Technische Universität München

Fakultät für Chemie

Lehrstuhl für Biochemie

X-ray crystallographic analysis of the ‘ancestral β-subunit protein’ (Anbu) from *Hyphomicrobium sp.* and the hetero-oligomeric protease ClpP1/2 from *Listeria monocytogenes*

Marie-Theres Vielberg

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigten Dissertation.

Vorsitzender: Prof. Dr. Aymelt Itzen

Prüfende der Dissertation:

1. Prof. Dr. Michael Groll
2. Prof. Dr. Ville Kaila

Die Dissertation wurde am 07.11.2017 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 01.12.2017 angenommen.

Für Mischa und meine Familie

Table of Contents

1 Summary

Proteolysis is not only essential for energy production and recycling of amino acids to synthesize new proteins but also has a crucial regulatory function in degradation of catalytically impaired or no longer required proteins. In bacteria, where no specialized organelles for protein degradation are found, a huge variety of proteases developed to undertake these tasks. Using X-ray crystallographic analysis, two representative candidates were characterised in detail: Anbu from *Hyphomicrobium sp.* strain MC1 and the ClpP1/2 heterocomplex from *Listeria monocytogenes*.

The 'ancestral β-subunit protein' (Anbu) was discovered in a recent study searching for bacterial progenitors of the 20S proteasome, a multi-subunit protease that performs the majority of non-lysosomal proteolysis in eukaryotes. It is also ubiquitously present in archaea, while its occurrence in bacteria is rare. Many of these organisms possess the homologous protease HslV instead. However, there is still a huge number of species, including all cyanobacteria, which encode neither one nor the other, questioning the evolutionary origin of this family of N-terminal threonine proteases. Studying the Anbu protein from *Hyphomicrobium sp.* strain MC1 resulted in growth of well-diffracting crystals, allowing us to solve its structure via SAD-phasing. Although it seems to assemble in defined particles in solution, Anbu formed helical superstructures in the crystal, contrary to the closed, barrel-shaped structures known from the 20S proteasome or HslV. Nevertheless, the fold of an individual subunit clearly identifies Anbu as an Ntn-hydrolase and the inter-subunit contacts within the structures of all three protein complexes are similar. Conservation of the Thr1O$^{\gamma}$ activating triad suggested that Anbu might also function as an endoprotease. However, using standard proteasomal substrates no such activity was detected so far. Potential reasons might be the missing activation of the Thr1-NH_2 terminus, the lack of a suitable substrate or interaction partner or the existence of only a structural, but no functional relationship between Anbu, the 20S proteasome and HslV.

The caseinolytic protease P (ClpP) was shown to be essential for the virulence of several facultative and obligate pathogens. Since there is a rapid increase in bacteria resistant to several commonly used antibiotics and the search for new targets has proven to be difficult, ClpP is listed as a potential candidate for drug development. In contrast to most bacteria, *Listeria monocytogenes* encodes two isoforms of ClpP, designated as *Lm*ClpP1 and

*Lm*ClpP2. Although the sequence identity of the two proteins amounts to less than 45 %, they can form a heterocomplex together. To understand how it differs from homomeric ClpPs, we crystallised and solved the structure of this protease via Patterson search methods. The data reveal that the overall architecture and stabilization of homo- and heteromeric ClpP particles adopt the same topology. However, there exist differences in the axial pore formation, active site composition and shape of the S1 pocket between *Lm*ClpP1 and *Lm*ClpP2. These results suggest a potential task sharing in the heterocomplex and provide valuable information for the design of isoform-specific inhibitors to characterise the function of the *Lm*ClpP1/2 heterocomplex in more detail.

2 Zusammenfassung

Die Bedeutung des Abbaus von Proteinen liegt nicht nur in der Gewinnung von Energie oder dem Recycling von Aminosäuren zum Aufbau neuer Proteine, sondern hat auch eine entscheidende regulatorische Komponente bei der Beschränkung der Lebensdauer von funktionell geschädigten oder redundanten Proteinen. In Bakterien, die keine auf die Proteolyse spezialisierten Organellen besitzen, übernimmt eine Vielzahl an Proteasen diese Aufgabe. Gegenstand dieser Arbeit ist die Röntgenstrukturanalyse zwei repräsentativer Vertreter dieser Enzyme: Anbu aus *Hyphomicrobium sp.* Stamm MC1 und der ClpP1/2 Heterokomplex aus *Listeria monocytogenes*.

Anbu, das „β-Untereinheiten-Vorläufer-Protein“, wurde in einer aktuellen Studie beschrieben, in der nach bakteriellen Vorfahren des 20S Proteasoms, der wichtigsten Protease im Cytosol von Eukaryoten, gesucht wurde. Auch in allen Archaeen findet man das 20S Proteasom, während es den meisten Bakterien fehlt. Stattdessen greifen viele von ihnen auf die homologe Protease HslV zurück. Es gibt jedoch immer noch eine Vielzahl an Organismen, wie zum Beispiel alle Cyanobakterien, in denen weder das eine noch das andere Protein vertreten ist, wodurch sich die Frage nach dem Ursprung dieser Protease-Familie stellte. Die Untersuchung des Anbu Proteins aus dem Organismus *Hyphomicrobium sp.* Stamm MC1 gipfelte im Wachstum gut-beugender Kristalle. Die Struktur des Proteins konnte mithilfe der SAD-Phasierungsmethode gelöst werden. Im Widerspruch sowohl zu den in Lösung gemessenen Daten als auch zu den zylindrischen Partikeln, die das 20S Proteasom oder HslV bilden, ordneten sich die Anbu Untereinheiten im Kristall als spiralförmige Superstrukturen an. Die Faltung der einzelnen Untereinheit hingegen zeigt deutlich, dass es sich bei Anbu um eine Ntn-Hydrolase handelt. Auch die Art der Kontakte zwischen verschiedenen Anbu-Molekülen lässt die Verwandtschaft mit dem 20S Proteasom und HslV erkennen. Die Entdeckung, dass die Aminosäuren, die die Nukleophil-Bildung am Threonin-Rest im aktiven Zentrum der Protease bewirken, in Anbu wiederzufinden sind, ließ vermuten, dass es sich bei diesem Protein auch um eine Endoprotease handelt. Bis jetzt konnte jedoch keine hydrolytische Aktivität gegenüber Standardsubstraten des Proteasoms oder HslV gemessen werden. Wie in der Kristallstruktur von Anbu zu sehen, kommt es zu keiner Aktivierung des N-Terminus von Thr1, was die Katalyse der Reaktion verhindern könnte. Weitere mögliche Erklärungen wären, dass der Ausgang der Versuche durch das Fehlen des richtigen Substrats oder eines

passenden Bindungspartners für Anbu negativ beeinflusst wurde. Zum jetzigen Stand der Untersuchungen lässt sich auch nicht ausschließen, dass es sich nur um eine strukturelle Verwandtschaft zwischen Anbu, dem 20S Proteasom und HslV handelt, die die proteolytische Funktion nicht miteinschließt.

Die caseinolytische Protease P (ClpP) ist verantwortlich für die Virulenz vieler Bakterien, sowohl fakultativ als auch obligat pathogener Spezies. Aufgrund der steigenden Zahl an Arten, die gegen eine Vielzahl an allgemein eingesetzten Antibiotika resistent sind, und da sich die Suche nach neuen Angriffspunkten als herausfordernd gestaltet, gilt ClpP als ein vielversprechender Kandidat für die Entwicklung neuer Wirkstoffe. Im Gegensatz zu den meisten Bakterien findet man in *Listeria monocytogenes* zwei verschiedene Isoformen von ClpP, bezeichnet als *Lm*ClpP1 und *Lm*ClpP2. Trotz einer Sequenzidentität der beiden Proteine von weniger als 45 % sind sie in der Lage, gemeinsam eine aktive Protease aufzubauen. Um zu verstehen, wie sich diese von ClpP Proteinen unterscheidet, die nur aus einer einzigen Untereinheit aufgebaut sind, wurde der Heterokomplex kristallisiert und seine Struktur durch molekularen Ersatz gelöst. Die Daten zeigen, dass die allgemeine Struktur des Partikels und die Stabilisierung des Komplexes die gleiche Topologie aufweisen wie in homomeren Clp Proteasen. *Lm*ClpP1 und *Lm*ClpP2 unterschieden sich jedoch sowohl in der Bildung der axialen Pore des jeweiligen Rings als auch in der Zusammensetzung des aktiven Zentrums und der Ausgestaltung der S1-Spezifitätstasche. Die Ergebnisse lassen auf eine mögliche Aufgabenteilung innerhalb des Heterokomplexes schließen und liefern wertvolle Hinweise für die Synthese von Inhibitoren, die sich ausschließlich gegen die eine oder die andere Untereinheit richten.

3 Introduction

The general function of proteins is catalysis of a huge variety of diverse chemical reactions to produce the vast majority of cellular compounds, however, they are also responsible for the stability and shape of a cell as well as regulating its motility and communication with the environment. To maintain the cellular integrity, only structurally and functionally impeccable proteins are of good use [1, 2]. So, misfolded or damaged proteins must be removed effectively. In the process of proteolysis, a protein is broken down into smaller peptides before it gets disassembled completely into the individual amino acids. In principal, the cleavage of a peptide bond is energetically favoured, but the time frame for its stochastic break down exceeds the half-life of a cell many times over [3]. Therefore, protein degradation needs to be catalysed by a specialized group of enzymes, called proteases [4].

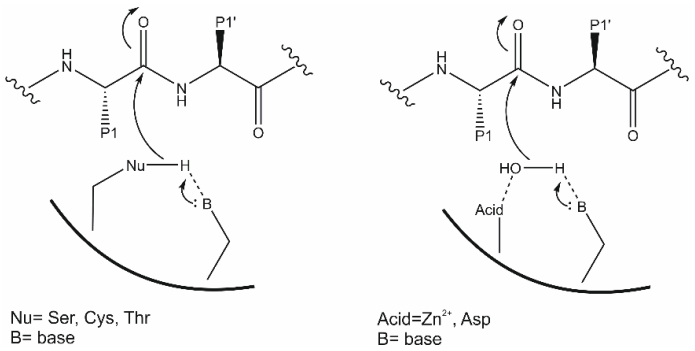


Figure 1: Mechanistic distinction of serine, cysteine and threonine proteases from aspartate and metalloproteases.

In the first group of proteases it is the active site residue (Ser/Cys/Thr) that attacks the carbonyl carbon atom of the scissile peptide bond forming a covalent link between the protease and its substrate, whilst in the second group this task is undertaken by a water molecule polarized by hydrogen bonding or coordination to the catalytic centre of the protease.

According to the composition of their active site, proteases can be divided into seven different classes: cysteine, serine, threonine, aspartic, glutamic and metalloproteases as well as asparagine peptide lyases (MEROPS database, [5]). With exception of the last, they can be further split up into two distinct subgroups based on differences in their catalytic mechanism (Figure 1) [6, 7]. In the case of cysteine, serine and threonine proteases, the active site residue – eponymous for each class of proteases – serves as the nucleophile that attacks the carbonyl carbon atom of the peptide bond [8]. Thereby, a covalent acyl-enzyme-complex is formed and the part of the peptide chain located downstream of the scissile bond is released (Figure 1, Figure 2). The shape of the active site of the protease promotes and stabilizes the intermediate state, in which the substrate is bound to the enzyme. The attack of a water molecule leads to its dissolution and release of the N-terminal part of the former

peptide chain as well as restoration of the catalytic centre. Generation of the active-site nucleophile is ensured by embedding it into an environment that allows efficient transfer of protons between different residues [8]. In contrast, aspartic, glutamic and metallo-proteases lack a covalent bond between the enzyme and its substrate during the catalytic cycle. The hydrolysis of the amide bond is catalysed by the nucleophilic attack of a water molecule polarized via hydrogen bonding or coordination to the reactive centre [6]. Independent of a protease's mode of action, formation of the two cleavage products requires consumption of a water molecule.

Another way to classify proteases is to distinguish between exo- and endo-proteases. The former remove either the N- or the C-terminal residue of a peptide chain, depending on their type, whereas the latter cleave a bond within. Using small peptide substrates, it could be shown that proteases of the second group favour specific residues to cut after *e.g.* basic or aromatic ones. Such preferences were found to have a structural explanation. Responsible are the so-called specificity pockets of a protease, in which the side chains of the amino acids located before and after the scissile peptide bond bind. To describe this phenomenon better, a special nomenclature was introduced by Schechter and Berger (Figure 2) [9]. The amino acid next to the cleavage site is designated as P1 (non-primed) residue, if it is located upstream and, thus, part of the N-terminal cleavage product, or P1' (primed) residue, if it is located downstream and part of the C-terminal one. The subsequent amino acids are then described as P2 or P2', P3 or P3' residues and so forth, respectively. The corresponding specificity pockets in the protease are called S1, S2, … or S1', S2', … and their shape primarily determines the cleavage preferences of a protease.

