Doctoral Thesis

The archaeal proteasome complex and small ubiquitin-like proteins

Author(s): Ranjan, Namit

Publication Date: 2010

Permanent Link: https://doi.org/10.3929/ethz-a-006325706

Rights / License: In Copyright - Non-Commercial Use Permitted
The archaeal proteasome complex and small ubiquitin-like proteins

A dissertation submitted to
ETH ZURICH

for the degree of
Doctor of Sciences

presented by
Namit Ranjan
M.Sc. Genomics, Madurai Kamaraj University, India
born on April 29th 1983
Indian

accepted on the recommendation of
Prof. Dr. Eilika Weber-Ban – examiner
Prof. Dr. Rudolf Glockshuber – co-examiner
Prof. Dr. Nenad Ban – co-examiner
Prof. Steven Burston – co-examiner

2010
# Contents

1 **Summary/Zusammenfassung**
   1.1 Summary  
   1.2 Zusammenfassung

2 **Introduction**
   2.1 Energy dependent protein turnover  
   2.2 Archaeal PAN-proteasome  
   2.3 Substrate recruitment for degradation  
   2.4 Protein modification by small ubiquitin-like proteins  
   2.5 Small archaeal modifier proteins (SAMPs)

3 **Aims of this study**

4 **In vitro reconstitution of the PAN-proteasome complex from Methanosarcina acetivorans**
   4.1 Introduction
   4.2 Experimental procedures  
      4.2.1 Cloning of constructs  
      4.2.2 Protein preparation  
      4.2.3 Analytical size-exclusion chromatography  
      4.2.4 Peptidase assay  
      4.2.5 Electron microscopy  
      4.2.6 ATPase activity of PAN  
      4.2.7 SDS-PAGE based analysis of model substrate degradation  
      4.2.8 Michaelis-Menten kinetics for GFP-ssrA and RepDn  
      4.2.9 FRET measurements using manual and stopped-flow mixing  
      4.3.10 Pulldowns
   4.3 Results
      4.3.1 Characterization of *Methanosarcina acetivorans* Proteasome  
      4.3.2 Recombinant PAN is fully assembled and exhibits ATPase activity  
      4.3.3 PAN stimulates degradation of model substrates  
      4.3.4 Substrate processing by the PAN-proteasome complex  
      4.3.5 Interaction of PAN with the proteasome is substrate independent  
      4.3.6 Identification of interacting partners for the PAN-proteasome complex
4.4 Discussion 31
4.5 Supplementary data 33

5 Structure and the activation of *Methanosarcina acetivorans* small archaeal modifier proteins (SAMPs) 35
5.1 Introduction 35

5.2 Experimental procedures
- 5.2.1 Construction of plasmids 36
- 5.2.2 Expression and purification of proteins 36
- 5.2.3 NMR measurement of SAMP1 (in collaboration) 37
- 5.2.4 Homology modeling of SAMP activating enzyme 38
- 5.2.5 Pulldown with Strep-SAMP1 38
- 5.2.6 Monitoring activation of SAMP using radiolabeled ATP 38
- 5.2.7 ESI mass-spectrometry analysis of adenylated SAMP 39
- 5.2.8 Phylogenetic analysis 39

5.3 Results
- 5.3.1 Bioinformatic analysis of *M. acetivorans* small modifier proteins 39
- 5.3.2 Solution structure of SAMP1 and comparison with known Ubls 41
- 5.3.3 SAMP activating enzyme (SAE) interacts with SAMPs 44
- 5.3.4 SAE activate and adenylates SAMPs 48

5.4 Discussion 51
5.5 Supplementary data 55

6 Concluding remarks and outlook 57

7 References 59

8 Curriculum vitae 69

9 Acknowledgments 71
Chapter 1

1.1 Summary

The energy-dependent protein degradation in cells across all three kingdoms of life is carried out by large cylindrical assemblies consisting of central protease- and outer ATPase rings. In archaea, the hexameric ATPase PAN binds to opposite ends of the central four-stacked proteasome core particle. PAN catalyzes the ATP-dependent unfolding and translocation of protein substrate into the central cavity of the proteasome, where degradation into small peptides takes place. In eukaryotes, the ubiquitin-proteasome system is well studied, whereas only recently evidence has emerged that small archaeal modifier proteins (SAMPs) with homology to ubiquitin could be involved in substrate tagging for degradation.

The first part of this thesis describes the in vitro reconstitution and characterization of the PAN-proteasome complex from a methanogenic archaeon, Methanosarcina acetivorans. It is demonstrated that recombinantly produced PAN and proteasome core particle from this organism interact to form the protease-active complex in vitro. Experimental evidence is presented showing that PAN exhibits ATPase activity, and the proteasome on its own can degrade a small fluorogenic model substrate Suc-LLVY-AMC. The comparison of the activities of M. acetivorans PAN and the proteasome with homologues from other archaea shows that the proteasome is similarly active, whereas PAN exhibits lower ATPase activity than its homologues. Experiments with various model substrate proteins show that the M. acetivorans PAN-proteasome complex recognizes and degrades ssrA-tagged model substrates originally designed for bacterial Clp proteases, like GFP-ssrA and λ-Rep-ssrA, most likely due to their unstructured nature. A fluorescence resonance energy transfer experiment using fluorescently tagged variants of an ssrA-tagged model substrate provides evidence for the directional transfer and processing of the model substrate starting at the unstructured ssrA tail.

In the second part of the dissertation, the activation mechanism of the recently identified small archaeal modifier proteins is investigated. Using bioinformatic analysis five SAMPs are identified from M. acetivorans that have the conserved C-terminal di-glycine motif and share homology to ubiquitin and ubiquitin-like proteins. Results are presented that demonstrate the activation of SAMP by a protein thus far annotated as a hypothetical protein of 27 kDa, and that is herein identified as SAMP activating enzyme (SAE) catalyzing the adenylation of SAMP in presence of ATP. The adenylation is specific for the conserved C-terminal glycine and requires ATP hydrolysis. The activation of SAMP occurs in analogous manner to the activation of ubiquitin and ubiquitin-like proteins involved in various cellular processes. Furthermore, in collaboration with the Allain-group the solution structure of SAMP1 is determined and it shows a four-stranded β-sheet opposing an α-helix, thereby classifying SAMP1 as a member of the β-grasp fold protein family. SAMP1 exhibits an extensive flexible surface loop between α3 and β3 with a surface exposed lysine residue, which is not observed in bacterial ubiquitin homologues or in many other archaeal SAMPs. In addition, SAMP1
possesses a convex groove with an “acidic” and “hydrophobic” face on opposite sides, which is a likely interaction site for other macromolecular binding partners.

1.2 Zusammenfassung

In Zellen aller drei Domänen des Lebens wird der energieabhängige Proteinabbau von grossen, zylindrischen Komplexen ausgeführt, die aus zentralen Protease- und äusseren ATPase-Ringen bestehen. In Archaeen bindet die hexamere ATPase PAN jeweils an die gegenüberliegenden Enden des aus vier Ringen bestehenden zentralen Zylinders des Proteasoms. PAN katalysiert die ATP-abhängige Entfaltung und Translokation von Substrat-Proteinen in die zentrale Kammer des Proteasoms, wo die Proteine dann zu kleinen Peptiden abgebaut werden. In Eukaryonten ist das Ubiquitin-Proteasom-System sehr gut untersucht, wohingegen sich für Archaeen erst kürzlich erste Hinweise ergeben haben, dass das zu Ubiquitin homologe Protein SAMP (Small Archaeal Modifier Protein) in Prozesse involviert ist, die Protein für die Degradation durch das Proteasom markieren.


Im zweiten Teil der vorliegenden Dissertation werden die Untersuchungen zur Aktivierung des erst kürzlich entdeckten Proteins SAMP beschrieben. Mittels bioinformatischer Methoden wurden fünf SAMPs in *M. acetivorans* identifiziert, die ein konserviertes Diglycin-Motiv am C-Terminus aufweisen und homolog zu Ubiquitin und ubiquitin-ähnlichen Proteinen sind. Im Rahmen dieser Arbeit konnte gezeigt werden, dass dieAktivierung von SAMP durch ein bis dahin als hypothetisch annotiertes Protein einer molekularen Masse von 27 kDa vermittelt wird. Dieses im Folgenden als „SAMP activating enzyme“ (SAE) bezeichnete Protein katalysiert in Gegenwart von ATP die Adenylierung von SAMP, wobei spezifisch das konservierte, C-terminale Glycin
Chapter 2

Introduction

2.1 Energy dependent protein turnover

Each cell has the capacity for energy-dependent turnover of proteins and removal of functionally impaired proteins (Bukau et al., 2006; Goldberg, 2003; Horwich et al., 1999; Sauer et al., 2004; Striebel et al., 2009b). The energy-dependent protein degradation process is conserved in all kingdoms of and is carried out by chaperone-protease complexes (Figure 1). Chaperone-proteases are responsible for maintaining cell homeostasis and are part of the post-translational protein quality control. The chaperone component belongs to the family of AAA+ ATPases and is responsible for unfolding and translocating substrates to the proteasome under the expense of ATP (Snider and Houry, 2008; Wickner et al., 1999). The biologically active state of the chaperone is a hexameric

Figure 1. Chaperone–proteases exist in all kingdoms of life. They consist of a barrel-shaped protease (blue) and a chaperone-ring belonging to the AAA+ family of proteins (beige). The chaperone recruits substrates (green) and uses the energy of ATP hydrolysis to unfold and translocate them into the protease chamber, where they are degraded to peptides. Eukaryotes contain the 20S proteasome capped by the complex 19S particle, harboring the ATPases Rpt (1–6) at its base. Substrates are mostly recognized by means of a polyubiquitin-tag (red barrel-shaped chain) which is removed from the substrate before degradation. Archaea contain a simpler version of the 20S proteasome that associates with PAN, the archaeal homolog of the Rpt-ATPases. As an example of a bacterial chaperone–protease ClpAP is shown, which recruits substrates often via N-terminal or C-terminal motifs (red extension).
ring with a central pore through which substrates are translocated into the proteasome. In bacteria and archaea, the chaperone is composed of six identical subunits, whereas in eukaryotes additional non-ATPase subunits combine with six ATPase subunits forming the 19S regulatory particle (Figure 1) (Schmidt et al., 2005; Smith et al., 2005).

The proteolytic core has a stacked-ring, cylindrical architecture. In bacteria it is composed of two back-to-back stacked homoheptameric rings (Kress et al., 2009; Striebel et al., 2009b), whereas the eukaryotic and archaeal 20S proteasome are composed four heptameric rings, the two outer \( \alpha \)-rings that bind to the ATPase and the two inner \( \beta \)-rings possessing the proteolytic active sites (Figure 1) (Maupin-Furlow et al., 2004; Striebel et al., 2009b; Zwickl et al., 1999b).

The process of protein degradation is initiated by the attachment of a degradation tag to substrate proteins that destines these protein for degradation. In eukaryotes and certain bacteria, the proteasome recognizes and degrades substrates covalently tagged with the small modifier proteins ubiquitin and Pup (prokaryotic ubiquitin-like protein), respectively (Elsasser and Finley, 2005; Finley, 2009; Striebel et al., 2009a). In archaea, ubiquitin-like small archaeal modifier proteins (SAMPs) have been identified, but this modification has not been unambiguously linked to degradation thus far (Humbard et al., 2010).

2.2 Archaeal PAN-proteasome

All archaea contain 20S proteasome particles that are built from stacked, heptameric rings arranged in an \( \alpha \beta \beta \alpha \) order (Figure 1), with the two inner \( \beta \) homoheptameric rings carrying the active sites and the two outer \( \alpha \)-rings containing the entrance pores into the degradation cylinder. Some archaea have two different isoforms of \( \alpha \)- or \( \beta \)-subunits resulting in several proteasome isoforms. The eukaryotic proteasome consists of different but structurally similar \( \alpha \) and \( \beta \)-subunits. During the assembly process the N-terminal \( \beta \)-propeptides are autocatalytically processed generating a catalytic active threonine (Thr) residue that is sequestered into the hollow interior of the 20S proteasome (Seemuller et al., 1996; Zwickl et al., 2000). Unlike in eukaryotes, where only three \( \beta \)-subunits possess a proteolytically active site, all \( \beta \)-subunits of the archaeal proteasome are proteolytically active (Maupin-Furlow et al., 2006; Zwickl et al., 1999b). The conserved catalytic mechanism of the \( \beta \)-subunits involves nucleophilic attack of the N-terminal threonine hydroxyl-group on the peptide bond of the substrate (Figure 2), resulting in cleavage and release of the peptide products generated (Kisselev et al., 2000). The active eukaryotic \( \beta 1, \beta 2, \) and \( \beta 5 \) proteasomal subunits display caspase-like (or peptidylglutamyl-peptide hydrolyzing activity), trypsin-like and chymotrypsin-like activity, whereas most archaeal proteasomes display chymotrypsin-like activity (Cardozo, 1993; Maupin-Furlow et al., 2003; Orlowski et al., 1993; Rivett, 1993). However, 20S proteasomes from methanogens apparently display a high caspase-like activity (Maupin-Furlow et al., 1998; Maupin-Furlow et al., 2003; Wilson et al., 2000).
Proteolysis by the 20S proteasome is carried out in co-ordination with an ATPase complex that unfolds the substrate and threads it into the proteasome. In archaea, the associated ATPase is PAN (proteasome associated nucleotidase), a hexamer belonging to the AAA+ (ATPase associated with various cellular activities) family. Like other members of the AAA+ family, PAN also contains a canonical AAA domain, P-loop.

**Figure 2. Mechanism of peptide bond hydrolysis by the proteosomes.** The hydroxyl group of the active site threonine or serine residue attacks the carbonyl of the peptide bond resulting in the formation of a tetrahedral intermediate (A). This intermediate decays with the generation of acyl-enzyme intermediate and the product. Finally nucleophilic attack by water on the acyl-enzyme intermediate yields the free proteasome and the second reaction product. The free proteasome recycles the whole process. The nucleophilic attacks are shown in RED.
motif, Walker A and Walker B motifs, and a second region of homology (SRH motif). PAN is the closest known homologue of the Rpt2 ATPase subunit of the eukaryotic 19S proteasome-regulatory complex, which shares 41-45% similarity. In most archaea, only one pan gene exists, but some possess two homologous genes, like for example *Haloferax volcanii*. Here, two gene products, PAN-A and PAN-B, are generated, the expression of which is differentially regulated under different growth phases (Kirkland et al., 2008).

Structural information revealed that PAN consists of six CC-OB (coiled-coil oligonucleotide binding) domains, linked to six AAA domains that dock onto the proteasome. The N-terminal coiled-coil domain is involved in substrate recognition. The OB ring is postulated to be the critical barrier by which the selectivity of proteasomes for unfolded proteins is enforced. In addition, it may also serve as a pivot for substrate unfolding. The conserved C-terminal hydrophobic-tyrosine-X motif (Hb-Y-X) of PAN is essential for association with and gate opening of the 20S cylinder upon ATP binding (Rabl et al., 2008; Smith et al., 2007) (Figure 3). The conformation of the narrow channel of α-rings regulates substrate entry into the proteasome. PAN, when mixed with the 20S proteasome and ATP, can unfold and degrade stable globular proteins and proteins lacking tight tertiary structure (Smith et al., 2005).

