engaging with a substrate is called S_1 and is used to illustrate the RP-architecture in Figure 1.3. Mechanistic models of the transition between these conformations are discussed in detail in chapter 1.5 and 1.6.

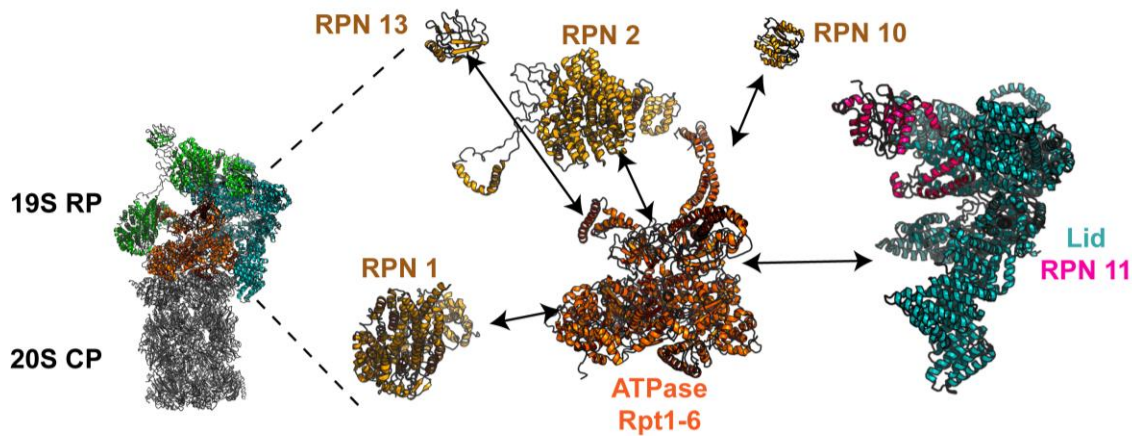


Figure 1.3 Architecture of the eukaryotic 19S regulatory particle. The RP consists of two major sub-assemblies: The heterohexameric AAA+ ATPase motor, comprising the six subunits Rpt1-6 (orange), is located at the base. The ubiquitin-receptors RPN1, RPN2 and RPN13 are also considered part of the base (orange). A lid structure comprising seven structural proteins (cyan) acts as a scaffold and also contacts the proteolytic core particle to maintain integrity of the full 26S proteasome. An additionally essential component of the lid is the Zn^{2+} dependent deubiquitinase RPN11 (pink), which removes polyubiquitin chains from proteasomal substrates. Figures are made based on a yeast proteasome cryo-EM model in the S_1 state (PDB: 6FVT).¹⁹

1.1.2 The 20S proteolytic core particle

The 20 S proteolytic core particle (20S CP), which is the catalytic subunit of the proteasome, exhibits a much simpler architecture (Figure 1.4).²² It comprises four stacked rings, which are arranged in a $\alpha_{1-7}\beta_{1-7}\beta_{1-7}\alpha_{1-7}$ fashion and form a hollow cylinder. Each of the rings is a heteroheptameric assembly of α_1 - α_7 and β_1 - β_7 , respectively. Inside the cylinder the 20S core particle is compartmentalized into a pair of distal ante-chambers and a central proteolytic core chamber. The proteolytic sites reside in the core chamber formed by the two inner β -rings. Only β_1 , β_2 and β_5 show proteolytic activity. They are unique threonine proteases, which are activated autocatalytically from propeptides upon full 20S assembly, to prevent uncontrolled activity and cell damage.²³ Two α -rings sandwich the two inner β -rings, providing a tight barrier for peptide entry to the proteolytic lumen. Upon insertion of the C-terminal tails of the regulatory particle ATPases the gates at the distal faces open up to allow for substrate entry into the protease chamber.¹⁹ The non-proteolytic ante-chamber has been found to maintain proteins in their unfolded state to allow for efficient proteolysis by actively interacting with unfolded polypeptides²⁴.

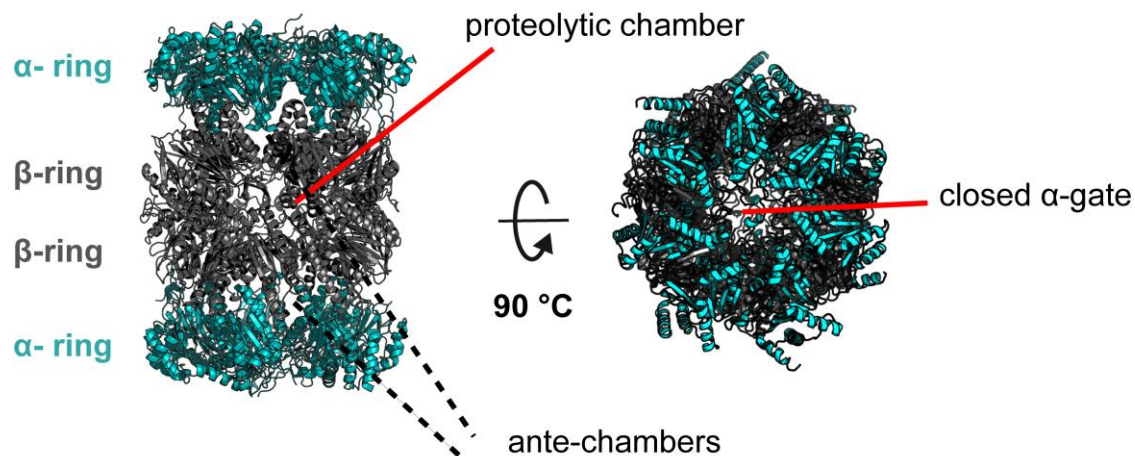


Figure 1.4 Architecture of the 20S proteolytic core particle. The eukaryotic 20S CP is assembled as cylinder composed of four stacked rings. Two central heteroheptameric β -rings are sandwiched between a pair of likewise heteroheptameric α -rings. The six threonine protease proteolytic sites (three on each β -ring) are sequestered in the core chamber of the CP. Two ante-chambers are formed between α - and β -rings and maintain polypeptides in an unfolded state to facilitate efficient proteolysis.²⁴ The α -gate entrance pore of the CP is closed in the S_1 state of the 26S proteasome. Structures represent the cryo-EM models from *S. cerevisiae* in the S_1 state (PDB: 6FVT).¹⁹

1.2 Archaeal proteasome systems

Archaea, as single cell organisms, are considered to be prokaryotes. They lack characteristic features of eukaryotes, like nuclei and other subcellular organelles. However, they also differ from bacteria in that they show unique properties similar to eukaryotes. The fact that many molecular machines of archaea are similar to those from eukaryotes and evolutionary relatively well conserved is exploited by molecular and structural biologists. Oftentimes archaeal systems are less complex homologues of their eukaryotic counterparts, making them easier to study.²⁵ Furthermore, the thermal stability of proteins from thermophilic archaea is of particular interest since these proteins are usually better behaved in *in vitro* studies.²⁶ Especially, NMR spectroscopy benefits from elevated temperatures tolerated by thermophilic proteins, since higher temperatures result in smaller correlation times τ_c , improving drastically the quality of NMR-spectra of larger proteins.^{25,27}

In the case of the archaeal proteasome, scientists have exploited all the above described advantages: The first high resolution crystal structure of a 20S proteasome core particle was published in 1995 and described the complex from the archaeon *Thermoplasma acidophilum*. Later on, archaeal proteasomes have contributed extensively to build the basis for the mechanistic understanding of these molecular machines.²⁸

The general architecture of archaeal proteasome systems is homologous to the eukaryotic ones; however, several important properties described in the following sections make them more accessible for biophysical and structural studies.

1.2.1 PAN - an archaeal regulatory particle homologue

PAN, the Proteasome Activating Nucleotidase was found to be the archaeal homolog of the eukaryotic 19S regulatory particle.²⁹ The most striking difference between the eukaryotic and archaeal proteasome activator is the absence of an equivalent of the lid and as well as of a

ubiquitin- or ubiquitin-homolog- processing machinery. Essentially, PAN is the homologue of the hexameric ATPase ring of the 19S particle without all regulatory elements. In contrast to the eukaryotic system, PAN is a homo-hexameric assembly formed by six copies of the same polypeptide chain.³⁰ However, the PAN amino acids sequence has 41-45 % sequence similarity to the Rpts of the 19S regulatory particle; consequently the archaeal PAN is considered to be the evolutionary precursor of the eukaryotic 19S RP. First high resolution structural information on proteasomal activators was obtained on PAN from the archaeon *Methanocaldococcus jannaschii* in 2009,^{31,32} the architecture of which was later found to be fairly conserved in the eukaryotic system: Likewise the 19S regulatory particle's ATPase PAN has three N-terminal coiled coils formed by two neighbouring subunits. To facilitate the formation of the CC from two adjacent N-terminal helices, a conserved trans-proline peptide bond is found in every other protomer around the ring, establishing the trimer of asymmetric dimer architecture for the homo-hexameric PAN. While the CCs in the 19S RP are involved in binding to several ubiquitin receptors and lid structural elements (see also chapter Figure 1.3), their function in archaea is not fully understood. Especially the role of a supposedly unstructured tail of ~ 40-50 residues upstream of the coiled coil forming region remains elusive. C-terminal to the CC, there is an OB-ring, which forms a 13 Å pore as a first constriction for folded substrates during the degradation.³¹ Mechanical work is generated by the C-terminal ATPase domain, which in PAN carries an HbYX motif for 20S interaction and α -gate opening on all six protomers.³³ PAN was found to process substrates with long flexible tails in both C- to N- and N- to C-terminal directions.³⁴ Further, it has been reported that PAN exhibits chaperone activity, both preventing the aggregation of unfolded proteins and promoting the refolding of unfolded proteins,³⁵ potentially acting as a therapeutic in protein-misfolding neurodegenerative diseases.³⁶ Due to its simple architecture, PAN has been studied extensively as a model system for proteasomal ATPases, establishing key properties which are conserved in the eukaryotic system, like asymmetric binding of nucleotides,³⁷ α -gate opening by tail insertion,³⁸ long-range allosteric coupling between subdomains,³⁹ and major conformational changes during active unfolding reactions.⁴⁰

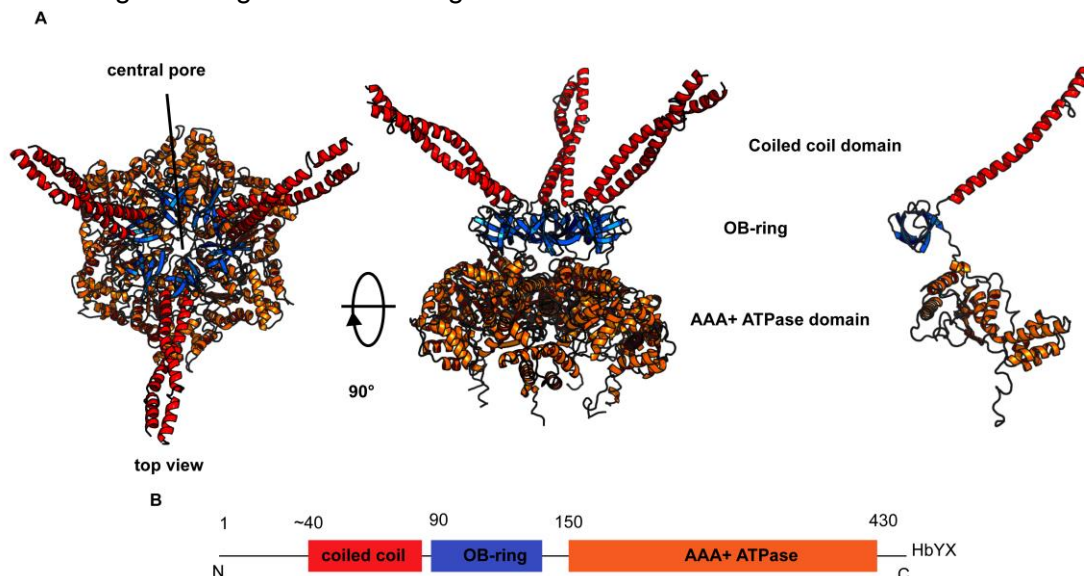


Figure 1.5 Architecture of the archaeal PAN. A) Top- and side view representations of the hexameric PAN (left and middle) and an isolated PAN-protomer (right). Three N-terminal coiled coil (CC) domains are formed by three pairs of helices of adjacent subunits and protrude from the top of the complex (red). The CC is followed by a rigid hexameric OB-ring (oligosaccharide-/oligonucleotide-binding) which forms a 13 Å pore as a first constriction for folded substrate proteins (left, top view). The C-terminal AAA+ ATPase ring, connected to the OB-ring via flexible linkers, binds and hydrolysis ATP to execute mechanical force on the folded substrates. Shown is a homology

model of PAN from *M jannaschii* based on cryo-EM reconstructions of PAN from *A. fulgidus* (PDB: 6HE4).⁴¹ B) Domain architecture of PAN: The first ~ 40 residues of PAN are considered to be disordered and have not been characterized structurally.

1.2.2 The archaeal 20S proteasome

The archaeal 20S proteasome displays a high structural similarity to the eukaryotic 20S core particle (Figure 1.4). It is likewise composed of two outer heptameric α -rings that sandwich a double β -ring. The archaeal 20S core particle, however, exhibits a simpler architecture: Both α - and β -rings are homoheptameric assemblies instead of the heteroheptameric assemblies in eukaryotes.⁴² Consequently, all 14 β -subunits harbour proteolytic sites for substrate degradation. Similar to the eukaryotic system, diffusion of protein substrates into the proteolytic chamber is blocked by the N-terminal extensions of the α -rings forming a gate.⁴³

1.2.3 Targeted protein degradation in archaea

Similar to the eukaryotic ubiquitylation archaea possess a posttranslational modifying system which was found to be involved in proteasomal degradation of proteins.¹⁸ The process named sampylation involves the covalent attachment of a small archaeal modifier protein (SAMP1 or SAMP 2) to substrate proteins. SAMP is considered to be a structural homologue of the eukaryotic ubiquitin and shares the prominent β -grasp fold of ubiquitin.⁴⁴ It also contains a hydrophobic patch similar to the one around Ile144 in ubiquitin, which is supposed to be involved in ubiquitin-like interactions with potential SAMP-interaction partners. Sampylation was first described in the archaeon *Haloferax volcanii* and is carried out by the enzyme UbaA (Ubiquitin-like activating protein of Archaea), a homologue of the eukaryotic E1 enzyme. Analogous to E1, UbaA catalyzes the activation of SAMP through ATP-dependent adenylation of its C-terminus. Although isopeptide bonds between SAMP and lysine- ϵ -amines have been found for several substrates, the homologues of the E2 enzymes and E3 ligases have not been identified yet, leaving the pathway downstream to SAMP-activation still not understood. However, archaeal enzyme-homologues to DUBs (deubiquitinases) have been found to remove SAMP from substrates proteins. The protease HvJAMM from *H. volcanii* was found to be a member of the JAMM/MPN+ family of protease, which also includes the 26S proteasome's DUB RPN11 in eukaryotes. In agreement, HvJAMM was reported to have proteolytic activity, removing isopeptide-linked SAMP from target proteins.⁴⁵ Despite this high similarity, implying a targeting system for proteasomal degradation similar to the eukaryotic UPS, no direct evidence for physical interactions between sampylated protein substrates and the archaeal proteasome have been described. Further, no sophisticated SAMP processing machinery, similar to RP in eukaryotes, has been described for archaeal proteasomes. Thus, it remains unclear whether sampylated proteins are directly targeted to the proteasome.

Most mechanistic studies on protein degradation by archaeal proteasomes employed substrates like the green fluorescent protein (GFP), which contained a genetically encoded ssrA-tag for targeting to the proteasome regulator PAN.^{35,40,46-49} This ssrA-tag originates from bacteria where it is incorporated into polypeptides during translation when the ribosome is stalled. These incomplete proteins are then targeted via their C-terminal ssrA-tag for degradation by bacterial AAA proteases.⁵⁰ It is well accepted that the eukaryotic proteasome degrades substrates also in a ubiquitin-independent process, where other degrons (degradation

tags) act as markers for superfluous proteins.⁵¹ By analogy, it is likely that archaeal proteasomes accept proteins marked with degrons other than ubiquitin or SAMP, but rather consisting of flexible unstructured tails similar to the *ssrA*. The efficiency of *ssrA* in targeting proteins to the degradation machinery in archaea has been demonstrated for the archaeal PAN.³⁴

A

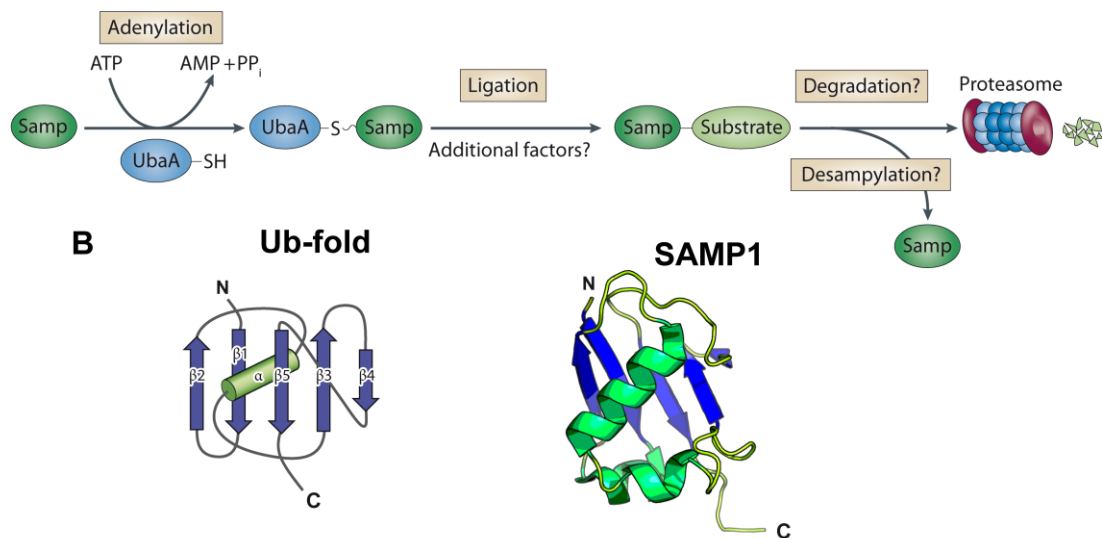


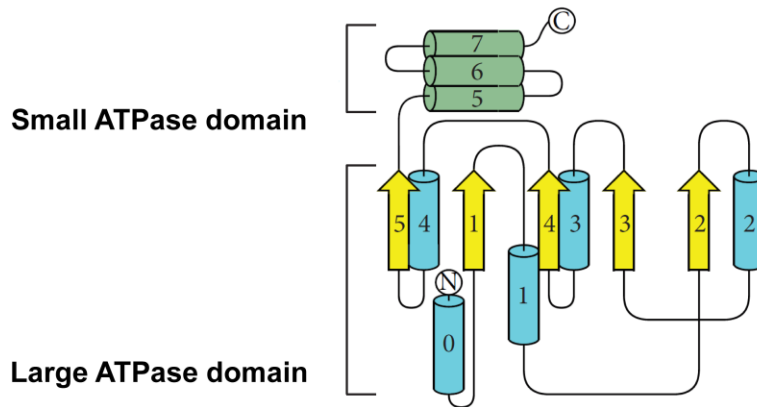
Figure 1.6 Ubiquitin like protein modifications in archaea A) For archaea a protein labelling pathway is proposed analogous to the eukaryotic ubiquitin proteasome system. SAMP, the structural analog of ubiquitin, was found to be linked to the E1 analog UbaA via thioester bonds in an ATP-dependent manner. However, the presence of additional factors, involved in the ligation of SAMP to proteins for proteasomal degradation, is yet to be discovered. B) The conserved β -grasp fold in ubiquitin (left) is conserved in the SAMP X-ray structure from the halophilic archaeon *Haloferax volcanii* (right, PDB: 3Po0).⁵² Panel A and left part of Panel B are adapted from MAUPIN-FURLOW et al.^{18,53}

1.3 Structural properties of AAA+ ATPases

AAA+ unfoldases, also termed AAA+ translocases, are members of the AAA+ ATPase (ATPases associated with various cellular activities) family of proteins which convert the chemical energy of ATP into mechanical force and apply it to substrates for unfolding and translocation. The hallmark of all AAA+ ATPases is a highly conserved ATPase domain with distinct sequence-specific and structural features, which identify this family as a subset of the larger p-loop NTPases superfamily.⁵⁴ The ATPase domain usually comprises 200-250 residues, which form a distinct $\alpha/\beta/\alpha$ architecture. At the center of this fold is a five-stranded β -sheet built by the β -strands β_5 - β_1 - β_4 - β_3 - β_2 . The central sheet is flanked by α -helices on both sides resulting in the $\alpha/\beta/\alpha$ -sandwich arrangement. (Figure 1.7 A). What distinguishes AAA+ ATPases from other p-loop NTPases is the insertion of β_4 between β_1 and β_3 , the absence of an antiparallel β -strand adjacent to β_5 and the lack of any further additional strand directly adjacent to β_2 and β_5 .⁵⁵ Most AAA+ ATPases contain an additional C-terminal helix bundle which forms a lid for the nucleotide binding site and further mediates interactions to other subunits of oligomeric complexes. The relative conformation of this lid to the $\alpha/\beta/\alpha$ -core is oftentimes dependent on the nucleotide state of the ATPase and consequently involved in

the conformational changes which execute mechanical work on substrates.⁵⁶ The lid or helix bundle is mostly referred to as the “small” ATPase subunit in contrast to the Walker A/B motifs carrying “large” ATPase subunit.

A



B

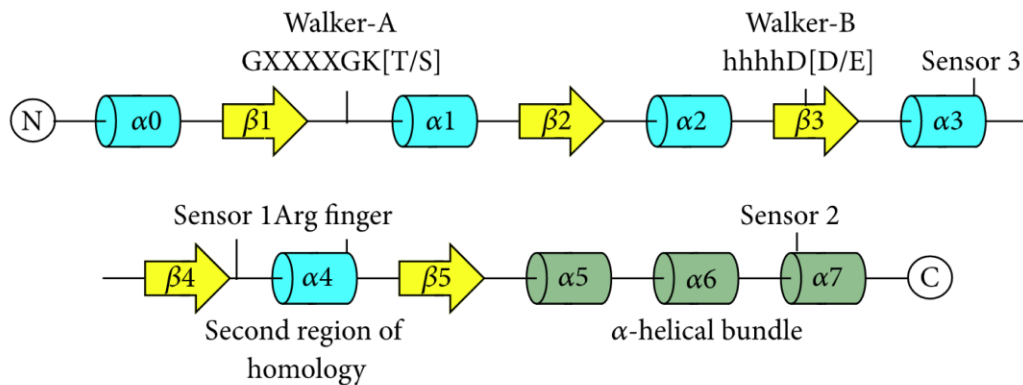


Figure 1.7 Topology of AAA+ ATPases A) AAA+ ATPases share a common canonical $\alpha/\beta/\alpha$ fold architecture. A central five-stranded β -sheet of the order $\beta_5-\beta_1-\beta_4-\beta_3-\beta_2$ is sandwiched between flanking α -helices on both sides. This large ATPase subdomain is accompanied by a C-terminal small ATPase subdomain, comprising a helix bundle. B) The canonical AAA+ domain contains a number of strictly conserved sequence features: The Walker A motif between β_1 and α_1 with a conserved lysine, the Walker B motif on β_3 harbouring the catalytic glutamate or aspartate and the second region of homology with a conserved trans-acting arginine finger. Schematics adopted from Miller et al.⁵⁶

The primary sequences of AAA+ ATPases share several distinct motifs common to all other p-loop NTPases (Figure 1.7 B): Most prominently are the Walker A and Walker B motifs named after the British biochemist John Walker (Nobel Prize in Chemistry, 1997). The two Walker motifs contain a couple of conserved residues critical for binding and hydrolysis of ATP. The Walker A motif comprises a GXXXXGK [T/S] sequence, where “X” is any residue and its C-terminus is either threonine or serine.⁵⁷ It forms a loop between β_1 and α_1 in the AAA+ topology. This canonical so called p-loop is the most conserved sequence motif in AAA+ proteins. Walker B motifs are composed of a hhhhD[D/E] sequence where “h” is a hydrophobic residue and the C-terminus either aspartate or glutamate. This glutamate at the C-terminus of β_3 is the catalytic base in the hydrolysis of ATP, where it activates a water molecule for the nucleophilic attack on the ATP- γ -phosphate. When mutated to glutamine or alanine ATP-hydrolysis is inhibited but ATP-binding still occurs (known as Walker B mutants). This is because the Walker B glutamate interacts with ATP only indirectly via the coordinated

water molecule whereas the conserved lysine residue in the Walker A motif directly interacts with the phosphate groups of the nucleotide.⁵⁸

What distinguishes the AAA+ ATPases from other p-loop NTPases is a motif called second region of homology (SRH).⁵⁸ It is located after the Walker B sequence and comprises 15-20 residues, spanning from part of β_4 , through all of α_4 to the loop connecting α_4 and β_5 . The functional features of the SRH motif include a conserved arginine finger and a so-called sensor 1. The sensor 1 motif is located at the N-terminal end of SRH in the loop which connects β_4 and α_4 and comprises a polar residue, like asparagine, serine, threonine or aspartate. Structurally the sensor resides between the Walker A and B motifs and helps to orient the nucleophilic water molecule which attacks the γ -phosphate of ATP. Thus the sensor 1 is also crucial for proper function of anything related to the hydrolysis of ATP.⁵⁹ The arginine finger as the second key feature of the SRH motif plays an exceptional role: While the Walker A/B and sensor 1 motifs act on the same subunit as cis-elements, the R-finger is trans-acting and reaches the ATP-binding site from the neighbouring subunit of the AAA+ ATPase. This is crucial for bridging the different subunits as well as for intersubunit allosteric signalling. The conserved arginine finger is located at the C-terminal end of SRH on the loop between α_4 and β_5 . Occasionally arginine is replaced by a lysine, fulfilling a similar function.⁵⁵ As trans-acting elements arginine fingers point into the nucleotide-binding and -hydrolysis-site of the neighbouring subunit, where they interact with the γ -phosphate of the bound nucleotide and are thought to have a role in stabilizing accumulated negative charges in a hydrolysis transition state.⁵⁸

Additional active site features of the AAA+ domain include sensor 2 and 3 residues. A conserved sensor 2 arginine located near the beginning of α_7 mediates conformational changes associated with the cycle of ATP-binding and hydrolysis by directly interacting with the α -phosphate of bound ATP. In AAA+ proteins with a C-terminal lid-structure the sensor 2 arginine is a cis-acting element, opposed to the arginine finger in the SRH motif. In ATPases without a C-terminal lid, however, sensor 2 can be transacting. A sensor 3 residue (either Arg or His) is located in some AAA+ ATPases on α_3 ; it has been proposed that this trans-acting residue plays a role in stabilizing a closed ATP-bound inter-subunit interface in hexameric AAA+ ATPases.⁵⁶

Since AAA+ ATPases do not act as monomers but form cyclical hexameric assemblies, inspection of the AAA+ specific intersubunit interactions and trans-acting motifs is of high relevance. In fact, the nucleotide binding pocket is located at the interface between two AAA+ subunits involving contribution from two monomers. Most notably, the above mentioned conserved arginine-finger is the major trans-acting element. It forms a tight bridge with the γ -phosphate of ATP on a neighbouring ATPase subunit in the counterclockwise direction. Thus, this interaction is dependent on ATP-binding in the neighbouring subunit and, consequently, this trans-acting interaction is abolished in ADP-bound or nucleotide-free binding pockets, where the R-finger is retracted from the nucleotide pocket. Further the so-called intersubunit signalling motif (ISS) involves several trans-acting interactions, involving intersubunit π -stacking of a phenylalanine aromatic ring with the guanidine-moiety of an arginine on the neighbouring subunits. The relative conformation and the presence of these intersubunit interactions is used as an indicator for the nucleotide state of different subunits in the context of structural studies on hexameric ATPase (Figure 1.8).^{41,60,61}

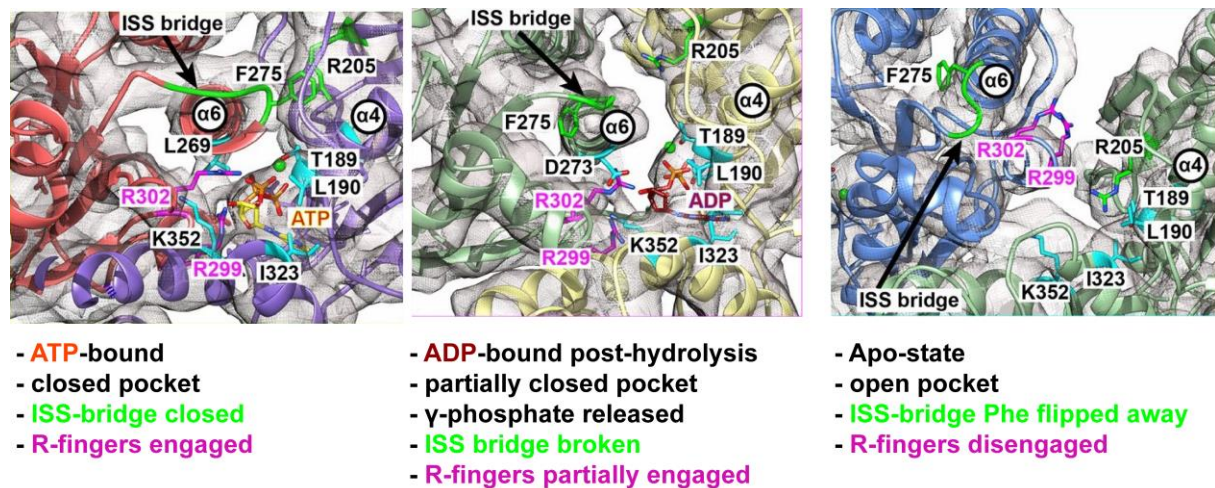


Figure 1.8 Intersubunit interactions in AAA+ ATPases depend on nucleotide state. Shown are three different nucleotide states in PAN from *A. fulgidus*.⁴¹ In the left panel an ATP-bound nucleotide pocket is displayed. The pocket is found in a closed conformation, in which it is covered by a lid formed by the so-called ISS-bridge (green, intersubunit-signalling via π -stacking of F275 with R205 on the ATP-bound right subunit). Further the conserved R-fingers R302 and R299 (magenta) are engaged with the γ -phosphates of ATP. The middle panel illustrates a post-hydrolysis, partially closed, nucleotide pocket with bound ADP. Helix 6 on the clockwise-neighbouring subunit carrying the ISS-bridge is retracted and the conserved F275 is flipped away from the binding pocket. Since the γ -phosphate is released the trans-acting R-fingers are also partially disengaged from the nucleotide. The right panel displays an apo-state nucleotide pocket, in which all intersubunit interactions are broken and the pocket adopts an open conformation. This figure was adapted from MAJUMDER et al.⁴¹

The family of AAA+ ATPases represents a large group of enzymes with a multitude of cellular functions.⁵⁵ Based on sequence and structural comparison they have been classified into several “clades” where members of each clade are derived from a common ancestor.⁶² While all AAA+ ATPases share the abovementioned core motifs, members of each clade are defined by insertions or alterations of special sequences or secondary structural elements inside this core.⁶³ In the following chapter, I will briefly characterize the features of the two clades relevant for this thesis containing protease associated translocating unfoldases.

1.4 The diverse architecture of protease associated AAA+ ATPases

AAA+ ATPase modules of the “classic clade” are the force-generating components in proteasomal- and FtsH-, Cdc48- unfoldases and in the D1-domain of ClpA/B/C. They all form closed ring hexamers and are involved in protein remodelling processes. Common to all classic AAA+ family members is a short α -helical insertion between β_2 and α_2 of the AAA+ domain, which forms a loop pointing towards the axial pore of hexameric AAA+ complexes (pore-1-loop) (Figure 1.9). Based on mutational studies⁶⁴ and also recent substrate bound structures^{60,65} this loop has been found to be heavily involved in substrate interaction and -processing. Further, the classical clade seems to lack the conserved sensor 2 arginine residue close to the beginning of α_7 .⁵⁶

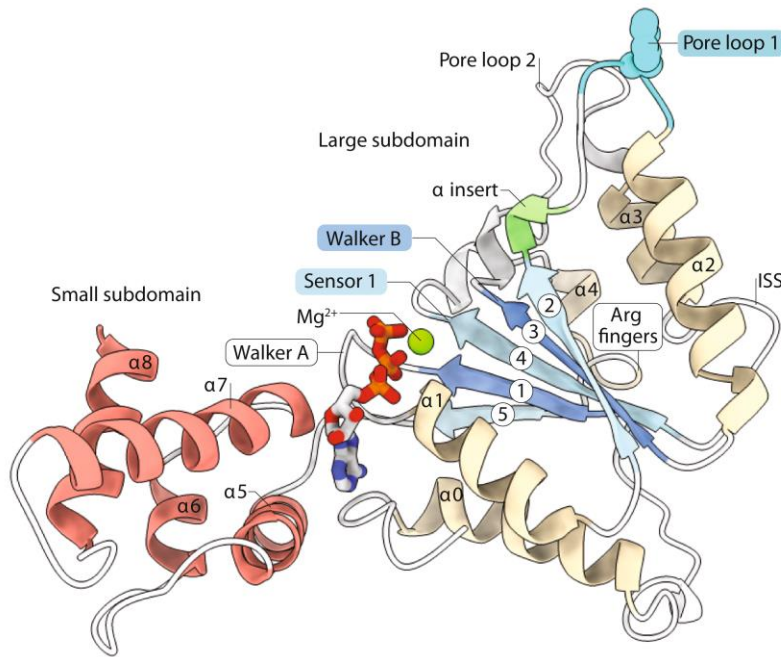


Figure 1.9 Three-dimensional topology of an ATPase domain of the classical clade. Displayed is the AAA+ ATPase subdomain of the mitochondrial protease YME1 which is a FtsH-like protease. All canonical AAA+ ATPase features (central 5-stranded β -sheet, Walker A/B, R-fingers etc.), are labelled accordingly. Unique for the classic AAA+ ATPases is the α -helical insertion (green) between β_2 and α_2 which carries the pore1-loop for substrate engagement (cyan). The figure is adapted from PUCHADES et al.⁶⁶

The second clade containing protease-associated ATPases is the so-called HCLR-clade named after their subfamilies HslU/ClpX, ClpA/B/C-CTD, Lon, and RuvB. The HCLR clade contains only one specific feature built in to the AAA+ core module: A β -hairpin insertion before the sensor 1 motif between α_3 and β_4 , dubbed the pre-sensor 1 β -hairpin (ps1 β h). Although crucial for proper functioning of HCLR clade ATPases, the ps1 β h insert was not found to interact with substrates during the translocation, but rather play a role in substrate recognition.⁶⁷

The functional diversity of all AAA+ unfoldases associated with proteases is further determined by diverse combinations of non-ATPase domains decorating the force-generating ATPase core.

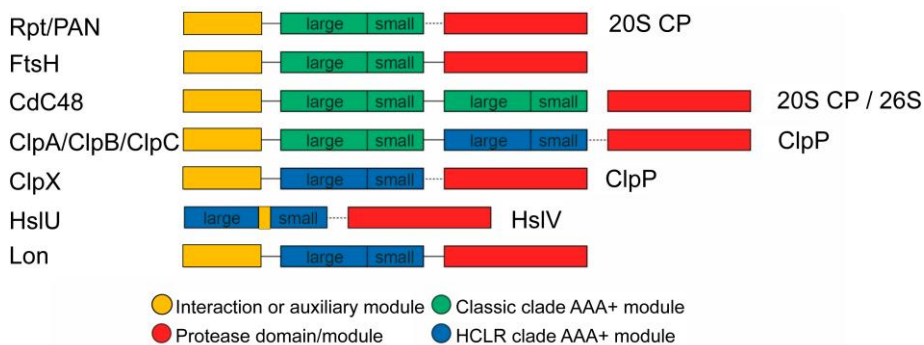


Figure 1.10 Domain architectures of protease associated AAA+ unfoldases. In addition to one or two conserved AAA+ modules, most unfoldases carry distinct additional modules, for example for substrate- or adapter protein interaction (yellow modules). Further, they interact with their associated proteases either by specific non-covalent interactions (Rpt/PAN, Clp-ATPases, HslU, dashed lines) or are part of one polypeptide carrying also the protease (FtsH, Lon, continuous lines). Cdc48 has been shown, to work together with the 20S proteasomal core particle, but it is also capable of transferring unfolded proteins to the 26S proteasome.⁶⁸ The schematic is adapted from ZHANG et al.⁶³

Generally, unfoldases carry different N-terminal non-ATPase domains which are involved in substrate recognition and -binding. Furthermore, several AAA+ unfoldases like Cdc48/VAT and ClpA/B/C contain not only one but two AAA+ modules in their primary sequence, which fulfill distinct functions. Unfoldases also differ significantly in the way they work together with their respective protease. While the AAA+ unfoldases in the proteasomal regulatory particle, ClpX, ClpA/B/C and HslU associate with their respective protease complexes via specific but non covalent interactions, the proteases of FtsH and Lon are essentially part of the same polypeptide chain as that of the unfoldase.⁶³

The overall architectures and high-resolution structures have been elucidated for most of the abovementioned proteolytic AAA+ complexes.^{41,60,66,69-72} Together with numerous biochemical and biophysical studies, they have contributed a great deal to the understanding of their function and mechanism. Herein, among others, the eukaryotic proteasome regulatory particle,^{60,65} its archaeal counterpart PAN^{41,73} and the bacterial AAA+ unfoldase ClpX^{69,70} have served as extensively studied model systems from the two above-mentioned clades. The following chapters will deal with their structural and functional properties, eventually discussing a potentially common mechanism for the processing of translocating AAA+ ATPases.

1.5 Structure and function of proteasomal ATPases

Proteasomes, as the major protease in eukaryotes, are composed of a barrel-shaped proteolytic core particle capped on both sides by hexameric AAA+ unfoldases. The unfoldases catalyze the unfolding of condemned proteins and translocate them into the proteolytic core particle in an ATP-dependent manner.⁷³ While the biochemical mechanism of the proteolytic cleavage is relatively simple,⁷⁴ unfolding of substrates prior to injection into the proteolytic chamber of the 20S CP requires highly coordinated hydrolysis of ATP in all six subunits with subsequent major conformational changes.⁶⁰ The current structural and functional understanding of proteasomal ATPases has been gained through studies on proteasomes in particular from *Homo sapiens*, *Saccharomyces cerevisiae* and the archaeon *Methanocaldococcus jannaschii*. Numerous biochemical studies together with recent high resolution cryo-EM models^{41,60,65} have contributed to the discovery of a conserved functional cycle of substrate processing, which I will review here. I will focus on the details of the archaeal proteasomal ATPase PAN (Proteasome activating nucleotidase) since it is the biological subject of this thesis. Where details in archaea are still unknown, insights from eukaryotic or bacterial systems will be referred to, providing a comprehensive picture of the structure and function of proteasomal ATPases.

Proteasomal ATPase hexamers can be structurally divided into several distinct subdomains (also see Figure 1.5): The ATPases from bacteria, archaea and eukaryotes carry N-terminal helical domains which dimerize to form three coiled-coils (CC) protruding from the complex. The CCs are followed by oligosaccharide- and oligonucleotide binding domains (OB) which form the so-called OB-ring with a ~13 Å central pore as the first constriction for folded substrates during proteasomal processing. As ATPases from the classic clade of AAA+ ATPases, proteasomes carry a C-terminal hexameric AAA+ domain composed of a large and small AAA+ subdomain. They contain the conserved signature features, such as Walker A/B motifs, trans-acting arginine fingers and pore 1/2 loops for substrate interaction.³¹

While the basic ATPase activity and unfolding function as part of the eukaryotic 26S proteasome has been described for the complex 19S regulatory particle throughout the

1990's,⁷⁵⁻⁷⁷ attempts to identify archaeal homologues have been hampered until the first full archaeal genome was sequenced.⁷⁸ As a result of the successful sequencing of the first archaeal genome from the thermophilic *Methanocaldococcus jannaschii*⁷⁹ ZWICKL and colleagues reported the first biochemical characterization of an archaeal 19S RP homologue.³⁰ Due to its ATP-dependent ability to activate the 20S core particle from *T. acidophilum* they named it PAN (Proteasome activating nucleotidase). In contrast to the hetero-hexameric ATPase from eukaryotes, the archaeal version assembled from a single gene product upon recombinant expression in *E. coli*, making an excellent model system for biochemical and structural studies on proteasomal ATPases. Following studies have established that PAN, like the 19S RP, serves multiple main functions in the proteasomal breakdown of substrate proteins: ATP-dependent unfolding and translocation of substrate proteins and gate-opening in the α -rings of the 20S CP.⁴⁶

Low resolution structures of both eukaryotic and archaeal proteasomes became available in 1993 and 2005, respectively (Figure 1.11).^{33,80} Based on sequence similarity to other AAA+ ATPases, whose structures were known at that time, it was suggested that proteasomal ATPases also adopt a hexameric ring-shaped assembly on top the 20S core particle. However, structure-based conclusions about the functional mechanism had been hampered until high resolution structures became available.

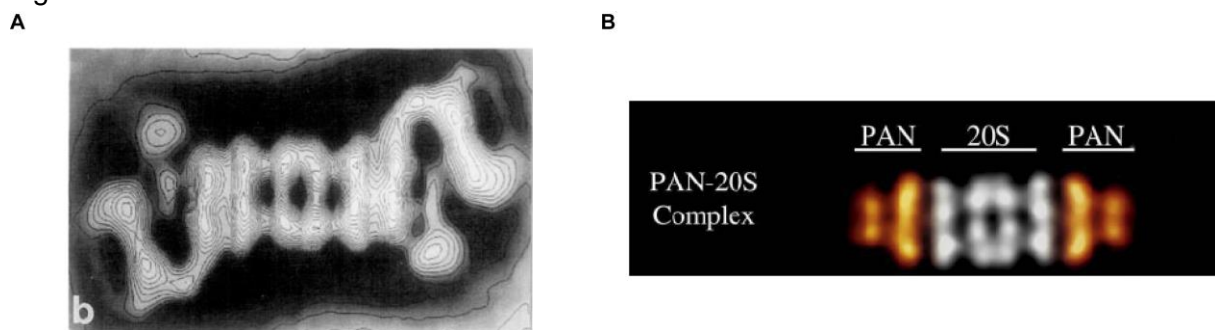


Figure 1.11 Low resolution electron micrographs of eukaryotic and archaeal proteasome. Early low resolution electron micrographs of the eukaryotic 26S proteasome (A) and the archaeal PAN-20S proteasomes (B), respectively, clearly show the compartmental architecture of several stacked rings. In the case of the eukaryotic 26S proteasome also distinct lid-densities are visible while the flexible PAN coiled coil domain is invisible. Panel A has been adapted from PETERS et al.⁸⁰ while panel B is adapted from Smith et al.³³

First X-ray crystallographic data on proteasomal ATPases had been reported in 2009 providing the structure of PAN-subdomains from *M jannaschii* and *A. fulgidus*.^{31,81} Since the full complex resisted crystallization, two conserved subdomains in PAN were studied (Figure 1.12): A planar N-terminal ring formed by six OB-domains (oligonucleotide/oligosaccharide binding) and a C-terminal ATPase domain, resembling the conserved canonical fold of AAA+ ATPases from the classic clade. Although PAN's ATPase domain did not crystallize as a defined hexameric ring-structure, high structural similarity of PAN's nucleotidase domain to the prototype AAA+ ATPase HslU justified building a hexameric ring model of PAN. The planar 6-fold symmetric AAA+-ring model of PAN resembled features of other protease-associated AAA+ ATPases like ClpX: The central pore was found to be narrow with a surface diameter of less than 5 Å and the pore was lined by aromatic Tyr-residues of the conserved pore-1 and pore-2-loops. An accompanying structure-guided mutagenesis study confirmed the functional importance of several conserved AAA+ key-features like the pore-1/2 loops and the catalytic glutamate 271. While the N-terminal OB-ring exhibits a pseudo-3-fold symmetry imposed by an unique cis-proline bond in three of the six subunits, the C-terminal planar AAA+-ring model was perfectly 6-fold symmetric.

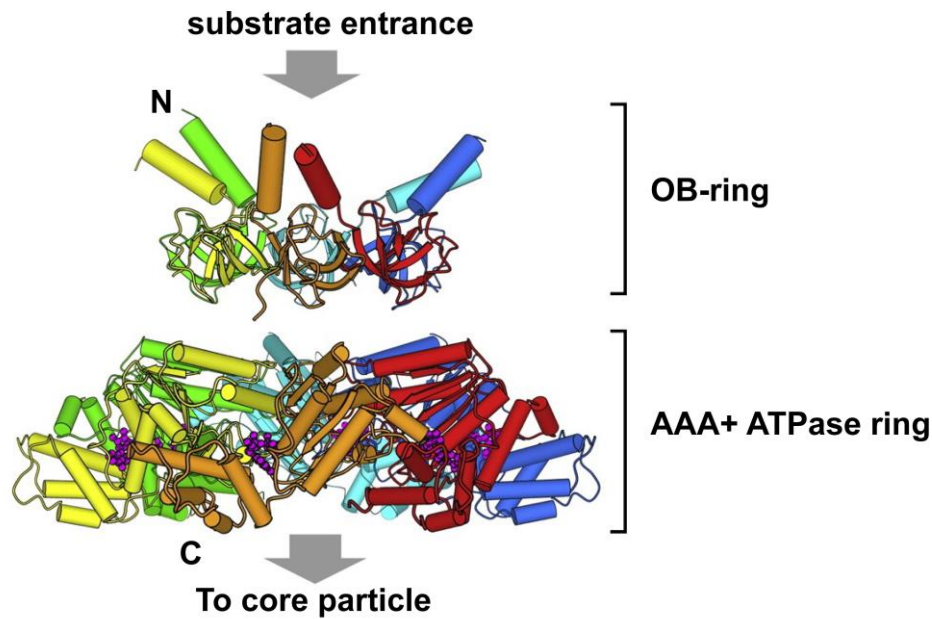


Figure 1.12 X-ray crystal structures of PAN subdomains from *M jannaschii*. Since full length PAN did not crystallize, ZHANG et al. solved the crystal structure of two separate domains. While the N-terminal subcomplex formed a hexameric OB-ring on its own, the C-terminal ATPase was not found to crystallize as a ring shaped assembly. However, a hexameric model could be obtained by superimposing six monomeric ATPase domains onto the hexameric crystal structure of the related bacterial AAA+ ATPase HslU. The figure is adapted from ZHANG et al.³¹

Further biochemical studies on PAN and the eukaryotic 26S proteasome, however, soon suggested that a functional state of PAN and the ATPases from the 26S proteasome exhibits functional asymmetries which are not compatible with a six-fold symmetric planar structure of the AAA+-ring.⁴⁸ In contrast to PAN, the AAA+-ATPase ring of the 26S proteasome is a heterohexameric assembly composed of the homolog but distinct subunits Rpt1-6, thus an asymmetric architecture seemed plausible. Indeed, early low resolution cryo-EM studies of the yeast 26S proteasome revealed the six ATPase subdomains to be arranged in a highly asymmetric spiral staircase arrangement, which had been reported only for other AAA+ proteins before.²⁰ However, for a long time functional asymmetries for the homohexameric PAN could be only demonstrated by various biochemical studies^{37,48} until a structure of full length PAN was reported in 2018.⁴¹

Since substrate processing by AAA+ molecular machines like the proteasome is a highly dynamic process involving major ATP-dependent conformational rearrangements static structures as “snapshots” contribute only in part to the mechanistic understanding. Consequently, in depth biochemical and biophysical analyses supported the structural models, where the proteasome was trapped in a range of several distinct states, thus providing profound insights into the functional mechanism of the proteasomal motor.

1.5.1.1 Importance of nucleotide binding to proteasomal ATPases

Mechanical work by proteasomal ATPases is fuelled by the hydrolysis of ATP. Each of the six subunits has a distinct nucleotide binding pocket and can potentially bind a nucleotide, either ATP or ADP. However, using slowly or non-hydrolysable ATP analogues it was dis-

covered that certain key functions of proteasomal ATPases do not depend on ATP-hydrolysis but only on its binding.³³ While ATP-hydrolysis is strictly required for successful unfolding and degradation of folded model substrates, binding of ATP (emulated by supplementation with the slowly hydrolysable ATP-analogue Adenosine 5'-(γ -thiotriphosphate), ATP γ S) facilitated binding of substrates, interaction of the ATPase subunits with the 20S core particle, simultaneous opening of the 20S α -gates and translocation of unstructured proteins like β -casein.⁸² These effects were not observed when no nucleotide or ADP was added indicating, that these processes do not need ATP to be hydrolysed but only to be bound to the proteasomal ATPases.

Although the PAN homohexamer contains six identical ATP-binding sites, further studies suggested that PAN only binds a maximum of four nucleotides at a given moment. Utilizing ATP γ S, SMITH and colleagues found that PAN, and also the 26S proteasome, have two distinct binding sites with different affinities for ATP. Proteasomal ATPases bind two ATP copies with high affinity (K_d 0.493 μ M) and two additional ATP molecules with a much lower affinity (K_d 113 μ M). In agreement, functions of PAN which are dependent on ATP binding (substrate binding, 20S association, gate-opening) also followed a biphasic behaviour with lowered activities at saturating nucleotide concentrations, suggesting that maximum activity was reached when two ATP molecules and two ADP molecules were bound.⁴⁸ A similar two-phase behaviour was also proposed for the 26S proteasome. Further, SMITH and colleagues deduced a first model of an ordered nucleotide hydrolysis pattern in the PAN ATPase ring, which essentially described the pairwise binding and hydrolysis of ATP in subunit opposite to each other in the ring. This asymmetry of nucleotide binding to non-equivalent subunits in the proteasomal ATPases was highly consistent with the findings of an asymmetric spiral staircase arrangements found in the 26S proteasome.²⁰ In the light of the spiral staircase structures of the 26S proteasome the nucleotide binding and hydrolysis model by SMITH was further refined in a study by his group in 2015, in which the two bound ATP molecules in ortho position to each other are followed by two ADP molecules and two empty protomers in the cycle (ATP-ATP-ADP-ADP-empty-empty). Further, the previously proposed role of the two conserved arginine-fingers in PAN's ATPase domains was revealed to play critical functions in the allosteric coupling between the ATPase subunits.³⁷

1.5.1.2 The conformational landscape of proteasome complexes

The functional asymmetries found in biochemical studies have been confirmed by a multitude of cryo-EM studies on both human and yeast 26S proteasomes^{20,60,65,83-90} and recently also PAN.⁴¹ Especially, the arrangements of the six AAA+ ATPase subdomains in the different structural models have attracted much attention, since they represent the origin of the ATP-hydrolysis-dependent conformational changes which generate mechanical force on substrates. The discovery of an asymmetric spiral staircase arrangement of the AAA+ modules caused speculations on an ordered ATP-hydrolysis pattern around the hexameric ring.²⁰ Proteasomal structures have been obtained in the presence of different nucleotide molecules which had a strong influence on the predominant structural arrangement of the AAA+ domains. Proteasome samples in the different nucleotide-bound states were found to contain several conformations simultaneously, representative of a highly dynamic conformational landscape of the regulatory particle.⁶³ The first subnanometer resolution structural model of

the yeast 26S proteasome displayed the abovementioned spiral staircase arrangement of the AAA+ module of the RP: While the small AAA+ subdomains of five of the six ring members were found to be in plane with respect to the core particle, the large AAA+ subdomains were lifted differentially to resemble a right handed staircase with Rpt3 at the top and Rpt2 at the lowest step, respectively. Rpt6 resided between Rpt2 and Rpt3 in the cyclical order and was found in a tilted orientation, bridging the highest (Rpt3) and the lowest (Rpt2) position of the staircase (Figure 1.13 A and B). Furthermore, this low resolution structure reported a previously observed 10 Å offset between the α -rings of the CP and the AAA+ ring. This offset was interpreted to prevent coaxial alignment of the AAA+ pore and the central CP core (blocking efficient substrate degradation) but on the other hand facilitate the engagement of the three HbYX motifs (on Rpt2/3/5) with their respective binding pockets in the core particle. While this structure was obtained in ATP saturating conditions and in the absence of substrate, samples supplemented with ATP γ S or a proteasomal substrate resulted in substantially changed conformational arrangements in the AAA-ring.^{83,84}

Further deep classification of cryo-EM data sets of both ATP- and ATP γ S-bound proteasomes⁸⁸ together with improved resolution, revealed the co-existence of several substantially different global conformational states of the proteasomal ATPases within given data sets. The proportional distribution of these conformations was found to be highly dependent on the nucleotide in the sample buffer and the presence or absence of ubiquitylated proteasomal substrates. Following the nomenclature for the structures of proteasomes from *S. cerevisiae* several states were distinguished based on a number of key criteria, unique to either a single or a subset of states: (1) relative position of CP α -ring, AAA+ ring, N-terminal N-ring and the lid to each other, (2) steepness of the spiral staircase, (3) relative positions of Rpt1-6 in the spiral staircase, (4) position of the deubiquitinase RPN11 active site with respect to the pore entrance, (5) nucleotide occupancy and nucleotide pocket conformation in each monomer (see also Figure 1.8), (6) openness of the α -ring gates in core particle and (7) presence of a substrate polypeptide inside the central pore.

The state predominant in the first ATP bound structure was later named S_1 . As described above, it is characterized by a pronounced offset between the AAA+ base of the regulatory particle and the CP. Consequently, the pore channels of the N-ring-domain, the ATPase-ring subdomain and the core particle are not coaxially aligned, rendering this conformation substrate translocation incompetent, which is supported by the fact that structures of actively translocating proteasomes are never found in this state.⁶⁰ Further, Rpn11, the proteasomal deubiquitinase with its isopeptidase active site is displaced from the central processing pore making cotranslocational deubiquitylation unlikely. The large AAA+ subdomains of all six ATPases are arranged in a steep spiral staircase with Rpt6 at the highest and Rpt2 at the lowest position, respectively. Rpt6 is disengaged from the spiral and resides at an intermediate altitude. All six nucleotide pockets contain nucleotide density. While five pockets are considered to be ATP bound, the binding pocket formed between Rpt2/6 at the spiral staircase "seam", however, displayed a more open conformation and less density, and is considered to be occupied by ADP (see also Figure 1.8).⁹¹ Although the three HbYX motifs of Rpt2/3/5 are docked into their binding pockets in the CP α -ring, the gates of the CP are closed in the S_1 state. This state is considered to be not actively translocating but competent to accept polyubiquitylated substrates.²¹

A second conformation is named S_2 as it is most closely related to S_1 and was found in various samples under both saturating conditions of ATP^{60,88} or ATP-analogues.^{88,91} The ATPase

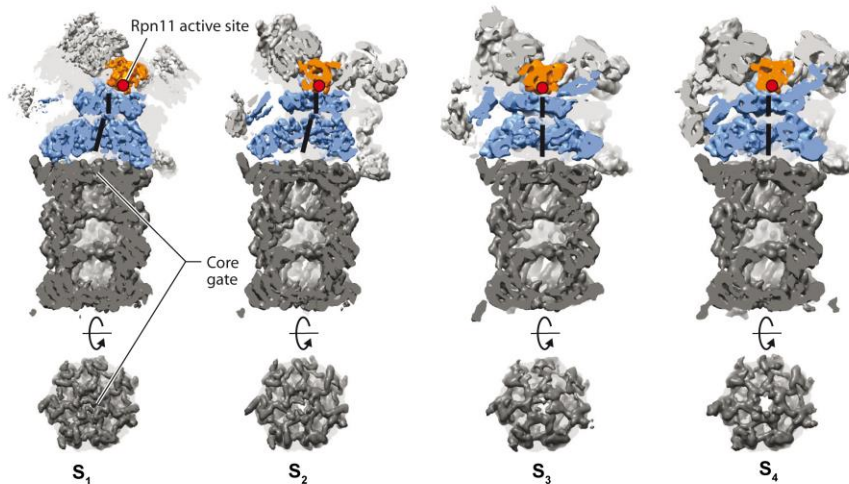
motor in S_2 is essentially unaltered compared to S_1 while a major 25° rigid-body rotation of the lid positions the Rpn11 active site towards the central processing pore, thus facilitating cotranslocational deubiquitylation.^{60,92} Because of the tight engagement of Rpn11 to the central pore, this S_2 is considered to be incompetent for substrate binding, placing it in the sequential order after substrate binding to the S_1 state.⁹² The pores of the N-ring, AAA+ ring and the CP are still not coaxially aligned in this state.

A major conformational rearrangement of the ATPase subunits is observed during the transition from S_2 to S_3 . Now, Rpn11, the N-ring, the AAA+ ring and the CP pores are perfectly coaxially aligned and form a much wider consecutive channel facilitating the translocation of a substrate polypeptide. The steepness of the spiral staircase now is much shallower and Rpt1-6 are arranged in a substantially altered fashion: Compared to the S_1 state now Rpt1 and Rpt4 are at the top and bottom, respectively, while Rpt5 bridges the seam between these two.^{60,83} In S_3 structures the Rpt-intersubunit interfaces are characterised by the formation of highly similar rigid-bodies composed of one small AAA+ subdomain with the large AAA+ subdomain of the clockwise-next neighbouring Rpt (Figure 1.13 D). These rigid body interfaces are not found in the S_1 state. The substrate engaging pore-1-loops residing on the large subdomains also adopt the staircase arrangement. Early structural models obtained from samples of stalled proteasomes in the presence of a substrate also adopted the S_3 state.⁸³ However due to limited resolution the substrate could not be modelled in the central pore, which precluded further interpretation until high-resolution structures of actively translocating proteasomes became available (chapter 1.5.1.4)^{60,65} confirming S_3 to be an actively processing state.

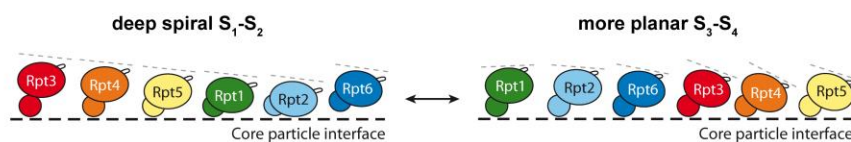
By means of deep classification of cryo-EM data sets recorded in the presence of ATP-analogues but absence of substrate a fourth minor state could be identified.^{91,93} The S_4 state was found in the presence of ATP γ S in the human proteasome⁹⁴ and in the presence of ADP-BeF_x, a non-hydrolysable ATP-analog in yeast proteasomes.⁹¹ This state is similar to the S_3 state in that it exhibits coaxial alignment of N-ring, AAA+-ring and the CP pores with the deubiquitinase Rpn11 active site residing at the pore entrance and also a shallower staircase arrangement compared to the S_1 and S_2 states. However, a striking difference is the rearrangement of the AAA+ motor towards a staircase in which Rpt5 and Rpt3 are at the top and bottom position, respectively, with Rpt4 residing at the seam. Further, the CP gate exhibits a low density consistent with an open-gate conformation in the crystal structure.⁹⁵ Interaction of the proteasomal ATPases with the CP and thus gate-opening is found to be not only dependent on the conserved HbYX motif at the C-terminal of Rpt2/3/5 but is also mediated by insertion of small C-terminal α -helices on Rpt1/6.^{19,60,89,91}

A fifth state, S_5 , has also been reported in an ATP γ S supplemented proteasome sample.⁹¹ In terms of its lid and ATPase-ring arrangement, it shares high similarity with the S_2 state. However, a prominent 6 Å downwards movement of Rpt6 towards the CP, appears to induce gate opening, as the gate density is clearly absent compared to the S_2 state. The same study also reported the existence of a sixth S_6 state. The S_6 , however could only be observed in proteasomes harbouring a Walker B glutamate to glutamine mutation in Rpt3, which allows for ATP-binding but not hydrolysis in this subunit. It generally resembles the ATPase arrangement found in the S_3 state, with the exception of a clearly open gate in the CP.

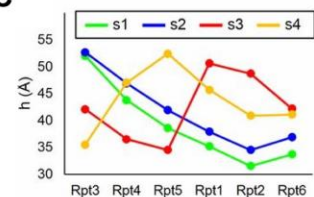
A



B



C



D

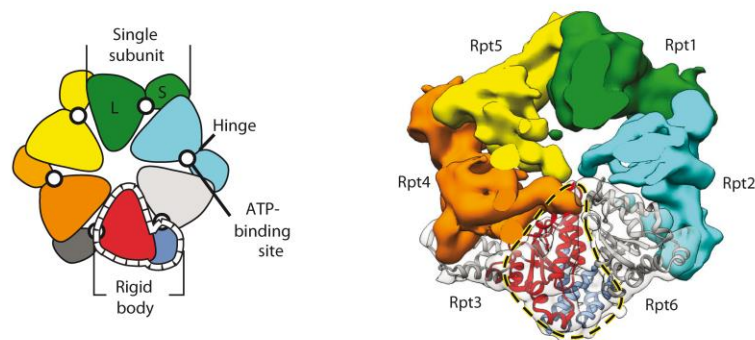


Figure 1.13 The 26S proteasome adopts 4 distinct conformational arrangements. A) Sequential progressing from the resting state S_1 to the translocation competent states S_3 and S_4 . While in S_1 and S_2 the deubiquitinase RPN11-active site, the AAA+ motor and the core particle are not aligned, they form a continuous channel in S_3 and S_4 indicative of a translocation competent state. Additionally, in S_4 the α -gate to the core particle is clearly open. During the transition from S_1 to $S_{3/4}$ the AAA+ ATPase motor undergoes major conformational changes: A deep spiral staircase arrangement in $S_{1/2}$ becomes more planar in $S_{3/4}$ (B) and the split site of the staircase ring switches from Rpt6/3 in S_1 to Rpt5/1 and Rpt4/5 in S_3 and S_4 , respectively (C). Intersubunit interactions within the spiral staircase become more uniform in state S_3 and S_4 , essentially forming rigid bodies between the small AAA+ subdomain of one Rpt with the large AAA+ subdomain of the counterclockwise neighbouring Rpt (D). Parts A), B) and D) right figure are adapted from BARD et al.²¹ C) is adapted from WEHMER et al.⁹¹ while D) left schematic is from MATYSKIELA et al.⁸³

1.5.1.3 Conformational landscape of archaeal proteasomes

Structural studies of the archaeal 19S RP analogue PAN in complex with an archaeal core particle have been long missing, due to the high structural of PAN-20S complexes.³³ To date, only two near-atomic resolution cryo-EM studies of full archaeal proteasomes have been reported.^{41,96} YU and colleagues focused on allosteric effects in the 20S core particle and achieved only a low resolution structure of PAN,⁹⁶ on the other hand, MAJUMDER et al discovered five distinct conformational states of the AAA+ motor in the PAN-20S proteasome

from *Archaeoglobus fulgidus*.⁴¹ In the presence of ATP γ S a stable complex of PAN and the 20S CP was formed. The overall structure was much simpler compared to eukaryotic 26S proteasomes since no lid or ubiquitin processing machinery is present in archaea. PAN was found to sit above the 20S core particle. The axial channels of PAN and the CP, however, were not aligned: PAN was positioned with an offset of 10 Å from the CP and PAN's axial channel was slightly inclined towards the CP. These findings resemble the misalignment found in S₁ state-structures of eukaryotic proteasomes.²¹

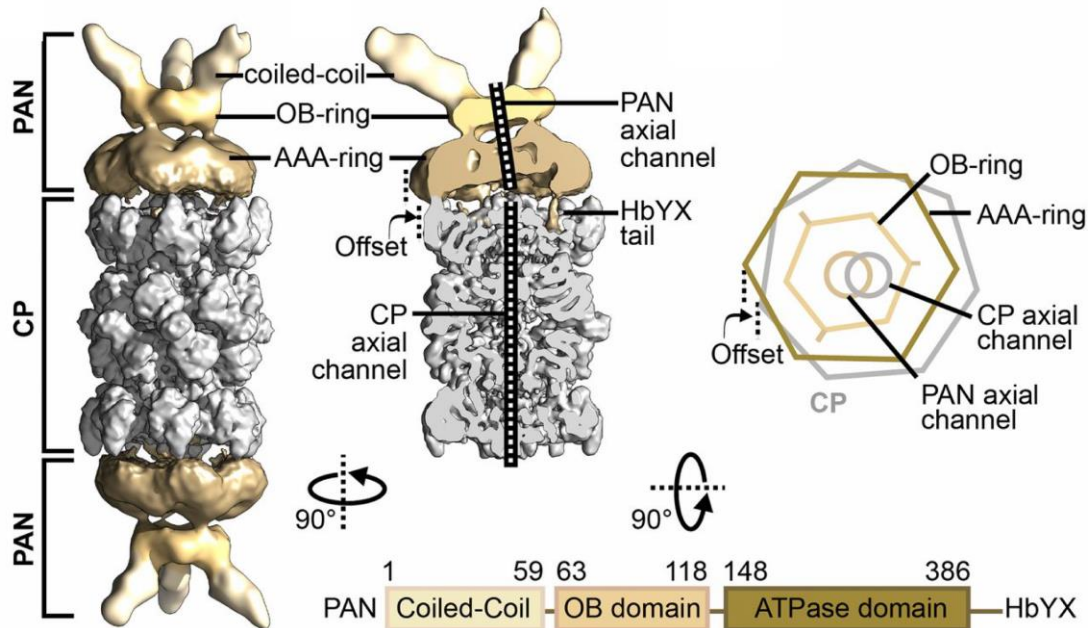


Figure 1.14 Cryo-EM reconstruction of the archaeal PAN-20S proteasome In the presence of ATP γ S the PAN-20S proteasome from *A. fulgidus* was found to assemble into doubly capped complexes (left). As for the eukaryotic 26S proteasomes the PAN ATPase motor interacts with the 20S core particle (CP) via its conserved C-terminal HbYX motif. In all identified particles, PAN sits above the CP with an 10 Å offset and the N-terminal coiled-coil and OB-ring are slightly tilted towards the CP. Consequently, OB-ring, AAA-ring and the CP axial channel are not coaxially aligned, obstructing a continuous substrate translocation channel (middle and right figure). This figure is adapted from MAJUMDER et al.⁴¹

Further features of the PAN model, however, do not allow for a definite classification according to the 26S proteasome nomenclature. While a spiral staircase adopted by the PAN AAA+ subdomains and their pore-1 loops displays a pitch similar to the steep S₁ staircase, nucleotide occupancies resemble the translocating competent S₃ state: The four highest PAN protomers in the staircase are found to be ATP- (ATP γ S-) bound, as judged by a closed binding pocket and tight intersubunit interactions. The pocket between the fourth and fifth protomer at the bottom of the staircase appears to be partially open, which is interpreted as a post-hydrolysis ADP-bound state. The last pocket at the staircase seam is clearly open and empty. Consequently, the sixth protomer is detached from the spiral staircase (Figure 1.15 A). While in 26S proteasomes the six ATPases are distinct and have consistent placements in the regulatory particle with respect to the lid and ubiquitin receptors, in PAN all ATPase monomers are equal. To differentiate the five distinct states the protomer sitting at the 10 Å offset with the 20S CP is considered as a fixed reference. The PAN structures of all five states are fairly similar, with the main difference being that in each state a different protomer reaches the top position in the staircase. In this way the five states can be arranged in an order, such that the position of ATP-hydrolysis (i. e. post-hydrolysis ADP-bound protomer) cycles through the ring, almost completing a full cycle. (Figure 1.15 B) This continuum of

nucleotide states around the cycle and the fact that the gate to the core particle is clearly open in all states fuelled later speculations that the found PAN states represent evidence for an around the cycle mechanism of ATP-hydrolysis-dependent substrate translocation⁶³ as it is also suggested based on substrate engaged AAA+ structures displaying a similar spiral staircase pattern.⁶⁵ However, the prominent misalignment of the PAN OB- and ATPase ring with the CP in all states contrasts these speculations, since current models of degradation by AAA+ proteases imply that the ATPases are aligned with their proteases to allow for substrate processing. Thus, the presented PAN models potentially represent intermediate states between a resting “offset”-state and an actively translocating state. Further studies, especially of substrate engaged PAN, will potentially solve this conflict and unravel the mechanism of activity of the evolutionary precursor to the 26S proteasome.