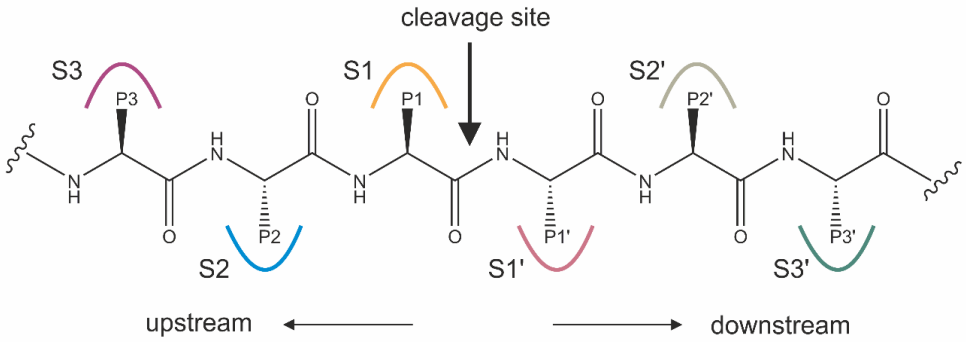


Figure 2: Classification of the specificity pockets of a protease after Schechter and Berger [9].

Once a peptide binds to the substrate channel of a protease, the side chains of the amino acids next to the scissile peptide bond are accommodated in the so-called specificity pockets S (depicted schematically in various colours), whereas the corresponding residues are labelled with P. Their precise naming depends on their location up- or downstream of the cleavage site and how many amino acids they are apart.

To ensure that no still required and functionally intact protein is prematurely degraded, proteolysis needs to be strictly regulated. For that, proteases have developed various strategies, *e.g.* pH-dependent or allosteric (in-)activation, production as zymogens, spatial separation and/or limitation to a specific substrate spectrum [4]. One widely used modus operandi is to sequester the catalytic centres of a protease inside a hollow, cylindrical structure. Multiple copies of one or more subunits assemble into different rings, which are then stacked one above the other [10]. Thereby, a proteolytic particle is formed with a catalytic chamber in the centre and axial pores on its top and bottom, which are normally sealed by a narrow gate restricting access to the active sites. To gain its full catalytic activity, the protease relies on the interaction with other, mostly regulatory proteins. Usually, these are members of the AAA^+-ATPase family [11]. Their task is not only to recognize and unfold the substrates of the protease but also to open the gate of its axial pores by binding to them. Then, under the consumption of ATP, the regulator threads the unfolded peptide chain into the hydrolytic chamber for processing. Such combinations of a protease and an ATPase are referred to as bipartite degradation machineries, emphasising the cooperative interaction of both proteins [12].

4 Objective

The 20S proteasome is the major non-lysosomal protease in eukaryotic and archaeal cells [13]. The activity of this barrel-shaped threonine protease is controlled and modulated by a wide range of ATPases and other regulators. Its bacterial homologue HslV must bind to the ATPase HslU to become fully functional. Recently, another potential member of this protease family was described. Biocomputational studies led to the discovery of the ancestral β-subunit protein (Anbu) [14]. Due to its occurrence in cyanobacteria, it was speculated whether Anbu might be the progenitor of these threonine proteases, though no ATPase suitable to regulate its activity was identified so far. Thus, we aimed to characterise Anbu from *Hyphomicrobium sp.* strain MC1 (*Hy*Anbu) at molecular resolution. Using biochemical analyses and X-ray crystallography, our goal was to elucidate both its structure and presumed function as well as the relationship to its proteasomal homologues as described in the first part of this thesis (see chapter 6). In parallel, two papers summarizing parts of our research on Anbu were submitted for publication, one in cooperation with the group of Prof. Dr. Matthias Bochtler from the International Institute of Molecular and Cell Biology in Warsaw (see chapter 10).

Particularly in prokaryotes, where no cellular compartmentalization exists, the formation of barrel-shaped proteases seems to be a preferred strategy for efficient regulation of intracellular proteolysis. A well characterised example is the caseinolytic protease ClpP found in a variety of bacterial species – many of them are human pathogens –, where it is often involved in virulence regulation. For protein degradation, it selectively interacts with one of the different Clp-ATPases.

In contrast to most bacteria, *Listeria monocytogenes* produces two isoforms of ClpP that can interact with each other to form the functional protease. Besides, one of the two subunits is also able to assemble into the proteolytic particle on its own. Studies of our cooperation partners from the group of Prof. Dr. Stephan A. Sieber at the Technical University Munich revealed that the proteolytic activity of the *Lm*ClpP1/2 heterocomplex was stimulated considerably better by interaction with the ATPase ClpX than that of the homomeric ClpP2 particle. In search of an explanation, we elucidated the structure of the *Lm*ClpP1/2-heterocomplex via X-ray crystallographic analysis. The results are presented in the second part of this thesis (see chapter 7) and were published in ‘Structure and

mechanism of the caseinolytic protease ClpP1/2 heterocomplex from *Listeria monocytogenes'* in *Angew. Chem. Int. Ed.* 2015, Vol. 54, p. 3598-3602 ([15], see also chapter 10)[1].

[1] The cooperation with the group of Prof. Dr. S. A. Sieber resulted in another publication on Clp proteases, where a non-covalent inhibitor for ClpP from *Staphylococcus aureus* was described and its binding mode was analysed via X-ray crystallography. These results, however, are not part of this thesis. They have been published in *Angew. Chem. Int. Ed.*, 54, 15892-6 ([16] and chapter 10).

5 Materials and Methods

This chapter describes the experimental background of the cloning, purification, biochemical and crystallographic analysis of *Hy*Anbu as well as the crystallization and structure elucidation of the *Lm*ClpP1/2 heterocomplex. Cross-references in this chapter are shown as paragraph numbers in parentheses.

5.1 Materials

5.1.1 Chemicals

Unless stated otherwise, all chemicals were purchased from one of the following companies: AppliChem (Darmstadt, DE), Biomol (Hamburg, DE), Merck (Darmstadt, DE), Sigma-Aldrich (St. Louis, US), Roth (Karlsruhe, DE) and VWR (Darmstadt, DE).

5.1.2 Antibiotics

To select bacterial clones carrying a plasmid with a specific resistance, LB-medium or LB-agar plates containing the corresponding antibiotic(s) were used.

Table 1: Antibiotics for positive selection of bacterial clones.

Antibiotic	Liquid culture	Agar plates
Ampicillin	50 µg/mL	180 µg/Ml
Kanamycin	50 µg/mL	100 µg/Ml

5.1.3 Bacterial strains

For cloning and expression of the genes of *Hy*Anbu and its presumed interaction partners, different strains of *Escherichia coli* were used. To guarantee consistent quality, the production of electrocompetent cells was in the responsibility of our technician Kathrin Gärtner.

Table 2: Genotype and origin of *E. coli* strains.

Strain	Genotype	Company
E. coli XL1 blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F´ *proAB lacI*qZΔ*M15* Tn*10* (Tetr)]	Stratagene (La Jolla, US)
E. coli XL-10 gold	*endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac* Hte Δ(*mcrA*)183 Δ(*mcrCB-hsdSMR-mrr*)173 tet^R F'[*proAB lacI*^{q}ZΔ*M15* Tn*10* (Tet^R Amy Cm^R)]	Agilent Technologies (St. Clara, US)

Table 2 (continued): Genotype and origin of *E. coli* strains.

Strain	Genotype	Company
E. coli BL21 (DE3)	*E. coli* str. B F⁻ *ompT gal dcm lon* $hsdS_B(r_B^-m_B^-)$ λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) $[malB^+]_{K\text{-}12}(\lambda^S)$	Novagen/Merck (Darmstadt, DE)
E. coli BL21-Gold (DE3)	*E. coli* str. B F⁻ *ompT gal dcm lon* $hsdS_B(r_B^-m_B^-)$ λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) $[malB^+]_{K\text{-}12}(\lambda^S)$ *endA1* The	Agilent Technologies (St. Clara, US)
E. coli SolubBL21	*E. coli* str. B F⁻ *ompT* $hsdS_B(r_B^-m_B^-)$ *gal dcm lon* λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*])*	Genlantis (San Diego, US)

*According to company information, there exist some uncharacterised mutations in this *E. coli* strain that are essential for its specific properties.

5.1.4 Buffer and Media

Commonly used buffer and media are listed below. For all solutions that were made in the laboratory double-distilled H_2O (ddH_2O) was used.

DNA loading buffer (10x)	10 mM	Tris/HCl, pH 8.2
	1 mM	EDTA
	50 % (v/v)	Glycerol
	0.25 % (w/v)	Bromophenol blue
	0.25 % (w/v)	Xylene cyanol
TAE buffer (50x)	2 M	Tris/Ac, pH 8.2
	100 mM	EDTA
Ethidium bromide solution	1 mg/L	Ethidium bromide
SDS sample buffer (5x)	60 mM	Tris/HCl, pH 6.8
	30 % (v/v)	Glycerol
	10 % (w/v)	Sucrose
	5 % (w/v)	SDS
	3 % (v/v)	β-Mercaptoethanol
	0.02 % (w/v)	Bromophenol blue
Separation gel buffer	1.5 M	Tris/HCl, pH 8.8
	0.4 % (w/v)	SDS

Stacking gel buffer	0.5 M	Tris/HCl, pH 6.8
	0.4 % (w/v)	SDS
APS (10 %)	10 % (w/v)	Ammonium persulfate
Electrophoresis buffer (1x)	25 mM	Tris
	192 mM	Glycine
	0.1 % (w/v)	SDS
Staining solution	0.05 % (w/v)	Coomassie Brilliant Blue
	25 % (v/v)	Isopropanol
	10 % (v/v)	Acetic acid
Destaining solution	10 % (v/v)	Acetic acid
LB_0 medium	1 % (w/v)	Peptone
	0.5 % (w/v)	Yeast Extract
	0.5 % (w/v)	NaCl
	(2 % (w/v)	Agar)
SOC medium	2 % (w/v)	Peptone
	0.5 % (w/v)	Yeast Extract
	20 mM	Glucose
	10 mM	$MgSO_4$
	10 mM	$MgCl_2$
	10 mM	NaCl
	2.5 mM	KCl
M9 mineral medium	33.7 mM	Na_2HPO_4*
	22.0 mM	KH_2PO_4*
	8.55 mM	NaCl*
	9.35 mM	NH_4Cl*
	0.4 % (w/v)	Glucose
	1 mM	$MgSO_4$
	0.3 mM	$CaCl_2$
	1 µg/L	Biotin

M9 mineral medium (continued)	1 µg/L	Thiamin
	1x	Trace elements solution

* Salts were set up as 10x stock solution with a pH of 7.2 adjusted with NaOH.

Trace elements solution (100x)	13.4 mM	EDTA
	3.1 mM	$FeCl_3*6H_2O$
	0.62 mM	$ZnCl_2$
	76 µM	$CuCl_2*2H_2O$
	42 µM	$CoCl_2*2H_2O$
	162 µM	H_3BO_3
	8.1 µM	$MnCl_2*4H_2O$

5.1.5 DNA and protein standards

peqGOLD DNA ladder mix (0.1-10 kb)	peqlab (Erlangen, DE)
Roti®-Mark STANDARD (14-200 kDa)	Roth (Karlsruhe, DE)

5.1.6 Enzymes

*Bam*HI-HF (20 U/µL)	New England Biolabs (Ipswich, US)
DNase I	Sigma-Aldrich (St. Louis, US)
*Dpn*I (20 U/µL)	New England Biolabs (Ipswich, US)
*Nco*I-HF (20 U/µL)	New England Biolabs (Ipswich, US)
PfuTurbo DNA Polymerase (2.5 U/µL)	Agilent (St. Clara, US)
Phusion HF DNA Polymerase (2 U/µL)	Finnzymes (Vantaa, FI)
*Pst*I-HF (20 U/µL)	New England Biolabs (Ipswich, US)
T4-DNA-Ligase (1 U/µL)	Invitrogen (Carlsbad, US)

5.1.7 Kits and pre-made solutions

Cloned *Pfu* reaction Buffer (10x)	Agilent (St. Clara, US)
CutSmart™ Buffer (10x)	New England Biolabs (Ipswich, US)
dNTP Mix (10 mM)	Bioline (Luckenwalde, DE)
peqGOLD Cycle-Pure Kit	peqlab (Erlangen, DE)
peqGOLD Plasmid Miniprep Kit I & II	peqlab (Erlangen, DE)
Pierce™ Fluoresc. Protease Assay Kit	Thermo Fisher Scientific (Waltham, US)
Phusion HF Buffer (10x)	Finnzymes (Vantaa, FI)

T4 DNA Ligase Buffer (5x)	Invitrogen (Carlsbad, US)
Z-GGL-AMC*	Bachem (Bubendorf, CH)
Suc-LLVY-AMC*	Bachem (Bubendorf, CH)
Boc-LLR-AMC*	Bachem (Bubendorf, CH)
Z-LLE-AMC*	Bachem (Bubendorf, CH)

*Fluorogenic substrates were dissolved in DMSO and stored at -20 °C.