![Figure 3. The closed and the open gate of the 20S proteasome.](image)
(a) Top view of the α-ring of the 20S proteasome without association of chaperone. The channel is closed. (b) Top view of the α-ring of the 20S proteasome bound to the chaperone or to peptides identical the ATPase C-termini with the hydrophobic-Tyr-X motif. The binding of the chaperone induces a conformational change leading to gate opening. Structural coordinates are taken from 1RYP (closed) and 1FNT (open).
2.3 Substrate recruitment for degradation

Substrates destined for degradation are recognized by the chaperone-moiety of the chaperone-protease. This recognition can be mediated or regulated through an additional control by adaptor proteins (Inobe and Matouschek, 2008; Striebel et al., 2009b). The recruitment of substrates can thus be either by direct recognition or through an adaptor-mediated process.

2.3.1 Recruitment of substrate in bacteria

In bacteria, most of the cytoplasmic substrates destined for degradation are recruited either to members of the Clp protease family or to Lon protease (Flynn et al., 2003; Gur and Sauer, 2008). ClpAP and ClpXP degrade ssrA-tagged proteins that are generated by tmRNA system. The ssrA tag is an 11-amino acid tail that is added to stalled nascent chains at the ribosome (Moore and Sauer, 2007). In addition to this, another well studied tagging signal in bacteria is the N-end rule system in which substrates are recognized by an aromatic residue, or leucine at the very N-terminus followed by an arginine or lysine (Erbse et al., 2006; Varshavsky, 1996).

In case of the ssrA-tagged substrates, recruitment can occur by direct recognition of the tag at the ClpA ATPase. However, recruitment of this class of substrates to ClpX is mediated by the adaptor protein SspB that increases the affinity of ssrA-tagged substrates towards ClpXP. In contrast, another adaptor protein, ClpS, switches the substrate specificity of ssrA tagged substrates towards N-end rule substrates for ClpA (Kress et al., 2009) by mediating the recognition of these substrates and by inhibiting degradation of other classes of substrates. These bacterial degradation signals are sufficient for localizing the substrates at the chaperone, and they serve at the same time as initiation sites for unfolding and translocation.

2.3.2 Prokaryotic pupylation

Pupylation is a process analogous to eukaryotic ubiquitination for tagging substrates for degradation in actinobacteria. In this process the small protein Pup is covalently attached to a lysine residue of substrates. Pup shares no overall sequence homology to ubiquitin except the glycine-glycine motif at the penultimate position of the Pup sequence. Pup is encoded with either a glutamate (Glu) or a glutamine (Gln) residue at its C-terminus depending on the organism. In the pupylation pathway, the ultimate Gln must be deamidated to Glu by the enzyme Dop (Deamidase of Pup), so that this Glu can then be conjugated to a lysine residue of the target substrate by PafA (Pearce et al., 2008; Striebel et al., 2009a; Sutter et al., 2010).

For proteasome-dependent degradation of the pupylated protein, the Pup-conjugated substrate must be unfolded by the bacterial proteasomal ATPase (Mpa) and transferred into the proteasome core. In this process the N-terminal coiled-coil (CC domain) domain of Mpa first interacts with Pup-conjugated substrate and the AAA
domain then unfolds the ppylated substrate under the expense of ATP. The interaction of the proteasome core with Mpa leads to the degradation of the Pup-conjugated substrate along with its modifier \textit{in vitro} (Striebel et al., 2010).

### 2.3.3 Ubiquitination in eukaryotes

In eukaryotes, substrates are targeted to the proteasome by ubiquitination, a process of covalently linking ubiquitin to substrates. Ubiquitin (Ub) is a small ubiquitous protein that is conjugated to a lysine residue of a substrate via an isopeptide-bond followed by formation of poly-ubiquitin through C-terminal Gly of one Ub to Lys48 of the preceding Ub (Hartmann-Petersen and Gordon, 2004; Kerscher et al., 2006). To be recognized by the proteasome, a substrate has to be modified with a chain of at least four Ubs (Elsasser and Finley, 2005). The whole polyubiquitination cascade needs the sequential action of

![Ubiquitin conjugation to substrates](image)

**Figure 4. Ubiquitin conjugation to substrates.** During activation, ubiquitin (orange) is converted to a C-terminal acyl adenylate by E1 (dark blue). Ubiquitin is then covalently attached to E1 through a thioester bond formed between the ubiquitin C-terminus and the conserved active site cysteine. It is then transferred to E2 (light blue) by trans-thioesterification. Finally, the protein substrate (black) is conjugated to ubiquitin by E3 (red).
three different enzymes, namely E1 (ubiquitin activating enzyme), E2 (ubiquitin conjugating enzyme) and E3 (ubiquitin-protein ligase). E1 activates ubiquitin by adenylation of the carboxylate of its C-terminal glycine and subsequently forms a covalent intermediate via a thioester-bond. Then the thioester bond is shuttled to E2 via a trans-thiolation reaction resulting in E2-ubiquitin thioester formation. The final conjugation step in ubiquitination is catalyzed by E3, resulting in formation of an isopeptide-bond between the substrate lysine residue and ubiquitin (Figure 4) (Elsasser and Finley, 2005; Finley, 2009; Striebel et al., 2009b).

Usually polyubiquitinated substrates are recognized by the 26S proteasome complex, however degradation of some substrates, like for example ornithine-decarboxylase, is independent of the ubiquitin pathway (Benaroudj et al., 2001). The Rpn10 and Rpn13 subunits of the 19S regulatory complex function as intrinsic receptors for polyubiquitin, whereas extrinsic receptors that function as adaptors can also recruit substrates for degradation. These extrinsic receptors contain a ubiquitin-binding (UBA, ubiquitin associated domains) and a proteasome-binding site (UBL, ubiquitin-like domain), and thus deliver substrates to the proteasome (Finley, 2009; Husnjak et al., 2008).

In addition to polyubiquitination, for efficient degradation, the substrate needs to contain an unstructured initiation site from which unfolding can occur. Therefore, polyubiquitination seems to function mainly in tethering the substrate to the proteasome until engagement for unfolding has occurred. Once the substrate is engaged for degradation, Rpn11 catalyzes the removal of polyubiquitin, resulting in recycling of Ub (Verma et al., 2002; Yao and Cohen, 2002).

2.4 Protein modification by small ubiquitin-like proteins

In eukaryotes, a major mechanism to regulate protein function involves the covalent attachment of monoubiquitin or ubiquitin-like proteins (UBLs) through an isopeptide bond to the Lys side chain of target proteins (Finley, 2009; Hochstrasser, 2000b; Kerscher et al., 2006). The hallmark of ubiquitin and UBL are the similar β grasp fold and a glycine-glycine motif at their C-terminus. UBLs comprise a diverse group of evolutionarily related small proteins that are involved in diverse biological processes and post-translational modifications with UBLs regulate numerous processes, including cell division, immune response and embryonic development. Defects in UBL pathways are associated with various diseases, particularly cancer, neurodegenerative disorders, and muscular dystrophy or cachexia (Ciechanover and Schwartz, 2004). Although UBLs vary in their degree of similarity to ubiquitin, they are all conjugated to target proteins by a similar enzymatic cascade of activation and conjugation (Figure 4) (Burroughs et al., 2009; Hochstrasser, 2000b, 2009; Leidel et al., 2009; Pedrioli et al., 2008; Schmitz et al., 2008).

Ubiquitin is universal in all eukaryotes and UBLs are prevalent in eukaryotes, but homologous proteins are also found in bacteria and archaea. The bacterial homologues of ubiquitin are MoaD and ThiS, and they have been shown to be involved in sulphur
transfer reactions for molybdopterin cofactor biosynthesis and thiamin biosynthesis, respectively. Both MoaD and ThiS have the conserved diglycine motif at their C-terminus, and the mechanism of their activation is similar as for ubiquitin (Begley et al., 1999a; Begley et al., 1999b; Lehmann et al., 2006; Rudolph et al., 2001; Schmitz et al., 2007). They are activated in an ATP-dependent manner by their cognate E1 enzymes, namely MoeB and ThiF (Duda et al., 2005; Lake et al., 2001; Lehmann et al., 2006; Leonardi and Roach, 2004; Schmitz et al., 2008; Zhang et al., 2010). Cysteine desulfurases have been shown to be involved in transferring sulfur from free cysteine to the adenylated MoaD intermediate forming its thioester form (Figure 5). IscS is one such cysteine desulfurase from E.coli (Zhang et al., 2010).

Homologs of the sulfur transfer protein are also present in eukaryotes, and one such known example is Urm1 from Saccharomyces cerevisiae. Urm1 shares sequence identity of 20% and 23% to MoaD and ThiS, respectively. However it lacks sequence homology with ubiquitin. Urm1 is shown to be activated by its cognate E1 enzyme, Uba4p, which shares sequence homology to MoeB and ThiF. In addition, Uba4p has a C-terminal rhodanese domain responsible for sulfur transfer to the adenylated Urm1 intermediate from its conserved cysteine of CRXGX(R/T) motif (Figure 5) (Bordo and Bork, 2002; Burroughs et al., 2009).

Figure 5. Model for sulphur transfer involved in various cellular functions. UBL (orange) is first activated by formation of a C-terminal acyl adenylate by the respective enzymes (dark blue). The adenylated UBL is converted to the thioester form by IscS in molybdopterin synthesis or, Uba4p in tRNA modification. The sulphur for thioesterification is proposed to be provided by free cysteine in the cell. The thioesterified UBL then transfers sulphur for various functions including molybdopterin co-factor biosynthesis, thiamine biosynthesis, tRNA modification, etc.
In summary, homologues of ubiquitin are present in all kingdoms of life, and they perform diverse functions ranging from sulfur transfer to posttranslational protein modification for regulatory roles other than degradation.

2.5 Small archaeal modifier proteins (SAMPs)

Archaea represent one of the three major evolutionary lineages of life. Archaea encode the 20S proteasome and an associated chaperone PAN, sharing homology to the eukaryotic ATPase subunits of the regulatory particle. Although it has been known that eukaryotes target their proteins for proteasomal degradation by ubiquitination, an equivalent mechanism has not definitively been shown in archaea. Archaeal ubiquitin-like proteins exist, but are less conserved and still poorly understood. Recently it was shown that two archaeal homologues of ubiquitin, SAMP1 and SAMP2, form protein-conjugates \textit{in vivo}. SAMPs are small proteins that exhibit homology to bacterial and eukaryotic ubiquitin-like proteins but little sequence homology to ubiquitin itself except a conserved C-terminal glycine-glycine motif. SAMP1 and SAMP2 were shown to make an isopeptide bond with the lysine residue of a target protein, analogous to the end product of the ubiquitination pathway. The levels of these conjugates were altered under different growth conditions and in proteasome gene knockout cells (Humbard et al., 2010). However, a direct link between SAMPylation and proteasomal degradation is still lacking.
Aims of the study
Chapter 3

Aims of this study

In archaea, the energy-dependent chaperone PAN associates with the 20S proteasome to carry out protein degradation. Previous studies from other archaeal PAN-proteasome complexes revealed that PAN can recognize, unfold and translocate protein substrates into the proteasome, where degradation takes place (Smith et al., 2006; Zwickl et al., 1999b). The first goal of this dissertation study was to characterize the PAN-proteasome complex from a methanogenic archaeon, *Methanosarcina acetivorans*. The aim was to reconstitute and biochemically characterize the PAN-proteasome complex *in vitro*, and to study the reaction steps involved in proteasome-mediated protein degradation in real-time using an *in vitro* fluorescence resonance energy transfer (FRET) based assay.

The second goal of this dissertation was to investigate mechanisms of substrate recruitment to the archaeal proteasome complex. A bioinformatic approach was taken to identify possible ubiquitin-like modifier proteins. Several such potential modifiers were identified, cloned and characterized *in vitro*. During the course of study a group working *in vivo* with another archaeon, *Halofexx volcanii*, showed that two such “small archaeal modifier proteins” (SAMPs) are covalently attached to protein lysines *in vivo*. These two SAMP proteins are homologues of the proteins studied in this dissertation. Using bioinformatic methods I also identified a potential enzyme of the SAMP modification pathway. A combination of assays, based on protein interactions using pulldowns and radioactivity were employed in order to study the SAMPylation pathway. Furthermore, the size and stability of *M. acetivorans* SAMP1 made it possible to utilize an NMR based approach in order to determine its solution structure.
Chapter 4

In vitro reconstitution of the PAN-proteasome complex from Methanosarcina acetivorans

4.1 Introduction

Energy-dependent protein degradation is carried out in all living cells and is responsible for rapid and selective turnover of many intracellular proteins (Baker and Sauer, 2006; Baumeister et al., 1998; Goldberg, 1992; Maupin-Furlow et al., 2004; Prakash and Matouschek, 2004). Protein degradation by the proteasome is carried out in association with a chaperone that unfolds the protein at the expense of energy and directs it into the proteasome for degradation (Smith et al., 2006; Smith et al., 2005; Zwickl et al., 1999b). In archaea, the proteasome is a large cylindrical complex composed of four stacked homoheptameric rings in an αββα arrangement (Figure 1). The α-rings act as gatekeepers, restricting the entry of folded proteins into the central chamber, while the inner β-rings carry the active sites where substrate degradation takes place. The N-termini of the β-subunits harbor pro-peptides that are autocatalytically processed generating the end-standing active site threonine (Thr) during the assembly process (Maupin-Furlow et al., 2003; Seemuller et al., 1996). In contrast to eukaryotes where only three β-subunits are active, all seven β-subunits of the archaeal proteasome are active.

PAN, a member of the AAA+ protein family is the chaperone that unfolds the substrate and threads it into the archaeal proteasome. PAN is a homolog of eukaryotic 26S ATPases (Zwickl et al., 1999a) and the crystal structure of PAN revealed that it adopts a donut-shaped structure with two subcomplexes, namely subcomplex I and II. Subcomplex I consists of a coiled-coil domain and the oligonucleotide/oligosaccharide-binding fold (OB fold), whereas subcomplex II represents the entire nucleotidase domain. It has been proposed that the translocation of substrate through the OB ring is the prerequisite for the selectivity of proteasomes for unfolded proteins, OB domain may serve as a fulcrum for substrate unfolding (Zhang F, 2009).

It was shown that PAN stimulates proteasomal degradation in presence of ATP of proteins lacking tight tertiary structure as well as globular proteins (Benaroudj et al., 2003; Navon and Goldberg, 2001). In addition, it has been shown that substrates tagged with the bacterial ssrA-tag are recognized and degraded by the archaeal PAN-proteasome complex (Benaroudj and Goldberg, 2000). However, there is no evidence for the existence of such a tagging system in archaea. Up to date, there are no natural substrates known for the archaeal PAN-proteasome complex.

This section of the thesis is about the in vitro reconstitution of the PAN-proteasome complex from a methanogenic archaeon, Methanosarcina acetivorans. We cloned and heterologously expressed in E. coli, purified and characterized the PAN-proteasome complex. We utilized an in vitro assay based on fluorescence resonance energy transfer (FRET) (Kolygo et al., 2009) to follow the degradation of substrates in real-time. The strong dependence of FRET efficiency on the distance between donor and
acceptor fluorophores (Stryer, 1978) can be used to follow the dynamic behavior of molecules in solution. We make use of the distance-dependence of FRET by engineering model substrates reporting in real-time on the steps involved in substrate processing by the chaperone-protease. Cysteine variants of the N-terminal domain of the cI repressor from bacteriophage λ (residues 1-92) carrying a C-terminal bacterial ssrA recognition tag (herein referred to as Rep) were chosen as model substrates. Furthermore, we set out to identify potential adaptors and natural substrates of the PAN-proteasome complex.