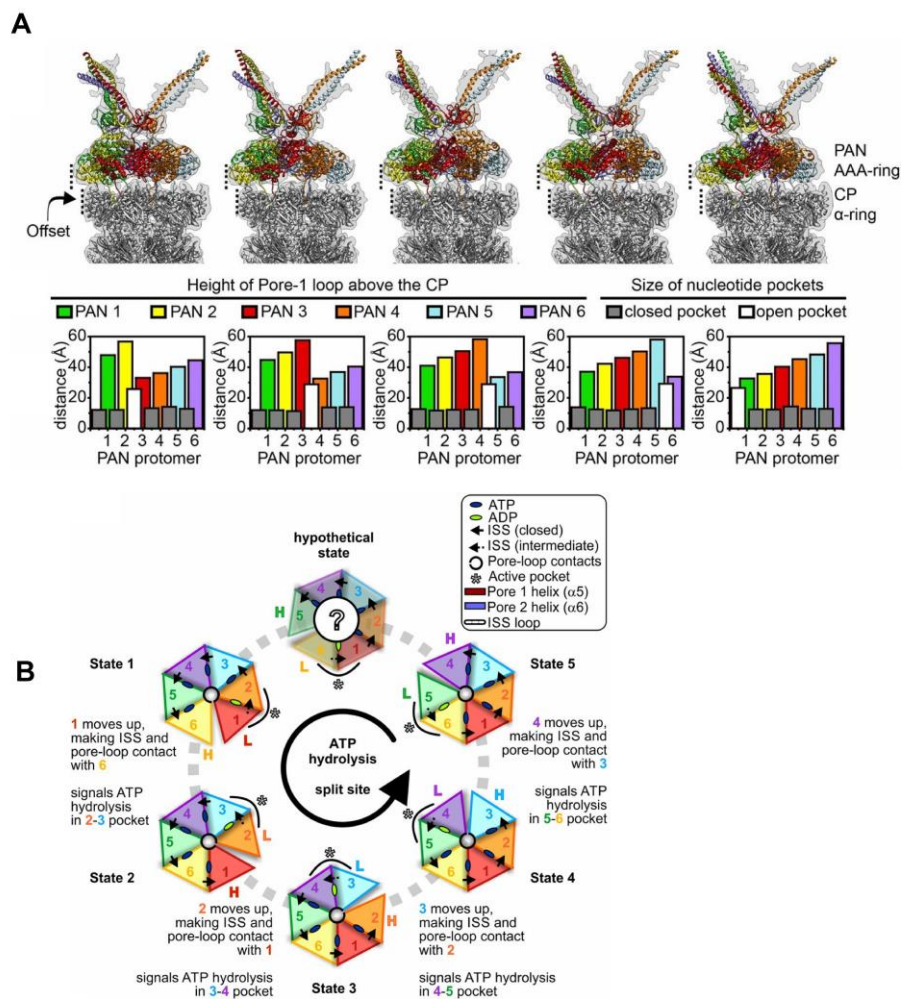


Figure 1.15 Conformational landscape and mechanism of ATP hydrolysis of the PAN-20S proteasome complex A) Setting the position of the offset as a reference, five distinct conformations of the PAN-20S proteasome can be identified. In all five states the PAN AAA+ motor is arranged in a spiral staircase, typical for hexameric AAA+ ATPases. With respect to the CP-offset site, the staircase seam (also known as split site) is found in five distinct subunits, such that the “seam” cycles continuously through the ATPase ring. Nucleotide occupancies of all five states show five bound nucleotides (4 ATP in the top four subunits and 1 ADP in the lowest subunit of the staircase, respectively). B) A model of sequential around-the-cycle ATP-hydrolysis is proposed, in which ATP-hydrolysis in the second to bottom subunit triggers the downwards movement of the spiral staircase. Consequently, the substrate binding pore-1-loops move unidirectionally from the top of the staircase towards the core particle. After ATP-hydrolysis and ADP-release has happened in the bottom subunit, it disengages the spiral staircase and reengages at the top in the next hydrolysis cycle. Figures are adapted from MAJUMDER et al.⁴¹

1.5.1.4 Mechanistic model of proteasomal substrate processing

The sequential conformational states S_1 - S_6 of the substrate-free 26S proteasome and the five PAN-20S states mentioned in the preceding subchapters present idle states of the AAA+ motor, since no external work input is applied (see chapter 1.5.1.2 and 1.5.1.3). High resolution structural studies on actively translocating proteasomes, however, have revealed further highly relevant snap-shots of substrate-processing and a mechanism of coupling ATP-hydrolysis to the translocation and unfolding of substrates could be proposed.^{60,65}

Cryo-EM analysis of substrate bound yeast proteasomes has revealed four distinct substrate-engaged conformations.⁶⁰ In this study the Rpn11 deubiquitinase was inhibited by o-phenantroline to stall the proteasome in a translocation competent and ATP-hydrolysing state.⁶⁰ The four substrate-engaged states carried a substrate polypeptide threaded through the pore from the N-terminal OB-ring entrance down to the CP α -gate, representative of an actively translocating proteasome. These four distinct states adopted a S_4 -state-like conformation in that, OB-ring, AAA+ ring, and CP were coaxially aligned, the α -gates in the CP were found to be open, and the ATPase spiral staircase had its split or “seam” on the Rpt4/5/1-face of the regulatory particle (Figure 1.16).

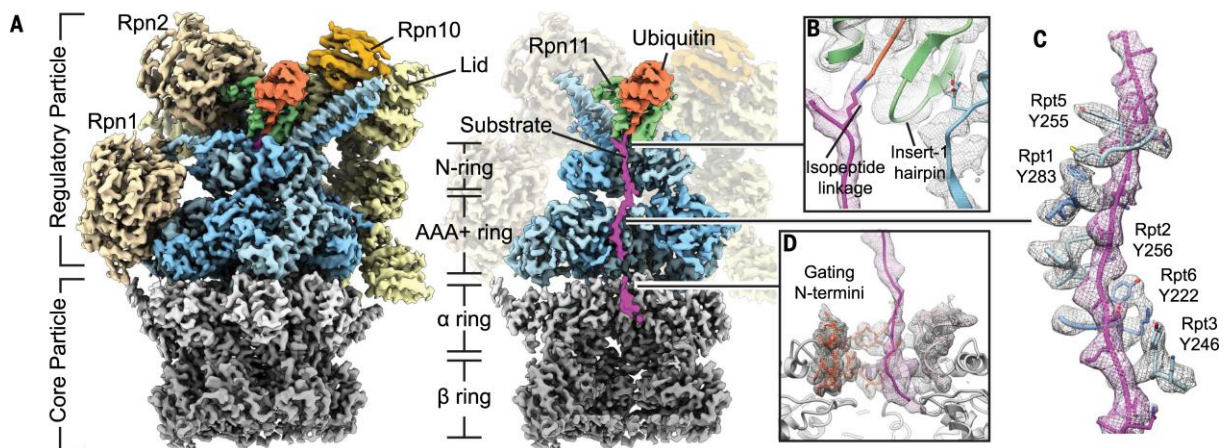
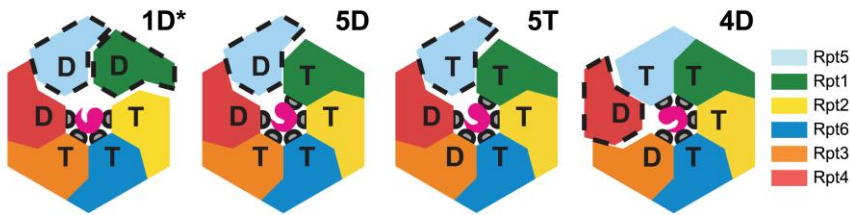


Figure 1.16 Structural model of the substrate engaged yeast 26S proteasome A) The substrate engaged 26S proteasome is found to be in a S_4 like state, with the Rpn11-active site, OB-ring, AAA+ ring and the CP-pore coaxially aligned and the α -gate open. Substrate density was found for one ubiquitin bound to Rpn11 and a substrate tail inserted into the translocation channel. B) The isopeptide bond between ubiquitin C-terminus and a substrate lysine- ϵ is visible in the active site of Rpn11. C) A substrate tail is found to be heavily engaged with pore-1-loop tyrosine of five Rpt-subunits which encircle the substrate and intercalate between side chains. D) In the S_4 like state the α -gate is clearly open and the gate termini reach towards the regulatory particle ATPases. The figure is adapted from DE LA PENA et al.⁶⁰

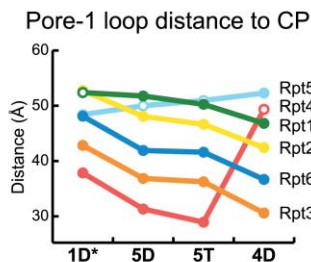
The substrate polypeptide was found to be encircled by and heavily interacting with a spiral staircase arrangement formed by pore1-loop tyrosines of five Rpt-subunits, leaving one Rpt subunit disengaged. Further, characteristic nucleotide binding pocket geometries with corresponding occupancies were identified for every subunit allowing for a placement of the different states in a sequential model for ATP-hydrolysis and translocation: DE LA PENA and colleagues named the four states based on the identity and nucleotide occupancy of the substrate-disengaged “seam”-subunit between the highest and lowest subunit in the spiral staircase. Rpt1-ADP, Rpt5-ADP, Rpt5-ATP, Rpt4-ADP were named as 1D*, 5D, 5T, and 4D, respectively.⁶⁰ In 1D* not only Rpt1 but also Rpt5 was disengaged and this state was consequently regarded as an “off-pathway” conformation. Placing the 5D, 5T and 4D states in a

sequential order revealed the consecutive processes of nucleotide hydrolysis, nucleotide exchange and conformational rearrangement of the hexameric ring, describing a full substrate translocation step (Figure 1.17):

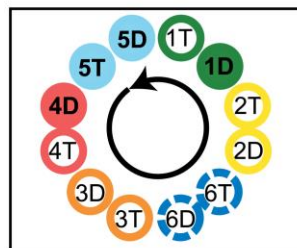
A



B



C



D

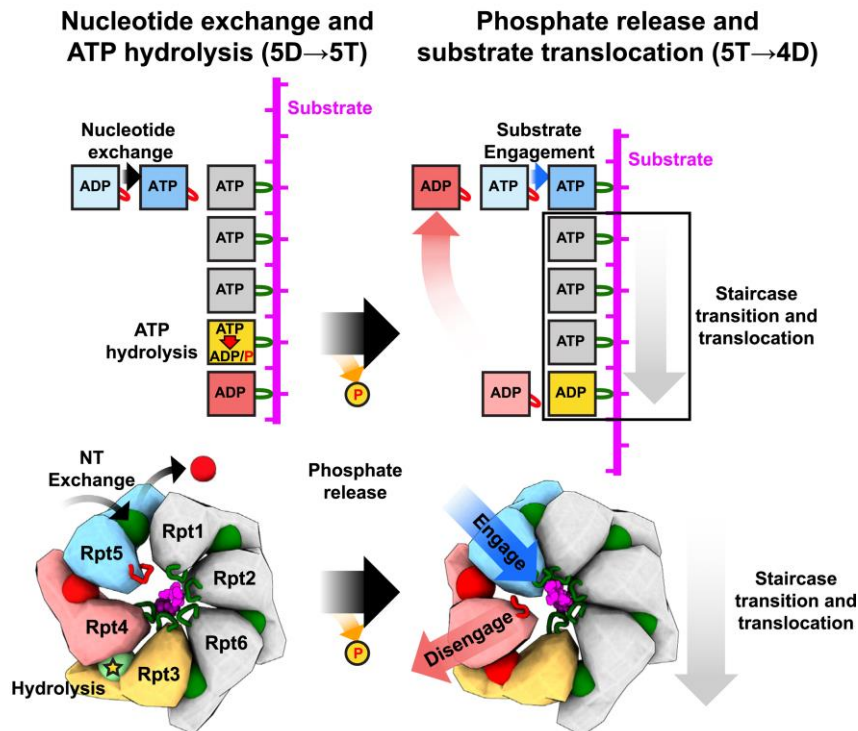


Figure 1.17 Mechanism of substrate translocation by the 26S proteasome. A) Four distinct conformational states of the 26S proteasome engaged with substrate polypeptide are discovered. Each state is defined by the nucleotide state of the Rpt which is not engaged with the substrate or the spiral staircase (T=ATP-bound, D=ADP bound). Placing the four states in sequential order describes one whole cycle of nucleotide hydrolysis, nucleotide exchange and conformational rearrangement. B) During the transition from state 5D to 4D the pore-1-loops of the substrate engaged subunits accomplish a 6 Å downwards movement, thereby translocating the substrate by two residues. C) Potentially existent states, in which the ATP hydrolysis progresses throughout the whole hexameric ring. D) A model of a full substrate translocation step is deduced from the sequentiality of the three states: ATP is hydrolysed in the second to bottom Rpt (here Rpt3). Simultaneously, ADP is exchanged for ATP in the disengaged Rpt 5 (5D → 5T). Subsequently, γ -phosphate is released from the penultimate post-hydrolysis subunit, while the bottom subunit disengages. Simultaneously, the previously disengaged ATP-bound subunit (Rpt5) engages with the substrate and the staircase at the top position while the all-ATP-bound staircase moves as a rigid body down by one register, translocating the substrate. Figures are adapted from DE LA PEÑA et al.⁶⁰

In the 5D state Rpt1 and Rpt4 are the highest and lowest substrate engaged subunits in the spiral staircase while, Rpt5 is disengaged. Rpt1/2/6/3/4 are engaged to the substrate and Rpt1/2/6/3 are ATP-bound, while Rpt4 is ADP bound. During the transition of 5D to 5T ATP is hydrolysed in Rpt3 which becomes ADP-bound (post-hydrolysis) and the ADP previously bound in Rpt5 is exchanged for ATP (though, Rpt5 is still not hydrolysis competent, since it lacks interaction with an arginine-finger of a clockwise-neighbouring subunit, which is needed for the complete closed conformation of the ATP-binding pocket, also see Figure 1.8). The 5D-5T transition does not involve huge conformational changes of the spiral staircase, while Rpt5 moves slightly up towards to the top of the staircase. The subsequent transition to the 4D state, however, entails several major rearrangements: As a result of the preceding ATP-hydrolysis the Rpt3 binding pocket opens, disrupting the intersubunit interacting with the neighbouring Rpt4, which induces the separation of Rpt4 from Rpt3 and most importantly the substrate, making Rpt4 the new ADP-bound “seam” subunit. Simultaneously, Rpt1/2/6/3, all engaged to the substrate via their pore-1 loops, move as a rigid body by one register downwards in the spiral staircase, translocating the substrate by 6 Å, which corresponds to roughly two amino acid residues (Figure 1.17 B & D). The ATP-bound Rpt5 moves to the top of the staircase, adopting a closed ATP-binding pocket and engaging with the substrate and in a newly formed rigid body of Rpt5/1/2/6 preparing the AAA+ motor for the next round of ATP-hydrolysis and translocation. Although, they only found this sequential progressing for the described combination of Rpt's, DE LE PENA and colleagues proposed that the corresponding 2T, 2D, 6T, 6D, 3T, 3D, 4T states and an additional 1T state exist completing a full cycle of ATP-hydrolysis around the heterohexamer (Figure 1.17 C).

In a structural ensemble of substrate-engaged human proteasomes DONG and colleagues reported two states E_{D1} and E_{D2} highly similar to states 5D and 4D in yeast, respectively, suggestive of a conserved mechanism in eukaryotes.⁶⁵ While the intermediate 5T state of ATP-rebinding before the large conformational change is missing in this ensemble, it additionally contained five distinct states of the proteasome, which illustrate the pathway from initial substrate-ubiquitin binding, deubiquitylation and initiation of substrate translocation. To facilitate these initial steps of substrate processing, nucleotide binding and hydrolysis patterns in the six ATPases are revealed which largely differ from the proposed model of sequential hydrolysis of a single nucleotide during processive translocation: In ATP-hydrolysis mode 1, the proteasome simultaneously hydrolyses two ATP in two Rpts opposite to each other in the AAA+ ring, to allow for widening of the ATPase pore and subsequent substrate tail insertion. This ATP-hydrolysis in opposite ATPase subunit in mode 1 is in agreement with findings by SMITH et al who proposed an ordered hydrolysis pattern for the archaeal PAN, in which ATP is hydrolysed in two subunits across the ring to each other.⁴⁸ Mode 2 is characterized by simultaneous ATP-hydrolysis in two adjacent subunits which is required for CP gating and initiation of translocation. Consequently, in this mode 2, two Rpt subunits are found to be disengaged from the spiral staircase allowing for a rearrangement in the ring to enter the subsequent processive translocation mode characterized by sequential hydrolysis one ATP at a time also found in the yeast proteasome and the archaeal PAN (mode 3).^{41,65}

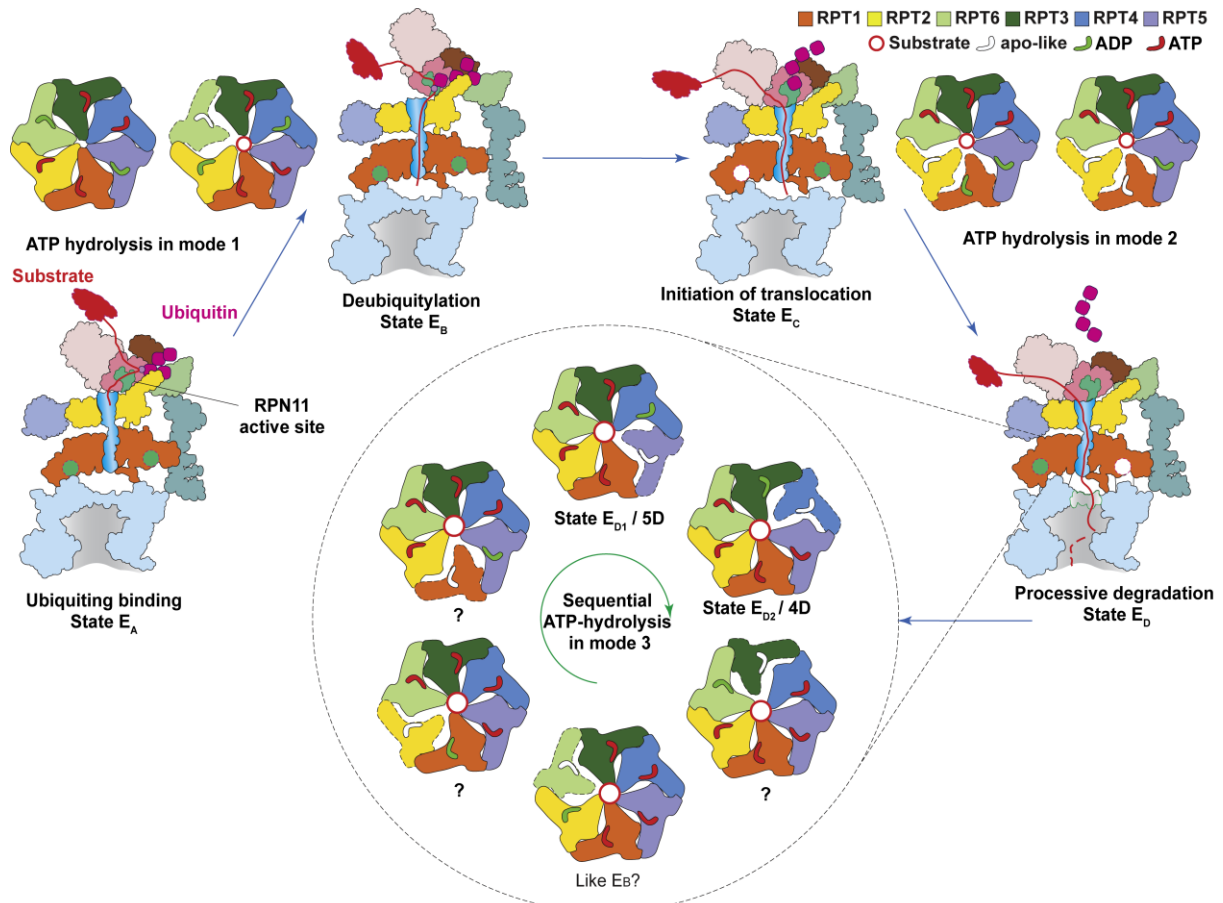


Figure 1.18 Current working model of substrate processing of the 26S proteasome. A structural ensemble of substrate engaged human 26S proteasome revealed the coordination of several critical steps prior to processive substrate translocation. The translocation incompetent proteasome in state E_A hydrolyses ATP in oppositely positioned Rpt subunits (ATP-hydrolysis mode 1) to allow for deubiquitylation and widening of the translocation pore. Subsequently, a flexible substrate tail can enter the pore and isopeptide-bound ubiquitin is cleaved. To initiate sequentially hydrolysis for processive translocation, ATP is hydrolysed in two adjacent subunits (ATP-hydrolysis mode 2). Finally, in mode 3, ATP is hydrolysed in a single subunit circles around the ring, consistent with the findings of DE LA PEÑA et al.⁶⁰ The State E_{D1} and E_{D2} in the human proteasome resemble fairly well states 5D and 4D in the yeast proteasome, respectively. The presented figure is adapted from DONG et al.⁶⁵

1.6 A conserved mechanism of translocating ATPases

Mechanistic models of the protein-unfolding and -translocation process by AAA+ unfoldases like the proteasome have been proposed based on various biochemical studies supported by structural snapshots from crystal structures,^{97,98} NMR studies,⁹⁹ cryo-EM reconstructions and single-molecule optical trapping experiments.¹⁰⁰⁻¹⁰⁵ However, a major leap towards the full understanding of how AAA+ ATPases couple ATP-hydrolysis to mechanical work was taken by the report of more than 40 high-resolution cryo-EM structures of AAA+ translocases bound to substrate polypeptide in their translocation channel.¹⁰⁶ In all structures the substrate polypeptide was found threaded through the AAA+ central pore and heavily engaged with five pore-1-loops of the ATPase subunits. Although, the described ATPases play various different functions, derive from different clades of AAA+ ATPases and have been analysed both as wild type and Walker B mutants, the arrangement of the ATPase in a spiral staircase is highly similar suggesting a common mechanism of action. While most studies only report a

single substrate-bound conformational state, a couple of structural ensembles of the same unfoldase in different consecutive states have been reported allowing the proposal of a sequential around-the-cycle ATP-hydrolysis mechanism with coordinated gripping of pore-1-loops while they move from the top to the bottom of the spiral staircase (see also Figure 1.17).^{60,65,107} However, this proposed mechanism and the claim that it is universal to all AAA+ unfoldases challenges a series of biochemical and single-molecule studies, which indicate a more probabilistic nucleotide hydrolysis mechanism.^{70,103,108} Particularly, the mechanism of the bacterial homohexameric AAA+ unfoldase ClpX has been extensively studied both structurally and biochemically,^{109,110} resulting in substrate engaged models from two different organisms and two independent research groups.^{69,70} Both studies report highly similar structures, resembling a spiral staircase arrangement of the ATPases (see also Figure 1.16 C), likewise a spiral of five tyrosine pore-loops engaging with a substrate polypeptide, rigid body formation of large and small AAA+ domains of neighbouring subdomains (see also Figure 1.13 D) and consistent nucleotide binding pattern (4-5 ATP in the 4-5 spiral staircase subunits, ADP or empty binding pocket in the disengaged subunit, see also Figure 1.17 D). However, the interpretation of these findings led to highly different models.¹⁰⁸ While RIPSTEIN and colleagues propose an around-the-cycle sequential mechanism, in which ATP is hydrolysed only in the subunit of the staircase leading to 6 Å translocation step per ATP hydrolysed,⁶⁹ FEI et al discuss their findings in the light of numerous mechanistic and biochemical studies on ClpX, previously reported from their group.⁷⁰ Especially a step size of 6 Å is not in agreement with their previous findings of ClpX basic step-sizes of ~ 20 Å and occasional kinetic burst steps of several nm length. Further, they invoke several relevant studies, in which they found that elimination of ATP-hydrolysis in 4/6 ClpX protomers did not abolish the unfolding and degradation abilities of the unfoldase.^{103,111} Similar observations have also been reported for the 26S proteasome, which was found highly defective, when ATP-hydrolysis is inhibited in Rpt3, Rpt4 and Rpt6, while inhibition of Rpt1, Rpt2 and Rpt5 only led to minor defects.¹¹² Moreover, biochemical studies on the archaeal proteasomal activator PAN indicate that binding of four ATP, as suggested by the recent structures, is rather unfavourable, implying a mechanism different from strictly sequential ATP-hydrolysis and translocation.³⁷ Although a lot of structural evidence is presented supporting a strictly sequential translocation mechanism, further functional and structural studies are highly needed, to explain biochemical findings in native, actively unfolding AAA+ ATPases. Especially, most structural techniques oftentimes report only on energetically favoured low-energy conformations, which are not necessarily the only biologically relevant conformations. To this end, biomolecular NMR-spectroscopy is a highly valuable technique, since it allows for the characterization of possibly relevant, low-populated states by means of relaxation dispersion experiments.^{113,114}

2 Methodological Background

2.1 NMR spectroscopy

2.1.1 Basic principles of NMR spectroscopy

The basic aspects of NMR spectroscopy from the following chapter have been adapted from the books *Protein NMR spectroscopy* by JOHN CAVANAGH,¹¹⁵ *NMR For Chemists and Biologists* by RODRIGO J. CARBAJO¹¹⁶ and *NMR in Chemistry* by WILLIAM KEMP.¹¹⁷

2.1.1.1 Nuclear magnetism and the origin of the NMR signal

Nuclear magnetic resonance (NMR) spectroscopy is based on the nuclear spin angular momentum, a quantum-mechanical property of nuclei of certain atomic isotopes. This property is characterized by the nuclear spin quantum number I . Depending on the nucleus I can have either positive integer, positive half-integer values or can be zero. Relevant for NMR spectroscopy are only nuclei with non-zero spin quantum numbers, such as ^1H , ^2H , ^{13}C , ^{15}N , ^{19}F and ^{31}P since NMR relies on the existence of a nuclear spin. Simple rules have been established empirically to relate the composition of a given nucleus to its nuclear spin quantum number.¹¹⁷ (1) If the numbers of protons and neutrons are both of even numbers, $I=0$. (2) If the number of protons plus the number of neutrons is odd, I is half integer ($1/2$, $1\ 1/2$, $2\ 1/2$...). (3) If the numbers of protons and neutrons are both odd I is integer (1 , 2 , 3 , ...).

The nuclear spin angular momentum I has a vector nature and is related to the nuclear spin quantum number I as follows¹¹⁵:

$$|I| = [I \cdot I]^{1/2} = \hbar[I(I + 1)]^{1/2} \quad (2-1)$$

Here, \hbar is the Planck's constant divided by 2π also called reduced Planck's constant. Due to the Heisenberg uncertainty restrictions in quantum mechanics only $|I|$ and one of its three Cartesian components can be specified simultaneously. By convention, in NMR it is the z -component of I defined as follows:

$$I_z = \hbar m \quad (2-2)$$

Here, m is the magnetic quantum number which can take values spaced by integer steps from $-I$ to $+I$. Consequently, for a spin- $1/2$ -system m can have the two discrete values of $+1/2$ and $-1/2$. Thus, the orientation of the z -component of the spin is quantised and the number of orientations is given by the relation $2I+1$. The allowed orientations of the spin's z -component are also known as the spin states. In equilibrium in the absence of an external magnetic field the spin states of a nucleus have the same energy.

As a consequence of the non-zero nuclear spin angular momentum, nuclei have a magnetic momentum. The nuclear magnetic momentum μ is given as:

$$\mu = \gamma I \quad (2-3)$$

Accordingly, the z -component of the magnetic moment is:

$$\mu_z = \gamma I_z = \gamma \hbar m \quad (2-4)$$

In equations (2-3) and (2-4) γ is a nucleus-specific constant, called gyromagnetic ratio (or sometimes magnetogyric ratio). Like the nuclear spin angular momentum I , the magnetic momentum is quantized.

In the presence of an external static magnetic field B_0 directed along the z -axis, as by convention the field in an NMR-spectrometer is, the energy of a spin state m follows the relation:

$$E_m = -\gamma I_z B_0 = -m \hbar \gamma B_0 \quad (2-5)$$

From equation (2-5) it follows that for a spin-1/2 system placed in an external magnetic field the two spin states (or spin orientations) adopt two discrete energy levels with opposite sign, the so-called high (β -) and low (α -) energy states or Zeemann-levels (Figure 2.1 B). The energy difference of the two separated levels is given by the following relation:

$$\Delta E = \hbar \gamma B_0 \quad (2-6)$$

According to the Boltzmann distribution, states of lower energy, which have their magnetic moment vector aligned parallel to the applied external magnetic B_0 field, are more populated than the high energy states with the magnetic moment aligned oppositely:

$$\frac{N_\alpha}{N_\beta} = e^{\Delta E/k_B T} \quad (2-7)$$

Here, N_α and N_β are the numbers of spins in the respective low and high energy states, k_B is the Boltzmann constant ($k_B = 1.380649 \times 10^{-23}$ J/K) and T the absolute temperature in Kelvin. Although this difference in population is relatively small it leads to a net magnetic moment as a sum of all magnetic moments of the sample. The z -component of this net magnetic moment vector is aligned in parallel with the B_0 field and is also referred to as the bulk magnetization M_z . Due to the relation in equation (2-6) the energy difference between the levels is larger the stronger the external magnetic field is. This fact is exploited in NMR spectroscopy using high field magnets to increase the sensitivity.

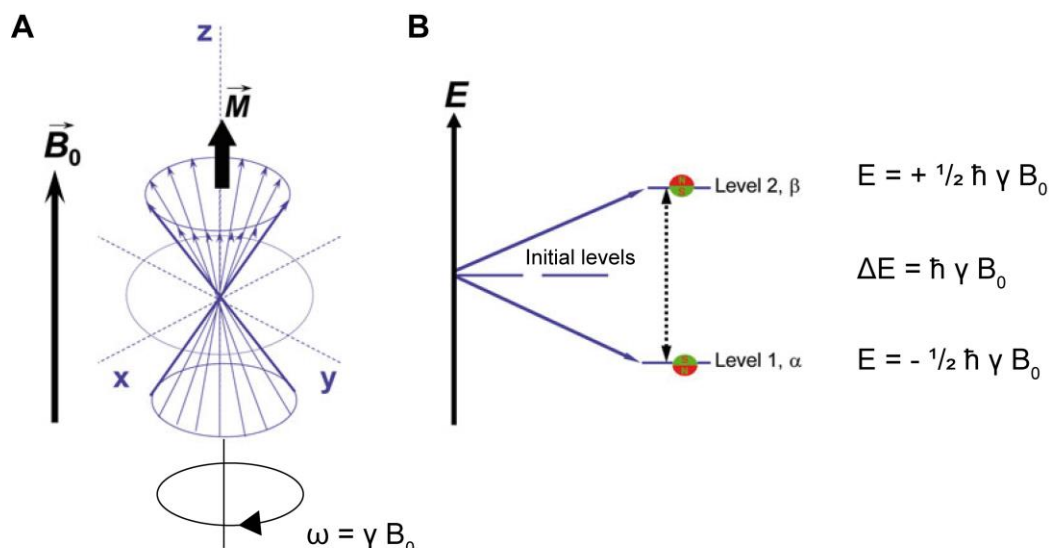


Figure 2.1 Nuclear spin states in an external magnetic field A) A classical physics vector model view on the precession of spins around the z -axis in an external magnetic field with Larmor frequency ω . The bulk magnetic moment resulting from a population difference of spins in low and high energy states is shown along the z -axis as the vector M . The x and y components of μ are unknown according to the Heisenberg Uncertainty principle and they sum up to 0 since they are uncorrelated, as indicated by the uniformly spread orientation of the spins in the xy -directions (thin arrows). Thus, $M_x=M_y=0$ at thermal equilibrium. B) The quantum mechanics representation of the separation of spin states of a spin 1/2 system in an external magnetic field B_0 . The energy and energy differences between the states are given by the equations on the side. This figure has been adapted and modified from, CARBAJO et al.¹¹⁶

In the presence of a B_0 field the bulk magnetization M precesses around the z -axis with an angular or precessional frequency known as the *Larmor frequency* ω which is directly related to the B_0 field strength and the gyromagnetic ratio γ :

$$\omega = \gamma B_0 \quad (2-8)$$

The *Larmor frequency* is given in radians per second (rad s^{-1}). Alternatively, it can be described in frequency units s^{-1} :

$$\text{Since } \nu = \frac{\omega}{2\pi} \quad (2-9)$$

$$\nu = \frac{\gamma B_0}{2\pi} \quad (2-10)$$

In equation (2-10) the *Larmor frequency* has the units s^{-1} or Hz.

In the absence of additional magnetic fields the x - and y - components of the magnetic moment μ are unknown, according to the *Heisenberg Uncertainty principle*; however, these components are not zero as $|I|^2 > I_z^2$. Because the μ_x and μ_y of individual spins are uncorrelated at thermal equilibrium, their sum is 0 ($M_x = M_y = 0$) (Figure 2.1 A).

The fundamental part of every NMR experiment is the irradiation of the sample with electromagnetic radiofrequency (RF) radiation. If the frequency of the RF radiation and thus the

energy is appropriately chosen, transitions between the different spin states can be induced. Thus, the resonance conditions are satisfied and a nucleus can transition from the low energy α -state to the high energy β -state. This change in the population balance between α - and β -states causes the bulk magnetization vector M to flip out of its collinear alignment with the z -axis; this results in the precession of the xy -component of the bulk magnetization M_{xy} around the z -axis in the xy -plane (*transverse plane*).

The transverse magnetization in the xy -plane precesses with the *Larmor frequency* ω ; a time-dependent signal which is recorded as an induced electrical current in detection coils positioned in the xy -plane consequently also oscillates with ω . The signal recorded as the time-dependent current is the *free induction decay* (FID) in the time domain. Fourier-transformation (FT) of this signal yields the typical basic NMR spectra in the frequency domain.

Equations (2-8) to (2-10) predict a single Larmor frequency for all spins of the same isotope within a given sample. However, due to the difference imposed by the chemical environment, individual spins in a molecule experience different effective magnetic fields and consequently rotate at different frequencies, according to equation (2-8). The differences in the size of the effective magnetic field at each spin position are due to the electron clouds surrounding each individual spin, which generate local magnetic fields opposing to the main field B_0 . These opposing fields are larger, the more electrons populate the cloud. Thus, each spin in a molecule feels its "own" effective magnetic field $B_{eff} = B_0 - B_{loc} = B_0 (1 - \sigma)$, and thus has its own frequency.

The absolute value of the nucleus frequency depends on the external B_0 field, meaning the higher the field strength, the higher is the resonance frequency. To obtain comparable values between spectrometers of different field strengths, the spin frequencies are by convention reported as *chemical shifts*, which is a dimensionless quantity on a ppm (parts per million) scale. The chemical shifts δ in ppm are calculated relative to the signal of an internal or external standard compound from the following equation, where ν and ν_{ref} are the frequencies of the nucleus of interest and reference signal, respectively:

$$\delta \text{ (ppm)} = \frac{\nu - \nu_{ref}}{\nu_{ref}} \quad (2-11)$$

2.1.1.2 Spin Relaxation

In NMR the spin state transitions are usually induced by short RF-pulses. Speaking in the abovementioned vector shifts model the equilibrium bulk magnetization M along the z -axis (also called *longitudinal magnetization*) is flipped away from the z -axis due to the fact that low energy spins absorb energy from the RF-field. This is equivalent to the transition of low energy spins from the low-energy α -state to the high-energy β -state. The x - and y -components of M at thermal equilibrium are 0. An appropriate RF-pulse, however, disturbs this equilibrium and generates magnetization in the xy -plane, the so-called *transverse magnetization*. The transverse magnetization then can be detected as the NMR signal (FID). However, after the RF-pulse is switched off, the system evolves to go back to the thermal equilibrium and the magnetization returns to its original state. This process is called spin relaxation and involves two fundamental processes:

- (1) The recovery of the original bulk magnetization along the z -axis is called *longitudinal relaxation*, *spin-lattice relaxation* or simply T_1 -relaxation. It is characterized by the recovery of the thermal equilibrium populations of the α - and β -states, whereby spins from the high energy β -states return to the α -states according to the Boltzmann distribution law. The corresponding energy transfer occurs through interactions with the surroundings (the *lattice*) of the spin; hence, the term spin-lattice relaxation. This relaxation process follows an exponential behaviour according to the following equation:

$$M_z = M_0 \cdot \left(1 - e^{-\frac{t}{T_1}}\right) \quad (2-12)$$

Here, M_0 is the magnetization along the z -axis at thermal equilibrium and T_1 the characteristic relaxation time. Often instead of the relaxation time T_1 the respective rate constant $R_1 = 1/T_1$ is used.

- (2) The second relaxation process describes the loss of magnetization in the xy - or *transverse plane*. It is termed the *transverse relaxation*, *spin-spin-relaxation* or simply T_2 -relaxation: In contrast to the T_1 -relaxation no energy transfer to the surroundings takes place. Simplified, a spin in the high-energy β -state changes to the α -state and the respective energy is transferred to a neighbouring spin which is moved to the β -state. As a consequence the participating spins lose their phase coherence in the xy -plane. This loss of coherence eventually results in a random orientation of all spins' transverse components and consequently in a loss of detectable NMR-signal. Like the spin-lattice relaxation also the spin-spin relaxation follows an exponential behaviour and has a characteristic relaxation time T_2 :

$$M_{xy} = M_0 \cdot e^{-\frac{t}{T_2}} \quad (2-13)$$

Both T_1 - and T_2 relaxation have important, albeit different impacts on the successful design and execution of specific NMR-experiments in the presented thesis: T_1 dictates the time until the bulk magnetization is back to equilibrium. Generally, a delay time between the repetitions of a pulse program of up to five times T_1 should be chosen to recover around 99 % of the initial equilibrium bulk magnetization along the z -axis. Thus, longer T_1 -relaxation becomes a limiting factor when fast repetition of excitation pulses is required (for fast pulsing techniques also see chapter 2.1.5). In contrast, increased T_2 relaxation results in broad linewidths of NMR-signals and thus decreased sensitivity and spectral resolution.

Essentially, the main cause for both relaxation phenomena are fluctuating local magnetic fields which stem from two main sources: *dipole-dipole interactions* (DD) and *chemical shift anisotropy* (CSA).

Dipole-dipole interactions are mutual interactions between the magnetic dipole moment of one spin with the magnetic dipole moment of a closely located spin. This interaction is dependent on the distance between the spins and on the orientation of this distance vector with respect to the external field B_0 . Due to molecular tumbling of molecules in solution, the orientation of the inter-spin vector with respect to the magnetic field B_0 is constantly altered, resulting in a fluctuating local magnetic field at each nucleus site. Fast spin-lattice relaxation oc-

curs when the oscillating frequency of the local magnetic fields is close to the nucleus' Larmor frequency. Spin-spin relaxation caused by DD is most efficient when the fluctuation frequency of the local magnetic field is slow, allowing the spins enough time to lose phase coherence. Since the molecular tumbling and its characteristic correlation time τ_c are dependent on the size of the molecule, spin relaxation is directly connected to the molecular weight of the molecule. This fact is especially important in the study of large biomolecules in solution since their slow tumbling results in fast T_2 relaxation and consequently weak and broad signals (also see chapter 2.1.2).

Chemical shift anisotropy arises from anisotropic B_{loc} in the surroundings of a given nucleus. As a consequence, the observed spin does not experience the same magnetic field from all cartesian directions and the chemical shift is of tensorial nature. Due to rapid molecular tumbling of the molecules, the chemical shift anisotropy tensor reorients with respect to the external magnetic field, which causes a time-dependent variation of the chemical shift and thus a loss of coherence, which results in enhanced relaxation. Simultaneously, the CSA contributions from the three dimensions are averaged resulting in a single measured isotropic chemical shift. The strength of the locally induced magnetic fields is proportional to B_0 and the resulting relaxation due to CSA is proportional to the square of the locally induced fields.

2.1.2 NMR of large protein complexes

Historically, NMR spectroscopy was routinely applied to the structural characterization of small molecules and biomolecules up to 10 kDa. Going towards NMR studies of larger proteins came with two general challenges: (1) NMR spectra get more crowded with increased molecular weight, simply due to the higher number of measurable spins. As a consequence spectral overlap can be a limiting factor. (2) A higher molecular weight correlates with slower molecular tumbling, longer rotational correlation times τ_c and consequently increased spin-relaxation via CSA and dipole-dipole interactions as discussed in chapter 2.1.1.2. Thus, with increasing molecular size, the acquisition of high-quality NMR spectra becomes challenging. Several developments, however, have made NMR spectroscopy of larger proteins possible. First, uniform isotopic labeling of proteins, using nutrients containing the naturally low-abundant spin-1/2 isotopes ^{13}C and ^{15}N as the sole carbon- and nitrogen source during the over-expression of proteins in cell cultures, together with three-dimensional assignment experiments (see chapter 2.1.3), have resulted in sufficient sensitivity and resolution, respectively, to access structures of proteins up to 30 kDa by NMR. Further improvements towards high-resolution studies of even bigger protein samples could be made by the (partial) replacement of protons by deuterium (^2H). This reduction in proton density within a protein sample leads to significantly improved relaxation properties, since relaxation mechanisms based on dipolar interactions of nearby protons are reduced. This in turn leads to higher sensitivity and resolution. Furthermore, the six times smaller gyromagnetic ratio γ of ^2H compared to ^1H reduces the relaxation of the heteronuclei ^{15}N and ^{13}C spins attached to ^2H and also results in narrower line-width and thus better spectral quality. The information loss due to the reduced proton number in deuterated samples is usually compensated by the gain in resolution and sensitivity.

For even bigger proteins, NMR-methodology has been developed to exploit phenomena that reduce the transverse (T_2) relaxation for a subset of NMR-lines.¹¹⁸ TROSY-NMR (Transverse Relaxation-Optimized Spectroscopy) relies on the interference between the two major relaxa-

tion mechanisms of chemical shift anisotropy and dipole-dipole coupling, a phenomenon termed cross-correlated relaxation. In a 2D correlation spectrum of a heteronuclear two-spin system without heteronuclear decoupling, the NMR-signal is split into four multiplet components due to the heteronuclear coupling between the two spins when decoupling is not applied. The different multiplet components of these four peaks have different T_2 relaxation times and thus different line-widths and intensities, due to different contributions of the CSA, DD cross-correlated relaxation to them. This contribution can either add to or subtract from the contributions of DD and CSA relaxation in itself: In the slowest relaxing multiplet component, the contribution of the DD and/or CSA interactions (auto-correlated relaxation) are almost cancelled by the contribution of the interference between DD and CSA interactions (cross-correlated relaxation), leading to slower T_2 and thus improved spectral properties. In a standard 2D correlation experiment, the four lines are mixed by decoupling, resulting in a single peak, where the contribution of cross-correlated relaxation is averaged out. The TROSY approach does not include such decoupling and instead selectively retains only the component for which the cross-correlated relaxation counteracts the auto-correlated relaxation. While dipole-dipole interactions are independent on the external B_0 field, CSA increases proportional with $(B_0)^2$. Consequently, almost perfect destructive interference occurs at a field strength at which dipole-dipole interactions and CSA are of similar magnitudes. For amide ^1H - ^{15}N TROSY experiments, this field strength is around 900 MHz.

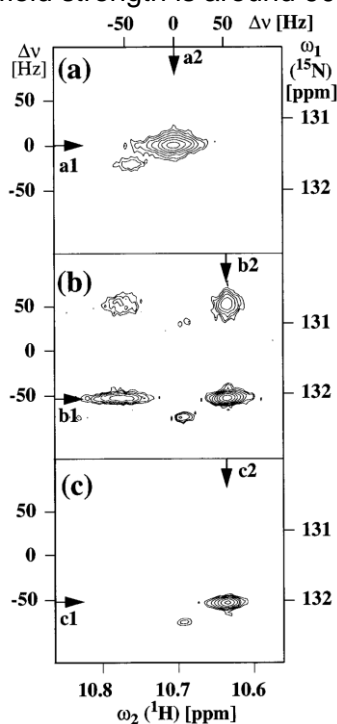


Figure 2.2 The TROSY principle. A) A conventional ^1H , ^{15}N - correlation spectrum with decoupling displays only a single peak for one amide. B) In a non-decoupled version the four multiplet components of one amide appear with different line-shapes and intensities due to different contributions of the CSA, DD cross-correlated relaxation to them. C) A ^1H , ^{15}N -TROSY spectrum of the same amides as in A) and B) only display a single peak with improved line-shape and intensity compared to the peaks in A) due to selectively retaining only the slowest relaxation multiplet component from B). This figure has been adapted from Pervushin et al.¹¹⁸

While the above-mentioned TROSY principle was originally developed for ^1H - ^{15}N spin systems, its extension to ^{13}C $^1\text{H}_3$ spin systems of isolated methyl groups has made high molecular weight proteins and complexes accessible by NMR. First introduced by the Kay group in 2003,¹¹⁹ it exploits several advantages of the ^{13}C $^1\text{H}_3$ spin system of methyl groups. (1) Enhanced sensitivity arises from the three chemically equivalent protons, which cause a three

times higher magnetization at the beginning of a pulse sequence and thus increased signal. (2) The rapid free rotation of methyl groups at the end of an amino acid moiety partially decouples their local rotational correlation time τ_c from the long τ_c of the large biomolecule. Transverse relaxation in such methyl group spin systems is dominated by dipole-dipole interactions and the chemical shift anisotropy contributions are very small. Thus, the methyl-TROSY effect does not arise from destructive interference between DD and CSA interactions but rather from the interference of ^1H - ^1H and ^1H - ^{13}C dipole-dipole interactions.¹²⁰ Deuteration of the protein except for the methyl groups is required to reduce dipolar interactions of the methyl group hydrogens with external, non-methyl ^1H -spins.

The successful execution of methyl-TROSY NMR experiments thus relies on the preparation of samples containing ^1H and ^{13}C labeled methyl groups in an otherwise highly deuterated background. Such labeling can be achieved by growing bacterial cultures in a medium containing a deuterated carbon source (either ^2H -glucose or ^2H -glycerol) and pure D_2O as a solvent. Before induction of protein expression, the supplementation of selectively ^1H , ^{13}C -methyl-labelled precursors for different methyl-carrying amino acids (Ile, Leu, Val, Met, Thr and Ala) allows for the selective incorporation of $^{13}\text{C}^1\text{H}_3$ probes into the recombinant protein.¹²¹ The relatively simple pulse sequence of a ^1H ,- ^{13}C HMQC 2D correlation experiment is already optimally prepared to exploit the described TROSY-effect in methyl groups. In summary, the methyl-TROSY approach is based on recording ^1H , ^{13}C HMQC 2D correlation experiments on highly deuterated proteins containing selectively ^1H ,- ^{13}C -labelled methyl groups.

Since backbone-directed assignment experiments (chapter 2.1.3) fail on large proteins or protein complexes due to the unfavorable relaxation behavior of backbone amides, assignment of methyl resonances follows a so-called “divide and conquer” approach: The large biomolecule or complex is physically divided into smaller, stable fragments (subcomplexes, or subdomains), which are accessible by standard NMR experiments. After assignment of methyl resonances in these smaller fragments the assignment is transferred to the large biomolecule or complex, which is then studied by TROSY-NMR. This strategy has been applied to several biological systems.^{25,27,122}

2.1.3 NMR assignment strategies for proteins

Assignment of spin resonances is a crucial step in each NMR study. At the beginning of every protein-NMR study the first spectrum acquired is a 2D correlation of ^1H - ^{15}N spin systems termed ^1H - ^{15}N HSQC (Heteronuclear Single Quantum Coherence). A ^1H - ^{15}N HSQC is considered as a “finger-print” of each protein sample and is usually recorded on a uniformly ^{15}N -labelled protein. It displays a peak for every proton attached to a nitrogen-15, such as backbone amides and side chain amides. Side chain amides do not always give rise to peaks in the ^1H - ^{15}N HSQC since their protons often rapidly exchange with water and are thus not detected. Prolines are also invisible, as they do not have an amide proton.

Based on the chemical shift dispersion (that is degree of spread of resonances along both ppm scales), this spectrum provides an estimate of the degree of foldedness of the protein. Further, the number of peaks inform on the completeness of the spectrum, while their shape information on the feasibility of the assignment.

The initial step is the assignment of the backbone amide resonances of the ^1H - ^{15}N HSQC to individual amino acids in a process termed sequential backbone resonance assignment. For this, a suite of 3D experiments has been established which allow to connect the backbone amide of one amino acid (i) with the backbone resonances of the preceding amino acid residue ($i-1$) (Figure 2.3). A common combination is the acquisition of the triple resonance experiments HN(CO)CA and HNCA on uniformly ^{15}N - and ^{13}C -labelled protein samples. The HN(CO)CA correlates the amide resonance of one amino acid (HN) with the C^α -carbon resonance of the preceding amino acid (one peak in the carbon-dimension of the 3D spectrum); it is an “inter-type” spectrum. The HNCA, in turn, correlates the amide of one amino acid with both the C^α of the same residue and that of the preceding residue (two peaks in the carbon-dimension); it is an “intra-type” spectrum. By matching C^α chemical shifts from intra- and inter-type spectra, one can generate sequential connectivities among most backbone amides in a protein.

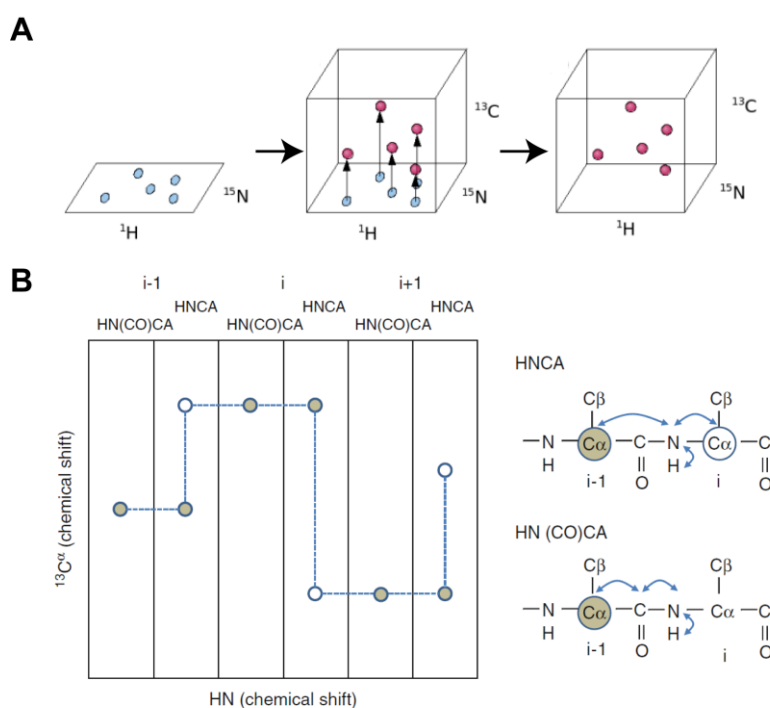


Figure 2.3 Principles of triple resonance backbone assignments A) Schematic of the introduction of a third dimension in 3D NMR-experiments: A ^1H , ^{15}N HSQC is represented by the spectrum on the left with blue peaks. In 3D spectra for backbone resonance assignment a third dimension (^{13}C) is introduced resulting in a three-dimensional resolution of the peaks in the ^1H , ^{15}N HSQC. B) The schematic principle of sequential backbone assignment is exemplified by the combination of the HN(CO)CA and the HNCA 3D experiments. The HNCA correlates the backbone amide with the C^α of the same (i) residue and the C^α of the preceding ($i-1$) residue resulting in two 3D-peaks in the 3D experiment at the NH position of residue i (scheme on the left): One with the $^{13}\text{C}^\alpha$ shift of the same residue and one with the $^{13}\text{C}^\alpha$ shift of the $i-1$ residue. The HN(CO)CA in turn correlates the amide at position NH, only with the $^{13}\text{C}^\alpha$ of the preceding $i-1$ residue resulting in a single peak with $^{13}\text{C}^\alpha$ shifts of the preceding residue. By matching the $^{13}\text{C}^\alpha$ shifts in the HN(CO)CA with $^{13}\text{C}^\alpha$ in the HNCA a sequential connectivity can be established and the ^1H , ^{15}N shifts of the preceding residue can be identified. Scheme A) is taken from protein-nmr.org.uk (retrieved on 28.02.2021) and scheme B) is adapted from INAGAKI.¹²³

Respective combinations of inter- and intra-type spectra are available to correlate the amide with C^γ (HNCO and HN(CA)CO)) and C^β (HN(COCA)CB and HN(CA)CB or HN(CO)CACB and HNCACB) spin. Experiments with magnetization transfer to the C^β carbon are particularly useful since they allow for the straight forward identification of several amino acid types via their unique C^β shifts (for example Ala, Thr, Ser). For large proteins, as presented in this thesis, deuteration helps to improve sensitivity, especially by reducing ^{13}C relaxation due to dipole-dipole interaction with attached protons. Furthermore, the TROSY-principle (see chapter

2.1.2) has been applied to the suite of backbone resonances described above, thus reducing transverse relaxation of the backbone amide nitrogen-15 and expanding the feasibility of backbone assignment on $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ -labelled samples of 30-50 kDa in size.^{124,125}

Once the backbone resonances have been identified, they can be connected to the amino acid side chain resonances. Specifically, spectra of the type H(CCCO)NH connect the ^1H - ^{15}N chemical shift of amino acids *i* with the H^α and further side chain proton shifts of the amino acid *i-1*. Together with the HCCH-TOCSY, one can obtain side chain proton and carbon chemical shifts. Further assignments can be achieved by recording several NOE-based 3D experiments.

Assignment of side-chain resonances in large proteins, is more challenging since the proteins are typically deuterated for improved relaxation properties and thus do not contain side-chain protons. For methyl-TROSY studies on large perdeuterated proteins a different approach has been developed to assign methyl ^1H and ^{13}C resonances: After backbone resonance assignment using TROSY-based experiments, ideally the C^α and C^β shift of most amino acids should have been identified. Several assignment experiments have been developed to create connectivity between the backbone resonances and the terminally protonated and ^{13}C labeled methyl groups of leucine (δ -methyls), valine (γ -methyls) and isoleucine ($\delta 1$ -methyl)¹¹⁹: Triple-labeled (^2H , ^{13}C , ^{15}N samples) with selective methyl protonation are prepared with the addition of α -keto-acid amino acid precursors for Ile, Leu and Val during the bacterial over-expression. These are deuterated, selectively ^1H , ^{13}C -labelled at one terminal methyl group and ^{13}C -labelled along the main chain (Figure 2.4 A & B). Isoleucine residues in the resulting sample are thus ^{13}C -labeled and entirely deuterated except for the terminal $\delta 1$ -methyl group, which is protonated. For Leu and Val residues only one of the δ - or γ -methyl groups is ^1H - ^{13}C -labelled to avoid the relaxation contribution from the dipole-dipole interaction between the protons of the geminal methyl groups (Figure 2.4 C). The carbon attached to the ^2H -methyl is non ^{13}C -labelled to eliminate unwanted pathways of magnetization transfer.

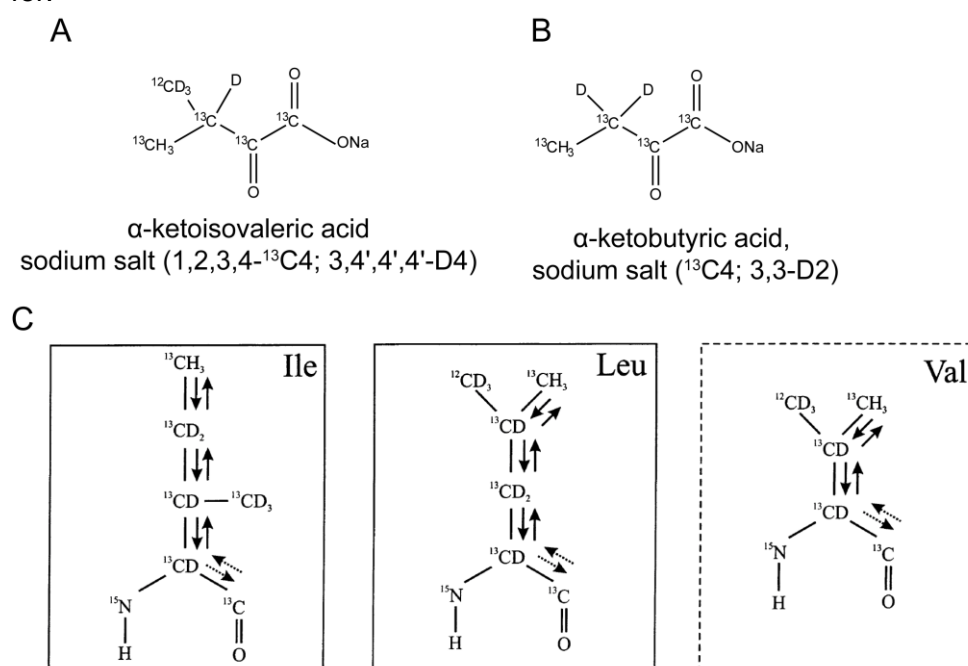


Figure 2.4 Labeling schemes and magnetization transfer steps for methyl assignment experiments in deuterated proteins. A) and B): Selectively protonated and ^{13}C -labeled amino acid precursors are used in the preparation of triple-labeled assignments samples. For leucine and valine assignments only one of the terminal methyl groups is ^1H and ^{13}C labeled while the second one is deuterated and ^{12}C -labeled; this is necessary to

ensure linear transfer of magnetization as described in C). The ILV-methyl resonances can be correlated to the C^α and C^β resonances obtained from earlier backbone assignments experiments employing a methyl-detected “out-and back” Ile, Leu-HMCM(CG)CBCA experiment for which the intra-residue magnetization flow is described by arrows in C). Correlation to the backbone C' is equally possible using Ile,Leu-HMCM(CG)CBCA)CO and Val-HMCM(CBCA)CO pulse schemes which transfer the magnetization further to the backbone C' , as indicated by the dotted arrows. Scheme C) has been adapted from TUGARINOV et al.¹¹⁹

2.1.4 Protein dynamics by NMR spectroscopy

One of the main advantages of biomolecular NMR spectroscopy is the ability to probe, in addition to the static structure, the dynamic processes of biomolecules, which are usually relevant for biological function. Looking at a protein as a single static structure oversimplifies the reality since proteins exhibit intrinsic dynamics, such as internal motions and conformational exchange processes that take place on a wide range of time scales ranging picoseconds (ps, 10^{-12} s) to seconds (Figure 2.5). NMR is perfectly suited to study these processes as it probes dynamics in a site-specific manner. In this chapter I focus on two selected protein-NMR dynamic techniques, which have been used in the present thesis work. First fast local dynamics on the pico- to nanosecond time scale can be revealed by studying the relaxation of the backbone amides. Second, CPMG-relaxation dispersion experiments reveal exchange processes that occur on slower time scales between high-microseconds and milliseconds. These slow dynamics processes are usually characterized by the exchange between at two or more distinct states, for example two different conformations of a protein or the ligand-bound and free forms.

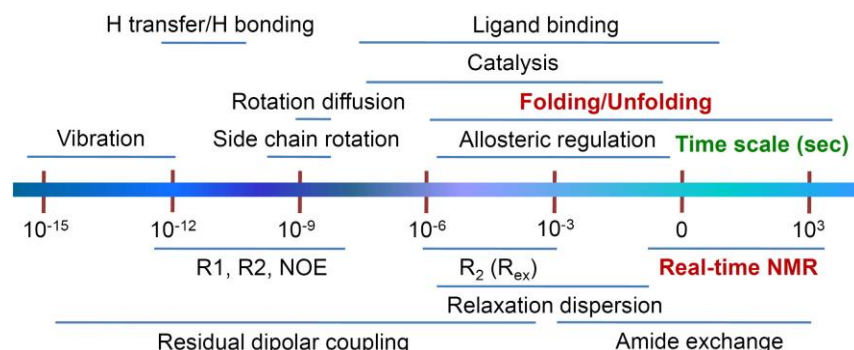


Figure 2.5 Dynamic processes in proteins and the respective techniques to study them. Adapted from Balbach et al.¹²⁶

2.1.4.1 Fast backbone dynamics from T_1 -, T_2 -relaxation experiments and hetNOE

A basic set of experiments for the characterization of fast backbone dynamics measure relaxation parameters for the backbone nitrogen-15 of each amide group, such as the spin-lattice- (T_1 or longitudinal) and spin-spin (T_2 or transverse) relaxation times (or their respective rate constants $R_1 = 1/T_1$ and $R_2 = 1/T_2$) and the heteronuclear Nuclear Overhauser Effect ($\{^1\text{H}\}$ - ^{15}N NOE, or here hetNOE).¹²⁷ The relaxation properties of ^{15}N in this regard are especially relevant as they mostly depend on dipolar interactions with the attached amide proton (also see chapter 2.1.1.2). Due to high frequency local motions in the protein backbone, which lead to fast reorientation of the ^1H - ^{15}N dipolar interaction with respect to the external B_0 -field, the relaxation of the ^{15}N is modulated. Thus, relaxation rates are excellent reporters

of high frequency molecular motion in the pico- to nanosecond time scale (that is up to the global correlation time of the molecule). The value of the hetNOE already gives a good initial estimate of the local flexibility and of the presence of secondary structure, since unstructured or unfolded regions display hetNOE of opposite signs compared to structured and folded core regions.¹²⁸ At this point it has to be noted, that measured R_2 values can also contain contributions from chemical exchange, which happens on much slower time scales (see chapter 2.1.4.2 below)

The two relaxation rates and the hetNOE parameter of each sampled amide give a good qualitative estimate on internal motions. However, they can also be related to quantitative dynamics in a framework developed by LIPARI and SZABO termed the *model free approach* because its application does not require the assumption of a specific model for internal motions.¹²⁹ The results of this analysis are site-specific quantitative order parameters S^2 . S^2 can take values between 0 (unrestricted motion, high flexibility) and 1 (highly restricted motion, totally rigid). Residues located in secondary structural elements, like helices, usually have values of $S^2 > 0.8$ while loops, turns and termini have reduced values of $0.5 < S^2 < 0.8$.¹³⁰ Apart from the local dynamics the described backbone relaxation experiments also allow for the estimation of the global parameters of molecular tumbling, the rotational correlation time τ_c which is directly related to the average ratio of R_1/R_2 of all sampled amides.

2.1.4.2 Chemical Exchange processes studied by relaxation dispersion experiments

Protein dynamics on the micro- and milliseconds time scale are typically studied using relaxation dispersion experiments. At this point it is useful to introduce the concept of chemical exchange since it is the basis of these fast to intermediate processes: A given spin in a protein can experience more than one chemical environment due to processes like conformational changes, ligand binding, protein folding and secondary structure rearrangements.¹³¹ Each two-site exchange process is characterized by several parameters which are interrelated (Figure 2.6): Considering two states A and B , the exchange between the two states is characterized by the velocity of the interconversion from state A to B and of the reverse process from B to A with respective rate constants k_A and k_B . Further, the two states have a relative population P_A and P_B , for example 75 % and 25 %. The exchange rate k_{ex} between the two populations is given as the sum of both k_A and k_B . From Figure 2.6 it is obvious that also the chemical shift difference between the two states $\Delta\nu$ determines the appearance and most importantly the line width of the respective NMR spectra. Three major scenarios, termed *exchange regimes* can be distinguished: (1) *Slow exchange*: If the exchange rate k_{ex} is much smaller than $\Delta\nu$, the spectrum contains two separate peaks at ν_A and ν_B with relative intensities proportional to their relative population. This originates from the fact that no significant exchange processes take place in the time needed to measure the chemical shift difference of the two species A and B . (2) *Intermediate exchange*: If $k_{ex} \approx \Delta\nu$ both peaks are broadened due to the exchange contribution R_2^{ex} to the observed overall transverse relaxation rate R_2^{Obs} . In this regime, depending on the exchange rate, peaks can even become undetectable due to severe broadening. (3) *Fast exchange*: In the fast exchange regime k_{ex} is larger (faster) than $\Delta\nu$. Consequently the nuclei exchange many times during the time needed to measure the chemical shift difference of the two species A and B and as a result one single peak appears at a chemical shift which is a population-weighted average of the chemical shift of the two states A and B .

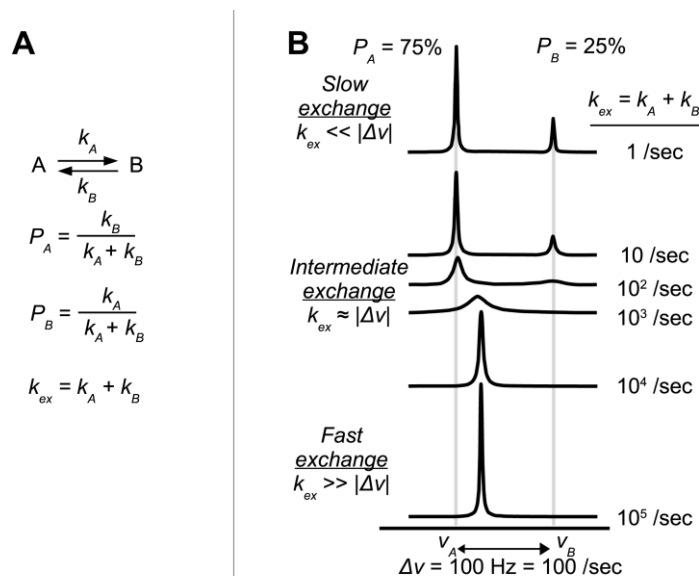


Figure 2.6 Chemical exchange. The basis for peak line shape and intensity of a two-site exchange system. A) The exchange rate k_{ex} is directly related to the interconversion rates k_a and k_b . B) Depending on the rate of chemical exchange relative to the chemical shift difference and the relative population of the two states, the spectrum shows two resolved peaks (slow exchange), broadened peaks (intermediate exchange) or a single peak with a chemical shift equal to the population-weighted average of the chemical shifts of the two exchanging states. This figure is adapted from KLECKNER et al.¹³²

Exchange processes in the intermediate regime are biologically interesting since they arise from events like binding and dissociation of ligands or conformational changes of protein domains. Further, if the population between the states is highly skewed, that is when $P_A \gg P_B$ the existence of the state B might not be directly visible in the spectrum but can be detected based on its exchange with the higher populated state. Relaxation dispersion experiments employing the CPMG- approach (Carr-Purcell Meiboom-Gill) allow for the determination of the exchange contribution R_2^{ex} to the observed transverse relaxation rate R_2^{obs} and thus help identifying and quantifying chemical exchange processes on the intermediate time scale from micro- to milliseconds. Using this approach, for example, it is possible to detect the kinetics of an exchange process (exchange rate), its thermodynamic parameters (i.e. the relative populations of the exchanging states) and the chemical shift difference between the ground (A) and the excited state (B) populated, even when P_B is as low as 1 %.¹³³

The general goal of CPMG relaxation dispersion experiments is the quantification of the contribution of exchange to the transverse relaxation rate R_2 . Transverse relaxation is the result of dephasing of the transverse magnetization during the evolution time of an NMR experiment, which essentially leads to loss of detectable signal in the xy -plane (see also chapter 2.1.1.2). Chemical exchange provides an extra contribution to this dephasing, in addition to the contribution from dipolar interactions and chemical shift anisotropy discussed in chapter 2.1.1.2. The central element of a CPMG experiment is a spin-echo pulse train which intends to suppress the exchange contribution to R_2 . The spin-echo principle is illustrated in Figure 2.7. After a time τ , a 180° pulse along the x -axis flips the transverse magnetization of two dephased signals negating their phase difference. Due to this flip the two signals regain phase coherence after a second time of duration τ . In the presence of chemical exchange with a rate faster than $1/\tau$, a spin does not have the same chemical shift in the two τ phases and thus coherence is not reached after 2τ , resulting in line-broadening and reduced signal intensity. The contribution of chemical exchange to the dephasing process is suppressed when $1/\tau \gg k_{ex}$.

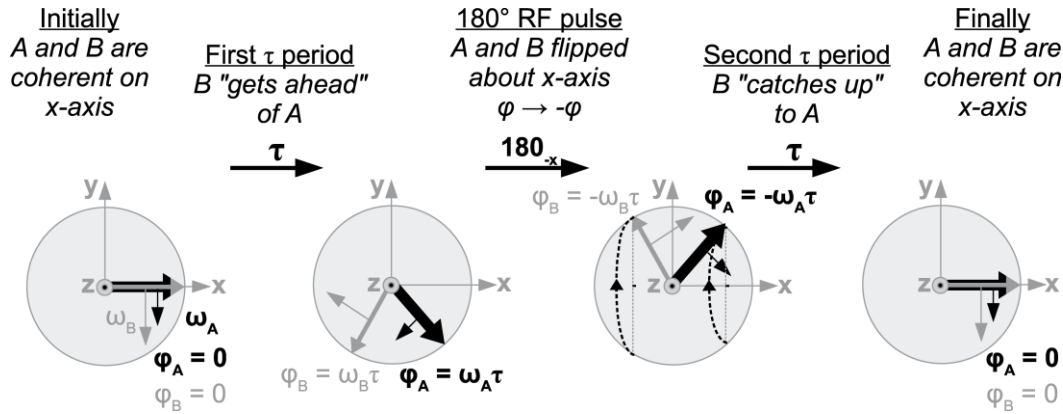


Figure 2.7 Spin-echo principle as the basis for a CPMG relaxation dispersion experiment. After an initial excitation pulse transverse spin magnetization of states A and B are in-phase. During an evolution time τ they lose their coherence and the spin of state B is ahead of the spin of state A due to their chemical shift difference. By applying a 180 ° pulse along the x -axis their magnetization is flipped bringing state A ahead of state B. After another period τ they regain their phase coherence. This figure is adapted from KLECKNER et al.¹³²

In a CPMG experiment the time τ is incrementally changed from one experiment to the other; all experiments contain spin-echo steps within the fixed CPMG relaxation delay time T_{CPMG} . The CPMG frequency $\nu_{CPMG} = 1/(4\tau)$ is a function of the time τ (Figure 2.8 A). The extent of suppression of the exchange contribution to the dephasing of state A and B during the spin-echo time T_{CPMG} is different in the different experiments, as it depends on the relative values of $1/\tau$ and k_{ex} . Thus, the R_2^{ex} is dependent on ν_{CPMG} . On the other hand, the contribution to the dephasing due to transverse spin relaxation mechanisms (excluding chemical exchange) R_2^0 does not depend on $1/\tau$.