5.1.8 Plasmids

The pETDuet™-1 plasmid was available in the laboratory and used without further modifications. It was originally purchased from Novagen (Darmstadt, DE). The plasmid contains two multiple cloning sites (MCS) enabling the simultaneous expression of two target genes. Each gene is under the control of its own T7 promoter, followed by a *lac* operator and the ribosome binding site. The plasmid also encodes the *lacI* repressor and the ampicillin resistance gene.

Table 3 lists all gene expressing plasmids cloned for this work. Details regarding the DNA sequence of the target gene or primers for amplification and insertion of appropriate restriction sites, are listed in the respective sections.

Table 3: Overview of all plasmids relevant for this work.

Name	Original plasmid	Target Gene	Restriction sites
pET_Anbu^{mut}(*hysp*)	pETDuet™-1	COG3848	*Nco*I, *Bam*HI
pET_Anbu^{wt}(*hysp*)*	pET_Anbu^{mut}(*hysp*)	-	-

*The plasmid pET_Anbu^{wt}(*hysp*) was derived via site directed mutagenesis from the plasmid pET_Anbu^{mut}(*hysp*), which was cloned by my Master's student Verena Bauer.

5.1.9 Primers

All primers were purchased from Eurofins Genomics (Ebersberg, DE). They were ordered as unmodified olignonucleotides, purified by either HPSF or HPLC, and delivered in lyophilized form. Primers were dissolved in ddH_2O to reach a final concentration of 100 pmol/µL and stored at -20 °C until further use.

Table 4: Primers for site directed mutagenesis or amplification of a target gene via PCR.

Name	Sequence (5'→ 3')*
anbu_hysp_mt_for	GAA GGA GAT ATA CC**A TGA CG**T ATG CGG TTG CCT T
anbu_hysp_mt_rev	AAG GCA ACC GCA TA**C GTC AT**G GTA TAT CTC CTT C

* Sites of mutagenesis or restriction sites are shown in bold.

5.1.10 Synthetic gene sequences

To facilitate recombinant protein production, the DNA sequence of *anbu* from *Hyphomicrobium sp.* strain MC1 (UniProt ID: F8JB59) was adapted to codon-usage of *E. coli* (software: GENEius, Eurofins Genomics). The customized sequence (see below) was then ordered for synthesis by Eurofins Genomics (Ebersberg, DE). Codon-optimization did not alter the primary sequences of the encoded proteins.

Codon-optimized gene sequence for *anbu* from *Hyphomicrobium sp.*:

ATGACGTATGCGGTTGCCTTTCGCTTAGAACGCGGCTTAGTATTCGCAGCTGATACTCGTACGAATGCGGGAGTCGACA
ACATCGCGCAGTACAAGAAACTCCAGCTCTGGCGTCAACCAGGTGAACGCGTCTGTGCTGCTTTCGGCGGGTAATTTAG
CAGCGACCCAAGCAGTTGTCAGCCTCATTAACGAACATCTGTCTCAGGAAACTGACGACGAGGTAACCACGCTGTTTAC
GGCGCCGAACATGTATCGCGCAGCTCGTGTTGTGGGTGATGCCGTTCGCGAAGCACGCTCCATTGATGGTGCGGCTCTG
GAAGCCTCCAAATTGGGGTTTAACACCAACTTCATCTTTGGCGGCCAGATCAAAGGCGAGCGTCCTCGCTTGTTTCAGA
TCTATCCGGAAGGAAACTTCATTGAAGCGACAGACGATACCCCGTTCTTTCAGATTGGCGAACACAAGTATGGCAAACC
GATTCTGGATCGCGTAGCACGAGCGATATGCGTCTTGGTGAAGCCGCCAAACTGATGCTGCTGTCATTCGACTCAACCC
TGCGTAGCAATCTGTCTGTGGGGATGCCCATTGATCTGGTGATCTACGAACGCGATACCTTTGACGTGACACGTGAGAA
ACGCATAGTGCCGATGACGAATACTTCCGGAATCTGAGCAATGCCTGGAGTGATGCTTTGCGTCAAGCGTTCTCGAAAA
TTGAGGAGTTTGATGTGTAA

5.1.11 Laboratory equipment

Balances

Analytical Balance TE124S	Sartorius (Göttingen, DE)
Precision Balance BP3100 P	Sartorius (Göttingen, DE)

Centrifuges

SIGMA 1-14	SIGMA Laborzentrifugen (Osterode am Harz, DE)
SIGMA 4K15 (rotor 11150/13350)	SIGMA Laborzentrifugen (Osterode am Harz, DE)
SIGMA 6-16K (rotor 12500)	SIGMA Laborzentrifugen (Osterode am Harz, DE)
SIGMA 8K (rotor 11805)	SIGMA Laborzentrifugen (Osterode am Harz, DE)

Chromatography systems and accessories

ÄKTAprime™ plus	GE Healthcare (Chalfont, St. Giles, GB)
ÄKTApurifier™	GE Healthcare (Chalfont, St. Giles, GB)
ÄKTApure™	GE Healthcare (Chalfont, St. Giles, GB)
HiTrap DEAE FF (5 mL)	GE Healthcare (Chalfont, St. Giles, GB)
HiTrap Q FF (5 mL)	GE Healthcare (Chalfont, St. Giles, GB)
MonoQ 5/50 GL (1 mL)	GE Healthcare (Chalfont, St. Giles, GB)

Superdex 200 10/300 GL	GE Healthcare (Chalfont, St. Giles, GB)
Q Sepharose Fast Flow resin*	GE Healthcare (Chalfont, St. Giles, GB)

* The column material was used to pack an anion exchange column with a 40-mL bed volume.

Crystallography

Crystal Phoenix Robot	Art Robbins Instruments (Sunnyvale, US)
MICROLAB® STARlet	Hamilton (Reno, US)
Oryx4 Robot	Douglas Instruments (Hungerford, GB)
Intelli-Plate® (96 wells)	Art Robbins Instruments (Sunnyvale, US)
RUMED Cooled Incubator 3000	Rubarth Apparate (Laatzen, DE)
Cooling Incubator IPP30Plus	Memmert GmbH + Co. KG (Schwabach, DE)
Quick Combi Sealer Plus	HJ-Bioanalytik (Mönchengladbach, DE)
Zoom Stereo Microscope SZX10	Olympus (Tokio, JP)
Mounted CryoLoop™	Hampton Research (Aliso Viejo, US)
CrystalCap HT™	Hampton Research (Aliso Viejo, US)
CrystalCap HT™ Vial	Hampton Research (Aliso Viejo, US)
CrystalWand Magnetic™	Hampton Research (Aliso Viejo, US)
Index HT/Additive Screen	Hampton Research (Aliso Viejo, US)
NeXtal Screens (1-5,8-9,11-12,15)	QIAGEN (Hilden, DE)
SuperClear Pregreased 24 well plate	Crystalgen (New York, US)
Siliconized Glass Cover Slides	Hampton Research (Aliso Viejo, US)
Vial Tongs	Molecular Dimensions (Newmarket, GB)
Box of Micro Tools	Molecular Dimensions (Newmarket, GB)
Foam Dewars	Spearlab (San Francisco, US)
Taylor Wharton CX100 ‘dry shipper’	tec-lab (Idstein, DE)
Taylor Wharton HC20 cryo tank	tec-lab (Idstein, DE)
UVEX microscope	JAN Scientific (Seattle, US)

Electrophoresis

Chamber and tray	Appligene (Watford,UK)
Electrophoresis Power Supply 6000	Pharmacia Biotech (Uppsala, SE)
Mini PROTEAN® Tetra Cell	Bio-Rad (Hercules, US)
PowerPac Basic Power Supply	Bio-Rad (Hercules, US)
Gel Documentation System G:BOX	Syngene (Cambridge, GB)
Digital Graphic Printer UP-D897	Sony (Minato, JP)

Additional equipment and material

NanoPhotometer™ Pearl	IMPLEN (München, DE)
NanoDrop 2000c	Thermo Fisher Scientific (Waltham, US)
Ultrospec 10 Cell Density Meter	Amersham Biosciences (Uppsala, SE)
Cary Eclipse Fluoresc. Spectrometer	Varian (Darmstadt, DE)
Spark™ 10M micropl. reader	Tecan (Männedorf, CH)
Nunc™ F96 MicroWell™ pl. (wht)	Thermo Fisher Scientific (Waltham, US)
inoLab® pH 720 pH-meter	WTW (Weilheim, DE)
Gene Pulser and Pulse Controller	Bio-Rad (Hercules, US)
Electroporation cuvette, 2 mm	peqlab (Erlangen, DE)
MyCycler™ Thermal Cycler	Bio-Rad (Hercules, US)
Thermomixer comfort	Eppendorf (Hamburg, DE)
Techne Dri-Block DB 2A	Bibby Scientific (Stone, UK)
Branson Digital Sonifier® 250	Branson Ultrasonic (Danbury, US)
Ultrasonic vessel	G. Heinemann (Schwäbisch Gmünd, DE)
MR Hei-Standard Magnetic stirrer	Heidolph (Schwabach, DE)
Vortex Genie 2	Scientific Industries (New York, US)
Incubator	Binder (Tuttlingen, DE)
Infors HT Multitron 2 Cell Shaker	INFORS HT (Bottmingen, CH)
Low Temperature Freezer Vials	VWR (Darmstadt, DE)
Locator Thermolyne Plus	Thermo Fisher Scientific (Waltham, US)
Dialysis Tubing	Roth (Karlsruhe, DE)
Varioklav®	HP Medizintechnik GmbH (München, DE)
Vivaspin® 6/20	Sartorius (Göttingen, DE)

5.1.12 Bioinformatic tools, software and databases

ApE- a plasmid editor v. 2.0.51	M. Wayne Davis (Salt Lake City, US)
CCP4 Software Suite v. 7.0	http://www.ccp4.ac.uk [17]
Citavi v. 5.4	Swiss Academic Software (Wädenswil, CH)
Coot v. 0.8.6	P. Emsley (Oxford, UK) [18]
CorelDRAW X5	Corel (Ottawa, CA)
DaliLite v. 3	Holm group (Helsinki, FI) [19]
GraphPad Prism v. 5	GraphPad Software Inc. (La Jolla, US)
JalView	http://www.jalview.org/ [20]

JPred server	http://www.compbio.dundee.ac.uk/jpred/index.html [21]
Microsoft Office 2016	Microsoft (Redmond, US)
MolProbity v. 4.3	Richardson Lab, Duke Univ. (Durham, US) [22]
MOLSCRIPT	P. J. Kraulis [23]
NEB Double Digest Finder	New England Biolabs (Ipswich, US)
NEB Tm Calculator	New England Biolabs (Ipswich, US)
PDBePISA	European Bioinformatics Inst. (Hinxton, UK) [24]
ProtParam	http://web.expasy.org/protparam/ [25]
PyMOL Molecular Graphic Systems	Schrödinger, LLC [26]
QuickChange® Primer Design	Agilent (St. Clara, US)
RCSB Protein Data Bank (PDB)	http://www.rcsb.org/pdb/home/home.do
SIAS server	http://imed.med.ucm.es/Tools/sias.html (Madrid, ES)
SMART server	http://smart.embl-heidelberg.de/ [27]
T-COFFEE Sequence Align. Server	http://www.tcoffee.org/ (Barcelona, ES) [28]
UCSF Chimera	http://www.rbvi.ucsf.edu/chimera/ [29]
UNICORN™ control software	GE Healthcare (Chalfont, St. Giles)
UniProt	http://www.uniprot.org [30]
XDS Programme Package	W. Kabsch (Heidelberg, DE) [31, 32]

5.2 Methods

5.2.1 Molecular biology

5.2.1.1 Cultivation and storage of *Escherichia coli*

All steps related to the cultivation and storage of *E. coli* cells were carried out under sterile conditions. The bacteria grew on LB-agar plates or -medium (5.1.4) containing the appropriate antibiotic(s) (5.1.2). By default, they were incubated at 37 °C and 140 rpm (only liquid cultures) overnight. Incubation temperature and time as well as shaking frequency were adapted to the requirements of the different strains. Glycerol stocks were made to store *E. coli* cells carrying a specific plasmid over a long period of time. For that, bacteria grew overnight to saturation, before up to 2 mL of the culture were spun down at 5000 *g* for 3 min. The supernatant was discarded and the pellet resuspended in 1 mL LB_0 medium with 30 % glycerol. The suspension was then transferred into special cryo vials (5.1.11) and stored in liquid nitrogen at -80 °C.