4.2 Experimental procedures

4.2.1 Cloning of constructs

All genes were obtained by PCR from *M. acetivorans* genomic DNA. The proteasome’s (CP) α- and β- subunit were cloned into pETDuet vector (Novagen) with NcoI/BamHI and Ndel/BglII restriction sites of multiple cloning site I and II, respectively. PAN was cloned via Ndel/BamHI restriction sites into pET20b(+) (Novagen).

Tagged inactive proteasome particles (iCP-Strep) were obtained by site directed mutagenesis from pETDuet harboring proteasomal genes, where the active site threonine (Thr) of β-subunit was replaced with alanine (Ala), and a Strep-tag was introduced at the C-terminus. PAN-Strep and Strep-PAN were cloned into pET20b(+) via NdeI/BamHI restriction sites.

4.2.2 Protein preparation

Active (CP), inactive (iCP) and tagged inactive (iCP-Strep) proteasome particles were overexpressed in BL21 (DE3) cells from IPTG-inducible plasmids at 25°C. PAN, Strep-PAN and PAN-Strep were overexpressed in Rosetta (DE3) cells from IPTG-inducible plasmids at 15°C.

Untagged proteasome particles were purified using ion-exchange (Q Sepharose) and hydrophobic interaction chromatography (Phenyl Sepharose) followed by size-exclusion chromatography (Superdex 200). Tagged proteasomes were purified by Strep-Tactin affinity chromatography (IBA) followed by size-exclusion chromatography (Superdex 200). All proteins were stored in 50 mM HEPES-KOH pH 7.5, 300 mM KCl, 1 mM DTT, 1 mM EDTA, 10% (v/v) glycerol.

PAN was purified using ion-exchange chromatography (Q Sepharose) and ammonium sulphate precipitation followed by dialysis against 50 mM HEPES-KOH pH 7.5, 300 mM KCl, 1 mM DTT, 50 mM MgSO₄, 10% (v/v) glycerol. Strep-PAN and PAN-Strep were purified by Strep-Tactin affinity chromatography (IBA) and stored in PAN storage buffer.

λ-Repressor-ssrA and GFP-ssrA were expressed and purified as described (Cranz-Mileva et al., 2008; Weber-Ban et al., 1999). The degradation substrates (RepDc and RepDn) were covalently modified with thiol-reactive fluorophores (1,5-IAEDANS and
Fluorescein-5-maleimide) forming a FRET donor-acceptor pair as described (Kolygo et al., 2009). RepDn carry the fluorophores at positions 4 and 13 close to the N-terminus, and RepDc carry the fluorophores at positions 77 and 92 close to the C-terminus, respectively.

Chromatographic material was purchased from GE healthcare, unless otherwise mentioned. Concentration of proteins was determined by absorption measurements at 280 nm. All purified proteins were verified by ESI-MS.

### 4.2.3 Analytical size-exclusion chromatography

20 μM purified proteasome and 10 μM PAN hexamer were analyzed on 2.4 ml Superose6 analytical size-exclusion chromatography in the absence and presence of 1 mM ATP in 50 mM Tris-HCl pH 7.5, 300 mM NaCl, 10% (v/v) glycerol. All runs were performed at room temperature. For assessing the molecular mass of the proteasome, ferritin (440 kDa) and thyroglobulin (667 kDa) were analyzed on Superose6 analytical size-exclusion chromatography in the same running buffer as standards.

### 4.2.4 Peptidase assay

The fluorogenic substrate Succinyl-Leu-Leu-Val-Tyr-AMC (AMC, 7-amino-4-methylcoumarin) was used to measure the peptide hydrolyzing activity of the proteasome. Therefore, 30 nM proteasome were added to different concentrations of the substrate (0 μM to 350 μM) in buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl) and the reaction was followed in a total volume of 400 μL. Fluorescence was measured in a Hitachi F-4500 fluorescence spectrometer with excitation wavelength at 380 nm and emission wavelength at 441 nm.

### 4.2.5 Electron microscopy

To confirm the integrity of the purified proteasome particles, we recorded negative stain electron microscopy images. Specimens for electron microscopy were prepared by applying 20 μL aliquots of 30 nM proteasome particles in 50 mM Tris-HCl pH 7.5, 300 mM KCl, 20 mM MgCl₂ and 10% (v/v) glycerol on freshly hydrophilized carbon-coated 400 mesh copper grids (Pelco) for 10 s, after which excess buffer was blotted off. The specimens were washed two times with 20 μL of water. 20 μL of a 1% (w/v) uranyl acetate aqueous solution was applied to the grids. Excess stain was blotted off, and grids were allowed to dry. Specimens were viewed in a Zeiss Leo 912 transmission electron microscope operating at a magnification of 31500x.
4.2.6 ATPase activity of PAN

The ATPase activity was measured using a continuous spectrophotometric assay at 291 nm coupled to inorganic phosphate production using 7-methylinosine (m7inosine) and the enzyme purine nucleoside phosphorylase (PNPase) (Rieger et al., 1997). 0.3 µM PAN hexamer was mixed with 1 mM m7inosine, 0.75 U/mL PNPase in presence and absence of 0.3 µM of 20S proteasome, 3 µM GFP-ssrA and 3 µM λ-Repessor-ssrA (λ-Rep-ssrA), where applicable. The reaction was started with the addition of ATP. To determine Michaelis-Menten parameters, different concentrations of ATP (0 mM to 8 mM) were used. Reactions were carried at 23°C in 50 mM HEPES-KOH pH 7.5, 300 mM KCl, 1 mM EDTA, 1 mM DTT, 50 mM MgSO4, 10% (v/v) glycerol, and the obtained data were analyzed using the following equation:

\[ V_0 = \frac{V_{max} \cdot [S]}{K_M + [S]} \]  

where \( V_0 \) is the initial velocity, \( V_{max} \) is the velocity at substrate saturation, \( S \) is the substrate concentration and \( K_M \) is the Michaelis constant.

4.2.7 SDS-PAGE based analysis of model substrate degradation

5 µM GFP-ssrA and λ-Rep-ssrA were mixed with 0.5 µM PAN-proteasome complex in 50 mM HEPES-KOH pH 7.5, 300 mM KCl, 20 mM MgCl2, 10% (v/v) glycerol supplemented with creatine phosphate/phosphocreatine kinase ATP-regeneration system (Sigma) at 23°C. The reaction was started by addition of 5 mM ATP and aliquots were taken at indicated time points (0 s aliquot was taken before addition of ATP) and mixed with 6X SDS sample buffer. All samples were subjected to SDS-PAGE using an 18% gel followed by Coomassie staining for analysis.

4.2.8 Michaelis-Menten kinetics for GFP-ssrA and RepDn

0.2 µM PAN-proteasome complex was mixed with different concentrations of GFP-ssrA (0 µM to 10 µM) or RepDn (0 µM to 7 µM) in 50 mM HEPES-KOH pH 7.5, 300 mM KCl, 20 mM MgCl2, 10% (v/v) glycerol supplemented with creatine phosphate/phosphocreatine kinase ATP-regeneration system (Sigma) at 23°C and fluorescence was measured using a Hitachi F-4500 fluorescence spectrometer. The reaction was started with 10 mM ATP and was followed till degradation was complete. Turnover (k, min\(^{-1}\)) for each substrate concentration was calculated, plotted and was fitted using equation (1). For GFP-ssrA the excitation and emission was set to 400 nm, 510 nm; for RepDn 340 nm, 475 nm, respectively.
4.2.9 FRET measurements using manual and stopped-flow mixing

Fluorescence measurements were carried out at 23°C using a PTI Quantamaster QM-7 spectrofluorimeter with excitation set at 340 nm and emission at 475 nm. Rapid kinetic measurements were performed at 23°C with a SX-20 MV stopped-flow spectrometer with excitation at 340 nm and emission at 475 nm. Final stopped-flow fluorescence traces are an average of 14-18 individual runs. For all FRET experiments, time-courses were sequentially measured with the donor-only substrate followed by the donor-acceptor-labeled substrate. FRET efficiency was calculated according to the following equation:

\[ \text{FRET}_{ef} = 1 - \frac{F_{DA}}{F_D} \]

where \( F_{DA} \) and \( F_D \) are the fluorescence intensities of the donor in presence of and absence of acceptor, respectively.

To follow the degradation kinetics of the PAN-proteasome complex, 3 µM PAN and 6 µM proteasome were preassembled in 5 mM ATP for 5 min and then mixed with 0.5 µM RepDc or RepDn either in a stopped-flow device or manually (PTI). The reaction was performed in 50 mM HEPES-KOH pH 7.5, 10% (v/v) glycerol, 400 mM KCl and 20 mM MgCl₂.

4.2.10 Pulldowns

For pulldowns using archaeal cell lysate as prey, PAN and iCP either as Strep-tagged, or coupled to amine reactive beads by reductive amidation (AminoLink, Pierce) were used. *M. acetivorans, Pyrococcus furiosus, Archaeoglobus fulgidus* and *Methanothermobacter thermautotrophicus* cell lysate was precipitated in 70% ammonium sulfate and the resuspended protein mixture was desalted using a PD10 column (GE Healthcare) into 50 mM Tris-HCl pH 7.5, 150 mM KCl, 5 mM MgCl₂, 10% (v/v) glycerol. Strep-tagged proteins or proteins coupled to beads were incubated with the treated lysate for different time incubations (1 h to 4 h) at either room temperature or 4°C. After three washing steps, bound proteins were eluted with SDS sample buffer and analyzed by Coomassie stained SDS-PAGE. The experiments were also performed in different buffers, with and without detergents.

For pulldowns using purified proteins, 0.3 µM iCP-Strep was incubated with 3 µM PAN hexamer in 50 mM Tris-HCl pH 7.5, 300 mM NaCl, 20 mM MgCl₂, 0.5 mM DTT, 10% (v/v) glycerol, 0.1% (v/v) Triton X-100 for 1 h at room temperature (reaction volume 200 µL). After three washing steps, bound PAN was eluted with SDS sample buffer, and analyzed on Coomassie stained SDS gel.
4.3 Results

4.3.1 Characterization of Methanosarcina acetivorans proteasome

In order to characterize the proteasome from *M. acetivorans*, we cloned, recombinantly overexpressed and purified the proteasome from *E. coli*. Analytical size-exclusion chromatography confirmed the assembly state of the purified proteasome and proteasome that the preparation was homogenous (Figure 6a). In order to assess the complex mass based on the size-exclusion chromatogram, the peak position of the proteasome was compared to molecular weight standards, ferritin and thyroglobulin. From the comparison, the molecular weight of the assembled proteasome was judged to be of correct size. The precise molecular mass of the individual subunits was confirmed by ESI-mass spectrometry (Supplementary figure 1). The purity of the preparation was confirmed by SDS-PAGE and subsequent Coomassie staining.

![Figure 6. Purified *M. acetivorans* proteasome is fully assembled. (a) Size exclusion-chromatogram of the purified proteasome. For comparison, molecular weight markers, ferritin (440 kDa, red) and thyroglobulin (667 kDa, green) were analyzed under the same conditions. (b) Negative stain electron microscopy images of the proteasome at 31500 x magnification shows the four stacks of the proteasome.](image)

For further analysis of the assembly state of the *M. acetivorans* proteasome, we performed negative stain electron microscopy. Fully assembled proteasomes were visible as a stack of four rings that are easily discerned. There were some top views visible as circular rings, probably because a small portion of the assembled proteasome decomposes into half-proteasomes during the specimen preparation (Figure 6b).

To characterize the proteolytic activity of the purified proteasome, we performed a peptidase assay using the small fluorogenic substrate Suc-LLVY-AMC. Though the
proteasome remains in a closed gate conformation in the absence of the chaperone (Rabl et al., 2008; Smith et al., 2007; Smith et al., 2005), and does not allow peptides >7 residues to enter the catalytic core, the fluorogenic substrate is small enough to diffuse into the active catalytic chamber. The free AMC released after cleavage was monitored spectrophotometrically. Enzyme assays were performed and the $K_M$ was calculated to be 16.5 µM (Figure 7) with an error of ±10%.

![Figure 7. Peptidase activity of M. acetivorans proteasome.](image)

The degradation of small fluorogenic substrate Suc-LLVY-AMC by the proteasome was monitored at different concentration. The fluorescence of released AMC after cleavage was followed, turnover $k$ is plotted against the concentration of substrate and $K_M$ is calculated using equation 1.

### 4.3.2 Recombinant PAN is fully assembled and exhibits ATPase activity

PAN belongs to the AAA$^+$ ATPase family, can unfold substrate at the expense of ATP and thereby directs the unfolded substrate to the proteasome (Smith et al., 2005; Zwickl et al., 1999a). To characterize the chaperone PAN from *M. acetivorans*, we cloned, overexpressed and purified PAN recombinantly in *E. coli*. Purified PAN was subjected to analytical size-exclusion chromatography. The appearance of a single peak confirmed the homogeneity of the preparation (Figure 8). However, in presence of ATP, the peak position of PAN shifted suggesting a conformational change upon addition of ATP (Figure 8). The purity of the purified PAN was assessed by SDS-PAGE followed by Coomassie staining, and the correct monomer molecular mass was confirmed by ESI-mass spectrometry (Supplementary figure 1).
Figure 8. PAN undergoes a conformational change in presence of ATP. The purified PAN was analyzed on analytical size-exclusion chromatography in absence of ATP (green) and in presence of ATP (black). Upon addition of ATP, PAN undergoes a conformational change as observed as a shift in peak position (black).

PAN, a member of AAA\(^+\) family shares the hallmark property of this family, containing a P-loop domain with the signature Walker A and B motifs for nucleotide binding and hydrolysis. We tested the ATPase activity of purified PAN by a continuous UV-spectrometric assay coupled to purine nucleoside phosphorylase (PNPase). From Michaelis-Menten analysis, the catalytic rate constant (\(k_{\text{cat}}\)) for ATP hydrolysis was calculated to be about 22 per min per PAN hexamer (Figure 9a), which is about 10 fold less than observed for other archaeal proteasomal ATPases (Zwickl et al., 1999a). To investigate the effect of the presence of the proteasome or model substrates on the PAN ATPase activity, we measured the ATPase activity of PAN in presence of equimolar proteasome, or 10-fold excess GFP-ssrA or \(\lambda\)-Rep-ssrA. No change in the ATPase activity of PAN was observed upon addition of either proteasome or excess of GFP-ssrA or \(\lambda\)-Rep-ssrA (Figure 9b).
Figure 9. PAN exhibits ATPase activity. (a) PNPase based ATPase activity of the purified PAN. For full Michaelis-Menten kinetics different concentrations of ATP (0 mM to 8 mM) were used. (b) ATPase activity of PAN is not affected by the addition of the proteasome and substrates like GFP-ssrA or λ-Rep-ssrA.
4.3.3 PAN stimulates degradation of model substrates

In archaea, no natural proteasomal substrates or any tagging system targeting proteins toward proteasomal degradation has been identified. PAN, when mixed with 20S proteasome and ATP, can stimulate unfolding and degradation of ssrA-tagged substrates and globular substrates. To characterize the stimulation of substrate unfolding and degradation by the *M. acetivorans* PAN-proteasome complex, we used GFP-ssrA and λ-Rep-ssrA as model substrates. The reaction was carried out under steady state conditions, where GFP-ssrA or λ-Rep-ssrA was mixed with the PAN-proteasome complex in presence or absence of ATP. Samples were taken at different time points and analyzed

![Figure 10. GFP-ssrA is recognized and degraded by the PAN-proteasome complex.](image)

The degradation of GFP-ssrA was analyzed on a coomassie stained SDS gel where the substrate is incubated with the PAN-proteasome complex in presence of ATP (a), or absence of ATP (b). Energy-dependent autodegradation of PAN was analyzed both in presence (a) and absence of substrate, GFP-ssrA (c).
by SDS-PAGE and subsequent Coomassie staining. We observed the complete degradation of GFP-ssrA after 4 hours in the presence of ATP, whereas no degradation was observed in the absence of ATP (Figure 10a, b). Therefore, the *M. acetivorans* PAN-proteasome complex can also recognize and degrade ssrA-tagged substrates, analogous to other archaeal PAN-proteasome complexes.