For every ν_{CPMG} an effective transverse relaxation rate R_2^{eff} (also referred to as R_2^{obs} = observed R_2) can be extracted from the following equation where I_0 and I_1 are the peak intensities of a given spin systems in a reference plane with no relaxation delay ($T_{CPMG}=0$) and of the experimental plane at ν_{CPMG} , respectively:

$$R_2^{eff}(\nu_{CPMG}) = -\frac{1}{T_{CPMG}} \ln \frac{I_1(\nu_{CPMG})}{I_0} \quad (2-14)$$

As the exchange contribution to the R_2^{eff} rate, R_2^{ex} , is dependent on ν_{CPMG} also R_2^{eff} is dependent on ν_{CPMG} . The plot of R_2^{eff} against ν_{CPMG} shows the characteristic relaxation dispersion curve. The maximum contribution of the exchange to the broadening is obtained when $1/\tau=0$ and is given by (Figure 2.8 C):

$$R_{ex} = R_2^{eff} - R_2^0 = R_2^{eff}(\nu_{CPMG}^0) - R_2^{eff}(\nu_{CPMG}^\infty) \quad (2-15)$$

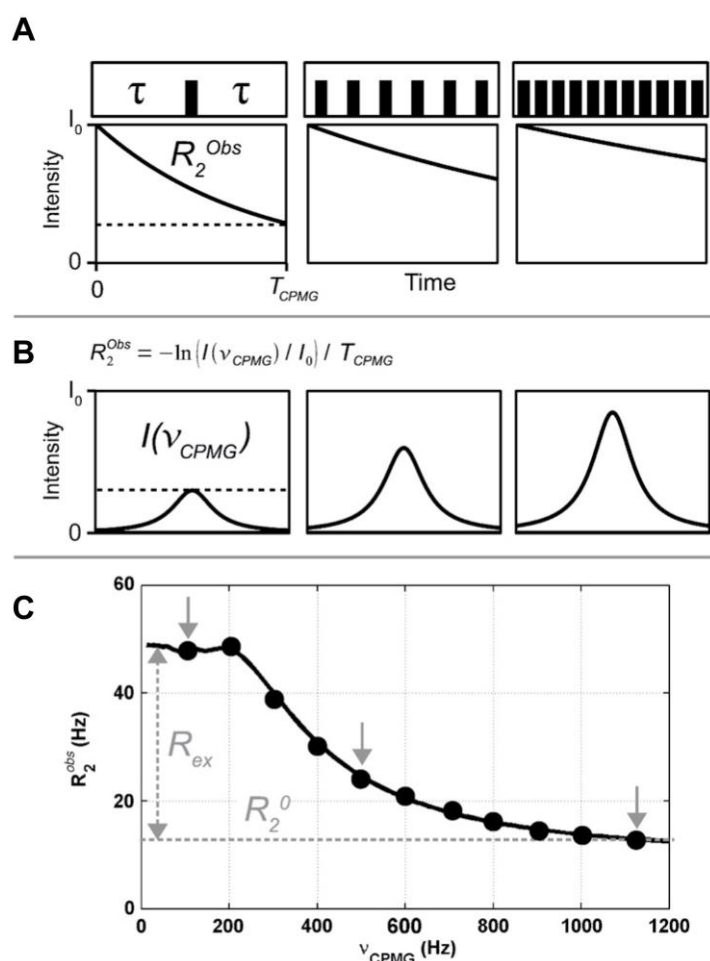


Figure 2.8 CPMG-relaxation dispersion experiments. A) A spin-echo sequence is repeated with increasingly increased repetition rate of the 180 degree pulses during T_{CPMG} , which manifests in the CPMG frequency ν_{CPMG} . B) For each ν_{CPMG} , an observed relaxation rate R_2^{obs} (or R_2^{eff} , for effective R_2) can be extracted from equation (2-14) based on the peak intensities at ν_{CPMG} and peak intensities where no relaxation delay is applied. C) Plotting the R_2^{obs} against ν_{CPMG} provides the characteristic relaxation dispersion profile, from which the exchange contribution to the observed relaxation rate can be extracted according to equation (2-15) for each ν_{CPMG} . From B it is apparent that ν_{CPMG} has a strong effect on the line-width of peaks due to partial or complete suppression of the exchange contribution to R_2^{eff} . This figure was adapted from KLECKNER et al.¹³²

Similarly, the R_2^{ex} for every ν_{CPMG} is:

$$R_2^{ex}(\nu_{CPMG}^i) = R_2^{eff}(\nu_{CPMG}^i) - R_2^{eff}(\nu_{CPMG}^\infty) \quad (2-16)$$

By fitting the relaxation dispersion curve using an appropriately chosen model one can obtain estimates of the population of the exchanging states, of the chemical shift differences between the states and of the interconversion rates k_{ex} .¹³³ In addition, the R_2^{ex} values of different experimental conditions (pH, temperature, ligand bound vs apo, etc...) might be compared, to study the effect of these conditions on the exchange process.

CPMG relaxation dispersion techniques were initially applied to the ^1H - ^{15}N -spin systems of the backbone; more modern techniques have been developed to address the relaxation and exchange behavior of ^1H - ^{13}C -methyl groups in an otherwise highly deuterated protein sample, allowing for exploiting the TROSY principle (see also section 2.1.2) and thus making high molecular weight protein available for dynamics analyses by relaxation dispersion.¹³⁴

2.1.5 SOFAST-NMR for Real-time NMR studies

While the NMR techniques discussed in the previous sections are suitable for the study of processes on the fast to intermediate time scales (picoseconds to microseconds), NMR is equally useful for investigations of much slower processes in the seconds to hour regime like slow protein folding as well as other slow-proceeding conformational rearrangements. The basis of all real-time NMR experiments is the repetitive acquisition of spectra, where each individual spectrum represents a time point of a kinetic process. Obviously, the repetition rate of the kinetic series must be chosen carefully, to achieve a compromise between sufficient sensitivity and a time-resolution good enough to study the process of interest. Early approaches to study protein folding by real-time NMR relied on the repetitive acquisition of one dimensional (1D) ^1H spectra during the unfolding reaction with a time-resolution of several seconds.¹³⁵ The interpretation of these experiments was limited to global parameters of the reaction kinetics of the folding, due to an inherent lack of atomic resolution of 1D NMR spectra. A similar study applied two-dimensional (2D) ^1H - ^{15}N -HSQC spectra to the same protein folding reaction.¹³⁶ While here the introduction of the second dimension allowed for a clear distinction between folded and unfolded protein, the time resolution was poor and required slowing down the reaction at low temperature and the use of high protein concentrations (> 1.5 mM). This is because the introduction of a second dimension in a heteronuclear NMR experiment like a ^1H - ^{15}N -HSQC comes along with increased time of each individual experiment, since the ^{15}N frequency in the indirect dimension is sampled by multiple repetitions of the pulse sequence with an incremented delay time. Additionally, each of the repetitions with the incremented delay time is also repeated several times (the number of scans) to achieve enough sensitivity. Three general strategies can be applied to reduce the time required for one full spectrum: (1) optimization of the sampling, to reduce the number of sampled points in the second dimension; (2) single-scan NMR; (3) reduction of the delay time between scans, which is required for the longitudinal magnetization to reach equilibrium (recycle delay) (also see chapter 2.1.1.2). SOFAST-HMQC NMR spectroscopy (band Selective Optimized Flip Angle Short Transient) follows the third approach, thereby yielding a good compromise between time resolution and sensitivity by shortening the inter scan delay (also termed (d_I)) as much as possible.¹³⁷ The d_I period in NMR experiments is critically required to regain the equilibrium magnetization along the z -axis by longitudinal relaxation, before starting the next repetition of the pulse sequence. To achieve a maximum of available z -magnetization at the beginning of each repetition, two major adjustments have been applied to the standard HMQC sequence (Figure 2.9): (1) The excitation of only a selected subset of spins using shaped excitation pulses leaves many proton spins unperturbed, which in turn enhance the T_I relaxation of the excited spins via dipolar interactions; (2) Partial excitation employing an optimized ^1H flip angle enhances the available steady-state magnetization along the z -axis for a given inter scan delay time. The concept of an optimized flip angle has been introduced by the later Nobel laureate Richard Ernst and it relates to the duration of the effective T_I relaxation delay as follows:

$$\beta_{opt} = \cos^{-1} \left(e^{-\frac{T_{rec}}{T_1}} \right) \quad (2-17)$$

Here, β_{opt} is the Ernst angle, T_1 the longitudinal relaxation time constant and T_{rec} is the effective ^1H - T_1 relaxation delay, including the duration of the full pulse sequence and the d_1 inter scan delay (Figure 2.9). A reduced flip angle thus yields a higher signal to noise for high repetition rate. SCHANDA et al expressed the relation between the signal to noise ratio and the flip angle for a given repetition rate as follows:¹³⁷

$$S/N \propto \frac{(1 - \exp(-T_{rec}/T_1)) \sin \beta}{1 - \exp(-T_{rec}/T_1) \cos \beta} \frac{1}{\sqrt{n T_{scan}}} \quad (2-18)$$

Here, $\beta = \alpha - 180^\circ$ is the excitation angle, taking into account the effect of the 180° ^1H refocusing pulse in the pulse sequence below (Figure 2.9), and T_{scan} is the time required for a single full scan. According to equation (2-18), the selective excitation pulses α are recommended to be between $120^\circ < \alpha < 150^\circ$.

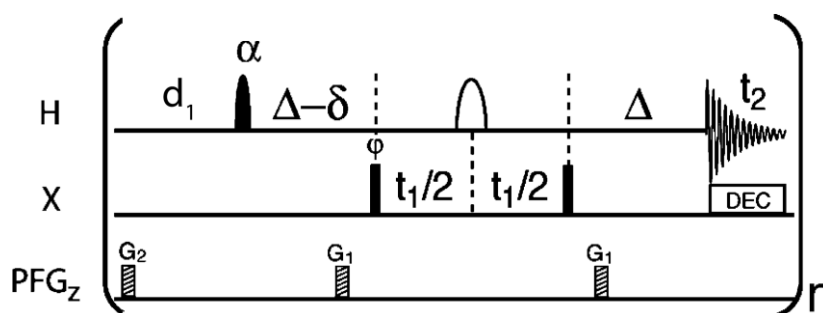


Figure 2.9 SOFAST-HMQC pulse sequence. The basic SOFAST-HMQC pulse scheme for a ^1H -X correlation experiment employs an optimized flip angle α and band selective proton excitation using shaped pulses as indicated by the non-rectangular pulse shapes. This figure has been adapted and modified from Schanda et al.¹³⁸

SOFAST-NMR employs the HMQC sequence which only requires few RF-pulses compared to HSQC sequences, and thus limits the loss of signal due to B_1 field inhomogeneities.

The SOFAST approach was initially implemented for ^1H - ^{15}N -correlations in non-deuterated protein samples which still contain a plethora of unperturbed proton spins after selective excitation of the amide protons.¹³⁷⁻¹³⁹ However, it has been also extended to ^1H - ^{13}C -correlations in methyl groups, which are interesting for larger molecular weight samples (also see chapter 2.1.2).¹³⁷ In methyl groups longitudinal spin-lattice relaxation is mostly mediated by dipolar interactions among the methyl ^1H and ^{13}C , which is additionally time-modulated by the fast rotation of the methyl group. The interaction with non-methyl protons is thus smaller and the enhancement of longitudinal relaxation in SOFAST-HMQC due to interactions with non-perturbed non-methyl protons is not as pronounced as for amides. Nevertheless, flip angle optimization in SOFAST results in significant signal increase at high repetition rates compared to a standard HSQC.¹⁴⁰ Combining the SOFAST approach with the TROSY effect in perdeuterated and selectively methyl- ^1H - ^{13}C labeled proteins has allowed for the successful recording of 2D correlation spectra on a > 400 kDa protein complex within a few seconds.¹⁴⁰

2.2 SEC-MALS

Prior to any high quality NMR study biochemical and biophysical characterization of the biological system is required to ensure the feasibility of high-resolution structural studies by NMR. Especially when working with oligomeric protein assemblies, knowledge of the molecular weight of the protein containing particle is of interest, since the NMR signal is dependent on the tumbling time constant of the particle and thus on its molecular weight (chapter 2.1.2). The standard technique for the characterization of the oligomeric state and associated molecular weights of a protein complex is size exclusion chromatography (SEC). SEC relies on the molecular sieve function of an appropriate chromatography material, which allows smaller particles to enter the material whereas bigger particles are excluded and elute from a SEC column earlier. This results in retention times based on the hydrodynamic radius of a particle. By comparison of the retention time to values obtained from a column calibration with standard proteins of known molecular weight, it is possible to obtain an estimation of the sample's molecular weight. However, this estimation generally assumes an ideal spherical shape of the protein particle under study and does not take into account potential interactions with the column packing material.

Combining SEC and MALS (Multi Angle Light Scattering) allows for the determination of the molecular weight of the particle, independent of its shape and its SEC retention time.

MALS exploits the angular dependence of the analyte's Rayleigh scattering intensity when irradiated by an incident laser beam. Central to the theoretical background of MALS is the reduced Zimm equation:¹⁴¹

$$\frac{c K}{R_{(\theta,c)}} = \frac{1}{M_w P(\theta)} + 2A_2 c \quad (2-19)$$

Where $R_{(\theta,c)}$ is the analyte's angle dependent excess scattering intensity, c the concentration of the analyte in mg/ml and M_w the molecular weight of the analyte. $R_{(\theta,c)}$ is essentially the difference between the scattered light of the analyte containing solution and that of the pure solvent. $P(\theta)$ describes the angular dependence of the scattered light based on a form factor of the analyte and is directly related to the radius of gyration R_g . A_2 is the second virial coefficient as a measure for non-ideal solution behavior of the analyte. The constant K is given as:

$$K = \frac{4\pi^2 \left(\frac{dn}{dc}\right)^2 n_0^2}{N_A \lambda_0^4} \quad (2-20)$$

where, (dn/dc) is the refractive index increment of the analyte n_0 the refractive index of the solvent, N_A the *Avogadro number* as 6.022×10^{23} and λ_0 is the vacuum wave-length of the incident laser beam.

In practice, the principle of MALS can be simplified with the following equations, which relate the experimental scattering intensity at angle θ $I_{(\theta)}$ to the excess Rayleigh ratio $R_{(\theta)}$.

$$R_{(\theta)} = \frac{(I_{\theta} - I_{\theta}^{solv}) \cdot r^2}{I_0 \cdot V} \quad (2-21)$$

From which follows,

$$R_{(\theta)} \propto I_{\theta} \quad (2-22)$$

Thus,

$$I_{\theta} \propto M_w \times c \times \left(\frac{dn}{dc}\right)^2 \quad (2-23)$$

where, $I_{(\theta)}^{solv}$ is the scattering intensity at angle θ of the pure solvent, V the illuminated volume of the scattering medium from which the detector at angle θ collects the light, r is the distance between the detector and the scattering volume and I_0 is the intensity of the incident beam. Equation (2-23) finally, simplifies the principle of the MALS-experiment, which relates the molecular weight M_w to the experimental parameters of scattering intensity $I_{(\theta)}$ (from the MALS detector) and concentration c (from the refractometer and the refractive index increment (dn/dc)). For biomolecules a refractive index increment of 0.1852 ml/g is a good approximation and is used in the calculation of the molecular weight. The light scattering intensity is obtained using a MALS detector which measures the scattering at multiple angles and the absolute concentration is routinely obtained from an in-line attached refractometer using the assumed (dn/dc) of 0.1852 ml/g.

As mentioned before SEC-MALS allows for the shape-independent determination of the molecular weight and as such is especially useful for protein samples which are assumed to be highly aspherical like the PAN coiled coil domain in the presented thesis (chapter 5.1.1). Further, the additional information on the molecular weight allows for detection of mixtures of oligomeric species within a single chromatographic peak as was done for the full length PAN in this thesis. As such, SEC-MALS is an invaluable tool to detect deviations from molecular weight monodispersity, which can be critical for high quality NMR experiments.

3 Aim of this Thesis Work

Proteolytic AAA+ unfoldases, like the proteasome, are at the center of protein quality control systems in bacteria, archaea and eukaryotes. High-resolution structural studies on close-to-native states of these molecular machines have long been exacerbated by their intrinsic dynamics and resulting conformational heterogeneity. A plethora of in-depth biochemical and biophysical studies, however, have laid a solid foundation for understanding the mechanism of protein-unfolding by these ATP-fuelled unfoldases.

In the recent years multiple cryo-EM models of various AAA+ unfoldases both in an apo state and bound to substrates have emerged and based on them a mechanism of unfolding common to all AAA+ unfoldases is proposed.¹⁰⁶ However, the proposed sequential rotary mechanism is in disagreement with key findings from previous structural, biochemical and biophysical experiments. While these structures undoubtedly, provide valuable high-resolution structural snap-shots and insights into the conformational landscape of these hexameric machines, the mechanistic interpretation of these and the generality of the proposed universal mechanism has started a debate whether all AAA+ unfoldases work in the same way.¹⁰⁸

In this PhD project I intended to exploit the advantages of solution-state NMR spectroscopy over other structural techniques (that are an atomic resolution and native-like sample conditions in solution) to study the structure and the unfolding mechanism of the proteasomal unfoldase PAN from the archaeon *Methanocaldococcus jannaschii*, contributing high-resolution insights to the understanding of the mechanism of protein unfolding by AAA+ ATPases.

Herein, two major approaches were pursued and several objectives for this thesis have been formulated:

1. Structural and functional studies on the archaeal PAN-20S proteasome including
 - Preparation and characterization of full-length PAN and individual PAN subdomains suitable for NMR studies
 - Obtaining NMR resonance assignments on individual PAN subdomains and full-length PAN
 - Characterization of the conformational states of full-length PAN
 - Investigating the interaction regions between PAN and a model substrate
2. Establishing the structural kinetics workflow of proteasomal degradation of the model substrate GFP-ssrA by time-resolved NMR-experiments on isotopically labeled model substrates during the proteasomal degradation by the PAN-20S proteasome.

4 Materials and Methods

4.1 Materials

4.1.1 Instruments

Table 4-1 Laboratory Instruments

	Instrument	Manufacturer
Centrifuges	Allegra 64R	Beckmann-Coulter, Carlsbad, CA, USA
	Heraeus Labofuge 400R	ThermoFisher Scientific, Dreieich, Germany
	Heraeus Fresco 21	ThermoFisher Scientific, Dreieich, Germany
	Sorvall Lynx 600	ThermoFisher Scientific, Dreieich, Germany
Incubators	Innova 44	New Brunswick Scientific, MA, USA
	Excella E24	New Brunswick Scientific, MA, USA
	Heratherm	ThermoFisher Scientific, Dreieich, Germany
Spectrophotometer	DS-11 (microvolume)	DeNovix Inc., Wilmington, DE, USA
	Photometer Bio	FoodALYT GmbH, Bremen, Germany
Plate reader	Spectramax M3	Molecular Devices, München, Germany
PCR cycler	T100 Thermal cycler	Bio-Rad, Hercules, CA, USA
Analytical Balance	Secura Analytical	Sartorius, Göttingen, Germany
SDS-PAGE	Powerpc Universal 500 V, 2.5 A 500W	Bio-Rad, Hercules, CA, USA
	MiniProteasn Tetra System	Bio-Rad, Hercules, CA, USA
Agarose gel electrophoresis	PowerPac Universal Power Supply	Bio-Rad, Hercules, CA, USA
	Mini-Sub Cell GT Cell	Bio-Rad, Hercules, CA, USA
Sonicator	SONOplus mini20	Bandelin, Berlin, Germany
High pressure homogenizer	Emulsiflex C5	Avestin Inc., Ottawa, Canada
RT-PCR machine	StepOne	ThermoFisher Scientific, Dreieich, Germany
MALS	miniDAWN TREOS	Wyatt Technologies, Goleta, CA, USA
	Optilab T-REX	Wyatt Technologies, Goleta, CA, USA
HPLC-MS	2797 separation module	Waters GmbH, Eschborn, Germany
	2545 pump system	Waters GmbH, Eschborn, Germany
	996 photodiode array detector	Waters GmbH, Eschborn, Germany
	Micro Mass quattro micro ESI mass detector	Waters GmbH, Eschborn, Germany

	Instrument	Manufacturer
NMR spectrometer	AVANCE III HD 600 with N ₂ -cooled inverse HCN triple-resonance cryogenic probehead	Bruker Biospin, Karlsruhe, Germany
	AVANCE III HD 850 with He-cooled inverse HCN triple-resonance cryogenic probehead	Bruker Biospin, Karlsruhe, Germany

4.1.2 Software

Table 4-2 List of used software programs

Program	Developer
Bruker Topspin 3.2	Bruker Biospin, Karlsruhe, Germany
nmrPipe	DELAGLIO et al ¹⁴²
CcpNMR analysis	VRANKEN et al ¹⁴³
PyMol 2.4	Schroedinger LLC.,
APBS tool 2.1	BAKER et al ¹⁴⁴
OriginPro 9.1	Origin Lab, Northhampton, MA, USA
UNICORN 7	Cytiva (Formerly GE Healthcare), USA
ASTRA 7.1.4	Wyatt Technologies, USA
Microsoft Office 2010 Professional	Microsoft Inc., Redmond, WA, USA
Illustrator CS2	Adobe, USA

4.1.3 Chemicals, Enzymes and Consumables

All standard laboratory chemicals, buffer components and solvents were purchased either from Sigma-Aldrich (Darmstadt, Germany), Carl Roth (Karlsruhe, Germany), VWR (Frankfurt, Germany) or Honeywell (Seelze, Germany). Table 4-3 lists all other chemicals, enzymes and consumables used in this thesis. Chromatography columns were all purchased from Cytiva (formerly GE Healthcare, München, Germany). Isotopically labelled compounds for protein labelling are listed in Table 4-4.

Table 4-3 Chemicals, Enzymes and Consumables used in this thesis

Product	Supplier
Acrylamide solution 40 % (37.5:1)	Carl Roth, Karlsruhe, Germany
Agarose	Sigma-Aldrich, Darmstadt, Germany
Amicon centrifugal filters 15 (3K, 10K, 30K, 100K MWCO)	Merck Millipore, Darmstadt, Germany
ADP	Sigma-Aldrich, Darmstadt, Germany

ATP	Carl Roth, Karlsruhe, Germany
ATPgS	Sigma-Aldrich, Darmstadt, Germany
β -ME	Carl Roth, Karlsruhe, Germany
Bromophenol blue	VWR, Frankfurt, Germany
BSA	Sigma-Aldrich, Darmstadt, Germany
Color Prestained Protein Standard	New England Biolabs, Frankfurt, Germany
cOmplete EDTA-free protease inhibitor cocktail tablets	Roche Diagnostics, Mannheim, Germany
Coomassie Brilliant blue G-250	Bio-Rad, Hercules, CA, USA
Dialysis tubing	Spectrum Chemicals, New Brunswick, NJ, USA
DNase I	Roche Diagnostics, Mannheim, Germany
dNTP mix	New England Biolabs, Frankfurt, Germany
DpnI endonuclease	New England Biolabs, Frankfurt, Germany
DTT	Carl Roth, Karlsruhe, Germany
HEPES	Carl Roth, Karlsruhe, Germany
Imidazole	Carl Roth, Karlsruhe, Germany
IPTG	Carl Roth, Karlsruhe, Germany
Kanamycinsulfate	Carl Roth, Karlsruhe, Germany
LB medium	Carl Roth, Karlsruhe, Germany
Lysozyme	Carl Roth, Karlsruhe, Germany
MES	Carl Roth
Microtiter plates for fluorescence measurements	Brand, Wertheim, Germany
Pefabloc sc protease inhibitor	Roche Diagnostics, Mannheim, Germany
Phusion polymerase	New England Biolabs, Frankfurt, Germany
Pfu plus polymerase	Roboklon, Berlin, Germany
Thermo-Fast PCR Plate, 48-well, clear	ThermoFisher Scientific, Dreieich, Germany
Laboratory plasticware	Sarstedt, Nümbrecht, Germany
3C-protease	made in-house
TEV-protease	made in-house
QIAprep spin iniprep plasmid isolation kit	Qiagen, Hilden, Germany
QIAquick Gel extraction kit	Qiagen, Hilden, Germany
RNase	Roche Diagnostics, Mannheim, Germany
pH-meter calibrating solutions	Hanna Instruments, Vöhringen, Germany
SDS	Carl Roth
SYBR orange Dye (5000x)	ThermoFisher Scientific, Dreieich, Germany
TRIS-base	Sigma-Aldrich, Darmstadt, Germany
Tris-HCl	Sigma-Aldrich, Darmstadt, Germany
Triton X-100	Sigma-Aldrich, Darmstadt, Germany
Unstained Protein Standard	New England Biolabs, Frankfurt, Germany

Table 4-4 Isotopically labeled compounds

Compound	Supplier
¹⁵ N Ammonium chloride	CIL, Tewksbury, MA, USA
D-Glucose (U- ¹³ C6)	CIL, Tewksbury, MA, USA
D-Glucose (1,2,3,4,5,6,6-D7)	CIL, Tewksbury, MA, USA
D-Glucose (U- ¹³ C6; 1,2,3,4,5,6,6-D7)	CIL, Tewksbury, MA, USA
Glycerol (D8, 99%)	CIL, Tewksbury, MA, USA
α-ketobutyric acid sodium salt (methyl- ¹³ C; 3,3-D2)	CIL, Tewksbury, MA, USA
α-ketobutyric acid sodium salt (¹³ C4; 3,3-D2)	CIL, Tewksbury, MA, USA
α-ketoisovaleric acid, sodium salt (3-methyl- ¹³ C; 3,4,4,4-D4)	CIL, Tewksbury, MA, USA
α-ketoisovaleric acid, sodium salt (1,2,3,4- ¹³ C4; 3,4',4',4'-D4)	CIL, Tewksbury, MA, USA
SLAM-methionine-ε labeling kit	NMR-Bio, Grenoble, France
D ₂ O (deuterium oxide)	Eurisotop, Saarbrücken, Germany

4.1.4 Media, Supplements and Buffers

Table 4-5 M9 medium composition

Volume	Stock
100 µL	1 M CaCl ₂
2 ml	1M MgSO ₄
6 ml	5 mg/ml Thiamine-HCl
1 ml	1000x Trace elements
4ml	0.25 mg/ml NH ₄ Cl (¹⁵ NH ₄ Cl)
20 ml	20 % carbon source (glucose or glycerol)
100 ml	10x M9 salts without NH ₄ CL
1 ml	50 mg/ml kanamycin
Ad 1000 ml	ddH ₂ O or D ₂ O

Table 4-6 Formulation of 10 x M9 salts without NH₄CL

Volume	Compound
75.2 g	Na ₂ HPO ₄ · 2 H ₂ O
30 g	KH ₂ PO ₄
5 g	NaCl
Ad 1000 ml	ddH ₂ O or D ₂ O
	Adjust pH to 7.4 (pD 7.0 in D ₂ O)

Table 4-7 Composition of the 1000x trace elements solution

Volume	Stock
50 ml	0.1 M FeCl ₃ in 0.12 M HCL
2 ml	1 M CaCl ₂
1 ml	1 M MnCl ₂ · 4 H ₂ O
1 ml	1 M ZnSO ₄ · 7 H ₂ O
1 ml	0.2 M CoCl ₂ · 6 H ₂ O
2 ml	0.1 M CuCl ₂ · 2 H ₂ O
1 ml	0.2 M NiCl ₂
2 ml	0.1 M Na ₂ MoO ₄ · 2 H ₂ O
2 ml	0.1 M Na ₂ SeO ₄
2 ml	0.1 M H ₃ BO ₄
36 ml	H ₂ O or D ₂ O

4.1.5 Plasmids and Oligonucleotides

4.1.5.1 Plasmids

Table 4-8 and Table 4-9 list the empty expression vectors used for subcloning and the expression plasmids used in this thesis, respectively. Amino acid sequences for the full length versions of PAN, 20S CP subunits, GFPuv-ssrA and frGFP-ssrA are listed in the appendix.

Table 4-8 Expression vectors used for subcloning of PAN and GFP genes

Name	Comments	Origin
pETM22	Multiple cloning site with N-terminal Trx-solubility tag, linker, His ₆ -tag and 3C cleavage site. Kanamycin resistance cassette. All constructs created using this vector contained the full Trx-His ₆ -3C-tag	EMBL collection
pETM11	Multiple cloning site with N-terminal His ₆ -tag and TEV-protease cleavage site. Kanamycin resistance cassette	EMBL collection

Table 4-9 List of expression plasmids used in this thesis

Name	Comments	Origin
pET30a_PAN_full-length	<i>M jannaschii</i> PAN, C-terminal His ₆ -tag	IBS Grenoble, synthetic gene from Genecust Europe
pETM22_PAN_full-length	<i>M jannaschii</i> PAN with cleavable N-terminal solubility tag	This thesis
pETM22_PAN_1-89	<i>M jannaschii</i> PAN CC-domain	This thesis

Name	Comments	Origin
pETM22_PAN_74-150	<i>M jannaschii</i> PAN OB-ring	This thesis
pETM22_PAN_1-150	<i>M jannaschii</i> PAN CC + OB-ring	This thesis
pETM22_PAN_74-430	<i>M jannaschii</i> PAN OB-ring+ATPase domain	This thesis
pETM22_PAN_74-430_F196A	<i>M jannaschii</i> PAN OB-ring+ATPase domain with F196A point mutation	This thesis
pETM22_PAN_150-430	<i>M jannaschii</i> PAN ATPase domain	This thesis
pETM22_PAN_150-430_F196A	<i>M jannaschii</i> PAN ATPase domain with F196A point mutation	This thesis
pET30a_GFPuv_ssrA	<i>A victoria</i> GFPuv with a A206K point mutation ^{145,146}	IBS Grenoble, synthetic gene from Genecust Europe
pETM11_GFPuv_ssrA	<i>A victoria</i> GFPuv with a A206K point mutation ^{145,146} , N-terminal His ₆ -tag	This thesis
pETM11_frGFP_ssrA	<i>A victoria</i> folding reporter GFP ¹⁴⁷ with a A206K point mutation ¹⁴⁵	This thesis
pET28a_20S_alpha	<i>M jannaschii</i> 20S core particle α -subunit with an N-terminal His ₆ -tag	IBS Grenoble, synthetic gene from Genecust Europe
pET30a_20S_beta	<i>M jannaschii</i> 20S core particle β -subunit without tag, including protease propeptide	IBS Grenoble, synthetic gene from Genecust Europe

4.1.5.2 Oligonucleotides

Oligonucleotide-primers were ordered from Sigma-Aldrich purified with the standard desalting procedure. The lyophilised oligonucleotides were dissolved in ddH₂O to a final concentration of 100 μ M and stored at -20 °C until use. Table 4-10 contains all primers used in this thesis with their respective 5'-3' sequence and application:

Table 4-10 Oligonucleotide primers used in this thesis

Name	Sequence (5'-3')	Application
PAN_into_pETM3C_fw	GAAGTTCTGTTCCAGGGGCC ATGGTCTTCGAGGAGTTC	Cloning PAN constructs into pETM22
PAN_into_pETM3C_rv	GCAAGCTTGTCGACGGAGCTC TTAGCGATACAGAACATCCAG	Cloning PAN constructs into pETM22
PAN_1-89_rv	GCAAGCTTGTCGAC- GGAGCTCTTAGACACGCAT- ACGGTCCAGTTC	Cloning PAN CC into pETM22
PAN_74_into_3C_fw	GAAGTTCTGTTCCAGGGGCC ATGAAAGAAAAC- GAAATCCTGCGTCGTG	Cloning PAN OB-ring into pETM22
PAN_into	GCAAGCTTGTCGAC-	Cloning PAN OB-ring into pETM22

Name	Sequence (5'-3')	Application
3C_150_rv	GGAGCTCTTATTCCGGCAG- TACGTCAACCAC	
PAN_into pETM3C_150_fw	GAAGTTCTGTTCCAGGGGCC GAAAACAAAGACTACCGTG- CAAAGCAATGG	Cloning PAN ATPase domain into pETM22
PAN_F196A_fw	GAAACACCCG- GAACTGGCCGAAAAAGTTGG- CATCG	Point mutation F196A
PAN_F196A_rv	CGATGCCAACTTTTTCGGCCA GTTCCGGGTGTTTC	Point mutation F196A
PAN_I7A_fw	TATGGTCTTCGAGGAG- TTCGCCTCTACGGAGCTGAA- GAAG	Point mutation I7A for assignment
PAN_I7A_rv	CTTCTTCAGCTCCG- TAGAGGCGAACTCCTCGAA- GACCATA	Point mutation I7A for assignment
PAN_I49A_fw	AGGAGCTGCAAGAAAAA- GCACGTGCCGCAGAAGT- GAAAG	Point mutation I49A for assignment
PAN_I49A_rv	CTTTCCAGTTCTGCGGCAC- GTGCTTTTTCTTGACGCTCCT	Point mutation I49A for assignment
PAN_I56A_fw	CGCAGAACTGGAAA- GCCGTGCCCTGAAACTG- GAACTGG	Point mutation I56A for assignment
PAN_I56A_rv	CCAGTTCAGTTTCAGGG- CACGGCTTTCAGTTCTGCG	Point mutation I56A for assignment
PAN_M74A_fw	AACGTGAAAACCTG- CAGCTGGCGAAAGAAAAC- GAAATCCTGC	Point mutation M74A for assignment
PAN_M74A_rv	GCAGGAT- TTCGTTTTCTTTCGCCAGCTG- CAGGTTTTACGTT	Point mutation M74A for assignment
PAN_M87A_fw	TGCGTCGTGAACTGGAC- CGTGCGCGTGTCCCAC	Point mutation M87A for assignment
PAN_M87A_rv	GTGGGACACGCGCAC- GGTCCAGTTCACGACGCA	Point mutation M87A for assignment
PAN_3K3E_fw	AAAGAGGAGGTTGAGGTTAAAGAAC- CGGCG CACCTG	Introduction of the point mutations K415E, K416E, K418E
PAN_3K3E_rv	AACCTCAAC- CTCCTCTTTCTCCATGAT- TTTTCTACCGCTTTGC	Introduction of the point mutations K415E, K416E, K418E
PAN_K414E_fw	CGGTAGAAAAATCATGGAG- GAAGAGGAGGTT- GAGGTTAAAG	Introduction of the point mutation K414 into the triple mutant K415E, K416E, K418E

Name	Sequence (5'-3')	Application
PAN_K414E_rv	TTAACCTCAAC- CTCCTCTTCCTCCATGAT- TTTTCTACCG	Introduction of the point mutation K414 into the triple mutant K415E, K416E, K418E
PAN_K406S_fw	CGCGATTATGTAAC- CATGGATGATTTTCG- CAGCGCGG- TAGAAAAAATCATG	Introduction of the point mutation K406S into the quadruple mutant K414E, K415E, K416E, K418E
PAN_K406S_rv	ATGATTTTTTCTAC- CGCGCTGCGAAAATCA- TCCATGGTTACATAATCGCG	Introduction of the point mutation K406S into the quadruple mutant K414E, K415E, K416E, K418E
GFP- ssrA_into_pETM 11_fw	CACTACTGA- GAATCTTTATTTTCAGGGCGCC ATGTCTAAAGGTGAA- GAACTGTTACCCGG	Cloning GFPuv_ssrA into pETM11
GFP- ssrA_into_pETM 11_rv	GAGCTCGAATTCGGATCCGG- TACTTACGCCGCCAGCG- CATAG	Cloning GFPuv_ssrA into pETM11
GFP_F64L_S65T _fw	GGCCGACCCTGGTTAC- CACCTGAC- CTATGGTGTTCAGTGTTC	Creating frGFP-ssrA from GFPuv-ssrA by two point mutations F64L and S65T
GFP_F64L_S65T _rv	GAGAAACACTGAACACCA- TAGGTCAGGGTGGTAAC- CAGGGTCGGCC	Creating frGFP-ssrA from GFPuv-ssrA by two point mutations F64L and S65T

4.1.5.3 Peptides

4.1.6 Bacterial Strains and competent cells

Table 4-11 Bacterial strains

Strain	Genotyp	Origin
<i>E. coli</i> BL 21 (DE3)	<i>fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS</i>	Prepared in-house
<i>E. coli</i> TOP 10	F- <i>mcrA</i> Δ(<i>mrr-hsdRMS-mcrBC</i>) Φ80 <i>lacZ</i> ΔM15 Δ <i>lacX74 recA1 araD139</i> Δ(<i>araleu</i>)7697 <i>galU galK rpsL</i> (StrR) <i>endA1 nupG</i>	Prepared in-house

4.2 Cloning and Mutagenesis

4.2.1 Cloning

All expression constructs designed and created in this work (Table 4-9) were constructed employing an oligonucleotide-based restriction-free ligation approach which did not need the introduction of endonuclease restriction sites.¹⁴⁸

In a first PCR-step the gene of interest was amplified using primers (Table 4-10) which created 5'- and 3'- overlaps with the insertion site in the target plasmid. PCR-mix and cyclor conditions are given in table Table 4-12 and Table 4-13:

Table 4-12 PCR mix RF-cloning – gene amplification

Component	Volume (µl)
Plasmid with gene of interest (50 ng/µl)	1
Phusion HF-Buffer (5x)	20
DMSO	3
Forward primer (10 µM)	2
Reverse primer (10 µM)	2
dNTPs (10 µM each)	2
ddH ₂ O	69
Phusion polymerase	1
Final mix	<u>100</u>

Table 4-13 Thermocycler conditions gene amplification

Step	Temp.	Time	
Initial denaturation	98 °C	30 s	} x45
Denaturation	98 °C	30 s	
Annealing	60 °C	30 s	
Elongation	72 °C	1 min/kb	
Final elongation	72 °C	7 min	
Storage	8 °C	∞	

The amplified and linearized genes were separated from the plasmid by gel electrophoresis in 1.5 % agarose-gels and subsequently extracted using the Qiagen gel extraction kit.

A second PCR utilized the previously amplified gene as a “mega-primer” and was carried out using the following conditions, while the master mix was split into three 25 µl reactions, which were used in reactions with three different annealing temperatures:

Table 4-14 Reaction mix for Megaprimer PCR

Component	Volume (µl)
Target plasmid (10-50 ng/µl)	1
Phusion HF-Buffer (5x)	20
DMSO	3
Gene-“Megaprimer” (50-150 ng/µl)	2

Component	Volume (μ l)
dNTPs (10 μ M each)	2
ddH ₂ O	71
Phusion polymerase	1
Final mix	<u>100</u>

Table 4-15 Cycler conditions for Megaprimer PCR

Step	Temp.	Time	
Initial denaturation	98 °C	30 s	} x 25
Denaturation	98 °C	30 s	
Annealing	58/61/64 °C	30 s	
Elongation	72 °C	1 min/kb	
Final elongation	72 °C	1 min	
Storage	8 °C	∞	

Next, the PCR mixes were incubated with the Dpn1 endonuclease for 3h at 37 °C to digest the original target plasmid. Finally, the digested Megaprimer PCR mix was used to transform chemically competent *E coli* Top 10 cells. Successful transformands were selected on agar plates prepared with 50 μ g/ml kanamycin. Plasmid-DNA of transformands was amplified in an overnight culture in LB, extracted with the Miniprep plasmid DNA extraction kit (Quiagen) and successful insertion of the gene was confirmed by Sanger sequencing.

4.2.2 Site-directed mutagenesis

Introduction of point mutations was carried out either following the QuikChange (Agilent Technologies) or the PIPE-cloning approach,¹⁴⁹ the latter of which allowed for the incorporation of several point mutations next to each other. Reaction mix and cycler conditions for the PIPE-cloning approach were identical to those for the Megaprimer PCR in Table 4-14 and Table 4-15 except for 1 μ l of both forward and reverse primers (10 μ M stock of each) instead of the Megaprimer. The QuikChange PCR was also conducted at three different annealing temperatures using mutagenesis primers designed with the QuikChange primer design tool from Agilent Technologies and the following reaction conditions were used:

Table 4-16 Reaction mix for QuikChange mutagenesis

Component	Volume (μ l)
Plasmid-DNA of wild-type gene (50 ng/ μ l)	1
Pfu plus polymerase buffer (10x)	10
DMSO	3
Forward primer (10 μ M)	4
Reverse primer (10 μ M)	4
dNTPs (10 μ M each)	4
ddH ₂ O	72
Pfu plus polymerase	2
Final mix	<u>100</u>

Table 4-17 Thermocycler conditions QuikChange mutagenesis

Step	Temp.	Time	
Initial denaturation	95 °C	1 min	} x 25
Denaturation	95 °C	1 min	
Annealing	58/61/64 °C	1 min	
Elongation	68 °C	11 min/kb	
Final elongation	68 °C	11 min	
Storage	8 °C	∞	

After the PCR was completed residual original plasmid DNA was digested with Dpn1 for 3h at 37 °C and successful mutants were selected on agar-plates prepared with 50 µg/ml kanamycin. Plasmid-DNA of transformands was amplified in an overnight culture in LB, extracted with the Miniprep plasmid DNA extraction kit (Quiagen) and successful insertion of the gene was confirmed by Sanger sequencing.

4.3 Protein expression and purification

4.3.1 Recombinant protein expression in LB medium

For expression of recombinant proteins in LB medium, 50 mL of LB medium supplemented with 50 µg/ml kanamycin were inoculated from agar-plates carrying *E coli* BL21 (DE3) freshly transformed with a plasmid containing the gene of interest. This preculture was incubated at 37 °C for 12-16 hours while shaking at 180 rpm in Erlenmeyer flasks. Next, the 1l main culture in LB medium (50 µg/mL kanamycin) was inoculated from the preculture at a ratio of 1:100 and incubated shaking at 220 rpm and 37 °C until an optical density at 600 nm (OD₆₀₀) of 0.6-0.9 was reached. Cells were allowed to cool down until they reached roughly 20 °C and expression of the target protein was induced by the addition of 0.1 mM (1 mM for full length PAN and proteasome 20S core particle components) isopropyl-β-d-1-thiogalactopyranoside (IPTG). The cultures were incubated at 20 °C and 220 rpm shaking for another 16-20 h (for GFP-ssrA constructs 8 h). Finally, cells were harvested by centrifugation at 5.000 x g, washed with PBS and kept frozen at -20 °C until use.

4.3.2 Recombinant expression of isotopically labelled protein and deuteration

Isotopically labeled proteins

For the production of non-deuterated isotopically labeled proteins (¹⁵N and ¹³C) a 50 mL preculture in M9 minimal medium (composition described in Table 4-5), was directly inoculated from an agar-selection plate and incubated over night at 37 °C. The next day, a main culture was prepared using the appropriately labeled nitrogen- (1 g /l ¹⁵NH₄Cl) and carbon (4 g/l D-Glucose (U-¹³C6)) source and was inoculated from the preculture in a ratio 1:50. The main culture was incubated at 37 °C while shaking at 230 rpm until an OD₆₀₀ of 0.6-0.9 was reached. Cultures were allowed to cool down to reach the desired expression temperature

(20 °C) and protein expression was induced by the addition of 0.1 mM IPTG. Induced cultures were further incubated at 20 °C and 230 rpm for 18 h (8 h for GFP-ssrA preparations). Expression was stopped by centrifugation at 5.000 x g and 4 °C, cells were washed with cold PBS and stored at -20 °C until further use.

Deuteration and site-specific labeling of methyl groups

For the uniform deuteration of recombinant protein, cells harboring the respective expression plasmid were adapted to 99 % D₂O by incrementally increasing the D₂O/H₂O ratio. Initially, a 5 ml LB culture was inoculated from an agar selection plate and grown at 37 °C and 180 rpm for 8-12 h until a visibly high cell density was reached. The next 5 ml culture was prepared in M9 medium using the deuterated carbon source (D-Glucose (1,2,3,4,5,6,6-D₇), D-Glucose (U-¹³C₆; 1,2,3,4,5,6,6-D₇) or glycerol-D₈) and 22 % D₂O and was inoculated with a ratio 1:25 from the LB-culture. Likewise, adaptation steps with 33, 55, 70 and 99% D₂O were carried out. Finally, a 50 ml preculture in 99 % D₂O was prepared, grown over night and used to inoculate the 500 ml final main culture by inoculating with all cells from the overnight culture. Cells were allowed to grow until an OD₆₀₀ 0.6-0.9 was reached and protein expression was induced as for non-deuterated preparations. Incorporation of selectively ¹H, ¹³C-labelled methyl groups was achieved by the addition of respectively labeled amino acid precursors before the induction with IPTG (60 mg/l α-ketobutyric acid, sodium salt (methyl-¹³C; 3,3-D₂) for terminal Ileδ₁-¹³C¹H₃-labeling, 120 mg/l α-ketoisovaleric acid, sodium salt (3-methyl-¹³C; 3,4,4,4-D₄ for terminal Leuδ- and Valγ- ¹³C¹H₃-labelling and 4 vials/l of the methionine-ε SLAM kit (NMR –Bio) for methionine- ε ¹³C¹H₃ labeling). For production of deuterated protein samples for the assignment of ILV-methyl groups, special amino acid precursor for full side-chain ¹³C labeling were used instead (60 mg/l α-ketobutyric acid, sodium salt (¹³C₄; 3,3-D₂) for full-chain Ileδ₁ and 120 mg/ml α-alpha-ketoisovaleric acid, sodium salt (1,2,3,4-¹³C₄; 3,4',4',4'-D₄) for Leuδ and Valγ). Cells were harvested and stored in the same way as their unlabeled counterparts.

4.3.3 Protein purification

Different preparations of the proteins used in this thesis have all been purified according to the following protocols in H₂O-based purification buffers, no matter if it was an unlabelled variant or an isotopically enriched or deuterated preparation. Protein identity and purity were routinely assessed by 15 % SDS-PAGE gel electrophoresis and coomassie staining.

4.3.3.1 Purification of recombinant archaeal *M jannaschii* proteasome components

Purification of the PAN coiled coil domain

Cells harbouring the expressed fusion protein Trx-His₆-PAN-1-90 were resuspended in 50 mL Histrap-buffer A (20 mM Tris-HCl, 500 mM NaCl, 10 mM Imidazole, 5 mM β-ME, pH 7.5) supplemented with DNase I, lysozyme, 1 tablet complete EDTA protease inhibitors and incubated for ten minutes shaking at RT on a nutating shaker. Cells were disrupted by sonication and cell debris was subsequently removed by centrifugation at 19.000 g. After filtration

through a 0.22 μ M syringe filter the cleared cell lysate was applied to a 5 ml HisTrap HP Ni-NTA column, preequilibrated in HisTrap buffer A. Unbound protein was removed by washing with 7 CV (column volume) HisTrap Buffer A. The Trx-fusion protein was eluted with a 10 CV gradient HisTrap-buffer A \rightarrow HisTrap-buffer B (same as A, but 500 mM Imidazole). Peak fractions were combined and buffer exchanged to HisTrap-buffer A. Next, the Trx-solubility tag was cleaved by incubating the fusion-protein with the 3C protease overnight at 4°C with mild shaking. Next, the cleavage mix was again passed over the Ni-NTA column to remove un-cleaved fusion protein. The flowthrough was concentrated using 3k MWCO-centrifugal filter units (Millipore) and the concentrated protein was further purified on a Highload 16/600 Superdex 75 pg size-exclusion chromatography column. Peak fractions containing the dimeric complex were collected and checked for purity and integrity by Tricin-SDS-PAGE. During the chromatographic purification steps after solubility tag cleavage, absorbance at 214 nm was used to follow the protein, since the PAN CC-domain does not contain any tryptophanes or tyrosine for detection at 280 nm. Purified protein concentrations were estimated by absorbance at 214 nm or by refractivity during MALS experiments (4.5.1).

Purification of PAN CC in conjunction with the PAN OB-ring (residues 1-150)

Cells harbouring the expressed fusion construct of PAN residues 1-150 C-terminally fused to the Trx-solubility tag, a hexahistidine tag and a 3C cleavage site were resuspended in HisTrap-buffer A (20 mM Tris-HCl, 500 mM NaCl, 10 mM MgCl₂, 10 mM imidazole and 5 mM β -ME at pH 7.5) supplemented with DNase I, lysozyme and one tablet of complete EDTA-free protease inhibitor cocktail. Lysis was allowed to proceed by incubation on a nutating shaker at RT for 30 min, after which the cells were sonicated for 20 min (amplitude 52 %, on/off: 5 s / 10 s) and cell debris was removed by centrifugation at 19.000 x g for 1 h at 4 °C. The cleared lysate was filtered with a syringe filter (0.45 μ M pore size) and loaded on a 5 ml HisTrap HP Ni-NTA-column preequilibrated in HisTrap-buffer A. Unbound protein was washed off the column by extensively washing with HisTrap-buffer A (10 CV) and 50 mM imidazole in HisTrap-buffer A (10 CV). The fusion protein was eluted from the column with a gradient from 50-500 mM imidazole in HisTrap-buffer A. Fractions containing the target protein were collected and buffer exchanged to HisTrap-buffer A on a HiPrep 26/10 Desalting column. Cleavage of the Trx-His6-tag was done overnight at RT using 1 mg 3C protease per estimated 10 mg of fusion protein. The next day, un-cleaved protein and the tag were removed by passing the cleavage mix over the HisTrap column again. The flowthrough containing the target protein was buffer exchanged to storage buffer (20 mM Tris-HCl, 100 mM NaCl, 10 mM MgCl₂, pH 7.5) and loaded onto a 5 ml HisTrap Q HP anion-exchange chromatography column preequilibrated in storage buffer. Unbound protein was washed with 10 CV of storage buffer and the PAN 1-150 was eluted with an NaCl gradient from 100-500 in storage buffer. The protein was concentrated with 3 kDa MWCO centrifugal filter units and further purified on a Superdex 200 increase 10/300 gl size-exclusion column. As for the PAN CC-domain, purification steps were monitored by absorbance at 214 nm, due to the lack of tryptophanes in the protein sequence. The purified protein was concentrated using 3k MWCO centrifugal concentrators and concentrations were estimated by absorbance at 214 nm or by refractivity during MALS experiments (4.5.1).

Purification of the PAN OB-ring (residues 74-150)

Cells harbouring the expressed fusion construct of PAN residues 74-150 C-terminally fused to the Trx-solubility tag, a hexahistidine tag and a 3C cleavage site were resuspended in Histrap-buffer A (20 mM HEPES, 500 mM NaCl, 10 mM MgCl₂, 10 mM imidazole and 5 mM β-ME at pH 7.0) supplemented with DNase I, lysozyme and one tablet of complete EDTA-free protease inhibitor cocktail. Lysis was allowed to proceed by incubation on a nutating shaker at RT for 30 min, after which the cells were sonicated for 20 min (amplitude 52 %, on/off: 5 s / 10 s) and cell debris was removed by centrifugation at 19.000 x g for 1 h at 4 °C. The cleared lysate was filtered with a syringe filter (0.45 μM pore size) and loaded on a 5 ml Histrap HP Ni-NTA-column preequilibrated in Histrap-buffer A. Unbound protein was washed off the column by extensively washing with Histrap-buffer A (10 CV) and 50 mM imidazole in Histrap-buffer A (10 CV). The fusion protein was eluted from the column with a gradient from 50-500 mM imidazole in Histrap-buffer A. Fractions containing the target protein were collected and buffer exchanged to Histrap-buffer A on a HiPrep 26/10 Desalting column. Cleavage of the Trx-His₆-tag was done overnight at RT using 1 mg 3C protease per estimated 10 mg of fusion protein. The next day, the cleavage mix was again loaded on the 5 ml Histrap HP column preequilibrated in Histrap buffer A. Although tag-cleaved, PAN 74-150 still bound to the column and after extensive washing with buffer A was eluted in a 10-500 mM imidazole gradient over 10 CV to separate it from the cleaved off TRx-His₆-tag and the His₆-tagged 3C protease. The target protein was exchanged to IEX-buffer A (20 mM HEPES, 100 mM NaCl, 10 mM MgCl₂, pH 7.5) and loaded on a Hitrap SP FF cation-exchange column, preequilibrated in IEX-buffer A. After extensive washing bound protein was eluted in a 100-1000 mM NaCl gradient over 10 CV. The target protein was concentrated (3k MWCO centrifugal filters) and further purified by size-exclusion chromatography on a Superdex 200 increase 10/300 GL column in storage buffer (20 mM Tris-HCl, 100 mM NaCl, 10 mM MgCl₂, pH 7.5). As for all N-terminal PAN constructs, the concentration of purified PAN 74-150 was estimated based on absorbance at 214 nm or by refractivity during MALS experiments (4.5.1).

Purification of the PAN ATPase domain (residues 150-430)

Cells harbouring the expressed fusion construct of the PAN ATPase domain (residues 150-430, F196A point mutation) C-terminally fused to the Trx-solubility tag, a hexahistidine tag and a 3C cleavage site were resuspended in Histrap-buffer A (20 mM Tris, 500 mM NaCl, 10 mM MgCl₂, 10 mM imidazole and 5 mM β-ME at pH 7.5) supplemented with DNase I, lysozyme and one tablet of complete EDTA-free protease inhibitor cocktail. Lysis was allowed to proceed by incubation on a nutating shaker at RT for 30 min, after which the cells were sonicated for 20 min (amplitude 52 %, on/off: 5 s / 10 s) and cell debris was removed by centrifugation at 19.000 x g for 1 h at 4 °C. The cleared lysate was filtered with a syringe filter (0.45 μM pore size) and loaded on a 5 ml Histrap HP Ni-NTA-column preequilibrated in Histrap-buffer A. Unbound protein was washed off the column by extensively washing with Histrap-buffer A (10 CV) and 50 mM imidazole in Histrap-buffer A (10 CV). The fusion protein was eluted from the column with a gradient from 50-500 mM imidazole in Histrap-buffer A. Frac-

tions containing the target protein were collected and buffer exchanged to Histrap-buffer A on a HiPrep 26/10 Desalting column. Cleavage of the Trx-His₆-tag was done overnight at RT using 1 mg 3C protease per estimated 10 mg of fusion protein. The next day, the cleavage mix was again loaded on the 5 ml Histrap HP column preequilibrated in Histrap buffer A. The target protein was in the flow through and was buffer-exchanged to IEX-buffer A (20 mM Tris-HCl, 500 mM NaCl, 10 mM MgCl₂, 5 mM DTT, pH 7.5) and loaded to a 5 ml Hitrap Q HP anion-exchange column preequilibrated in IEX-buffer A. After extensive washing with IEX-buffer A, the PAN ATPase domain was eluted with a 100-500 NaCl gradient in 10 CV. The protein was concentrated (10k MWCO) and further purified on a Hiload 16/600 Superdex 75 pg size-exclusion chromatography column in storage buffer (20 mM Tris-HCl, 500 mM NaCl, 10 mM MgCl₂, 2 mM TCEP, pH 7.5). Protein concentration was measured via absorbance at 280 nm and the purified protein was snap frozen in liquid nitrogen and stored at -80°C until use. For NMR-assignment experiments the PAN ATPase domain was buffer exchanged to NMR-buffer (20 mM MES, 500 mM NaCl, 10 mM MgCl₂ and 2 mM TCEP, pH 6.1) on a 5 ml Hitrap desalting column.

Purification of full-length PAN

For most experiments in this thesis, I used a PAN construct containing a C-terminal non-cleavable His₆-tag, as described in the literature:⁴⁰ Cells harboring the expressed fusion protein were resuspended in lysis buffer (20 mM Tris-HCl, 100 mM NaCl, 5 mM imidazole, 10 mM MgCl₂, pH 7.5) supplemented with DNase I, RNase I, lysozyme, 0.1 % (v/v) Triton-X-100, one tablet complete EDTA free protease inhibitor cocktail and 1mg/ml AEBSF (or Pe-fabloc) protease inhibitors. The resuspended cells were lysed by sonication and the cell debris was removed by centrifugation at 19.000 g at 4 °C for 1h. The cleared lysate was filtered (0.45 µm syringe filters) and loaded onto a 5 ml Histrap HP Ni-NTA column preequilibrated in lysis buffer. Unbound protein was removed by washing with 10 CV lysis buffer, 10 CV lysis buffer with 20 mM imidazole and 10 CV with 50 mM imidazole. The His₆-tag fusion protein was eluted in a step of 500 mM imidazole in lysis buffer. Imidazole was removed by buffer exchange (HiPrep 26/10 Desalting column in storage buffer (20 mM Tris-HCl, 100 mM NaCl, 10 mM MgCl₂, pH 7.5)). Next, PAN was loaded to a Hitrap Q HP anion-exchange column, equilibrated in storage buffer. After washing off unbound protein PAN was eluted in a 100-500 mM NaCl gradient in 10 CV. PAN was concentrated using 100k MWCO centrifugal filters and further purified on a Hiload 16/600 Superdex 200 pg size-exclusion column in storage buffer. PAN was concentrated again and protein concentration was estimated from absorbance at 280 nm. The purified protein was snap frozen in liquid N₂ and stored at -80 °C.

For unfolding and degradation studies including the 20S proteasome core particle in chapter 5.2.4 PAN was prepared from a Trx-His₆-3C-PAN fusion construct with a free C-terminus. Cells harboring the PAN fusion construct were resuspended in lysis buffer (20 mM Tris-HCl pH 7.5, 500 mM NaCl, 10 mM MgCl₂, 10 mM imidazole and 5 mM β-ME) supplemented with DNase I, RNase, complete EDTA-free protease inhibitor cocktail and lysozyme. After enzymatic lysis for 20 min cells were disrupted by sonication. Next, cell debris was removed by centrifugation (19.000 x g, 4°C, 1 h) and the cleared lysate was loaded to a 5 ml HisTrap HP Ni-NTA column preequilibrated in lysis buffer. Unbound protein was washed off the column with lysis buffer and then with 10 CV 50 mM imidazole in lysis buffer. The Trx-His₆-PAN fu-

sion protein was eluted with a gradient from 50 to 500 mM imidazole in 10 CV, and peak fractions were desalted to lysis buffer on a HiPrep 26/10 desalting column. The Trx-His₆-tag was cleaved with 3C protease overnight. The cleavage mix was re-passed over the HisTrap column. Unbound PAN was desalted to PAN storage buffer (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 10 mM MgCl₂) and loaded to a 5 ml HiTrap Q HP column for anion-exchange chromatography. After extensively washing with PAN storage buffer, PAN was eluted with 10 CV linear gradient (100-500 mM NaCl). Finally, PAN was purified on a HiLoad 16/600 Superdex 200 pg size-exclusion chromatography column in storage buffer, concentrated using a 100k MWCO Amicon centrifugal filter (Millipore) and stored at -80°C.

Purification and assembly of the 20S proteasome core particle

The *M jannaschii* 20 S proteasome was assembled from the individual domains following a protocol previously reported for the *T acidophilum* 20S CP.¹⁵⁰ Individual subunits were expressed and purified separately and afterwards assembled into the full 20S core particle.

Cells harboring the expressed α -subunits with an N-terminal His₆-tag were resuspended in lysis buffer (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 10 mM MgCl₂, 5 mM imidazole) supplemented with complete EDTA-free protease inhibitor cocktail, DNase I, lysozyme incubated at room temperature for 15 minutes and lysed by sonication. Cell debris was removed by centrifugation at 19.000 x *g* and 4 °C. The cleared lysate was loaded to a HisTrap HP 5 ml column preequilibrated in lysis buffer. Unbound protein was washed off the column with 10 CV lysis buffer. 20S- α -subunits were eluted in a step to 500 mM imidazole in lysis buffer and buffer exchanged to storage buffer (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 10 mM MgCl₂) on a HiPrep 26/10 desalting column. The α -subunits were further purified by anion-exchange chromatography on a HiTrap Q HP 5 ml column with elution by a gradient of 100-500 mM NaCl in 10 CV. Finally, the protein was concentrated (10k MWCO centrifugal filters) and passed over a HiLoad 16/600 Superdex 200 pg size-exclusion chromatography column in storage buffer. Fraction corresponding to single α -rings were collected and kept at 4 °C until assembly of the 20S.

Cells harboring the β -subunits were resuspended in storage buffer supplemented with protease inhibitors, DNase and lysozyme, incubated at room temperature for 15 min and lysed by two passages through a high-pressure homogenizer at 15.000 psi. The lysate was spun down at 19.000 x *g* at 4°C for 1h and loaded onto a 5 ml HiTrap SP HP cation exchange column preequilibrated in storage buffer. Unbound protein was removed by washing with 10 CV storage buffer and eluted with a gradient 100-1000 mM NaCl in 20 CV. β -subunits were concentrated and further purified by size-exclusion chromatography on a HiLoad 16/600 Superdex 75 pg column in storage buffer.

The 20S CP was assembled by mixing purified, but not concentrated β -subunits and α -rings and slowly shaking at 37 °C (volume \approx 30 ml). To prevent microbial growth in the preparation 0.02 % NaN₃ was added. After 6 h the mix was concentrated to 2 ml (100k MWCO centrifugal filters) at ambient temperature and further incubated at 37°C for 16h. Finally the assembled 20S proteasome was purified by size-exclusion chromatography on a HiLoad 16/600 Superdex 200 pg column in storage buffer, fractions corresponding to the $\alpha_7\beta_7\beta_7\alpha_7$ were collected, concentrated and stored at -80 °C until further use. Propeptide autocatalysis of the β -subunits was confirmed by ESI-mass spectrometry.

4.3.3.2 Purification of recombinant green fluorescent protein preparations

Purification of GFPuv-ssrA by ethanol extraction

For initial experiments including establishment of the activity assay, a first set of methyl-TROSY spectra and the resonance assignment, GFPuv was purified by ethanol extraction as described before.^{1,40,151}

E coli cells containing the expressed GFPuv protein were resuspended in lysis buffer (50 mM Tris-HCl, 100 mM NaCl, pH 7.5) supplemented with DNase I, RNase I, lysozyme, complete EDTA-free protease inhibitors, 0.1 % Triton X-100 and were lysed by sonication on ice. Cell debris was removed by centrifugation for 30 min at 19.000 x *g* and the cleared lysate was filtered (0.45 μ M syringe filters. 100 mM TEA (triethanolamine base) and 1.6 M (NH₄)₂SO₄ (dry ammonium sulfate) was added to reach 40 % ammonium sulfate saturation. The mix was incubated at 4 °C for 1h while being stirred vigorously. Next, precipitated proteins were removed by centrifugation at 3.000 x *g* for 20 min at 4 °C and more ammonium sulfate was added to achieve 70 % saturation (2.8 M (NH₄)₂SO₄). The mix was extracted with 4 volumes pure ethanol by shaking vigorously for 1 min and aqueous and ethanol phases were separated by centrifugation for 5 min at 4.000 x *g*. The aqueous phase was removed and re-extracted with 16 volumes pure ethanol. The combined ethanol phases in turn were extracted with 4 volumes n-butanol and the formed phases were separated by centrifugation at 4 000 *g* for 5 min. The final extraction was carried out with an equal volume of chloroform to remove residual organic solvents. The resulting organic phase was extracted with an equal volume of 1.2 M (NH₄)₂SO₄. The two resulting aqueous phases were combined and loaded to a 5 ml phenyl-sepharose column preequilibrated in buffer A (20 mM Tris-HCl, 0.8 M (NH₄)₂SO₄ pH 7.5). Unbound protein was removed by washing the column until a stable UV signal at 280 nm was reached. GFP was eluted in a step elution in buffer B (20 mM Tris-HCl, 100 mM NaCl, pH 7.5) and concentrated with a 10 MWCO centrifugal filter. Finally, GFPuv was purified on a Superdex 16/600 75 μ g size-exclusion chromatography column in lysis buffer, snap frozen in liquid nitrogen and stored at -80 °C.

Purification of GFPuv-ssrA and frGFP-ssrA by Ni-affinity chromatography

Initial GFPuv time-course studies and all frGFP experiments were conducted on GFP preparations from His₆-tagged recombinant protein.

Cells harboring the recombinant His₆-TEV-GFP fusion protein were resuspended in Histrap buffer A (50 mM Tris-HCl, 100 mM NaCl, 10 mM imidazole, 5 mM β -ME, pH 7.5) supplemented with DNase I, lysozyme, complete EDTA-free protease inhibitors and incubated on ice for 30 min on a nutating shaker. Next, the cells were lysed by sonication for 20 min with intervals on/off 5/10 sec (52 % amplitude settings) and subsequently the cell debris was removed by centrifugation at 19.000 x *g* for 1h at 4 °C. The cleared lysate was filtered (0.45 μ M syringe filter) and loaded onto a 5 ml Histrap HP Ni-NTA column preequilibrated in Histrap buffer A. Unbound protein and impurities were washed from the column with 15 CV buffer and the his-tagged GFP was eluted in an elution step of 500 mM imidazole in Histrap buffer A. The eluate was diluted with an equal volume of Histrap buffer A and 1 mg of TEV

protease was added to 10 mg of GFP fusion protease. Tag-cleavage was allowed to proceed at 4 °C overnight while dialyzing against a 100-fold volume of Histrap buffer A to remove excess imidazole. The next day, the cleavage mix was passed over the Ni-NTA column again to remove the His₆-tag and uncleaved fusion protein. The flow-through containing the cleaved GFP was concentrated (10k MWCO centrifugal filters) and further purified on a Superdex 16/600 75 µg size-exclusion chromatography column in storage buffer (20 mM Tris-HCl, 100 mM NaCl, pH 7.5), concentrated again, snap frozen in liquid N₂ and stored at -80 °C until use.

4.4 Biochemical assays

4.4.1 Fluorescence based GFP-unfolding assays

The unfolding and degradation activities of PAN and the PAN-20S proteasome were measured based on the loss of intrinsic fluorescence of green fluorescent proteins (either GFPuv-ssrA or frGFP-ssrA). Loss of GFP fluorescence was monitored either in a SpectraMax M3 plate reader (Molecular Devices) in fluorescence kinetics mode (excitation 405 nm, emission 509 nm) in 100 µl reactions or in a StepOne realtime-PCR system (Thermo Scientific) in 25 µl using the ROX-dye filter. In both cases ATP was added to the reaction mix at the last moment and the reaction was started in the machines, which had been preheated to 60 °C (55°C). All activity experiments were recorded as triplicate measurements. Activity curves were obtained by normalizing each data point to the initial fluorescence intensity which was set to 100 % and plotting the fractional decay over time. Samples typically contained an excess of GFP-ssrA variants over PAN/PAN-20S (5:1 up to 20:1). Further compositions of exact samples of respective measurements are indicated in the results section (chapter 5.2).

4.5 Bioanalytical methods

4.5.1 Analytical SEC and SEC-MALS

Analytical size exclusion chromatography and SEC-MALS were carried out on AEKTA pure chromatography systems (Cytiva) using different size-exclusion chromatography columns as follows:

Table 4-18 Analytical size-exclusion chromatography columns for SEC and SEC-MALS used in this thesis

Sample	Estimated Molecular weight (kDa)	Column
Full length PAN variants	300-600	Superose 6 increase, 10/300 gl
PAN coiled coil	21	Superdex 75 increase, 10/300 gl
PAN CC-OB-ring	100	Superdex 200 increase, 10/300 gl
PAN OB-ring	50	Superdex 200 increase, 10/300 gl
PAN ATPase wild type	60-180	Superdex 200 increase, 10/300 gl
PAN ATPase F196A	31	Superdex 75 increase, 10/300 gl,

		Superdex 200 increase, 10/300 gl
20S proteasome CP	700	Superose 6 increase, 10/300 gl
GFP variants	26-28	Superdex 75 increase, 10/300 gl

All analytical columns were run at a flow of 0.5 ml/min in H₂O based buffers. SEC-MALS experiments were run on an AEKTA pure system connected inline to a mini-DAWN TREOS MALS detector (Wyatt) and an Optilab T-rEX refractometer for scattering detection and accurate concentration determination, respectively. For all proteins the manufacturer-recommended refractivity increment dn/dc of 0.1852 ml/g was used to calculate the absolute the protein concentration of each chromatography volume section. The SEC-MALS experimental set-up was verified by a calibration run of bovine serum albumin (BSA, 1 mg/ml in sample buffer), to ensure detector calibration and proper alignment of UV-detector cell, MALS-detector and refractometer. All data analyses including baseline correction, Zimm-plots and molecular weight calculations were conducted in the proprietary MALS software ASTRA 7.1.4 (Wyatt).

4.5.2 ESI-protein mass spectrometry

Protein mass spectrometry of purified proteins was carried out on an HPLC-MS (Waters 2797 separation module and Waters 2545 pump system) system without chromatographic separation. UV absorbance was monitored using Waters 996 photodiode array detector and masses were identified with a Waters Micro Mass quattro micro ESI mass detector. Typical protein concentrations were between 0.1 and 2 mg /ml and 20 μ l of sample were injected using the Waters 2797 auto sampler. 100 % H₂O as a solvent and a flow of 0.2 ml/min were used for studies on GFP variants. MS-data was analysed using the system's Waters Mass-Lynx software suite.

4.6 NMR spectroscopy

All NMR experiments of the presented thesis were acquired either on a Bruker Avance III HD spectrometer with 600 MHz proton frequency and a N₂-cooled triple resonance HCN 'inverse' probe head or a Bruker Avance III HD spectrometer with 850 MHz proton frequency and a He-cooled triple resonance HCN 'inverse' probe head. Both spectrometers were operated using Centos Linux workstations running Bruker Topspin 3.2 software. Data analysis of multi-dimensional NMR spectra was carried out in CcpNMR analysis version 2.4.¹⁴³

4.6.1 Resonance assignments

All resonance assignment spectra were processed in NMRpipe¹⁴² and analysed using CcpNMR analysis version 2.4.¹⁴³ List of resonance assignments are deposited in the appendix.

PAN coiled-coil

Backbone resonance assignment of the PAN coiled-coil domain comprising residues 1-89 was carried out at a slightly acidic pH of 6.5 to reduce solvent exchange of backbone amides, which led to a reduced number of NH-cross-peaks at 7.5. Further, to improve the relaxation properties of the sample and increase the sensitivity for broader peaks all experiments were recorded at 45 °C. Sample integrity was monitored by ^1H - ^{15}N -HSQC spectra and once substantial degradation was observed, a fresh sample was used for further experiments. Sample concentrations were estimated to be between 500 and 1000 μM monomer by integrating the methyl proton signals in 1D proton spectra. A set of backbone triple resonance experiments on uniformly ^{13}C and ^{15}N labeled protein consisted of a ^1H - ^{15}N HSQC, HNCO, 3D HNCO, 3D HN(CA)CO, 3D HN(CO)CA, 3D HNCA, 3D HN(CO)CACB, 3D HNCACB.¹⁵²⁻¹⁵⁶ All assignment experiments for the PAN CC were recorded at 600 MHz proton frequency in NMR buffer (20 mM MES, 100 mM NaCl, 10 mM MgCl_2 , 0.1 % NaN_3 , 10 % D_2O and pH 6.5).

PAN OB-ring

Initial experiments on the 50 kDa PAN-ring were carried out on ^{15}N -labelled and non-deuterated samples at 600 MHz proton frequency to assess the chemical shift dispersion and feasibility of the backbone resonance assignment. Since the protein exhibits a high thermal stability, experiments were all recorded at 60 °C. Backbone resonance experiments were recorded on triple-labelled samples (uniformly ^2H , ^{13}C and ^{15}N labelled) as a set of TROSY-based experiments (^1H - ^{15}N -TROSY-HSQC,¹¹⁸ 3D TROSY-HNCO, 3D TROSY-HN(CA)CO, 3D TROSY-HNCA, 3D HN(CO)CA, 3D TROSY HN(COCA)CB, 3D TROSY-HNCACB, 3D TROSY HN(CA)CB).^{118,124,125,157} Samples were prepared at 600-700 μM monomer (100-120 μM PAN OB-ring hexamer) in NMR buffer (20 mM Tris-HCl, 100 mM NaCl, 10 mM MgCl_2 , 0.05 % NaN_3 , pH 7.5).

Assignment of methyl resonances was carried out on triple labelled samples (^2H , ^{13}C , ^{15}N) in which isoleucine $\delta 1$ -, leucine $d\delta 1/2$ - and valine $\gamma 1/2$ methyl groups were selectively protonated ("full chain labelled").¹¹⁹ The methyl resonances were connected to the C^α and C^β resonances of the respective residue using a HMCM[CG]-CBCA COSY experiment.¹¹⁹

PAN ATPase domain

Backbone assignment of the PAN ATPase domain was conducted on a monomeric monodisperse F196A mutant construct. Further, five C-terminal point mutations were included which reduce the formation of dodecamers in full length PAN (K406S, K414E, K415E, K416E, K418E, see also chapter 5.1.3). Initial backbone assignment experiments were recorded at pH 7.5 (20 mM Tris-HCl, 500 mM NaCl, 10 mM MgCl_2 , 2 mM TCEP, 0.05 % NaN_3 , pH 7.5) but yielded insufficient coverage of the backbone to achieve full assignment. A second set of experiment were recorded at pH 6.1 (20 mM MES, 500 mM NaCl, 10 mM MgCl_2 , 2 mM TCEP, 0.05 % NaN_3 , pH 6.1) to reduce solvent exchange of labile backbone amide proton. Both sets were recorded on triple labeled samples (^2H , ^{13}C , ^{15}N) as TROSY-version backbone experiments consisting of ^1H - ^{15}N -TROSY-HSQC,¹¹⁸ 3D TROSY-HNCO, 3D TROSY-HNCACB.^{124,125} Inter-type spectra (3D TROSY-HNCO and 3D TROSY-

HN(CO)CACB) were recorded at 600 MHz and intra-type spectra (3D TROSY-HNCACB) were recorded at 850 MHz.