5.2.1.2 Preparation of plasmid DNA from *E. coli*

To extract large quantities of plasmid DNA, a liquid culture with bacteria carrying the desired plasmid was grown overnight to saturation. Dependent on the type of plasmid (high- or low-copy), the volume of the culture amounted to up to 15 mL. Cell disruption as well as isolation and purification of the plasmid DNA took place according to the manual of the Plasmid Miniprep Kit I or II (5.1.7). The eluted DNA was stored at -20 °C.

5.2.1.3 PCR Amplification

A DNA sequence was selectively amplified by polymerase chain reaction (PCR). The standard protocol was adapted to recommendations in the manufacturer's manual of the Phusion High-Fidelity DNA Polymerase (5.1.6). Codon-optimized, synthetic gene fragments (5.1.10) served as templates for the reaction. By default, primers were designed according to the following rules: each sequence started with four to six adenine nucleotides followed by the selected restriction site and between 15 to 30 bases complementary to the gene of interest. To ensure a solid binding, the sequence of the primer was chosen so that it preferably ended with either a guanine or a cytosine. Depending on the restriction site, some additional bases had to be inserted in front of the sequence complementary to the gene of interest, avoiding a shift within the open-reading frame. The melting temperature for a pair of primers was calculated with the NEB Tm Calculator (5.1.12). All components listed below were pipetted into a PCR Eppendorf tube and mixed thoroughly.

ddH_2O	75 µL
Phusion HF Buffer	20 µL
Template DNA	1 µL
dNTP Mix	2 µL
Forward primer	0.5 µL
Reverse primer	0.5 µL
Phusion HF DNA Polymerase	1 µL
total:	100 µL

Subsequently, the reaction vessel was placed in a thermocycler. The standard amplification protocol is shown in Table 5.

Table 5: Standard amplification protocol for a PCR.

Phase	Temperature	Duration	Number of cycles
Initial denaturation	98 °C	30 s	1
Denaturation	98 °C	10 s	20-30
Annealing	T_M(primer)	20 s	
Elongation	72 °C	15 s/1 kb	
Final elongation step	72 °C	10 min	1

The annealing temperature as well as the length of the elongation step and the number of amplification cycles were adapted to the specific requirements of each reaction. To validate the outcome of the amplification reaction, a 5 µL sample was analysed by agarose gel electrophoresis, whereas the rest of the DNA was purified for further processing.

5.2.1.4 Site-directed mutagenesis

Codons from a given gene sequence were deleted or mutated or new ones inserted via this technique. Mutagenesis primers were designed with the QuickChange® Primer Design tool (5.1.12) and ordered for synthesis by Eurofins Genomics (Ebersberg, DE). Instructions for the mutagenesis reaction were taken from the manual of the QuickChange II Site-Directed Mutagenesis Kit (Agilent Genomics). Normally, four different amounts of template DNA (5, 10, 20 and 50 ng) were tested simultaneously. After the cycling reaction was finished, the reaction mixture cooled down to room temperature, before 1 µL of *Dpn*I (5.1.6) was added and the sample was incubated at 37 °C overnight. Transformation of *E. coli* XL-1 blue or XL-10 gold cells (5.1.3) followed, which were spread out on agar plates containing the appropriate antibiotic later. Not less than two clones were used to inoculate overnight cultures for extraction of the plasmid DNA. Presence of the desired mutation was verified by DNA sequencing.

5.2.1.5 Agarose gel electrophoresis

To analyse the content of DNA samples, they were separated on 1 % agarose gels (1 % (w/v) agarose in 1x TAE buffer (5.1.4)). Prior to loading the gel, DNA samples were diluted in DNA loading buffer (5.1.4). Next to the samples, 10 µL of a DNA ladder mix (5.1.7) were pipetted into the gel to accurately determine the size of the separated DNA fragments. The gel ran at 120 V for 30-40 min in 1x TAE buffer. Then, it was stained in ethidium bromide solution (5.1.4) for about 20 min before DNA bands were visualized under UV light (λ = 365 nm).

5.2.1.6 DNA purification

DNA samples were purified with the peqGOLD Cycle-Pure Kit (5.1.7) to remove remaining reaction components from previous cloning steps. Instructions were followed as stated in the manufacturer's protocol. The purified DNA was stored at -20 °C.

5.2.1.7 DNA concentration measurement

The concentration of a DNA sample was determined measuring its absorption at 260 nm and using the following relation: A_{260} = 1 is equivalent to 50 ng/µL of double-stranded DNA. A wave scan in the range of 220 to 320 nm controlled the purity of the sample.

5.2.1.8 DNA sequencing

DNA sequencing by the GATC Biotech AG (Konstanz, DE) ensured that a gene was inserted correctly into the respective expression plasmid and no frameshift or mutation occurred during the cloning procedure. Information about the required amount and concentration of DNA as well as specific demands for the corresponding primers are given on the company's website. Insertions with a length ≥1 kb were controlled by both forward- and reverse-sequencing. Results were reviewed by comparing the decoded with the expected sequence using the 'T-Coffee Multiple Sequence Alignment' tool (5.1.12, [28]).

5.2.1.9 Restriction digest of DNA

To digest a DNA molecule for either analytical or preparative purposes, it was incubated with the appropriate restriction endonucleases (5.1.6) in their recommended buffer at 37 °C. Normally, a double digest was performed, in which two different restriction endonucleases cut the DNA at the same time. Their compatibility with respect to buffer, specific activity and digestion temperature was checked before, using the NEB Double Digest Finder tool (5.1.12).

An analytical restriction digest verified that the gene of interest had been incorporated into the respective expression plasmid. All components were mixed together as stated below and incubated at 37 °C for 1-2 h. Then the sample was analysed by agarose gel electrophoresis.

In contrast, the aim of a preparative restriction digest was to prepare both the target gene DNA, amplified by PCR, and the appropriate expression plasmid for the following ligation

reaction. They were cut with the same restriction enzymes, so that the complementary ends of the DNA pieces resulting thereof could be linked together later. All components were mixed and incubated at 37 °C for at least 3 h. Afterwards, the digested DNA fragments were purified and stored at -20 °C until further use.

Analytical digest:		**Preparative digest:**	
DNA	3 µL	DNA	20 µL
CutSmart™ Buffer	1 µL	CutSmart™ Buffer	5 µL
Enzyme 1	0.5 µL	Enzyme 1	2 µL
Enzyme 2	0.5 µL	Enzyme 2	2 µL
ddH_2O	5 µL	ddH_2O	11 µL
total:	10 µL	total:	50 µL

5.2.1.10 Ligation of DNA fragments

In the process of ligation, the DNA of the target gene is inserted into a suitable expression plasmid by linking together complementary ends. By default, 50 ng of plasmid DNA were used per ligation. The molar ratio of insert to plasmid DNA was chosen to be in the range of 5:1 to 3:1. Insert and plasmid DNA were mixed with ddH_2O to reach a final volume of 7.5 µL. The reaction approach was heated at 55 °C for 10 min and cooled down on ice for 5 min before 2 µL of T4-ligase buffer and 0.5 µL of T4-DNA-Ligase were added (5.1.6/ 5.1.7). The ligation was either incubated at room temperature for 1 to 2 h or at 4 °C overnight. Afterwards, 1 µL of the ligation product was used for transformation of the appropriate *E. coli* strain (5.1.3).

5.2.1.11 Transformation of electro-competent *E. coli* cells

Electroporation served to incorporate a particular plasmid into the suitable bacterial strain. Per reaction, 1 µL of DNA was added to 30 µL of electro-competent *E. coli* cells (5.1.3), thawed on ice before. The mixture was pipetted into an ice-cooled electroporation cuvette, which was then put into the electroporator (5.1.11). The cells were exposed to an electrical pulse of 2,500 V for about 5.4 to 5.8 ms. Immediately, 1 mL of SOC-medium were given to the cells (5.1.4), and they were transferred into a fresh Eppendorf tube. To allow them developing their newly acquired antibiotic resistance, the bacteria were incubated for 1 h at 37 °C and 1,000 rpm. Subsequently, the cells were either spread out on agar plates and/or used to inoculate a liquid culture, each containing the appropriate antibiotic(s).

5.2.2 Protein chemistry

5.2.2.1 Heterologous gene expression in *E. coli*

To find out which expression strain and what temperature were the most suitable to obtain high yields of soluble, recombinant protein, small-scale expression tests were performed first. Freshly transformed *E. coli* cells grew in a 5 mL-overnight culture. The next day a larger culture of about 50 to 100 mL was inoculated and incubated at 37 °C. The growth of the bacteria was monitored by measuring the optical density of the culture at λ=600 nm (OD_{600}). Once reaching an OD_{600} of 0.5 to 0.8, heterologous gene expression was induced by addition of IPTG to a final concentration of 100 µM. Prior to induction a 1 mL-sample was taken, the cells were spun down and stored at -20 °C. Further samples were collected during the following growth phase, in which the optical density was measured periodically. Meanwhile, the incubation temperature for the culture was decreased, if necessary. Finally, all cells were harvested and the resulting pellet was also kept at -20 °C.

The pellets of all 1 mL-samples were solved in a certain amount of saline (0.9 % (w/v) NaCl), equivalent to their optical density, to take the varying cell densities into consideration. Afterwards, they were analysed by SDS-PAGE (5.2.2.9) to detect overexpression of the gene of interest during the growth of the *E. coli* cells. In addition, the remaining pellet of the expression culture was also dissolved in saline. The bacteria were disrupted by sonication and the lysate was separated into pellet and supernatant via centrifugation. Analysing the protein content of both fractions by SDS-PAGE verified that the target protein had been produced in a soluble form.

After suitable parameters to produce large quantities of soluble, recombinant protein were determined, large-scale expression was performed. Pre-cultures were grown overnight and then used to inoculate the main culture(s) (up to 18 L) in a ratio of 1:20 to 1:100. Main cultures were incubated under the conditions determined beforehand. The pellets of the harvested cells were stored at -20 °C.

5.2.2.2 Production of L-selenomethionine-labelled protein

To experimentally determine the phase information for a yet unknown protein structure, anomalous scatterers must be present at defined positions in the crystal. The simplest way to ensure this is to replace the amino acid methionine with L-selenomethionine (SeMet), thereby, directly introducing heavy atoms at specific, pre-determined sites in the protein.

For production of selenomethionine-labelled protein, main cultures were grown in M9 (5.1.4) instead of LB medium but at the same conditions that were established before in the expression tests. Fifteen minutes prior to induction with IPTG, 0.1 g/L lysine, threonine, phenylalanine, 0.05 g/L leucine, isoleucine, valine and 0.05 g/L SeMet were added to the cultures, according to standard protocols [33]. At the end of the growth period, cells were harvested by centrifugation and pellets were frozen at -20 °C until further use.

5.2.2.3 Cell disruption by sonication

The frozen cell pellet was thawed on ice and homogenously dissolved in a sufficient amount of lysis buffer (for purification of *Hy*Anbu: 50 mM HEPES/NaOH, pH 8.0, 25 mM NaCl). A tip of a spatula of DNase I (5.1.6) was added to the cell solution before the cells were lysed by sonication (5.1.11). For a better cell disruption, dissolved cell pellets were transferred into ultrasonic vessels. Usually, the largest tip of the sonicator with a diameter of 0.5 cm was used. The applied pulse with an amplitude of 40 % was alternately switched on and off for 1 s each. The overall duration of sonication depended on the number of cells that were lysed and varied between 3 and 5 min. Subsequently, the lysate was spun down at 41,415 *g* and 4 °C for 30 min. In the following purification process, the supernatant was loaded onto a chromatographic column, whilst the pellet was discarded.

5.2.2.4 Chromatographic purification of *Hy*Anbu

All purification steps were carried out on an ÄKTA FPLC system (5.1.11) allowing a continuous gradient elution. To monitor elution of the target protein, both the UV absorption signal at 280 nm and the conductivity of the eluate were recorded during each run. The eluate was collected in several aliquots that were analysed by SDS-PAGE thereafter (5.2.2.9). *Hy*Anbu was purified on two different anion exchange columns at 4 °C followed by a size exclusion chromatography (SEC) at 25 °C. Prior to each run, the anion exchange columns were rinsed with 1 M NaCl to remove debris and potential remains from previous purifications. Afterwards, all columns were washed extensively and stored in ddH_2O with 0.02 % NaN_3. For purification of SeMet-labelled *Hy*Anbu, all buffers were supplemented with 5 mM DTT.