Interestingly, we observed a very slow energy-dependent auto-degradation of PAN in the presence and absence of substrates (Figure 10a, c over-night). This suggests that PAN is recognized itself as a substrate.

To further characterize the purified PAN-proteasome complex, we performed enzyme assays and fit the data using equation (1). For GFP-ssrA, the $K_M$ was calculated to be 0.9 µM and 1.3 µM for RepDn, respectively (Figure 11).

![Figure 11. Michaelis-Menten kinetics for degradation of GFP-ssrA and RepDn by the PAN-proteasome.](image)

The degradation of different concentrations of GFP-ssrA (a) and RepDn (b) were followed by mixing with PAN-proteasome. For each concentration the calculated turnover, $k$ (min$^{-1}$) was plotted against the substrate concentration and $K_M$ was calculated using equation (1).
4.3.4 Substrate processing by the PAN-proteasome complex

After binding of a substrate protein, the chaperone-protease unfolds the substrate, translocates it from the chaperone ring into the proteolytic chamber and cleaves it into small peptides. To follow degradation of substrates in real-time, PAN and the proteasome were assembled in the presence of ATP to form the PAN-proteasome complex that was then mixed with either of two FRET substrates (Figure 12a and b). Both carried a cysteine pair located close enough together on the primary sequence to obtain efficient FRET transfer, but far enough apart that cleavage in the proteolytic chamber would result in a cut between the modified cysteine residues. In one substrate, the reporter pair was located near the N-terminus, in the other near the C-terminus. Fluorescence time-courses were recorded either after manual mixing or after rapid mixing in a stopped-flow device (first 6 min). A decrease in the FRET signal was observed for the RepDc probe after a lag-phase of about 40 s (Figure 12c, inset, blue trace). In the case of RepDn (Figure 12c, inset, red trace), the decrease in FRET starts later (after a lag-phase of about 80 s), indicating a later arrival of the N-terminus in the degradation chamber. This supports a model of directional processing of substrates starting at the tag. The ssrA recognition tag is presumably recruited in this case due to its unstructured nature. Interestingly, RepDc showed a signal increase in the phase preceding degradation (between 15 and 40 s), presumably due to unfolding or threading of the substrate towards the proteolytic active sites located in the center of the 20S particle. Complete degradation of the model substrates was observed after about 25 min for RepDc and after about 35 min for RepDn. Both degradation curves reach the same final FRET level suggesting a similar peptide size distribution with a fraction of peptides still bearing both fluorophores.
Figure 12. Substrate degradation by PAN-proteasome followed by FRET. Fluorescently-labeled cysteine variants of the ssrA-tagged model substrate used to probe degradation activity (a) and (b). The model substrate (N-terminal domain of λ repressor) is shown in ribbon representation (PDB entry 1lmb), the C-terminal ssrA tag shown in blue is not part of the x-ray structure, yellow and green balls represent the positions of engineered cysteines. c) Degradation of RepDc (blue trace) or RepDn (red trace) was followed after manual or stopped-flow (first 6 min) mixing with PAN-proteasome complex preassembled in the presence of ATP. Inset shows the first 4 min of the reaction.

4.3.5 Interaction of PAN with the proteasome is substrate independent

PAN when mixed with the substrate and ATP can unfold and translocate the substrate into the proteasome where degradation takes place. However, the interaction between hexameric PAN and heptameric proteasome was difficult to observe (Smith et al., 2005). In order to investigate the interaction between PAN and the proteasome, we performed
Results

Pulldown experiments. Inactive C-terminally Strep-tagged \textit{M. acetivorans} proteasome (iCP-Strep) was incubated with PAN in the presence of ATPγS for 1 h. The bound PAN was eluted and analyzed on Coomassie stained SDS gel. We observed a strong interaction of PAN with the proteasome (Figure 13, lane 1) and the interaction was found to be independent of substrates. As a control, PAN was incubated without iCP-Strep, and we did not observe non-specific interaction of PAN with the Strep-Tactin beads (Figure 13, lane 2).

![Figure 13. PAN interacts with the proteasome.](image)

**Figure 13. PAN interacts with the proteasome.** iCP-Strep was incubated with PAN for 1 h at room temperature, bound PAN was eluted and analyzed on Coomassie stained SDS gel (lane 1). As a control PAN was incubated without iCP-Strep (lane 2).

4.3.6 Identification of interacting partners for the PAN-proteasome complex

From this and other studies, it has been shown that archaeal PAN-proteasome complex can recognize, unfold and degrade ssrA-tagged model substrates like GFP and \textit{\lambda}-Repressor. However, no adaptors or natural substrates are known for the archaeal PAN-proteasome complex. To find potential interacting partners for the PAN-proteasome complex, we performed pulldown experiments using iCP-Strep and PAN or proteins (PAN, proteasome) coupled to amine reactive beads as bait and various archaeal cell lysates as prey. After incubation of prey with the bait, we eluted the proteins associated with the bait protein and analyzed them by Coomassie stained SDS gel. As a control, we incubated the prey without bait in the same buffer and the eluant was analyzed by SDS-PAGE and Coomassie staining. The identity of the potential interacting partners was confirmed by mass spectrometry. Under different conditions used for the pulldowns, we were not able to identify any promising interacting partners because the identified proteins were also present in the controls.

30
4.4 Discussion

Over last decades the insight into the world of proteolysis has expanded and proteases are now known to be involved in various biological processes like maintaining homeostasis between anabolic and catabolic pathways, or having a key regulatory role in many cellular pathways (Maupin-Furlow et al., 2005; Maupin-Furlow et al., 2004; Zwickl et al., 1999b). In archaea, the ATPase PAN associates with the proteasome for the breakdown of protein substrates (Maupin-Furlow et al., 2005; Maupin-Furlow et al., 2006; Zwickl et al., 1999b).

In this study, we have reconstituted and characterized the PAN-proteasome complex \textit{in vitro} from a methanogenic archaean, \textit{Methanosarcina acetivorans}. In order to characterize the organism’s proteasome, we cloned the respective coding genes and heterologously expressed them in \textit{E. coli} and purified the proteasome to homogeneity. Making use of analytical size-exclusion chromatography and peptidase assay, we showed that the purified proteasome is fully assembled and active (Figure 6, 7). The calculated $K_M$ ($\sim 17 \mu M$) for Suc-LLVY-AMC (Figure 7) is comparable to the $K_M$ observed for \textit{Pyrococcus furiosus} proteasome (Madding et al., 2007), suggesting a similar mechanism of substrate degradation by the \textit{M. acetivorans} proteasome. The unfolding and threading of substrates to the proteasome is carried out by the chaperone PAN in an energy-dependent process requiring ATP (Striebel et al., 2009b; Zwickl et al., 1999a). Our results from \textit{M. acetivorans} also show that the purified PAN is active, exhibits ATPase activity (Figure 9a) and is not affected by the presence of either the proteasome or substrates (Figure 9b), which is in contrast to the observation made from \textit{Methanocaldococcus jannashii} (Zwickl et al., 1999a). It is still an open question why \textit{M. acetivorans} PAN behaves different from other archaeal PAN. The archaeal chaperone PAN shows extensive homologies to the six ATPases of the eukaryotic 19S regulatory particle that are associated with the 20S proteasome, and is therefore the most likely evolutionary precursor of the 19S regulatory particle. Despite the large evolutionary distance, the primary sequence and domain organization of PAN and those of human and yeast 26S ATPases are indeed very similar (Zwickl et al., 1999a), and therefore, probably function quite similarly in promoting protein breakdown by the proteasome particle.

Fluorescence resonance energy transfer is a powerful tool for studying not only protein-protein interaction, but also intramolecular conformational changes (Klostermeier and Millar, 2001; Rye, 2001; Yang and Narlikar, 2007). The key advantage of the FRET method is that it mainly reports on the distance between fluorophores, thus providing information on the inter- or intramolecular proximity (Van Der Meer, 1994). In order to obtain a fluorescence based, real-time readout of substrate processing by the PAN-proteasome complex, we generated a model substrate $\lambda$-Rep-ssrA, which is labeled with the donor and acceptor fluorophore, thereby forming a FRET donor-acceptor pair. Our data show that PAN-proteasome recognizes, unfolds and degrades the model substrate under the expense of ATP (Figure 12c). A delay between the onsets of RepDc and RepDn degradation by PAN-CP complex (Figure 12c, inset) provides a rough estimate of the rate with which the polypeptide is spooled through the proteasome entrance pores. The
cysteine pair on RepDn is about 75 amino acids upstream compared to the equivalent cysteine pair on RepDc leading to a difference in the lag-phases of about 40 s. This suggests a translocation rate of roughly two amino acids per second. This is slower than what was observed for bacterial ClpAP and ClpXP (Kolygo et al., 2009). There, the delay between the arrival of the C-terminally located acceptor fluorophore at the donor fluorophore inside the ClpP cavity versus the arrival of the N-terminally located acceptor fluorophore is about 1.5 s. This suggests a rate of about 50 amino acids per second. However, the ATPase activities of the respective PAN-proteasome, ClpAP and ClpXP complexes also differ greatly. While ClpA shows the highest ATPase rate of about 8 s\(^{-1}\) in the presence of ClpP and substrate, ClpX is less active with a rate of around 2 s\(^{-1}\), but PAN-proteasome exhibits an even slower rate of only 0.4 s\(^{-1}\). Still even considering the amount of ATP molecules hydrolyzed per substrate molecule, the PAN-proteasome exhibits a lower efficiency per ATP hydrolyzed. In addition, our SDS-PAGE based degradation assay also confirmed that *M. acetivorans* can recognize and degrade ssrA-tagged substrates despite the fact that the PAN ATPase activity is not stimulated in their presence (Figure 10, 9b). Interestingly, we observed an energy-dependent autodegradation of PAN analogous to the autodegradation of *E. coli* ClpA by ClpAP in the presence of ATP (Maglica et al., 2008), where the C-terminal stretch is recognized as a degradation tag. We propose a similar mechanism for PAN autodegradation and this mechanism of self-destruction might be crucial for maintaining the homoeostasis of cell during stress.

To identify potential interacting partners of the PAN-proteasome complex, we performed pulldown experiments using different strategies (Strep-tagged proteins and proteins coupled to amine reactive beads) and under various conditions, but we were not able to identify any candidates. A possible explanation could be the weak interaction between interacting partners and PAN or the proteasome. As the biochemical approach was not successful, we applied bioinformatical means in order to identify interacting partners for the PAN-proteasome. This will be discussed in the next chapter of this thesis.

In summary, we reconstituted the methanogenic PAN-proteasome complex and characterized it biochemically. We could show that the purified proteasome behaves similar to other archaeal proteasomes. We could also show that the chaperone, PAN is ATPase active and can recognize and unfold ssrA-tagged substrates in presence of ATP. The degradation of ssrA-tagged substrates by the PAN-proteasome is directional, proceeding from the C-terminal tag to the N-terminus and we could show that PAN interacts strongly with the proteasome even in the absence of substrates.
### Supplementary table 1: Primer details for the genes amplified for this study. fw stands for the forward primer, rev for the reverse primer. CP stands for the proteasome core particle.

<table>
<thead>
<tr>
<th>Primer Type</th>
<th>Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>NcoI_CPa_fw</td>
<td>5’ ACTCCATGGCACCACAGATGGG 3’</td>
</tr>
<tr>
<td>BamHI_CPa_rev</td>
<td>5’ CCTGACGGATCCCTATTATCGCTGCTTTGAG 3’</td>
</tr>
<tr>
<td>NdeI_CPB_bw</td>
<td>5’ ATGGATGATGACAAATATCTAAAGGG 3’</td>
</tr>
<tr>
<td>BglII_CPB_rev</td>
<td>5’ GCACGAAGATCTTAGTTATAATGAATCTCTCTTCTGAG 3’</td>
</tr>
<tr>
<td>NdeI_CPB_T1A_fw</td>
<td>5’ GCGGCCGAACATATGGCAACTACCGTAGGAGTAGT 3’</td>
</tr>
<tr>
<td>BglII_CPB_T1A-Strep_rev</td>
<td>5’ GGAAGATCTTATTTCTGAACTGGGCTGCCAGCTGCTTG 3’</td>
</tr>
<tr>
<td>NdeI_PAN_bw</td>
<td>5’ GGCAGACATATGCGCAGTCATCTTG 3’</td>
</tr>
<tr>
<td>BamHI_PAN_rev</td>
<td>5’ GCTGGATCTTACACGAAATTGTCTC 3’</td>
</tr>
<tr>
<td>NdeI_StrepPAN_bw</td>
<td>5’ ATAGAACATATGGGAGCCACCGCAGTCGAAAAGCCTAGCAGATGCGCAGTCACCTTGTGA 3’</td>
</tr>
<tr>
<td>NdeI_PANStrep_rev</td>
<td>5’ CACGAACATATGCGCAGTCACCTTGTGAAGCCCGGATCT 3’</td>
</tr>
</tbody>
</table>

*M. acetivorans* PAN-proteasome complex
Supplementary figure 1: Electronspray ionization (ESI)-mass spectrometry analysis of *M. acetivorans* proteasome and proteasome associated nucleotidase (PAN). The theoretically expected mass for the α-subunit is 27019.6 Da, for β-subunit is 21813.1 Da and for PAN is 46965.3 Da. We observed a mass difference of 131.7 Da and 203.1 Da for the α-subunit. This mass difference could be caused by removal of the first Met, and the removal of Ala with the Met.
Chapter 5

Structure and the activation of *Methanosarcina acetivorans* small archaeal modifier proteins (SAMPs)

The determination of the SAMP1 solution structure results from a collaboration with the group of Frédéric Allain. I am responsible for all the biochemistry and preparation of all samples.

5.1 Introduction

In eukaryotes, ubiquitin is covalently attached to proteins either as a tag for protein degradation or as a regulatory post-translational modification (Hershko and Ciechanover, 1982; Hochstrasser, 2000b, 2009; Huang et al., 2004; Pickart and Fushman, 2004). The process of ubiquitin conjugation to proteins involves the sequential action of three different enzymes, namely E1 (ubiquitin activating enzyme), E2 (ubiquitin conjugating enzyme) and E3 (ubiquitin-protein ligase). E1 activates ubiquitin by adenylation of its C-terminal glycine and subsequently forms a covalent intermediate via a thioester-bond. The thioester bond is then shuttled to E2 via a trans-thiolation reaction resulting in E2-ubiquitin thioester formation. In the final conjugation step catalyzed by E3, an isopeptide-bond is formed between a substrate lysine residue and ubiquitin (Finley, 2009; Hochstrasser, 2000a, b; Schulman and Harper, 2009; Welchman et al., 2005).