Resonance assignment of Ile- δ 1-, Leu- δ 1/2-, Val- γ 1/2-methyl groups was carried out on similar triple labeled samples as for the backbone assignment, however with selective protonation of the respective methyl groups. A HMCM[CG]-CBCA COSY experiment was utilised to connect the methyl resonances with the known C $^{\alpha}$ and C $^{\beta}$ resonances from the backbone assignment.¹¹⁹

Methionine ϵ -methyl groups were assigned using a NOESY-based approach. For this experiment a PAN ATPase-domain sample was prepared, which was uniformly ^{15}N -, ^2H -labelled except for the Ile- δ 1-, Leu- δ 1/2-, Val- γ 1/2 and Met- ϵ , which were ^1H -, ^{13}C -labelled. A 3D ^{13}C -HMQC-NOESY- ^{13}C -HMQC^{158,159} was recorded to obtain inter-methyl NOEs. Theoretical expected inter-methyl NOEs were calculated from the PAN ATPase-domain X-ray structure (PDB entry 3H4M) with the help of a bash script written in-house by Dr. John K Kirkpatrick and used together with the experimental inter-methyl NOEs to identify methionine resonance assignments.

PAN full length

Isoleucine and methionine methyl resonances in full-length PAN were assigned by superimposing ^1H - ^{13}C -TROSY-HMQC spectra of Ile and Met-labeled full length PAN with respective spectra of the individual domains (PAN OB-ring and PAN ATPase-domain). Methionine resonances in PAN's coiled coil and OB-ring were assigned by mutating the respective residue to alanine and comparing resulting methyl-TROSY spectra with the wildtype. In a similar manner Ile resonances of the coiled coil domain have been assigned.

GFPuv-ssrA

Resonance assignment of GFPuv-ssrA focused on the methyl groups of isoleucine, leucine, valine and methionine (ILVM) for use in time-resolved unfolding studies on deuterated GFP containing selectively protonated and ^{13}C -labelled methyl groups. First, the backbone resonances of all ILV-residues and the following residues were identified with the help of a set of backbone assignment experiments (^1H - ^{15}N -HSQC, 3D HNCO, 3D HN(CO)CACB, 3D HNCACB) recorded on a uniformly ^{15}N and ^{13}C labeled sample (600 μM GFPuv-ssrA in 20 mM Tris-HCl, 100 mM NaCl, 10 mM MgCl_2 , 0.05 % NaN_3 , pH 7.5, 10 5 D_2O 42 °C). Further, confirmation was obtained from comparison to a suitable template from the BMRB (Biological Magnetic Resonance Data Bank) under entry 5666, where partial backbone assignment of GFPuv was reported.¹⁶⁰ Assignment of side-chain and methyl resonances of ILV-residues was aided by a 3D H(CCCO)NH TOCSY spectrum^{161,162} to obtain the H $^{\alpha}$ -shifts and a 3D HCCH-TOCSY spectrum^{163,164} in conjunction with a constant time ^1H - ^{13}C -HSQC to obtain side-chain ^1H and ^{13}C shifts. A 3D ^{13}C -HMQC-NOESY- ^{13}C -HMQC^{158,159} was recorded to obtain inter-methyl NOEs. Theoretical expected inter-methyl NOEs were calculated from the GFPuv X-ray structure (PDB entry 1B9C) with the aid of a bash script written in-house by Dr. John K Kirkpatrick and used together with the experimental inter-methyl NOEs to identify methionine resonance assignments. Assignment transfer from 42 °C to 60 °C was carried out by recording a temperature series of constant time ^1H - ^{13}C -HSQC spectra.

frGFP-ssrA

Backbone assignment of frGFP-ssrA was carried out on uniformly ^{15}N and ^{13}C labeled samples (600 μM frGFP-ssrA, buffer: 20 mM Tris-HCl, 100 mM NaCl, 0.1 % NaN_3 , 10% D_2O and pH 7.5) at 42 °C. A set of ^1H - ^{15}N -HSQC, 3D HNCO, 3D HN(CO)CACB, 3D HNCACB was recorded and assignment of all backbone resonances was done. Using a 3D H(CCCO)NH-TOCSY in conjunction with a 3D HCCH-TOCSY spectrum and a constant time ^1H - ^{13}C -HSQC, side-chain resonances of all methyl containing residues except for methionine (Ala, Ile, Leu, Val, Thr) was conducted. Methionine ϵ -methyl resonances could be unambiguously transferred from the assignment of the GFPuv-ssrA resonances. Assignment transfer from 42 °C to 60 °C was carried out by recording a temperature series of constant time ^1H - ^{13}C -HSQC spectra.

4.6.2 Backbone dynamics experiments on PAN's coiled coil domain

Backbone dynamics experiments on PAN's homodimeric coiled coil domain (residues 1-89) were recorded as a set of $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE, ^{15}N T_1 -relaxation and $T_{1\rho}$ experiments, on uniformly ^{15}N labeled samples (buffer: 20 mM MES, 100 mM NaCl, 10 mM MgCl_2 , 0.1 % NaN_3 , 10 % D_2O pH 6.5). Twelve T_1 -longitudinal recovery delays were chosen from 20 to 1600 ms. For $T_{1\rho}$ studies two sets were recorded with the ^{15}N field strength and relaxation delay set to 1.25 kHz and up 200 ms or 2.5 kHz and up to 100 ms, respectively, to obtain both the $T_{1\rho}$ of broad and sharp peaks, respectively.

Pseudo 3D spectra were processed in NMRPipe and integration of peaks was carried out in the NMRpipe component FuDA, where peaks were manually classified into well-separated, partially-overlapped and fully-overlapped. An initial guess of the peak's linewidth was obtained in CcpNMR analysis and used as input for FuDA. Data analysis including the fitting of peak volumes to obtain R_1 and $R_{1\rho}$ relaxation rates and the calculation of intensity ratios for the hetNOE was carried out using bash scripts written in-house by Dr. John P. Kirkpatrick.

4.6.3 NMR-titrations with the ssrA-peptide

Titrations of the ssrA-peptide to the PAN CC and PAN OB ring were done using ^{15}N labelled and ^2H , ^{15}N -labeled proteins, respectively. The lyophilised ssrA-peptide was dissolved in NMR buffer (for the CC: 20 mM MES, 100 mM NaCl, 10 mM MgCl_2 , 2 mM EDTA, 0.1 % NaN_3 , pH 6.5. For the OB-ring: 20 mM Tris-HCl, 100 mM NaCl, 10 mM MgCl_2 , 0.1 % NaN_3 , pH 7.5) and the pH was adjusted to the desired value using NaOH.

An ^1H - ^{15}N -HSQC spectrum of PAN CC at 300 μM monomer (corresponding to 15 μM dimer) concentration was recorded as an „apo“-spectrum at 45 °C. An ssrA-bound spectrum was recorded on a sample of 300 μM CC monomer concentration and 3 mM ssrA-peptide. The additional ionic strength included in the peptide stock solution due to the pH-adjustment resulted in an increase of the pI pulse of the ssrA-containing sample. To eliminate potential shift changes occurring due to different ionic strengths in the apo and holo samples, the apo sample was re-measured after the addition of NaCl to bring the pI to a similar level as for the holo sample. The final NaCl concentration in PAN CC samples was 140 mM.

For the PAN OB-ring ^1H - ^{15}N -TROSY-HSQC spectra were recorded for apo (300 μM monomer corresponding to 60 μM hexamer) and holo (300 μM monomer 1.9 mM ssrA-

peptide). Similar to the case for the ssrA-peptide in the CC-NMR buffer the ionic strength in PAN-OB-ssrA holo samples was significantly increased, such that the respective apo sample was re-measured after adjusting the ionic strength accordingly. The final NaCl concentration for all PAN-OB ring samples of the titration was 300 mM.

4.6.4 PAN/PAN-GFP CPMG relaxation dispersion experiments

CPMG-relaxation dispersion experiments on PAN and PAN/frGFP-ssrA were recorded as two sets of experiments obtained at 60 °C:

PAN apo:

20 μ M PAN-His₆ (uniformly deuterated, Ile δ 1 ¹H, ¹³C and Met ϵ ¹H, ¹³C), 50 mM ADP in buffer (20 mM Tris-HCl, 100 mM NaCl, 20 mM MgCl₂, 0.05 % NaN₃, pH 7.5 (pD 7.1))

PAN:frGFP-ssrA 1:10

20 μ M PAN-His₆ (uniformly deuterated, Ile δ 1-methyl and Met ϵ -methyl ¹H,-¹³C-labelled), 50 mM ADP, 200 μ M frGFP-ssrA (no isotopic labelling) in buffer (20 mM Tris-HCl, 100 mM NaCl, 20 mM MgCl₂, 0.05 % NaN₃, pH 7.5 (pD 7.1))

Both datasets were obtained using the same parameters of a multiple quantum relaxation dispersion experiment¹³⁴ with a constant 16 ms CPMG delay period and 14 planes including one reference plane without relaxation delay and 13 with CPMG frequencies ν_{CPMG} from 62.5 to 1750 Hz. Resulting pseudo 3D spectra were processed in NMRpipe,¹⁴² peak intensities were extracted with FuDA and R_2^{eff} relaxation rates and an approximated exchange rate $R_2^{ex,approx.}$ were fitted using bash scripts written in-house by Dr. John P Kirkpatrick.

4.6.5 Time-resolved NMR-experiments on the unfolding and degradation of GFP-ssrA

Time-resolved NMR experiments of proteasomal GFP-unfolding were carried out exclusively on the 850 MHz spectrometer. Sample were prepared freshly before each run from components previously kept frozen at – 80 °C to ensure maximal sample integrity. All components were slowly thawed at room temperature and carefully degassed under vacuum to avoid formation of micro bubbles in the NMR-sample which would otherwise cause severe signal loss due to inhomogeneities. 180 μ l reaction mixtures were combined in the following order (final sample concentrations are given in parentheses):

1. Buffer (20 mM Tris-HCl, 100 mM NaCl, pH 7.5),
2. D₂O (10 %),
3. MgCl₂ (100 mM),
4. frGFP-ssrA (100 μ M, uniformly ¹⁵N and ¹³C labelled),
5. full-length PAN (5 μ M hexamer),
6. 20S CP (5 μ M).

At the last moment ATP (100 mM, in buffer, pH adjusted to 7.5) was added and the 1.5 ml test tube was briefly centrifuged to bring the full mix to the bottom of the tube. Next, the mix was transferred to a 3 mm NMR glass tube with a glass Pasteur pipette, briefly shaken to the bottom of the tube, sealed with laboratory film and transferred to the NMR spectrometer with

a probe head heated to 60 °C (55 °C). Tuning and matching and pulse calibrations were carried out before on a “dummy” sample with identical buffer compositions. After a stable lock signal was reached the magnet was shimmed quickly to account for minor inhomogeneities. The accumulated dead-time after insertion into the preheated NMR probe was typically 2.5 min.

Finally, a time-series of ^1H - ^{13}C -SOFAST HMQC spectra was acquired to follow the reaction course. SOFAST-HMQC parameters optimized for the 100 μM ^{15}N , ^{13}C - frGFP-ssrA sample included a 120° flip angle, a recycle delay d_1 of 150 ms and an acquisition time t_2 of 50 ms. Each spectrum was recorded with 2 scans and took 100 s to be acquired. After 18 2D SOFAST spectra had been recorded a 1D ^1H proton spectrum was recorded, to allow for the monitoring of ATP hydrolysis to ADP and to control the magnet shimming. Typically, a full time series consisted of six blocks of 18 spectra. Each condition at 60 °C was conducted as triplicate measurements (at 55 ° duplicate measurements)

All 2D SOFAST HMQC of a given time series were combined into a pseudo-3D data-set and were processed in Bruker Topspin.

4.6.6 Data analysis of NMR-time series

Time dependent peak intensity changes

Pseudo 3D spectra from the time-series on frGFP-ssrA were analysed in CcpNMR analysis. Peaks were automatically picked in all pseudo-3D planes using the tool “Follow intensity changes”. Peak intensities were obtained as peak heights and the spectral noise was taken as the experimental error. Intensity changes for all peaks from three (two for 55 °C data) replicates were fitted to three different exponential decay models (equations: (4-1), (4-2), (4-3)) using the NLF-fitting tool (non-linear curve fitting) in OriginPro 9.1:

$$y = A_1 e^{(-x/t_1)} + y_0 \quad (4-1)$$

$$y = A_1 e^{(-x/t_1)} + mx + y_0 \quad (4-2)$$

$$y = A_1 e^{(-x/t_1)} + A_2 e^{(-x/t_2)} + y_0 \quad (4-3)$$

For each sampled peak the triplicate measurements were used to achieve statistical significance: Whether a single exponential (ExpDec1, equation (4-1)), double exponential (ExpDec2, equation (4-3)) or a single exponential combined with a linear decay (ExpLin, equation (4-2)) is appropriate for the triplicates or duplicates for each peak was determined by an F-test. Statistics of the NLF results of all three models were used to calculate a combined average F-ratio F_{avg} :

$$F_{avg} = \frac{\left((SS_{null,1} + SS_{null,2} + SS_{null,3}) - (SS_{alt,1} + SS_{alt,2} + SS_{alt,3}) \right) / (SS_{alt,1} + SS_{alt,2} + SS_{alt,3})}{\left((DF_{null,1} + DF_{null,2} + DF_{null,3}) - (DF_{alt,1} + DF_{alt,2} + DF_{alt,3}) \right) / (DF_{alt,1} + DF_{alt,2} + DF_{alt,3})} \quad (4-4)$$

Where $SS_{null,1/2/3}$ are the fit errors (sum of squares) of the less restrictive model (i.e. Null-hypothesis, either ExpDec1 or ExpLin) and $SS_{alt,1/2/3}$ are the fit errors of the more complicated model (i.e. alternative hypothesis, either ExpLin or ExpDec2). Accordingly, DF are the respective degrees of freedom. A (right-tailed) F probability distribution was calculated in Excel using the combined average F-ratio F_{avg} and the DF of numerator and denominator, respectively. Taking into account the triplicate measurements (or duplicate, where the denominator of equations (4-5) and (4-6) was 2) the DF of the numerator was taken from equation (4-4) as:

$$DF_{num} = \frac{\left((DF_{null,1} + DF_{null,2} + DF_{null,3}) - (DF_{alt,1} + DF_{alt,2} + DF_{alt,3}) \right)}{3} = 1 \quad (4-5)$$

Accordingly, the DF of the denominator was:

$$DF_{denom} = \frac{(DF_{alt,1} + DF_{alt,2} + DF_{alt,3})}{3} \quad (4-6)$$

If the resulting p -value was <0.05 , the model with more parameters was considered to be justified. The parameters of the appropriate model, for example average exponential time t_1 , were extracted as average of the parameters of the triplicates (or duplicates).

CSP analysis

Absolute cumulative chemical shift perturbations for unfolding and degradation runs have been calculated according to equation (4-7): $\delta_{C/H}^0$ are the reference shifts taken from the first SOFAST-HMQC of a control run without PAN or 20S. $\delta_{H/C}^\infty$ represent shifts at the end of the exponential unfolding/degradation process and was obtained from calculating a sum of all SOFAST-HMQC in a time series after the exponential process reached its asymptotic value (i.e. after the first 15 SOFAST-HMQC, roughly 45 min).

$$\Delta\delta \text{ (ppm)} = \sqrt{[\delta_H^0 - \delta_H^\infty]^2 + \left[\frac{\delta_C^0 - \delta_C^\infty}{4} \right]^2} \quad (4-7)$$

Time-dependent CSPs have been calculated by correcting for contributions of the small slowly evolving shift changes in the absence of PAN or 20S with the following equation (4-8).

$$\Delta\delta^{corr}(t) \text{ (ppm)} = \sqrt{[\delta_H^{run}(t) - \delta_H^{ctrl}]^2 + \left[\frac{\delta_C^{run}(t) - \delta_C^{ctrl}(t)}{4} \right]^2} \quad (4-8)$$

Here, $\delta_{H,C}^{run}(t)$ are the time-dependent shifts from an active unfolding/degradation run whereas $\delta_{H,C}^{ctrl}(t)$ are the respective shifts of a control run without PAN/PAN-20S.

Time-resolved CSP data were fitted using the same workflow and statistics used for the peak intensity decay data above.

5 Results

The current understanding of the mechanism of substrate processing by proteasomal ATPases is mostly based on a multitude of biochemical studies, various crystal structures of active or inactive ATPases and more recently high-resolution cryo-EM models. In the first part of the following results section I mostly used solution NMR-techniques for structural studies on the proteasomal AAA ATPase PAN from the archaeon *Methanocaldococcus jannaschii*. The second part, in turn, focuses on time-resolved solution NMR-studies on a model substrate during the processing by the PAN-20S proteasome.

5.1 NMR studies on the archaeal PAN-20S proteasome system

To make large molecular assemblies of multimeric protein complexes like the proteasome accessible to NMR studies, structural biologists often follow a so called “divide-and-conquer” approach: Large multi-domain and multimeric protein are divided into smaller fragments which retain biological relevance but are accessible by standard triple resonance NMR experiments. Information gained from studies of the isolated smaller domains is then transferred to higher molecular weight assemblies.^{25,27,122,165} In the case of PAN from *Methanocaldococcus jannaschii* ZHANG et al. identified suitable starting points for dividing the 300 kDa hexameric complex into two smaller sub-domains, of which they solved the crystal structure.³¹ After subjecting full length PAN to limited proteolysis they found that residues 74-150 formed a stable hexameric assembly (dubbed OB-ring). The isolated ATPase domain comprising residues 150-430 was reported to be a not well defined multimeric particle, which dissociated on a gel filtration column and eluted as a mixture of species ranging from \approx 180 kDa to 30 kDa. Besides the C-terminal ATPase and the N-terminal OB-ring, PAN contains an N-terminal coiled coil domain, which extends the small coiled coil part of the OB-ring. In this PhD project I set out to dissect PAN into similar sub-units (Figure 5.1), clone, express and purify the respective constructs, assign the methyl-resonances of the isolated domains and transfer them to the full length complex to facilitate NMR studies of the full-length active complex. Owing to its origin from a thermophilic archaeon, NMR experiments on PAN have been recorded at elevated temperature (mostly 60 °C).

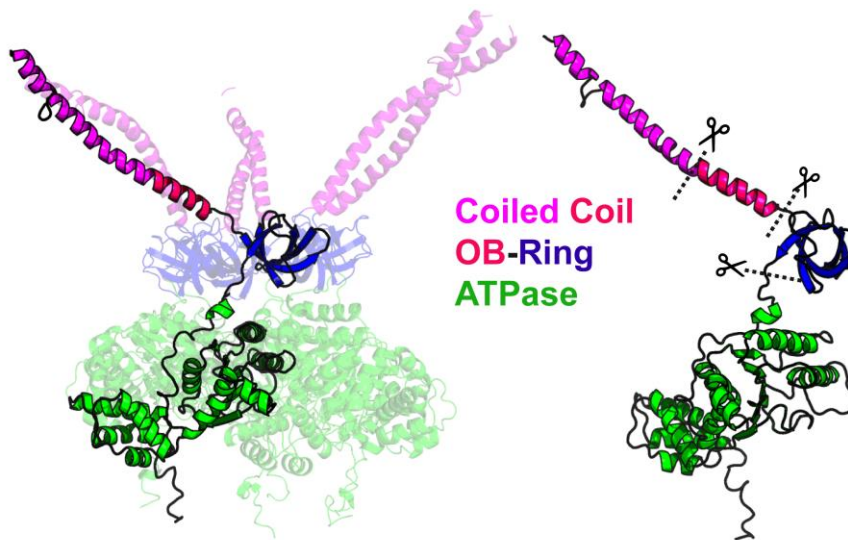


Figure 5.1 Homology model of the *M jannaschii* PAN hexamer highlighting the fragmentation sides of the “divide and conquer approach” used in this study. The boundaries of the constructs are as follows: Coiled coil residues 1-89 (magenta and pink), OB-ring residue 74-150 (pink and blue) and ATPase domain residues 150-430 (green). The homology model was created based on a cryo-EM model of PAN from *A. fulgidus* (PDB: 6HE8)⁴¹ using Swiss modeler.¹⁶⁶

5.1.1 PAN’s coiled coil domain

Proteasomes from all three domains of life share the existence of N-terminal coiled coil (CC) regions which are involved in binding to substrates and in eukaryotes also to further regulatory factors like the ubiquitin recognition- and processing machinery.⁶⁰ For PAN the N-terminal boundary of these CCs has been predicted to be around residues 45-50 reaching up to residue 89.³⁹ Contrary, residues 1-45, are considered to be unstructured and their function is not understood yet. To gain further insights into the structure and function of the full N-terminus I set out to study the full N-terminal domain by solution NMR since this technique is especially suitable for flexible and poorly structured proteins, in contrast to X-ray crystallography and cryo-EM, which have only provided low resolution information on these domains.^{31,41} As part of the “divide and conquer” approach (Figure 5.1) I generated a construct of PAN’s N-terminal domain covering residues 1-89, thus having a 15 amino acid overlap with the OB-ring construct. SEC-MALS analysis of the resulting protein revealed that the construct assembled into a stable complex of 21.75 kDa, confirming the formation of a dimer (theoret. mass.: 21.823 kDa) (Figure 5.2. A & B). Next, I investigated the general structural appearance by recording ¹H-¹⁵N-HSQC spectra of the CC-dimer (Figure 5.2 C & D). At a first glance it is clear that the spectrum contains peaks with two distinct behaviors: (1) Around 30 relatively sharp, intense and little dispersed peaks cluster towards the middle of the spectrum, indicative of regions with little structure. (2) 40-50 peaks are broader and much less intense; however they show a higher degree of chemical shift dispersion, indicative of more structured regions. To understand the identities of the respective sets of peaks, I assigned the backbone resonances of the complex. I could obtain assignments of all amide resonances except for Glu51, for which I could only assign C^γ-, C^α-, and C^β-resonances. Interestingly, although the protein is present as a dimer only one set of peaks was observed, indicating a symmetric dimer-arrangement. After peak-assignment, it became apparent that the rather

weak, more dispersed peaks mostly belong to residues located in the C-terminal half of the chain, namely the part which was proposed to have a high probability of forming a coiled-coil structure.

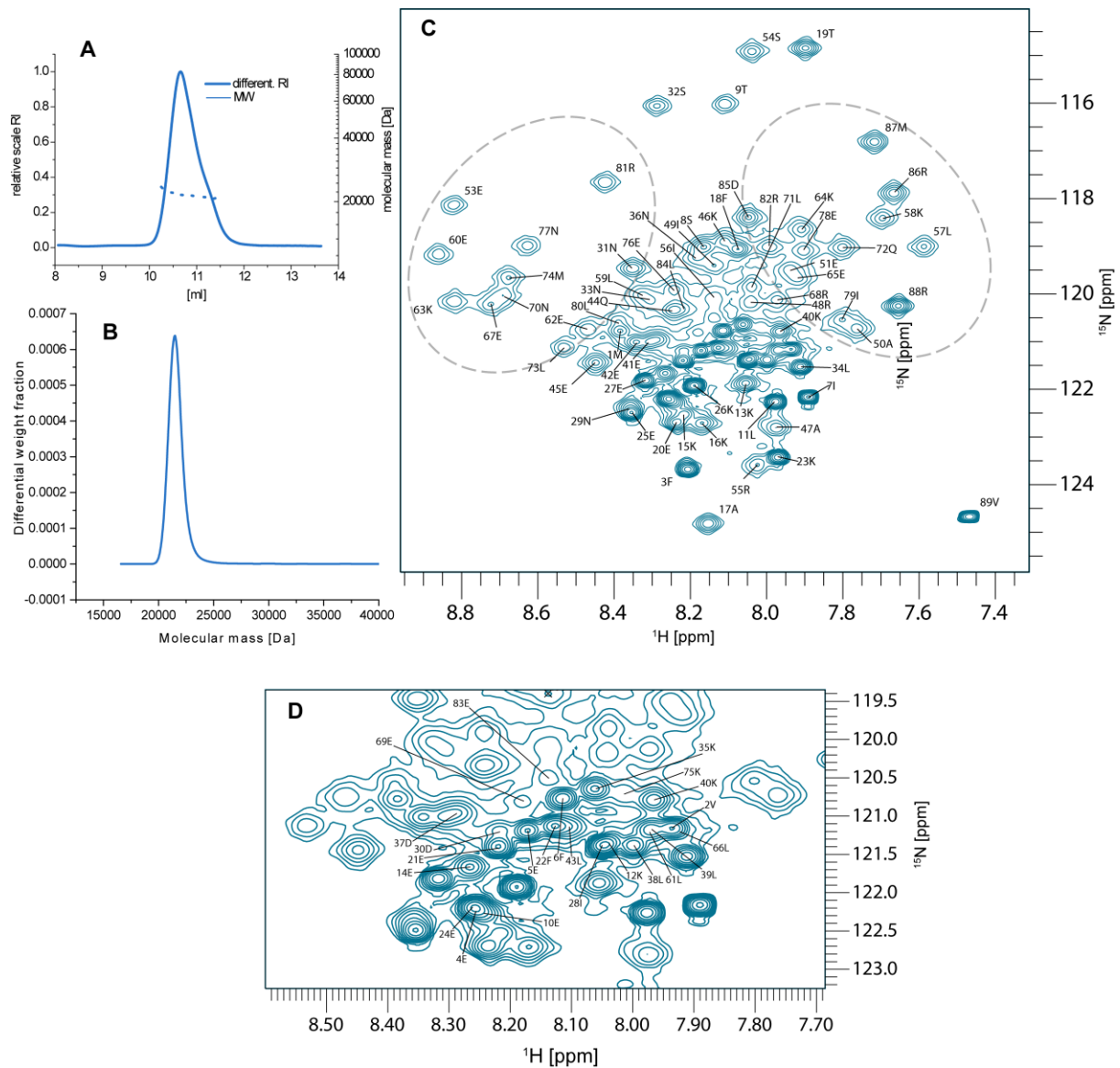


Figure 5.2 Characterization of PAN's coiled coil domain. (A & B) SEC-MALS analysis of the construct 1-89 confirms the formation of a 21.75 kDa complex, consistent with a homodimeric coiled coil. (C) ^1H - ^{15}N -HSQC spectrum of PAN's coiled coil displaying two clearly different sets of peaks: around 30 sharp and intense peaks in the center and up to 40 broader, weaker and more dispersed peaks, as shown in encircled areas. Annotations represent assignments of backbone amides. (D). Expanded region of the ^1H - ^{15}N -HSQC spectrum of (C). Generally, less dispersed and strong peaks were assigned to residues towards the N-terminal part of the sequence and more dispersed and weaker peaks to residues in the C-terminal half of the sequence, corresponding to the unstructured and structured regions, respectively. Note that for each residue only one amide peak was observed, suggesting a symmetric arrangement of the two protomers. All assignment spectra on the CC domain were recorded at 45 °C and 600 MHz.

Analysis of the secondary structure elements based on the C^γ -, C^α - and C^β -chemical shifts using the program DANGLE¹⁶⁷ revealed the boundaries of helical regions, which potentially engage in a coiled coil structure (Figure 5.3 A). According to this analysis a nearly continuously helical structure starts at Asp37 and reaches until the C-terminus. Residues 1-37 exhibit no continuous structural features with only occasional helical propensity. Additional backbone dynamics experiments on ^{15}N -labeled protein revealed consistent results: Meas-

measurements of $\{^1\text{H}\}$ - ^{15}N heteronuclear NOEs (hetNOE) for each backbone amide, which are excellent reporters on internal motions, confirm that residues up to Glu41 exhibit high flexibility, which is in good agreement with the anticipated motions of random coil regions predicted from DANGLE. Furthermore, analysis of rotating frame relaxations rates (R_2^p) reveals that the regions after residue Asn36 relax significantly faster than the flexible N-terminal tail. For all structured residues, i. e. residues with hetNOE ≥ 0.55 , an average global correlation time τ_c of 19 ns was calculated. Considering the molecular weight of 21.75 kDa, this τ_c is unusually high and reflects the anticipated asymmetric shape of the elongated coiled-coils; this τ_c also accounts for the broadness of the peaks in the ^{15}N -HSQC. All these findings point towards the fact that the coiled coil domain of PAN is fairly disordered up to residues 36/37 and transitions into an α -helical coiled-coil after this position, consistently with earlier predictions of the coiled-coil propensities made from the amino acid sequence.³⁹

Besides its assumed function in substrate engagement, the CC-domain also seems to be crucial for PAN's stability. Truncation of this domain from full-length PAN led to merely unstable protein preparations, which could not be used in further studies (section 5.1.4.2).

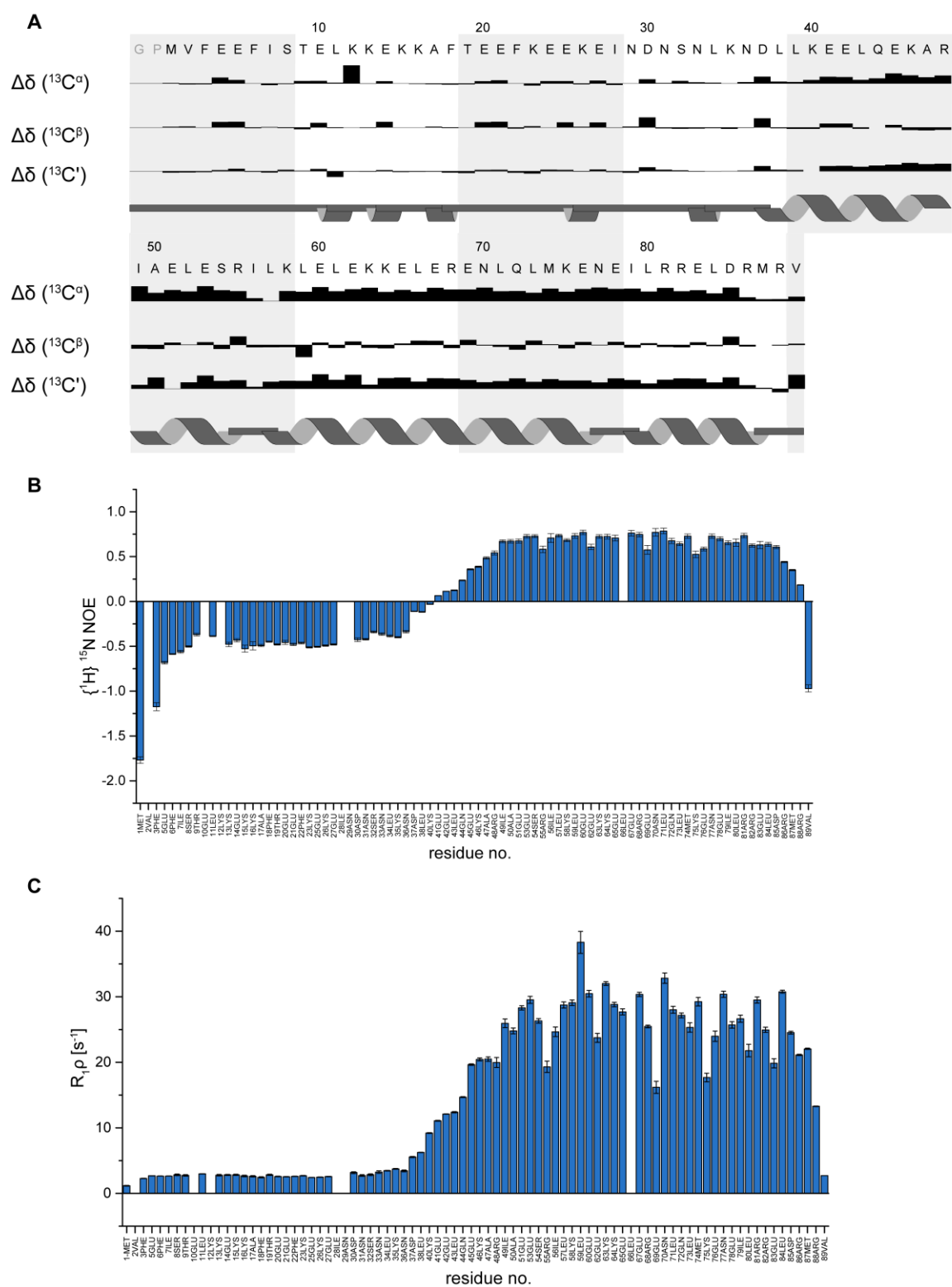


Figure 5.3. Secondary structure and backbone dynamics of PAN's coiled-coil domain. (A) Prediction of secondary structure done by the program DANGLE, confirms the findings from the backbone assignment, namely that the domain splits into an almost unstructured N-terminal region and an α -helical C-terminal region. This is further confirmed by measurements of ^1H - ^{15}N heteronuclear NOEs, which report on highly flexible regions up to residue Glu41 (B). Consistently, analysis of rotating frame relaxation rates $R_1\rho$ indicates much faster relaxation for residues in the proposed structured regions (C). An average rotational correlation time τ_c of 19 ns was calculated from all peaks with hetNOE > 0.55. This value of the correlation time is unusually high for a 21 kDa complex and indicates that the PAN-CC domain adopts a highly asymmetric shape.

5.1.1.1 Effect of the N-terminal coiled coil domain on PAN constructs

As described in the previous chapter, the full N-terminal domain of PAN comprises an unstructured followed by a structured helical region that dimerizes to form a coiled-coil structure (Figure 5.3). The biological function of the unstructured region is not entirely understood and PAN variants lacking this sequence also seem to have biological activity.³⁰ However, for the system studied here, N-terminal truncation led to reduced stability and high susceptibility to precipitation (see also chapter 5.1.4.2). To study the influence of the extended coiled-coil on hexameric PAN complexes, I generated a construct of the OB-ring (chapter 5.1.2) decorated with the full coiled-coil including the unstructured N-terminal tail (in total residues 1-150). Upon recombinant expression in *E. coli* this construct formed homogenous hexameric assemblies of ~100 kDa, as assessed by SEC-MALS (Figure 5.4 A), consistent with the findings that the OB-ring itself contributes to the hexamerization of full-length PAN (chapter 5.1.2).³¹ By means of ¹⁵N-TROSY NMR spectroscopy, we analysed the structural features of this construct: Unexpectedly, the spectrum of the hexamer of the 150-residue protein displayed only 28 backbone amide peaks (Figure 5.4 B). Further, a low chemical shift dispersion in both ¹H- and ¹⁵N-dimension indicated that the observed peaks arise from amides in region with little secondary structure. Indeed, overlay of this spectrum with ¹⁵N-HSQC and ¹⁵N-TROSY spectra of the isolated coiled-coil domain (1-89) and the OB-ring (74-150), respectively, suggests that only a part of the unstructured N-terminal regions of the construct are detected in the 1-150 construct (Figure 5.4 C & D). Due to the relatively slow tumbling of the 100 kDa complex it is not surprising that ¹⁵N-TROSY spectra do not display high spectral quality and strong signal intensities, especially, when spectra are acquired at ambient temperature. However, considering that the 1-150 sample was also perdeuterated, one could have expected at least some of the structured regions to be detectable in this experiment. Analysis of the translational diffusion coefficient of the complex resulted in a relatively small diffusion coefficient $D = 3.25 \times 10^{-11} \text{ m}^2/\text{s}$. Assuming a spherical globular shape this would correspond to a molecular mass of 1 MDa, roughly ten times bigger than expected. However, SEC-MALS confirmed the mass to be 100 kDa. These results indicate that the flexible tail and the extended coiled coil, residues 1-89, highly influence the diffusion behaviour of the complex, which adopts the peculiar shape of a rigid ring with multiple highly disordered extensions. While residues 74-150 form a rigid hexameric ring of 50 kDa, which results in NMR-spectra of decent quality (chapter 5.1.2), extending it by the full CC domain, including the disordered region, drastically changes its translational diffusion behaviour, while only doubling the molecular weight. The molecular tumbling, as described by a much longer rotational correlation time τ_c , becomes much slower than expected for a complex of this size. A longer τ_c in turn correlates with faster transverse relaxation and thus reduced signal. As a result, only the highly flexible regions in the 1-150 multimer can be detected, since their relaxation behaviour is much more favourable in the context of the slowly tumbling complex. The combination of a rigid ring with multiple highly flexible extensions is thus unfavourable for NMR studies. This is further reflected by the fact that methyl-TROSY NMR-spectra of full-length PAN exhibit only modest quality, potentially due to the peculiar shape of the rather compact OB-ring and ATPase domain, decorated with the long and highly flexible N-terminal domains (chapter 5.1.4.2).

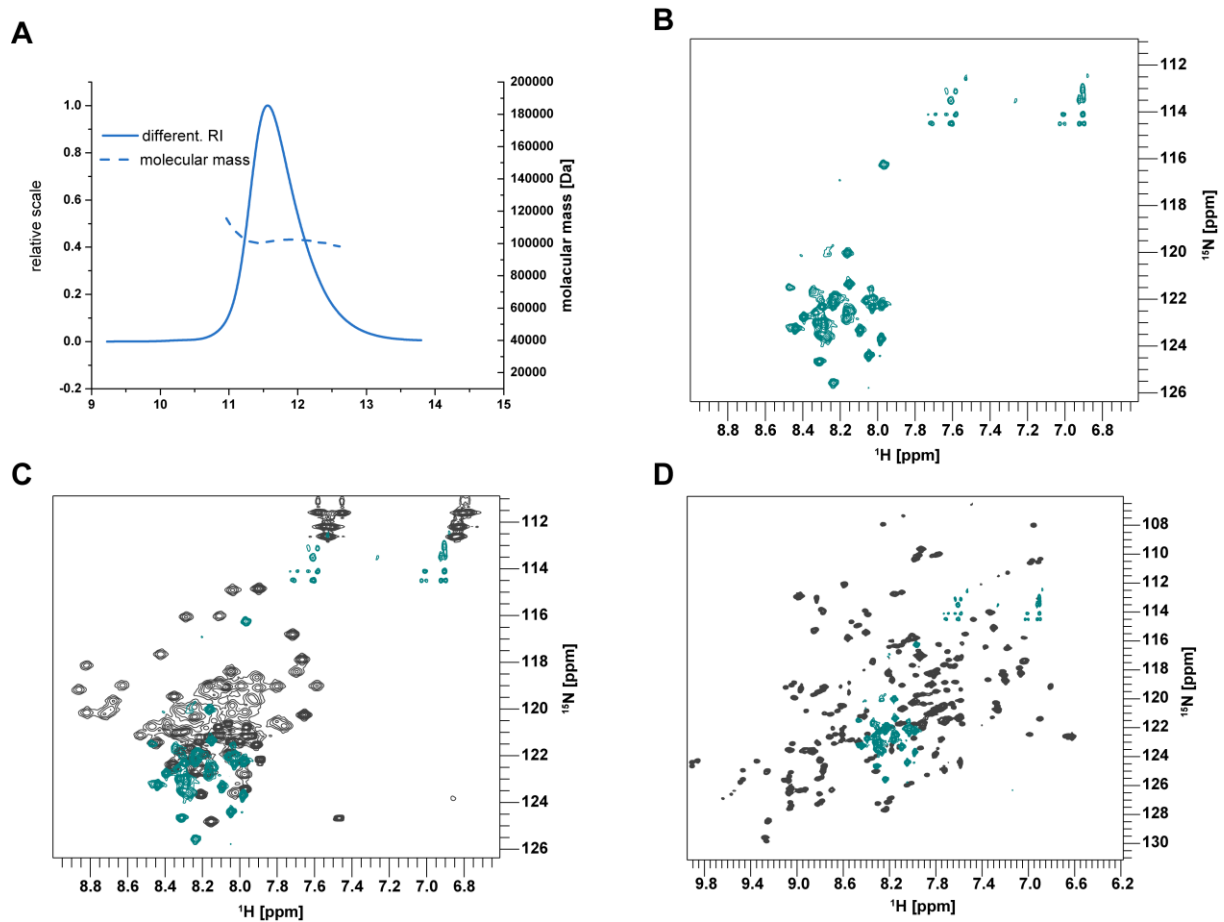


Figure 5.4 Effect of the N-terminal coiled coil domain on PAN constructs. A construct comprising the N-terminal coiled coil domain and the OB-rings is a homogenous hexamer as judged by SEC-MALS (A). ^{15}N -TROSY spectra of the 150-amino-acid construct only display 28 backbone amide peaks with little chemical shift dispersion, indicative of little secondary structure. (B). Overlay with ^{15}N -HSQC and ^{15}N -TROSY spectra (grey in C & D) of the isolated coiled-coil domain (C) and the OB-ring (D), respectively, reveals that only the far N-terminal disordered regions are detectable in the 1-150-aa-construct. An unexpectedly small diffusion coefficient of $D = 3.25 \times 10^{-11} \text{ m}^2/\text{s}$ for the 100 kDa complex suggests a highly asymmetric shape, which results in a slow molecular tumbling and thus loss of most signals, due to fast transverse relaxation. Spectra of the coiled-coil domain and OB-ring have been recorded at 45 °C/850 MHz and 60 °C/850 MHz, respectively. Spectra of the 1-150-aa-construct have been recorded at 25 °C/850 MHz.

5.1.2 The OB-ring is a hexamer with up to 6 conformations

The second domain in PAN, following the coiled coil (chapter 5.1.2), is the so-called OB-ring, owing its name to the presence of an oligonucleotide/oligosaccharide-binding fold (OB-fold). When produced as a construct of residues 74-150, whereby residues 74-89 are overlapping in sequence with the C-terminal part of the CC domain, the OB-ring has been reported to form a stable hexamer, suggesting that it has a pivotal role in the hexamerization of the full-length complex³¹. The crystal structure of this complex revealed a pseudo 6-fold symmetry of this complex, although, two monomers identical in sequence form an asymmetric dimer, which in turn is assembled into a hexameric trimer of dimers. Residues 74-89 in the asymmetric dimers form three protruding coiled coils in the hexameric donut-shaped ring composed of the six β -strand rich OB-fold (residues 90-150). In this PhD project I intended to assign the ILV-methyl resonances of this domain and subsequently transfer the assignment to the full-length protein. First, I expressed the isolated complex in *E. coli* and confirmed its mass to be 50 kDa by SEC-MALS, consistent with the formation of a hexamer (monomer mass: 8.73 kDa). Next, I studied the feasibility of the resonance assignment, by examining

the ^1H - ^{15}N -HSQC of a ^{15}N -labelled sample. The chemical shift dispersion was reasonably good, albeit I could observe 143 cross-peaks which are almost twice as many as the residues in one unit of the homohexamer. I then proceeded with the backbone assignment: Due to the relatively big size of the complex of 50 kDa, assignment experiments were recorded on uniformly ^{15}N - ^{13}C -labelled and perdeuterated samples using TROSY versions of the triple resonance experiments (chapter 2.1.2). Furthermore, elevated temperatures of 60°C have been beneficial for better sensitivity of rather weak peaks. Upon careful analysis of the HNCO, as the most sensitive triple resonance experiment providing a third dimension and thus further resolution to the ^1H - ^{15}N -HSQC, I could find 216 correlations, which would correspond to three or more different protomer conformations within the hexamer, considering that one protomer contains 71 non-proline residues. However, during the course of the backbone assignment I was challenged by the low signal-to-noise of numerous spin systems preventing the assignment of all amide peaks to a particular residue in the sequence. On the other side the final backbone assignment did cover most of the full sequence leaving only residues 74-78 and a loop-region (res. 111-114) unassigned (Figure 5.1 B). Nonetheless, for many sequence regions I could assign up to six spin systems to the same residue suggesting that parts of the OB-ring are effectively existent in up to six conformations. Generally, peaks of the same residue clustered in the same region of the ^{15}N -TROSY (Figure 5.5 A & C) suggesting that the chemical environment in the different conformations is similar. While certain sister peaks exhibited modest to severe overlap and where almost indistinguishable in triple resonance experiments, others would be well resolved and even distant in both ^1H - and ^{15}N -dimensions. For residue 143T, for example, which gave rise to six different peaks, peaks were enough well-resolved to assign them unambiguously, however four residues clustered into two pairs of two ($^{\text{C-F}}143\text{T}$) (Figure 5.5 C). Besides the multitude of sister-residues clustering in the same region of the spectrum, two residues revealed a different pattern: of Leu92 (diamonds in Figure 5.5 A), for which I could assign three distinct peaks, only two peaks were in the same region, while the third was found far away in the spectrum, suggesting a significantly different chemical environment. A second distinct feature was shown by residue Leu142 (asterisks in Figure 5.5 A): for this, three assigned peaks clustered around 118 ppm (^{15}N) whereas one was located at 122 ppm. Furthermore, the latter one was potentially a cluster of three peaks, of which two were completely lacking correlations in the HNCACB, HNCA, HN(CA)CB spectra, while they appeared at almost identical shifts in the HNCO and thus were tentatively assigned (Figure 5.5 D). Common to Leu142 and Leu92 is the close proximity in space to residue Pro 91 which had been found to be the underlying key to the OB-ring's asymmetry (Figure 5.5 E&F): Its preceding peptide bond (Pro90-Pro91) adopts alternating cis-/trans configurations in adjacent protomers thereby imposing alternating backbone configurations and allowing the formation of the coiled coil structure by the protruding α -helices of residues 74-89.³¹

Although not every residue gave rise to six different peaks it can be assumed that the hexameric complex is mostly present in six different conformations. This is further supported by the heavily asymmetric appearance of many peaks in the ^{15}N -TROSY suggestive of more than just one underlying peak, which could not be separately assigned mainly due to insufficient resolution in the ^1H -dimension of triple resonance experiments. The presence of six different conformations in the OB-ring is equally evident from ^1H - ^{13}C -HMQC spectra recorded on the ILV-methyl groups which are relevant for subsequent transfer of assignments to the full-length PAN (section 5.1.4.4). Since I only focussed on Ile and Met residues in the context

of the full length protein, I only assigned the OB-ring's two isoleucines using an HmCmCC-COSY-type experiment¹¹⁹. Thus, connecting the methyl-group resonances to the C α and C β resonances from the sequential backbone assignment I could assign 5 IleC δ 1-methyl peaks to Ile93 and two peaks to Ile79.

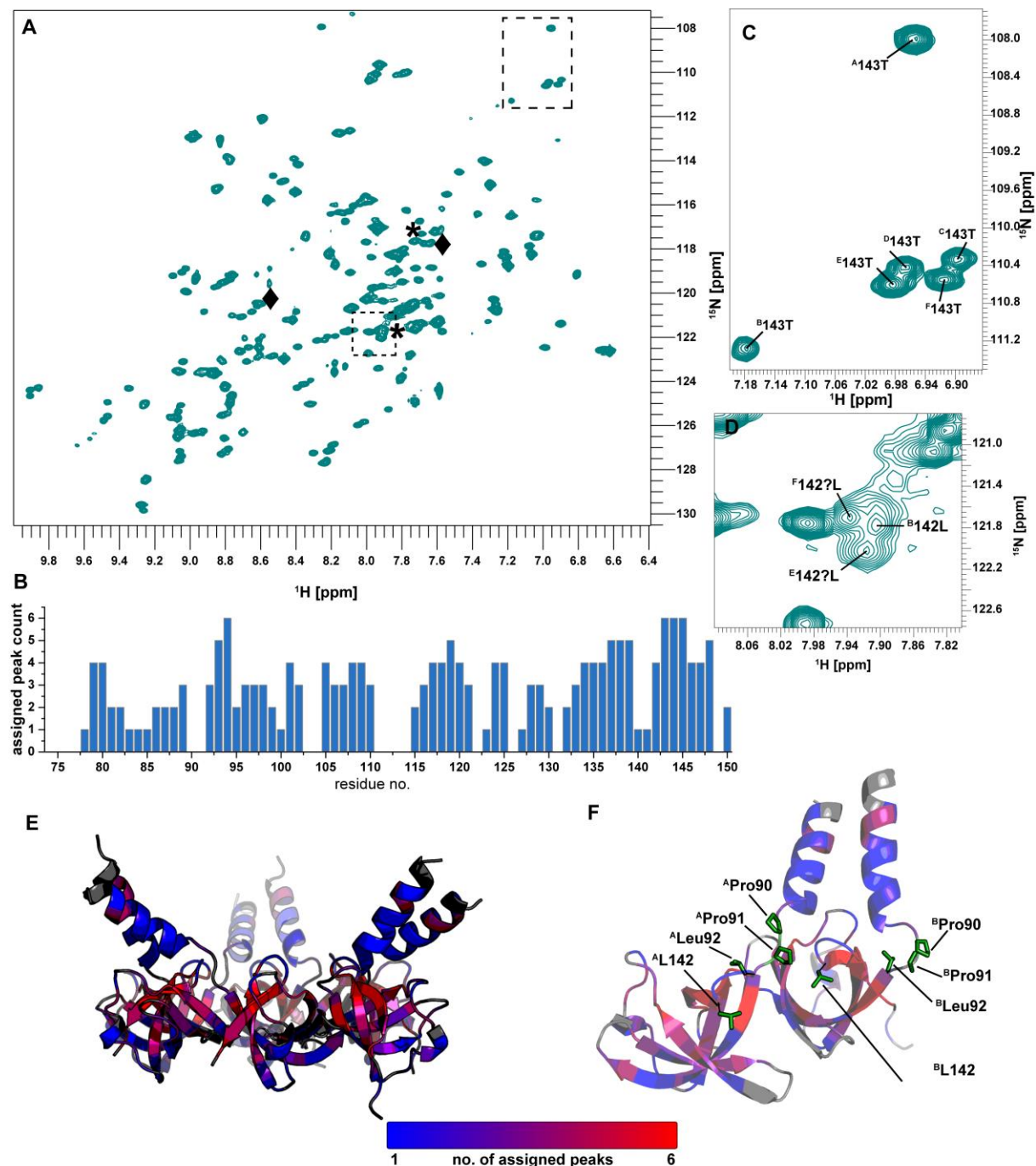


Figure 5.5 PAN's OB-ring is a heterohexameric assembly. The 77 residue OB-ring domain assembles into a hexameric ring of 51 kDa, as assessed by SEC-MALS. A ^{15}N -TROSY spectrum (A) reveals an intrinsic asymmetry in the ring, suggested by the presence of 216 assigned cross peaks. Through sequential backbone assignment it becomes evident that the OB-ring contains six distinct conformations, since up to six different peaks could be assigned for each residue (B). Generally, peaks of corresponding residues cluster in the same region of the spectrum, as exemplified by threonine 143 in (C). However, asterisks and diamonds in (A) indicate 142Leu and 92Leu, respectively, for which the respective peaks cluster into two significantly distanced clusters, suggesting entirely different chemical environments of the respective amide groups from the same residue. Indeed, the OB-ring is expected to form an asymmetric hexamer (trimer of asymmetric dimers) (E), in which an alternating cis-

/trans configuration of Pro91 is responsible for distinct backbone conformations in adjacent protomers, enabling the formation of the coiled coil structure by the α -helices in res. 74-89. Residue 142Leu and 92Leu which show the most prominent chemical shift dispersion among their conformations (diamonds and asterisks in (A)) are close to this cis-/trans-hinge of Pro91 (F) and thus experience different chemical environments in dependence of the conformation of Pro91. Although only counted as four unambiguously assigned peaks, 142Leu most likely exist also in six conformation since ^{13}C 142Leu clusters with two additional peaks with almost identical HNCO-resonances (D), which however lack any resonances in HNCA, HNCACB, HN(CA)CB and were thus not considered in the sequential backbone assignment. All spectra on the PAN OB-ring were recorded at 60 °C and 850 MHz.

5.1.3 ATPase domain

Conversion of chemical energy into mechanical force for the unfolding of substrate proteins is carried out by PAN's ATPase domain, which hydrolyses the high energy phosphoric anhydride bond of ATP and subsequently undergoes large scale conformational changes which facilitate substrate binding, unfolding and translocation. The X-ray crystal structure of this nucleotidase domain was solved by ZHANG et al. and resembles the overall canonical fold of AAA ATPases, comprising an α/β fold for nucleotide binding and a small helical domain at the C-terminal end.³¹ The nucleotide binds between the α/β fold and the helical domain implying a potential rearrangement of these two subdomains relative to each other upon binding and hydrolysis of ATP (Figure 1.9). Although the X-ray structure was solved by molecular replacement using a template (p97 AAA ATPase) which crystalized as a hexamer, this nucleotidase domain did not form a stable hexameric complex on its own.

To enable NMR studies of the full-length PAN I set out to assign the backbone and methyl resonances of the ATPase domain which spans from residue 150 up to the C-terminus. Unfortunately, as described above the domain on its own is notoriously instable, which manifested in precipitation of the target protein upon the first step of chromatographic purification. Precipitates could be dissolved by heating to 60 °C, however reprecipitated soon after. Nevertheless, I was able to analyse the oligomeric state of this instable domain by SEC confirming the findings of ZHANG et al. who found the complex to dissociate on the gel filtration column. As a result, the sample eluted as mixture of hexamers and smaller oligomeric states down to monomers (Figure 5.6 B). In this state acquisition of high-quality NMR-data did not seem feasible and optimization of sample quality was required to obtain a monomeric ATPase domain of 31 kDa, which would allow traditional triple resonance-based assignment experiments. Testing of a multitude of buffers conditions by thermal shift assays and alternative construct boundaries (150-430, 155-430, 166-430 and 176-430) was not successful. Upon thorough literature research on hexameric AAA+ ATPases and mutations leading to monomers thereof, I found that CAILLAT et al. could obtain pure monomers of the AAA+ ATPase MsVps4 from a previously poorly defined oligomeric state introducing a Phe126 to Ala mutation which abolishes inter-protomer interactions.¹⁶⁸ Since the ATPase domains of PAN and MsVps4 share a high sequence similarity (42 %) and a homologous phenylalanine residue (Phe196) is present in PAN, testing this point mutation on PAN appeared to be promising. Furthermore, careful examination of inter-domain contacts in the PAN ATPase crystal structure revealed a hydrophobic interaction network around Phe196 between two monomers, emphasizing the relevance of this residue for inter-protomers interactions (Figure 5.6 A). Indeed, after introducing the point mutation Phe196Ala into the ATPase domain all precipitation/aggregation issues were resolved and the domain could be obtained in highly soluble (up to 1 mM or 30 mg/ml) quantities. The oligomeric state was confirmed by SEC-MALS to be perfectly monomeric with a molecular weight of 32.8 kDa (calculated mass: 31.4 kDa).

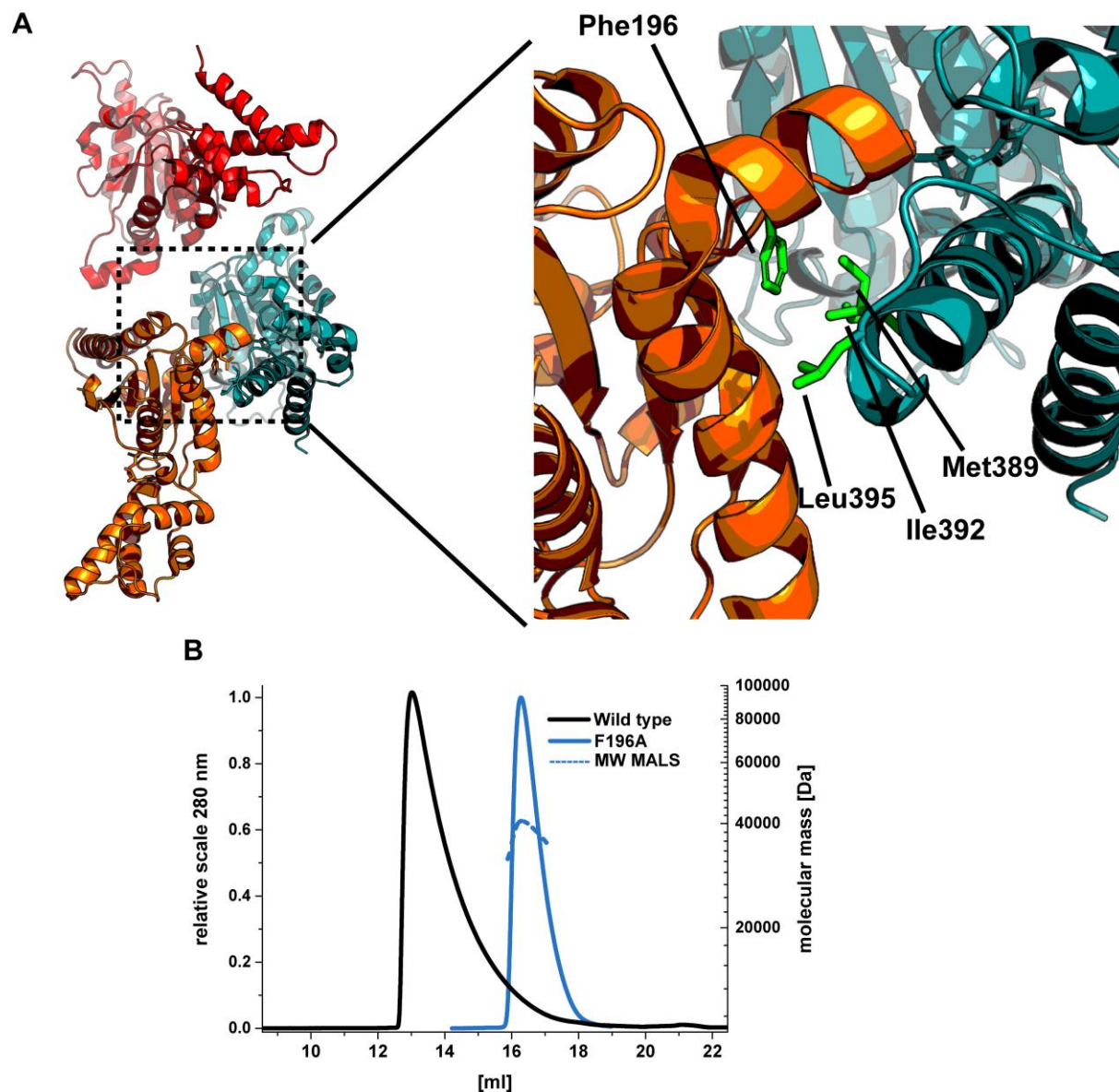


Figure 5.6 Inter-protomer interactions in PAN's ATPase depend on a hydrophobic network around Phe196. (A) The PAN ATPase domain was crystallized with three monomers in the asymmetric unit (PDB: 3h4m) which make contacts via a hydrophobic network (indicated by green highlighted side chains in the right inset) around Phe196 on one monomer and Met389, Ile392 and Leu395 on the adjacent monomer. A homologous interaction network was also found in the crystal structure of MsVps4 (PDB: 4d80) where mutation of the Phe to Ala led to exclusively monomers of the respective ATPase. Correspondingly, mutating Phe196 in PAN's ATPase domain to Ala resulted in the formation of pure monomers as judged by SEC-MALS (B). The wild type construct under the same conditions elutes from the gel filtration column as a mixture of hexamers and smaller oligomers (black trace). SEC- and SEC-MALS runs were carried out on a Superdex S200 10/300 column at 500 mM NaCl.

With a monomeric version of PAN's ATPase domain in hand, assignment of backbone resonances appeared feasible and was carried out on a ^{15}N - ^{13}C -labelled and perdeuterated sample. Due to its excellent thermal stability the experiments were acquired at 60°C , which resulted in high quality ^{15}N -TROSY spectra with little spectral overlap and uniform peak intensities (Figure 5.7). Initial trials led to only incomplete backbone resonance assignment, lacking several stretches of the protein sequence, which comprises several loops. After re-recording triple resonance experiments at a lower pH (6.1) ^{15}N -TROSY spectra displayed 32 more peaks compared to the higher pH. Overall, 3D spectra were of equally good quality, which facilitated assignment of $> 94\%$ of backbone amide resonances. Subsequently, reso-

nances of Ile δ 1, Val γ and Leu δ -methyl-resonances were assigned aided by an HmCmCC-COSY experiment, which links the methyl resonances to the previously assigned C $^{\alpha}$ - and C $^{\beta}$ -resonances of the same residue.¹¹⁹ With most ILV-methyl resonances assigned, assignment of Met ϵ -methyl-resonances was carried out using a ^{13}C -HMQC-NOESY experiment. In total, 8 out of 9 Met ϵ , 22 out of 24 Ile δ -1-, 41 out of 50 Leu δ - and 40 out of 40 Val γ -methyl-resonances could be assigned unambiguously.

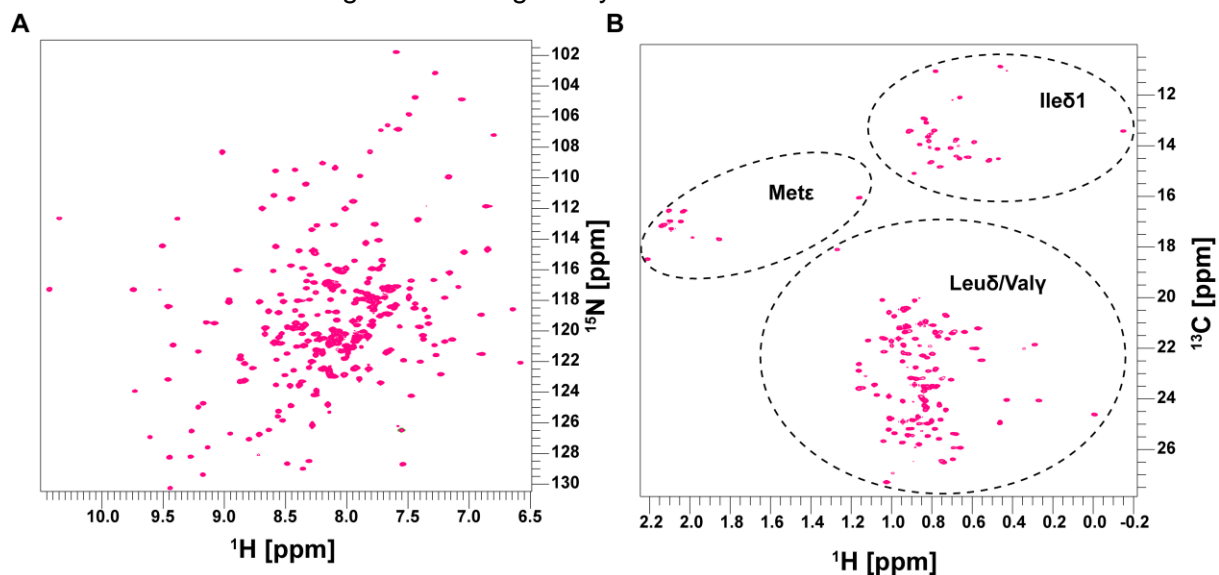


Figure 5.7 2D NMR spectra of PAN's ATPase. ^{15}N -TROSY (A) and ^{13}C -HMQC (B) of PAN's ATPase exhibit excellent resolution and homogeneously peak intensities, indicative of a well-behaved protein. 94 % of all backbone amide resonances of the 31 kDa ATPase domain were assigned using triple resonance experiments. Subsequently, assignment of > 90 of ILVM-methyl resonances could be achieved. Spectra were recorded at 60 °C and 850 MHz ^1H frequency and samples of ~ 600 μM .

5.1.4 Studies on functionally working full length PAN

5.1.4.1 Full length PAN exists in a dynamic equilibrium between hexameric and dodecameric species

As for all the single domains of PAN, NMR studies of the full length PAN require a soluble monodisperse and homogenous sample. Early studies of full length PAN from different archaea report the unfoldase to form complexes of 500-600 kDa in its isolated form as assessed by analytical size exclusion chromatography and appropriate calibration of the size exclusion column.^{29,30,169} Because a monomer has a molecular weight of 47 kDa, these 500-600 kDa would correspond rather to dodecameric than hexameric assemblies. However, PAN is thought to function as a hexamer which sits on top of the 20S core particle's heptameric ring of alpha-subunits, which is also consistent with cryo-EM models of the PAN-20S proteasome from *A. fulgidus* and the evolutionary related regulatory particles from eukaryotes.^{41,65} A recent study on PAN in which IBRAHIM et al. conducted small angle neutron scattering experiments on PAN reported a protocol to "hexamerize" the complex by supplementing the sample with 4 mM ATP and 10 mM MgCl_2 (buffer otherwise: 20 mM Tris-HCl, 100 mM NaCl) and incubation at 60 °C for 30 min before injecting it into an analytical size exclusion chromatography column.⁴⁰ By means of size exclusion chromatography coupled to multi angle light scattering (SEC-MALS) we found these condition not to be sufficient to obtain a monodisperse hexameric PAN (Figure 5.8). For projected methyl-TROSY NMR studies of the full-length PAN monodispersity was a strict requirement, at a minimum required complex

concentration of 20 μM , corresponding to 120 μM of PAN-protomer. Starting with a mixture of dodecamers (600 kDa) and hexamers (300 kDa) we set out to optimize sample conditions in favour of PAN hexamers using the aforementioned 120 μM PAN protomer throughout this project.

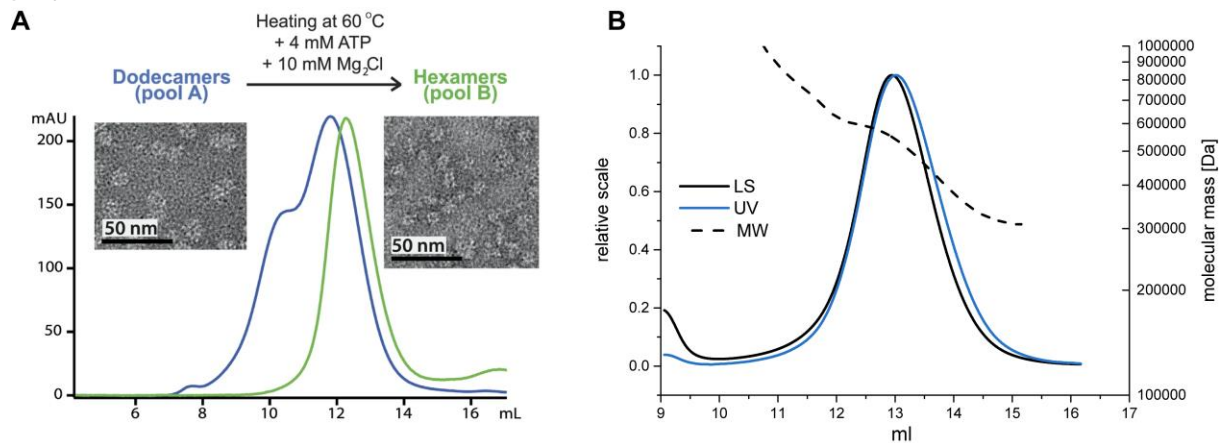


Figure 5.8 10 mM ATP and heat treatment is not sufficient to obtain a homogeneously hexameric PAN sample. A) Hexamerization protocol from Ibrahim et al. requires heating to 60 $^{\circ}\text{C}$ in presence of ATP and MgCl_2 , resulting in a shift to the right on a Superose 6 10/300 GL and a disappearing of a shoulder indicative of dodecamer removal. Negative stain electron microscopy also showed the presence of mostly smaller species. (Figure adapted from IBRAHIM et al.⁴⁰). (B) Using SEC-MALS we found in our lab that the same conditions yield a mixture of dodecamers (600 kDa) and hexamers (300 kDa) as indicated by the molecular weight measurement (dashed line, ranging from 600 to 300 kDa within the peak). Furthermore, the offset of the light scattering-(LS) and UV-trace is an indicator of molecular weight inhomogeneity.

The effect of buffer ionic strength on the oligomeric state of protein complexes is a widely accepted concept.¹⁷⁰ I thus screened buffer conditions with several different monovalent salt concentrations by SEC-MALS. Indeed, increasing the NaCl concentration from 100 up to 1000 mM induced a change in the SEC-MALS-profiles (Figure 5.9): Higher salt concentrations resulted in a delayed elution from the SEC-column and a better separation of higher oligomeric species (dodecamers and bigger) and hexamers. Furthermore, the mass distribution shifted drastically to lower molecular weight species. While more than 70 % of detected PAN species were bigger than 500 kDa at 100 mM NaCl, at 500 mM more than 80 % was smaller than 400 kDa. Still, the homogeneity was not satisfactory.

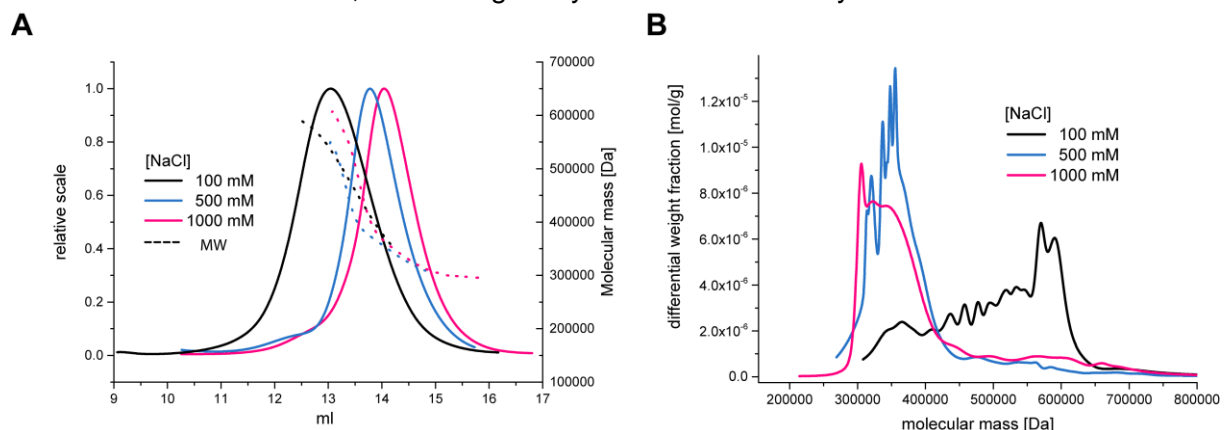


Figure 5.9 The oligomeric state of PAN is dependent on ionic strength. (A) With increasing NaCl concentration PAN elutes later on an analytical SEC column with a big leap going from 100 to 500 mM NaCl. (B) The mass distribution of PAN species changed from > 70 % larger than 500 kDa at 100 mM NaCl to > 80 % smaller than 400 kDa at 500 mM NaCl. Further addition of salt to 1000 mM NaCl led to only smaller shifts in elution time and no significant improvements in PAN's mass distribution.