Dependent on its volume, the supernatant of the cell lysate was loaded either onto two 5 mL HiTrap Q FF columns in series or a 40 mL self-packed Q Sepharose column (5.1.11) equilibrated in 50 mM HEPES/NaOH, pH 8.0, 20 mM NaCl. The column was washed with

150 mM NaCl (10 CVs) before bound proteins were eluted by gradually increasing the NaCl concentration to 450 mM (7.5 to 10 CVs) followed by two stepwise elutions at 600 mM and 1 M NaCl (5 CVs each). Fractions containing *Hy*Anbu were pooled and dialysed against 25 mM Bis-Tris, pH 6.0, 25 mM NaCl overnight (5.2.2.5). The next day, the protein solution was loaded onto a 5 mL Hi-Trap FF DEAE column (5.1.11) equilibrated in 50 mM Bis-Tris, pH 6.0, 20 mM NaCl. Proteins were eluted by a linear gradient of 15 CV length to a final concentration of 500 mM NaCl and a subsequent stepwise elution using 1 M NaCl (5 CVs).

Peak fractions of *Hy*Anbu were combined and concentrated with a 50,000 MWCO centrifugal filter unit (5.2.2.6). Shortly before it was loaded onto a Superdex 200 10/300 GL column (equilibrated in 20 mM Tris, pH 7.5, 150 mM NaCl), the sample was filtered to remove larger aggregates formed potentially during concentration. All fractions of the *Hy*Anbu peak were pooled and concentrated. For crystallization trials, the buffer was changed to 20 mM Tris, pH 7.5, 40 mM NaCl. Purified protein was stored either at -20 °C or at 4 °C (5.2.2.8).

5.2.2.5 Buffer exchange by dialysis

Dialysis was used to gently exchange the buffer of the protein sample by one more suitable for the next purification step or analysis. For this purpose, the protein was transferred into a dialysis tubing with an MWCO of 14 kDa (5.1.11) equilibrated in the starting buffer. The dialysis tubing was then given into a vessel containing a sufficient amount of the target buffer, normally about 5 L, and was incubated there under slow stirring at 4 °C overnight. Afterwards, the protein was removed from the dialysis tubing for further processing.

5.2.2.6 Concentration of a protein sample

The aim of this technique was to increase the concentration of a protein solution with the help of centrifugal filter units (5.1.11) up to one required for subsequent characterization steps. The MWCO of the filter was chosen according to the molecular weight of the protein of interest. Filters were rinsed beforehand with the same buffer the protein was stored in. The centrifugation speed was chosen as recommended in the manufacturer's protocol. The length of a centrifugation step depended on how aggregation prone a protein was and, usually, lasted not longer than 5 min. Between two steps the protein solution was mixed thoroughly and fresh solution was added, if necessary. When the desired concentration was

reached eventually, the protein solution was transferred into an Eppendorf tube and spun down at 16,000 *g* for 10 min to remove potential aggregates. The cleared protein solution was stored on ice or at 4 °C.

5.2.2.7 Measurement of protein concentration

To determine the concentration of a protein sample, its absorption signal at 280 nm was measured with a Nanophotometer (5.1.11). A spectrum in the range of 220 to 320 nm was recorded to detect possible impurities as *e.g.* remaining DNA. Based on the primary sequence of the protein, the ProtParam tool (5.1.12) computed its specific extinction coefficient. The protein concentration in the sample was calculated using the Lambert-Beer law.

5.2.2.8 Protein storage

To store a protein for a short amount of time, it was kept in the fridge at 4 °C. For long-term storage, however, the solution was diluted with an equal volume of 100 % glycerol and frozen at -20 °C. Prior to further use, it was once more purified by SEC.

5.2.2.9 SDS-PAGE

The protein content of a sample was analysed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). By default, 12 % SDS-gels were used. Table 6 lists the composition of both the stacking and the separating gel. APS and TEMED were added shortly before the gel was cast.

Table 6: Composition of a 12 % SDS-gel.

Component	Separating gel	Stacking gel
ddH_2O	2.25 mL	1 Ml
Separating gel buffer	1.25 mL	-
Stacking gel buffer	-	1.25 Ml
40 % Acrylamide/bisacrylamide (1:29)	1.5 mL	0.25 Ml
10 % APS	75 µL	50 µL
TEMED	7.5 µL	5 µL

To prepare the samples, 20 µL of protein solution were mixed with 5 µL of SDS-sample buffer (5.1.4) before they were heated at 95 °C for approximately 5 min. 10 µL of each sample and a protein length standard (5.1.5) were loaded into the pockets of the gel. It ran at 40 to 45 mA for about 30 min until the SDS-sample buffer reached the lower end of the

gel. Afterwards, the gel was incubated in staining solution for roughly 20 min or overnight before it was decolourized again with destaining solution (5.1.4). The gels were stored in ddH_2O and a picture was taken for documentation.

5.2.3 Activity assays

Aim of these experiments was to determine the proteolytic activity of *Hy*Anbu against both the model protease substrate FITC-casein and the standard proteasomal peptide substrates Z-GGL-AMC, Suc-LLVY-AMC, Boc-LLR-AMC and Z-LLE-AMC. All assays were measured with a Spark^{TM} 10M multimode microplate reader (5.1.11) at 25 °C, unless stated otherwise.

The Pierce Fluorescent Protease Assay Kit (5.1.7) was used to monitor the cleavage of FITC-casein by *Hy*Anbu in the microplate format over the course of 2 h (FRET modus, λ_{ex}=485 nm, λ_{em} =538 nm). Yeast 20S proteasome (yCP) alone or preincubated with 1 mM bortezomib for 1 h served as control.

The peptidolytic activity of *Hy*Anbu was also analysed in the 96-well plate format. The change in fluorescence was measured for about 1 h using excitation and emission wavelengths of 360 and 460 nm, respectively. In a 150 µL drop, 66 nM of protein were mixed with 200 µM substrate (final concentrations) in the buffer of choice, 20 mM Tris, pH 7.5, 150 mM NaCl by default. All samples were measured in triplicates and yCP was used as a positive control. Next to the standard reaction, the cleavage of Z-GGL-AMC was analysed under varying buffer conditions. For measuring the effect of different amounts of salts, the concentration of NaCl was varied between 0 and 1 M. In the pH screen, all of the following buffers were used at a concentration of 100 mM: Citrate, pH 4.5 and 5.5; MES, pH 6.5; Tris, pH 7.5 and 8.5; Bicine, pH 9.5. Besides, the influence of the additives SDS (0.01 %), DTT and glutathione (5 mM each) and the impact of an elevated temperature (37 °C) on the cleavage properties of *Hy*Anbu were tested.

5.2.4 Protein crystallization and structure determination

5.2.4.1 Crystallization

Initial crystallization trials were performed via the sitting drop vapour diffusion technique. To estimate the crystallization tendency of a protein in general and specifically at the

selected concentration, NeXtal Screens 3 to 5 were found to be the most suitable to start with (5.1.11). Next to a reservoir of 40 µL, drops composed of 0.2 µL protein mixed with 0.2 µL reservoir solution were set up in 96 well-plates by a liquid-handling robot system (5.1.11). Plates were incubated at 20 °C and crystal growth was monitored periodically using light microscopy (5.1.11). To decide whether a potential hit might be a protein or a salt crystal, the plates were examined under a UV microscope (5.1.11). In the event that a tendency for crystals to grow particularly well under specific conditions, *e.g.* PEG or MPD, was found, the appropriate screen was used for further fine screening. Additional parameters to be varied were protein concentration, drop size and/or temperature (4 °C, 37 °C). Once a suitable condition was found, customized fine screens were set up manually, potentially switching to the hanging drop vapour diffusion method to enlarge the drop volume and, thereby, the crystal size. Suitable crystals were mounted into cryogenic loops, soaked in cryoprotectant to avoid formation of ice crystals and vitrified in liquid nitrogen at 100 K for transport and storage prior to data collection.

5.2.4.2 Data collection and processing

Diffraction data were measured both at the beamline X06SA at the Swiss Light Source (SLS), Villingen, Switzerland [34] and at the beamline ID23-1 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France [35]. For structure determination via experimental phasing, a fluorescence scan confirmed the presence and type of the anomalous scatterer. In addition, the exact wave lengths for peak and inflection were determined. A first data set at the peak wave length was measured for single-wavelength anomalous dispersion (SAD). All other data sets were recorded with synchrotron radiation of $\lambda = 1.0$ Å.

X-ray intensities were processed, scaled and integrated with the programme packages XDS and XSCALE [31, 32]. For evaluation of anomalous data sets, Friedel pairs were treated as individual reflections. Further data analysis and structure determination were performed using the ccp4i/ccp4i2 programme suite [17, 36]. To assign the screw axes and thereby the space group, the recorded images were checked for systematic absences with HKLview. With the help of AIMLESS the general data quality was examined and the data were tested for twinning [37]. By calculating the Matthew's coefficient, the probable number of molecules within the asymmetric unit and the solvent content of the crystal were determined. Calculation of self-rotation functions by the programme GLFR revealed the

existence of non-crystallographic rotational symmetry [38]. The further procedure was dependent upon the kind of phase information available to solve the structure.

5.2.4.3 Experimental phase determination

Single-wavelength anomalous dispersion proved to be successful for experimental phasing of SeMet-labelled *Hy*Anbu. The automated structure solution pipeline CRANK2 [39] was used to phase the anomalous data set at a resolution of 2.2 Å (PDB ID: 5NYF). After a single solution for the heavy atom structure was found and the correct hand was determined, the resulting phases were already of sufficient quality to generate an electron density map, in which most of the secondary structure elements and side chain residues could be traced automatically.

5.2.4.4 Patterson search techniques

All remaining data sets of *Hy*Anbu as well as the structure of *Lm*ClpP1/2 were solved by Patterson search methods with help of the PHASER software [40]. While the experimentally phased data set (PDB ID: 5NYF) served as a starting model for the different *Hy*Anbu structures, the PDB entries 4JCT and 4JCQ turned out to be suitable search ensembles for the heterocomplex from *L. monocytogenes*. Prior to the first search, all solvent molecules present were taken out and the models were pruned by removing terminal, flexible and/or regions sticking out of the molecule. Depending on the Matthew's coefficient calculated before, the number of copies to search for was defined. A solution was considered to be reasonable if the R_{free}-factor dropped below 45 %. In such a case, it was refined once again against the experimental data using REFMAC5 [41] before model building took place.

5.2.4.5 Model building and refinement

To further improve the solutions obtained by either phasing method, iterative rounds of model building and refinement were performed using COOT [18] and REFMAC5 [41]. After all residues resolved in the electron density map had been included into the protein model and provided the resolution was high enough (≤ 2.8 Å), the hydration shell of the protein was modelled automatically by the programme ARP/wARP [42]. If identifiable, other solvent molecules as *e.g.* ions or small molecules were included into the model. Final TLS-refinement resulted in good R_{work} and R_{free}-values and excellent stereochemistry for all structures as evaluated by calculation of the Ramachandran plot with SFcheck [43] and

Molprobity [22]. Details regarding data collection and refinement are summarised in Table 8 or Table 9 in the result sections, respectively.

5.2.5 Graphic representation and biocomputational analyses

All figures illustrating the molecular architecture of a protein were made with PyMOL [26], MOLSCRIPT [23] or UCSF Chimera [29] and then exported as PNGs for further processing with CorelDraw to produce high quality figures. Different sequences were aligned with the help of the T-Coffee Multiple Alignment Server [28]. The alignment was either used directly or loaded into JalView [20] for further graphical modifications, from which it was also exported as PNG. To calculate the identity or similarity of two protein sequences, the afore alignment was uploaded to the SIAS server, which then determined the corresponding numbers using default settings. ProtParam was used to calculate the most important properties of a protein required for purification as *e.g.* molecular weight, pI or extinction coefficient based on its primary sequence. The secondary structure prediction by JPred [21] and the domain analysis by SMART [27] also relied on it. The 'in-solution' stability and interaction surface of potential oligomeric complexes seen in the crystal structure was analysed by PDBePISA [24]. GraphPad Prism was used to evaluate the row data of the activity assays measured for *Hy*Anbu graphically. Further references for all programmes mentioned here are listed in section 5.1.12.

6 Structural and functional analysis of Anbu from *Hyphomicrobium sp.* strain MC1[2]

6.1 Introduction

6.1.1 The 20S proteasome in eukaryotes and archaea

The viability of all eukaryotic and archaeal cells depends on the proper functioning of the 20S proteasome, responsible for the majority of cytosolic protein degradation [44, 45]. Also referred to as 'prosome' or 'multicatalytic proteinase', the 20S proteasome core particle (CP) is a barrel-shaped protease with a height of ~150 Å and a diameter of ~110 Å [46–48]. It is composed of 28 subunits, arranged in four seven-membered rings, stacked on top of each other [49]. The two outer rings of the particle are formed by α-subunits, whilst the inner ones consist of β-subunits, resulting in the overall complex having a stoichiometry of $\alpha_{1\text{-}7}\beta_{1\text{-}7}\beta'_{1\text{-}7}\alpha'_{1\text{-}7}$ (Figure 3a) [44, 50, 51]. Dividing the CP along its longitudinal axis shows that inside a central channel with three larger cavities connected via two narrow passages exists. The so-called antechambers between the α- and β-rings are slightly smaller than the proteolytic chamber formed by the two β-rings [48]. On both ends, the access to the central channel is restricted by axial pores.