The role of ubiquitin and ubiquitin-like proteins (Ubls) as post-translational protein modifiers is well established. However, an analogous role of their prokaryotic homologues in protein modification has for a long time remain elusive. In *Escherichia coli*, two sulphur transfer proteins with homology to ubiquitin, MoaD (molybdopterin synthase small subunit) (Lake et al., 2001; Rudolph et al., 2001; Schmitz et al., 2007; Zhang et al., 2010) and ThiS (involved in thiamine biosynthesis) (Begley et al., 1999a; Begley et al., 1999b; Duda et al., 2005; Lehmann et al., 2006; Leonardi and Roach, 2004; Wang et al., 2001), are activated by MoeB and ThiF respectively in an ATP-dependent manner analogous to the first step in ubiquitin-activation. Similarly, yeast Uba4p activates Urm1, a ubiquitin-related modifier that resembles the prokaryotic members (Leidel et al., 2009; Pedrioli et al., 2008; Schmitz et al., 2008; Xu et al., 2006). Apart from sharing the same activation mechanism with ubiquitin, all known Ubls, both the members involved in protein modification as well as the ones involved in sulfur transfer, are members of the β-grasp fold protein family which spans all three domains of life (Burroughs et al., 2007; Burroughs et al., 2009; Iyer et al., 2006).

The first prokaryotic ubiquitin-like modifier, Pup (prokaryotic ubiquitin-like protein), was discovered in mycobacteria and does not exhibit sequence or structural homology to ubiquitin except for a di-glycine motif near its C-terminus (Pearce et al., 2008). The modification pathway, termed pupylation, is related but chemically and enzymatically distinct. In the mycobacterial pupylation pathway, the C-terminal Gln of Pup is first deamidated by Dop (deamidase of Pup), before Pup can be covalently
Results

attached to a substrate lysine residue by the Pup ligase PafA (Striebel et al., 2009a; Sutter et al., 2010). In contrast, true ubiquitin homologs in bacteria have not been implicated in post-translational modification of proteins, but appear to function only in a role as sulfur transfer proteins. However, in archaea two small modifier proteins with homology to ubiquitin (SAMP1 and SAMP2) from Halofex volcanii were recently shown to be covalently conjugated to proteins in vivo (Darwin and Hofmann, 2010; Humbard et al., 2010), suggesting that they might play a role similar to ubiquitin in eukaryotes. Although SAMP1 and SAMP2 from H. volcanii were found to be covalently attached to proteins in vivo, (Humbard et al., 2010), the steps involved in the conjugation pathway are not known. To date, no structures have been reported for any members of the SAMP family of proteins. In this study we determined the solution structure of SAMP1 from another euryarchaeota, the acetate-utilizing methanogen Methanosarcina acetivorans, to place SAMPs in structural context with other ubiquitin-like modifier proteins. In addition, to understand the mechanism by which SAMPs are activated and conjugated to protein substrates, we cloned and characterized SAMP activating enzyme (SAE), a hypothetical protein that shows homology to eukaryotic E1 enzyme.

5.2 Experimental procedures

5.2.1 Construction of plasmids

All genes were amplified from Methanosarcina acetivorans genomic DNA. SAMP1, SAMP1ΔGG, SAMP2aΔGG and SAMP activating enzyme were cloned with NdeI and BamHI restriction sites into a modified version of the pPROEx vector (Invitrogen), in which the EheI site was replaced with an NdeI site. Strep-SAMP1, Strep-SAMP2a and Strep-SAMP2b were cloned with NdeI and BamHI restriction sites into the pET20b(+) vector. Primer sequences can be found in Supplementary Table 1. Sequences for all clones were confirmed by DNA sequencing (Microsynth).

5.2.2 Expression and purification of proteins

All SAMPs were expressed in E. coli BL21(DE3) from IPTG inducible plasmids at 20°C. SAMP activating enzyme was expressed in E. coli BL21(DE3) from an IPTG-inducible plasmid at 25°C.

His-tagged proteins were purified by Ni²⁺-affinity chromatography. After cleavage of the fusion protein with TEV protease (Invitrogen), His₆-TEV and TEV protease were removed by Ni²⁺-affinity chromatography. Further purification was attained by size-exclusion chromatography on a Superdex75 column (GE Healthcare) and the purified proteins were stored in 50 mM HEPES-KOH pH 7.5, 300 mM NaCl, 1 mM EDTA, 1 mM DTT, 10% (v/v) glycerol at -20°C.

Strep-SAMP1, Strep-SAMP2a and Strep-SAMP2b were purified using Strep-Tactin affinity chromatography (IBA) followed by size-exclusion chromatography on a
Superdex-75 (GE Healthcare). Proteins were stored in 50 mM HEPES-KOH pH 7.5, 300 mM NaCl, 1 mM EDTA, 1 mM DTT, 10% (v/v) glycerol at -20°C. For NMR measurements, SAMP1 was expressed in M9 minimal medium supplemented with 99% 15N ammonium chloride and for 13C, 15N-labeled SAMP1 with 99% 13C6 glucose. Purification of SAMP1 was carried out as described. Protein was stored in 20 mM sodium phosphate pH 6.8, 150 mM NaCl, 0.1 mM EDTA. With the exception of the amide proton exchange experiments which were conducted with 0.6 mM 15N-labeled SAMP1, all experiments were measured with 1 mM 13C, 15N-labeled SAMP1.

5.2.3 NMR measurements and structure determination

All NMR measurements were performed at 25°C on Bruker Avance 500, 600, 700, and 900 spectrometers equipped with triply tuneable cryogenic probeheads. Backbone resonance assignments were obtained automatically using the mimetic algorithm MATCH (Volk et al., 2008) within the Unio’10 package using peaklists obtained with GAPRO from 4D APSY-HACANH, 5D APSY-CBCACONH and 5D APSY-HACACONH experiments (Hiller et al., 2005) and completed manually using a 15N-resolved NOESY and two 3D 13C-resolved NOESY spectra, the latter being recorded with the 13C carrier in the aliphatic and aromatic regions respectively (Talluri and Wagner, 1996). All NOESY spectra were measured with 120 ms mixing time. Resonance assignments of aliphatic sidechains were obtained automatically using the ALASCA algorithm within GAPRO using a peaklist obtained with a 5D APSY-HC(CC-TOCSY)CONH experiment (Hiller et al., 2008). These resonance assignments were confirmed and extended manually using the peaklist derived from a 4D APSY-HCCH-COSY (Hiller et al., 2008) together with the 3D NOESY datasets. Aromatic Hδ-Cδ resonances were correlated to the Hβ-Cβ resonances with an 4D APSY-HBCB(CG)CDHD experiment (Krähenbühl and Wider, in preparation) based upon the standard 3D (HB)CB(CG)CDHD experiment (Yamazaki et al., 1993). Resonance assignments were extended to the remainder of the aromatic spin system using a 4D APSY-HCCH-COSY adapted to aromatic spin-systems (Hiller et al., 2008). The aromatic resonance assignments were then confirmed and completed with the aromatic 3D 13C-resolved NOESY.

The complete resonance assignments and the three 3D NOESY spectra furnished the input for automated peak-picking and NOE assignment using the ATNOS/CANDID program (Herrmann et al., 2002a, b) within the UNIO suite (Unio’10) together with CYANA 3.0 for structure calculation (Guntert et al., 1997). The 20 cyana conformers with the lowest target function were energy-minimized in an explicit water bath with the program AMBER 8.0 using the rna.ff99 amber forcefield (D.A. Case et al., 2006) and the generalized born solvent model (D.A. Case et al., 2006). Structures were validated with Procheck-NMR (Laskowski et al., 1993; Laskowski et al., 1996) and analyzed with MOLMOL (Koradi et al., 1996). 15N{1H}-NOEs were measured at a 1H frequency of 500 MHz using a 2 sec saturation period following a 3 sec recycle delay as described previously (Grzesiek and Bax, 1993; Renner et al., 2002). Slowly exchanging amides
were monitored by measuring a series of $[^{15}\text{N},^{1}\text{H}]-\text{HSQC}$ experiments of a sample of $^{15}\text{N}$-labeled SAMP1 lyophilized from $\text{H}_2\text{O}$ and redissolved in $^{2}\text{H}_2\text{O}$. The first measurement was made approximately 10 min after dissolution in $^{2}\text{H}_2\text{O}$ and the last experiment was measured 7 days later. Exchange rates were measured by fitting intensity curves to a monoexponential. Protection factors were determined by dividing the exchange rates by the intrinsic exchange rates of unstructured peptides corrected for pH and temperature (Bai et al., 1993).

NMR spectra were analyzed using CARA (www.nmr.ch). All figures were generated using MOLMOL and ray traced using the POV-Ray program (www.povray.org).

5.2.4 Homology modeling of SAMP activating enzyme

The amino acid sequence of SAE and the coordinates of MoeB from MoaD-MoeB complex (PDB ID:1JW9) were submitted to Easy Modeller v1.0. The resulting coordinates for SAE were used to align the protein with our determined SAMP1 structure using MOLMOL (Koradi et al., 1996).

5.2.5 Pulldown with Strep-SAMP1

10 µM Strep-SAMP1, Strep-SAMP2a or Strep-SAMP2b was incubated with 5 µM SAE in the absence or presence of 5 mM ATP for 2 hours at room temperature. Then, 20 µL Strep-Tactin Sepharose (IBA) was added followed by an additional incubation for 30 minutes at room temperature. After three washing steps with reaction buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 10% (v/v) glycerol, 20 mM MgCl$_2$), bound proteins were eluted with 1X SDS sample buffer and were analyzed on 15% SDS-PAGE followed by Coomassie staining.

5.2.6 Monitoring activation of SAMP using radiolabeled ATP

A 60 µL reaction containing 20 µM SAMP, 12 µM SAE and 100 µM $\alpha$-$^{32}\text{P}$-ATP or $\gamma$-$^{32}\text{P}$-ATP (Hartmann Analytic) in 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 20 mM MgCl$_2$ and 10% (v/v) glycerol were incubated at room temperature for 20 minutes. Following incubation, 2 µL aliquots were added to a 2 mL scintillator. The remainder of the reaction mixture was loaded onto a protein desalting spin column (Pierce) equilibrated in 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 20 mM MgCl$_2$ and 10% (v/v) glycerol to remove unbound nucleotide. After desalting, a 2 µL aliquot was added to 2 mL scintillator and scintillation counts were measured. The amount of either $\alpha$-$^{32}\text{P}$ or $\gamma$-$^{32}\text{P}$ bound per active site of SAE was calculated. Controls in which SAE was omitted were used to correct for residual free radionucleotide that was not efficiently removed by the procedure. The number of counts in these control experiments indicated that maximally 5% of initially used free radionucleotides remained in the reaction mixture.
5.2.7 ESI mass-spectrometry analysis of adenylated SAMP

10 µM SAMPs were incubated with 10 mM ATP with or without 10 µM SAE in reaction buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 10% (v/v) glycerol, 20 mM MgCl₂) for 20 minutes at room temperature. 20 µL samples were desalted with ZipTip C4 pipette tips (Millipore). Desalted samples were eluted from C4-coated pipette tips with 20 µL 50% acetonitrile in water and loaded into the mass spectrometer. Spectra were recorded on a Micromass Q-TOF Ultima API (Waters, Milford Massachusetts).

5.2.8 Phylogenetic analysis

A total of 28 representative sequences from various identified protein modifier families were obtained from the GenBank. The amino acid sequences were aligned using ClustalX2 and the phylogenetic unrooted tree was generated using the NJ method. Details about sequences are provided in Supplementary Table 2.

5.3 Results

5.3.1 Bioinformatic analysis of M. acetivorans small modifier proteins

In order to identify proteins in archaia that might act as post-translational modifiers conjugated to substrate lysine residues, we analyzed the protein coding genes of the M. acetivorans genome, selecting proteins that are both small (less than 100 residues) and contain a conserved C-terminal di-glycine motif. We identified five proteins that meet these criteria, which share different levels of homology to known Ubls (MA1595, MA1713, MA2752 that is present in duplicate also as MA2950, MA3300 and MA4086). Sequence alignment (Figure 14a) shows that three of them are M. acetivorans homologues of the recently identified ubiquitin-like modifiers from H. volcanii SAMP1 and SAMP2), and we therefore refer to them as SAMP1 (MA4086), SAMP2a (MA2752) or SAMP2a’ (MA2950) and SAMP2b (MA3300). The SAMP proteins exhibit low levels of homology to each other and to other ubiquitin-like proteins along most of their sequence, only the C-terminal stretch of residues displays higher similarity. The different SAMPs vary in length, with the SAMP1 members being about 30-40% longer than the SAMP2 members. The sequence preceeding the conserved di-glycine motif at the C-terminus also differs between the different SAMP members. SAMP1 proteins exhibit a more extended conserved motif including a notable di-proline (PPVxGG), which they share with the sulfur transfer protein MoaD involved in molybdopterin biosynthesis. SAMP2 proteins on the other hand lack this extended conserved motif.
Figure 14. Sequence analysis and secondary structure elements of *M. acetivorans* SAMP1. (a) Multiple sequence alignment of *M. acetivorans* and *H. volcanii* SAMPs with ubiquitin and ubiquitin-like proteins. The sequences were aligned using ClustalX2. Ma, *Methanosarcina acetivorans*; Hv, *Haloferax volcanii*; Hs, *Homo sapiens*; Sc, *Saccharomyces cerevisiae*; Ec, *Escherichia coli*. Residues are numbered according to *M. acetivorans* SAMP1. (b) Positions of secondary structures in SAMP1 where β strands, α helices and a 3_{10} helix are indicated by cyan arrows and red cylinders, respectively. A disordered extended loop is indicated by a broken line. (c) Deviations of $^{13}$Ca shifts from random values (Wishart et al., 1995). (d) Amide $^1$H protection factors of SAMP1. (e) $^{13}$N{H} NOE values. The negative value for Gly97 (-2.0) extends beyond the vertical axis scale.
5.3.2 Solution structure of SAMP1 and comparison with known Ubls

SAMP1 was shown to be covalently conjugated to proteins in vivo via an isopeptide bond to lysine residues of substrate proteins (Humbard et al., 2010). Like ubiquitin and related modifier proteins, SAMPs are small, ranging from 70 to 100 residues in length, and are, therefore, amenable to NMR structural analysis. To study the SAMP proteins in vitro, we cloned SAMP1, SAMP2a and SAMP2b from M. acetivorans, expressed them heterologously in E. coli and purified them. For NMR analysis, SAMP1 was expressed in isotopically labeled form and showed well-dispersed \([^{15}N,^{1}H]\)-HSQC and \([^{13}C,^{1}H]\)-HSQC spectra indicative of a folded domain. On the basis of APSY (Hiller et al., 2005) spectroscopy and automated assignment using the algorithm MATCH (Volk et al., 2008), nearly complete backbone resonance assignments were obtained. Sidechain assignments were obtained with sidechain APSY experiments. The assignments were extended and completed with 3D \(^{15}N\)- and \(^{13}C\)-resolved NOESY experiments. This resulted in complete resonance assignments with the exception of the exchangeable side chain resonances of Lys and Arg residues and the side chain resonances of the two N-terminal residues Gly and His which are due to the vector-derived N-terminal expression tag. The chemical shifts of the \(C_{\beta}\) and \(C_{\gamma}\) resonances in the five Pro residues indicated that they are with high probability in the trans conformation (Schubert et al., 2002) and this was confirmed by strong \(H_{N}-H_{\delta}\) and \(H_{\alpha}-H_{\delta}\) NOEs (Wuthrich, 1986). The upfield \(C_{\beta}\) shifted position of the single Cys60 at 27.0 ppm indicated that it was reduced (Sharma and Rajarathnam, 2000). \(^{13}C_{\alpha}\) shift deviations from random coil values \(\Delta^{13}Ca\) are consistent with the secondary structure boundaries determined by structure determination with alpha helical segments indicated for 11-14 and 28-38 and \(\beta\)-strands identified for residues 2-8, 20-25, 64-67 and 86-91 (Figure 14b,c). In addition, a 3\(_{10}\) helix comprising residues 40-42 is present in SAMP1. With the exception of \(\alpha1\), the amide groups of these secondary structure elements show strongly retarded \(^{1}H\) exchange rates (Figure 14d). \(^{15}N\{^{1}H\}\)-heteronuclear NOEs indicate that the backbone has limited picosecond to nanosecond mobility except for the N-terminal dipeptide, the C-terminal tetrapeptide and residues 49-58 (Figure 14e).