Generally, high-ionic-strength-buffers should be avoided in NMR experiments, since they result in higher electrical resistance and thereby cause compromised sensitivity.¹⁷¹ Instead of increasing the buffers ionic strength the introduction of point mutations which change the surface charge at selected positions in the protein represents also a valid strategy to prevent self-association into oligomers. Assuming that formation of dodecamers occurs by stacking of two PAN hexamer rings on top of each other, three orientations are conceivable: C to C, N to N, C to N. ZHANG et al. found in their crystallographic work on PAN's isolated domains that the proximal face of the hexameric ATPase-ring is highly basic. Based on the calculated electrostatic surface in water we identified five lysine residues which are mainly responsible for this positive charge in the C-terminal region of PAN. To invert these charge hot spots we replaced four lysines with negatively charged glutamates and one with serine (K406S, K414E, K415E, K416E, K418E) (Figure 5.10).

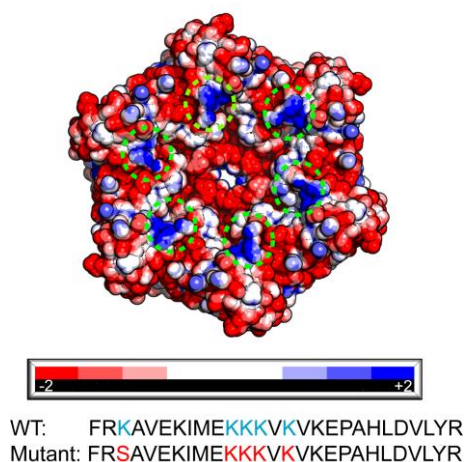


Figure 5.10 PAN's ATPase ring exposes basic hot spots on its C-terminal face. Electrostatic surface of PAN's ATPase ring calculated by the program APBS¹⁴⁴ unravels basic patches at the C-terminus. Five lysine residues on all of the six protomers were identified most responsible for the basic character of this face of the ATPase ring (green encircled). A mutant was generated replacing four lysines by acidic glutamate and one lysine by the non-charged hydrophilic amino acid serine.

SEC-MALS analysis of the PAN mutant revealed a significant improvement compared to the wildtype (Figure 5.11). The molecular weight at 100 mM NaCl is shifted to smaller PAN species with 90 % of the sample being smaller than 450 kDa, opposed to the wildtype which is mostly larger than 450 kDa. At 500 mM NaCl the mutant, like the wildtype, appears to associate in much more homogenous hexamers with the majority displaying a size between 300 and 400 kDa. Nevertheless, the difference in the molecular weight distribution between mutant and wildtype is not as strong as at 100 mM: The mutant displays less than 10 % higher molecular weight species, while the wild-type still contains more than 15 % higher molecular weight species. Although this ratio does not represent a perfectly homogenous and monodisperse hexamer, less than 10 % non-hexameric impurities in the sample were considered to be acceptable for later NMR experiments. High ionic strength buffers seemed to be required even for the mutant and were thus also used for initial NMR experiments.

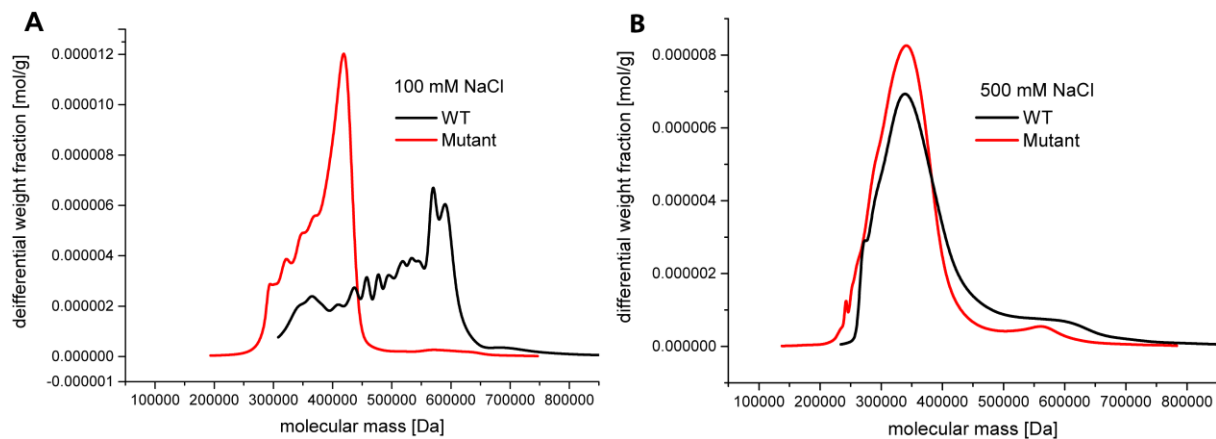


Figure 5.11 SEC-MALS analysis of PAN and PAN mutant. Introduction of five C-terminal mutations inverting the local electrostatic surface improved the homogeneity of the sample (A). In high salt (500 mM NaCl) conditions, the mutant is a clearly more homogenous hexamer compared to the wild type (B).

5.1.4.2 Optimization of full length PAN constructs to improve sample quality for NMR experiments

To investigate the feasibility of structural studies of PAN and its substrates by NMR we prepared an initial NMR sample of the PAN mutant (see section 5.1.4.1) in the optimized buffer conditions. PAN was fully deuterated except for the terminal methyl groups of isoleucine ($\delta 1$), leucine ($\delta 1/\delta 2$) and valine ($\gamma 1/\gamma 2$) (see also section 2.1.2) which were ^1H - and ^{13}C labelled. Unexpectedly even for a complex of this big size, the ^1H - ^{13}C - HMQC spectrum was only of modest quality: Most peaks were broad, weak and significantly overlapped except for some very sharp intense ones in the middle of the spectrum (Figure 5.12). Spectral overlap was not completely unanticipated, since PAN is present as a homohexameric assembly, with possibly up to six different conformations and signals for each of the expected 192 methyl groups. The translational diffusion coefficient of this sample however, was measured to be $D \approx 4.5 \cdot 10^{-10} \text{ m}^2 \text{ s}^{-1}$, which corresponds to molecular weights of $\approx 350 \text{ kDa}$ applying a shape factor for non-symmetric assemblies. Although this is roughly consistent with the expected molecular weight of the complex, a clear deviation of the Gaussian fit from the experimental peak intensities indicates a substantial heterogeneity with respect to the diffusion coefficient of species contained in the sample (Figure 5.12 B). Addition of the slowly hydrolysable ATP-analogue ATP γ S, which PAN is thought to bind but not hydrolyse and thus restrict its conformational landscape, did not alter the characteristics of the spectrum. Furthermore, spectra in phosphate buffer were acquired since thermal shift assays indicated that PAN is most thermodynamically stable in presence of phosphate. Thus, by binding to PAN's nucleotide binding pocket phosphate might also be able to block PAN's dynamics and improve the spectral quality. However, phosphate has not proven to change the spectral appearance.

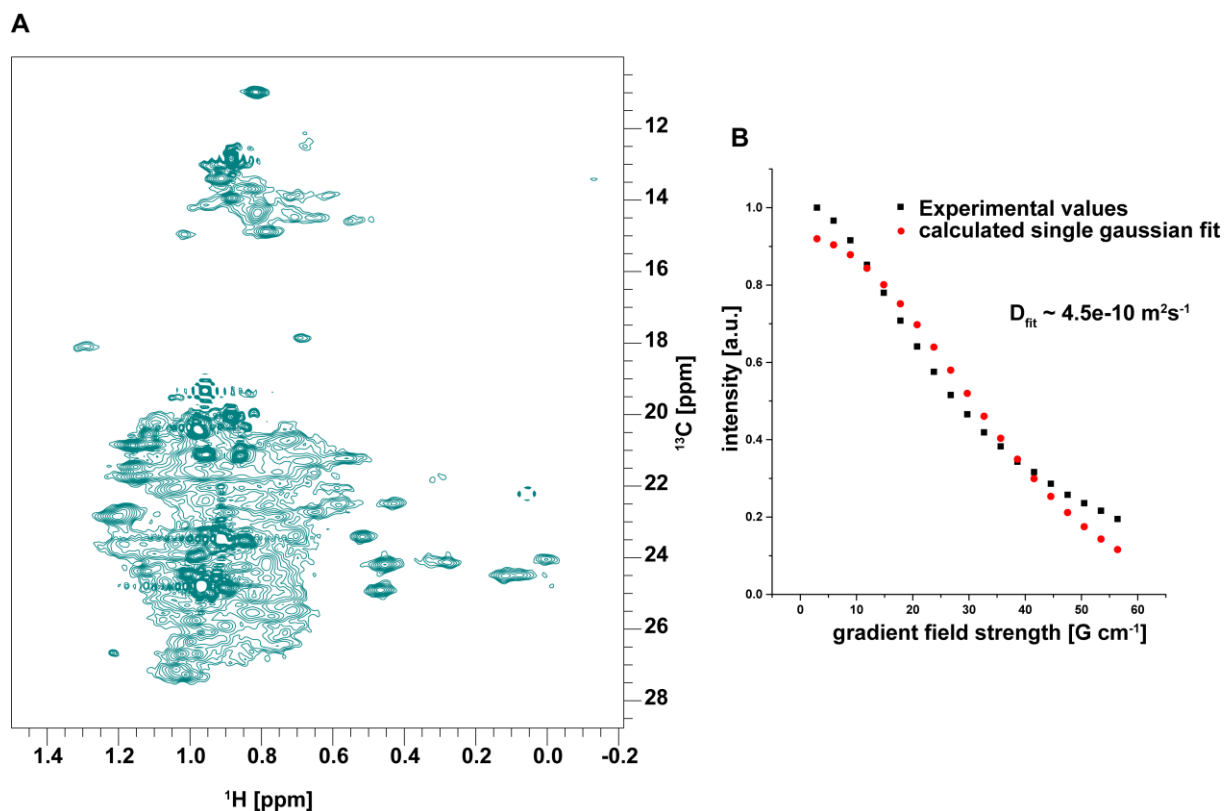


Figure 5.12 NMR-analysis of the full-length PAN mutant. (A) A ^1H - ^{13}C -HMQC spectrum was recorded on an ILV-labelled sample of PAN containing five C-terminal point mutations, which invert its surface charge to minimize dodecamer-formation. The general appearance of most peaks is broad and results in poor resolution. (B) Analysis of the sample's diffusion coefficient indicated a strong inhomogeneity with respect to the diffusion coefficient, as illustrated by the deviation of the single Gaussian fit (red) to the experimental diffusion peak intensities (black). The fit resulted in a D of $\sim 4.5 \times 10^{-10} \text{ m}^2\text{s}^{-1}$ with a corresponding hydrodynamic radius $R_h \sim 5.2 \text{ nm}$ and molecular weight of $\sim 350 \text{ kDa}$. However, these only represent weighted averages of all species with different diffusion coefficients present in the sample.

Since high quality structural and dynamic studies of proteins crucially depend on well resolved two-dimensional NMR spectra, I set out to identify the reasons for this unusual relaxation behaviour and find strategies to optimize sample conditions even further than described in section 5.1.4.1:

Apparently, initial optimization trials to obtain a nearly homogenous hexameric PAN had not been sufficient. Despite optimized buffer conditions and beneficial point mutations described in section 5.1.4.1 changed PAN's oligomeric state towards hexamers, some part or property of PAN kept causing the unusual relaxation behaviour leading to low quality methyl-TROSY spectra. Removal of potentially conformationally exchanging parts of PAN appeared to be a solution to this problem. As illustrated in Figure 5.1 PAN comprises an extended coiled coil domain towards its N-terminus (residues 1-73), the role of which is not yet understood. Constructs comprising this coiled coil domain and the OB-ring, without the ATPase domain, yielded ^{15}N -TROSY spectra of poor quality and an unusually slow translational diffusion, suggesting an abnormal relaxation behaviour induced by the extended coiled coil-domain (see also 5.1.1.1). Further, the far-N-terminal region of PAN was found to be highly flexible, potentially causing severe conformational exchange in the context of full-length PAN. Removal of this domain represented a promising strategy to influence the overall relaxation behaviour and reduce chemical exchange effects. Furthermore, studies have shown that the CC domain is not strictly required for PAN to form hexamers and PAN ($\Delta 1-73$) might even be

biologically relevant, since the native PAN gene contains a second translation initiation site leading to mixed hexamers of full length and $\Delta 1-73$ variants when expressed from non-codon-optimized genes in *E. coli*.^{29,30} These findings together allowed me to design an improved PAN ($\Delta 1-73$) variant which retains functional activity. Expression trials of this variant yielded about ten times more soluble protein after affinity purification compared to the wild-type protein. Unfortunately, this protein was extremely prone to aggregation at concentrations > 1 mg/ml, which rendered it unsuitable for NMR experiments under these conditions. However, heating to 60°C or addition of 50 mM ATP – but not ADP – to concentrated protein solutions rapidly dissolved macroscopically visible aggregates. I hypothesized that ATP binding induces conformational changes to a state in which inter-PAN interactions are minimized, thus yielding a soluble sample. Since purification on a preparative size-exclusion chromatography column including 50 mM ATP at high temperature was not feasible, further optimization of expression- and buffer conditions was necessary. Precipitation and instability can also be caused by too high concentration of expressed proteins in the bacterial cell; thus, optimization trials also included shorter expression times at lower temperatures to decrease the amount of total overexpressed protein. Additionally, expression without an N-terminal thioredoxin-solubility tag was tested. Nonetheless in both cases no significant improvement of protein stability could be found. High and low salt (100 to 1000 mM NaCl), and addition of up to 1 % detergent like Tween-20 or Triton X-100 did also not help keeping PAN ($\Delta 1-73$) in solution during purification. Furthermore, like for the full length protein, a thermal shift assay suggested 50 mM sodium phosphate as an improved buffer, which unfortunately also did not alter the protein stability. Besides these trials only mutational studies appeared to be promising to obtain a well-behaved PAN ($\Delta 1-73$):

Mutation of phenylalanine 196 to alanine (F196A) led to substantial solubility improvement of the isolated ATPase domain. As described in section 5.1.3 Phe196 resides at the interface between two PAN protomers. Thus, this mutation should lead to altered inter-subunit interactions and thereby might also decrease PAN's dynamic structural exchanges leading to a more stable protein sample. Indeed, the introduction of the point mutation F196A into PAN ($\Delta 1-73$) drastically improved its solubility and a hexameric complex could be prepared by size-exclusion chromatography in high yields, which did not precipitate up to 30 μ M complex concentration (180 μ M monomer). SEC-MALS analysis confirmed the formation of a monodisperse PAN ($\Delta 1-73$) hexamer with a calculated molecular weight of 253.5 ± 0.4 kDa (Figure 5.13). Interestingly, the complex elutes ca. 1.4 ml later on a Superose 6 increase 10/300 GL SEC column, which is more than what was expected from a change in molecular weight less than 100 kDa, suggesting that PAN ($\Delta 1-73$) F196A is not only smaller in molecular weight but also significantly different in its shape. These findings raised the hope that this construct would yield methyl-group NMR spectra of good quality. I prepared an ILV-methyl-labelled and otherwise perdeuterated sample and recorded ^1H - ^{13}C -HMQC spectra (Figure 5.13 C). The spectra were of excellent quality with good resolution and sharp peaks for most of the signals. Furthermore, superimposition of the ILV-spectrum of PAN ($\Delta 1-73$) with the ILV-spectra of the OB-ring construct and of the isolated ATPase domain, respectively, indicated satisfactory agreement between the isolated domains and the hexameric ATPase, which is crucial for further assignments transfer from the smaller constructs to the full-length protein. Unfortunately, activity assays, conducted monitoring the unfolding of the model substrate GFP-ssrA, revealed that this truncated PAN variant is inactive, which also rendered it not suitable for further investigations (see Appendix Figure 1).

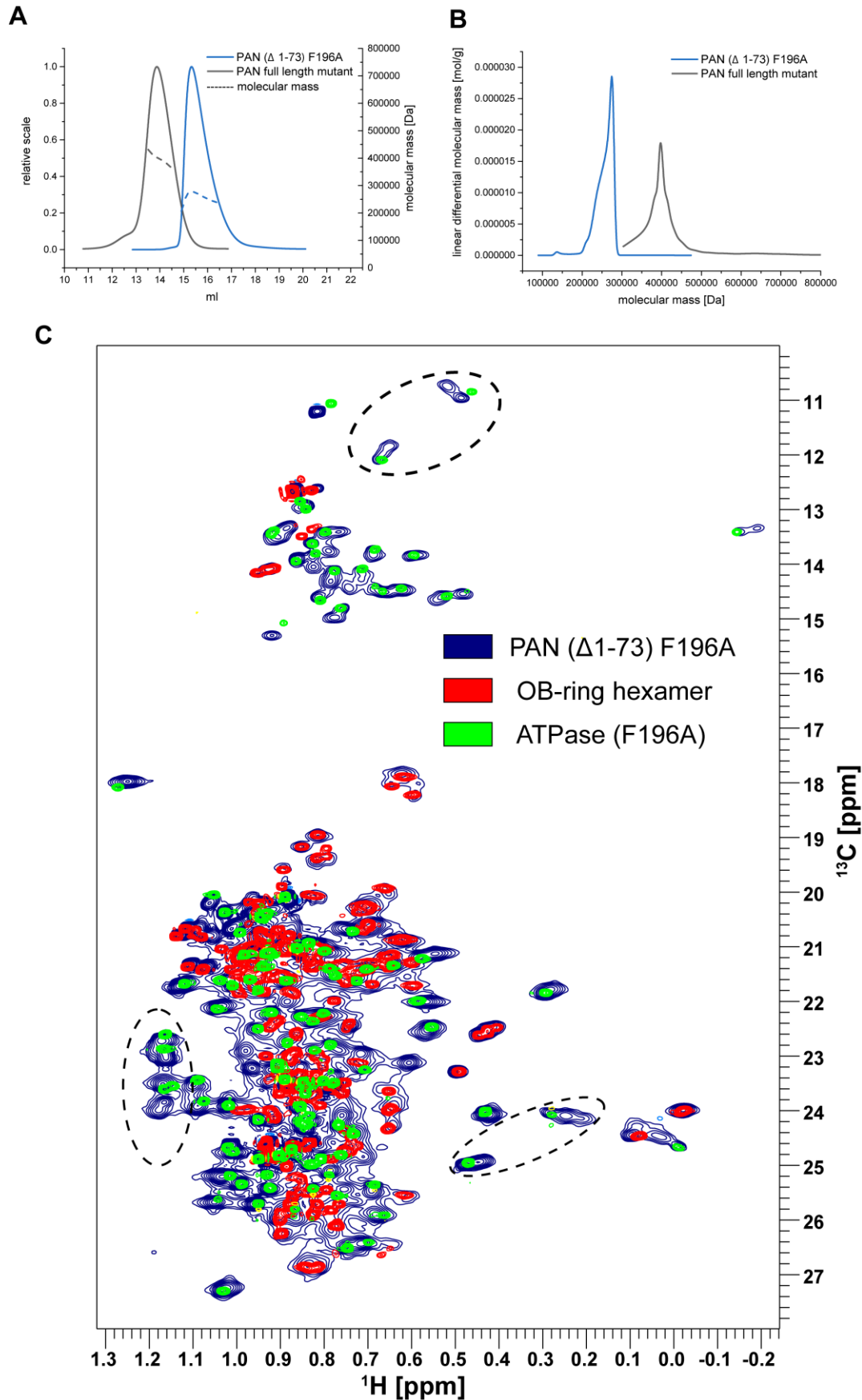


Figure 5.13 SEC-MALS analysis and ^1H , ^{13}C -HMQC spectra of PAN ($\Delta 1-73$) F196A. Truncation of PAN's extended coiled coil (residues 1-73) and introduction of the point mutation F196A renders PAN a monodisperse hexameric complex with a molecular weight of 253.5 ± 0.4 kDa (blue traces) as assessed by SEC-MALS (A & B).

The complex elutes on a Superose 6 10/300 GL increase SEC-column 1.4 ml later than the optimized full length PAN mutant (grey) (5.1.4.1). (C) ^1H - ^{13}C -HMQC spectra of the truncated F196A construct (ILV-labelled) exhibit much improved signal-to-noise, better chemical shift dispersion and much narrower line-width, indicative of a non-exchanging, rather homogenous sample (blue traces). Furthermore, superposition with ILV-methyl HMQC spectra of the isolated OB-ring (red traces) and the ATPase domain (F196A-mutant), respectively, suggests good structural conservation between the full complex and the isolated domain: For every peak of the isolated domains corresponding peak intensities in the assembled complex can be detected. Interestingly, single peaks of the monomeric ATPase domain split into two or three peaks in the context of the full-length protein (encircled regions) indicating that the ATPase domain also adopts at least three different conformations in the full complex, similar to the hexameric OB-ring (section 5.1.2).

5.1.4.3 High concentrations of nucleotides improve spectral quality of PAN substantially

During the second part of this PhD thesis covering studies on the processing of the model substrate GFP-ssrA by the unfoldase PAN (section 5.2) I made an unexpected observation: While titrating unlabelled (thus, NMR-invisible) wild-type PAN onto the IM-labelled (δ 1- and ϵ of Ile and Met, respectively) GFP-ssrA, substantial spectral intensity emerged in regions, not corresponding to the IM-methyl groups of GFP-ssrA (Figure 5.14). Interestingly, peak intensities only appeared when the sample was supplemented with 50 mM ATP, which was included to start the ATP-dependent unfolding of GFP-ssrA. This finding is notable in that PAN, as a particle of several hundreds of kDa should not give rise to significant signals, if not ^{13}C -labelled and perdeuterated at all hydrogen positions but the methyl groups. In any case, the appearance of the new signals was certainly related to the presence of the high concentrations of ATP.

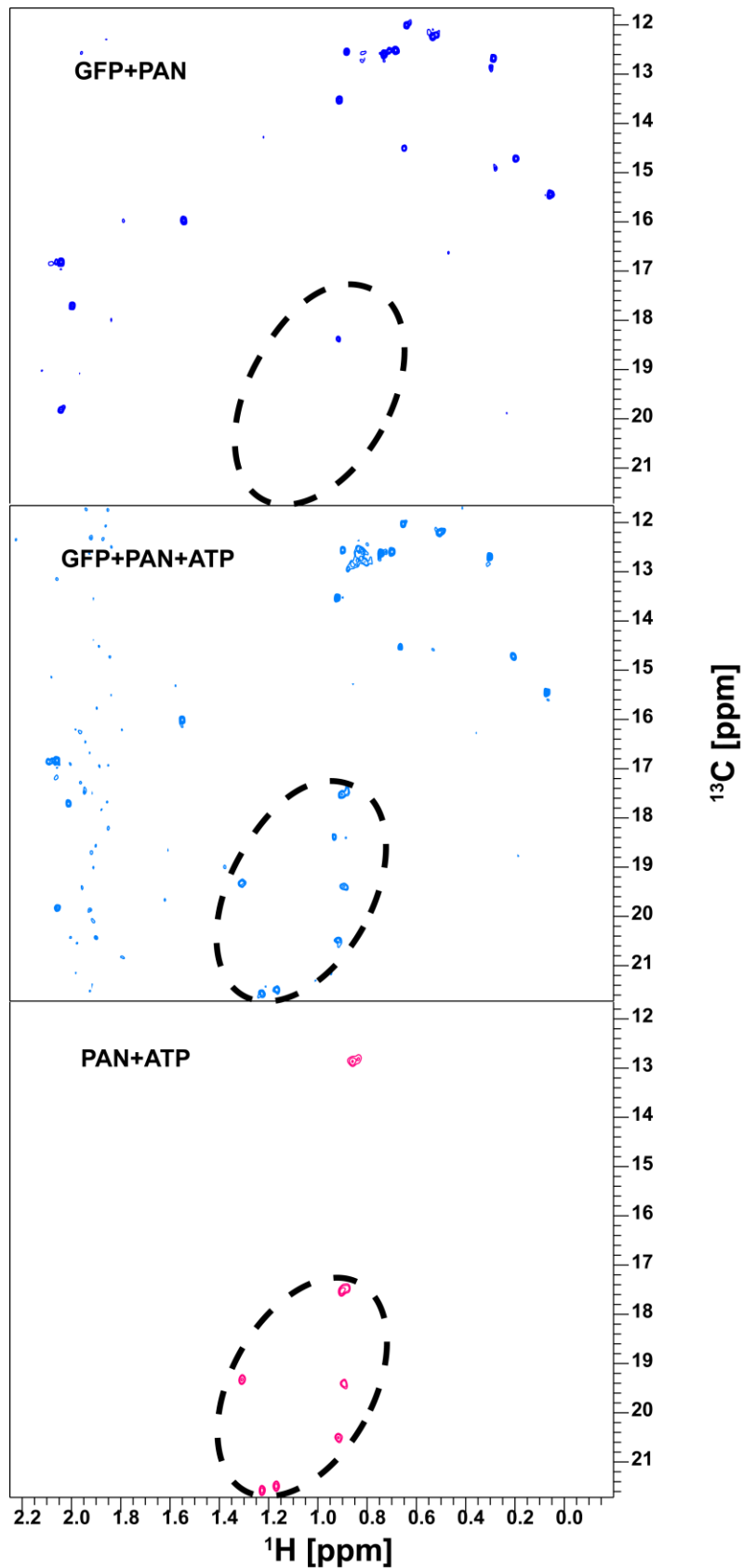


Figure 5.14 Addition of 50 mM ATP makes intensities of unlabelled PAN appear in ^1H , ^{13}C -HMQC spectra. In the absence of nucleotide (top panel) non-labelled PAN does not give rise to peak intensity when titrated onto IM-labelled GFP-ssrA. Upon addition of 50 mM ATP to the sample new peaks appear (middle panel, encircled). These new peaks arise from unlabelled PAN, as confirmed by the control spectrum with only PAN and ATP present (bottom panel).

Consequently, I tested wild type PAN's molecular weight by SEC-MALS using this high ATP concentration both in the injected sample and in the running buffer, to maintain the same molarity of ATP throughout the entire size-exclusion column. Indeed, this experiment confirmed that 50 mM ATP changes the elution profile and molecular weight distribution of PAN towards >95 % monodisperse hexameric PAN with a calculated molecular mass of 315.5 ± 2.3 kDa (Figure 5.15 A); this result is slightly better than that obtained for the PAN mutant at 500 mM NaCl (see section 5.1.4.2). This high ATP concentration allowed us to work with the wild-type protein rather than with the mutant and to maintain moderate concentrations of salt (100 mM NaCl), as desirable to maximize the sensitivity of the NMR-experiment.¹⁷¹ Next, I prepared a 1M-labelled sample of wild-type full length PAN to investigate the effect of this improvement on the quality of $^1\text{H}, ^{13}\text{C}$ -HMQC spectra. As expected for a significantly smaller and more homogenous complex, the quality of $^1\text{H}, ^{13}\text{C}$ -HMQC spectra was improved substantially after addition of 50 mM ATP, as illustrated in Figure 5.15. B/C. Since subsequent titration experiments, described in section 5.1.5, were designed to be done under non-unfolding conditions, we also analysed the effect of ADP on the spectral quality. Similarly to ATP, ADP improved the quality of PAN methyl spectra substantially. Titration of ADP onto PAN established that 50 mM ADP is sufficient to optimize the spectral quality, as 100 mM of ADP did not lead to further changes.

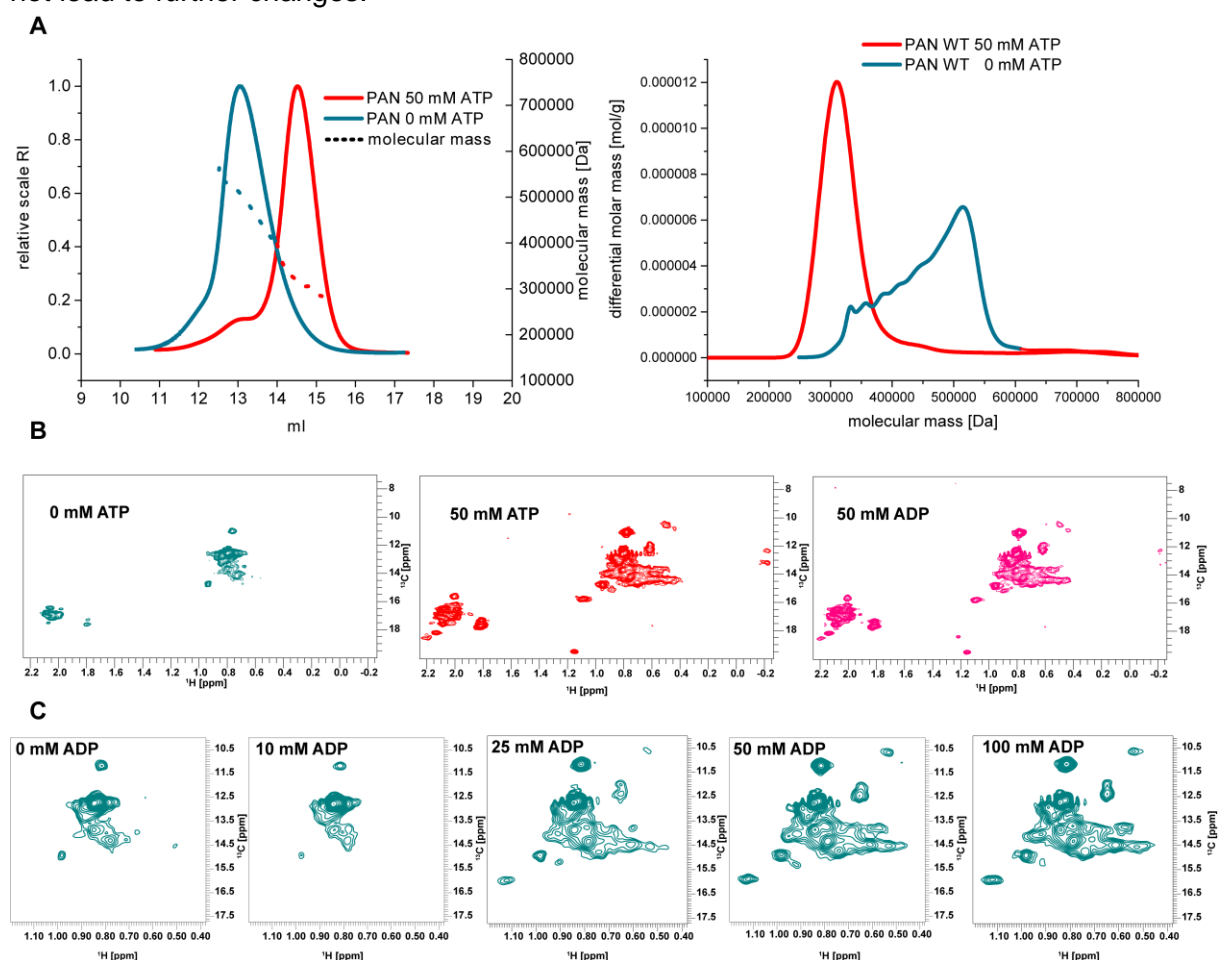


Figure 5.15 50 mM nucleotide improve the spectral quality of wild-type PAN substantially. A) 50 mM ATP in sample and running buffer shift wild-type PAN's oligomeric state towards almost exclusively hexamer as detected by SEC-MALS. B) $^1\text{H}, ^{13}\text{C}$ HMQC spectra of 1M-labelled wild-type PAN were recorded with and without supplementation of 50 mM ATP. Addition of the nucleotide leads to the appearance of many more methyl peaks in both Ile and Met region, resulting in a much more detailed spectrum. Replacement of ATP by ADP had similar effect on spectral quality. C) Excerpt of the Ile-region of spectra acquired during the titration of ADP onto PAN to establish that 50 mM ADP is sufficient to improve the spectral quality. Titrating 100 mM ADP did not lead to further changes

in the spectrum. These improvements in spectral quality are consistent with the substantial decrease in molecular weight as illustrated in A). PAN concentration was 120 μM for all experiments and all NMR spectra were recorded at 60°C and 850 MHz.

5.1.4.4 Assignment of methyl resonances in an asymmetric full-length PAN

Structural studies of large molecular assemblies by NMR strictly depend on the assignment of methyl resonances in the context of the full complex. As described in sections 5.1.4.2 and 5.1.4.3 spectral quality of full-length PAN was limited even under optimized conditions and especially in the Leu/Val region methyl peaks were severely overlapped (also see Figure 5.16. B). To facilitate the transfer of methyl-resonance assignments from the individual domains (sections 5.1.2 and 5.1.3), I consequently focused on the ϵ -methyl group of methionines and the δ 1-methyl group of isoleucines. Superimposing ^1H , ^{13}C HMQC methyl spectra of the individual ATPase and OB-ring domains with a spectrum of the full-length wild-type PAN demonstrates the general conservation of structural features of the individual domains in the context of the full complex (Figure 5.16), since many isolated and well dispersed peaks are positioned similarly both in the isolated domains and the full complex. Furthermore, Figure 5.16 B illustrates prominently the spectral complexity of PAN's full-length spectrum with a multitude of Leu and Val peaks buried in unresolved regions, rendering resonance assignment thereof challenging. Interestingly, I also observed that several of the well dispersed peaks of the ATPase do appear as triplets in the full complex, which equally resembles the conformational heterogeneity which I already discovered in the isolated OB-ring (5.1.2). These findings suggest that PAN is present as an asymmetric assembly with at least three different protomer conformations, which is consistent with findings of multiple conformations of PAN in cryo-EM studies.⁴¹

Including several residues with multiple conformations a total of 40 Met- and Ile-cross peaks from 25 different residues were assigned in full length PAN (Figure 5.17 A and B): Ile and Met-methyl-resonance assignments of the ATPase domain were transferred by superimposing the assigned spectrum (section 5.1.3, Figure 5.7 B) of the isolated construct with the full length version (Figure 5.16). 8/9 Met- and 10/ 24 Ile-resonances could be assigned unambiguously. For the OB-ring hexamer, only three different Ile-resonances of Ile93 could be transferred from the individual domain unambiguously. The remaining Ile79 is likely to be buried in the severely overlapped Ile-region of the spectrum. Met-resonances of the OB-ring and likewise Ile- and Met-resonances of the coiled coil domain were assigned by point mutation of the respective residues to alanine and subsequent comparison of wild type and mutant spectra (Figure 5.17. C-G).

Afterwards, I moved to the characterization of the interaction of PAN with the model substrate GFP-ssrA (following chapter 5.1.5).

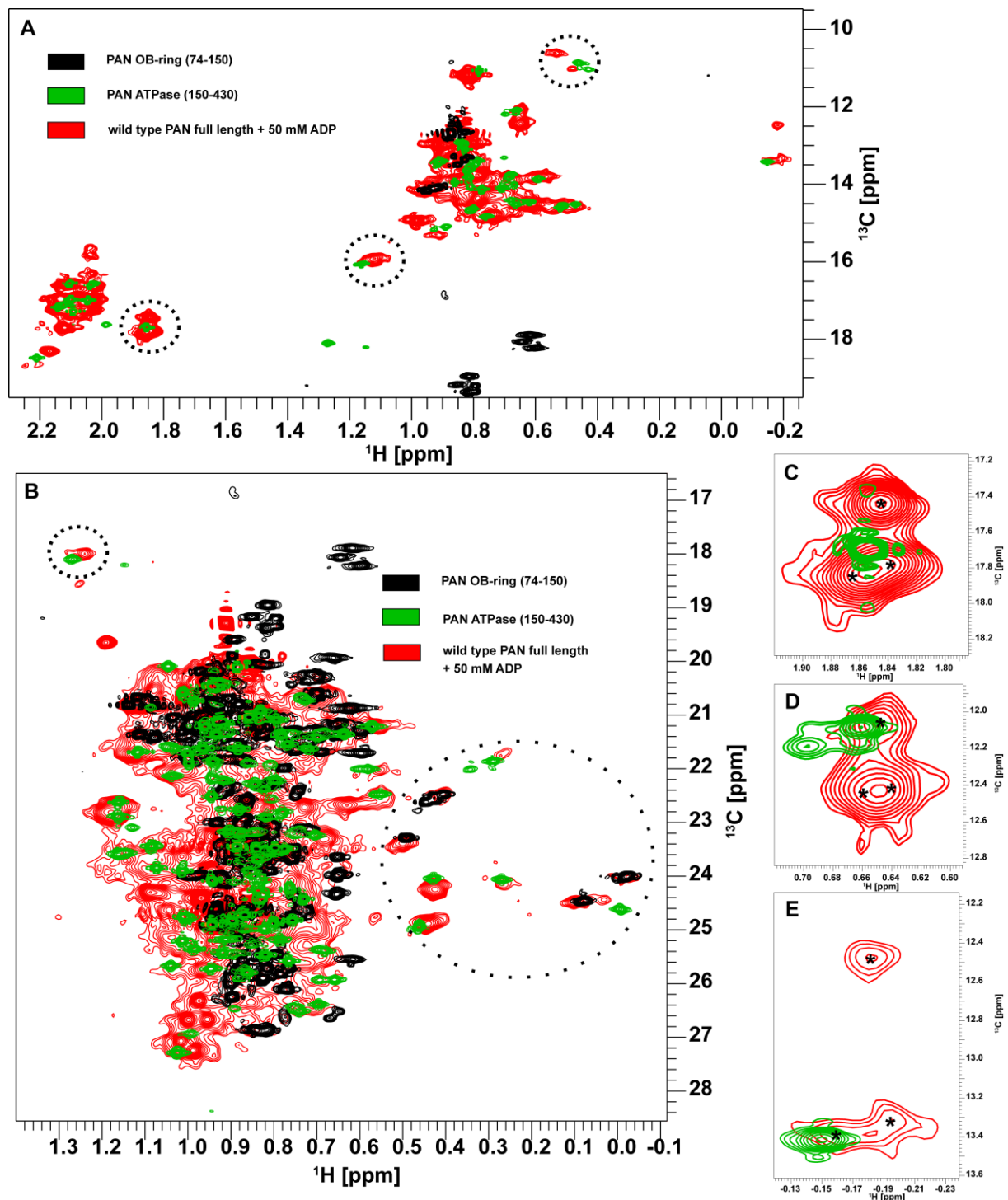


Figure 5.16 Overlay of ^1H - ^{13}C -HMQC spectra of PAN's individual domains and the full-length complex. A) Excerpt of the Ile- and Met-region of an overlay of the OB-ring hexamer (ILV-labelled, black), monomeric ATPase domain (ILVM-labelled, F196A mutant, green) and full length wildtype PAN (IM-labelled, supplemented with 50 mM ADP, red). B) Excerpt of the Leu-/Val-region of the same spectra as in A) except for red being a full length version of PAN containing a point mutation I7A used to assign this residue. Encircled non-overlapped regions depict, that structural features of the individual domains are conserved in the context of the full length complex. Note also, that most of the Leu-/Val-peaks are severely overlapped and thus were not used in subsequent experiments (5.1.5). (C), D), and E): Several isolated peaks split into up to three different peaks (indicated by asterisks) when transferred from monomeric ATPase domains (green) to the full length hexameric PAN (red) suggesting an asymmetric assembly of PAN. Shown are excerpts of Met- (C) and Ile-regions (D/E). All presented spectra were recorded at 60 °C and 850 MHz.

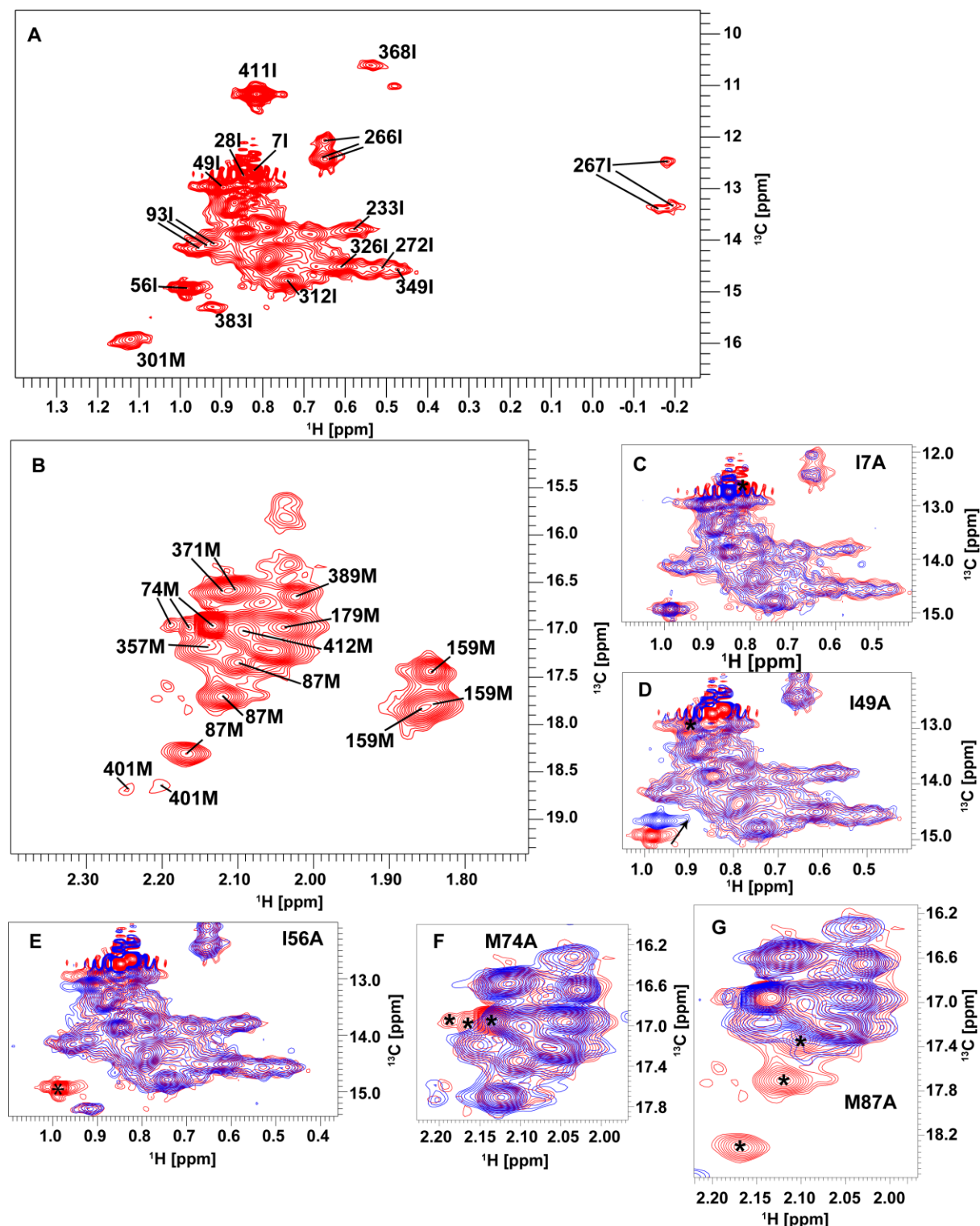


Figure 5.17 Ile- and Met- methyl resonance assignment of full-length PAN. A total of 40 Ile (A) and Met (B) cross-peaks were assigned including several residues which exist in multiple conformations in the hexameric full length complex. (C-G) Met-resonances of the OB-ring as well as Ile- and Met-resonances of PAN's extended coiled were assigned by point mutations to alanine and comparison of mutant (blue) and wild type (red) spectra. Asterisks indicate mutated residues which disappear in mutant spectra. Note that for I49A (D) not only the original peak disappears but also I56 (assigned in (E)) shifts significantly (indicated by arrow). All spectra were recorded at 60 °C and 850 MHz.

5.1.5 Characterization of PAN-substrate interactions by NMR-spectroscopy

The first step in a functional cycle of protein processing by the PAN-20S proteasome is the interaction of the substrate protein's degradation tag or disordered region with PAN. Studying this functional interaction in an intact and active unfoldase by solution NMR is a challenging endeavour due to several reasons: (1) in the presence of ATP the substrate is unfolded during the time necessary to record one 2D experiment. (2) It has been reported, that ADP – in contrast to non-hydrolysable analogues – is not sufficient for PAN to efficiently engage with its substrates.⁴⁸ (3) Non hydrolysable analogues are also hydrolysed by PAN under NMR-conditions. Here, we chose to use ADP seemed for the following NMR-studies, since it supports good spectral quality of PAN and the quantities necessary for an NMR-sample at 50 mM concentration are inexpensive compared to slowly hydrolysable analogues like ATP γ S or AMP-PNP.

First, I characterized the binding of the model substrate GFP-ssrA to PAN in the presence of 50 mM ADP. To this end, I titrated unlabelled GFP-ssrA onto IM-labelled PAN. Unexpectedly, no chemical shift perturbations (CSP) were observed. The spectra appeared identical to those of PAN in the absence of GFP-ssrA, except for intensity of natural abundant GFP-ssrA, which was added in high excess (Figure 5.18).

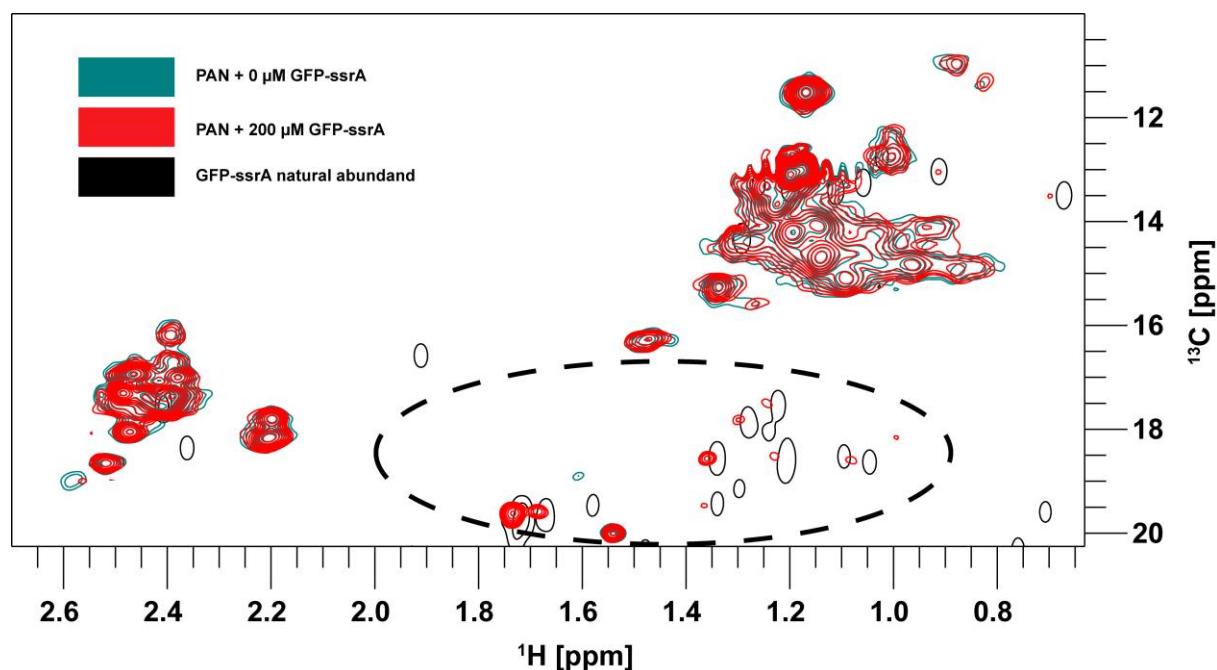


Figure 5.18 GFP-ssrA interaction with PAN. GFP-ssrA was titrated onto IM-labelled PAN and ^1H , ^{13}C -HMQC spectra were recorded of PAN alone (cyan) and PAN with 10-fold excess of GFP-ssrA (red). No significant differences were observed except for the appearance of peaks of unlabelled GFP-ssrA (the black contours are overlaid here from a spectrum of ^{13}C -labelled GFP-ssrA alone, as a control). The encircled regions indicates the region of Ala/Thr/Val-methyl groups of GFP-ssrA: of note are the intense peaks around 1.7 ppm (^1H), which arise from several alanines of the GFP disordered C-terminal ssrA tail. PAN was used at 20 μM hexamer (120 μM monomer) and GFP-ssrA at 200 μM , respectively. All samples were supplemented with 50 mM ADP and spectra were recorded at 60 $^\circ\text{C}$ and 850 MHz.

Previous studies have shown that substrates harbouring exposed disordered regions, like the ssrA-tag are processed by PAN.^{34,94} Though, it is assumed that PAN does not form a stable complex with its substrates and that substrate binding is transient and dependent on ATP-binding.⁴⁸

Despite the fact that it was postulated that ADP binding is not sufficient for PAN to bind the substrate, I aimed at analysing potential first encounters of PAN with the model substrate GFP-ssrA by NMR. The NMR-technique of Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion is a unique tool to investigate processes like these weak and transient unfoldase-substrate interactions since it can probe exchange processes with lifetimes ranging from ~ 50 μs to ~ 10 ms and populations of the excited or intermediate states (here substrate bound to or interacting with PAN) as low as 0.5–1%.¹¹⁴ Thus, we recorded two sets of CPMG experiments to characterize chemical exchange processes in PAN induced by substrate interaction: One reference experiment of PAN, IM-labelled and perdeuterated in the presence of 50 mM ADP and one set of IM-labelled PAN supplemented with 50 mM ADP and 200 μM GFP-ssrA. In both cases the PAN concentration was 20 μM hexamer (120 μM monomer). After extraction of the effective relaxation rates R_2^{eff} for each peak at every CPMG pulsing rate, we calculated an approximate exchange rate $R_2^{\text{ex, approx}}$ as the difference between the R_2^{eff} at the minimum and at the maximum CPMG field strength: these values are displayed in Figure 5.19 A. $R_2^{\text{ex, approx}}$ can be roughly correlated to the extent of conformational exchange processes at each methyl group site; slow local dynamics induced by PAN-substrate interactions (i.e. conformational exchange between the bound and unbound state) would result in changes of this parameter. However, several aspects of the extracted data are apparent: (1) The exchange rates exhibit high errors; (2) data could not be obtained for each methyl group in both conditions; (3) several exchange rates are of negative sign. Only spin systems 23 and 60 display a significant difference of exchange rates between the two sets of experiments, potentially indicative of substrate induced changes. However, careful examination of the respective relaxation dispersion curves rather suggests that this apparent difference is an artefact of extracting data from these poor-quality-relaxation dispersion curves, which do not show the expected profile. Whether a result of insufficient data quality or a reflection of the incompetence of PAN to engage the substrate in the absence of ATP or an ATP-analogue, we were not able to detect any significant changes between bound and unbound PAN by CPMG relaxation dispersion experiments. Thus, the given experimental set up and sample quality cannot result in a reliable conclusion as to whether PAN interacts with GFP-ssrA also in the absence of ATP- or ATP-analogues.

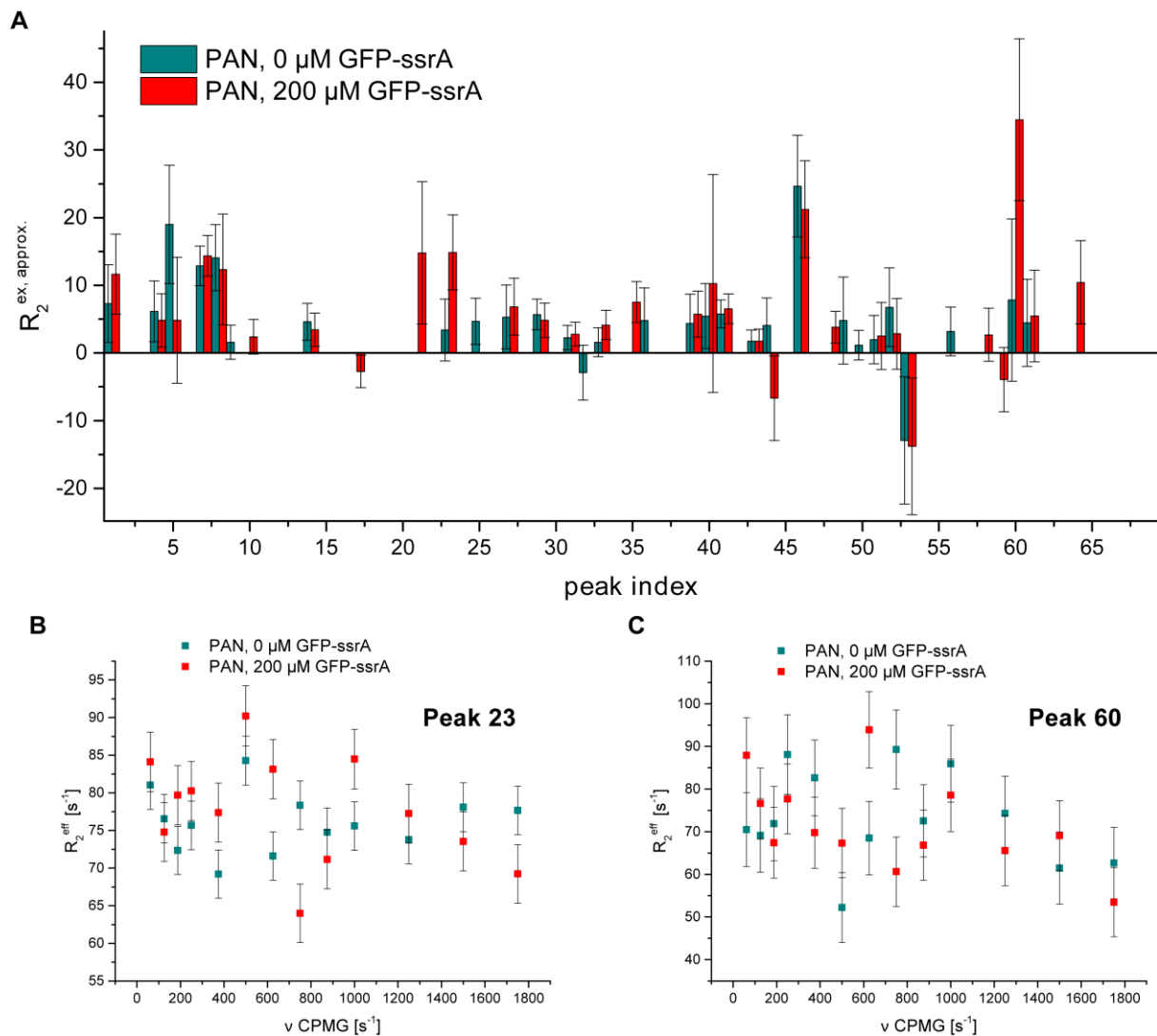


Figure 5.19 PAN-substrate interactions studied by CPMG-relaxation dispersion. (A) Approximate exchange rates $R_2^{\text{ex, approx.}}$ for PAN alone (cyan) and PAN supplemented with excess of GFP-ssrA (red) do not indicate significant differences in dynamics between substrate-bound and unbound. Peak 23 and 60 are the only exceptions, for which the extracted $R_2^{\text{ex, approx.}}$ differ significantly considering the respective error bars. However, relaxation dispersion profiles of these selected peaks (C & D, respectively) do not show the typical profile, suggesting insufficient data quality. As a reflection of poor data quality and low signal-to-noise, binding of GFP-ssrA to PAN under the given conditions could not be assessed.

5.1.5.1 Interaction of the substrate with PAN N-terminal domains

Studying PAN-substrate interactions in the context of full-length PAN constructs proved challenging as described in the previous section: The limited spectral quality occluded potential interactions. It is well established, that proteasomal substrates enter the unfoldase via its N-terminal pore and make substantial contacts with pore-1-loops within the ATPase subdomain.⁶⁰ In eukaryotes, proteasomes have a sophisticated ubiquitin-processing machinery, which recognises proteasomal substrates and directs them for unfolding and degradation; however, such a system is unknown in archaea. Instead, it has been proposed that PAN's N-terminal domains, comprising the unstructured tail, the coiled coil domain and the OB-ring are involved in tethering substrates to PAN by interacting with unstructured regions on substrates.³² To investigate this early encounter of substrate with PAN, I did NMR-titrations using

a synthetic *ssrA*-peptide as a substrate mimic (NH_3^+ -AANDENYALAA-COOH). To this end I recorded ^{15}N -HSQC and ^{15}N -TROSY spectra for PAN-CC and PAN-OB, respectively, both in their apo form and supplemented with excess of the synthetic *ssrA* peptide. After adding *ssrA*-peptide in ten-fold excess (300 μM PAN-CC monomer, 3 mM *ssrA*-peptide) to the PAN coiled-coil domain, no chemical shift changes could be observed: the spectra appeared identical, which is indicative of no binding (Figure 5.20 A). Titration of the *ssrA*-peptide in six-fold excess to the OB-ring-hexamers (300 μM OB-ring, 1.9 mM *ssrA*-peptide) led to a similar result: the spectra looked identical, with only minor chemical shift changes <0.01 ppm deriving from non-specific effects (Figure 5.20 B).^{172,173} Considering the large excess of *ssrA*-peptide in both cases, shift changes should have been observed if (specific) binding of the peptide to the respective proteins had been present. Consequently, the *ssrA*-tag does not seem to interact with either PAN-CC or the PAN-OB-ring in their isolated form in any specific manner.

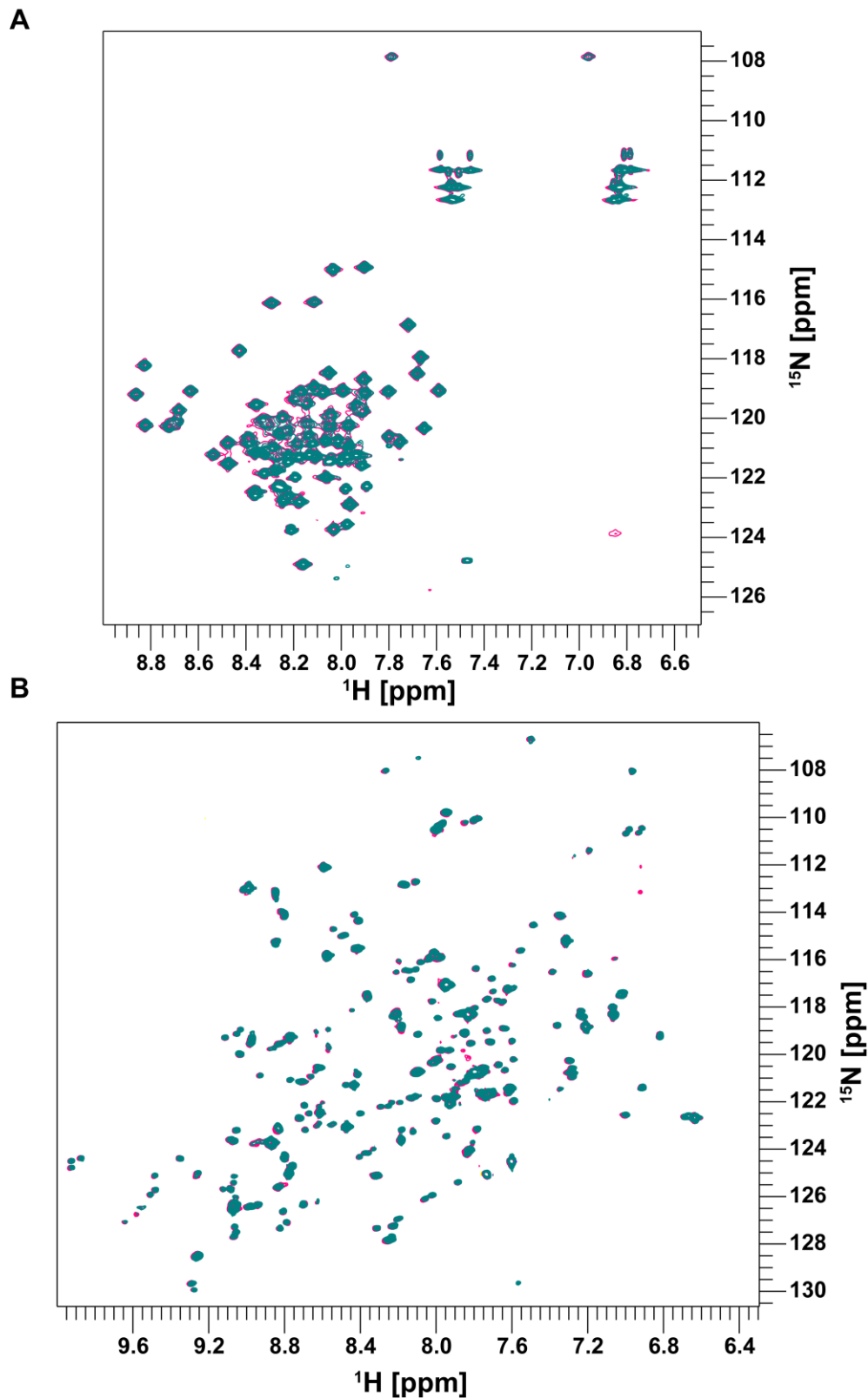


Figure 5.20 PAN N-terminal domains do not interact specifically with a ssrA-peptide. Synthetic ssrA-peptide as a substrate mimic was titrated onto the PAN-CC- (A) and the PAN OB-ring-domains (B), respectively. A) ^{15}N -HSQC spectra of the PAN-CC domain in the apo form (300 μM monomer, 150 μM dimer, pink) and the PAN-CC domain with 3 mM ssrA-peptide (cyan) perfectly overlay, indicating no specific interaction of the PAN-CC domain with the peptide. Likewise in B) only subtle changes are visible in an overlay of apo (pink) and ssrA-supplemented (cyan) ^{15}N -TROSY spectra of the PAN OB-ring (OB-ring was at 60 μM hexamer or 300 μM monomer, while ssrA-peptide was added at 1.9 mM). Spectra of PAN-CC and the PAN OB-ring domains were recorded at 45 $^{\circ}\text{C}$ and 60 $^{\circ}\text{C}$, respectively, at 850 MHz.

5.2 Real-time NMR studies on the processing of a model substrate by the archaeal PAN-20S proteasome

Studying the time-course of an unfolding- or degradation reaction by the proteasome can assist in drawing conclusions about the mode of action of this molecular machine. To this point kinetic studies on the mechanism of action of AAA+ proteases mostly relied on monitoring fluorescent protein-substrates by fluorescence spectroscopy,^{49,64,174} nucleotide hydrolysis kinetics,¹⁷⁵ single molecule force spectroscopy with optical tweezers^{100,101,103,105} and lately also time-resolved SANS (small angle neutron scattering).^{1,40} However, common to all these techniques is their inherent lack of atomic resolution. Solution NMR spectroscopy, is capable of simultaneously probing many different sites in protein substrates at atomic resolution and has been used in static⁹⁹ and also dynamic time-resolved fashion¹⁷⁶ to investigate the unfolding of a protein substrate by a proteasomal unfoldase and the hydrolysis of a substrate protein by the 20S proteasome, respectively. Here, I aimed to combine the atomic resolution of two dimensional NMR with time-resolution of repeated measurements of 2D-SOFAST-HMQC spectra¹³⁷ to obtain residue specific and real-time information of the proteasomal processing of the model substrate GFP-ssrA. The following chapter covers the preparation and testing of suitable substrate proteins and the establishment of experimental conditions for an NMR-based real-time assay which resulted in a thorough time-resolved and residue specific analysis of the PAN-20S proteasome processing of a fluorescent protein substrate.

5.2.1 Initial studies on the processing of GFP-ssrA substrates by the PAN-20S proteasome system

The green fluorescent protein C-terminally modified with an ssrA-degron (GFP-ssrA) has been widely used as a model substrate for PAN and other AAA+ unfoldases due to the convenient spectroscopic detection of its degradation by fluorescence decay and its superb stability in solution in a wide range of conditions.¹⁷⁷ Initial studies on the unfolding and degradation by the PAN-20S proteasome were carried out on a GFPuv-variant containing a C-terminal ssrA-tag,^{1,40} which however, turned out to be not suitable for NMR-studies on isotopically enriched samples and was replaced by an optimized variant for final NMR-time course experiments (section 5.2.2).

Generally, GFP is targeted by the ssrA-degron to unfolding and hydrolysis by the PAN-20S proteasome in an ATP-dependent manner which can be detected by monitoring the loss of GFP-fluorescence at 509 nm and by SDS-PAGE of reaction mixes, respectively (Figure 5.21 A & B). However, due to its very compact and highly stable β -barrel structure GFP-ssrA tends to refold after successful unfolding attempts¹⁷⁸ regaining some of its native fluorescence, especially when the 20S CP is not included in the reaction set-up. Thus, fluorescence based assays need to be fine-tuned so that refolding is maximally avoided and a sufficient overall conversion of the substrate is achieved. Indeed, I found that the stoichiometry of PAN over the substrate is substantially critical, since strong reappearance of fluorescence was observed when PAN was used in equimolar or excess quantities (Figure 5.21). Since the

final goal was to monitor the unfolding in real-time inside an NMR tube assay conditions were chosen to meet the following requirements:

- $\geq 100 \mu\text{M}$ of GFP-ssrA to record 2D NMR spectra with a time-resolution and sufficient signal-to-noise (signal intensities decay during the unfolding/degradation reaction, thus a concentration of $100 \mu\text{M}$ is effective only at the start of the reaction).
- Duration of the initial fast decay $\gg 10$ min to record sufficient data points in this critical phase
- The build-up of potentially inhomogeneous aggregates of unfolded GFP-ssrA should be slow to avoid problems from inhomogeneities in the NMR sample, especially within the first 10-20 minutes of the time-course.

Consequently, the reaction mix needs to be scaled up accordingly to the required high GFP-ssrA concentration. For initial NMR-studies I found that $5 \mu\text{M}$ PAN, 50 mM ATP, 50 mM MgCl_2 and incubation at 60°C were suitable conditions for a long-lasting reaction without severe precipitates forming after 30 minutes.

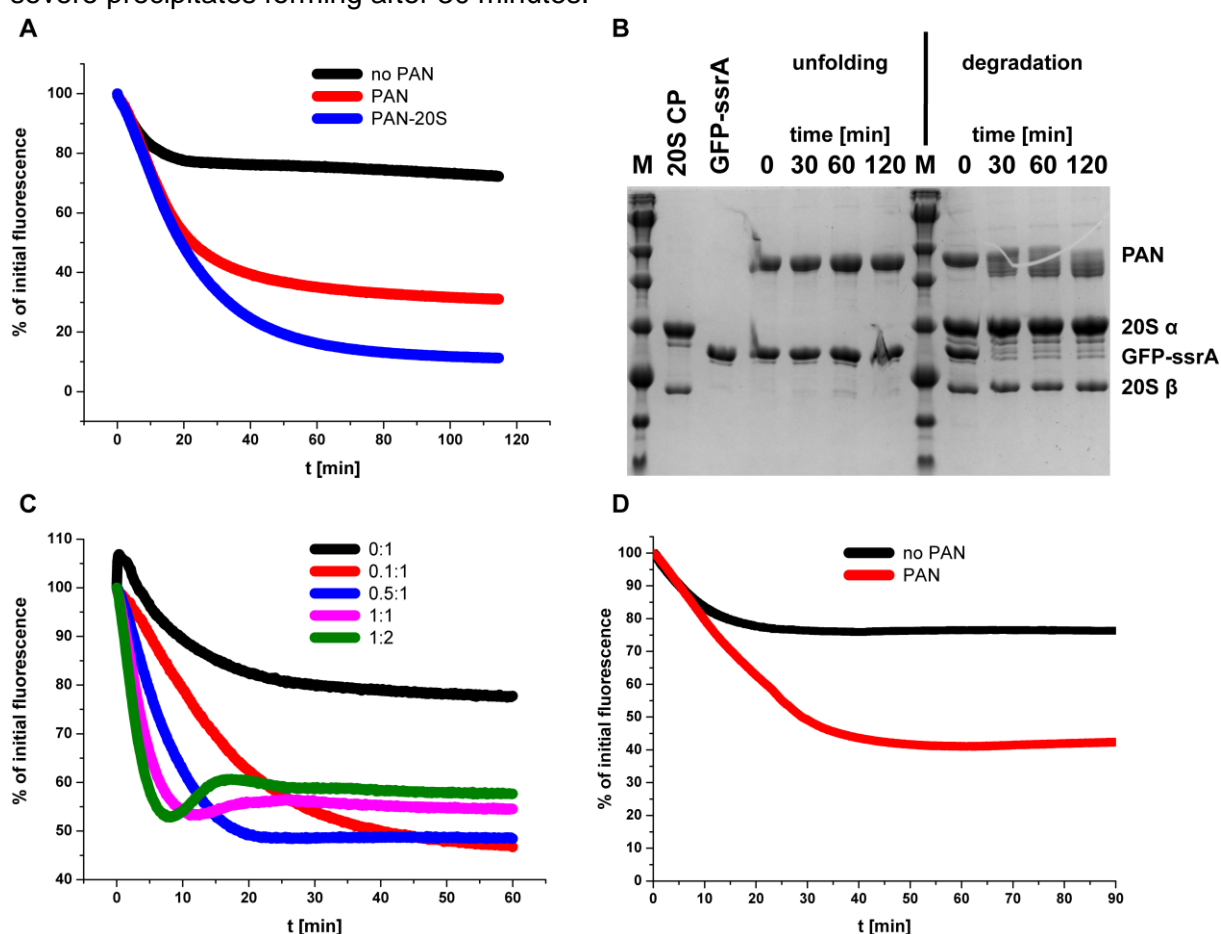


Figure 5.21 Initial experiments on the unfolding and degradation of GFPuv-ssrA by the proteasome and optimization for NMR-experiments (A) GFPuv-ssrA unfolding and proteolysis by the PAN-20S proteasome can be detected by monitoring the decay of native GFP-fluorescence at 509 nm (here, $1 \mu\text{M}$ GFPuv-ssrA, $0.5 \mu\text{M}$ PAN, $0.5 \mu\text{M}$ 20S CP, 10 mM ATP and 10 mM MgCl_2 at 60°C). Addition of the 20S CP leads to a substantially lower final fluorescence indicative of proteolysis of unfolded GFPuv-ssrA. Note that fluorescence in controls without PAN also drops by $\approx 20\%$, which is a well-known phenomenon induced by the elevated temperature of the assay^{1,49,94}. (B) Proteolysis of GFP can be further confirmed by running reaction mix on SDS-PAGE gels, which display the disappearance of most of the GFPuv-ssrA band when the 20S CP is included in the reaction mix. (C) Using excess of substrate is critical to obtain maximal conversion of folded to unfolded non-fluorescent GFPuv-ssrA. In the figure PAN:substrate molar ratios are indicated (concentrations of GFPuv-ssrA, PAN, ATP and MgCl_2 were: $1 \mu\text{M}$, $0.1\text{--}2 \mu\text{M}$, 10 mM and 10 mM , respectively). PAN (D) Optimized and scaled up assay for initial time-resolved experiments: $100 \mu\text{M}$ GFPuv-ssrA, $5 \mu\text{M}$ PAN, 50 mM ATP, 50 mM MgCl_2 at 60°C . All fluorescence kinetics presented here were carried out in a 96 well plate reader with reaction volumes of $100 \mu\text{l}$.