The most sophisticated proteasome types are found in higher eukaryotes. In yeast the CP is composed of seven different α- and seven different β-subunits, all of which take their specific place in the fully assembled particle (Figure 4) [52, 53]. Archaeal proteasomes are in many aspects less complex than eukaryotic ones. Normally, their CP consists of only one type of α- and β-subunit each, as illustrated by the archaeon *Thermoplasma acidophilum* (*Ta*) [51, 54], but there also exist a few archaea that encode two α- and/or β-subunits or even more [55–57]. All α- and β-subunits were found to have a common evolutionary ancestor and, because of their fold, they are classified as members of the family of N-terminal nucleophile (Ntn)-hydrolases (Figure 3b) [58, 59]. The common core of this fold is an αββα motif, that is two layers of β-sheets flanked on top and bottom by

[2] In parallel to writing this PhD thesis, the results described in this section were submitted for publication under the title 'On the trails of the proteasome fold: structural and functional analysis of the ancestral β-subunit protein Anbu' by Vielberg, M.-T. *et al.* and 'The *Y. bercovieri* Anbu crystal structure sheds light on the evolution of highly (pseudo)symmetric multimers' by Piasecka, A. *et al.* (*manuscripts under revision*, see also section 10).

one layer of α-helices each. The α- and β-subunits differ not only in their location in the CP, but also in their function. This is reflected structurally by major differences in their N-terminal regions (Figure 3) [48, 53].

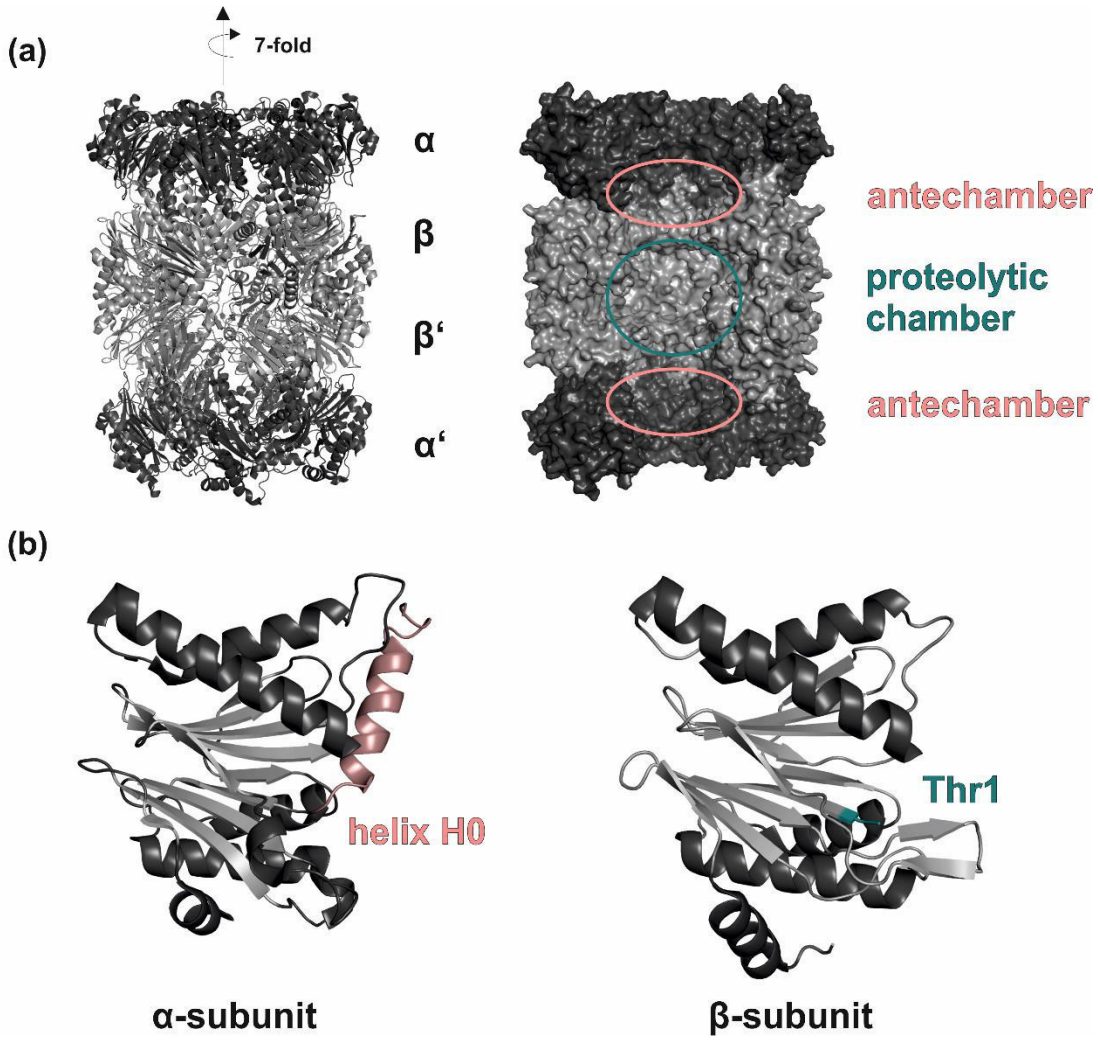


Figure 3: General structure and composition of the 20S proteasome core particle.

(a) The CP has an overall $\alpha_{1\text{-}7}\beta_{1\text{-}7}\beta'_{1\text{-}7}\alpha'_{1\text{-}7}$ stoichiometry, consisting of four seven membered rings formed by two different types of subunits (left). The cross section of the CP shows that there are three chambers located inside connected by a central channel (right). **(b)** The α- and β-subunits share the same fold. Major structural variations are limited almost exclusively to the N-termini of both subunits. The α-subunits possess an additional helix H0 (rose), whereas in the β-subunits Thr1 is the N-terminal nucleophile (teal) (PDB ID: 1PMA [48]; α- and β-subunits of the CP: dark grey and light grey; α- and β-subunit alone: coloured by secondary structure elements, various grey scales).

The α-subunits fulfil three different tasks: they are essential for the assembly of the complex by providing a scaffold for the β-subunits, they serve as a binding platform for a variety of regulatory or accessory proteins and they control access to the catalytic sites [60]. For this, the α-subunits possess an N-terminal helix H0, which is absent in the β-subunits (Figure 3) [48]. The crystal structure of the yeast proteasome shows these helices to narrow down the axial pores at the top and bottom of the CP. In addition, the residues located in front of the

H0 helices form a tight network of interactions, a gate, that closes the entrance of the pore, thereby, preventing the arbitrary degradation of proteins [53]. The most important part in this process plays the N-terminal tail of the α3-subunit illustrated by the finding that its deletion results in opening of the axial pores and a significant enhancement of the basal peptidase activity of the CP [60, 61].

Initially, there seemed to exist a large discrepancy between eukaryotic and archaeal proteasomes regarding their gating mechanism. In structures of the CP from *T. acidophilum* and *Archaeoglobus fulgidus*, the N-termini of the α-subunits were found to be disordered leading to the impression of significantly wider axial pores with unrestricted access to the central channel [48, 62]. However, the residues required for closing the pores in the eukaryotic system are strictly conserved in archaea and their deletion has similar effects as in eukaryotes [62, 63]. Further studies could prove eventually that there exists an axial gate in the archaeal system, but the corresponding regions are quite flexible, constantly switching between an open and a closed conformation [64]. These dynamical changes create the impression of open axial pores as seen in the crystal structures.

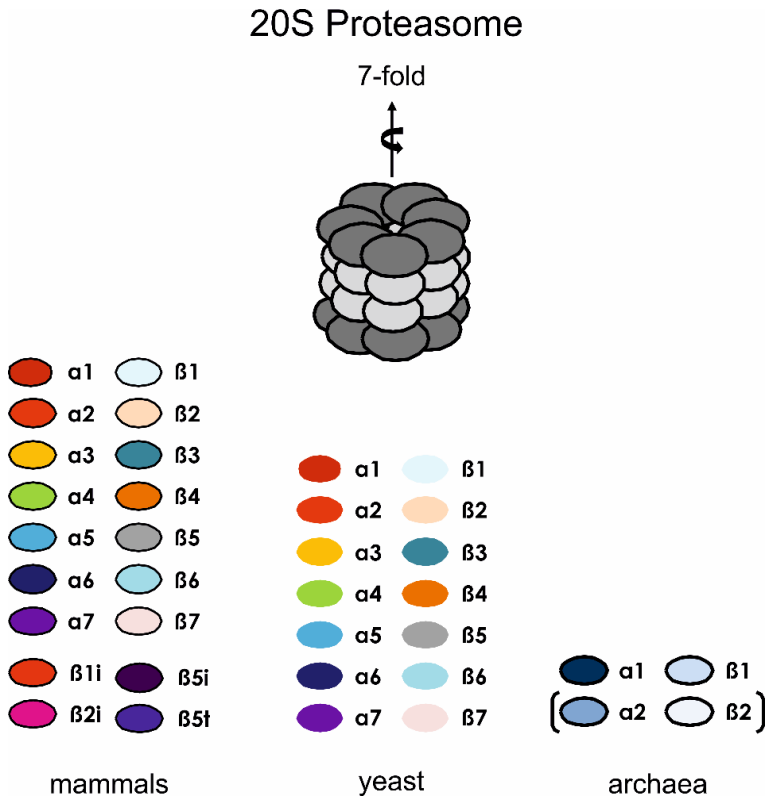


Figure 4: The number of different α- and β-subunits in the CP is species-dependent.

In eukaryotes, the CP is composed of seven different α- and seven different β-subunits. In contrast to unicellular organisms as yeast, mammals also replace the active β-subunits by cell-specific isoforms. Archaeal CPs are usually less complex with only one, sometimes two types of α- and β-subunits.

In contrast to the α-subunits, the β-subunits harbour the active centres of the protease (Figure 3). Mutational and structural studies showed that the CP is an N-terminal threonine protease with the catalytic nucleophile located at the tip of strand S1 of the β-subunits [48, 65]. Out of the seven different β-subunits in eukaryotic proteasomes only three are proteolytically active: β1, β2 and β5 [53, 65]. Using mutational studies and *in vitro* substrate assays, distinct peptide cleavage preferences could be assigned to each of the active subunits, designated as chymotrypsin-, trypsin- and casein-like, whilst *in vivo* experiments demonstrated the ability of the proteasome to hydrolyse almost every peptide bond, independently of its amino acid context [66, 67]. In mammals, the three active β-subunits of the constitutive CP can be replaced by cell-specific isoforms, called β1i, β2i, β5i or β5t (Figure 4). Their incorporation leads to the formation of two additional types of proteasomes, the immuno- and the thymoproteasome with certain responsibilities in cellular homeostasis [68]. Since archaeal CPs normally include only one type of β-subunit, their activity towards peptidic substrates is in most cases reduced to the chymotrypsin-like cleavage preference, but again this has no considerable impact on the hydrolysis of proteins [69].

Catalytically active β-subunits are produced as inactive precursors with a pro-sequence ahead of Thr1 [70]. These sequences offer protection against N-terminal modification, which would impede or eliminate the catalytic activity, but are also discussed to play a role in the folding of the β-subunits as well as in maturation of the proteolytic particle in general [71]. With the exception of the halophilic archaea, archaeal pro-sequences are rather short compared to eukaryotic ones [72]. Despite variations in their length and sequence, they all bear the characteristic Gly(-1)Thr1 motif, which is an essential requirement for the autocatalytic cleavage of the pro-sequences [73, 74]. In addition, it was found that the pro-sequences bind into the substrate channel of the β-subunits by forming an anti-parallel β-sheet, the same way a substrate would do [53]. So, most likely there exists a common mechanism in the CP for cleavage of a peptide bond of either a substrate or the own pro-sequence.

The proteolysis reaction starts with the nucleophilic attack of Thr1O$^{\gamma}$ on the carbonyl carbon atom of the substrate's scissile peptide bond resulting in formation of the first cleavage product and the covalently bound acyl-enzyme intermediate [53]. The latter is cleaved by addition of a nucleophilic water molecule releasing the remaining peptide

fragment and regenerating the active site. In contrast, at the beginning of autolysis the initial attack of Thr1O^{γ} on its own pro-sequence leads to formation of a hydroxy-oxazolidine ring, because the Thr1 N-terminus is still involved in peptide bond formation with Gly(-1) [75]. Protonation of the transition state then leads to freeing of the Thr1 N-terminus and binding of the pro-sequence to ThrO^{γ} in the same way as found in the acyl-enzyme intermediate during substrate cleavage [75]. The following steps take place accordingly.