The nearly complete resonance assignments formed the basis for automated peak-picking and NOE assignment of the three 3D NOESY experiments using UNIO-ATNOS/CANDID (Herrmann et al., 2002a, b) combined with structure calculation using CYANA (Guntert et al., 1997). The 20 structures with the lowest target functions were energy minimized with AMBER (D.A. Case et al., 2006) resulting in a well-defined structure of SAMP1 as documented by Table 1 and Figure 15a. The SAMP1 structure (Figure 15b) consists of a four-stranded \(\beta\)-sheet packing against an \(\alpha\)-helix classifying SAMP1 as a member of the \(\beta\)-grasp fold family. The strand order within the sheet is 2143, which is the same as in ubiquitin or other ubiquitin-like proteins. SAMP1 possesses two \(\alpha\)-helical segments. Helix \(\alpha1\) connects strands \(\beta1\) and \(\beta2\), and helix \(\alpha2\) extends across the face of the \(\beta\)-sheet. It is flanked by strand \(\beta2\) at its N-terminal end and a short 3\(_{10}\) helical segment at its C-terminal end. The remaining strands \(\beta3\) and \(\beta4\) are connected by
extended loops with irregular coil conformation. The fold is stabilized by interacting sets of hydrophobic residues in the interface between the β-sheet and α-helix 2.

We submitted the NMR conformer of SAMP1 with lowest target function energy to the fold recognition program DALI (Dietmann et al., 2001; Holm and Sander, 1993) in a search for structural homologues. The result showed that SAMP1 has a global fold similar to MoaD and Urm1 (Table 2). Structural alignment of SAMP1 with MoaD, Urm1, CysO and ubiquitin revealed that SAMP1 possesses the same topology as these.

**Figure 15: NMR structure of SAMP1.** (a) Superposition of the backbone of the 20 best energy-refined CYANA conformers; β-sheets in cyan, α-helices in orange and the extended loop of residues 49-58 in yellow. (b) Ribbon representation of SAMP1. (c) Ribbon representation of structural homologues of SAMP identified by a DALI alignment. PDB codes for the coordinates are given in Table 2.
Small archaeal modifier proteins (SAMPs)

Table 1. Input for the structure calculations and characterization of the energy-minimized NMR structures of SAMP1 at pH 6.8 and 25°C

<table>
<thead>
<tr>
<th>NMR upper distance limits(^a)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>3946</td>
</tr>
<tr>
<td>Intraresidual</td>
<td>614</td>
</tr>
<tr>
<td>Sequential</td>
<td>(</td>
</tr>
<tr>
<td>Medium range</td>
<td>(1&lt;</td>
</tr>
<tr>
<td>Long range</td>
<td>(</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Structure statistics (^b)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Residual NOE violations</td>
<td></td>
</tr>
<tr>
<td>Number (\geq 0.1 \text{ Å})</td>
<td>28 ± 4</td>
</tr>
<tr>
<td>Maximum, Å</td>
<td>0.24±0.09</td>
</tr>
<tr>
<td>Deviations from idealized geometry</td>
<td></td>
</tr>
<tr>
<td>Bond lengths, Å</td>
<td>0.00375 ± 0.00008</td>
</tr>
<tr>
<td>Bond angles, °</td>
<td>1.449±0.017</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>rmsd to the mean coordinates, Å (^c)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>bb (1–93)</td>
<td>0.45 ± 0.24</td>
</tr>
<tr>
<td>All heavy atoms (1–93)</td>
<td>0.79 ± 0.28</td>
</tr>
<tr>
<td>bb (1–48,59–93)(^d)</td>
<td>0.20 ± 0.05</td>
</tr>
<tr>
<td>All heavy atoms (1–48,59–93)</td>
<td>0.52 ± 0.04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ramachandran plot statistics (^e)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Most favored regions (%)</td>
<td>70.9</td>
</tr>
<tr>
<td>Additionally allowed regions (%)</td>
<td>25.5</td>
</tr>
<tr>
<td>Generously allowed regions (%)</td>
<td>3.0</td>
</tr>
<tr>
<td>Disallowed regions (%)</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\(^a\) NOE upper distance restraints are divided into categories based on the residue positions of the two atoms involved in each restraint.
\(^b\) The average value for the 20 energy-minimized conformers with the lowest residual CYANA target function values and the standard deviation among them are given.
\(^c\) bb indicates the backbone atoms N, C^\(\alpha\) and C'. The numbers in parentheses indicate the residues for which the rmsd was calculated. RMSD values were calculated with MOLMOL (Koradi et al., 1996).
\(^d\) The ordered polypeptide segments excluding the extended loop of residues 49-58.
\(^e\) As determined by PROCHECK (Laskowski et al., 1993).
homologues with structural differences apparent mostly in the loop regions (Figure 15c).
In *M. acetivorans* SAMP1, there is an extended loop located N-terminally of strand β3,
which carries two surface exposed lysine residues, one at the start (Lys49) and one in the
middle (Lys53). Interestingly, in SAMP1 from *H. volcanii* this loop is predicted to be
somewhat shorter based on sequence comparison and it does not feature lysines at the
equivalent positions, suggesting that this loop might play a unique role in *M. acetivorans*.

**Table 2. SAMP1 structural homologues**

<table>
<thead>
<tr>
<th>PDB code</th>
<th>Z-score</th>
<th>rmsd</th>
<th>annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1V8C</td>
<td>15.6</td>
<td>1.3</td>
<td><em>Thermus thermophilus</em> MoaD</td>
</tr>
<tr>
<td>3DWG</td>
<td>13.3</td>
<td>1.9</td>
<td><em>Mycobacterium tuberculosis</em> CysO</td>
</tr>
<tr>
<td>1FMA</td>
<td>10.3</td>
<td>1.8</td>
<td>MoaD from MoaE-MoaD complex</td>
</tr>
<tr>
<td>2PKO</td>
<td>10.1</td>
<td>2.1</td>
<td><em>Saccharomyces cerevisiae</em> Urm1</td>
</tr>
<tr>
<td>1JW9</td>
<td>9.7</td>
<td>1.8</td>
<td>MoaD from MoeB-MoaD complex</td>
</tr>
<tr>
<td>1ZUD</td>
<td>4.2</td>
<td>2.2</td>
<td>ThiS from ThiF-ThiS complex</td>
</tr>
<tr>
<td>1UBQ</td>
<td>4.1</td>
<td>2.7</td>
<td><em>Homo sapiens</em> Ubiquitin</td>
</tr>
</tbody>
</table>

In addition the surface of *M. acetivorans* SAMP1 features a convex groove
involving α2 and three loops located between 3₁₀-β3, β2-α2 and β3-β4. The floor of the
groove include residues Ile29, Asp30, Leu33, Ser34, Asp37 from α2, and Leu59 and
Leu78 from the loops between 3₁₀-β3 and β3-β4, respectively (Figure 16a,b). Ile72,
Glu76, Glu79 and Glu26, Lys27 from the loops between α2-β2 and β3-β4 make the wall
on one side, whereas residues Glu55, Ile56 and Leu57 from the loop between 3₁₀-β3 form
the other wall of the groove. Interestingly, this convex groove is not present in MoaD, a
structural homolog of SAMP1 based on DALI alignment. Due to the shorter loop
segments in SAMP1 from *H. volcanii*, the surface there might also not feature such a
pronounced groove. In ubiquitin, the so-called “hydrophobic patch” is one of the surface
features that mediate binding to interaction partners. The structure of the bacterial MoaD-
MoeB complex also shows interaction between the ubiquitin-like MoaD and the sulfurase
MoeB via hydrophobic surface residues. Inspection of the corresponding location on the
surface of SAMP1 shows a similar hydrophobic patch suggesting that SAMP1 may form
a similar interaction with its activating enzyme. We return to this observation later.

### 5.3.3 Putative SAMP activating enzyme interacts with SAMPs

Ubiquitin or ubiquitin-like modifiers as well as the small ubiquitin-like proteins involved
in sulfur transfer (MoaD or ThiS), must be enzymatically activated to either be
conjugated to substrate proteins or to accept sulfur. In eukaryotes, ubiquitin and Ubls are
activated by E1 family proteins (Hochstrasser, 2000b), which also exhibit homology to
the activating enzymes in the sulfur transfer pathways (MoeB or ThiF). The *M. acetivorans*
genome codes for one protein, MA0255 (27 kDa) that shows sequence
Small archaeal modifier proteins (SAMPs)

similarity to eukaryotic E1 and to bacterial MoeB (molybdopterin synthase sulfurylase) (Figure 16c).

In order to investigate the role of this putative SAMP-activating enzyme (SAE) in the SAMP protein conjugation pathway in vitro, we cloned the SAE protein from *M. acetivorans*, expressed it heterologously in *E. coli* and purified it. To test for possible interaction between SAE and the various SAMP proteins, we generated Strep-tagged versions of the potential modifiers, namely Strep-SAMP1, Strep-SAMP2a, Strep-SAMP2b and Strep-SAMP3 (MA1713) and performed pulldown assays where the purified Strep-SAMPS were incubated with purified SAE in presence or absence of ATP. After incubation with SAE, the Strep-tagged SAMPS were immobilized on Strep-Tactin-decorated sepharose beads and, after multiple washing steps, eluted using SDS. The eluants were tested for retained SAE protein by coomassie-stained SDS-PAGE. As a control for non-specific interaction of SAE with the Strep-Tactin-decorated sepharose beads, SAE was incubated with the resin in the absence of any Strep-SAMP protein under the same conditions. Both in the absence and in the presence of ATP, SAE interacts strongly with all SAMPS (Figure 17a,b). Interestingly, upon incubation of Strep-SAMP1 with SAE in the presence of ATP, we observed an additional band in SDS-PAGE running just below the Strep-SAMP1 band (Figure 17a, lane 1). We predicted that this additional band might represent the adenylated (activated) form of SAMP1. The interaction of SAE with Strep-SAMP2a appears to be stronger in the presence of ATP (Figure 17b, lane 3) compared with experiments conducted in the absence of nucleotide (Figure 17b, lane 4). However, for Strep-SAMP1 or Strep-SAMP2b no such difference is detected (Figure 17a, lanes 1,2, Figure 17b, lanes 6,7). Additional bands were not observed in the presence of ATP for Strep-SAMP2a, 2b and 3, possibly because the migration behavior of the potential activated forms is the same as for the non-activated forms. The second band observed above the Strep-SAMP2a and 2b bands stems from an impurity in the protein preparation.
Figure 16: SAMP1 surface properties and interaction model with modeled SAE. (a) Surface representation of the SAMP1 convex groove in light blue. (b) Electrostatic surface representation of SAMP1. The surface color represents the magnitude of electrostatic potential: red, negative; blue, positive; white, neutral. (c) Sequence alignment of *M. acetivorans* SAMP-activating enzyme with *E. coli* MoeB (molybdopterin synthase sulfurylase). Alignment was performed using ClustalX2. * indicates conserved residues. Red triangles indicate the conserved active site residues responsible for coordination and binding of Mg$^{2+}$ and ATP. (d) Model of SAMP1 interaction with modeled SAE obtained by structure
Small archaeal modifier proteins (SAMPs)

homology modelling using the MoaD-MoeB complex crystal structure (PDB ID: 1JW9). The SAE model is shown in gold, SAMP1 interacting residues in magenta with the C-terminal extension in red. Grey line connects Cα atoms of residues 171 and 179 of SAE. These residues are excluded from the model because no electron density was reported for these residues of MoeB in the crystal structure of the complex. (e) Ribbon representation of SAMP1 and modeled SAE in the same color-coding and orientation as in d. The sidechains of interacting residues from SAMP1 are shown in stick representation.

Based on the results described above, SAE can form a complex with SAMP1, which due to the high homology of SAE to MoeB would be expected to be closely similar to the MoaD-MoeB complex solved for the bacterial system. We therefore homology-modeled SAE using the coordinates of MoeB from the MoaD-MoeB complex (PDB ID: 1JW9) and aligned our determined SAMP1 structure with the coordinates of MoaD in the complex (Figure 16d,e). A close inspection of the interface region of this model and comparison to the MoaD-MoeB complex shows that SAMP1 features a related, pronounced hydrophobic patch including residues Phe9, Ala10, Leu66, Leu91 that in addition with Asn11, Gly61, Ser62, Asn64, Asn71, Arg73 and His74 might serve as an interface for interaction with SAE. At the center of this hydrophobic patch the residues Phe9, Ala10 and Leu91 make contacts with SAE. Two of these three residues at corresponding positions are identical in MoaD. In that complex, the C-terminal extension from residue 92-97 of MoaD extends into the active site cavity formed by the activating enzyme during the activation process. SAMP1 possesses a similar extension containing mostly hydrophobic residues which fits into the corresponding cavity of SAE. Furthermore, Arg13 and Glu14 of SAMP1 are in corresponding positions to residues Arg11 and Glu12 of MoaD, which make ionic interactions and a hydrogen bond with MoeB (Lake et al., 2001).
Results

Figure 17. SAMP activating enzyme interacts with *M. acetivorans* SAMPs. Strep-SAMP1, Strep-SAMP2a or Strep-SAMP2b (10 µM each) were each incubated with SAE (5 µM) ± ATP (10 µM) and purified using Strep-tactin sepharose beads. SAE coeluting with Strep-SAMP1, Strep-SAMP2a or Strep-SAMP2b was analyzed by coomassie-stained SDS-PAGE. As a control SAE was incubated in absence of any Strep-SAMP. The observed faint band is due to non-specific binding of SAE towards streptavidin beads. (a) Interaction of SAE with Strep-SAMP1 with or without ATP (lanes 1 and 2), (b) Interaction of SAE with Strep-SAMP2a and Strep-SAMP2b ± ATP (lanes 1, 2, 4 and 5). The band observed above the Strep-SAMP2a and Strep-SAMP2b band is due to impurity from the protein preparation. (c) Time course of formation of adenylated Strep-SAMP1. Strep-SAMP1 (10 µM) was incubated with SAE (10 µM) and ATP (10 mM), and samples were taken at the indicated time points and analyzed by coomassie-stained SDS-PAGE.