5.2.2 GFPuv-ssrA is not a suitable fluorescent protein for NMR-studies on isotopically enriched samples

After selection of optimized conditions I set out to test the feasibility of time-resolved NMR measurements on the unfolding of GFPuv-ssrA by PAN. To this end I prepared a perdeuterated sample of GFPuv-ssrA with only the isoleucine $\delta 1$ and methionine ϵ -methyl groups ^1H - ^{13}C -labelled (IM-labelling scheme, hereafter referred to as dGFP), a labelling scheme which should result in relaxation properties and thus sharp and intense peaks in every spectrum of a time series, even at 100 μM protein and short acquisition time per spectrum. In addition, the reduced peak count simplifies data analysis. Components of the reaction mix were mixed before loading the sample into the NMR and the reaction was started by adding ATP at the final moment and inserting the tube into the preheated spectrometer probe (60 $^\circ\text{C}$). The time-dependent changes of dGFP were monitored by recording a series of ^1H , ^{13}C -SOFAST-HMQC spectra which allow measurements with enhanced sensitivity at high repetition rates.¹³⁷ During these initial experiments I made several unexpected observations:

1. Several isoleucines appeared as double peaks with unequal intensities, potentially representing a major and a minor conformation (Figure 5.22 B).
2. Intensities of most peaks showed significant decay even in control experiments with no PAN present or no ATP included in the reaction mix (Figure 5.22 A).
3. The less intense peaks of each peak pair showed a decay, while the most intense peak showed near-to-constant intensities throughout both unfolding runs and controls.
4. Formation of precipitates/aggregates in control runs was also observed by visual inspection of the sample after the time-series, suggesting instabilities of the dGFP sample.

After assignment of GFPuv-ssrA methyl-resonances (5.2.5) I discovered that all peaks showing two different conformations were clustered next to the GFP chromophore (Figure 5.22 C) suggesting conformational heterogeneity in the chromophore region, which might lead to instability of the GFP core at the elevated temperatures of the unfolding assay. The maturation of the GFP chromophore is known to follow specific kinetics at different temperatures and conditions of protein biosynthesis.¹⁷⁹⁻¹⁸¹ Since the perdeuterated sample was prepared from *E. coli* cultures grown in defined medium based on 99% D_2O and a deuterated carbon source, the conditions of the growth were different from preparations in LB medium for protonated GFP (hGFP) samples used in previous “offline”-test (section 5.2.1). Indeed, careful comparison of UV-Vis traces of hGFP and dGFP suggested substantially less native GFP chromophore absorbance at 396 nm for dGFP samples, as determined by the ratio between 280 nm and 396 nm (Figure 5.22 D). Following the hypothesis that the fluorophore of dGFP was not fully matured, potentially by expression at low temperature in M9 minimal medium, I incubated the deuterated GFP sample at a mildly warmer temperature (42 $^\circ\text{C}$) and monitored the changes both by NMR and by UV/Vis-spectroscopy. As expected, the intensities of the peaks representative of the minor conformation decreased, whereas those of the peaks of the major conformation increased (Figure 5.22 E). Consequently, the relative absorbance of GFP at 396 nm increased consistent with the maturation of the GFP chromophore (Figure 5.22 F). This conversion is also consistent with the findings that peaks of conformation A did

not decay equally in time-resolved NMR measurements since their intensity was constantly replenished by the converted conformation B.

Although the chromophore maturation was not complete after 24 hours these findings motivated me to subject dGFPuv-ssrA to elongated heat treatment to fully convert the chromophore into the mature form. Although long-term stability of hGFP-ssrA at 42 °C was previously confirmed by ESI-MS to ensure intactness, the C-terminal disordered ssrA-tag was proteolysed after the elongated heat treatment of dGFPuv-ssrA, as assessed by ESI-MS and undetectable unfolding in activity assays. These drawbacks rendered this approach useless for unfolding and degradation studies on dGFP. Thus, I decided to test the feasibility of time series studies on ^{13}C -labelled, non-deuterated GFPuv-ssrA, despite the potential compromises in resolution and sensitivity, due to the less favourable relaxation behaviour. Unfortunately, even in hGFPuv-ssrA batches expressed in protonated M9 medium with ^{13}C -labelling, the chromophore did not form completely, calling for an alternative strategy.

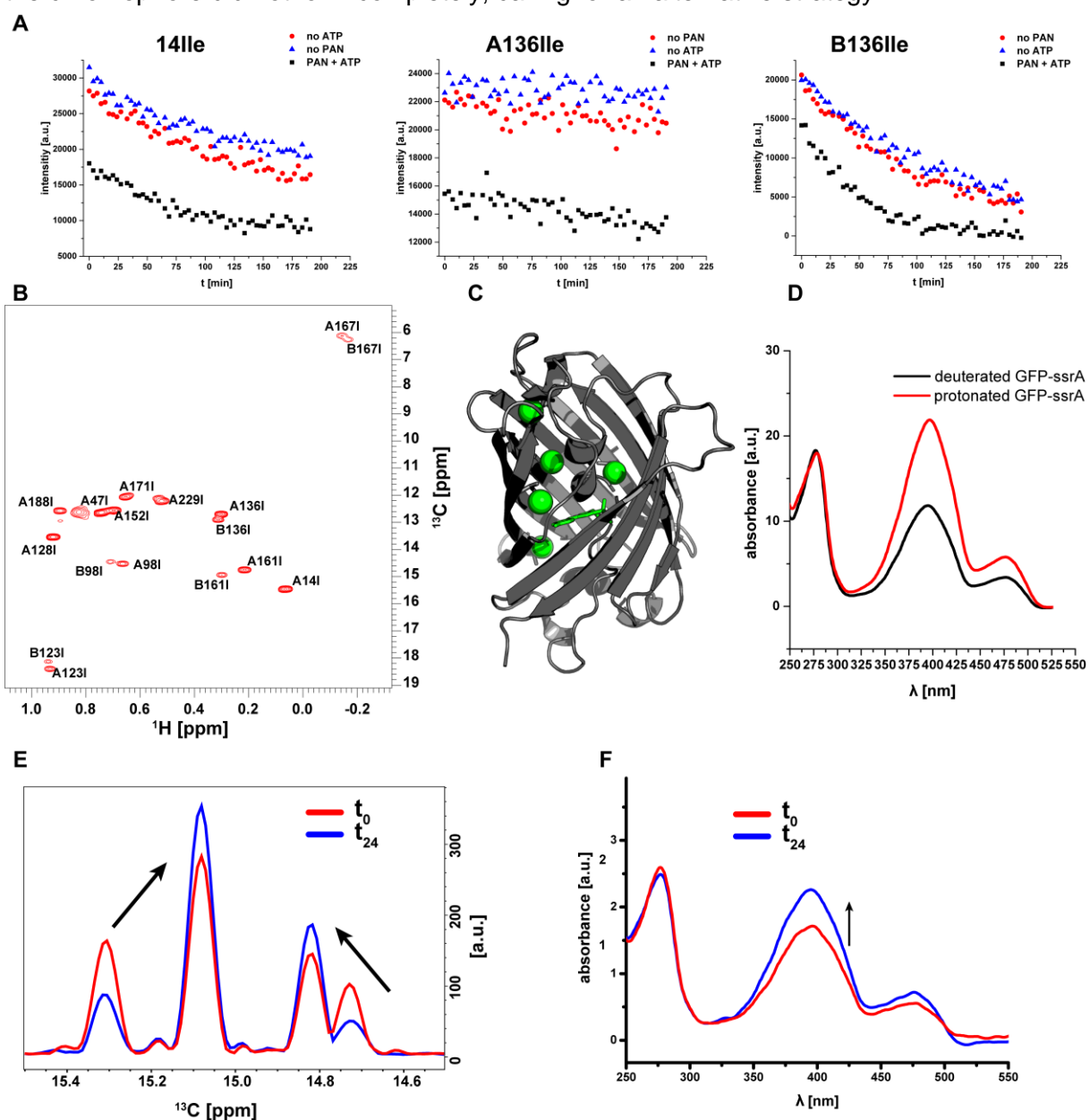


Figure 5.22 The chromophore of GFPuv-ssrA does not fully mature in deuterated protein preparations (A) Intensity decays of selected Ile δ 1 methyl peaks extracted from series of ^1H , ^{13}C -SOFAST HMQC during the unfolding of dGFPuv-ssrA by PAN at 60 °C. Single conformation peaks, like 14Ile, displayed a decay of similar kinetics both for reactions in which PAN was included and controls. Double conformation peaks like 136Ile

showed slow, linear decay for the major conformation A and fast, exponential decay for conformation B. (B) ^1H , ^{13}C -SOFAST HMQC spectrum of dGFPuv-ssrA showing Ile δ 1-methyl groups with a second conformation, which are highlighted as green spheres in the GFP crystal structure (PDB 1b9c) and cluster next to the chromophore (green sticks) (C). (D) Comparison of UV/Vis spectra of hGFPuv-ssrA and dGFPuv-ssrA indicates significantly less GFP chromophore specific absorbance at 396 nm for dGFP. This demonstrates that in dGFP the chromophore has not fully matured. (E) Incubating dGFPuv-ssrA at intermediate temperature (42 °C) interconverts the second minor conformation into the major conformation as illustrated by 1D ^{13}C -projections of ^1H , ^{13}C -HMQCs at the start (red) and after 24 h (blue). During this time also the relative chromophore absorbance of GFP at 396 nm has increased significantly (F).

5.2.3 A folding reporter GFP variant as a substrate for the PAN-20S proteasome

Recombinant expression of GFPuv-ssrA in defined media and at lower temperature repeatedly led to incomplete maturation of the GFP chromophore resulting in unstable and conformationally heterogeneous samples. In the past 25 years substantial protein engineering has generated a multitude of GFP-variants for various applications. Engineering goals were mostly the tuning of spectroscopic properties like wavelength combinations for fluorescence spectroscopy, quantum yields, brightness and also optimized expression and maturation kinetics at various temperatures for experiments in recombinant eukaryotes.¹⁷⁹ Folding reporter GFP (frGFP) is a variant which differs from the above described GFPuv in only two point mutations near the chromophore (F64L and S65T), which improve folding substantially.^{182,183} and thus speed up the maturation by a factor of two.¹⁷⁹ These properties made frGFP a promising candidate for a substrate protein in the proposed time-resolved NMR experiments, since it can be obtained at a low cost by introduction of the two mutations by site-directed mutagenesis into the plasmid already available. Indeed, mutagenesis and recombinant expression of frGFP in minimal medium resulted in a five times higher yield of purified protein with excellent spectroscopic properties. Most importantly the protein was conformationally homogeneous, as concluded from the methyl region of the ^1H , ^{13}C -SOFAST-HMQC spectrum (Figure 5.23), making it a suitable substrate for time-resolved NMR-studies on the unfolding and degradation by the PAN-20S proteasome.

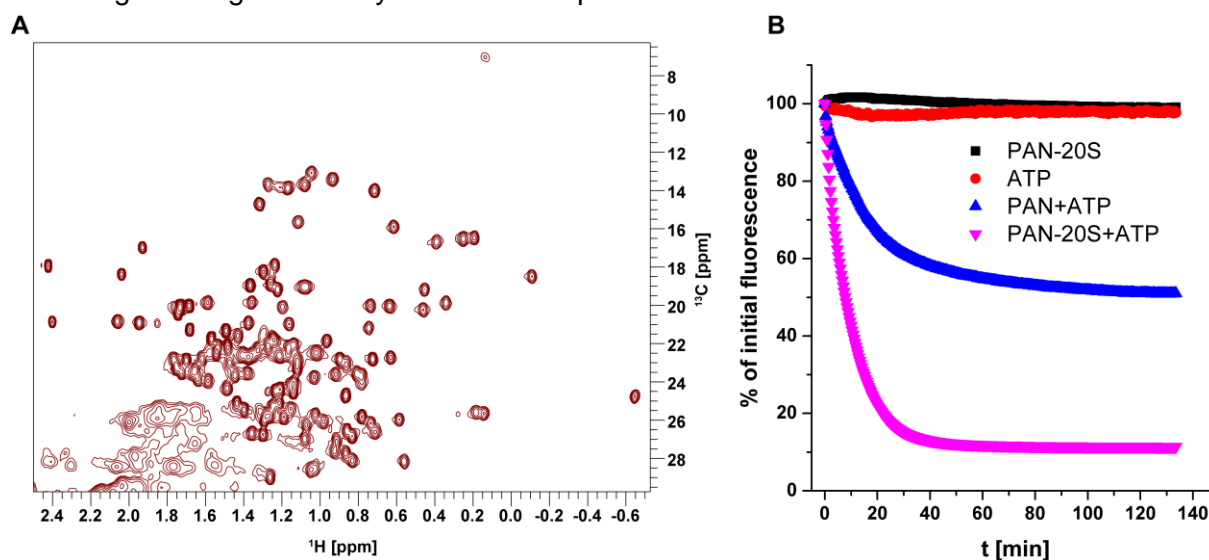


Figure 5.23 frGFP-ssrA is a suitable substrate for time-resolved NMR studies (A) The ^1H , ^{13}C -methyl-SOFAST HMQC of protonated and uniformly ^{13}C -labelled frGFP-ssrA recorded at 60 °C and 850 MHz does not show conformational heterogeneity, which was instead detected for dGFPuv-ssrA (section 5.2.2). Furthermore, it is well accepted as a substrate by the PAN-20S proteasome, as determined by the fluorescence assay (B).

5.2.4 Real-time NMR studies on the unfolding and degradation of GFP by the PAN-20S proteasome system

With a suitable substrate in hand I set out to record NMR time-series of the unfolding and degradation reaction. Initial experiments using the previously selected assay conditions suggested that further optimization was needed to improve the time resolution during the exponential phase of the reaction. The assay conditions were optimized by increasing both the ATP and MgCl₂ concentrations to 100 mM keeping PAN at 5 μM and GFP at 100 μM, respectively (Figure 5.24). To obtain high-quality NMR-data on the processing of frGFP-ssrA by the PAN-20S careful preparation of the individual NMR-samples was crucial: Every component of the sample was thoroughly degassed to avoid formation of small bubbles which could form in the final mix upon transferring the NMR-tube to the preheated (60 °C) spectrometer probe. ATP was added at the last moment before transferring the final mix to the NMR-tube and the reaction was started by heat-activating PAN inside the NMR-spectrometer. A typical dead time of two minutes between inserting the tube and acquisition of the first FID was required to stabilize the temperature inside the probe. For rigorous data analysis a set of triplicate repeats of both unfolding by PAN and degradation by the PAN-20S was recorded at 60 °C. An additional duplicate data set was obtained at 55 °C to potentially improve the data quality by slowing down the reaction of the thermally activated PAN-20S proteasome system.

Parts of the following chapters are adapted from a manuscript KRÜGER, G. et al. 2021 which was originally written by myself.

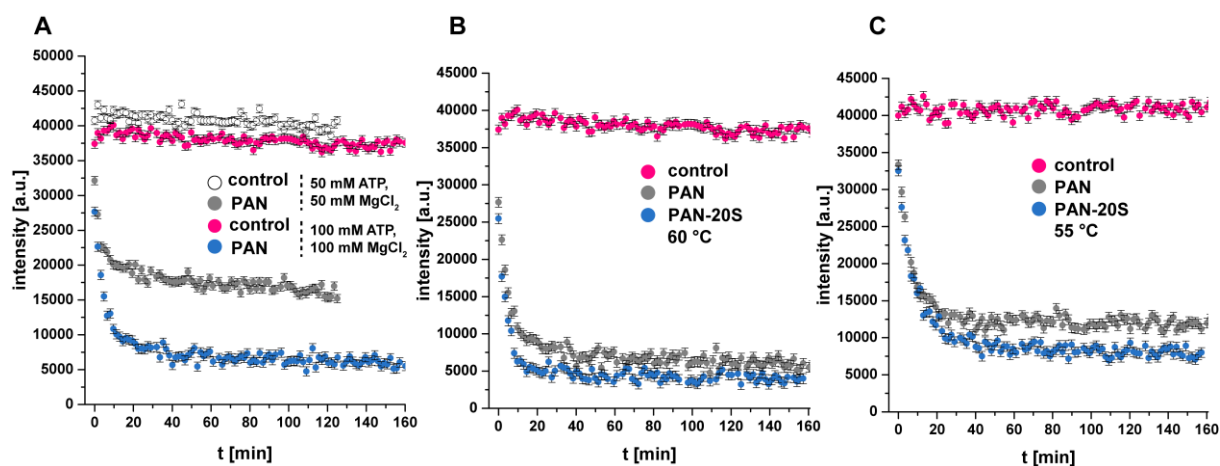


Figure 5.24 Effect of ATP- and MgCl₂ concentration and temperature on NMR-time series (A) Increasing the ATP and MgCl₂ concentration from 50 mM to 100 mM each, results in a much higher conversion of frGFP-ssrA during the unfolding. To improve the time-resolution of the initial exponential phase, NMR-time series of unfolding and degradation have not only been recorded at 60 °C close to the activity maximum of PAN (B) but also at 55 °C (C). Lowering the temperature resulted in a slowing down of both unfolding and 20S-proteolysis but also to a total conversion of frGFP-ssrA smaller than at 60 °C. All curves in this figure represent intensities of the Leu220γ1 peak from time-series of ¹H,¹³C-SOFAST-HMQC spectra recorded at 850 MHz. In all reactions frGFP-ssrA and PAN concentrations were 100 μM and 5 μM, respectively. ATP and MgCl₂ were used at 100 mM each in (B) & (C).

5.2.4.1 Residue-specific analysis of time-resolved NMR-data of the unfolding and degradation reaction

Unlike most other techniques protein NMR-spectroscopy is well-suited to obtain atomic-resolution information on a protein state by monitoring a multitude of resonances simultaneously in one experiment. Combining this spatial resolution with time-resolution by recording 2D-SOFAST HMQC at a high repetition rate represents a powerful tool to gain insights into the kinetics and dynamics of processes like the unfolding of a substrate by the PAN-20S proteasome. Thus, having the assignment of frGFP-ssrA in hand (section 5.2.5), I could monitor the changes of up to 100 resolved methyl groups of frGFP during the unfolding and degradation by the PAN-20S proteasome at a time-resolution of 100 seconds. To this end, I subjected uniformly ^{13}C -labelled frGFP-ssrA to unfolding and degradation by unlabelled PAN and the PAN-20S proteasome, respectively, which allowed me to observe exclusively the resonances of the fluorescent protein during the reaction while the degradation machinery and ATP were essentially invisible in the $^1\text{H}, ^{13}\text{C}$ -SOFAST-HMQC experiments. Time-dependent changes in the series of spectra are analysed in detail in the following sections and include the intensity-decay of frGFP resonances, chemical shift changes experienced by a multitude of methyl-peaks, transient appearance of new resonances and conversion of natively folded frGFP-ssrA to hydrolysed peptide fragments.

GFP unfolds cooperatively

In the absence of PAN or the PAN-20S proteasome, peak intensities of frGFP supplemented with 100 mM ATP and 100 mM MgCl_2 essentially do not decay, consistent with a stable protein at the assay conditions of 60 °C (Figure 5.24). Upon heat-activated and ATP-dependent unfolding by PAN, the intensities of most frGFP-ssrA methyl-group peaks decay rapidly at the start; later the decay slows down and eventually the intensities reach a plateau (Figure 5.24, blue traces). As a first approximation this process can be described as a biexponential decay with two characteristic exponential time constants t_1 and t_2 :

$$y = A_1 e^{(-x/t_1)} + A_2 e^{(-x/t_2)} + y_0 \quad (5-1)$$

However, after extracting peak intensities for all time-points and peaks it was apparent that not all decay curves would need a biexponential model, as the fitting of some of the curves to this model resulted in high errors of the non-linear curve fit parameters (NLC-fit). In order to examine which model describes the different decay curves best I fitted every curve to a biexponential (ExpDec2), a single exponential (ExpDec1) decay model as well as a combination of an exponential and a linear decay (ExpLin). Based on a combined F-test comparing the triplicate fit results of the different models (see Methods, section 4.6.6), a model was chosen for every sampled methyl group. As a result, most peak intensities followed the ExpLin model, which describes an initial characteristically fast exponential with a subsequent slow linear decay; this is equivalent to approximating the slow second exponential term in ExpDec2 with a linear curve. Overall, unfolding of frGFP-ssrA by PAN at 60 °C proceeded with an average time constant $t_1 = 314 \pm 30$ s. As illustrated in Figure 5.25 A & B >90 % of all sampled methyl groups decayed with a time constant not significantly different from the mean time constant, considering the respective errors. Nevertheless, time constants of twelve peaks displayed a

significantly faster or slower decay, respectively. Peaks were considered to be significantly differing when their t_1 was outside the boundaries formed by the standard deviation of the average t_1 (red bars in Figure 5.25 A & C). As long as unambiguous assignments of the respective peaks were available, they are highlighted in the frGFP structure as blue spheres for faster- and magenta for slower decaying peaks, respectively. Among the ten faster decaying peaks there are three peaks which could not be assigned unambiguously to a particular residue (located in the random coil region of Ala β and Val γ). Similarly, two faster decaying peaks (47Ile δ 1 and 220Leu γ 1) do overlap severely with unassigned peaks in random coil regions, which emerges during the first few spectra of the time series. Since this overlap occludes the true intensity of the underlying assigned peaks, their fast decay is considered to be an artefact. Faster decaying assigned peaks not affected by overlap arise from methyl groups randomly scattered over frGFP's structure with no clearly identifiable pattern, (14Ile δ 1, 18Leu δ 2, 65Thry2, 68Val γ 2 and 120Val γ 2). Also the two peaks decaying significantly slower than average (136Ile δ 1 and 188Ile γ 2) were not related to each other. Yet, it has to be noted that 65Thry2, as part of the covalently formed chromophore, as well as 14Ile δ 1 undergo considerable chemical shift changes during the time series (The nature of these shift changes is covered in the following sections below). Upon addition of the 20S core particle frGFP is processed significantly faster with a mean t_1 of 227 ± 27 s while the obtained data of all but four individual peaks required fitting to a single exponential model (Figure 5.25 C). As for the unfolding no significant deviation from the mean t_1 was detected for most peaks except for the three faster decaying peaks 150Val γ 2, 98Ile γ 2 and 125Leu γ 1 (Figure 5.25 D).

To test whether a better time-resolution during a slower degradation reaction would reveal more amino-acid specific differences, I recorded a second data set of the unfolding and degradation reaction at 55 °C. At this temperature the activity of the archaeal PAN-20S proteasome is slowed down, despite at the expenses of a loss of total turnover compared to 60 °C (also refer to Figure 5.24). Due to limited sample quantities this data set could only be obtained as duplicate repeats of the reaction with PAN and PAN-20S, respectively.

As expected, analysis of the 55 °C-dataset confirmed a slower processing of frGFP-ssrA with an average unfolding- $t_1 = 426 \pm 67$ s and degradation- $t_1 = 419 \pm 78$ s, respectively, consistent with the higher activity of the PAN-20S proteasome at elevated temperatures. Similar as for the data obtained at 60 °C, most of the sampled peaks decayed within one standard deviation of the average t_1 (Figure 5.25 E & G). Likewise, during the unfolding without the 20S CP, intensities of 47Ile δ 1 and the same unassigned intensity in the Val γ -random-coil-region decayed significantly faster, as described above for data at 60 °C. Interestingly, also the chromophore methyl group 65Thry2 decayed faster, as it did at 60 °C. As illustrated in Figure 5.25 F, four residues decay significantly slower than average, which are also randomly distributed within the frGFP β -barrel structure (150Val γ 2, 18Leu δ 1, 123Ile γ 2, 62Thry2). During the PAN-20S mediated degradation at 55 °C no peaks decayed significantly slower than average, three peaks, however, showed a significantly faster decay behaviour (229Ile δ 1, 150Val γ 2, 14Ile δ 1, Figure 5.25 H). Although the kinetics of the reaction is obviously slower resulting in a slightly improved time resolution, the data from the studies at 55 °C do not provide any new insights. Thus, we conclude that frGFP does not unfold in a strictly sequence specific but rather in a cooperative manner when being processed by the PAN-20S proteasome.

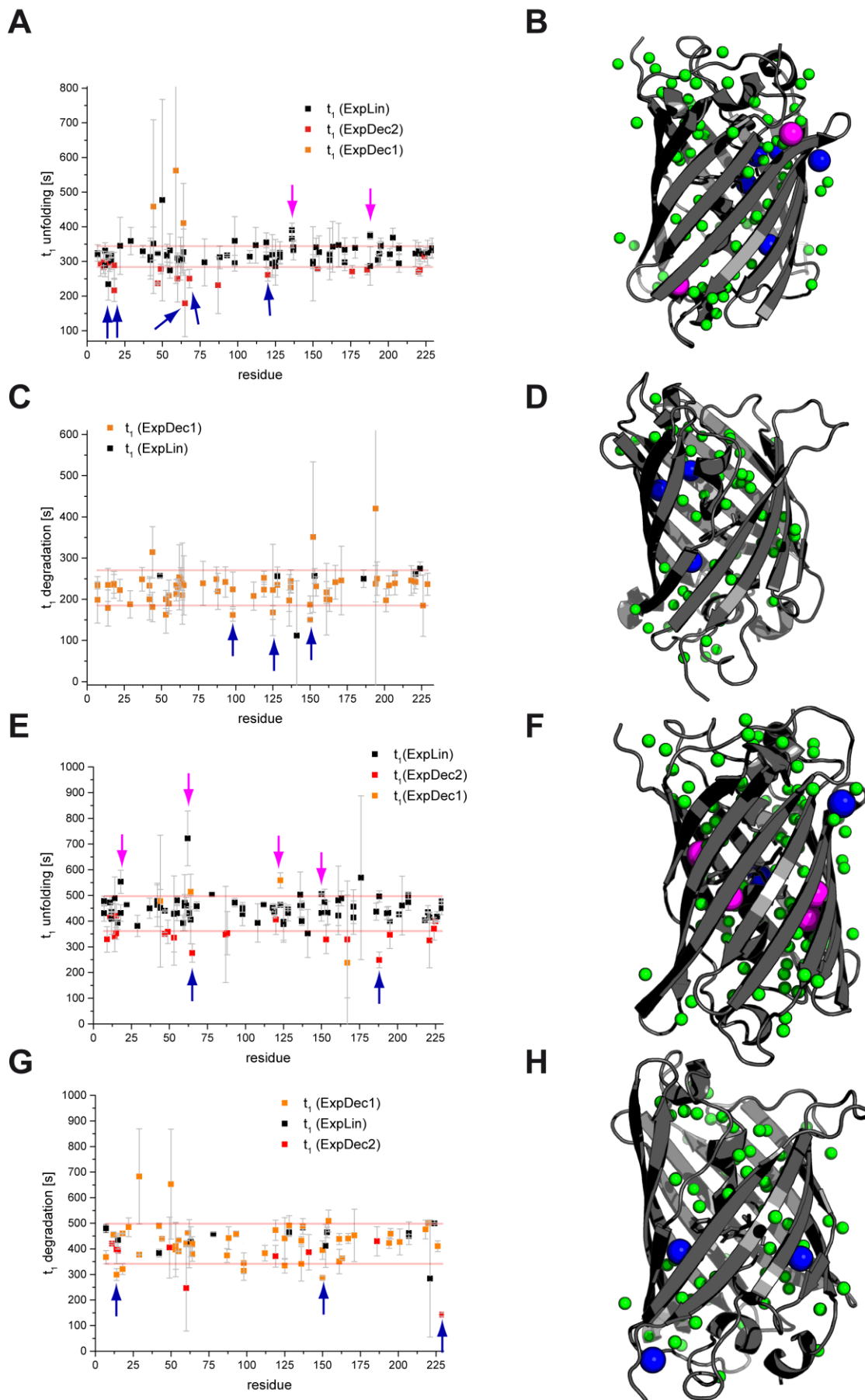


Figure 5.25 frGFP-ssrA is unfolded in a non-sequence specific manner Whether at 60 °C (A-D) or 55 °C (E-H) the methyl peaks of nearly all residues sampled in unfolding- and degradation NMR-time series decay with a

time constant t_1 which lies within one SD of the mean t_1 (pink boundaries in A, C, E, G, green spheres in B, D, F, H). During the unfolding in the absence of the proteasome core particle, five methyl peaks decay significantly faster at 60 °C (blue spheres in B) and three at 55 °C (blue spheres in F). Consistent with a higher activity of the thermophilic archaeal proteasome unfolding at 60 °C is on average significantly faster ($t_1 = 314 \pm 30$ s) than at 55 °C ($t_1 = 426 \pm 76$ s). In the presence of the proteasome core particle frGFPssrA is processed significantly faster with an average $t_1 = 227 \pm 42$ s (60 °C) and $t_1 = 419 \pm 78$ s (55 °C), respectively. Similar to unfolding the t_1 of only few methyl peaks deviates significantly from the mean t_1 during the degradation by the PAN-20S (C, D, G and H): Only three peaks decay significantly faster at either 60 °C or 55 °C (blue spheres in D and H, respectively). The random spread of the methyl groups with faster or slower degradation rates over the frGFP-ssrA crystal structure (PDB: 2B3Q) does not allow to suggest a particularly sequence-specific unfolding by PAN and the PAN-20S. frGFP-ssrA rather unfolds in a cooperative manner before translocation through PAN's central pore into the 20S core particle for proteolysis. Data from 60 °C and 55 °C was acquired as triplicate- and duplicate repeats, respectively. Error bars in A, C, E, G represent one standard deviation of the respective mean of the repeats. A modified version of this figure will appear in KRÜGER G. et al. (*manuscript in preparation*).

frGFP precipitates/aggregates quickly after being processed by PAN

Each ^1H , ^{13}C -SOFAST-HMQC spectrum of an NMR-time series does only show peaks of frGFP-ssrA-species which have a favourable relaxation behaviour. Thus, one can assume that after unfolding precipitated and aggregated species are essentially NMR-invisible. Thus, the conversion of entirely natively folded frGFP-ssrA into unfolded frGFP-ssrA and subsequent aggregation/precipitation leads to the decay of the overall peak intensity during the time-course, as described above. However, besides the apparent decay of most signals several subtle changes can be observed in the first few spectra of the unfolding time series: As mentioned above new peak intensity emerges in the random coil regions of Ile δ 1, Ile γ 2 Val γ , Leu γ and Ala β methyl groups which cannot be assigned to a particular residue (Figure 5.26 A). Most of this additional peak intensity is severely overlapped with peaks of native frGFP and cannot be resolved in the given NMR-spectra. One peak, however, is resolved well enough to extract peak intensities and fit them with an exponential function. Sitting in the random coil region of Val γ methyl groups this peak decays with a consistently small t_1 of 179 ± 10 s at 60 °C (331 ± 11 s at 55 °C) making it essentially one of the fastest decaying peaks among all sampled peaks. As mentioned above, several of the significantly faster decaying peaks similarly overlap with the newly emerging peak intensities. They all share a small t_1 together with extraordinary small standard deviations: 47Ile δ 1 ($t_1 = 236 \pm 2$ s), 220Leu δ 1 ($t_1 = 270 \pm 3$ s), ssrA-Ala β -I ($t_1 = 277 \pm 2$ s), ssrA-Ala β -II ($t_1 = 278 \pm 2$ s) (data from the data set at 60 °C). Thus, the fast decay of these overlapped peaks reflects the mixed contributions of the assigned peaks and the decay of the additional random coil intensities. In the case of 47Ile δ 1 the peak intensity in the first spectrum of a run with PAN is in fact 20 % higher compared to the control run (Figure 5.26 B), whereas for peaks not affected by these random coil intensities the respective intensities are around 35-40 % lower compared to the control, which reflects the drop of intensity during the 100 s dead time of the NMR-measurement (Figure 5.26 C).

Appearance of random coil peak intensities at the beginning of the unfolding reaction is a clear indicator for the presence of unfolded or partially unfolded frGFP-ssrA in the reaction mix. The fact that this random coil intensity decays with a fast time constant suggests that unfolded frGFP-ssrA is unstable at 60 °C and precipitates quickly after being unfolded and released by PAN, consequently becoming NMR-invisible.

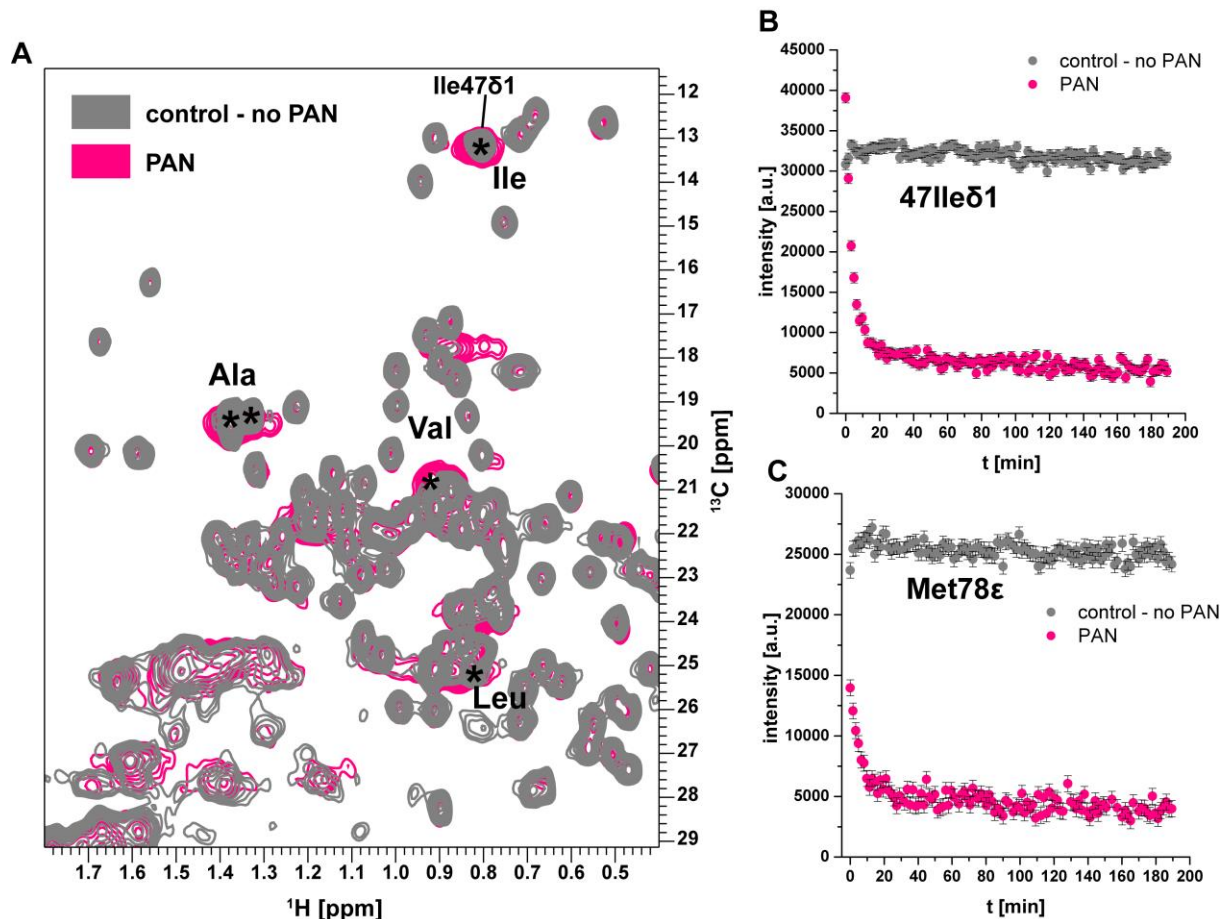


Figure 5.26 An unfolded frGFP-ssrA intermediate can be detected and precipitates quickly (A) The overlay of the first measured ^1H , ^{13}C -SOFAST-HMQC of frGFP-ssrA from a control series without PAN (grey) and an unfolding time series (magenta) display the appearance of new peak intensities in the random coil regions of Ala-/Ile-/Val- and Leu-methyl groups. Asterisks depict peaks sampled and analysed as described in the previous section and displayed in Figure 5.25 A & B: Two alanine peaks assigned to the C-terminal ssrA-tag, 47Ile δ 1, 220Leu δ 1 and a well resolved peak in the Val random coil region, which is unassigned. They all overlap with the newly appeared peak intensity and decay significantly faster than average with respect to their time constant. As exemplified by 47Ile δ 1 in (B) the affected peaks can have intensities 20 % higher than the control in the first spectrum of the time series as a result of the contribution of the underlying random coil intensity. This is particularly noticeable when comparing to an entirely unaffected peak in (C) which drops by 40 % in the first spectrum compared to the control as a result of unfolding happening in the dead time of the NMR-time series. The quick decay of the newly emerged random coil intensity reflects the precipitation of unfolded frGFP-ssrA at 60 °C, which renders it NMR-invisible. A modified version of this figure will appear in KRÜGER G. et al. (*manuscript in preparation*).

A close-to-native intermediate of frGFP-ssrA can be assessed by NMR chemical shift perturbation during the unfolding process

During an NMR time-series the peaks chemical shift perturbations are other important parameters to be considered. Indeed, besides the decay of intensities, I could notice that both during time courses of proteasomal-processing and, though to a much lower, control experiments without the PAN-20S, most of frGFP peaks shifted relative to the reference position (Figure 5.27 B). These chemical shift perturbations (CSPs) represent a sensitive indicator of a side-chain environment and molecular structural integrity and complement the data from peak intensity decays described above. The measured CSPs were not uniformly spread throughout the frGFP sequence ranging from 0.01 to 0.24 ppm (Figure 5.27 A). High CSPs clustered on β -strands 7, 8 and 10 and the central α -helix and affected methyl groups pointing towards the covalently attached chromophore (Figure 5.27 C & D). In the frGFP crystal structure (PDB: 2B3Q) 167Ile γ 1 (highest CSP with $\Delta\delta=0.24$ ppm) is located on strand β 8 and

is in very close proximity (3.9 Å) to 62Thry2 ($\Delta\delta=0.08$ ppm) on the central helix. 203Thry2 ($\Delta\delta=0.09$ ppm), the second furthest shifting peak, has a hydrophobic contact with 150Valy1 ($\Delta\delta=0.07$ ppm) at a distance of 3.9 Å. After having analysed the quality and quantity of these changes I was interested in their kinetics over the time course. Shift changes grew exponentially in unfolding runs, while they followed a linear function in control experiments with only ATP/MgCl₂ and no PAN or (Figure 5.27 E). Assuming that the process underlying the linear CSPs in control runs is present also in unfolding time series, the CSPs observed in unfolding experiments were corrected by the CSPs of control experiments before being fitted to an exponential (described in Material and Methods section 4.6.6). Finally, I extracted the exponential time constants for 26 peaks with an overall CSP > 0.02 ppm (Figure 5.27 F). With an average $t_1=636 \pm 324$ s at 60 °C (897 ± 334 s at 55 °C), the process underlying the shift changes is on a similar time scale as the intensity decay and therefore the unfolding reaction. Time-dependent CSPs could also be observed in PAN-20S degradation runs with an identical pattern, however quantitative analysis of time-resolved CSPs was hampered due to too much noise. Considering that the chemical shift changes are still relatively small, one can assume that frGFP-ssrA observed at each time point is still close-to-native, and only slightly altered. Thus, while the overall intensity of folded, soluble GFP decays over time the amount of the close-to-native species over native frGFP accumulates. Since the unfolding by PAN proceeds on a similar time-scale as the evolution of the CSPs, these processes are possibly closely related to each other. But where does this structurally slightly altered frGFP-ssrA species originate? Interestingly, the same peaks which exhibit strongest shift changes during the ATP-dependent unfolding or degradation do also exhibit strongest shifts upon addition of ATP and MgCl₂ to isolated frGFP, indicating a role of these additives in the shifting of multiple peaks. These additives contained at 100 mM each in controls and active proteasome runs contribute substantial ionic strength to the buffer. Further, PAN catalyzed ATP-hydrolysis increases the ionic strength even further through the release of ADP, inorganic phosphate and a proton. Thus, we conclude that the altered GFP state derives either from ADP binding or from the increased ionic strength. In any case, it seems to be related to ATP hydrolysis.

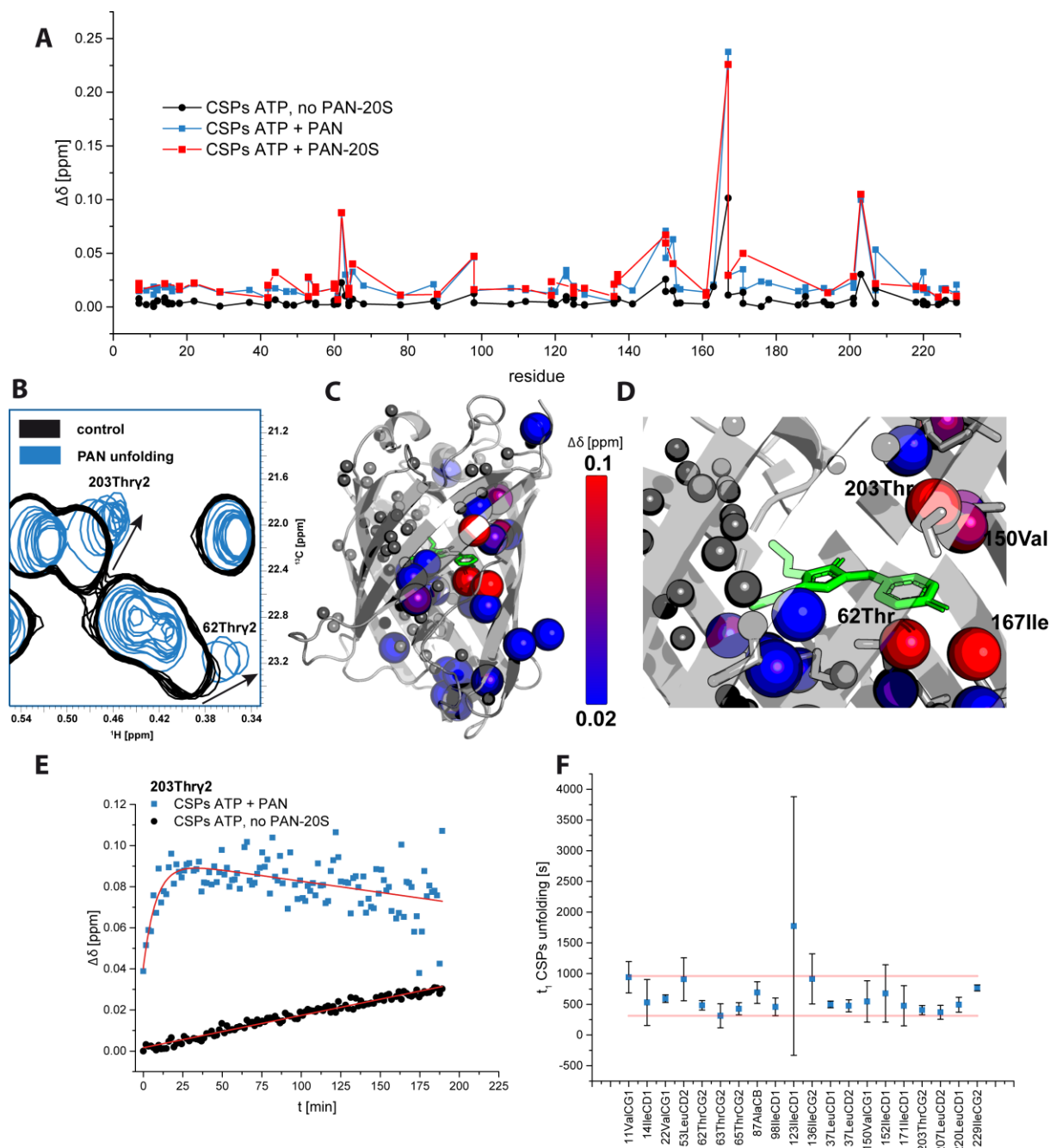


Figure 5.27 A close-to native intermediate can be observed during the unfolding (A) Cumulative chemical shift changes in GFP are not uniformly spread over GFP and are much more pronounced during the PAN-20S catalysed unfolding reaction. (B) An overlay of the first ten spectra of a control (black) and unfolding (blue) time series exemplifies the substantial chemical shift changes of two selected peaks during the first 15-20 minutes of the unfolding reaction. Each individual spectrum is displayed as one contour. (C) A cluster of high CSPs is located on the GFP face formed by β -strand 7,8,10 and the central helix. Grey spheres represent sampled peaks with a total CSP lower than 0.02 ppm. The four peaks with the highest CSPs (red) form two pairs of hydrophobic interactions as illustrated in (D): 167Ile δ 1 ($\Delta\delta$: 0.23 ppm) and 62Thry2 ($\Delta\delta$: 0.08 ppm), 203Thry2 ($\Delta\delta$: 0.09 ppm) and 150Val γ 1). (E) Time-resolved view of the CSPs of an unfolding and of a control time series reveals a linear kinetic of chemical shift changes for the control (black) and an exponential kinetic of chemical shift changes for the unfolding run (blue), respectively, as exemplified by 203Thry2. (F) Fitting of the time-resolved CSPs of methyl groups highlighted in (C) & (D) to an exponential function yielded time constants for data recorded at 60 °C between 300 and 1800 s (blue squares) with a mean $t_1 = 636 \pm 324$ s (pink boundaries) which is on a similar time scale as the unfolding reaction. A modified version of this figure will appear in KRÜGER G. et al. (*manuscript in preparation*).

The PAN-20S proteasome rapidly degrades substrate proteins into small unstructured oligopeptides

During the unfolding of PAN in the absence of the 20S core particle unfolded, frGFP-ssrA is released from the unfoldase and subsequently aggregates and precipitates. However, when PAN is combined with the 20S CP it interacts with the proteasome α -ring via its C-terminal HbYX motif and translocates the unfolded and linearized polypeptide towards the proteolytic cleavage sites buried in the lumen of the core particle. Here the peptide chain is hydrolysed into small fragments of around ten amino acids length and thus formation of aggregates of the unfolded protein is prevented.^{1,184} Consistent with this, due to precipitation, near complete loss of frGFP-ssrA methyl intensities can be observed at the end of the time-series when frGFP-ssrA is only unfolded by PAN but not hydrolysed (Figure 5.24 B and Figure 5.28 A, blue contours): Unfolded frGFP-ssrA forms aggregates, which become essentially NMR-invisible (also Figure 5.28 E, middle tube). Addition of the 20S core particle, however, leads to even faster and higher loss of methyl intensities (Figure 5.24) so that most dispersed methyl peaks are entirely vanished at the end of the reaction (Figure 5.28 A burgundy contours, E right tube). As opposed to the unfolding reaction the chemical shift dispersion in the ^1H -dimension collapses and only spectral intensity in the methyl-random coil regions can be detected, consistent with the absence of any structured ^{13}C -labelled species. Furthermore, several overlapped peaks in these random-coil regions emerge compared to the control experiments without PAN or the 20S. Analysis of the intensities of these rising peaks shows an exponential growth, as exemplified by four peaks representative of different methyl groups (Figure 5.28 B-E). Fitting of 20 growth curves with an exponential model resulted in an average growth- $t_1 = 355 \pm 155$ s. Thus, the growth of the unstructured random coil intensity, representative of small unstructured oligopeptides, is tightly related to the loss of frGFP-ssrA as described above, suggesting a well regulated cooperation between the unfoldase and the proteolytic core particle. The formation of aggregates, which is usually harmful in cells, is thereby avoided.

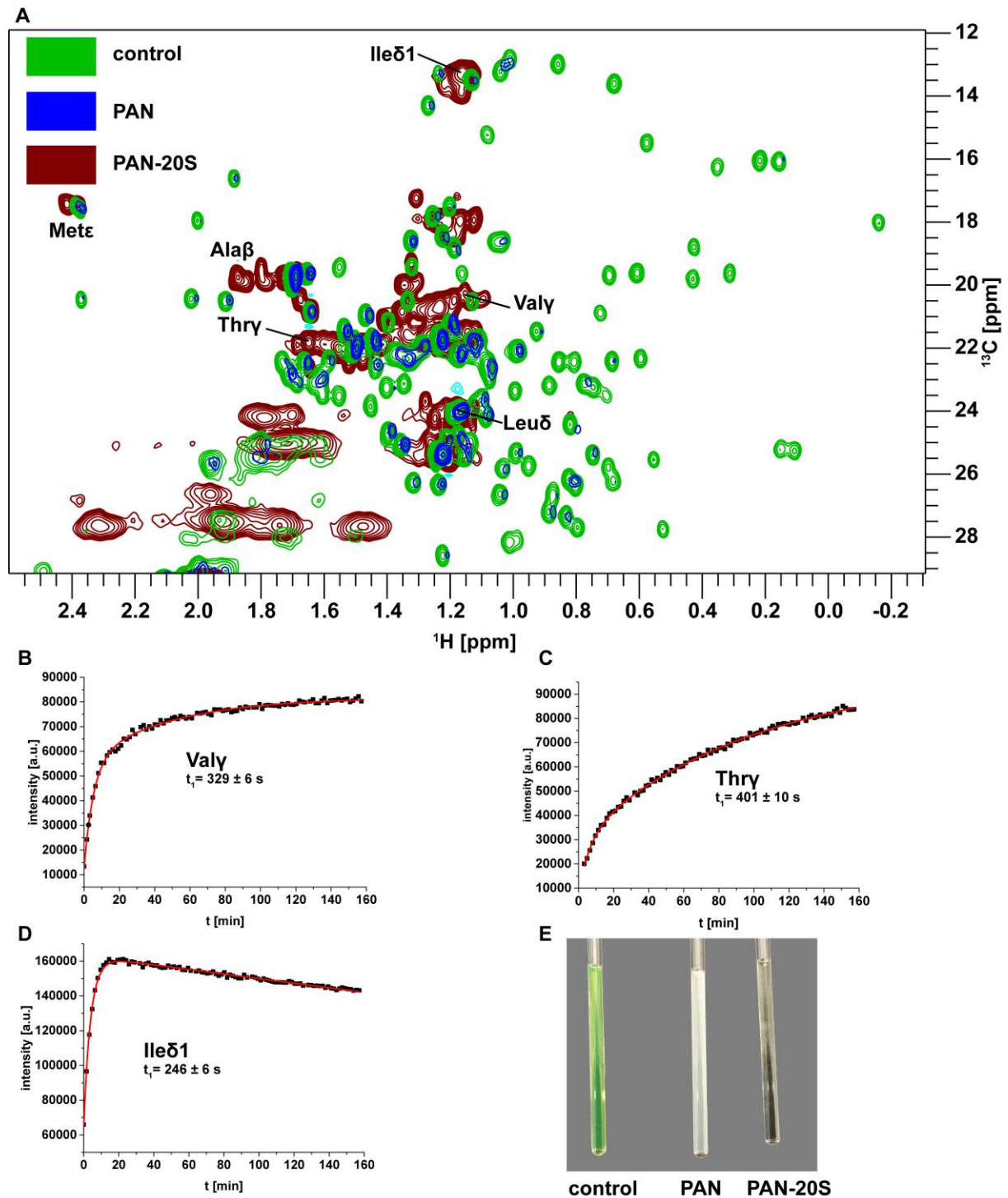


Figure 5.28 The PAN-20S proteasome hydrolyses frGFP-ssrA yielding unstructured peptide fragments (A) Superposition of ^1H , ^{13}C -SOFAST HMQC of frGFP-ssrA at the end of the respective NMR series, shows that frGFP-ssrA preserves its fold throughout the control series without PAN or the 20S CP (green contours, left tube in (E)). During the unfolding time-series (blue contours) most of frGFP-ssrA-intensity vanishes, consistent with formation of high molecular weight aggregates of unfolded protein (middle tube in (E)). Upon addition of the 20S CP, dispersed peaks of frGFP-ssrA disappear completely, leaving only newly appeared peaks in the random coil regions of the ILVMAT-methyl groups. Fitting the growth of intensities of the peaks in the the random coil region with an exponential function (B-D) yields time constants similar to the loss of peak intensities in the dispersed regions of the spectrum, consistent with a tight correlation between disappearing of folded protein and appearance of unstructured proteins. Visual inspection of the NMR-tube after degradation of frGFP-ssrA by the PAN-20S proteasome indeed does not show a sign of precipitation, consistent with the proteolysis of unfolded frGFP-ssrA into small oligopeptides (E, right tube).[†] A modified version of this figure will appear in KRÜGER G. et al. (*manuscript in preparation*).

5.2.5 Resonance-assignments of different GFP variants

To facilitate the amino acid resolved studies on the unfolding of GFP variants described in sections 5.2.4.1 knowledge of the methyl-groups chemical shift was crucial. Backbone resonance assignments of fluorescent proteins have been published for several fluorescent proteins including, GFPuv,^{160,185} and Venus.¹⁸⁶ All these assignments were carried out on GFP variants which tend to dimerize at higher concentration leading to significant line broadening in the dimer interface, preventing the assignment of the backbone resonances in these regions. In the present work, I initially used a variant of GFPuv harbouring the mutation A206K which completely prevents the dimer formation. Since time-resolved studies were to be carried out on the ILVM-methyl groups only, I used the published backbone assignment as a guide. Backbone-amide resonances of Ile, Val and Leu were identified using the reported backbone resonance assignment and confirmed by standard triple resonance backbone experiments recorded on GFPuv-ssrA. ILV-methyl resonance assignment was obtained with the help of H(CCCO)NH and HCCH-TOCSY spectra. Met ϵ -resonance assignment was achieved using a 3D-NOESY-¹³C-HSQC and extraction of inter methyl distances from the crystal structure (PDB: 1B9C). All experiments were recorded on GFPuv-ssrA at 42 °C to mimic conditions used by KHAN et al. in their assignment of GFPuv.¹⁶⁰ Finally, a total of 28/28 Ile- 36/40 Leu-, 30/32 Val- and 3/5 Met-methyl resonances could be assigned. Missing Leu- and Met- were located within the flexible C-terminal tail of the GFPuv-ssrA which presumably was degraded during the long duration of the backbone experiments.

After initial experiments on Ile-Met labelled GFPuv-ssrA I decided to use the optimized GFP variant frGFP. The introduction of the two point mutations F64L and S65T altered the methyl-¹H,¹³C-HSQC significantly so that no simple transfer of assignments was possible (Figure 5.29). Thus, I carried out the full backbone assignment and subsequently the assignment of Ala-, Ile-, Leu-, Val- Met- and Thr-methyl resonances. No unambiguous assignment could be obtained for the GFP C-terminal flexible tail and the C-terminal ssrA-tag, due to insufficient chemical shift dispersion of this unstructured region and also proteolytic degradation of this exposed region during the long duration of assignment experiments at the elevated temperature. Backbone amide resonances of 200/239 non-proline residues could be obtained. 18 amide resonances of residue 231 until 247 could not be assigned since this region comprises a C-terminal unstructured and flexible tail of GFP, which is further elongated by the ssrA-degradation tag. Among the other unassigned backbone regions are the stretches 143-149 and 202-205, which are part of adjacent β -strands that show distortions from the ideal β -barrel structure in the crystal structure (2B3Q), in agreement to what has been reported for the assignment of GFPuv.¹⁶⁰ However, assignment of methyl resonances, excluding the C-terminal flexible regions, was achieved to near completeness: 8/8 Ala-, 24/24 Ile-, 38/38 Leu- 30/32 Val-, 3/5 Met- and 17/17 Thr- methyl resonances were obtained.

For NMR-time series on the unfolding and degradation by the PAN-20S proteasome at 55 °C and 60 °, respectively, assignments of methyl resonances were transferred to the higher temperature by a temperature series of ¹H,¹³C-HSQC from 42 °C to 60 °C.

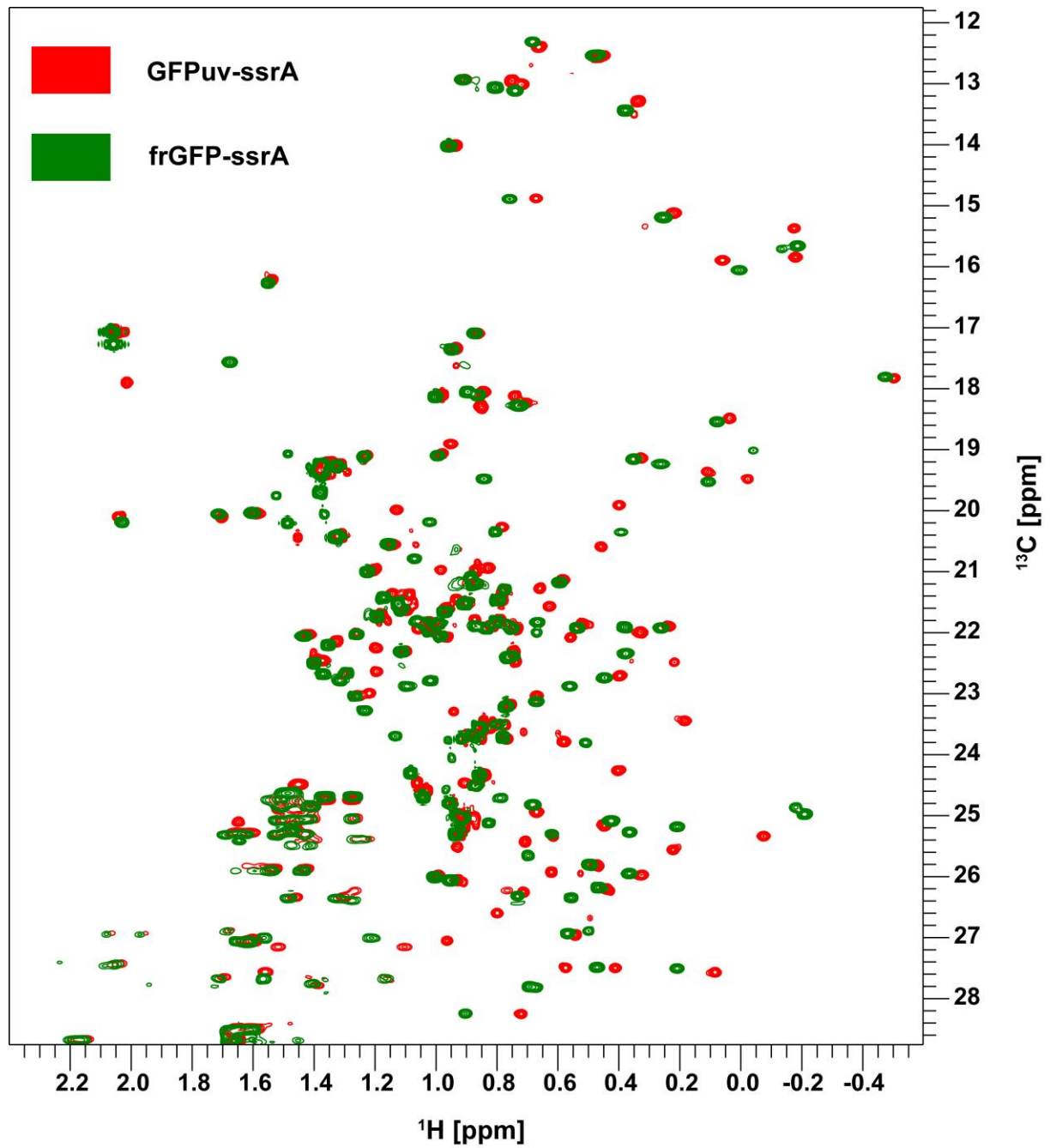


Figure 5.29 Methyl-region excerpt from ^1H - ^{13}C -HSQC spectra of GFPuv-ssrA (red) and frGFP-ssrA (green) recorded at 42 °C and 850MHz.

6 Discussion, Conclusions and Outlook

The proteasomal ATPase PAN is a member of the large AAA+ family of molecular machines which are involved in various cellular processes, involving the structural remodelling of biomolecules.⁵⁶ Common to all these processes are substantial conformational rearrangements in the machines which apply mechanical force to their substrate, which in turn also undergo large structural changes like unfolding, refolding, assembly and disassembly. The classical structural technique of X-ray crystallography contributed various landmark high-resolution structures of AAA+ ATPases setting the foundation for the functional understanding of the diverse AAA+ family. Further, a plethora of recent near-atomic resolution cryo-EM structures have shed light on the ATP-dependent mechanism of substrate processing by various AAA+ ATPases.^{63,106} However, common to these structural techniques is the requirement to restrict the molecular internal motions and dynamics range to allow for high-resolution structure determination. Common approaches for this include the use of slowly- or non-hydrolysable ATP-analogs, point mutations of ATPase catalytic residues or the truncation of flexible regions, which may result in heterogeneous or instable samples. Solution NMR spectroscopy, however, does not require crystallization or low-temperature vitrification of the biological samples and allows for the study of wild-type ATPases in close to native environments. Since the advent of NMR-methodologies optimized for large biological complexes (methyl-TROSY) and improved instrumentation, like cryogenic probes and high-field spectrometers, large molecular machines, including also AAA+ ATPases, have become in reach of NMR-spectroscopy.

In this thesis I studied the archaeal proteasome activating nucleotidase (PAN) and the processing of a model substrate by this unfoldase employing solution state NMR.

6.1 NMR studies on the archaeal proteasome activator PAN

The PAN-hexamer - a challenging subject for NMR-spectroscopy

At the beginning of this PhD project the high resolution structural information on archaeal PAN complexes was limited to crystal structures of the isolated N-terminal OB-ring-hexamer^{31,81} and a monomeric C-terminal AAA+ subdomain.³¹ While structural models of the 26S proteasome as the eukaryotic counterpart suggested a similar assembly of the individual subdomains in PAN, no structural model of full length PAN was available, leaving the exact interplay between substrate recognition at the N-terminus and ATP-hydrolysis-driven unfolding obscure. NMR spectroscopy promised to be the technique of choice to study the highly dynamic PAN complex both in its isolated form and in association with a substrate or the 20S proteolytic core particle. To date, no NMR-study has reported on a full length wild-type AAA+ ATPase with resonance assignments covering all subdomains.

Here, I pursued a “divide-and conquer approach”, based on obtaining the methyl-resonance assignments of PAN domains by standard NMR-methodologies and then transferring these assignments to the full protein for structural studies of the fully functional unfoldase. While resonance assignment of the OB-ring and a monomeric ATPase domain proved fairly straight forward (see below), NMR studies of full-length PAN represented a major challenge in terms of spectral quality.

The major challenge in NMR-studies of large biomolecular systems is the reduced sensitivity and line-broadening, due to faster relaxation of the transverse magnetization caused by longer rotational correlation times τ_c . However, methyl-TROSY based experiments and advanced isotopic labelling protocols have allowed acquiring well resolved spectra of complexes as big as 1 MDa.¹⁸⁷ A monodisperse sample of hexameric PAN weighs ~ 300 kDa (6x 49 kDa) and thus should be accessible by methyl-TROSY based experiments. However, although proposed to function as a hexameric ATPase, PAN from different archaeal species has been found to form as a 600 kDa complex by analytical size-exclusion chromatography, corresponding to a self-assembly of two hexameric rings into a dodecameric double ring. Using SEC-MALS, I established that wild-type PAN exists in a dynamic equilibrium between dodecamer and hexamer, and that this equilibrium can be shifted towards the hexameric species adding high concentrations of ATP or ADP. Indeed, as a consequence of a clearly reduced molecular weight and a homogeneous assembly state, methyl-TROSY spectra of isoleucine- δ 1 and methionine- ϵ residues improved substantially upon the addition of ADP to the sample. Still, the spectral resolution and sensitivity was of only modest quality and many resonances remained broad and overlapped. Interestingly, after transfer of resonance assignments from the isolated N- and C-terminus it became evident that the homomeric PAN protomers are present in at least three different conformations. A multitude of recent cryo-EM models and crystal structures of homohexameric AAA+ ATPases including PAN from *A. fulgidus* revealed highly asymmetric arrangements of the six identical ATPase protomers.^{41,106} Thus, conformational heterogeneity within each hexamers is more than likely and is in strong agreement with a functional asymmetry necessary for a rotary sequential mechanism of ATP-dependent translocation. A spiral staircase arrangement of the ATPase subunits has been established by multiple cryo-EM reconstructions¹⁰⁶ as the structure underlying a rotary sequential mechanism of ATP-dependent translocation. However, reports of conformational heterogeneity detected by NMR are ambiguous. In a study of p97, a multifunctional mammalian AAA+ ATPase, SCHUETZ and coworkers reported a single set of assigned methyl-TROSY-peaks,¹⁸⁸ although p97 is known to adopt an asymmetrical arrangement.¹⁰⁶ Contrary, a study from the same research group on the archaeal VAT, a homolog of p97, clearly establishes the split-ring spiral staircase architecture also by NMR.¹⁸⁹ In this study HUANG et al introduced single methyl-group labels via cysteine-chemistry to establish that residues at protomer-protomer interfaces clearly experience locally different chemical environments. They report two distinct states for the methyl label, which was in agreement with accompanying cryo-EM analyses. They suggested that two protomers at the split ring interface are disengaged from the ring-structure (state A) while the remaining four protomers are conformationally uniform. Although in the publication the authors do not describe any further heterogeneity within the four units of the ring except for the split site, careful inspection of the published spectra suggests the existence of heterogeneity also in the peaks of state B. Introducing the methyl-label in a region of VAT which was predicted to be more symmetric led to the observation of only one peak. From this is it evident that different regions in the hexameric AAA+ ATPases display various degrees of conformational heterogeneity. This is also seen in this work: for example, Met87 was found to give rise to three distinct well resolved peaks, while the conformational heterogeneity was less clear or absent for other peaks. M87 is located at the C-terminal end of the PAN's coiled coil domain close to the junction to the OB-ring. Intriguingly, this residue has been exploited in a cross-linking study by Snoberger et al. to analyse the conformational landscape of PAN. By cross-linking single cysteine residues

introduced at position 87 they found that only one third of PAN formed dimers, suggesting that the other two thirds of the coiled coil around Met87 are in a distinct non-cross-linkable conformation. Finally, they found that the three PAN coiled coils can indeed adopt three distinct conformations: One symmetrically “zipped”, one “unzipped” and one “zipped” out-of-register. The three distinct resonances of Met87 in the here presented methyl-TROSY spectra are in good agreement with these three coiled-coil conformations. Further conformational heterogeneity was also found for several methyl-probes of PAN’s ATPase subdomains (159Met ϵ , 266Ile δ 1, 267Ile δ 1). However, both in the N-terminal CC and OB-ring subdomains and the C-terminal ATPase domain most methyl-residues only gave rise to a single peak, indicating less heterogeneity or averaging of multiple highly flexible conformations. However, broad line-widths of most peaks in the Ile δ 1 region and the resulting spectral overlap obstructed their unambiguous assignment in the context of the full complex. Thus, drawing a full picture of PAN’s conformational landscape was limited a subset of peaks.

In conclusion, although the PAN oligomeric state was optimized to obtain mostly hexameric species, residual higher oligomeric species may still be present in the optimized conditions, as judged by SEC-MALS. Potentially, these bigger species are in exchange with the hexameric species resulting in the observed line broadening. Furthermore, full length PAN carries six 30-40-residues N-terminal tails which were found to be highly flexible and unstructured. They extend from the three coiled-coils which protrude from the almost spherical OB-ring-ATPase subcomplex. This asymmetric global shape, consisting of a rather compact core decorated with the coiled coils and additional disordered regions may cause slower molecular tumbling, which in turn yields to substantial line-broadening. Unfortunately, PAN constructs lacking the disordered region could not be obtained in a stable state, precluding NMR-studies of truncated and thus more compact PAN variants. Of note, AAA+ ATPases in the abovementioned studies do not contain any of the long protruding structures found in PAN and adopt a more compact spherical global shape. Additionally, the systems used in these studies allowed working at much higher protein concentrations, with respect to that achievable with PAN.

Nevertheless, although the spectral quality of full length PAN precluded sophisticated structural modelling of PAN, I could confirm the presence of a functional asymmetry, as it has been reported to be crucial for the proper functioning of force-generating AAA+ ATPases.

Conformational landscape of PAN-subdomains

AAA+ ATPase domain

A high-resolution crystal structure of the 30 kDa AAA+ ATPase domain had been available at the beginning of this PhD project.³¹ However, the PAN wild type ATPase domain has a tendency to form instable assembly up to the size of hexamers reflecting partial formation of hexameric ATPase rings. Thus, to allow full assignment of backbone and ILV-methyl resonance I had to find a stably monomeric variant. Modifying subunit building blocks to enable robust structural characterization of defined sizes is a valid strategy which has previously been exploited to allow for the transfer of resonance assignments from small subunits like the 21 kDa 20S core particle α -subunit to the full 670 kDa 20S core particle.²⁷ Careful investigation of the three ATPase subunits interfaces in the crystal structure as well as literature search¹⁶⁸ led to the discovery of a single point mutation which rendered the C-terminal

ATPase domain highly soluble and stable. With this mutant, I could achieve near complete backbone- and ILVM-methyl group assignment. Notably, the structure of the AAA+ subdomain appeared to be conserved in the context of the full-length protein, emphasising the relevance of assigning building blocks and transferring assignments to larger protein assemblies.

N-terminal OB-ring

Similarly, I intended to study the N-terminal OB-ring of PAN. Contrary to the AAA+ subdomains, this domain readily assembles into a hexamer, driven by the formation of three intermolecular coiled coils. Consequently, it is also considered to be the hexamerization scaffold for full-length PAN. The X-ray crystal structure of the OB-ring has already revealed that the structure is asymmetric due to a trans-proline-linkage in every other protomer around the ring, resulting in a pseudo-homohexamer that can be described as a trimer of dimers. However, by assigning most backbone-resonances of this domain in the hexameric oligomerization state, I discovered the existence of six conformations of the protomers within the ring, which are also confirmed in the assignment of isoleucine- δ 1 methyl groups. Intriguingly, peak intensities of different conformations of the same residue were fairly similar, ruling out the presence of major and minor alternate conformations. However, not every backbone amide gave rise to six resolved peaks: Many peaks obviously displayed an asymmetric appearance suggestive of more than one conformation, though they were often times not further resolved in 3D experiments to confidently assign distinct conformations. Consequently, one can assume that monomers within the ring adopt subtly different folds which might also imply a role of the OB-ring in mediating asymmetry in the full-length context of PAN. While the OB-ring is generally considered rigid, to build a constriction for folded protein elements during the unfolding, it must also allow for allosteric transmission of nucleotide states from the C-terminal AAA+-subdomain to the N-terminal CC-domain. The flexible linkers between the AAA+-ring and the OB-ring allow dynamic interfaces; however, conformational heterogeneity within the OB-ring, except for the prominent trans-proline linkage, had not been studied in detail. SNOBERGER and colleagues suggested that a conformational heterogeneity of PAN's coiled coils is also linked to the activity of AAA+ domains,³⁹ further corroborating the role of the OB-ring in allosterically transmitting asymmetries from the C-terminus to the N-terminus, potentially through distinct conformations of its individual.