Two potential candidates for initial activation of Thr1O^{γ} via deprotonation were identified: Lys33 and Thr1NH_2 [48, 74]. However, involvement of the latter would require two different mechanisms for auto- and proteolysis and so it is the concerted action of Asp17 and Lys33 that activates Thr1O^{γ} in its function as Brønsted base [76]. In contrast, the Thr1 N-terminus is protonated by the cooperative effect of Asp166 and Ser129 and acts as Brønsted acid for incoming peptide substrates to generate a proper leaving group of the emerging free amine in the first cleavage product. Subsequently, it also prepares the attack of the nucleophilic water molecule by its deprotonation in both auto- and proteolysis reactions. Hence, the CP possesses two catalytic triads, Thr1O^{γ}-Lys33N^{ε}-Asp17$O^{\delta 1}$ and Thr1NH_3^+-Ser129O^{γ}-Asp166O^{δ} – likewise indispensable for its proper function [76]. Notably, in a few archaeal CPs, the active site Asp17 is replaced by Glu17, which, however, has no influence on their activity [48, 62].

The β-subunits only gain their proteolytic activity after full maturation of the CP, a complex process that can be divided into various steps. First, the α-rings are formed, to which the pro-β-subunits attach to constitute a so-called half-proteasome. Afterwards, binding of two of these half-proteasomes via their β-rings leads to assembly of the entire particle. Last, the pro-sequences are removed by the autocatalytic cleavage reaction described above, completing the maturation of the proteolytic chamber and the CP as a whole [77]. In eukaryotes, where seven different α- and seven different β-subunits need to find their specific place within each half-proteasome, assembly chaperone complexes assist the process. Indeed, heterologous overexpression of specific α-subunits leads to spontaneous double α-ring formation indicating traces of self-assembly [78], but a high binding tendency between α-subunits of the same kind prevents heterogenous ring composition. This is one thing, the assembly factors need to monitor closely; just as they also must impede the too tight association of α-rings with each other. Eukaryotic β-subunits have specific C-terminal extensions that mediate their contact to others and, thus, guide the

assembly of the β-rings on the basic framework the fully assembled α-ring provides [79, 80]. As mentioned before, catalytically active β-subunits are usually produced as precursor forms. For both the completion of the β-ring assembly and the following maturation of the entire proteolytic particle out of the two half-proteasomes, the pro-sequence of the β5-subunit seems to be of particular importance. Its ability to promote these processes, even if expressed in *trans*, illustrates its chaperone function for proteasome assembly [70].

In archaea, the lower number of different subunits simplifies the assembly process, making it lesser to not at all dependent on auxiliary proteins. It could be shown that heterologous expression of the α-subunits of the *Ta*CP alone leads to spontaneous ring-formation, whilst the β-subunits on their own remain monomeric, unprocessed and inactive. Mixed together both subunits are able to form a fully assembled CP [81]. The role the pro-sequences of the archaeal β-subunits play in this process is still under debate. They are not essential for the *in vitro*-assembly of the *Ta*CP [82], which is in line with the recent discovery that some archaeal β-subunits of the Asgard archaea do not encode a pro-sequence, whilst *in vivo*-experiments with halophilic archaea seem to indicate the opposite [72].

On its own, the CP has only a basal peptidase activity, which is generally higher in the archaeal proteasomes than in the eukaryotic ones, quite likely related to the distinct gate formation described before. For full proteolytic activity, the CP must interact with regulatory proteins that are responsible for selecting the appropriate substrates, opening the gate in the α-rings and (actively) translocating the substrates into the central channel and further into the proteolytic chamber. The main interaction partner of the eukaryotic proteasome is the 19S regulatory particle, which binds on top and bottom of the CP forming the so-called 26S proteasomes. The regulatory particle, consisting of 19 subunits in total, is divided into two different parts, the lid and the base [83]. Part of the latter are six AAA^+-ATPases, Rpt1-6, which form a hexameric ring structure and interact directly with the α-subunits of the CP, so they can open the axial pores and thread the unfolded substrate proteins into the proteolytic particle [84]. All other subunits are involved in substrate recognition and binding as well as removal of ubiquitin, a small regulatory protein that works as a degradation signal in eukaryotic cells [85].

The best characterised archaeal adaptor protein is PAN, discovered due to its homology to the eukaryotic Rpt-ATPases of the 19S particle [86]. PAN also forms ring-shaped, hexameric particles and interacts with the α-subunits via a conserved HbYX motif [87].

Some archaea do not encode PAN, but other AAA$^+$-ATPases with a similar proteasome-interacting region leading to the assumption that the proteolytic activity of the archaeal CPs might be controlled by a network of different adaptor proteins [88, 89]. In addition, at least in some archaea, there seems to exist a system resembling the eukaryotic ubiquitin-tagging mechanism [90].

6.1.2 Proteasomes and proteasome-like particles in bacteria

6.1.2.1 Actinobacteria and the 20S proteasome

In contrast to eukaryotes and archaea, the presence of the CP is extremely rare in bacteria, probably limited to species belonging to the class of Actinobacteria, *e.g. Frankia* sp. [91], *Streptomyces coelicolor* [92], *Rhodococcus erythropolis* [93] and *Mycobacterium* spp. [94] as well as a few of the order *Nitrospirales* [95]. The CP is dispensable for the normal growth of these organisms [96], but at least for *M. tuberculosis* it was found to be essential for the virulence of the bacterium [97]. This and the sparse occurrence of the CP suggest that the corresponding genes were originally transmitted by horizontal gene transfer to a common ancestor of these bacteria [98]. Although the architecture of the bacterial CP is as simple as that found in archaea, some features like the length of the pro-sequences or certain steps in the assembly of the complex strongly resemble those of eukaryotic proteasomes [91].

Bacterial CPs consist of one type of α- (*prcA*) and one type of β-subunit (*prcB*), with one exception in the *R. erythropolis* strain NI86/21 (Figure 5) [93]. This bacterium possesses two different copies of each subunit organized in two separate operons, one of which was probably obtained by lateral gene transfer from a close relative [98]. The CPs formed in *Rhodococcus* were found to be composed of all four subunits [93], whereas *in vitro* experiments showed that all possible combinations of one α- and one β-subunits are able to assemble into proteolytically active proteasomes. In contrast to studies with the archaeal α-subunits [82], heterologous expression of the *prcA* genes alone did not result in spontaneous α-ring-formation [99]. By themselves, the β-subunits also remained monomeric and unprocessed. After mixing the individual subunit types, rapid assembly of proteolytically active particles was observed suggesting a different first step for the proteasome formation in bacteria. Half-proteasomes seem to be rather built up by heterodimers than via initial assembly of α-rings as docking platforms for the β-subunits [100]. Further maturation of the CP proceeds the same way as seen in archaea and eukaryotes. The importance of the bacterial β-pro-sequences for this process was found to

be species-dependent. So, the pro-sequences in *M. tuberculosis* impede proteasome maturation [101], whereas in *Rhodococcus* they promote it, even if expressed in *trans* [100, 102].

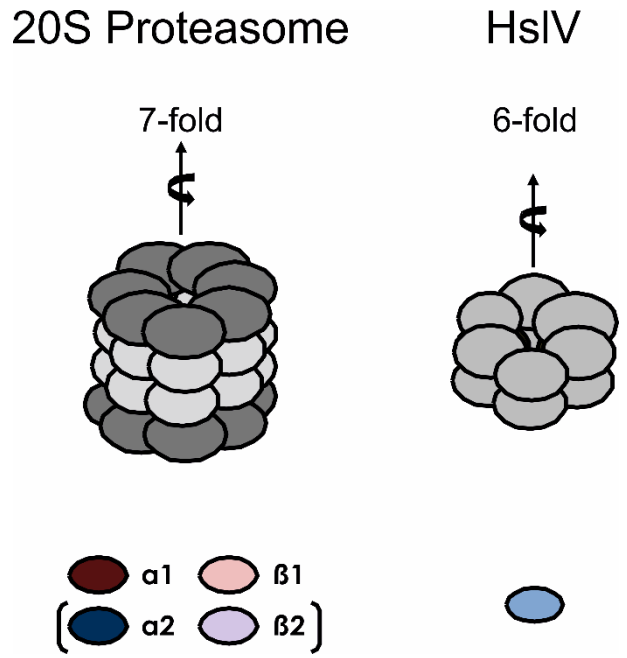


Figure 5: Structure of the two N-terminal threonine proteases found in bacteria.

Only a very few bacteria, mostly actinomycetes, encode a CP, generally consisting of one α- and one β-subunit as illustrate schematically above. In other species, the proteasome-like protease HslV is found. The subunits of HslV are closely related to the proteasomal β-subunits, but no equivalent to the α-subunits exists here. HslV is a dodecamer formed of two six-membered rings, thus, the radial symmetry is altered compared to the CP.

The presence of only one kind of β-subunit in the bacterial CP reduces its cleavage preference for peptidic substrates to one specificity: the chymotryptic-like, the same as in archaeal proteasomes [91, 92]. Remarkably, for the CP of *M. tuberculosis* it was shown that, due to the architecture of its substrate binding channel, it exhibits all three different peptide cleavage activities as found in eukaryotic proteasomes [101]. Nonetheless, the full catalytic function of bacterial CPs depends on their interaction with regulatory proteins. In the crystal structure of the mycobacterial CP, the axial pores are closed [103] and deletion of the first eight N-terminal residues of the α-subunits clearly enhanced the peptidolytic activity [101] indicating that these proteasomes rely on an N-terminal gating mechanism as well. The bacterial homologue of PAN and the eukaryotic Rpt-ATPase subunits is called Mpa or ARC [104, 105]. A recent crystallographic study showed that, in contrast to its archaeal and eukaryotic homologues, Mpa possesses an additional C-terminal domain, which might interfere with robust proteasome binding or require the assistance of additional co-factors [106]. Nonetheless, there also exists a bacterial targeting system for proteasomal

substrates and it was proven that Mpa recognizes such proteins and assists their degradation by the mycobacterial CP [107].

6.1.2.2 HslV – a bacterial homologue of the CP

Many of the bacteria that do not possess a CP encode the related, proteasome-like threonine protease heat shock locus V (HslV) [108]. A sequence comparison revealed the relationship of HslV with the β-subunits of the 20S proteasome – the sequence identity between them amounts to 25 to 30 % [109].

HslV forms a dodecameric particle composed of two hexameric rings stacked one above the other (Figure 5). No equivalent to the proteasomal α-subunit exists in this system [110]. The nature of the $\beta_6\beta'_6$-complex was analysed both by electron microscopic [111] and X-ray crystallographic studies [112]. The interaction of the two hexameric rings is similar to that between the $\beta_7\beta'_7$-rings in the CP enclosing a central cavity, where the active centres of the protease are located. This resemblance is also reflected in the fold of the individual monomers, apart from strand S10 and helix H5; both are missing in HslV [112]. Their lack probably enables the tighter subunit packing during ring-formation explaining the difference in symmetry compared to the CP [113]. Due to a small α-helical insertion in the loop connecting helix H2 and strand S5, the apical pores are narrower in HslV than in the proteasomal β-rings, a structural adjustment potentially compensating for the absence of α-subunit pendants in the bacterial system [112].

Although the active site residues are the same as in the CP [114, 115], HslV shows negligible hydrolytic activity against peptidic substrates and none against proteins [110, 116]. To gain its proteolytic activity, HslV must interact with the AAA^+-ATPase HslU, encoded in the same operon and co-transcribed with it [108, 109]. *In vivo* complex formation was proven by co-immunoprecipiation from cellular extracts [115]. In presence of ATP, HslU assembles into hexameric rings, which are then able to bind to both ends of the double-hexameric HslV particle forming a bipartite destruction complex – in contrast to others, without a symmetry-mismatch [111, 113, 116]. Despite this difference, substrate selection, unfolding and translocation into the proteolytic particle remain the principal tasks of HslU. The importance of the interaction between the protease and the ATPase is not only expressed in the increase of the proteolytic activity of HslV by about two magnitudes but also in the significant stimulation of the ATP hydrolysis rate of HslU [117].