5.3.4 SAE activate and adenylates SAMPs

In the molybdopterin synthesis pathway, MoeB activates and adenylates MoaD in the presence of ATP (Lake et al., 2001; Schmitz et al., 2007). SAE shares homology to MoeB and contains conserved residues essential for proper coordination of Mg\(^{2+}\) and ATP during the activation process (Figure 1b, arrows). To assess if SAE can activate and adenylate SAMPs in presence of ATP, we analyzed the activation of SAMPs using radiolabeled \(\alpha\)\(^{32}\)P-ATP or \(\gamma\)\(^{32}\)P-ATP. SAMP1, Strep-SAMP2a, Strep-SAMP2b or Strep-SAMP3 was incubated with SAE in presence of radiolabeled ATP and, following the incubation, unbound nucleotide was removed by gel filtration through a desalting spin column. As a control SAE was incubated with radiolabeled ATP in the absence of a SAMP binding partner. In the presence of SAMP1, Strep-SAMP2a or Strep-SAMP2b
and α-\(^{32}\)P-ATP, there was an increase of \(^{32}\)P bound per active site of SAE compared to experiments conducted in the absence of SAMPs, confirming that SAE can activate SAMPs in presence of nucleotide (Figure 18a,b; black bars). There was a small amount of ATP bound to SAE in absence of any SAMP binding partner (Figure 18a,b), which was expected as SAE is predicted to feature an ATP binding site according to sequence alignment with other E1 enzymes. However, when SAE and SAMPs were incubated in the presence of γ-\(^{32}\)P-ATP, there was no increase in \(^{32}\)P bound per active site of SAE, confirming that during the activation of the SAMPs, the γ-phosphate is released, presumably in the form of pyrophosphate (Figure 18a,b; grey bars). Together these data conclusively show that MA0255 acts as the SAMP-activating enzyme by adenylation the SAMPs.

In molybdopterin synthesis or in the ubiquitination pathway, adenylation occurs at the conserved C-terminal glycine (Finley, 2009; Lake et al., 2001; Schmitz et al., 2007; Schulman and Harper, 2009). As the SAMPs also contain the conserved di-glycine motif at the C-terminus, the same mode of activation would be expected. To confirm this, we generated truncated versions of SAMP1 and SAMP2a (SAMP1ΔGG, SAMP2aΔGG), where the conserved diglycine motif was removed. Upon incubation of either SAMP1ΔGG or SAMP2aΔGG in the presence of SAE and \(α-\(^{32}\)P-ATP, there was an increase of \(^{32}\)P bound per active site of SAE compared to experiments conducted in the absence of SAMPs, confirming that SAE can activate SAMPs in presence of nucleotide (Figure 18a,b; black bars). There was a small amount of ATP bound to SAE in absence of any SAMP binding partner (Figure 18a,b), which was expected as SAE is predicted to feature an ATP binding site according to sequence alignment with other E1 enzymes. However, when SAE and SAMPs were incubated in the presence of γ-\(^{32}\)P-ATP, there was no increase in \(^{32}\)P bound per active site of SAE, confirming that during the activation of the SAMPs, the γ-phosphate is released, presumably in the form of pyrophosphate (Figure 18a,b; grey bars). Together these data conclusively show that MA0255 acts as the SAMP-activating enzyme by adenylation the SAMPs.

In molybdopterin synthesis or in the ubiquitination pathway, adenylation occurs at the conserved C-terminal glycine (Finley, 2009; Lake et al., 2001; Schmitz et al., 2007; Schulman and Harper, 2009). As the SAMPs also contain the conserved di-glycine motif at the C-terminus, the same mode of activation would be expected. To confirm this, we generated truncated versions of SAMP1 and SAMP2a (SAMP1ΔGG, SAMP2aΔGG), where the conserved diglycine motif was removed. Upon incubation of either
SAMP1ΔGG or SAMP2aΔGG with SAE and α-32P-ATP, there was no increase in 32P bound per active site of SAE (Figure 18a,b), confirming that activation of SAMPs occurs at the C-terminal glycine as for ubiquitin or MoaD. While the experiments with radiolabeled ATP demonstrated activation of SAMP1 and SAMP2a/2b, they do not provide information about the exact nature of the activated intermediate. Based on our results with γ-32P-ATP and in analogy to ubiquitin activation, we expect SAMP adenylation to take place. To confirm the formation of the adenylated intermediate, we performed ESI mass-spectrometry of the products of the adenylation reaction described previously. SAMP1, Strep-SAMP2a and Strep-SAMP2b were incubated with ATP and SAE, and analyzed by ESI mass-spectrometry. In the mass spectra of samples obtained with adenylation reactions of SAMP1 and Strep-SAMP2b,

![Figure 19. Formation of an adenylated intermediate shown by ESI mass-spectrometry.](image_url)

SAMP1ΔGG or SAMP2aΔGG with SAE and α-32P-ATP, there was no increase in 32P bound per active site of SAE (Figure 18a,b), confirming that activation of SAMPs occurs at the C-terminal glycine as for ubiquitin or MoaD. While the experiments with radiolabeled ATP demonstrated activation of SAMP1 and SAMP2a/2b, they do not provide information about the exact nature of the activated intermediate. Based on our results with γ-32P-ATP and in analogy to ubiquitin activation, we expect SAMP adenylation to take place. To confirm the formation of the adenylated intermediate, we performed ESI mass-spectrometry of the products of the adenylation reaction described previously. SAMP1, Strep-SAMP2a and Strep-SAMP2b were incubated with ATP and SAE, and analyzed by ESI mass-spectrometry. In the mass spectra of samples obtained with adenylation reactions of SAMP1 and Strep-SAMP2b,
we observed an additional peak whose mass corresponded to AMP bound to SAMP1 (Figure 19a) or Strep-SAMP2a (Figure 19b), respectively. However, when SAMP1 or Strep-SAMP2a were incubated with ATP in the absence of SAE, the peak corresponding to the adenylated form of SAMP was not observed (Figure 19a and b). This confirmed that the nucleotide-bound states of SAMP1 and SAMP2a are the adenylated form. As a control, we incubated SAMP1ΔGG and SAMP2aΔGG with SAE and ATP, and analyzed the product with ESI mass-spectrometry. The absence of the adenylated peak for SAMP1ΔGG and SAMP2aΔGG (Figure 19a,b) supported our conclusion that the C-terminal glycine of SAMPs is activated and adenylated by SAE.

The appearance of the adenylated intermediate of Strep-SAMP1 can be followed by coomassie-stained SDS-PAGE, since the migration behavior of the sampylated form of Strep-SAMP1 is different from the non-activated form. Strep-SAMP1 was incubated with SAE and ATP, and samples were taken at the indicated time points and were analyzed using coomassie-stained SDS-PAGE. The adenylated Strep-SAMP1 runs slightly below the SAMP1 band. The band corresponding to the adenylated form of Strep-SAMP1 was not observed when ATP or SAE was omitted from the reaction, or when the deletion variant SAMP1ΔGG was used (Figure 17c).

5.4 Discussion

Post-translational modifications of proteins occur in all cells and serve as a means of diversification and regulation of many cellular functions (Walsh and Jefferis, 2006; Wickliffe et al., 2009). The use of macromolecular tags such as ubiquitin, SUMO (Morris, 2010; Thomson and Guerra-Rebollo, 2010; Wilkinson and Henley, 2010) or NEDD8 (Parry and Estelle, 2004; Xirodimas, 2008) has been well-established as a feature of eukaryotic cells. The existence of analogous small protein modification pathways has only recently been shown for prokaryotes. In mycobacteria and related species, the small protein Pup (prokaryotic ubiquitin-like protein) is conjugated to lysine residues of substrate proteins by a pathway chemically and evolutionarily distinct from ubiquitination (Striebel et al., 2009a; Sutter et al., 2010). It has been suggested that the evolutionary precursors of ubiquitin and ubiquitin-like proteins in prokaryotes are small proteins involved in sulfur transfer, whose activation by adenylation of a C-terminal diglycine motif bears resemblance to the activation by E1-type enzymes (Duda et al., 2005; Huang et al., 2004; Lake et al., 2001; Lehmann et al., 2006; Leidel et al., 2009; Schmitz et al., 2008). Strikingly, two archaeal protein modifiers with homology to prokaryotic sulfur transfer proteins and to ubiquitin, SAMP1 and SAMP2 from H. volcanii, were found to be covalently conjugated to proteins in vivo (Humbard et al., 2010), indicating the existence of ubiquitin-like protein modification in the archaeal lineage. In our study, we identified homologs of ubiquitin-like small archaeal modifier proteins (SAMPs) from the methanogenic archaeon M. acetivorans, confirmed by NMR structure determination that these proteins exhibit the ubiquitin-like beta-grasp fold and reconstituted the energy-dependent activation of several SAMPs by an E1-like enzyme.
we termed SAE (SAMP activating enzyme). This provides the first experimental evidence of the initial steps involved in the sampaylation pathway in archaea. However, it is still an open question whether SAMPs, like their bacterial homologs, might not also be involved in sulfur transfer pathways. For example, in *M. acetivorans*, the gene coding for SAMP1 is located directly upstream of the gene coding for the molybdopterin biosynthesis protein MoeA2, and phylogenetic analysis of Ubls places SAMP1 in the same branch with MoaD. *M. acetivorans* SAMP2a and SAMP2b are the closest homologues of *H. volcanii* SAMP2, and they cluster together with ThiS, the sulfur transfer protein involved in thiamine biosynthesis (Wang et al., 2001) (Figure 20). This suggests that *M. acetivorans* SAMP1 and SAMP2a/2b might be involved in different pathways, although they share the same activation mechanism. Interestingly, the protein conjugates identified for SAMP1 and SAMP2 in *H. volcanii* largely belong to pathways associated with sulfur metabolism, for example the putative sulfur-activating enzyme MoeB (the homologue of SAE in our study), the homolog to bacterial MobB involved in formation of molybdopterin nucleotide, the homologue of bacterial MoeA that mediates ligation of Mo to molybdopterin or homologs of SuB/D involved in iron-sulfur cluster formation (Daniels et al., 2008; Outten et al., 2003). Notably, it has been shown previously that bacterial MoaE, the archaeal homologue of which has been identified to be conjugated to SAMP1 in *H. volcanii*, can form a covalent bond with its binding partner MoaD (a close SAMP1 homologue) via MoaD’s activated C-terminal glycine and a lysine residue in MoaE after prolonged incubation *in vitro*. The structure of the covalent MoaE/MoaD complex of *E. coli* has been solved and was shown to be highly similar to the native MoaE/MoaD non-covalent complex (Rudolph et al., 2001). It is possible that the conjugation of ubiquitin-like proteins to substrate proteins developed as a side reaction from the interaction of sulfur-transfer activated carrier proteins with their non-covalent binding partners, evolving into a regulatory feature on a specific subset of proteins related to sulfur metabolism. It remains to be established if archaeal SAMPs might have retained both types of activities, sulfur transfer and protein conjugation. This question will only be answered when enzymes interacting with activated SAMPs downstream of the SAE enzyme are identified.

In order to place *M. acetivorans* SAMP1 in the structural context of ubiquitin-like protein modifiers, we determined the solution structure of SAMP1 by NMR. Our analysis revealed that the overall topology of SAMP1 assigns it as a member of the β-grasp fold family with a four-stranded anti-parallel β-sheet packing against an α-helix. The strand order within the sheet is 2143, which is the same as in ubiquitin or other ubiquitin-like proteins such as Urm1 or MoaD. Given the structural homology and the close phylogenetic relationship of SAMP1 with bacterial MoaD as well as the homology of SAMP activating enzyme SAE to MoeB, we predict that the interaction of the SAMP1 C-terminus with SAE is likely analogous to the interaction of MoaD with MoeB during the activation pathway (Lake et al., 2001; Schmitz et al., 2007). However, certain amino acid sequence and structural features of *M. acetivorans* SAMP1 differ from MoaD, and are not even conserved between SAMP1 and SAMP2 from the same organism, or
Small archaeal modifier proteins (SAMPs)

Figure 20. Phylogenetic relationship between SAMPs and ubiquitin-like proteins. The unrooted tree was generated using the NJ method. Ag, Anopheles gambiae; An, Aspergillus nidulans; At, Arabidopsis thaliana; Ce, Caenorhabditis elegans; Dd, Dictyostelium discoideum; Dm, Drosophila melanogaster; Ec, Escherichia coli; Gg, Gallus gallus; Hs, Homo sapiens; Hv, Haloferax volcanii; Il, Idiomarina loihiensis; Mm, Mus musculus; Ma, Methanosarcina acetivorans; Mm, Methanosarcina mazei; Pf, Pyrococcus furiosus; Sc, Saccharomyces cerevisiae; Tb, Trypanosoma brucei; Tt, Thermus thermophilus.

between *M. acetivorans* SAMP1 and *H. volcanii* SAMP1. One such feature is an extended loop comprising residues 50-60 of *M. acetivorans* SAMP1 with a surface exposed lysine residue (Lys51) at its tip. This loop exhibits a range of conformations
consistent with both the rapid exchange of its backbone amides and the reduced $^{1}H^{15}N$-NOE values. Such differentiating features might play an important role in determining the nature of the downstream events after the activation. Another characteristic surface feature of SAMP1 is the presence of a convex groove with an “acidic” face on one side and a “hydrophobic” on the other. Many Ubls like the small ubiquitin-like modifier (SUMO) and yeast autophagy-related protein 8 (ATG8) possesses a groove involving the α-helix and the β-sheet (Winget and Mayor, 2010). The α/β groove of SUMO has been shown to function as a binding surface for SUMO interacting motif (SIM), involved in different processes like DNA repair or transcription repression (Hecker et al., 2006), whereas protein p62 interacts with the groove of LC3 (human homolog of ATG8) during the autophagy process. Interestingly, ubiquitin has not been shown to use the α/β groove binding site extensively for ubiquitin-binding motif (Winget and Mayor, 2010). However, whether the groove of SAMP1 serves as a binding site for interacting partner involved in different pathways is yet to be investigated.

To investigate the potential specificity of SAMP activating enzyme toward individual SAMPs, we tested the interaction of SAE with the various Strep-tagged SAMPs. While SAE interacted with all four SAMPs tested (SAMP1, SAMP2a, SAMP2b, and SAMP3), even in the absence of nucleotide, the adenylated intermediate was only detected for SAMP1 and SAMP2a/2b both by mass spectrometry and in the assay with radiolabeled ATP. This suggests that SAE either shows specificity toward a subset of SAMPs or that the adenylated intermediate is considerably less stable in case of SAMP3. SAMP4 was not tested as the expression clone did not produce protein. We did not observe ATP dependent activation of SAMP1ΔGG or SAMP2aΔGG by SAE, confirming that activation takes place at the C-terminal glycine residue, and accordingly, we could not detect the adenylated intermediate with the deletion variants by mass spectrometry. For SAMP2a, there is an additional non-adenylated form with a molecular mass of 18 Da less than the calculated mass for SAMP2a (Figure 19b). The reason for the two different populations of SAMP2a observed by ESI mass-spectroscopy is not clear. Interestingly, the additional non-adenylated peak of SAMP2a is dependent on the presence of SAE (Figure 19b), suggesting the enzymatic release of a water molecule from SAMP2a. One possible explanation would be the formation of an internal thioester between the C-terminal carboxylate and an internal Cys residue, a product that could only be formed after adenylation of the C-terminal glycine by SAE.