PAN N-terminal coiled coil and disordered region

The extreme N-terminus of PAN is formed of three coiled coils which are decorated with assumed disordered regions of 30-50 residues. The formation of the coiled coils between helices of adjacent subunits is facilitated by a trans-proline peptide bond in three of the six subunits, thereby positioning the helices parallel to each other.³¹ While the original crystal structure of PAN reported a solely parallel and symmetric arrangement of the coiled coils, evidence has emerged, that in proteasomal ATPases also other arrangements are possible: In one cryo-EM model of the yeast proteasome the coiled coils formed by Rpt3/6 are not fully "zipped" and Rpt1/2 form a CC where the helices are displaced by 14 residues from a symmetric arrangement. Similar findings in PAN, obtained from cross-linking studies, indicate the evolutionary conservation from archaea to eukaryotes. However, the functional role of the

coiled coils in archaea is not yet fully understood, while in eukaryotic proteasomes they fulfill several regulatory functions through base-lid interactions. Similar regulatory components are not known in archaea, thus, it is speculated that the CC and the disordered region are mostly involved in substrate recruiting. Indeed, point mutations in both the PAN CC and the terminal residues of the substrate degradation tags appear to influence substrate processing.⁹⁴ For the disordered region at the N-terminal end of the CC no functional or structural studies have been reported.

Here, I intended to use solution NMR-spectroscopy to study a construct comprising both the disordered region and the coiled coils. Assignment of backbone-resonances of a coiled coil dimer indicated a highly symmetric arrangement, since only one set of peaks was observed. Consequently, the dynamic conformational switching reported by SNOBERGER et al for the CC in the context of full PAN or the 26S proteasome is not present in this isolated subdomain, highlighting the importance of allosteric regulation of these conformational heterogeneity by the C-terminal AAA+ subdomain. Initial visual inspection of peak line widths, secondary structure prediction based on backbone chemical shifts and subsequent backbone dynamics experiments revealed that the N-terminal boundary of the coiled coil domain is around residue 40. This represents the first experimental evidence that the PAN-N-terminus is indeed disordered. Previous studies on PAN and the 26S proteasome did not show any resolvable density for this region.^{19,31,41}

NMR titrations studies using a synthetic *ssrA*-peptide as a model substrate mimic did not show any specific binding to the PAN isolated CC domain. Thus, specific substrate interactions with the CC or the disordered region are unlikely. Substrate recruitment functionalities of the N-terminus are probably restricted to highly transient and unspecific interactions between disordered regions on both the substrate and PAN. This is in agreement with the reported abilities of the PAN N-terminus to prevent aggregation or induce refolding of unfolded and disordered proteins.^{35,81}

Intriguingly, I discovered that the full N-terminal domain of PAN contributes substantially to the overall stability of PAN. In attempts to find a PAN construct better behaved in NMR experiments I prepared several N-terminally truncated PAN versions. Common to all of them was a highly reduced stability which resulted in a high tendency to precipitate or aggregate. Either high concentrations of ATP - but not ADP - or the introduction of the point mutation F196A which destabilizes intersubunit contacts in the AAA+ subdomain promotes the solubility of the truncated construct. Whether there is a direct connectivity between the N-terminally disordered region and the C-terminal AAA+ subdomain is uncertain. However, comparison of low resolution cryo-EM reconstructions of a similarly truncated PAN version and full length PAN suggests that residues form the disordered region and parts of the coiled coil domain can also be contacting the sides of the PAN hexameric rings, instead of protruding to the top.⁴⁹ Thus, my results suggest that the coiled-coil domains and in particular the disordered N-terminus has the function to stabilize the structure of the complex (at least in absence of the 20S core particle), rather than to recruit the substrate.

6.1.1 PAN-substrate interactions

It is widely accepted that the most important interaction of AAA+ unfoldases with their substrates is the engagement of aromatic pore-1-loops buried within the AAA+ motor with unstructured flexible tails of substrates.¹⁰⁶ Consequently, even highly ubiquitylated globular pro-

teins are not processed by the 26S proteasome as long as they do not carry an unstructured initiation site long enough to reach into the AAA+ pore.¹⁹⁰ While in eukaryotes ubiquitylated proteins bind to specific receptors on proteasomes no such complementary system is known in archaea and the location of initial substrate interactions in archaeal proteasomes remains elusive.

In this thesis, I intended to exploit the residue specific resonance assignments in both full-length PAN and the isolated N-terminal domains of the OB-ring and the extended coiled coil domain, to study PAN substrate interactions. Although not native to archaea, the green fluorescent protein C-terminally fused to an *ssrA*-degradation tag originating from *E. coli* has long served as a convenient model substrate in studies of archaeal proteasomes.^{35,37,39,191}

Titration of GFP-*ssrA* into a solution of full-length PAN and 50 mM ADP did not result in any observable chemical shift changes, indicative of either very weak or no binding. Also CPMG relaxation dispersion experiments were unable to confirm binding. This was not completely unexpected since it has been suggested before that ADP-bound PAN is incompetent of binding *ssrA*-tagged substrates.⁴⁸ This is especially relevant for interactions of the disordered *ssrA*-tag to the pore-1-loops, which are assumed to become only accessible in an ATP bound state by coaxial alignment of the OB-ring and the AAA+-ring.²¹ However, the binding and hydrolysis of ATP is thought to be allosterically triggered and enhanced by substrate engagement suggesting that substrates also interact with not fully ATP-bound PAN where they are able to shift the unfoldase from an idle to a processing state.⁴⁸

One could argue that using ADP in such experiments is not wise, since it lowers the chances of detecting binding substantially. ATP or the slowly hydrolysable ATP-analogue ATP γ S, however, would not have constituted a valid alternative: PAN hydrolyses ATP under the NMR conditions at 60 °C within the order of minutes leaving PAN essentially ADP bound⁴⁸ likewise ATP γ S was found to be hydrolysed by AAA+ ATPases at elevated temperatures precluding the acquisition of long CPMG-experiments.

Besides the binding of *ssrA*-tagged substrates to pore-1-loops of AAA+ ATPases which has been confirmed experimentally,¹⁷⁵ the location of initial substrate engagement with PAN has remained speculative. However, the architecture of PAN with a C-terminal binding interface to the 20S proteolytic core particle and the N-terminal disordered region on the opposite site suggests involvement of the N-terminus in the first encounters of PAN with its substrates. To localize and study the nature of substrate interactions with the N-terminal subdomains of PAN I titrated a synthetic *ssrA*-peptide onto both the OB-ring and the extended coiled coil with its N-terminal disordered region. Even a large excess of peptide over the PAN subdomains did not result in any detectable chemical shift changes, suggesting that the peptide does not bind these domains specifically.

The above findings suggest that the binding of *ssrA*-tagged proteins to PAN does not involve a specific interaction with a certain surface region. Rather, it is attractive to speculate that the high flexibility of the six disordered tails and the conformational heterogeneity of the coiled coil sample a large conformational volume, interact transiently with disordered regions tails like the *ssrA*-tag on a substrate and thus increase the local concentration of proteasomal substrates leading to a higher probability of tail insertion into the AAA+ unfoldase motor. Further, this highly dynamic interaction of the PAN N-terminus with disordered and unstructured protein regions is in good agreement with multiple independent reports of PAN's N-terminus

exhibiting chaperone-like activities in refolding or prevention of aggregation of unfolded proteins.^{35,81}

6.2 Time-resolved NMR-studies on the processing of a model substrate by the PAN-20S proteasome

The following part of the discussion is drafted extensively considering findings from a collaborative publication on time-resolved SANS as part of the French-German project PROT-stretch within which this PhD thesis was written.¹

The processing of substrate proteins by AAA+ proteases like the PAN-20S proteasome is found to be a highly coordinated process involving substrate recognition, ATP-dependent cotranslocational unfolding by an unfoldase and subsequent proteolysis in a compartmentalized protease. The final step of peptide hydrolysis in proteasomes has been extensively studied by X-ray crystallography^{22,42} and solution NMR dynamics studies.^{24,27,43}

Substrate processing by AAA+ unfoldases that cap the 20S core particle, involves major conformational rearrangements which allow the unfoldase engage substrates, apply force and exert mechanical work on substrates, whereby the substrates' protein fold breaks and the protein unfolds.⁶³ The coupling of ATP-hydrolysis to these conformational changes in various AAA+ unfoldases has been elucidated by high-resolution cryo-EM of substrate engaged complexes¹⁰⁶ and a conserved sequentially ordered mechanism of ATP-hydrolysis around the cycle has been suggested for both the eukaryotic 26S proteasomes^{60,65} and the archetypal AAA+ unfoldase ClpX.^{69,70} Intriguingly, the understanding of the unfolding mechanism gained from previous extensive studies including several single-molecule mechanical unfolding experiments¹¹⁰ is challenged by the cryo-EM models resulting in a debate whether a strictly sequential rotary mechanism really describes the unfolding process.¹⁰⁸ Although they often come as an ensemble of several conformations, which can be manually put into a sequential order, relevant cryo-EM structures essentially represent structural snap shots lacking true temporal resolution. Furthermore, most studies lack appreciable spatial resolution of substrates outside the translocation channel thus not revealing (un-) folding intermediates prior the translocation.

In this part of my PhD project I applied solution NMR-spectroscopy to study the unfolding process of the model substrate folding reporter GFP-ssrA (hereafter named GFP-ssrA for simplicity) by the archaeal PAN-20S proteasome in real-time, thereby exploiting the uniquely high spatial resolution of this structural biology technique combined with the conservation of enzymatic activity and conformational dynamics in solution. The experimental approach was based on the repetitive recording of NMR spectra focusing on components involved in the unfolding reaction. Since NMR spectra of the unfoldase PAN suffered from low sensitivity, time-resolved experiments observing PAN signals were unfeasible. Consequently, I focused on the substrate of the unfolding reaction, the PAN model substrate GFP-ssrA.

Employing the fast repetition rate of sensitivity enhanced ^{13}C -SOFAS-HMQC NMR spectra I was able to follow the states that GFP-ssrA experienced during the processing by the PAN-20S proteasome with a high temporal resolution. NMR spectroscopy of selectively isotopically labelled GFP-ssrA allowed detecting exclusively signals of the substrate, since the unfoldase and the other components of the reaction mix were unlabelled and thus essentially NMR-invisible. Exploiting 100 assigned methyl group resonances as molecular probes I discovered that during the ATP-dependent processing by PAN, GFP-ssrA is cooperatively unfolded without detectable partially unfolded intermediates. This conclusion is based on the fact that more than 90 % of all methyl peaks decayed with a uniform time constant during the unfolding process. Very few peaks decayed either faster or slower than the average; however, these peaks were scattered on the 3D structure of GFP and thus not indicative of differential unfolding kinetics for specific structural regions of GFP. Similar observations were made when the proteolytic core particle was included in the reaction. However, consistent with previous studies,^{1,100} the loss of folded GFP-ssrA signals was significantly faster in the presence of the 20S proteasome, indicating a more efficient unfolding process. The similarity of the unfolding mechanisms with or without the protease is consistent with findings reported for the ClpX unfoldase.¹⁰⁰

Cooperative unfolding of substrates by AAA+ unfoldases has been previously described for the single domain protein GB1: AUGUSTYNYIAK and colleagues, for example, monitored the unfolding of substrates by cross-linking substrates to the archaeal AAA+ unfoldase VAT complexes. Cross-links to the substrates were designed to trap the unfoldase-substrates at defined stages of the unfolding process.⁹⁹ Employing methyl-TROSY NMR focused on selectively isotopically labelled substrates, they found that the single domain protein GB1 was entirely unfolded once engaged with the unfoldase, indicating that the first encounter with the gripping motions of the unfoldase motor resulted in complete unfolding. A classical double domain protein calmodulin, however, retained most of the native structure in one domain while the other domain was entirely unfolded, indicating a sequential threading of the substrate from one to the other terminus through the unfoldase pore.

In turn, the folding and unfolding of the single domain protein GFP has also been studied by various techniques, including hydrogen-deuterium exchange (HDX) NMR experiments,¹⁹² single molecule force spectroscopy^{193,194} as well as single-molecule optical tweezer studies using the AAA+ unfoldase ClpX.^{100,101,103,105} Based on these studies several unfolding pathways with different intermediates have been postulated. The presence and nature of these intermediates strongly depended on the unfolding method: A study of denaturant induced unfolding by HDX NMR experiments suggested an intermediate state formed after initial unfolding of β -strands 7-10, which were found to be least stable and most flexible.¹⁹² These findings could be confirmed in an unfolding study based on single molecule optical tweezers using the unfoldase ClpX: SEN and colleagues found that a first short-lived intermediate is formed by extraction of the extreme C-terminal β -strand 11 (lifetime: 45 ms) followed by another intermediate (lifetime: 130 ms) consisting of β -strands 1-6, after the unravelling of β -strands 7-10.¹⁰¹ In the current time-resolved NMR study I could not detect the presence of such unfolding intermediates. However, the temporal resolution of the NMR studies does not reach the ms timescale and thus the presence of these short-lived intermediates cannot be excluded. Interestingly, for ClpX a nonlinear relationship between successful unfolding of GFP and the ATP concentrations has been reported: At low ATP hydrolysis rates ClpX cannot efficiently translocate the $\Delta\beta$ 11 intermediate further fast enough to prevent its refolding

and consequently requires several attempts for complete unfolding.^{101,177} A similar behaviour for PAN, however, was not found, suggesting that a potential $\Delta\beta_{11}$ is of much lower frequency.¹⁹¹

Besides the decay of peak intensity of natively GFP I could also observe the transient appearance of new peaks in random coil regions of the spectrum during unfolding time courses without the 20S protease. Random coil peaks are a good indicator of poorly structured or unfolded proteins suggesting the appearance of transient unfolded protein species during the time-course of the NMR experiments. The random coil peaks could be detected only during the first 10 minutes of the unfolding reaction and decayed with a time-constant significantly smaller than average: this time-constant was consistent throughout all repetitions of the experiments, even at different temperatures. Potentially, intensities of unfolded GFP can arise from two distinct species: (1) A transient GFP species which has been pulled on by PAN and cooperatively unfolded but not yet fully translocated through the PAN-ring, similar to species characterized by inter bead extension after unfolding in single molecule experiments.¹⁰¹ The presence of a species was also predicted, albeit at low abundances, from time-resolved SANS (tr-SANS) data in our collaborators' study.¹ (2) Fully unfolded GFP released after complete translocation through PAN. The first species, potentially still bound to PAN, is unlikely to give rise to detectable signal, due to both its low abundance (GFP was used in 20 fold excess over PAN, consequently only a small fraction of GFP can be bound at any given time) and the fact that the PAN-bound species would have a worse relaxation behaviour in the hypothetical >300 kDa complex. Thus, the random coil peaks are likely to result from unfolded GFP which has been readily translocated through PAN. The early appearance and subsequent rapid decay of these peaks can be explained by the aggregation and subsequent precipitation of unfolded GFP molecules. A similar lag time between loss of native GFP signal and build-up of GFP aggregates could also be observed in the abovementioned tr-SANS study.¹ Here, the increase of the forward scattering intensity, I_0 , (which is proportional to the squared molecular weight of the particles in solution and thus extremely sensitive to the presence of oligomers) occurred with a slight with respect to the loss of natively folded GFP signal (measured by online fluorescence measurement). Thus, both tr-NMR and tr-SANS data are compatible with the transient presence of soluble unfolded GFP, which rapidly aggregates; the formation of the aggregates causes a rapid loss of random-coil NMR-intensity in NMR and an increase of the I_0 forward scattering intensity in SANS.

Potentially, unfolded and soluble GFP molecules could refold after they have been released from PAN. Such refolding has been a concern in unfolding studies using GFP^{195,196} but also other substrates⁹⁹ when AAA+ unfoldases were used without their associated proteases. In our experiments, the refolding of GFP, if present at all, is likely to be a minor event because the assay conditions were optimized to achieve maximal unfolding even without the 20S CP. Furthermore, the clear formation of visible aggregates indicates this aggregation is the major fate of unfolded GFP, especially as the reactions were carried out at either 55 or 60 °C, where GFP refolding should be less favoured than at 37 °C or ambient temperatures.

Addition of the 20S core particle prevented the formation of visible aggregates, consistent with efficient translocation of unfolded GFP polypeptides into the proteolytic chamber and hydrolysis to small soluble polypeptides. Similarly, in tr-SANS experiments with the full PAN-20S proteasome the forward scattering intensity, I_0 , decreases rapidly to zero, indicating the almost complete disappearance of high-molecular weight- GFP species.

Equally, I could detect the build-up of strong peaks exhibiting little chemical shift dispersion; these peaks are indicative of the formation of unstructured GFP-derived peptides, which can be well-detected by NMR (in contrast to the GFP-aggregates, which are NMR-invisible). Quantitative analysis of the time-dependence of these peaks revealed that they grow with an average time-constant similar to that of the intensity decay of folded GFP, indicating a tight coupling of the two processes. This is consistent with the results from the tr-SANS study¹: MAHIEU et al found that during the PAN-20S degradation of GFP the respective SANS curves showed a time-dependent decrease of I_0 (due to the decay of full-length GFP) and a concomitant increase of the scattering intensity at high q-angles $I(q)$ (which has a higher contribution from small molecular-weight rather than high molecular-weight particles). At each time point the SANS curve could be fitted to a mixture of natively folded GFP and decapeptide species, whereby the volume fraction of the decapeptide increased whereas that of natively folded GFP decreased over time. Furthermore, MALDI-TOF mass spectrometry conducted on PAN-20S degradation mixtures confirmed the presence of such decapeptides.¹

The efficient translocation of unfolded polypeptides from PAN to the proteolytic chambers in the 20S is an essential function of the proteasome because it prevents the release of unfolded proteins into the cytoplasm, which could form harmful protein aggregates. In eukaryotic proteasomes, association of the AAA+ unfoldase, residing in the 19S regulatory particle, with the 20S CP is ensured by the docking of multiple C-terminal tails including two loops and HbYX motifs of three AAA+ subunits which also induce gate-opening to allow passage of unfolded proteins.¹⁹ Additional lid-20S interactions hold the 26S proteasome together.⁸⁷ Although the archaeal PAN carries C-terminal HbYX motifs on all six subunits, the PAN-20S interaction has been described as very transient in the presence of either ATP or ADP.^{33,41} Nevertheless, our experiments demonstrate that unfolded polypeptides are translocated with high efficiency. The fact that the time constants of the decay of native GFP and raise of the peptides' concentration are similar suggests that the unfolded GFP is translocated directly from PAN to the 20S proteasome, without being first released in solution. Interestingly, the tr-SANS study used a PAN variant with a C-terminal His₆-tag following the HbYX motifs, consequently, PAN-20S interactions might have been weakened even further. This, however, did not affect the efficiency of translocation, in agreement with the hypothesis that the unfolded polypeptide chain itself acts as a tether between PAN and the 20S.¹ In this regard it would be interesting to characterize the of PAN-20S interactions during an unfolding experiment, potentially using FRET probes both in PAN and the 20S α -rings.

The time-resolved NMR study present here together with the published tr-SANS data highlight the unique complementarity of the two techniques: While NMR spectroscopy reports on the structure and dynamics at an atomic level in solution, contrast matched SANS reports on the overall structural integrity and state of the substrate at the different stages of unfolding. The NMR-time series data have a slightly lower time-resolution compared to the SANS data and a quite long dead time in the beginning of each time-course. After heat-activating the PAN-20S machinery by inserting the NMR tube into a preheated probe, roughly 2.5-3 min passed until the NMR-magnet and probe reached thermal equilibrium, locking and shimming were done and the first signal was acquired. During this dead time the 180 μ l sample was gradually heated by an N₂-airstream and the thermophilic enzyme was activated. As a result, GFP-signals in the first 2D SOFAST-HMQC spectrum of each time-series had already

dropped by ~ 40-50 %. This loss could not be avoided using the available instrumentation. IN principle, several instrumental set ups have been reported to initiate (un-) folding kinetics inside the NMR-spectrometer with minimal dead-time, including short laser pulses transferred to the NMR-tube using quartz optical fibres to release a chemical folding trigger from a chelator,^{197,198} or micro-injection devices to induce a pH jump.¹³⁹ Such systems in combination with SOFAST-NMR would have reduced the dead-time of unfolding experiments.

In addition, the time resolution of the NMR-experiments was limited by the relatively long duration of each time-point (100 sec) because of the low substrate concentration. In fact, while GFP can be concentrated beyond 1.5 mM, its concentration in the sample was limited by the necessity to keep an optimized molar ratio to PAN. Further, the effective observable concentration of GFP during the degradation reaction was decreased from 100 μ M (50-60 μ M after the dead-time) to < 20 μ M, as a result of unfolding, thus leading to the decrease of the signal-to noise ratio during the time-course.

7 Bibliography

- 1 Mahieu, E. *et al.* Observing Protein Degradation by the PAN-20S Proteasome by Time-Resolved Neutron Scattering. *Biophys. J.* **119**, 375-388, doi:https://doi.org/10.1016/j.bpj.2020.06.015 (2020).
- 2 Labbadia, J. *et al.* The Biology of Proteostasis in Aging and Disease. *Annu. Rev. Biochem.* **84**, 435-464, doi:10.1146/annurev-biochem-060614-033955 (2015).
- 3 Dever, T. E. *et al.* Mechanism and Regulation of Protein Synthesis in *Saccharomyces cerevisiae*. *Genetics* **203**, 65, doi:10.1534/genetics.115.186221 (2016).
- 4 Vihervaara, A. *et al.* Molecular mechanisms driving transcriptional stress responses. *Nature Reviews Genetics* **19**, 385-397, doi:10.1038/s41576-018-0001-6 (2018).
- 5 Yim, W. W.-Y. *et al.* Lysosome biology in autophagy. *Cell Discovery* **6**, 6, doi:10.1038/s41421-020-0141-7 (2020).
- 6 Dikic, I. Proteasomal and Autophagic Degradation Systems. *Annu. Rev. Biochem.* **86**, 193-224, doi:10.1146/annurev-biochem-061516-044908 (2017).
- 7 Huang, Q. *et al.* Ubiquitin/proteasome pathway impairment in neurodegeneration: therapeutic implications. *Apoptosis* **15**, 1292-1311, doi:10.1007/s10495-010-0466-z (2010).
- 8 Hegde, A. N. *et al.* Perturbations of Ubiquitin-Proteasome-Mediated Proteolysis in Aging and Alzheimer's Disease. *Front. Aging Neurosci.* **11**, doi:10.3389/fnagi.2019.00324 (2019).
- 9 Voutsadakis, I. A. The ubiquitin-proteasome system in colorectal cancer. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1782**, 800-808, doi:https://doi.org/10.1016/j.bbadis.2008.06.007 (2008).
- 10 Finley, D. Recognition and Processing of Ubiquitin-Protein Conjugates by the Proteasome. *Annu. Rev. Biochem.* **78**, 477-513, doi:10.1146/annurev.biochem.78.081507.101607 (2009).
- 11 Zhang, X. *et al.* Drug Development Targeting the Ubiquitin-Proteasome System (UPS) for the Treatment of Human Cancers. *Cancers (Basel)* **12**, doi:10.3390/cancers12040902 (2020).
- 12 Komander, D. *et al.* The Ubiquitin Code. *Annu. Rev. Biochem.* **81**, 203-229, doi:10.1146/annurev-biochem-060310-170328 (2012).
- 13 Hershko, A. *et al.* THE UBIQUITIN SYSTEM. *Annu. Rev. Biochem.* **67**, 425-479, doi:10.1146/annurev.biochem.67.1.425 (1998).
- 14 Nakayama, K. I. *et al.* Ubiquitin ligases: cell-cycle control and cancer. *Nature Reviews Cancer* **6**, 369-381, doi:10.1038/nrc1881 (2006).
- 15 Mattern, M. *et al.* Using Ubiquitin Binders to Decipher the Ubiquitin Code. *Trends Biochem. Sci.* **44**, 599-615, doi:10.1016/j.tibs.2019.01.011 (2019).
- 16 Xu, P. *et al.* Quantitative Proteomics Reveals the Function of Unconventional Ubiquitin Chains in Proteasomal Degradation. *Cell* **137**, 133-145, doi:https://doi.org/10.1016/j.cell.2009.01.041 (2009).
- 17 Vijay-Kumar, S. *et al.* Structure of ubiquitin refined at 1.8Å resolution. *J. Mol. Biol.* **194**, 531-544, doi:https://doi.org/10.1016/0022-2836(87)90679-6 (1987).
- 18 Maupin-Furlow, J. Proteasomes and protein conjugation across domains of life. *Nature Reviews Microbiology* **10**, 100-111, doi:10.1038/nrmicro2696 (2012).
- 19 Eisele, M. R. *et al.* Expanded Coverage of the 26S Proteasome Conformational Landscape Reveals Mechanisms of Peptidase Gating. *Cell reports* **24**, 1301-1315 e1305, doi:10.1016/j.celrep.2018.07.004 (2018).
- 20 Lander, G. C. *et al.* Complete subunit architecture of the proteasome regulatory particle. *Nature* **482**, 186-191, doi:10.1038/nature10774 (2012).
- 21 Bard, J. A. M. *et al.* Structure and Function of the 26S Proteasome. *Annu. Rev. Biochem.* **87**, 697-724, doi:10.1146/annurev-biochem-062917-011931 (2018).

- 22 Groll, M. *et al.* Structure of 20S proteasome from yeast at 2.4Å resolution. *Nature* **386**, 463-471, doi:10.1038/386463a0 (1997).
- 23 Chen, P. *et al.* Autocatalytic Subunit Processing Couples Active Site Formation in the 20S Proteasome to Completion of Assembly. *Cell* **86**, 961-972, doi:https://doi.org/10.1016/S0092-8674(00)80171-3 (1996).
- 24 Ruschak, A. M. *et al.* The proteasome antechamber maintains substrates in an unfolded state. *Nature* **467**, 868-871, doi:10.1038/nature09444 (2010).
- 25 Lapinaite, A. *et al.* The structure of the box C/D enzyme reveals regulation of RNA methylation. *Nature* **502**, 519-523, doi:10.1038/nature12581 (2013).
- 26 Savchenko, A. *et al.* Strategies for structural proteomics of prokaryotes: Quantifying the advantages of studying orthologous proteins and of using both NMR and X-ray crystallography approaches. *Proteins: Structure, Function, and Bioinformatics* **50**, 392-399, doi:10.1002/prot.10282 (2003).
- 27 Sprangers, R. *et al.* Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* **445**, 618-622, doi:10.1038/nature05512 (2007).
- 28 Humbard, M. A. *et al.* Prokaryotic proteasomes: nanocompartments of degradation. *Journal of molecular microbiology and biotechnology* **23**, 321-334, doi:10.1159/000351348 (2013).
- 29 Wilson, H. L. *et al.* Biochemical and physical properties of the Methanococcus jannaschii 20S proteasome and PAN, a homolog of the ATPase (Rpt) subunits of the eucaryal 26S proteasome. *J. Bacteriol.* **182**, 1680-1692 (2000).
- 30 Zwickl, P. *et al.* An Archaeobacterial ATPase, Homologous to ATPases in the Eukaryotic 26 S Proteasome, Activates Protein Breakdown by 20 S Proteasomes. *J. Biol. Chem.* **274**, 26008-26014, doi:10.1074/jbc.274.37.26008 (1999).
- 31 Zhang, F. *et al.* Structural insights into the regulatory particle of the proteasome from Methanocaldococcus jannaschii. *Mol. Cell* **34**, 473-484, doi:10.1016/j.molcel.2009.04.021 (2009).
- 32 Djuranovic, S. *et al.* Structure and Activity of the N-Terminal Substrate Recognition Domains in Proteasomal ATPases. *Mol. Cell* **34**, 580-590, doi:<http://dx.doi.org/10.1016/j.molcel.2009.04.030> (2009).
- 33 Smith, D. M. *et al.* ATP Binding to PAN or the 26S ATPases Causes Association with the 20S Proteasome, Gate Opening, and Translocation of Unfolded Proteins. *Mol. Cell* **20**, 687-698, doi:<http://dx.doi.org/10.1016/j.molcel.2005.10.019> (2005).
- 34 Berko, D. *et al.* The direction of protein entry into the proteasome determines the variety of products and depends on the force needed to unfold its two termini. *Mol. Cell* **48**, 601-611, doi:10.1016/j.molcel.2012.08.029 (2012).
- 35 Benaroudj, N. *et al.* PAN, the proteasome-activating nucleotidase from archaeobacteria, is a protein-unfolding molecular chaperone. *Nat. Cell Biol.* **2**, 833-839 (2000).
- 36 Brooks, C. *et al.* Archaeal Unfoldase Counteracts Protein Misfolding Retinopathy in Mice. *The Journal of Neuroscience* **38**, 7248, doi:10.1523/JNEUROSCI.0905-18.2018 (2018).
- 37 Kim, Y. C. *et al.* ATP binding to neighbouring subunits and intersubunit allosteric coupling underlie proteasomal ATPase function. *Nature communications* **6**, 8520, doi:10.1038/ncomms9520 (2015).
- 38 Smith, D. M. *et al.* Docking of the proteasomal ATPases' carboxyl termini in the 20S proteasome's alpha ring opens the gate for substrate entry. *Mol. Cell* **27**, 731-744, doi:10.1016/j.molcel.2007.06.033 (2007).
- 39 Snoberger, A. *et al.* Conformational switching in the coiled-coil domains of a proteasomal ATPase regulates substrate processing. *Nature communications* **9**, 2374, doi:10.1038/s41467-018-04731-6 (2018).
- 40 Ibrahim, Z. *et al.* Time-resolved neutron scattering provides new insight into protein substrate processing by a AAA+ unfoldase. *Sci. Rep.* **7**, 40948, doi:10.1038/srep40948 (2017).

- 41 Majumder, P. *et al.* Cryo-EM structures of the archaeal PAN-proteasome reveal an around-the-ring ATPase cycle. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 534-539, doi:10.1073/pnas.1817752116 (2019).
- 42 Lowe, J. *et al.* Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science* **268**, 533, doi:10.1126/science.7725097 (1995).
- 43 Religa, T. L. *et al.* Dynamic Regulation of Archaeal Proteasome Gate Opening As Studied by TROSY NMR. *Science* **328**, 98, doi:10.1126/science.1184991 (2010).
- 44 Ranjan, N. *et al.* Solution Structure and Activation Mechanism of Ubiquitin-Like Small Archaeal Modifier Proteins. *J. Mol. Biol.* **405**, 1040-1055, doi:https://doi.org/10.1016/j.jmb.2010.11.040 (2011).
- 45 Hepowitz, N. L. *et al.* Archaeal JAB1/MPN/MOV34 metalloenzyme (HvJAMM1) cleaves ubiquitin-like small archaeal modifier proteins (SAMPs) from protein-conjugates. *Mol. Microbiol.* **86**, 971-987, doi:10.1111/mmi.12038 (2012).
- 46 Smith, D. M. *et al.* Proteasomes and their associated ATPases: a destructive combination. *J. Struct. Biol.* **156**, 72-83, doi:10.1016/j.jsb.2006.04.012 (2006).
- 47 Benaroudj, N. *et al.* ATP Hydrolysis by the Proteasome Regulatory Complex PAN Serves Multiple Functions in Protein Degradation. *Mol. Cell* **11**, 69-78, doi:[http://dx.doi.org/10.1016/S1097-2765\(02\)00775-X](http://dx.doi.org/10.1016/S1097-2765(02)00775-X) (2003).
- 48 Smith, D. M. *et al.* ATP binds to proteasomal ATPases in pairs with distinct functional effects, implying an ordered reaction cycle. *Cell* **144**, 526-538, doi:10.1016/j.cell.2011.02.005 (2011).
- 49 Medalia, N. *et al.* Architecture and molecular mechanism of PAN, the archaeal proteasome regulatory ATPase. *J. Biol. Chem.* **284**, 22952-22960, doi:10.1074/jbc.M809643200 (2009).
- 50 Gottesman, S. *et al.* The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev.* **12**, 1338-1347 (1998).
- 51 Erala, J. *et al.* Ubiquitin-independent proteasomal degradation. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1843**, 216-221, doi:https://doi.org/10.1016/j.bbamcr.2013.05.008 (2014).
- 52 Jeong, Y. J. *et al.* Crystal structure of ubiquitin-like small archaeal modifier protein 1 (SAMP1) from *Haloflex volcanii*. *Biochem. Biophys. Res. Commun.* **405**, 112-117, doi:https://doi.org/10.1016/j.bbrc.2011.01.004 (2011).
- 53 Maupin-Furlow, J. A. Prokaryotic ubiquitin-like protein modification. *Annu. Rev. Microbiol.* **68**, 155-175, doi:10.1146/annurev-micro-091313-103447 (2014).
- 54 Hanson, P. I. *et al.* AAA+ proteins: have engine, will work. *Nature Reviews Molecular Cell Biology* **6**, 519-529, doi:10.1038/nrm1684 (2005).
- 55 Erzberger, J. P. *et al.* EVOLUTIONARY RELATIONSHIPS AND STRUCTURAL MECHANISMS OF AAA+ PROTEINS. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 93-114, doi:10.1146/annurev.biophys.35.040405.101933 (2006).
- 56 Miller, J. M. *et al.* Fundamental Characteristics of AAA+ Protein Family Structure and Function. *Archaea* **2016**, 9294307, doi:10.1155/2016/9294307 (2016).
- 57 Hanson, P. I. *et al.* AAA+ proteins: have engine, will work. *Nat. Rev. Mol. Cell Biol.* **6**, 519-529, doi:10.1038/nrm1684 (2005).
- 58 Wendler, P. *et al.* Structure and function of the AAA+ nucleotide binding pocket. *Biochim. Biophys. Acta* **1823**, 2-14, doi:10.1016/j.bbamcr.2011.06.014 (2012).
- 59 Karata, K. *et al.* Dissecting the Role of a Conserved Motif (the Second Region of Homology) in the AAA Family of ATPases: SITE-DIRECTED MUTAGENESIS OF THE ATP-DEPENDENT PROTEASE FtsH. *J. Biol. Chem.* **274**, 26225-26232 (1999).
- 60 de la Peña, A. H. *et al.* Substrate-engaged 26S proteasome structures reveal mechanisms for ATP-hydrolysis-driven translocation. *Science* **362**, eaav0725, doi:10.1126/science.aav0725 (2018).

- 61 Puchades, C. *et al.* Structure of the mitochondrial inner membrane AAA+ protease YME1 gives insight into substrate processing. *Science* **358**, eaao0464, doi:10.1126/science.aao0464 (2017).
- 62 Iyer, L. M. *et al.* Evolutionary history and higher order classification of AAA+ ATPases. *J. Struct. Biol.* **146**, 11-31, doi:https://doi.org/10.1016/j.jsb.2003.10.010 (2004).
- 63 Zhang, S. *et al.* AAA+ ATPases in Protein Degradation: Structures, Functions and Mechanisms. *Biomolecules* **10**, doi:10.3390/biom10040629 (2020).
- 64 Zhang, F. *et al.* Mechanism of Substrate Unfolding and Translocation by the Regulatory Particle of the Proteasome from *Methanocaldococcus jannaschii*. *Mol. Cell* **34**, 485-496, doi:http://dx.doi.org/10.1016/j.molcel.2009.04.022 (2009).
- 65 Dong, Y. *et al.* Cryo-EM structures and dynamics of substrate-engaged human 26S proteasome. *Nature* **565**, 49-55, doi:10.1038/s41586-018-0736-4 (2019).
- 66 Shin, M. *et al.* Structural basis for distinct operational modes and protease activation in AAA+ protease Lon. *Science Advances* **6**, eaba8404, doi:10.1126/sciadv.aba8404 (2020).
- 67 Chang, C.-W. *et al.* Structural Elements Regulating AAA+ Protein Quality Control Machines. *Frontiers in molecular biosciences* **4**, doi:10.3389/fmolb.2017.00027 (2017).
- 68 Olszewski, M. M. *et al.* The Cdc48 unfoldase prepares well-folded protein substrates for degradation by the 26S proteasome. *Communications Biology* **2**, 29, doi:10.1038/s42003-019-0283-z (2019).
- 69 Ripstein, Z. A. *et al.* A processive rotary mechanism couples substrate unfolding and proteolysis in the ClpXP degradation machinery. *eLife* **9**, e52158, doi:10.7554/eLife.52158 (2020).
- 70 Fei, X. *et al.* Structures of the ATP-fueled ClpXP proteolytic machine bound to protein substrate. *eLife* **9**, e52774, doi:10.7554/eLife.52774 (2020).
- 71 Botos, I. *et al.* Cryo-EM structure of substrate-free *E. coli* Lon protease provides insights into the dynamics of Lon machinery. *Current Research in Structural Biology* **1**, 13-20, doi:https://doi.org/10.1016/j.crstbi.2019.10.001 (2019).
- 72 Bochtler, M. *et al.* The structures of HslU and the ATP-dependent protease HslU–HslV. *Nature* **403**, 800-805, doi:10.1038/35001629 (2000).
- 73 Majumder, P. *et al.* Proteasomes: unfoldase-assisted protein degradation machines. *Biol. Chem.* **401**, 183-199, doi:10.1515/hsz-2019-0344 (2019).
- 74 Huber, E. M. *et al.* A unified mechanism for proteolysis and autocatalytic activation in the 20S proteasome. *Nature communications* **7**, 10900, doi:10.1038/ncomms10900 (2016).
- 75 DeMartino, G. N. *et al.* PA700, an ATP-dependent activator of the 20 S proteasome, is an ATPase containing multiple members of a nucleotide-binding protein family. *J. Biol. Chem.* **269**, 20878-20884 (1994).
- 76 Dubiel, W. *et al.* Subunit 4 of the 26 S protease is a member of a novel eukaryotic ATPase family. *J. Biol. Chem.* **267**, 22699-22702 (1992).
- 77 Glickman, M. H. *et al.* The Regulatory Particle of the *Saccharomyces cerevisiae* Proteasome. *Mol. Cell. Biol.* **18**, 3149, doi:10.1128/MCB.18.6.3149 (1998).
- 78 Ehlers, C. K., Friedrich; Dannemann, Burkhardt. Screening for Molecules Interacting with Proteasomes in *Thermoplasma acidophilum*. *Biol. Chem.* **378**, 249-254, doi:https://doi.org/10.1515/bchm.1997.378.3-4.249 (1997).
- 79 Bult, C. J. *et al.* Complete Genome Sequence of the Methanogenic Archaeon, *Methanococcus jannaschii*. *Science* **273**, 1058, doi:10.1126/science.273.5278.1058 (1996).
- 80 Peters, J.-M. *et al.* Structural Features of the 26 S Proteasome Complex. *J. Mol. Biol.* **234**, 932-937, doi:https://doi.org/10.1006/jmbi.1993.1646 (1993).

- 81 Djuranovic, S. *et al.* Structure and activity of the N-terminal substrate recognition domains in proteasomal ATPases. *Mol. Cell* **34**, 580-590, doi:10.1016/j.molcel.2009.04.030 (2009).
- 82 Smith, D. M. *et al.* ATP binding to PAN or the 26S ATPases causes association with the 20S proteasome, gate opening, and translocation of unfolded proteins. *Mol. Cell* **20**, 687-698, doi:10.1016/j.molcel.2005.10.019 (2005).
- 83 Matyskiela, M. E. *et al.* Conformational switching of the 26S proteasome enables substrate degradation. *Nat. Struct. Mol. Biol.* **20**, 781-788, doi:10.1038/nsmb.2616 (2013).
- 84 Śledź, P. *et al.* Structure of the 26S proteasome with ATP-γS bound provides insights into the mechanism of nucleotide-dependent substrate translocation. *Proceedings of the National Academy of Sciences* **110**, 7264-7269, doi:10.1073/pnas.1305782110 (2013).
- 85 Beck, F. *et al.* Near-atomic resolution structural model of the yeast 26S proteasome. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 14870-14875, doi:10.1073/pnas.1213333109 (2012).
- 86 Huang, X. *et al.* An atomic structure of the human 26S proteasome. *Nat. Struct. Mol. Biol.* **23**, 778-785, doi:10.1038/nsmb.3273 (2016).
- 87 Lasker, K. *et al.* Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 1380-1387, doi:10.1073/pnas.1120559109 (2012).
- 88 Unverdorben, P. *et al.* Deep classification of a large cryo-EM dataset defines the conformational landscape of the 26S proteasome. *Proceedings of the National Academy of Sciences* **111**, 5544, doi:10.1073/pnas.1403409111 (2014).
- 89 Chen, S. *et al.* Structural basis for dynamic regulation of the human 26S proteasome. *Proceedings of the National Academy of Sciences* **113**, 12991, doi:10.1073/pnas.1614614113 (2016).
- 90 Lu, Y. *et al.* Conformational Landscape of the p28-Bound Human Proteasome Regulatory Particle. *Mol. Cell* **67**, 322-333.e326, doi:10.1016/j.molcel.2017.06.007 (2017).
- 91 Wehmer, M. *et al.* Structural insights into the functional cycle of the ATPase module of the 26S proteasome. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 1305-1310, doi:10.1073/pnas.1621129114 (2017).
- 92 Bard, J. A. M. *et al.* The 26S Proteasome Utilizes a Kinetic Gateway to Prioritize Substrate Degradation. *Cell* **177**, 286-298 e215, doi:10.1016/j.cell.2019.02.031 (2019).
- 93 Zhu, Y. *et al.* Structural mechanism for nucleotide-driven remodeling of the AAA-ATPase unfoldase in the activated human 26S proteasome. *Nature communications* **9**, 1360, doi:10.1038/s41467-018-03785-w (2018).
- 94 Zhang, F. *et al.* Mechanism of substrate unfolding and translocation by the regulatory particle of the proteasome from *Methanocaldococcus jannaschii*. *Mol. Cell* **34**, 485-496, doi:10.1016/j.molcel.2009.04.022 (2009).
- 95 Forster, A. *et al.* The 1.9 Å structure of a proteasome-11S activator complex and implications for proteasome-PAN/PA700 interactions. *Mol. Cell* **18**, 589-599, doi:10.1016/j.molcel.2005.04.016 (2005).
- 96 Yu, Z. *et al.* Allosteric coupling between α-rings of the 20S proteasome. *Nature communications* **11**, 4580, doi:10.1038/s41467-020-18415-7 (2020).
- 97 Stinson, B. M. *et al.* Nucleotide binding and conformational switching in the hexameric ring of a AAA+ machine. *Cell* **153**, 628-639, doi:10.1016/j.cell.2013.03.029 (2013).
- 98 Glynn, S. E. *et al.* Structures of asymmetric ClpX hexamers reveal nucleotide-dependent motions in a AAA+ protein-unfolding machine. *Cell* **139**, 744-756, doi:10.1016/j.cell.2009.09.034 (2009).
- 99 Augustyniak, R. *et al.* Cotranslocational processing of the protein substrate calmodulin by an AAA+ unfoldase occurs via unfolding and refolding intermediates.

- Proc. Natl. Acad. Sci. U. S. A.* **115**, E4786-E4795, doi:10.1073/pnas.1721811115 (2018).
- 100 Maillard, R. A. *et al.* ClpX(P) generates mechanical force to unfold and translocate its protein substrates. *Cell* **145**, 459-469, doi:10.1016/j.cell.2011.04.010 (2011).
- 101 Sen, M. *et al.* The ClpXP protease unfolds substrates using a constant rate of pulling but different gears. *Cell* **155**, 636-646, doi:10.1016/j.cell.2013.09.022 (2013).
- 102 Bustamante, C. *et al.* Revisiting the central dogma one molecule at a time. *Cell* **144**, 480-497, doi:10.1016/j.cell.2011.01.033 (2011).
- 103 Cordova, J. C. *et al.* Stochastic but highly coordinated protein unfolding and translocation by the ClpXP proteolytic machine. *Cell* **158**, 647-658, doi:10.1016/j.cell.2014.05.043 (2014).
- 104 Lu, Y. *et al.* Substrate degradation by the proteasome: A single-molecule kinetic analysis. *Science* **348**, 1250834, doi:10.1126/science.1250834 (2015).
- 105 Aubin-Tam, M. E. *et al.* Single-molecule protein unfolding and translocation by an ATP-fueled proteolytic machine. *Cell* **145**, 257-267, doi:10.1016/j.cell.2011.03.036 (2011).
- 106 Gates, S. N. *et al.* Stairway to translocation: AAA+ motor structures reveal the mechanisms of ATP-dependent substrate translocation. *Protein Sci.* **29**, 407-419, doi:10.1002/pro.3743 (2020).
- 107 Gates, S. N. *et al.* Ratchet-like polypeptide translocation mechanism of the AAA+ disaggregase Hsp104. *Science* **357**, 273-279, doi:10.1126/science.aan1052 (2017).
- 108 Tsai, F. T. F. *et al.* Same structure, different mechanisms? *eLife* **9**, e56501, doi:10.7554/eLife.56501 (2020).
- 109 Olivares, A. O. *et al.* Mechanistic insights into bacterial AAA+ proteases and protein-remodelling machines. *Nature reviews. Microbiology* **14**, 33-44, doi:10.1038/nrmicro.2015.4 (2016).
- 110 Olivares, A. O. *et al.* Mechanical Protein Unfolding and Degradation. *Annu. Rev. Physiol.* **80**, 413-429, doi:10.1146/annurev-physiol-021317-121303 (2018).
- 111 Martin, A. *et al.* Rebuilt AAA + motors reveal operating principles for ATP-fuelled machines. *Nature* **437**, 1115-1120, doi:10.1038/nature04031 (2005).
- 112 Beckwith, R. *et al.* Reconstitution of the 26S proteasome reveals functional asymmetries in its AAA+ unfoldase. *Nat. Struct. Mol. Biol.* **20**, 1164-1172, doi:10.1038/nsmb.2659
- <http://www.nature.com/nsmb/journal/v20/n10/abs/nsmb.2659.html#supplementary-information> (2013).
- 113 Sekhar, A. *et al.* NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proceedings of the National Academy of Sciences* **110**, 12867, doi:10.1073/pnas.1305688110 (2013).
- 114 Korzhnev, D. M. *et al.* Probing Invisible, Low-Populated States of Protein Molecules by Relaxation Dispersion NMR Spectroscopy: An Application to Protein Folding. *Accounts of Chemical Research* **41**, 442-451, doi:10.1021/ar700189y (2008).
- 115 Cavanagh, J. *et al.* in *Protein NMR Spectroscopy (Second Edition)* (eds John Cavanagh *et al.*) 1-28 (Academic Press, 2007).
- 116 Carbajo, R. J., & Neira, J. L. *NMR for Chemists and Biologists*. (Springer Netherlands, 2013).
- 117 Kemp, W. in *NMR in Chemistry: A Multinuclear Introduction* (ed William Kemp) 14-28 (Macmillan Education UK, 1986).
- 118 Pervushin, K. *et al.* Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences* **94**, 12366, doi:10.1073/pnas.94.23.12366 (1997).
- 119 Tugarinov, V. *et al.* Ile, Leu, and Val Methyl Assignments of the 723-Residue Malate Synthase G Using a New Labeling Strategy and Novel NMR Methods. *J. Am. Chem. Soc.* **125**, 13868-13878, doi:10.1021/ja030345s (2003).

- 120 Tugarinov, V. *et al.* Cross-Correlated Relaxation Enhanced ^1H - ^{13}C NMR Spectroscopy of Methyl Groups in Very High Molecular Weight Proteins and Protein Complexes. *J. Am. Chem. Soc.* **125**, 10420-10428, doi:10.1021/ja030153x (2003).
- 121 Schütz, S. *et al.* Methyl TROSY spectroscopy: A versatile NMR approach to study challenging biological systems. *Progress in nuclear magnetic resonance spectroscopy* **116**, 56-84, doi:https://doi.org/10.1016/j.pnmrs.2019.09.004 (2020).
- 122 Danilenko, N. *et al.* Histone chaperone exploits intrinsic disorder to switch acetylation specificity. *Nature communications* **10**, 3435, doi:10.1038/s41467-019-11410-7 (2019).
- 123 Inagaki, F. in *Encyclopedia of Biophysics* (ed Gordon C. K. Roberts) 2033-2037 (Springer Berlin Heidelberg, 2013).
- 124 Salzmann, M. *et al.* TROSY-type Triple-Resonance Experiments for Sequential NMR Assignments of Large Proteins. *J. Am. Chem. Soc.* **121**, 844-848, doi:10.1021/ja9834226 (1999).
- 125 Salzmann, M. *et al.* TROSY in triple-resonance experiments: New perspectives for sequential NMR assignment of large proteins. *Proceedings of the National Academy of Sciences* **95**, 13585, doi:10.1073/pnas.95.23.13585 (1998).
- 126 Kumar, A. *et al.* Real-time protein NMR spectroscopy and investigation of assisted protein folding. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1850**, 1965-1972, doi:https://doi.org/10.1016/j.bbagen.2014.12.003 (2015).
- 127 Kay, L. E. *et al.* Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **28**, 8972-8979, doi:10.1021/bi00449a003 (1989).
- 128 Renner, C. *et al.* Practical aspects of the 2D ^{15}N - $\{^1\text{H}\}$ -NOE experiment. *J. Biomol. NMR* **23**, 23-33, doi:10.1023/A:1015385910220 (2002).
- 129 Lipari, G. *et al.* Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104**, 4546-4559, doi:10.1021/ja00381a009 (1982).
- 130 Kempf, J. G. *et al.* Protein dynamics from solution NMR. *Cell Biochem. Biophys.* **37**, 187-211, doi:10.1385/CBB:37:3:187 (2002).
- 131 Kovermann, M. *et al.* Protein dynamics and function from solution state NMR spectroscopy. *Q. Rev. Biophys.* **49**, e6, doi:10.1017/S0033583516000019 (2016).
- 132 Kleckner, I. R. *et al.* An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1814**, 942-968, doi:https://doi.org/10.1016/j.bbapap.2010.10.012 (2011).
- 133 Korzhnev, D. M. *et al.* Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* **430**, 586-590, doi:10.1038/nature02655 (2004).
- 134 Korzhnev, D. M. *et al.* Probing Slow Dynamics in High Molecular Weight Proteins by Methyl-TROSY NMR Spectroscopy: Application to a 723-Residue Enzyme. *J. Am. Chem. Soc.* **126**, 3964-3973, doi:10.1021/ja039587i (2004).
- 135 Balbach, J. *et al.* Following protein folding in real time using NMR spectroscopy. *Nat. Struct. Biol.* **2**, 865-870, doi:10.1038/nsb1095-865 (1995).
- 136 Balbach, J. *et al.* Protein Folding Monitored at Individual Residues During a Two-Dimensional NMR Experiment. *Science* **274**, 1161, doi:10.1126/science.274.5290.1161 (1996).
- 137 Schanda, P. *et al.* SOFAST-HMQC Experiments for Recording Two-dimensional Heteronuclear Correlation Spectra of Proteins within a Few Seconds. *J. Biomol. NMR* **33**, 199-211, doi:10.1007/s10858-005-4425-x (2005).
- 138 Schanda, P. *et al.* Very Fast Two-Dimensional NMR Spectroscopy for Real-Time Investigation of Dynamic Events in Proteins on the Time Scale of Seconds. *J. Am. Chem. Soc.* **127**, 8014-8015, doi:10.1021/ja051306e (2005).
- 139 Schanda, P. *et al.* Protein folding and unfolding studied at atomic resolution by fast two-dimensional NMR spectroscopy. *Proceedings of the National Academy of Sciences* **104**, 11257, doi:10.1073/pnas.0702069104 (2007).

- 140 Amero, C. *et al.* Fast Two-Dimensional NMR Spectroscopy of High Molecular Weight Protein Assemblies. *J. Am. Chem. Soc.* **131**, 3448-3449, doi:10.1021/ja809880p (2009).
- 141 Zimm, B. H. The Scattering of Light and the Radial Distribution Function of High Polymer Solutions. *The Journal of Chemical Physics* **16**, 1093-1099, doi:10.1063/1.1746738 (1948).
- 142 Delaglio, F. *et al.* NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277-293, doi:10.1007/BF00197809 (1995).
- 143 Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins: Structure, Function, and Bioinformatics* **59**, 687-696, doi:https://doi.org/10.1002/prot.20449 (2005).
- 144 Baker, N. A. *et al.* Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences* **98**, 10037, doi:10.1073/pnas.181342398 (2001).
- 145 Zacharias, D. A. *et al.* Partitioning of Lipid-Modified Monomeric GFPs into Membrane Microdomains of Live Cells. *Science* **296**, 913-916, doi:10.1126/science.1068539 (2002).
- 146 Cramer, A. *et al.* Improved Green Fluorescent Protein by Molecular Evolution Using DNA Shuffling. *Nat. Biotechnol.* **14**, 315-319, doi:10.1038/nbt0396-315 (1996).
- 147 Pedelacq, J. D. *et al.* Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24**, 79-88, doi:10.1038/nbt1172 (2006).
- 148 van den Ent, F. *et al.* RF cloning: A restriction-free method for inserting target genes into plasmids. *J. Biochem. Biophys. Methods* **67**, 67-74, doi:https://doi.org/10.1016/j.jbbm.2005.12.008 (2006).
- 149 Klock, H. E. *et al.* in *High Throughput Protein Expression and Purification: Methods and Protocols* (ed Sharon A. Doyle) 91-103 (Humana Press, 2009).
- 150 Velyvis, A. *et al.* An economical method for production of (2)H, (13)CH₃-threonine for solution NMR studies of large protein complexes: application to the 670 kDa proteasome. *PLoS One* **7**, e43725, doi:10.1371/journal.pone.0043725 (2012).
- 151 Noor, S. S. M. *et al.* Purification of Recombinant Green Fluorescent Protein from Escherichia Coli Using Hydrophobic Interaction Chromatography. *Journal of Liquid Chromatography & Related Technologies* **37**, 1873-1884, doi:10.1080/10826076.2013.825847 (2014).
- 152 Kay, L. E. *et al.* Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance (1969)* **89**, 496-514, doi:https://doi.org/10.1016/0022-2364(90)90333-5 (1990).
- 153 Ikura, M. *et al.* A novel approach for sequential assignment of proton, carbon-13, and nitrogen-15 spectra of larger proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry* **29**, 4659-4667, doi:10.1021/bi00471a022 (1990).
- 154 Grzesiek, S. *et al.* Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *Journal of Magnetic Resonance (1969)* **96**, 432-440, doi:https://doi.org/10.1016/0022-2364(92)90099-S (1992).
- 155 Clubb, R. T. *et al.* A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue ¹HN, ¹⁵N, and ¹³C' chemical shifts in ¹⁵N - ¹³C-labelled proteins. *Journal of Magnetic Resonance (1969)* **97**, 213-217, doi:https://doi.org/10.1016/0022-2364(92)90252-3 (1992).
- 156 Wittekind, M. *et al.* HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha- and Beta-Carbon Resonances in Proteins. *Journal of Magnetic Resonance, Series B* **101**, 201-205, doi:https://doi.org/10.1006/jmrb.1993.1033 (1993).
- 157 Eletsky, A. *et al.* TROSY NMR with partially deuterated proteins. *J. Biomol. NMR* **20**, 177-180, doi:10.1023/A:1011265430149 (2001).
- 158 Vuister, G. W. *et al.* Increased Resolution and Improved Spectral Quality in Four-Dimensional ¹³C/¹³C-Separated HMQC-NOESY-HMQC Spectra Using Pulsed Field

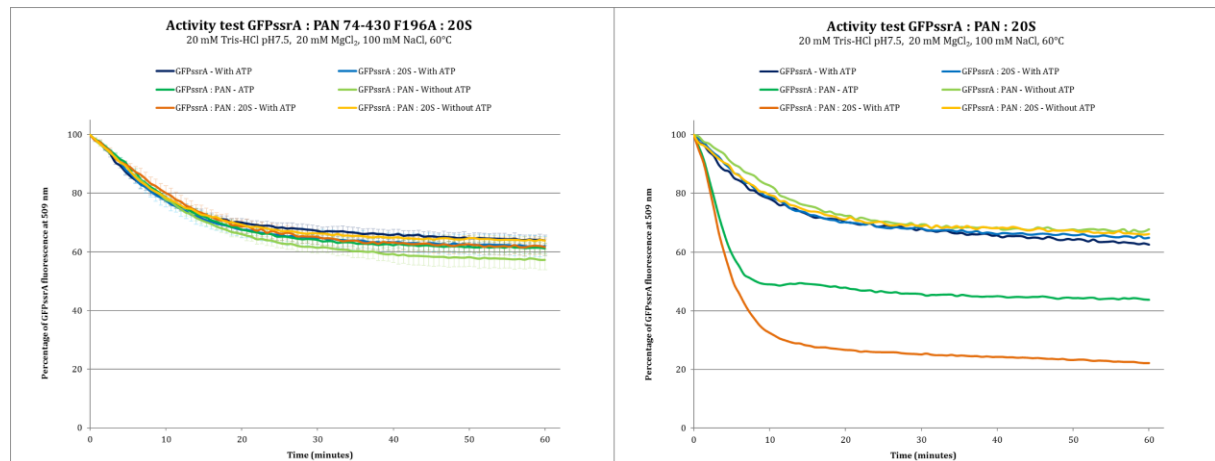
- Gradients. *Journal of Magnetic Resonance, Series B* **101**, 210-213, doi:https://doi.org/10.1006/jmrb.1993.1035 (1993).
- 159 Clore, G. M. *et al.* Four-dimensional carbon-13/carbon-13-edited nuclear Overhauser enhancement spectroscopy of a protein in solution: application to interleukin 1.β. *Biochemistry* **30**, 12-18, doi:10.1021/bi00215a002 (1991).
- 160 Khan, F. *et al.* Letter to the Editor: ¹H, ¹⁵N and ¹³C backbone assignment of the Green Fluorescent Protein (GFP). *J. Biomol. NMR* **26**, 281-282, doi:10.1023/A:1023817001154 (2003).
- 161 Logan, T. M. *et al.* Side chain and backbone assignments in isotopically labeled proteins from two heteronuclear triple resonance experiments. *FEBS Lett.* **314**, 413-418, doi:https://doi.org/10.1016/0014-5793(92)81517-P (1992).
- 162 Montelione, G. T. *et al.* An efficient triple resonance experiment using carbon-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. *J. Am. Chem. Soc.* **114**, 10974-10975, doi:10.1021/ja00053a051 (1992).
- 163 Bax, A. *et al.* ¹H - ¹H correlation via isotropic mixing of ¹³C magnetization, a new three-dimensional approach for assigning ¹H and ¹³C spectra of ¹³C-enriched proteins. *Journal of Magnetic Resonance (1969)* **88**, 425-431, doi:https://doi.org/10.1016/0022-2364(90)90202-K (1990).
- 164 Kay, L. E. *et al.* A Gradient-Enhanced HCCH-TOCSY Experiment for Recording Side-Chain ¹H and ¹³C Correlations in H₂O Samples of Proteins. *Journal of Magnetic Resonance, Series B* **101**, 333-337, doi:https://doi.org/10.1006/jmrb.1993.1053 (1993).
- 165 Graziadei, A. *et al.* The guide sRNA sequence determines the activity level of box C/D RNPs. *eLife* **9**, e50027, doi:10.7554/eLife.50027 (2020).
- 166 Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296-W303, doi:10.1093/nar/gky427 (2018).
- 167 Cheung, M.-S. *et al.* DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *J. Magn. Reson.* **202**, 223-233, doi:https://doi.org/10.1016/j.jmr.2009.11.008 (2010).
- 168 Caillat, C. *et al.* Asymmetric ring structure of Vps4 required for ESCRT-III disassembly. *Nature communications* **6**, 8781, doi:10.1038/ncomms9781 (2015).
- 169 Prunetti, L. *et al.* Structural and biochemical properties of an extreme 'salt-loving' proteasome activating nucleotidase from the archaeon *Haloferax volcanii*. *Extremophiles* **18**, 283-293, doi:10.1007/s00792-013-0615-8 (2014).
- 170 Majhi, P. R. *et al.* Electrostatically Driven Protein Aggregation: β-Lactoglobulin at Low Ionic Strength. *Langmuir* **22**, 9150-9159, doi:10.1021/la053528w (2006).
- 171 Voehler, M. W. *et al.* Performance of cryogenic probes as a function of ionic strength and sample tube geometry. *J. Magn. Reson.* **183**, 102-109, doi:10.1016/j.jmr.2006.08.002 (2006).
- 172 Vinogradova, O. *et al.* in *NMR of Proteins and Small Biomolecules* (ed Guang Zhu) 35-45 (Springer Berlin Heidelberg, 2012).
- 173 Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Progress in nuclear magnetic resonance spectroscopy* **73**, 1-16, doi:https://doi.org/10.1016/j.pnmrs.2013.02.001 (2013).
- 174 Medalia, N. *et al.* Functional and structural characterization of the *Methanosarcina mazei* proteasome and PAN complexes. *J. Struct. Biol.* **156**, 84-92, doi:10.1016/j.jsb.2006.03.015 (2006).
- 175 Martin, A. *et al.* Pore loops of the AAA+ ClpX machine grip substrates to drive translocation and unfolding. *Nat. Struct. Mol. Biol.* **15**, 1147-1151, doi:10.1038/nsmb.1503 (2008).
- 176 Ukmar-Godec, T. *et al.* Proteasomal degradation of the intrinsically disordered protein tau at single-residue resolution. *Science Advances* **6**, eaba3916, doi:10.1126/sciadv.aba3916 (2020).

- 177 Martin, A. *et al.* Protein unfolding by a AAA+ protease is dependent on ATP-hydrolysis rates and substrate energy landscapes. *Nat. Struct. Mol. Biol.* **15**, 139-145, doi:10.1038/nsmb.1380 (2008).
- 178 Kim, Y.-I. *et al.* Dynamics of Substrate Denaturation and Translocation by the ClpXP Degradation Machine. *Mol. Cell* **5**, 639-648, doi:https://doi.org/10.1016/S1097-2765(00)80243-9 (2000).
- 179 Balleza, E. *et al.* Systematic characterization of maturation time of fluorescent proteins in living cells. *Nat Methods* **15**, 47-51, doi:10.1038/nmeth.4509 (2018).
- 180 Shaner, N. C. *et al.* A guide to choosing fluorescent proteins. *Nat. Methods* **2**, 905-909, doi:http://www.nature.com/nmeth/journal/v2/n12/supinfo/nmeth819_S1.html (2005).
- 181 Piston, D. W. *et al.* Fluorescent protein FRET: the good, the bad and the ugly. *Trends Biochem. Sci.* **32**, 407-414, doi:10.1016/j.tibs.2007.08.003 (2007).
- 182 Cormack, B. P. *et al.* FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33-38, doi:https://doi.org/10.1016/0378-1119(95)00685-0 (1996).
- 183 Heim, R. *et al.* Improved green fluorescence. *Nature* **373**, 663-664, doi:10.1038/373663b0 (1995).
- 184 Kisselev, A. F. *et al.* The Sizes of Peptides Generated from Protein by Mammalian 26 and 20 S Proteasomes: IMPLICATIONS FOR UNDERSTANDING THE DEGRADATIVE MECHANISM AND ANTIGEN PRESENTATION. *J. Biol. Chem.* **274**, 3363-3371 (1999).
- 185 Georgescu, J. *et al.* Letter to the Editor: Backbone HN, N, C α and C β assignment of the GFPuv mutant. *J. Biomol. NMR* **25**, 161-162, doi:10.1023/A:1022296413190 (2003).
- 186 Hsu, S. T. *et al.* ^1H , ^{15}N and ^{13}C assignments of yellow fluorescent protein (YFP) Venus. *Biomolecular NMR assignments* **3**, 67-72, doi:10.1007/s12104-009-9143-y (2009).
- 187 Mas, G. *et al.* Structural investigation of a chaperonin in action reveals how nucleotide binding regulates the functional cycle. *Science Advances* **4**, eaau4196, doi:10.1126/sciadv.aau4196 (2018).
- 188 Schütz, A. K. *et al.* Exploiting conformational plasticity in the AAA+ protein VCP/p97 to modify function. *Proceedings of the National Academy of Sciences* **114**, E6822, doi:10.1073/pnas.1707974114 (2017).
- 189 Huang, R. *et al.* Unfolding the mechanism of the AAA+ unfoldase VAT by a combined cryo-EM, solution NMR study. *PNAS* **113**, E4190-E4199, doi:10.1073/pnas.1603980113 (2016).
- 190 Peth, A. *et al.* ATP-Dependent Steps in the Binding of Ubiquitin Conjugates to the 26S Proteasome that Commit to Degradation. *Mol. Cell* **40**, 671-681, doi:https://doi.org/10.1016/j.molcel.2010.11.002 (2010).
- 191 Snoberger, A. *et al.* The Proteasomal ATPases Use a Slow but Highly Processive Strategy to Unfold Proteins. *Frontiers in molecular biosciences* **4**, 18, doi:10.3389/fmolb.2017.00018 (2017).
- 192 Huang, J. R. *et al.* Stable intermediate states and high energy barriers in the unfolding of GFP. *J. Mol. Biol.* **370**, 356-371, doi:10.1016/j.jmb.2007.04.039 (2007).
- 193 Bertz, M. *et al.* Navigating the Folding Energy Landscape of Green Fluorescent Protein. *Angew. Chem. Int. Ed.* **47**, 8192-8195, doi:https://doi.org/10.1002/anie.200802987 (2008).
- 194 Schlierf, M. *et al.* Complex Unfolding Kinetics of Single-Domain Proteins in the Presence of Force. *Biophys. J.* **99**, 1620-1627, doi:https://doi.org/10.1016/j.bpj.2010.06.039 (2010).
- 195 Weber-Ban, E. U. *et al.* Global unfolding of a substrate protein by the Hsp100 chaperone ClpA. *Nature* **401**, 90-93, doi:10.1038/43481 (1999).
- 196 Blythe, E. E. *et al.* Ubiquitin- and ATP-dependent unfoldase activity of P97/VCP•NPLOC4•UFD1L is enhanced by a mutation that causes multisystem

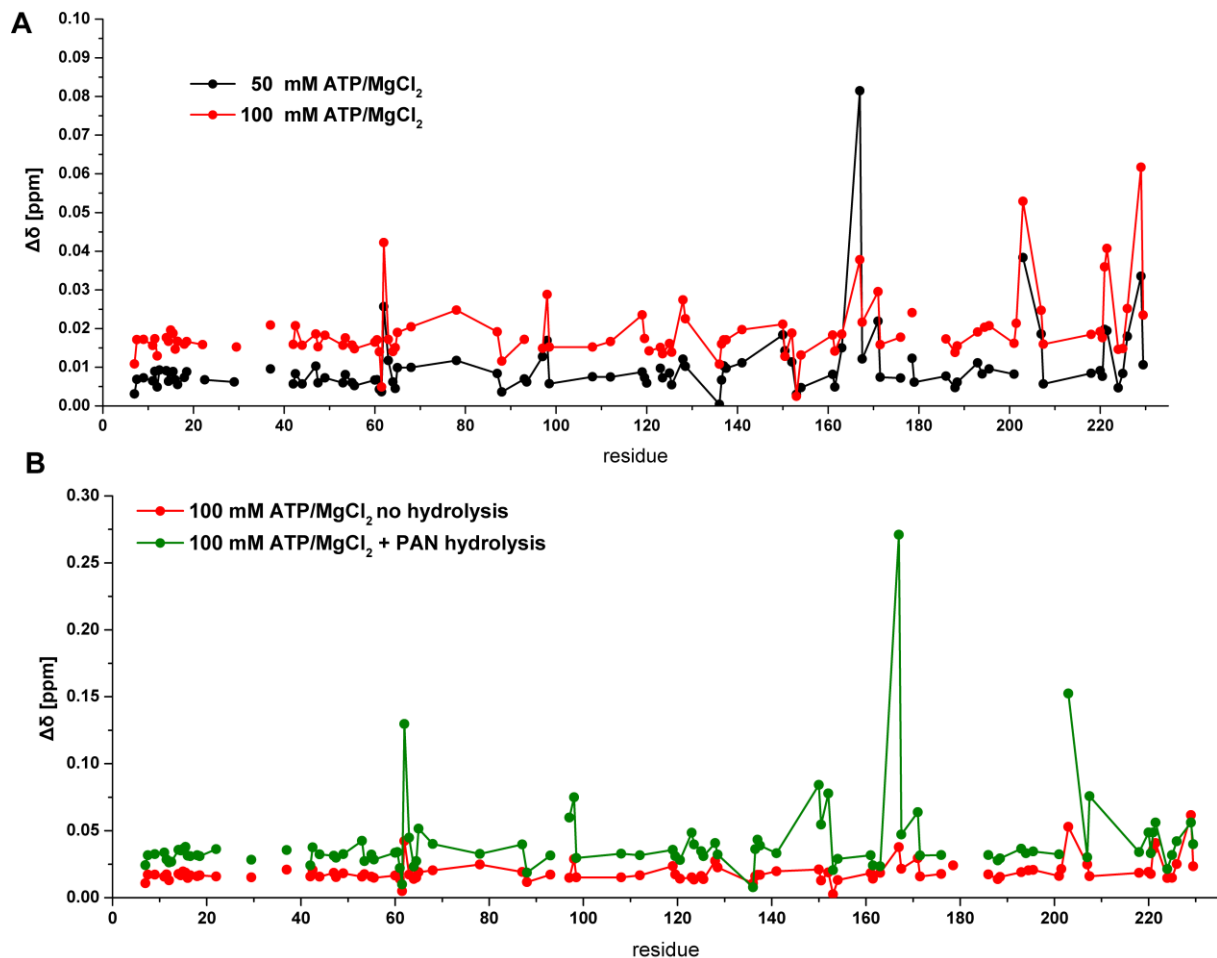
- proteinopathy. *Proceedings of the National Academy of Sciences* **114**, E4380, doi:10.1073/pnas.1706205114 (2017).
- 197 Kühn, T. *et al.* Monitoring the Kinetics of Ion-Dependent Protein Folding by Time-Resolved NMR Spectroscopy at Atomic Resolution. *J. Am. Chem. Soc.* **122**, 6169-6174, doi:10.1021/ja994212b (2000).
- 198 Fürtig, B. *et al.* Conformational Dynamics of Bistable RNAs Studied by Time-Resolved NMR Spectroscopy. *J. Am. Chem. Soc.* **129**, 16222-16229, doi:10.1021/ja076739r (2007).