In conclusion, SAMP1 has a β-grasp fold as revealed by the solution structure, and based on structural alignment and phylogenetic tree relationships SAMP1 is closely related to E.coli sulfur transfer protein MoaD. The identified SAMP activating enzyme (SAE) interacts with SAMPs strongly independent of ATP. SAMPs are activated by adenylation at their C-terminal glycine by SAMP activating enzyme in presence of ATP. However, downstream steps involved in protein modification at lysine residues in a process called SAPtylation or in pathways involving sulfur transfer have yet to be elucidated, to shed further light on the role of SAMPs in archaea.
Small archaeal modifier proteins (SAMPs)

**Supplementary**

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAMP1 fp NdeI</td>
<td>5’ ATA TCA <strong>CAT</strong> ATG GCT GAA GTA AAA GTG AAG TTA TTT GCA AAT CTA C<strong>G</strong>T <strong>GAG</strong> 3’</td>
</tr>
<tr>
<td>SAMP1 rp BamHI</td>
<td>5’ TGT <strong>GGA</strong> TCC TCA GCC CCC CGA AAC AGG TGG TAG A<strong>AAT</strong> C<strong>G</strong>T <strong>GAG</strong> 3’</td>
</tr>
<tr>
<td>SAMP1ΔGG rp BamHI</td>
<td>5’ TGT <strong>GGA</strong> TCC TCA CGA A<strong>AAC</strong> AGG TGG TAG A<strong>AAT</strong> CCC GAT TTC ATC <strong>ATC</strong> 3’</td>
</tr>
<tr>
<td>SAMP2a fp NdeI</td>
<td>5’ ATT GCC <strong>CAT</strong> ATG GTG AGT AAA AAA GTA CAC GTG CAA ATC CAC GCT GG <strong>GG</strong> 3’</td>
</tr>
<tr>
<td>SAMP2aΔGG rp BamHI</td>
<td>5’ GAT <strong>GGA</strong> TCC CTA TGA A<strong>AAC</strong> AAC TCT CAT GAT CGT A<strong>AAT</strong> CCC GAT TTC ATC <strong>ATC</strong> 3’</td>
</tr>
<tr>
<td>Strep-SAMP1 fp NdeI</td>
<td>5’ ATT GCA <strong>CAT</strong> ATG TGG AGC CAC CCG CAG TTC GAA AAA GCT AGC AGT ATG GCA GAA GTA AAA GTG AAG 3’</td>
</tr>
<tr>
<td>Strep-SAMP2a fp NdeI</td>
<td>5’ ATT GCC <strong>CAT</strong> ATG TGG AGC CAC CCG CAG TTC GAA AAA GCT AGC AGT GTG AGT AAA AAA GTA CAC GTG CAA 3’</td>
</tr>
<tr>
<td>Strep-SAMP2a rp BamHI</td>
<td>5’ GAT <strong>GGA</strong> TCC CTA ACC CCC TGA A<strong>AAC</strong> AAC TCT CAT GAT CGT A<strong>AAT</strong> CCC GAT TTC ATC <strong>ATC</strong> 3’</td>
</tr>
<tr>
<td>Strep-SAMP2b fp NdeI</td>
<td>5’ ATT GCC <strong>CAT</strong> ATG TGG AGC CAC CCG CAG TTC GAA AAA GCT AGC AGT TTG AGT AAA AGG TTA TAC <strong>TAC</strong> 3’</td>
</tr>
<tr>
<td>Strep-SAMP2b rp BamHI</td>
<td>5’ TGA <strong>GGA</strong> TCC TCA GCC ACC TGT GAC CAC TTT AAG TAT TGT GAG TCT ATC GGA <strong>GGA</strong> 3’</td>
</tr>
<tr>
<td>Strep-SAMP3 fp NdeI</td>
<td>5’ ATT GTA <strong>CAT</strong> ATG TGG AGC CAC CCG CAG TTC GAA AAA GCT AGC AGT CTG AAA ATC CAC GTA AAG TTC C<strong>CTT</strong> 3’</td>
</tr>
<tr>
<td>Strep-SAMP3 rp BamHI</td>
<td>5’ TAT <strong>GGA</strong> TCC TCA GCC CCC GGC A<strong>AAC</strong> TGG TGG T<strong>TAC</strong> 3’</td>
</tr>
<tr>
<td>SAMP activating enzyme fp NdeI</td>
<td>5’ ATT GCC <strong>CAT</strong> ATG AAT GAT TTA GAA CGT GAA AAA TAC AGC AGG <strong>AGG</strong> 3’</td>
</tr>
<tr>
<td>SAMP activating enzyme rp BamHI</td>
<td>5’ GAT <strong>GGA</strong> TCC TCA TTT ATT T<strong>TTC</strong> ATC CGA TTT GTT AAG AAA <strong>AAA</strong> 3’</td>
</tr>
</tbody>
</table>

**Supplementary table 1:** Primer sequences of proteins used in this study. Restriction enzymes used are highlighted in bold. fp and rp stands for the forward primer and the reverse primer, respectively.
Supplementary table 2: Ubiquitin protein superfamily members used in this study

<table>
<thead>
<tr>
<th>Protein ID</th>
<th>Species</th>
<th>Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ma SAMP1</td>
<td><em>Methanosarcina acetivorans</em></td>
<td>Q8TIQ6</td>
</tr>
<tr>
<td>Ma SAMP2a</td>
<td><em>Methanosarcina acetivorans</em></td>
<td>Q8TMB2</td>
</tr>
<tr>
<td>Ma SAMP2b</td>
<td><em>Methanosarcina acetivorans</em></td>
<td>Q8TKU4</td>
</tr>
<tr>
<td>Hv SAMP1</td>
<td><em>Halofex volcanii</em></td>
<td>D4GUF6</td>
</tr>
<tr>
<td>Hv SAMP2</td>
<td><em>Halofex volcanii</em></td>
<td>D4GZE7</td>
</tr>
<tr>
<td>Ec MoaD</td>
<td><em>Escherichia coli</em></td>
<td>CAA49864.1</td>
</tr>
<tr>
<td>Hs MoaD</td>
<td><em>Homo sapiens</em></td>
<td>NP_789776.1</td>
</tr>
<tr>
<td>Mm MoaD</td>
<td><em>Mus musculus</em></td>
<td>Q9Z224</td>
</tr>
<tr>
<td>An MoaD</td>
<td><em>Aspergillus nidulans</em></td>
<td>XP_682412.1</td>
</tr>
<tr>
<td>Tt MoaD</td>
<td><em>Thermus thermophilus</em></td>
<td>1V8C_D</td>
</tr>
<tr>
<td>Pf MoaD</td>
<td><em>Pyrococccus furiosus</em></td>
<td>AAL80667.1</td>
</tr>
<tr>
<td>Ec ThiS</td>
<td><em>Escherichia coli</em></td>
<td>O32583</td>
</tr>
<tr>
<td>Ii ThiS</td>
<td><em>Idiomarina lothiensis</em></td>
<td>YP_155158.1</td>
</tr>
<tr>
<td>Pf ThiS</td>
<td><em>Pyrococccus furiosus</em></td>
<td>NP_578790.2</td>
</tr>
<tr>
<td>Hs Urm1</td>
<td><em>Homo sapiens</em></td>
<td>AAH03581.1</td>
</tr>
<tr>
<td>Mm Urm1</td>
<td><em>Mus musculus</em></td>
<td>1XO3_A</td>
</tr>
<tr>
<td>Gg Urm1</td>
<td><em>Gallus gallus</em></td>
<td>CAG31999.1</td>
</tr>
<tr>
<td>Dm Urm1</td>
<td><em>Drosophila melanogaster</em></td>
<td>NP_996018.1</td>
</tr>
<tr>
<td>Dd Urm1</td>
<td><em>Dictyostelium discoideum</em></td>
<td>XP_638993.1</td>
</tr>
<tr>
<td>Ag Urm1</td>
<td><em>Anopheles gambiae</em></td>
<td>EAA11560.2</td>
</tr>
<tr>
<td>Sc Urm1</td>
<td><em>Saccharomyces cerevisiae</em></td>
<td>NP_012258.1</td>
</tr>
<tr>
<td>Tb Urm1</td>
<td><em>Trypanosoma brucei</em></td>
<td>AAX79740.1</td>
</tr>
<tr>
<td>Ce Smt3</td>
<td><em>Caenorhabditis elegans</em></td>
<td>AAK18969.1</td>
</tr>
<tr>
<td>At Smt3</td>
<td><em>Arabidopsis thaliana</em></td>
<td>NP_194414.1</td>
</tr>
<tr>
<td>Sc Smt3</td>
<td><em>Saccharomyces cerevisiae</em></td>
<td>AAS56500.1</td>
</tr>
<tr>
<td>Hs SUMO1</td>
<td><em>Homo sapiens</em></td>
<td>IA5R</td>
</tr>
<tr>
<td>Hs Ub</td>
<td><em>Homo sapiens</em></td>
<td>P62988</td>
</tr>
<tr>
<td>Sc Ub</td>
<td><em>Saccharomyces cerevisiae</em></td>
<td>P61864</td>
</tr>
</tbody>
</table>
Chapter 6

Concluding remarks and outlook

In this thesis a characterization of the PAN-proteasome complex and small archaeal modifier proteins (SAMPs) from *Methanosarcina acetivorans* was carried out. The obtained results show that *M. acetivorans* PAN and the proteasome can form a complex that degrades ssrA-tagged model substrates *in vitro*. Using a FRET based approach, substrate processing by the PAN-proteasome complex is followed in real-time. The data show that ssrA-tagged substrates are engaged from their C-termini, presumably because of the unstructured nature of the ssrA-tag. In addition this dissertation demonstrates the activation of recently identified small archaeal modifier proteins (SAMPs) by a novel protein, which we named SAMP activating enzyme (SAE). SAMPs share homology to ubiquitin and ubiquitin-like proteins (Ubls) and the presented data demonstrate that SAMPs are activated in a manner analogous to the first step of the ubiquitination pathway, where the C-terminal glycine of ubiquitin is adenylated. By NMR we could solve the solution structure of SAMP1. It is similar to the stably folded β-grasp proteins like ubiquitin and Ubls with respect to the overall fold.

The primary function of ubiquitin is to target substrates for degradation, whereas Ubls are involved in various cellular activities including, for example, sulfur transfer or RNA modification. It was observed recently that *Halofexx volcanii* SAMP1 and SAMP2 are covalently conjugated to proteins via an iso-peptide bond involving lysine residues of target proteins and itself *in vivo*. However, the precise role of the recently discovered SAMPs in the cell are yet to be investigated. To this end it is important to study the steps downstream of the activation of the SAMPs by SAE (Figure 21). It is suggested that it is possible to make poly-SAMP as archaeal SAMPs have lysines that can be linked to chains. It remains to be seen whether the surface exposed lysine in the extended loop of SAMP1 that we identified could be involved in poly-SAMPylation. It will also be important to identify binding partners of SAMPs to gain information about the pathways where SAMP might be involved in.

![Figure 21](image)

**Figure 21:** Schematic representation of the adenylation process of SAMPs by SAMP activating enzyme, SAE in presence of ATP. Upon interaction of SAMP and ATP with SAE, the C-terminal carboxylate of SAMP is activated by formation of the mixed with the AMP group while PPi is released. The adenylated intermediate, SAMP-AMP, presents an activated species that can be involved in protein modification at Lys residues or in pathways involving sulfur transfer.
Chapter 7

References


References


References


Curriculum vitae

Namit Ranjan

Date of birth: 29.04.1983  
Nationality: Indian  
Address: Institute of Molecular Biology and Biophysics,  
ETH Hoenggerberg,  
Schafmattstr. 20, CH-8093 Zurich

Education

2006 – 2010  
Doctoral studies at Institute of Molecular Biology and Biophysics, ETH Zürich.  
Thesis supervisor: Prof. Dr. Eilika Weber-Ban  
PhD committee: Prof. Dr. Rudi Glockshuber (ETH Zürich), Prof. Dr. Nenad Ban (ETH Zürich), Prof. Steve Burston (Univ. of Bristol)

2004 – 2006  
M.Sc. in Genomics from Center for Excellence in Genomic Sciences, School of Biological Sciences, Madurai Kamaraj University, India.  
M.Sc. thesis title: “The Journey about Alp$^+$ phenotype in E.coli: Further Molecular and Genetic studies”, under the supervision of Dr. M. Hussain Munavar, Department of Molecular Biology, Madurai Kamaraj University, India.

2000 – 2003  
Bachelor’s Degree in Biotechnology, Chemistry and Botany from T. John College, Bangalore University, India.

1998-2000  
“All India Senior Secondary School Examination” from Central Board of Secondary Education (CBSE), India.

1996-1998  
“All India Secondary School Examination” from Central Board of Secondary Education (CBSE), India.

Awards, Merits and Memberships

- Awarded “Certificate for Excellence” for securing highest marks in Bachelor studies by T. John College, Bangalore University, India.
Curriculum vitae

- Awarded the Second Prize in Paper Presentation competition conducted at Biotechcellence’06, organized by Centre for Biotechnology, Anna University, Chennai, India.
- Secured 1st rank in All-India level Entrance Examination for M.Sc. Genomics, conducted by Center for Excellence in Genomic Sciences, School of Biological Sciences, Madurai Kamaraj University, India.
- Member of Zurich Molecular Life Science PhD program of Zurich graduate school from 2006.

Presentations

- Selected for an oral presentation at Biotechcellence’06 between February 24th and 26th, 2006, organized by Centre for Biotechnology, Anna University, Chennai, India.
- Poster presentation in yearly PhD retreat of Zurich Molecular Life Science PhD program.

Teaching Experience

- Supervision of undergraduate student course during 4-week block practicum for ‘Protein folding, assembly and degradation’ course of Department of Biology, ETH Zürich.
- Supervision of undergraduate students in the 2nd semester Biology Practical Course.

Research Experiences

- One year Master’s thesis on “The Journey about Alp+ phenotype in E.coli: Further Molecular and Genetic studies”, under the supervision of Dr. M. Hussain Munavar, Department of Molecular Biology, Madurai Kamaraj University, India.

Publications

Acknowledgments

I am thankful to my supervisor Prof. Eilika Weber-Ban for her advice, support and encouragement throughout my PhD studies. Her great and innovative scientific idea kept my motivation high and has influenced my way of scientific thinking. Thanks to her for the freedom that she gave to develop and pursue my own ideas.

I am grateful to my PhD thesis committee members Prof. Rudi Glockshuber, Prof. Nenad Ban and Prof. Steven Burston for their critical analysis of my project progress and scientific guidance. Their ideas from different scientific angles really helped me to make successful progress of my PhD project.

I would like to thank Fred Damberger and Markus Sutter for the successful collaboration on the SAMP project. Thanks to Markus for his help in SAMP project from bioinformatics side, his experience has really helped a lot to think about SAMP from a different angle. Thanks to Fred for his wonderful and successful collaboration for NMR studies. I am particularly thankful to him for his incredible patience with me.

Thanks to Kristina Kolygo for the collaboration on PAN-proteasome substrate processing using FRET studies. This collaboration has boosted my scientific motivation with my first scientific article as co-author.

Thanks to Wolfgang Kress, Frank Striebel, Frank Imkamp and Željka Maglica for taking time out whenever needed to discuss the scientific things and for their constant motivation throughout my PhD study.

Thanks to Dennis Özcelik, Ethan Guth and Jannis Warweg for the helpful and fun atmosphere in the lab.

I am extremely grateful to my family especially my parents and my brother Satish, for believing in me and for always being with me with their blessings. I therefore, dedicate this thesis to them.

I would like to thank Amit Sikdar, Neel, Surendra, Swati, Tariq, Aditi, Shishir, and all friends for their support and encouragement.

Last but not least, thank to God for always being with me.