Appendix

Additional figures



Appendix Figure 1 Activity assay of PAN(1-73) F196A The truncated PAN construct PAN (Δ 1-73) (left panel) does not show activity in unfolding of the model substrate GFP-ssrA compared to the wildtype protein (right panel). Activity assays for this construct have been performed by Dr. Emilie Mahieu, IBS Grenoble, France.



Appendix Figure 2 Chemical shift changes in frGFP-ssrA as a response to ATP/MgCl₂. In the presence of ATP/MgCl₂ peaks of residues in the region of β_{7/8/10} undergo chemical shift changes, which are further increased upon ATP-hydrolysis dependent unfolding by PAN. The charts display GFP-ssrA CSPs at respective conditions calculated versus peak positions in GFP-ssrA without added ATP/MgCl₂. Peak positions in PAN-time-courses were obtained by combining SOFAST-spectra of time-course experiments from the end of the experiment, when the exponential phase has finished, representing the steady state-state. A modified version of this figure will appear in KRÜGER G. et al 2021 (manuscript in preparation).

Amino acid sequences

Listed here are the amino acid sequences of all full-length proteins prepared in this thesis.

M jannaschii full-length PAN with C-terminal His₆-tag

MVFEEFISTELKKEKAFTEEFKEEKEINDNSNLKNDLLKEELQEKARIAELESRIKLELEKKE
LERENLQLMKENEILRRELD RMRVPPLIVGT VVDKVG ERKVVVKSSTGPSFLVNVSHFVNP
DLAPGKRVCLNQQTLTVVDVLPENKDYRAKAMEVDERPNVRYEDIGGLEKQMQEIREVVEL
PLKHPELF EKVGIEPPK GILLYGPPGTGK TLLAKAVATETNATFIRVVGSELVKKFIGEGASLV
KDIFKLAKEKAPSIIFIDEIDAIAAKRTDALTGGDREVQRTLMQLLAEMDGF DARGDVKIIGATN
RPDILDPAILRPGRFDRIIEVPAPDEKGRLEILKIHTRKMNLAEDVNLEEIAKMTEGCVGAE LK
AICTEAGMNAIRELRDYVTMDDFRKAVEKIMEKKKVKVKEPAHLVDVLYRLEHHHHHH

M jannaschii proteasome core particle alpha subunit with N-terminal His6-tag

MGHHHHHPPSAYDRAITVFSPEGRLYQVEYAREAVRRGTTAIGIACKDGVVLAVDRRITSK
 LVKIRSIEKIFQIDDHVAATSGLVADARVLIDRARLEAQIYRLTYGEEISIEMLAKKICDIKQAY
 TQHGGVRPFGVSLIAGIDKNEARLFETDPSGALIEYKATAIGSGRPVVMELLEKEYRDDITL
 DEGLELAITALKANEDIKPENVDVCIITVKDAQFKKIPVEEIKKLIKVKKKLNEENKKEEENR
 EETKEKQEE

M jannaschii proteasome core particle beta subunit

MDVMKGTTTVGLICDDAVILATDKRASLGNLVADKEAKKLYKIDDYIAMTIAGSVGDAQAIVR
 LLIAEAKLYKMRTGRNIPPLACATLLSNILHSSRMFPFLTQIIIGGYDLLEGAKLFSLDPLGGMN
 EEKFTTATGSGSPIAYGVLEAGYDRDMSVEEGIKLALNALKSAMERDTFSGNGISLAVITKDG
 VKIFEDEEIEKILDSMKAKPKKKTTRSRRSK

GFPuv-ssrA

MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTCLKFICTTGKLPVPWPTLVT
 TFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRI
 ELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQKNGIKANFKIRHNIEDGSVQLADHYQQNT
 PIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMLLEFVTAAGITHGMDELAANDENYALAA

frgFP-ssrA

MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTCLKFICTTGKLPVPWPTLVT
 TLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRI
 ELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQKNGIKANFKIRHNIEDGSVQLADHYQQNT
 PIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMLLEFVTAAGITHGMDELAANDENYALAA

Resonance assignments

Appendix Table 1 Backbone resonance assignment of the PAN CC-domain (res. 1-89)

Residue index	Amino acid	H ^N (ppm)	N (ppm)	C' (ppm)	C ^α (ppm)	C ^β (ppm)
1	Met	8.38	120.75	175.76	55.65	33.11
2	Val	7.93	121.17	175.54	62.11	33.09
3	Phe	8.2	123.68	175.58	57.78	39.74
4	Glu	8.27	122.25	176.06	57.79	30.48
5	Glu	8.17	121.22	176.16	56.9	30.55
6	Phe	8.12	120.78	175.54	57.96	39.54
7	Ile	7.89	122.21	175.97	61.14	38.99
8	Ser	8.17	119.04	175.24	58.5	63.9
9	Thr	8.11	116.01	174.94	62.28	69.85
10	Glu	8.24	122.29	176.66	56.87	30.22
11	Leu	7.98	122.29	175.66	55.63	42.28
12	Lys	8.04	121.35	176.54	62.15	-
13	Lys	8.06	121.92	176.51	56.62	33.12
14	Glu	8.26	121.67	176.32	56.65	30.51
15	Lys	8.22	122.55	176.33	-	-
16	Lys	8.17	122.74	175.99	56.37	33.2

Residue index	Amino acid	H ^N (ppm)	N (ppm)	C' (ppm)	C ^α (ppm)	C ^β (ppm)
17	Ala	8.16	124.85	177.26	52.41	19.48
18	Phe	8.08	119.04	175.82	57.82	39.58
19	Thr	7.9	114.84	174.22	61.79	69.99
20	Glu	8.24	122.68	176.19	56.77	30.44
21	Glu	8.22	121.39	176.03	56.87	30.55
22	Phe	8.13	121.12	175.34	-	39.43
23	Lys	7.97	123.43	175.87	56.11	33.52
24	Glu	8.26	122.19	176.37	56.68	-
25	Glu	8.36	122.5	176.34	56.75	30.39
26	Lys	8.19	121.93	176.38	56.39	33.28
27	Glu	8.32	121.84	176.41	56.69	30.33
28	Ile	8.05	121.39	175.92	61.25	38.84
29	Asn	8.359	122.39	174.78	53.27	39.32
30	Asp	8.22	121.21	176.21	54.49	41.31
31	Asn	8.35	119.46	175.87	53.58	38.82
32	Ser	8.29	116.05	174.71	59.63	63.67
33	Asn	8.33	120.02	175.25	53.71	38.77
34	Leu	7.91	121.53	177.4	55.71	42.27
35	Lys	8.06	120.65	176.52	56.7	32.79
36	Asn	8.19	119.23	175.36	53.73	39.02
37	Asp	8.29	120.95	176.67	55.17	41.03
38	Leu	8	121.38	177.9	56.19	42.24
39	Leu	7.97	121.19	177.77	55.79	42.11
40	Lys	7.97	120.79	-	57.76	32.79
41	Glu	8.31	121.01	177.54	58.03	30
42	Glu	8.34	121.02	177.62	57.89	30.02
43	Leu	8.11	121.14	178.72	56.56	42.07
44	Gln	8.25	120.35	177.77	57.71	28.89
45	Glu	8.45	121.45	177.96	59.06	29.61
46	Lys	8.11	118.91	178.81	59.08	32.43
47	Ala	7.98	122.81	179.71	54.55	18.37
48	Arg	8.04	120.19	178.55	58.78	30.12
49	Ile	8.14	119.41	177.17	65.9	37.79
50	Ala	7.76	120.73	181.05	55.22	18.06
51	Glu	7.94	119.5	-	59.48	29.54
52	Leu	-	-	179.31	58.34	42.52
53	Glu	8.82	118.14	180.12	60.49	29.45
54	Ser	8.04	114.92	176.82	61.86	62.93
55	Arg	8.03	123.6	178.76	59.4	33.47
56	Ile	8.14	120.08	176.83	66.14	37.96
57	Leu	7.59	119	179.45	-	41.62
58	Lys	7.7	118.42	179.04	59.68	32.57
59	Leu	8.3	120.19	179.73	58.2	38.77
60	Glu	8.86	119.18	180.41	60.26	29.53
61	Leu	7.97	121.21	180.1	58.41	41.55
62	Glu	8.47	120.76	180.27	59.58	29.88
63	Lys	8.82	120.17	177.78	60.7	32.33
64	Lys	7.91	118.66	179.76	59.35	32.18
65	Glu	7.92	119.66	179.33	59.35	29.27
66	Leu	7.93	121.19	179.55	58.26	43.7
67	Glu	8.72	120.24	179.22	60.06	30.1
68	Arg	7.97	120.14	179.39	59.89	29.99
69	Glu	8.18	120.81	178.4	59.55	30.38

Residue index	Amino acid	H ^N (ppm)	N (ppm)	C' (ppm)	C ^α (ppm)	C ^β (ppm)
70	Asn	8.69	120.03	177.26	57.34	38.93
71	Leu	8.04	119.89	179.72	58.59	41.89
72	Gln	7.8	119.03	179.16	59.2	28.1
73	Leu	8.53	121.13	179.57	58.14	43.74
74	Met	8.68	119.65	179.22	59.46	33.07
75	Lys	8.02	120.71	179.09	59.33	32.2
76	Glu	8.24	119.91	178.29	59.68	29.72
77	Asn	8.63	118.98	177.05	57.34	39
78	Glu	7.9	119.1	179.05	59.75	29.75
79	Ile	7.8	120.53	178.92	65.34	38.11
80	Leu	8.39	120.59	179.02	58.28	42.95
81	Arg	8.42	117.67	179.08	60.41	30.35
82	Arg	7.99	119.02	179.39	59.36	30.2
83	Glu	8.14	120.51	179.04	59.24	29.65
84	Leu	8.23	120.32	179.19	57.85	42.12
85	Asp	8.05	118.4	178.58	57.07	40.81
86	Arg	7.67	117.88	177.54	57.96	30.45
87	Met	7.72	116.82	176.41	56.14	33.42
88	Arg	7.66	120.25	175.28	56.79	30.77
89	Val	7.47	124.67	181	63.87	33.18

Appendix Table 2 Backbone resonance assignment of the PAN OB-ring (res. 74-150), conformation A

Residue index	Amino acid	H ^N (ppm)	N (ppm)	C' (ppm)	C ^α (ppm)	C ^β (ppm)
78	Glu	7.59	119.35	174.5	58.6	-
79	Ile	7.59	120.12	178.24	63.94	36.96
80	Leu	7.91	119.69	178.6	57.43	41.25
81	Arg	7.97	118.38	178.51	59.17	29.08
82	Arg	7.72	118.86	179.14	58.55	29.07
83	Glu	7.99	120.26	179.2	58.65	28.83
84	Leu	7.97	119.77	179.47	57.88	41.26
85	Asp	8.09	119.06	179.78	57.1	40.06
86	Arg	7.83	118.99	178.24	58.12	29.11
87	Met	7.55	115.53	176.33	56.56	32.23
88	Arg	7.48	114.5	-	54.76	30.37
89	Val	6.81	119.16	173.52	60.02	31.46
92	Leu	7.64	117.73	177.28	53.08	42.8
93	Ile	8.61	122.87	176.04	61.19	40.04
94	Val	8.19	126.88	176.6	62.07	31.16
95	Gly	8.85	115.26	171.44	44.27	-
96	Thr	8.41	113.9	174.64	60.47	71.16
97	Val	8.93	126.24	175.7	64.5	30.98
98	Val	9.26	129.85	174.61	63.57	32.15
100	Lys	8.87	123.67	175.13	55.1	32.92
101	Val	7.69	121.36	175.17	61.28	31.89
102	Gly	7.98	110.34	173.22	44.38	-
105	Lys	7.04	117.46	175.64	55.16	35.42
106	Val	8.53	114.69	174.05	58.52	35.02
107	Val	8.67	122.48	176.32	61.07	31.6
108	Val	9.47	124.96	173.54	58.23	34.68
109	Lys	9.06	127.17	177.29	54.53	31.7

Residue index	Amino acid	H ^N (ppm)	N (ppm)	C' (ppm)	C ^α (ppm)	C ^β (ppm)
110	Ser	9.04	123.1	177.28	57.72	62.79
115	Ser	8.16	112.74	172.49	56.86	64.32
116	Phe	8.82	119.5	174.61	55.32	44.19
117	Leu	9.07	125.64	175.9	54.5	42.88
118	Val	9.02	118.87	174.86	58.93	35.2
119	Asn	8.14	116.36	174.83	51.95	40.11
120	Val	7.63	118.85	175.6	61.07	32.86
121	Ser	8.12	121.74	175.38	58.46	63.73
124	Val	7.28	120.28	174.03	60.54	32.74
125	Asn	8.79	127.08	175.18	49.88	38.19
127	Asp	7.93	117.01	177.19	55.85	39.5
128	Asp	7.84	118.22	176.2	54.86	40.6
129	Leu	7.22	118.3	173.89	53.58	40.78
130	Ala	6.66	122.55	173.69	49.52	18.87
132	Gly	8.99	112.91	174.65	44.59	-
133	Lys	7.84	121.06	176.19	55.58	31.76
134	Arg	8.76	124.61	176.14	55.86	30.85
135	Val	8.78	114.03	174.06	57.78	35.03
136	Cys	8.08	116.05	173.76	56.51	29.94
137	Leu	9.91	124.65	176.97	52.16	41.82
138	Asn	8.39	122.43	176.2	54.43	40.81
139	Gln	9.35	124.32	175.57	59.14	28.73
140	Gln	7.99	115.77	177.99	57.99	28.76
141	Thr	8.25	107.92	173.59	60.92	69.81
142	Leu	7.7	116.74	172.97	56.24	37.09
143	Thr	6.96	108.01	175.57	62.01	69.46
144	Val	7.94	123.4	177.22	63.12	31.02
145	Val	8.65	120.84	173.39	60.78	33.03
146	Asp	7.69	117.33	174.4	53.05	43.29
147	Val	9.1	119.22	174.49	60.35	33.86
148	Leu	7.88	125.28	174.89	52.22	40.61
150	Glu	7.73	124.89	181.15	57.79	30.56

Appendix Table 3 Backbone resonance assignment of the PAN OB-ring (res. 74-150), conformation B

Residue index	Amino acid	H ^N (ppm)	N (ppm)	C' (ppm)	C ^α (ppm)	C ^β (ppm)
79	Ile	7.66	120.32	178.42	64.13	37
80	Leu	7.91	120.36	178.83	57.57	41.51
81	Arg	8.01	117.7	178.78	59.54	29.08
82	Arg	7.84	118.99	-	58.81	29.05
86	Arg	7.8	119.48	178.64	58.82	29.73
87	Met	7.78	116.24	175.98	56.62	33.18
88	Arg	7.37	116.42	176.08	55.98	30.23
89	Val	6.99	122.45	174.03	60.47	31.71
92	Leu	8.61	120.47	176.76	52.55	43.71
93	Ile	8.53	122.83	175.9	61.51	40.06
94	Val	8.22	127.15	176.59	62.03	31.08
96	Thr	8.39	114.17	174.6	60.51	71.14
97	Val	8.97	126.33	175.73	64.43	30.95
98	Val	9.28	129.59	174.62	63.51	32.16
99	Asp	7.33	114.01	174.84	53.14	44.05

101	Val	7.57	121.75	174.83	61.26	32.09
102	Gly	7.95	110.09	173.24	44.37	-
105	Lys	7.07	118.21	175.64	55.36	35.43
106	Val	8.48	114.93	173.81	58.46	34.91
107	Val	8.72	122.63	176.18	61	31.69
108	Val	9.49	125.78	173.73	58.54	34.76
109	Lys	9.06	127.58	177.08	54.57	31.66
110	Ser	9.07	123.5	177.13	57.77	62.8
115	Ser	8.1	112.63	172.55	56.87	64.36
116	Phe	8.86	119.71	174.64	55.27	44.19
117	Leu	9.12	125.63	175.88	54.49	42.86
118	Val	9.05	119.04	174.86	58.97	35.25
119	Asn	8.21	116.42	-	-	-
124	Val	7.26	120.83	174.31	60.77	32.49
125	Asn	8.82	127.29	175.19	49.9	38.12
128	Asp	7.86	118.21	176.22	54.84	40.59
129	Leu	7.2	118.74	173.84	53.56	40.65
130	Ala	6.63	122.6	173.73	49.51	18.84
135	Val	8.79	113.82	174.16	57.75	35.01
136	Cys	8.01	115.63	174	56.35	30.14
137	Leu	9.91	124.4	176.94	52.43	42.12
138	Asn	8.24	122.11	176.33	53.95	40.3
139	Gln	9.26	125.09	175.88	59.15	28.79
142	Leu	7.9	121.79	174.76	55.23	37.96
143	Thr	7.18	111.27	174.49	60.67	71.13
144	Val	8.82	125.53	175.47	-	31.75
145	Val	8.61	122.46	174.31	61.42	32.47
146	Asp	7.57	117.07	-	-	-
148	Leu	8.02	125.84	174.93	52.32	40.41
150	Glu	7.59	124.39	180.98	57.69	30.58

Appendix Table 4 Backbone resonance assignment of the PAN OB-ring (res. 74-150), conformation C

Residue index	Amino acid	H ^N (ppm)	N (ppm)	C' (ppm)	C ^α	C ^β
79	Ile	7.73	120.42	176.98	62.5	37.52
80	Leu	7.76	122.8	-	56.06	-
89	Val	6.9	121.39	173.58	60.19	31.84
94	Val	8.23	123.32	176.52	60.68	-
95	Gly	8.59	112.12	171.45	44.18	-
96	Thr	8.41	115.41	174.53	60.88	70.54
97	Val	9.07	126.47	176	64.02	30.82
98	Val	9.26	128.51	174.86	62.87	32.36
99	Asp	7.3	115.05	174.56	53.06	43.6
101	Val	7.59	121.28	174.89	61.36	31.86
102	Gly	7.93	109.66	173.09	44.26	-
105	Lys	7.04	117.33	175.67	55.18	35.44
106	Val	8.56	115.73	173.19	58.61	34.74
107	Val	8.47	122.98	176.19	61.14	32.04
108	Val	9.06	126.22	173.36	58.96	34.54
109	Lys	8.77	124.9	176.53	54.59	33.83
117	Leu	9.05	125.07	176.41	54.51	42.69
118	Val	8.92	120.82	174.68	58.59	35.51
119	Asn	8.12	116.77	175.46	52.22	39.72
120	Val	8.1	120.7	175.86	61.69	31.89

Residue index	Amino acid	H ^N (ppm)	N (ppm)	C' (ppm)	C ^α	C ^β
121	Ser	7.99	122.73	-	58.86	64.81
124	Val	7.27	120.51	174.19	60.71	32.54
125	Asn	8.7	126.32	-	-	-
129	Leu	7.22	118.1	173.9	53.59	40.92
135	Val	8.78	113.91	174.12	57.72	35.03
136	Cys	8.04	115.93	173.95	56.22	30.04
137	Leu	9.86	124.28	176.8	52.48	-
138	Asn	8.22	121.97	176.84	53.55	39.69
139	Gln	9.07	125.37	175.93	58.84	28.63
142	Leu	7.79	118.01	173.43	55.89	-
143	Thr	6.9	110.34	175.48	61.8	70.22
144	Val	8.32	123.96	176.14	63.22	31.02
145	Val	8.61	122.07	-	61.07	32.73
146	Asp	7.62	117.22	-	53.01	-

Appendix Table 5 Backbone resonance assignment of the PAN OB-ring (res. 74-150), conformation D

Residue index	Amino acid	H ^N (ppm)	N (ppm)	C' (ppm)	C ^α (ppm)	C ^β (ppm)
79	Ile	7.76	120.64	176.57	61.78	37.43
80	Leu	7.81	123.84	-	55.44	-
92	Leu	8.62	120.42	176.61	52.53	43.69
93	Ile	8.68	123.09	175.92	61.38	39.97
94	Val	8.3	127.25	176.59	62.2	31.1
101	Val	7.73	121.61	175.22	59.29	31.89
105	Lys	7.08	117.87	175.77	55.39	35.46
108	Val	9.47	125.59	173.7	58.5	34.78
109	Lys	9.05	127.43	177.08	54.56	31.66
110	Ser	9.06	123.57	177.05	57.75	62.8
119	Asn	8.1	116.3	174.83	51.97	40.08
120	Val	7.7	119.41	175.72	61.24	32.6
121	Ser	7.99	121.75	175.13	58.71	63.94
123	Phe	7.18	116.5	175.32	58.37	38.78
124	Val	7.35	118.68	173.76	60.09	33.1
125	Asn	8.81	126.57	175.29	49.78	38.25
128	Asp	7.82	118.24	-	-	-
133	Lys	7.8	120.68	175.86	55.33	31.57
134	Arg	8.89	123.62	176.06	55.25	30.83
135	Val	8.83	113.02	173.87	57.73	34.71
136	Cys	8.21	118.19	174.02	56.3	28.52
137	Leu	9.57	126.49	-	53.04	-
142	Leu	7.73	117.62	173.33	56.05	38.03
143	Thr	6.97	110.43	175.56	62	70.09
144	Val	8.37	124.12	176.59	63.29	31.12
145	Val	8.62	120.52	173.86	60.45	32.91
146	Asp	7.67	117.76	174.31	53.08	43.49
147	Val	9.02	119.93	174.67	60.48	33.78
148	Leu	8.05	125.97	174.99	52.37	40.4

Appendix Table 6 Backbone resonance assignment of the PAN OB-ring (res. 74-150), conformation E

Residue index	Amino acid	H ^N (ppm)	N (ppm)	C' (ppm)	C ^α (ppm)	C ^β (ppm)
93	Ile	8.54	122.87	175.91	-	-
94	Val	8.18	123.57	176.55	60.66	-
116	Phe	8.76	119.16	173.58	55.68	41.68
117	Leu	8.32	125.02	176.61	54.64	42.01
118	Val	8.96	119.24	174.52	59.08	34.38
119	Asn	8.18	118.66	175.17	52.29	40.11
120	Val	8.08	120.67	175.42	61.96	32.16
132	Gly	8.97	112.9	174.75	44.58	-
133	Lys	7.82	120.86	176.17	55.57	31.71
134	Arg	8.8	124.26	176.04	55.97	30.9
137	Leu	9.64	126.89	176.92	52.4	42.07
138	Asn	8.28	122.15	176.3	53.97	40.33
139	Gln	9.26	124.97	175.88	59.15	28.8
143	Thr	6.98	110.59	175.57	62.14	70.05
144	Val	8.4	124.28	176.56	63.28	31.03
145	Val	8.59	121.93	174.25	61.04	-
147	Val	8.76	121.01	174.73	60.88	32.95
148	Leu	8.24	127.67	175.53	52.88	40.35

Appendix Table 7 Backbone resonance assignment of the PAN OB-ring (res. 74-150), conformation F

Residue index	Amino acid	H ^N (ppm)	N (ppm)	C' (ppm)	C ^α (ppm)	C ^β (ppm)
93	Ile	8.6	122.82	174.58	59.19	-
94	Val	8.18	123.14	176.55	60.71	32.69
119	Asn	8.16	116.34	-	52.05	40.1
134	Arg	8.93	123.52	-	56.36	-
138	Asn	8.18	121.95	176.51	53.55	39.66
139	Gln	9.09	125.57	-	58.84	28.59
143	Thr	6.92	110.54	175.45	61.98	70.18
144	Val	8.35	124.11	175.54	63.28	30.96
145	Val	8.69	122.01	174.05	61.24	-
147	Val	8.72	121.09	174.81	60.89	32.97
148	Leu	8.22	127.57	-	52.91	40.26

Appendix Table 8 Assignment of methyl group resonances in the PAN OB-ring domain

Methyl group	¹ H (ppm)	¹³ C (ppm)
A79IleCδ1	0.80342	12.66609
A93IleCδ1	0.88782	14.08405
A106ValCy1	0.79881	21.38591
A120ValCy2	0.72269	21.65478
A129LeuCδ1	0.62653	23.66059
A129LeuCδ1	0.62335	23.98847
A135ValCy1	0.5709	18.23452
A137LeuCδ1	0.75724	24.90307
A142LeuCδ1	0.05321	24.47809
A142LeuCδ2	0.46579	23.30171
B92LeuCδ1	0.75816	25.83856
B93IleCδ1	0.90269	14.11697

Methyl group	¹ H (ppm)	¹³ C (ppm)
B97ValCγ1	0.59464	20.88417
B101ValCγ1	0.68325	20.2682
B101ValCγ2	0.67322	20.60949
B129LeuCδ1	0.62754	24.35249
C137LeuCδ1	0.73	24.72694
C137LeuCδ2	0.71795	24.76314
C142LeuCδ1	0.37912	22.49192
D79IleCδ1	0.84653	12.6825
D93IleCδ1	0.9257	14.17591
D124ValCγ1	0.63439	21.19202
D124ValCγ2	0.63986	21.34224
D142LeuCδ1	0.40128	22.56747
D142LeuCδ2	-0.04732	24.01892
E118ValCγ1	0.57065	21.73043
E118ValC γ2	0.57422	21.3061
F93IleCδ1	0.80246	13.37586

Appendix Table 9 Assignments of backbone resonances of the PAN ATPase domain (residues 150-430)

Residue index	Amino acid	H ^N (ppm)	N (ppm)	C ^α (ppm)	C ^β (ppm)
154	Tyr	7.9	120.37	58.28	37.6
155	Arg	7.95	121.09	56.68	29.61
156	Ala	7.82	122.49	52.57	18.44
157	Lys	7.83	117.61	56.08	32.18
158	Ala	7.9	119.19	52.73	18.58
159	Met	7.42	112.74	53.9	35.22
160	Glu	8.77	122.43	54.77	32.01
161	Val	8.46	124.92	61.28	32.73
162	Asp	9.45	128.25	52.64	42.2
163	Glu	8.84	123.25	57.6	29.71
164	Arg	8.41	118.03	53.05	30.12
166	Asn	8.01	115.95	52.47	37.96
167	Val	7.93	122.41	61.67	33.57
168	Arg	8.47	123.59	53.29	30.59
169	Tyr	8.35	120.98	60.84	36.14
170	Glu	8.29	113.42	57.77	28.05
171	Asp	7.79	119.59	-	-
172	Ile	7.37	118.61	60.67	37.87
173	Gly	10.43	117.27	43.57	-
174	Gly	9.02	108.32	46.71	-
175	Leu	8.24	120.17	53.1	42.3
176	Glu	8.32	120.65	61.06	28.6
177	Lys	8.52	118.07	58.92	31.08
178	Gln	8.59	120.8	60.51	25.67
179	Met	8.96	117.94	58.94	32.51
180	Gln	7.82	118.05	58.78	26.97
181	Glu	7.68	118.87	59.51	28.96
182	Ile	8.31	117.96	64.39	36.95
183	Arg	8.58	121.89	60.01	28.96
184	Glu	7.77	118.27	59.26	29.11
185	Val	7.67	106.53	62.35	31.83
186	Val	8.04	116.16	63.98	32.52
187	Glu	7.44	118.22	59.62	29.92

Residue index	Amino acid	H ^N (ppm)	N (ppm)	C ^α (ppm)	C ^β (ppm)
190	Leu	7.04	114.86	56.74	41.96
191	Lys	8.16	116.77	57.6	33.5
192	His	8.39	114.69	52.65	29.26
194	Glu	9.46	118.4	58.13	27.38
195	Leu	7.68	121.4	56.61	40.9
196	Ala	7.83	120.31	54.33	17.23
197	Glu	7.79	115.91	58.1	29.11
198	Lys	7.66	118.89	57.79	32.05
199	Val	7.85	114.23	62.67	31.36
201	Ile	7.49	117.71	59.72	38.91
202	Glu	8.28	126.09	52.93	29.06
205	Lys	9.46	123.17	57.64	32.07
206	Gly	7.28	103.15	44.21	-
207	Ile	8.49	119.71	58.78	-
208	Leu	9.14	127.6	52.56	44.14
209	Leu	9.22	124.98	52.48	42.68
210	Tyr	8.42	118.67	53.96	41.91
211	Gly	8.1	109.36	44.31	-
214	Gly	10.35	112.67	47.07	-
215	Thr	8.2	109.05	61.21	69.61
216	Gly	8.42	109.46	46.22	-
217	Lys	9.73	123.92	61.04	29.34
219	Leu	7.54	121.91	57.69	41.26
220	Leu	8.26	117.49	57.64	41.56
221	Ala	7.34	119.1	54.68	19.1
222	Lys	8.33	116.74	59.55	31.96
223	Ala	8.53	125.85	53.65	17.78
224	Val	7.87	117.91	64.71	30.59
225	Ala	7.73	123.4	54.91	16.89
226	Thr	8.15	115.69	66.48	68.37
227	Glu	8.08	120.43	-	28.32
228	Thr	7.5	105.84	61.4	69.14
229	Asn	7.71	117.21	54.5	36.93
230	Ala	8.6	120.75	54.47	20.07
231	Thr	8.72	118.11	64.18	68.52
232	Phe	8.49	128.69	56.91	40.92
233	Ile	8.36	129.01	59.64	39.17
233	Ile	-	-	-	-
234	Arg	8.8	127.09	54.14	32.02
235	Val	9.27	126.52	59.74	33.87
236	Val	8.72	126.76	61.7	30.75
237	Gly	8.9	116.03	48.05	-
238	Ser	7.95	111.53	61.24	-
239	Glu	7.14	120.55	57.73	29.21
240	Leu	-	-	-	-
240	Leu	7.78	118.63	57.98	41.34
241	Val	7.17	109.96	63.09	31.15
242	Lys	7.45	118.52	57.29	32.59
243	Lys	7.83	116.84	56.83	32.51
244	Phe	7.64	117.32	57.24	38.61
245	Ile	7.41	119.74	62.12	37.1
245	Ile	-	-	-	-
246	Gly	8.46	111.37	45.44	-

Residue index	Amino acid	H ^N (ppm)	N (ppm)	C ^α (ppm)	C ^β (ppm)
247	Glu	8.08	119.28	55.75	29.85
248	Gly	8.34	110.42	47.63	-
249	Ala	8.29	122.01	54.89	-
250	Ser	7.77	113.04	60.51	62.31
251	Leu	8.11	122.99	57.62	40.9
252	Val	8.12	117.17	67.09	30.62
253	Lys	7.64	118.43	60.06	31.15
254	Asp	8.01	120.19	57.37	39.95
255	Ile	8.49	121.46	65.23	37.21
255	Ile	-	-	-	-
256	Phe	7.8	117.75	62.33	38.3
257	Lys	-	-	59.64	31.7
258	Leu	8.05	121.35	57.39	40.88
259	Ala	8.16	118.76	-	18.36
260	Lys	7.71	115.75	58.86	31.71
261	Glu	-	-	58.54	29.07
262	Lys	7.82	116.52	54.13	30.38
263	Ala	6.59	122.11	51.51	17.38
265	Ser	8.61	116.13	57.12	66.66
266	Ile	8.87	121.64	58.27	39.82
266	Ile	-	-	-	-
267	Ile	8.56	125.59	59.95	38.24
267	Ile	-	-	-	-
268	Phe	9.45	130.28	54.57	42.08
269	Ile	8.95	126.71	59.22	36.99
269	Ile	-	-	-	-
270	Asp	8.64	126.47	52.55	42.35
271	Glu	8.08	120.81	56.02	26.68
272	Ile	8.33	117.21	63.99	37.23
273	Asp	8.97	118.13	54.8	39.75
274	Ala	7.93	120.53	54.13	17.46
275	Ile	6.8	107.2	60.49	39.34
276	Ala	7.79	119.35	50.16	19.5
277	Ala	7.27	121.6	52.12	18.77
278	Lys	8.41	121.02	56.47	32.22
279	Arg	8.39	123.52	55.4	30.33
282	Ala	8.02	122.88	52.48	18.41
283	Leu	8.12	119.93	55.47	41.49
284	Thr	8.02	112.03	61.31	69.85
287	Asp	8.1	120.36	54.94	-
288	Arg	8.25	121.25	58.06	29.38
289	Glu	8.42	119.48	58.56	28.42
290	Val	8.01	120.95	65.9	31.04
291	Gln	7.95	119.38	58.62	27.65
292	Arg	8.09	118.76	59.18	29.33
293	Thr	7.92	117.92	67.2	29.33
294	Leu	8.21	122.5	57.98	39.33
295	Met	8.46	115.74	57.99	30.98
296	Gln	7.92	119.62	57.65	27.45
297	Leu	8.1	120.5	58.41	40.63
298	Leu	8.3	116.35	57.59	39.62
299	Ala	7.83	120.28	54.68	16.99
300	Glu	8.08	117.02	57.28	26.4

Residue index	Amino acid	H ^N (ppm)	N (ppm)	C ^α (ppm)	C ^β (ppm)
301	Met	7.88	117.47	59.11	32.08
302	Asp	8.37	119	56.14	39.34
303	Gly	7.45	104.74	45	-
304	Phe	7.38	120.53	56.67	39.49
307	Arg	8.27	114.74	55.37	28.51
308	Gly	7.58	106.81	44.87	-
309	Asp	8.08	115.95	53.14	39.57
310	Val	7.16	116.22	60.63	33.78
311	Lys	-	-	60.66,52.18	33.78
312	Ile	8.19	120.49	59.92	39.45
313	Ile	9.61	126.94	58.97	40.41
314	Gly	9.38	112.67	42.88	-
315	Ala	8.31	122.42	50.2	22.99
316	Thr	8.25	113.12	59.99	69.19
317	Asn	8.87	121.84	53.15	39.54
318	Arg	8.67	119.81	52.65	30.62
320	Asp	8.58	114.5	55.07	39.14
321	Ile	6.86	111.84	59.69	37.93
322	Leu	6.9	118.96	54.18	40.69
323	Asp	8.51	122.9	51.94	41.97
325	Ala	9.21	121.35	54.59	17.98
326	Ile	7.89	109.87	62.09	37.61
327	Leu	7.29	120.93	53.11	40.55
328	Arg	6.65	118.58	54.04	29.6
330	Gly	9.51	114.43	44.74	-
331	Arg	8.02	121.8	55.37	28.07
332	Phe	8.24	121.29	57.4	36.21
333	Asp	7.73	119.05	54.15	42.04
334	Arg	7.48	117.2	53.8	31.62
335	Ile	8.56	125.23	59.96	39.13
336	Ile	9.28	128.21	60.37	40.76
337	Glu	8.31	128.52	55.35	29.63
338	Val	9.18	129.38	59.44	30.77
342	Asp	8	121.22	51.85	39.88
343	Glu	8.44	119.85	61.25	29.3
344	Lys	7.76	117.84	58.68	30.94
345	Gly	8.58	109.55	47.22	-
346	Arg	9.15	119.44	60.38	29.31
347	Leu	7.81	119.63	58.13	40.3
348	Glu	7.92	117.37	59.48	28.98
349	Ile	8.24	118.96	65.95	36.9
350	Leu	8.53	120.17	58	41.47
351	Lys	8.37	120.22	59.96	31.12
352	Ile	7.57	119.72	65.21	37.92
353	His	7.94	115.41	58.91	30.74
354	Thr	7.73	106.88	61.49	70.3
355	Arg	7.24	122.87	59.6	30.02
356	Lys	8.14	117.94	59.88	38.21
359	Leu	8.27	124.22	54.12	41.56
360	Ala	9.18	124.73	51.19	19.69
361	Glu	8.64	118.74	58.28	28.78
362	Asp	7.74	114.08	53.01	39.77
363	Val	7.38	120.28	64.6	30.25

Residue index	Amino acid	H ^N (ppm)	N (ppm)	C ^α (ppm)	C ^β (ppm)
364	Asn	7.55	128.72	50.71	38.49
366	Glu	8.34	119.61	60.04	27.94
367	Glu	7.33	119.52	58.6	28.62
368	Ile	7.2	117.83	62.68	34.56
369	Ala	8.58	123.54	55.18	16.7
370	Lys	7.87	120.05	59.79	31.61
371	Met	7.74	115.97	57.34	34.01
372	Thr	7.06	104.87	59.65	67.65
373	Glu	6.9	121.51	58.56	29
374	Gly	7.71	115.37	44.93	-
375	Cys	8.18	116.93	60.05	28.76
376	Val	8.69	112	59.18	33.08
377	Gly	8.6	111.15	47.82	-
378	Ala	8.83	122.14	55.59	19.02
379	Glu	7.09	117.11	58.16	29.24
380	Leu	7.79	118.45	57.58	39.97
381	Lys	8.06	120.01	59.66	30.74
382	Ala	7.4	121.99	55.03	16.99
383	Ile	8.41	119.83	65.45	37.47
384	Cys	7.75	118.15	64.24	26.28
385	Thr	8.36	115.9	66.04	68.33
386	Glu	8.44	122.55	58.13	27.83
387	Ala	8.37	123.31	55.05	15.35
388	Gly	7.6	101.81	46.62	-
389	Met	8.06	120.58	-	-
390	Asn	8.27	120.19	56.2	37.94
391	Ala	7.2	120.73	55.29	18.16
392	Ile	7.59	117.09	64.29	-
393	Arg	8.25	121.55	59.06	-
394	Glu	7.29	116.67	55.41	28.67
395	Leu	-	-	55.54	37.2
396	Arg	8.03	117.44	55.18	33.93
398	Tyr	6.85	114.65	54.33	39.68
399	Val	8.56	120.41	60.72	33.04
400	Thr	9.75	117.32	59.83	72.16
401	Met	9.42	120.93	57.44	30.24
402	Asp	8.56	118.81	57.7	39.94
403	Asp	7.64	121.24	58.04	39.92
404	Phe	7.67	117.11	62.75	39.61
405	Arg	9.08	119.49	60.72	29.08
406	Ser	8.1	113.07	61.05	62.25
407	Ala	7.48	124.26	55.03	19.93
408	Val	8.54	117.62	66.68	31.25
409	Glu	7.48	116.73	58.96	28.7
410	Lys	-	-	58.53	31.97
411	Ile	7.94	118.2	61.86	35.93
412	Met	8.27	116.85	57	31.47
413	Glu	7.62	118.52	56.85	29.2
414	Glu	7.93	120.67	-	29.78
416	Glu	8.2	121.42	-	-
417	Val	7.93	120.37	62.03	32.22
418	Glu	8.25	123.93	56.14	29.78
419	Val	7.99	121.42	61.94	31.97

Residue index	Amino acid	H ^N (ppm)	N (ppm)	C ^α (ppm)	C ^β (ppm)
420	Lys	8.14	125.29	55.52	32.36
421	Glu	7.88	121.06	53.5	30.36
425	Leu	7.83	122.21	55.08	41.66
426	Asp	8.22	120.54	54.46	40.68
427	Val	7.69	117.78	62.03	31.99
428	Leu	7.9	123.12	55.05	41.62
429	Tyr	7.79	119.6	57.23	37.83
430	Arg	7.56	126.49	30.89	-

Appendix Table 10 Assignment of methyl-group resonances in the PAN ATPase domain

Question marks “?” indicate tentative assignments

Methyl-group	¹ H (ppm)	¹³ C (ppm)
159MetCε	1.85435	17.69315
161Valγ1	0.72719	20.72315
161Valγ2	0.7683	21.5375
167Valγ1	0.63703	21.35502
167Valγ2	0.69785	21.41207
172Ileδ1	0.66266	14.51133
175Leuδ1	1.03993	25.6668
?179MetCε	2.0444	16.98847
182IleCδ1	0.81512	13.81552
185Valγ1	1.00054	21.71301
185Valγ2	1.04561	20.09117
186Valγ1	0.84408	22.28614
186Valγ2	0.9453	22.51847
?188Leuδ2 δ1	0.69561	26.38526
190Leuδ1	0.82201	24.97223
190Leuδ2	0.84519	24.00199
195Leuδ1	0.81396	23.56664
195Leuδ2	0.82453	24.34711
199Valγ1	0.96017	21.21487
199Valγ2	1.01252	20.58092
201?Ileδ1	0.8277	13.09068
207Ileδ1	0.80129	14.63453
208Leuδ1	0.42793	24.03969
208Leuδ2	0.46254	24.95345
209Leuδ1	0.83774	23.68655
209Leuδ2	1.02479	27.29102
219Leuδ1	-0.00617	24.61477
219Leuδ2	0.26918	24.06242
220Leuδ1	0.83066	24.07831
224Valγ1	0.93373	21.35672
224Valγ2	1.16256	22.88631
233Ileδ1	0.58933	13.84706
235Valγ1	0.78262	21.3832
235Valγ2	0.83139	20.92826
236Valγ1	0.92619	20.39692
236Valγ2	1.03846	22.13734
240Leuδ2	0.76228	24.25165
240Leuδ1	0.80174	24.88085

Methyl-group	¹ H (ppm)	¹³ C (ppm)
241Valγ1	0.96502	21.61524
241Valγ2	0.98691	20.73092
245Ileδ1	0.91561	13.44412
251Leuδ1	0.7303	24.42599
251Leuδ1	0.75672	24.82047
252Valγ1	0.88187	21.61944
252Valγ2	1.08485	23.44088
255Ileδ1	0.70705	14.08352
258Leuδ1	0.76433	25.57627
258Leuδ2	0.83679	24.26824
266Ileδ1	0.66087	12.08505
267Ileδ1	-0.14998	13.41622
269Ileδ1	0.78813	13.39642
272Ileδ1	0.51845	14.59585
275Ileδ1	0.7717	14.12451
283Leuδ1	0.88588	23.44363
283Leuδ2	0.94388	24.90184
290Valγ1	1.11665	21.69291
290Valγ2	1.16195	22.62496
294Leuδ1	0.88002	22.75346
294Leuδ2	0.98492	25.35365
295MetCε	2.09195	17.27543
297Leuδ1	0.91929	25.42231
297Leuδ2	1.01289	23.90034
298Leuδ1	0.72167	21.62523
298Leuδ2	0.78482	25.18002
301MetCε	1.15972	16.05047
310Valγ1	0.57053	21.21532
310Valγ2	0.79401	21.08756
312Ileδ1	0.75988	14.83159
313Ileδ1	0.90808	13.39382
321Ileδ1	0.86027	13.94727
322Leuδ1	0.55065	22.47099
322Leuδ2	0.74083	26.51057
326IleCd1	0.61879	14.44999
327Leuδ1	0.77987	22.80261
327Leuδ2	0.81983	25.46388
335Ileδ1	0.82016	13.6458
336Ileδ1	0.6779	13.73946
338Valγ1	0.79733	22.22316
338Valγ2	0.81813	22.90067
347Leuδ1	0.77534	23.48897
347Leuδ12	0.94687	25.72596
?349Ileδ11	0.46975	14.51079
350Leuδ11	0.65883	25.92798
350Leuδ1	0.70413	23.24097
352Ileδ1	0.67942	14.40652
357MetCε	2.13999	17.16571
359Leuδ1	1.01382	25.20867
359Leuδ2	1.07284	23.84399
363Valγ1	0.58173	22.00541
363Valγ2	1.16211	23.60119
365Leuδ1	0.28976	21.85427
365Leuδ2	0.68169	25.38669

Methyl-group	¹H (ppm)	¹³C (ppm)
368Ileδ1	0.46043	10.86965
371MetCε	2.10321	16.55217
376Valγ1	1.03485	21.62143
376Valγ2	1.26856	18.09669
380Leuδ1	0.82318	22.37404
383IleCd1	0.88799	15.08817
389MetCε	2.03449	16.60092
392Ileδ1	0.84017	12.91274
395Leuδ1	0.87659	23.2163
395Leuδ2	0.91228	25.17956
399Valγ1	0.90419	23.16791
399Valγ2	0.94169	21.85451
401MetCε	2.20939	18.47805
408Valγ1	0.91725	22.22283
408Valγ2	1.14353	23.56228
411Ileδ1	0.78149	11.05292
412MetCε	2.09797	16.97492
417Valγ1	0.88497	20.10568
417Valγ2	0.94264	20.46903
419Valγ1	0.9366	21.13237
?425Leuδ2 δ1	0.8408	23.4932
?425Leuδ2 δ1	0.79705	23.47315
427Valγ1	0.85787	21.0408
427Valγ2	0.92734	21.12646
?428Leuδ2 δ1	0.8968	24.83463
?428Leuδ2 δ1	0.86758	24.7207

Appendix Table 11 Assignment of Ile δ 1 and Met ϵ methyl resonances in full-length PAN

Residue index	Amino acid	Conformer	IleH δ 1 (ppm)	IleC δ 1	Meth ϵ (ppm)	MethC ϵ (ppm)
7	Ile	A	0.82	12.65	-	-
28	Ile	A	0.85	12.74	-	-
49	Ile	A	0.9	12.98	-	-
56	Ile	A	0.98	14.93	2.13	-
74	Met	A	-	-	2.133	16.96
74	Met	B	-	-	2.19	16.95
74	Met	C	-	-	2.16	16.97
87	Met	A	-	-	2.12	17.7
87	Met	B	-	-	2.1	17.35
87	Met	C	-	-	2.17	18.31
93	Ile	A	0.95	14.16	-	-
93	Ile	B	0.92	14.08	-	-
93	Ile	C	0.94	14.11	-	-
159	Met	A	-	-	1.84	17.79
159	Met	B	-	-	1.85	17.83
159	Met	C	-	-	1.85	17.44
179	Met	A	-	-	2.04	16.98
233	Ile	A	0.58	13.8	-	-
266	Ile	A	0.64	12.42	-	-
266	Ile	B	0.66	12.44	-	-
266	Ile	C	0.65	12.06	-	-
267	Ile	A	-0.2	13.32	-	-
267	Ile	B	-0.16	13.39	-	-
267	Ile	C	-0.18	12.48	-	-
272	Ile	A	0.53	14.56	-	-
301	Met	A	-	-	1.12	15.92
312	Ile	A	0.74	14.78	-	-
326	Ile	A	0.61	14.51	-	-
349	Ile	A	0.47	14.56	-	-
357	Met	A	-	-	2.14	17.17
368	Ile	A	0.54	10.61	-	-
371	Met	A	-	-	2.1	16.58
371	Met	B	-	-	2.25	16.59
383	Ile	A	0.92	15.29	-	-
389	Met	A	-	-	2.09	16.65
401	Met	A	-	-	2.2	18.63
401	Met	B	-	-	2.25	18.69
411	Ile	A	0.81	11.17	-	-
412	Met	A	-	-	2.1	17.02

Appendix Table 12 Assignments of ILVM methyl-group resonances in GFPuv-ssrA

Methyl group	^1H (ppm)	^{13}C (ppm)
7LeuC δ 1	0.30662	25.90681
7LeuC δ 2	0.32267	22.10917
11ValCy1	0.80111	21.91163
11ValCy2	0.97716	21.98128
12ValCy1	1.22076	22.03736
12ValCy2	1.36255	22.05753
14IleC δ 1	0.04797	15.87606
14IleC δ 2	0.70991	18.19102

Methyl group	¹ H (ppm)	¹³ C (ppm)
15LeuCδ1	0.99828	24.64317
15LeuCδ2	1.02707	24.56018
18LeuCδ1	0.54353	27.44639
18LeuCδ2	0.68182	28.28083
22ValCy1	0.56509	21.12165
22ValCy2	0.71753	22.36252
29ValCy2	1.19557	23.05364
42LeuCδ1	0.51257	26.99987
42LeuCδ2	0.58569	26.02394
44LeuCδ1	0.38419	24.42561
44LeuCδ2	0.76493	26.63279
47IleCδ1	0.72555	13.07001
47IleCy2	0.82062	18.29689
60LeuCδ1	0.15697	23.46863
60LeuCδ2	0.38072	27.41428
61ValCy1	0.37422	19.97402
78MetCε	1.52555	16.2499
88MetCε	2.00814	20.1323
93ValCy2	1.07217	22.29976
98IleCδ1	0.64177	14.92371
98IleCg2	-0.52551	17.80125
112ValCy1	0.63601	22.96863
112ValCy2	0.83519	21.74002
119LeuCδ1	0.40854	23.34295
119LeuCδ2	0.61985	25.12366
120ValCy2	0.87536	21.40223
120ValCy2	0.92855	21.59208
123IleCδ1	0.91172	18.8084
123IleCy2	0.75021	20.22893
128IleCδ1	0.90671	13.99307
128IleCg2	0.95467	18.22771
136IleCδ1	0.28597	13.16465
136IleCy2	0.02268	18.41626
137LeuCδ1	0.4887	22.04884
137LeuCδ2	0.67186	26.19373
141LeuCδ1	0.41678	26.15334
141LeuCδ2	0.44821	25.84981
152IleCδ1	0.67276	12.926
152IleCy2	0.66032	18.29836
161IleCδ1	0.20114	15.14785
161IleCg2	0.06453	19.40197
167IleCδ1	-0.19629	6.32977
167IleCy2	-0.18313	15.51234
171IleCδ1	0.63701	12.42299
171IleCy2	0.95409	19.0978
176ValCy1	0.76534	21.5186
176ValCy2	0.84216	20.98304
188IleCδ1	0.87443	12.95915
188IleCy2	0.83685	17.1822
201LeuCδ1	0.58817	25.37586
207LeuCδ1	0.41135	25.18379
207LeuCδ2	0.54453	23.91622
218MetCε	1.9763	18.02751
220LeuCδ2	0.93217	27.10736

Methyl group	¹ H (ppm)	¹³ C (ppm)
221LeuCδ1	0.74775	23.91621
221LeuCδ2	0.88586	26.00483
224ValCγ1	0.62505	20.99833
229IleCδ1	0.52672	12.70705
229IleCγ2	0.89436	17.53451

Appendix Table 13 Assignments of backbone resonances of frGFP-ssrA at 315K

Residue index	Amino acid	H ^N (ppm)	N (ppm)	C ^α (ppm)	C ^β (ppm)
2	Ser	7.99	115.17	57.73	65
3	Lys	8.56	122.14	57.64	32.66
4	Gly	8.46	108.81	47.08	-
5	Glu	7.94	119.2	59.82	29.78
6	Glu	7.86	117.66	58.48	29.59
7	Leu	7.55	117.19	-	41.3
8	Phe	7.54	113.62	58.03	39.49
9	Thr	7.41	113.5	63.9	69.08
10	Gly	8.15	111.67	44.11	-
11	Val	8.2	120.81	63.33	32
12	Val	9.09	131.74	59.14	37
14	Ile	8.15	119.19	59.48	42.86
15	Leu	8.72	128.95	54.07	46.52
16	Val	8.81	122.84	60.19	35.73
17	Glu	8.87	126.01	55.64	34.02
18	Leu	9.21	125.84	54.04	45.55
19	Asp	8.26	128.3	54.03	43.53
20	Gly	8.44	110.68	43.96	-
21	Asp	7.12	120.35	53.58	43.74
22	Val	8.65	123.7	60.83	34.32
23	Asn	9.07	127.3	54.56	37.11
24	Gly	8.19	104.39	45.83	-
25	His	8.1	121.17	55.38	29.37
26	Lys	8.62	126.19	55.8	33.47
27	Phe	8.42	118.47	55.89	41.58
28	Ser	7.82	114.64	57.68	66.14
29	Val	9.15	123.61	60.31	37.52
30	Ser	9.1	122.25	56.76	66.19
31	Gly	9.77	111.54	46.07	-
32	Glu	7.91	117.29	54.43	33.99
33	Gly	8.01	110.12	46.19	-
34	Glu	8.75	121.69	55.15	33.65
35	Gly	8.83	108.04	44.31	-
36	Asp	9.06	122.78	52.34	42.8
37	Ala	10.72	129.73	54.45	20.07
38	Thr	9.09	116.27	66.56	68.17
39	Tyr	6.97	115.72	57.72	40.77
40	Gly	8.21	112.08	47.07	-
41	Lys	7.7	118.96	55.32	37.48
42	Leu	8.84	124.89	54.11	46.08
43	Thr	8.09	114.51	59.93	70.46
44	Leu	8.62	124.7	54.59	50.04
45	Lys	7.68	119.09	55.22	35.03

Residue index	Amino acid	H ^N (ppm)	N (ppm)	C ^α (ppm)	C ^β (ppm)
46	Phe	9.17	124.81	55.99	43.3
47	Ile	9.53	117.5	59.39	42.34
48	Cys	8.95	124.85	58.41	27.48
49	Thr	8.51	120.94	63.52	68.15
50	Thr	8.02	109.96	61.46	69.12
51	Gly	7.36	110.42	45.38	-
52	Lys	8.46	126.05	55.19	33.13
53	Leu	9.48	132.87	53.35	42.44
55	Val	6.12	106.94	56.5	32.43
59	Thr	7.75	103.72	64.68	69.09
60	Leu	7.95	118.33	54.48	43.27
61	Val	6.97	120.98	68.83	31.91
62	Thr	8.12	106.38	63.81	68.31
63	Thr	7.52	119.97	66.86	68.21
64	Leu	7.85	119.09	57.45	41.83
65	Thr	8.6	107.62	-	-
70	Cys	7.24	116.15	57.27	27.81
71	Phe	7.44	116.92	60.66	38.72
72	Ser	6.71	112.67	59.53	63.95
73	Arg	8.4	125.49	55.72	30.22
74	Tyr	9.08	132.4	55.49	38.99
78	Met	8.1	117.59	54.76	33.84
79	Lys	7.28	121.22	60.59	32.49
80	Arg	8.23	116.31	57.26	28.67
81	His	7.58	115.76	55.04	31.2
82	Asp	6.49	118.33	52.19	38.18
83	Phe	8.2	126.29	60.27	39.94
84	Phe	6.79	113.1	59.87	38.19
85	Lys	7.27	112.25	59.85	32.52
86	Ser	6.9	114.7	60.7	62.89
87	Ala	6.71	121.2	52.05	19.11
88	Met	7.54	116.63	52.85	31.06
90	Glu	9.58	126.44	61.78	28.97
91	Gly	8.6	100.6	45.34	-
92	Tyr	8.68	111.51	55.15	41.26
93	Val	9.54	121.01	61.15	33.45
94	Gln	9.61	128.61	55.14	33.5
95	Glu	9.69	132.65	55.15	33.98
96	Arg	8.71	118.06	55.99	35.15
97	Thr	8.23	115.16	63.46	70.96
98	Ile	9.54	127.69	59.78	39.69
99	Ser	8.79	122.12	56.47	63.78
100	Phe	8.6	128.39	58.32	38.42
101	Lys	7.85	127.09	58.66	32.5
102	Asp	9.16	122.05	56.34	39.95
103	Asp	8.58	122.99	53.7	44.03
104	Gly	7.68	108.17	44.28	-
105	Asn	7.8	111.59	51.5	43.49
106	Tyr	9.58	116.9	52.95	40.41
107	Lys	9.4	122.78	55.81	34.66
108	Thr	9.31	115.05	59.53	71.01
109	Arg	8.73	123.76	56.82	32.63
110	Ala	9.06	131.13	49.86	25.4

Residue index	Amino acid	H ^N (ppm)	N (ppm)	C ^α (ppm)	C ^β (ppm)
111	Glu	8.56	119.23	55.78	32.61
112	Val	8.62	125.62	60.62	32.01
113	Lys	8.6	119.95	55.07	35.26
114	Phe	7.6	117.99	59.39	40.23
115	Glu	9.18	124.45	55.81	29.94
116	Gly	8.82	115.81	47.17	-
117	Asp	8.52	126.23	54.32	41.45
118	Thr	7.66	115.14	62.27	71.5
119	Leu	8.55	130.77	52.72	41.21
120	Val	8.93	125.19	60.14	34.25
121	Asn	8.43	122.4	51.09	41.14
122	Arg	8.62	125.39	56.16	31.8
123	Ile	9.25	123.96	61.47	43.95
124	Glu	9.01	127.36	55.53	32.84
125	Leu	8.62	126.95	54.37	46.69
126	Lys	9.12	129.07	54.76	35.6
127	Gly	10.3	115.42	45.58	-
128	Ile	9.11	124.82	60.02	43.09
129	Asp	8.58	115.71	55.8	38.31
130	Phe	8.51	115.28	59.44	39.46
131	Lys	9.79	123.5	55.77	33.47
133	Asp	8.24	113.99	52.44	40.02
134	Gly	7.14	104.57	44.32	-
135	Asn	9.21	114.54	55.69	38.51
136	Ile	7.12	116.18	64.14	36.07
137	Leu	8.9	117.77	56.62	39.7
138	Gly	7.1	103.05	45	-
139	His	7.44	116.03	58.21	28.32
140	Lys	7.89	114.97	54.99	32.65
141	Leu	7.22	118.91	55.7	42.01
142	Glu	8.02	121.24	56.48	31.05
150	Val	-	-	61.66	-
151	Tyr	7.5	127.11	59.96	38.25
152	Ile	8.74	126.53	60.34	39.96
153	Thr	8.65	118.37	59.64	72.01
154	Ala	8.97	125.04	52.86	20.08
155	Asp	8.77	122.3	41.44	-
156	Lys	8.78	123.2	59.66	32.28
158	Lys	7.21	116.08	55.07	32.77
159	Asn	7.86	118.26	54.27	38.35
160	Gly	7.26	99.29	44.91	-
161	Ile	7.53	110.21	58.3	43
162	Lys	9.1	120.42	54.75	36.9
163	Ala	8.66	119.62	50.76	23.06
164	Asn	8.58	117.51	53.15	42.8
165	Phe	8.22	116.6	56.72	40.03
166	Lys	8.7	120.15	55.28	35.01
167	Ile	7.91	123.91	55.67	37.13
168	Arg	9.75	126.82	54.78	31.18
169	His	8.96	124.81	54.42	30.05
170	Asn	9.08	125.46	55.05	37.97
171	Ile	8.83	126.66	59.45	36.96
172	Glu	8.53	124.69	59.25	29.74

Residue index	Amino acid	H ^N (ppm)	N (ppm)	C ^α (ppm)	C ^β (ppm)
173	Asp	7.4	117.99	53.84	40.37
174	Gly	8.27	108.76	45.31	-
175	Ser	8.09	117.38	58.18	64.43
176	Val	8.17	116.75	61.04	35.68
177	Gln	9.33	127.56	53.35	31.8
178	Leu	8.91	130.37	54.7	44.06
179	Ala	9.11	125.5	50.58	21.3
180	Asp	8.83	131.81	55.96	42.27
181	His	9.18	123.21	56.02	29.78
182	Tyr	8.46	123.22	57.52	40.02
183	Gln	8.68	127.12	53.72	33.66
184	Gln	9.12	123.95	55.24	33.17
185	Asn	8.83	121.44	52.33	41.13
186	Thr	8.66	113.97	58.61	71.39
188	Ile	8.78	124.07	64.21	38.84
189	Gly	9.29	109.82	44.16	-
190	Asp	8.28	117	53.77	41.75
191	Gly	8.33	109.7	-	-
193	Val	7.99	115.33	59.34	35.4
194	Leu	8.02	121.99	53.86	41.09
195	Leu	7.97	121.23	50.21	41.21
197	Asp	9.34	120.74	54.02	41.58
198	Asn	8.87	121.1	54.32	37.44
199	His	8.32	117.45	56.25	28.08
200	Tyr	8.38	114.38	55.22	42.16
201	Leu	8.55	116.11	52.59	43.14
206	Lys	9.33	123.84	55.27	35.42
207	Leu	9.06	127.91	53.62	42.23
208	Ser	9.46	115.56	57.74	65.45
215	Arg	9.01	120.5	56.5	31.04
216	Asp	9.26	125.37	55.75	41.61
217	His	8.41	124.3	56.76	33.12
218	Met	7.9	116.63	54.98	37.28
219	Val	7.64	127.59	62.43	31.8
220	Leu	9.07	129.1	53.99	48.31
221	Leu	8.35	129	54.02	45.07
222	Glu	8.79	127.69	54.27	-
223	Phe	8.89	122.15	56.4	-
224	Val	9.32	123.14	61.03	35.19
225	Thr	8.3	121.49	60.22	71.9
226	Ala	8.36	128.72	51.42	19.19
227	Ala	8.69	122.19	51.69	22.04
228	Gly	8.02	103.94	45.36	-
229	Ile	7.4	118.92	60.72	39.45

Appendix Table 14 Assignments of methyl-group resonances of frGFP-ssrA at 333K

Methyl-group	¹ H (ppm)	¹³ C (ppm)
7Leuδ1	0.34995	25.89223
7Leuδ2	0.36226	22.08346
9Thry2	1.33786	22.15042

Methyl-group	¹ H (ppm)	¹³ C (ppm)
11Valy2	0.84894	21.92522
12Valy1	1.25725	22.06367
12Valy2	1.40351	22.08598
14Ileδ1	0.03353	15.95263
14Iley2	0.72027	18.30172
15Leuδ1	1.03656	24.72713
15Leuδ2	1.07442	24.3833
16Valy1	1.05718	21.88978
16Valy2	1.22548	23.20278
18Leuδ1	0.56326	26.87883
18Leuδ2	0.68815	27.84622
22Valy1	0.60493	21.12455
22Valy2	0.75584	22.30951
37Alaβ	1.69827	20.11011
38Thry2	0.86875	21.14844
42Leuδ1	0.50733	27.01396
42Leuδ2	0.55693	26.36762
43Thry2	1.12289	21.38723
44Leuδ1	0.3751	25.42329
44Leuδ2	0.66628	27.75813
47Ileδ1	0.80887	13.15407
47Iley2	0.89989	18.12985
49Thry2	1.36468	22.75257
50Thry2	1.29153	22.65822
53Leuδ1	-0.22219	24.92208
53Leuδ2	-1.00696	24.02615
55Valy1	-0.16723	15.73642
55Valy2	0.26893	22.00072
59ThrCg2	0.79578	23.45849
60Leuδ1	0.44854	22.82106
60Leuδ2	0.4699	27.39183
61Valy1	0.2805	19.30582
61Valy2	0.79666	24.68516
62Thry2	0.41446	23.3175
63Thry2	1.1249	23.60006
64Leuδ1	-0.17659	24.86536
64Leuδ2	0.55523	22.88664
65Thry2	1.08134	22.92432
68Valy1	1.00998	20.20542
78Metε	1.56732	16.26442
87Alaβ	1.22798	19.1152
88Metε	2.04104	20.12792
93Valy1	1.11902	21.60431
93Valy2	1.1124	22.26289

Methyl-group	¹H (ppm)	¹³C (ppm)
97Thry2	1.17485	21.44234
98Ileδ1	-0.46979	17.7511
98Iley2	0.75478	14.91878
108Thry2	1.07003	20.86428
110Alaβ	1.63962	25.39513
112Valγ1	0.66945	23.01595
112Valγ2	0.85979	21.82095
118Thry2	1.04539	21.81296
119Leuδ1	0.42294	22.96402
119Leuδ2	0.66205	25.00023
120Valγ1	0.90871	21.52527
120Valγ2	0.96373	21.6094
123Ileδ1	0.80187	20.23405
123Iley2	0.83653	19.34081
125Leuδ1	0.19934	27.44823
125Leuδ2	0.22626	25.24239
128Ileδ1	0.9558	13.97712
128Iley2	1.00565	18.23599
136Ileδ1	0.09141	18.44708
136Iley2	0.35466	13.28146
137Leuδ1	0.53668	22.06025
137Leuδ2	0.71956	26.21887
141Leuδ1	0.4704	26.10914
141Leuδ2	0.49829	25.81159
150Valγ1	-0.0216	19.12802
150Valγ2	0.39031	20.4458
152Ileδ1	0.72228	12.98136
153Thry2	1.13971	20.59894
154Alaβ	1.58867	20.18479
161Ileδ1	0.25069	15.16401
161Iley2	0.10312	19.47607
163Alaβ	1.27768	23.13464
167IleCδ1	-0.2218	6.32444
167Iley2	-0.11286	15.76914
171Ileδ1	0.68402	12.3446
171Iley2	0.99693	19.10554
176Valγ1	0.80776	21.46586
176Valγ2	0.8866	21.03358
178Leuδ1	0.85945	24.37209
178Leuδ2	0.93515	25.29003
179Alaβ	0.77655	21.32623
186Thry2	1.38999	22.488
188Ileδ1	0.91062	12.94342
188Iley2	0.87576	17.16186

Methyl-group	¹H (ppm)	¹³C (ppm)
193Valγ1	0.80624	21.83175
193Valγ2	0.86115	18.45499
194Leuδ1	0.71089	25.52805
194Leuδ2	0.75955	22.58315
195Leuδ1	0.77571	23.29863
195Leuδ2	0.99682	25.95309
201Leuδ1	0.62156	25.32219
201Leuδ2	0.75763	22.11139
207Leuδ1	0.42185	25.08882
207Leuδ2	0.50272	24.02007
218Metε	1.68107	17.64772
219Valγ1	0.91589	21.31088
219Valγ2	0.98981	21.86693
220Leuδ1	0.82987	25.0988
220Leuδ2	0.90458	28.26012
221Leuδ1	0.77932	23.82857
221Leuδ2	0.93584	25.99904
224Valγ1	0.65179	21.73616
225Thry2	1.21089	21.02048
226Alaβ	0.37512	19.29833
227Alaβ	1.00271	22.03263
229Ileδ1	0.54344	12.61745
229Ileγ2	0.93672	17.45793

Acknowledgements

First, and foremost, I would like to thank Professor Teresa Carlomagno for the great opportunity of working in her research group and on this highly challenging but also competitive research project. I am grateful for all the possibilities for the experimental work and the opportunity to work together in an international research collaboration including the participation in overseas conferences. Thank you very much for all the input and the constant motivation and stimulation to progress this research further.

I would like to thank Prof. Dietmar Manstein and Prof. Thomas Brüser for being part of my thesis committee and taking interest into my research by reading my PhD thesis.

As part of the French-German collaboration on the PAN-project I would also like to thank Dr. Frank Gabel, Dr. Bruno Franzetti and Dr. Emilie Mahieu for their great input, all the exciting discussions on PAN and the successful publication on the SANS experiments of the unfolding of GFP by PAN.

Special thanks go to Dr. John Kirkpatrick, who introduced me to the theoretical, but also practical, aspects of protein NMR spectroscopy. Without your tireless and patient explanations I would have never grasped the NMR background, necessary to finish this PhD project. Further, I would like to thank you for your invaluable support in setting up NMR experiments and assisting in the data analysis using your bash scripts.

Further, I would like to thank Dr. Luca Codutti, Dr. Juliane Buschmann and Dr. Petra Hinse for taking care of the entire organizational lab management tasks, resulting in a smooth working atmosphere in the lab.

All present and past members of the Carlomagno lab I would like to thank for the very nice atmosphere. It has been a great time and I am really grateful for all the coffee-break discussions, the relentless believe in science and the confidence that every research project will turn out successful eventually.

On a more personal note, I would like to thank Dr Michelangel Marasco, MD, PhD, for the great friendship we developed over the times. I do miss our coffee break and the verbal diarrhea of these conversations as much as you do. You really helped me get through this PhD.

Last, I would like to thank my little family, my wife Hannah, my son Kalle and my daughter Charlotte, for supporting me through these almost five years of my PhD. I am grateful that you tolerated me and my unsteady mood especially during difficult times during the lab work part of the project but also during the COVID-19 lockdown and me writing the thesis at home. Further, I would like to thank my parents for always believing in me and supporting my throughout my studies.

Curriculum Vitae Georg Krüger

Georg Ernst Heinrich Krüger, M. Sc.
Hansestrasse 85,
21337 Lüneburg
georg_krueger@me.com
0176 70569990
Date of Birth 08.08.1991 in Schorndorf, Germany
Married, two children (3 and 0 y/o)

Professional Experience

08/2016-04/2020 **Leibniz University Hannover**
Center of Biomolecular Drug Research (BMWZ)
Doctoral Researcher
Biomolecular NMR spectroscopy

Higher Education

08/2016- today **Leibniz University Hannover**
Doctoral student in Biomolecular NMR spectroscopy
Supervisor: Prof. Teresa Carlomagno

10/2013-06/2016 **Ernst-Moritz-Arndt-University Greifswald**
Master's studies Biochemistry

03/2015-12/2015 **University of British Columbia, Vancouver, Canada**
Guest researcher and Master's student
Lab of Prof Dr. Stephen Withers
Thesis title: "Towards chemo-enzymatic polysialylation of therapeutic relevant proteins"

10/2010-09/2013 **Ernst-Moritz-Arndt-University Greifswald**
Bachelor's studies
Thesis with Prof. Dr. Uwe Bornscheuer: Thesis title: „Studien über Acylasen als Quorum-Quenching-Enzyme“

School education

09/2003-06/2010 **Altes Gymnasium Oldenburg**
Abitur,

10/2007-05/2008 Crew member of "High Seas High School" on the sail training ship Johan Smidt (crossing the Atlantic Ocean on a sailing vessel twice)

Internships

03/2012 & 09/2012 **Grünenthal Pharma Aachen**
Dept. of Molecular Pharmacology, Medicinal Chemistry and Analytical Chemistry

Voluntary Services

- 03/2011-02/2015 Organizing committee of the Greifswald International Students Festival (GrIStuF), Board Member of the Student's club GrIStuF
- 10/2010-09/2011 Support of an autistic fellow student on different concerns

Scholarships

- (12/2011-06/2016) Performance scholarship of the Foundation of German Business der (Sdw), "Begabtenförderwerk Klaus Murmann", including travel grants for a 10 month research stay in Vancouver, Canada

Language skills

German (native speaker)
English (proficient)
Spanish (basic skills)

Contributions to scientific publications

Mahieu E, Covès J, **Krüger G**, Martel A, Moulin M, Carl N, Härtle M, Carlomagno T, Franzetti B, Gabel F. Observing Protein Degradation by the PAN-20S Proteasome by Time-Resolved Neutron Scattering. *Biophys J.* 2020 Jul 21;119(2):375-388. doi: 10.1016/j.bpj.2020.06.015. Epub 2020 Jun 24. PMID: 32640186; PMCID: PMC7376118.

Last D, **Krüger GH**, Dörr M, Bornscheuer UT. Fast, Continuous, and High-Throughput (Bio) Chemical Activity Assay for N-Acyl-L-Homoserine Lactone Quorum-Quenching Enzymes. *Appl Environ Microbiol.* 2016 Jun 30;82(14):4145-54. doi: 10.1128/AEM.00830-16. PMID: 27208131; PMCID: PMC4959188.