In the light of the strong structural resemblance between HslV and the proteasomal β-subunits and the functional equivalence between HslU and the proteasomal ATPases, it is remarkable that these two kinds of regulatory proteins differ both in their sequence and structure. HslU shares about 50 % sequence identity with the ATPase ClpX from the Clp family instead (see also section 7.1.4) and they form a separate ATPase family together [115, 118]. The differences on the structural level are reflected in an altered mode of action as seen in crystallographic studies on HslUV from *Haemophilus influenzae* [119] [120] . As mentioned above, the binding of regulatory proteins to the CP is responsible for opening of the pores in the α-subunits. The pore widths of HslV alone or in complex with HslU are though almost the same [120, 121]. Nevertheless, the double-sided binding of HslU leads to a relocation of the apical helices, a movement that affects the immediate surroundings of the active site. The rearrangement of residues in the substrate binding channel of HslV are inevitable to enable binding of a potential peptide ligand or inhibitor by antiparallel β-sheet formation [120]. In addition, the carboxyterminal tails of HslU stretch out to bind in the outer clefts between adjacent HslV protomers [119], participating in the formation of both the S1 and the S3 pocket of the protease [120]. Taken together, the proteolytic activity of HslV is allosterically activated by HslU via complex formation.

6.1.3 Anbu - the missing link in proteasomal evolution?

All proteasomal α- and β-subunits must result from the same evolutionary origin, they also share with HslV. For quite some time, HslV was seen as the 'simpler' version and, thus, ancestral to the CP [122], also owed to the fact that, until recently, no other homologue of the family of N-terminal threonine proteases was discovered. From an evolutionary point of view this was problematic since HslV is indeed found in many bacterial species, but absent from others. This includes all cyanobacteria, which, according to fossil records, are believed to be amongst the oldest living organisms on earth [122, 123]. The ostensible lack of proteasomal relatives in this bacterial phylum raised questions about the structural and functional roots of the N-terminal threonine proteases.

A study that claimed the existence of two HslV paralogous both in *Enterococcus faecium* and *Magnetospirillum magnetotacticum* provided the initial hint for a way out [122]. In the first case, the organism probably obtained the second copy of HslV by lateral gene transfer, whilst in the second, gene duplication seemed to be the most plausible explanation. However, in 2008 an elaborate biocomputational analysis of around 240 bacterial genomes

showed that one of the two proteins in *M. magnetotactium* was an, until then unknown, homologue of the CP and HslV [14]. It was named Anbu, short for ancestral beta-subunit protein. Cluster analysis indicated Anbu forming an autonomous group within the proteasome family just like HslV as well as the α- and the β-subunits of the CP, which it seems to resemble the most. Sequence alignments and structure prediction assumed that Anbu has a fold similar to the β-subunits with extensions in some loop regions and at its C-terminus. Besides, pendants to Thr1, Asp17 and Lys33 could be identified indicating a potential proteolytic activity, but no sign of a pro-sequence was detected. Remarkably, this new proteasomal homologue is found in several cyanobacteria as well as in α-, β- and γ-proteobacteria, whereby it is often present in one species and missing in the closest relative. The authors of the study interpreted this as a hint of Anbu being a quite ancient protein, possibly the ancestor of the whole proteasome family. Analysis of the genetic environment of Anbu showed that it is normally encoded in one operon together with three other, yet uncharacterised proteins. However, neither of them seems to be a member of the HslU/ClpX or the proteasomal ATPase family and also no other suitable candidate for this task was detected to date [14].

It was our goal to characterise both the function and structure of Anbu via biochemical and X-ray crystallographic analysis to achieve molecular insights into the protein itself as well as about the relationship with its homologues and their evolutionary roots.

6.2 Results

To select the most promising candidate for structural and functional analysis, different Anbu proteins from several eubacteria belonging to distinct phyla were studied in a parallel approach. Out of these, the structure of the alpha-proteobacterium *Hyphomicrobium sp.* strain MC1 (*Hy*Anbu) could be solved successfully.

6.2.1 Cloning and purification of *Hy*Anbu [3]

The *anbu* gene in *Hyphomicrobium sp.* has a length of 732 bp, encoding a 27,500 Da large protein. Sequence analysis and secondary structure prediction by the XtalPred Server revealed that only one longer region around residue number 70 is expected to be disordered [124]. For the rest of the protein, secondary structural elements could be assigned and it was classified as sufficiently stable – beneficial conditions for purification and crystallization. To facilitate expression in *E. coli*, a codon-optimized variant of the gene was cloned into the pET-Duet vector (see sections 5.1.8 and 5.1.10), which did not have any impact on the primary sequence of *Hy*Anbu. Apart from the wild type protein ($Anbu^{wt}$), a Gly(-1) mutant ($Anbu^{mut}$) was cloned to test Anbu's ability of removing a pro-sequence in front of Thr1 with the aim to analyse whether $Anbu^{mut}$ could remove both the methionine and the glycine residue via autolysis.

Addition of a C-terminal affinity tag to simplify purification turned out to be disadvantageous for several Anbu proteins as it resulted in spontaneous aggregation while concentrating the protein. Since Anbu was predicted to be an N-terminal threonine protease, use of a tag at this position was excluded a priori. Hence, after recombinant expression in *E. coli* BL21 (DE3) cells, $Anbu^{wt}$ and its mutant were purified in their native forms, using anionic exchange chromatography and SEC (see section 5.2.2.4). The retention volumes of both Anbu variants implied that they form larger complexes of about ten to twelve subunits in solution (Figure 6). To exclude a concentration-dependent effect, different amounts of protein were loaded onto the column, all leading to the same outcome. Even after freezing, *Hy*Anbu eluted as a multi-subunit particle suggesting it must form a reasonably stable

[3] Cloning and expression of the distinct *anbu* genes as well as first purification and crystallization attempts for the proteins were initiated by Dr. Anja List, a previous PhD student in the group of Prof. Dr. Michael Groll and performed by Verena Bauer during her Master's thesis under Prof. Groll's and my supervision. Although she obtained crystals of $Hy Anbu^{mut}$ diffracting to about 3 Å, the structure could not be solved due to twinning and lack of a suitable model for Patterson search methods. Eventually, changes in the purification strategy, crystallization and use of SeMet-labelled protein allowed us to solve the structure of *Hy*Anbu via SAD-phasing.

complex in solution. The identity of the purified proteins was confirmed by mass spectroscopic analysis. Moreover, it showed that the start methionine was missing in both peptide chains, releasing Thr1 as N-terminal nucleophile in Anbu^{wt}, whereas the Gly(-1) in Anbu^{mut} was still present.

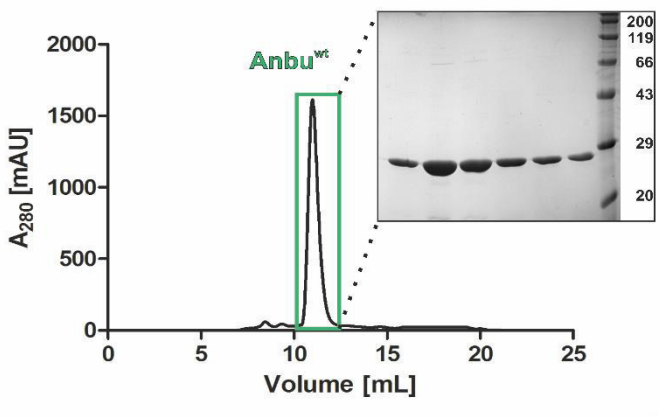


Figure 6: Analysis of the oligomeric state of *Hy*Anbu via size exclusion chromatography.

*Hy*Anbu was purified on a Superdex 200 10/300 GL column. According to the retention volume of approximately 11 mL, it must form a multi-subunit complex 10- to 12-times larger than the monomer with a molecular weight of about 27 kDa. The corresponding sections of the SDS-gel are depicted on the right (for copyright information see section 10).

6.2.2 Activity measurements

The presumption that Anbu might be another member of the family of N-terminal threonine proteases was clearly supported by the finding that residues complementary to Thr1, Asp17 and Lys33 are conserved in Anbu's primary sequence. Since the sequence similarities in Anbu, the CP and HslV are mostly limited to the first sixty amino acids following Thr1, it was not possible to identify potential counterparts of Ser129 and Asp166 based on the sequence alignment. Nonetheless, the activity of Anbu against standard proteasomal substrates was analysed. Unexpectedly, *Hy*Anbu was neither active against peptidic AMC-substrates as *e.g.* Z-Gly-Gly-Leu-AMC (Figure 7a-e) nor FITC-labelled casein (Figure 7f). To exclude a limitation of the experimental set up, several parameters as temperature, salt or pH were varied as well as the impact of some additives were tested (Figure 7b-e). Furthermore, 20S proteasome from yeast (yCP) was used as an internal, positive control in each measurement.

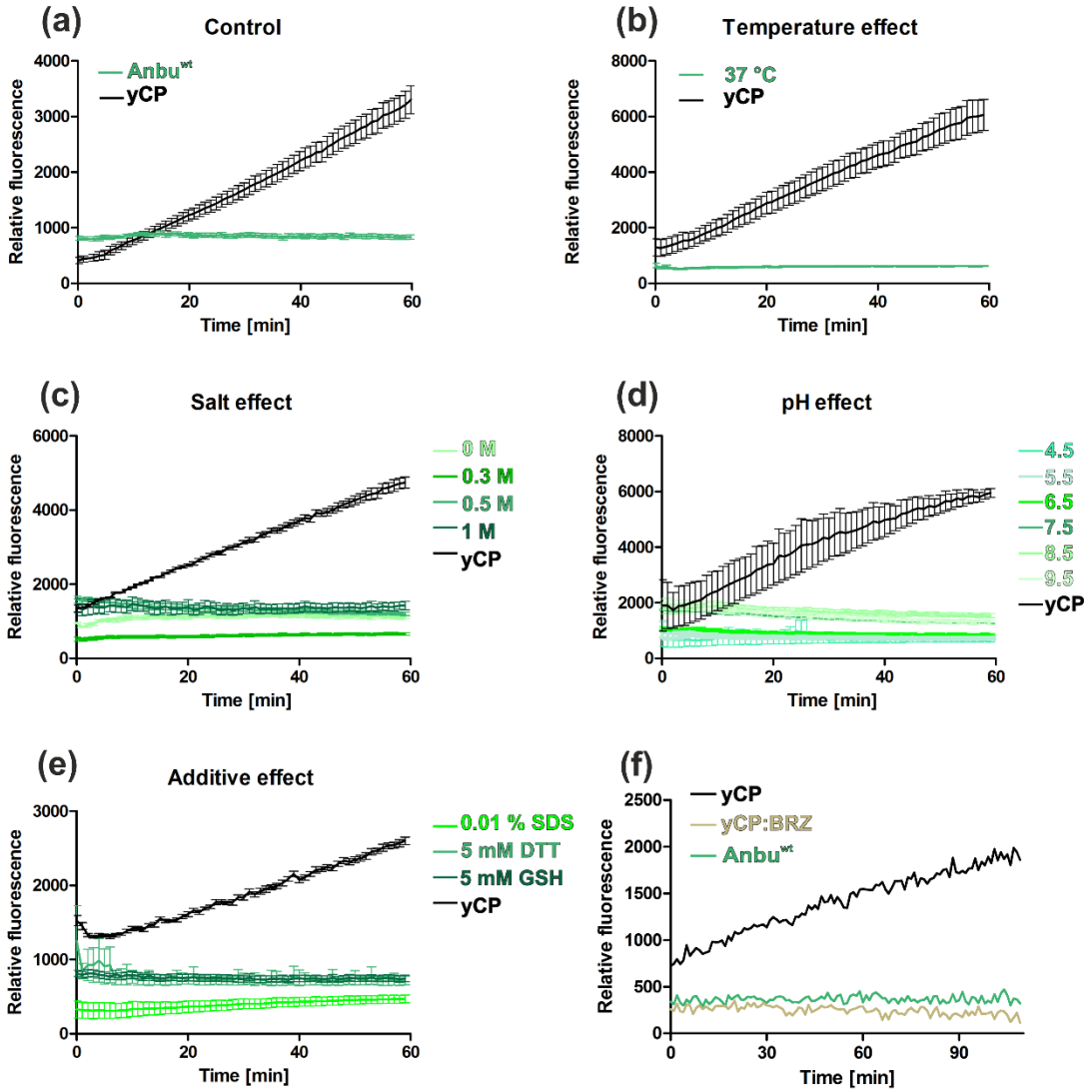


Figure 7: *Hy*Anbu does not cleave peptidic AMC-substrates nor FITC-casein.

(a) The catalytic activity of Anbu^{wt} against Z-Gly-Gly-Leu-AMC in 20 mM Tris, pH 7.5, 150 mM NaCl was measured as change in fluorescence over time. Variations in temperature **(b)**, salt concentration **(c)**, pH **(d)** or additives **(e)** had the same result: *Hy*Anbu did not cleave the substrate (green curves). Yeast 20S proteasome (yCP) served as positive control (black curve). Three replicates are shown as mean ± SD. **(f)** Monitoring the cleavage of FITC-casein led to a similar outcome. Despite its low basal peptidase activity, yCP (black curve) was able to degrade the substrate, whereas Anbu (green curve) behaved like bortezomib-inhibited, catalytically blocked yCP (yCP:BRZ, golden curve, for copyright information see section 10).

6.2.3 Crystallization and structure determination

To define the oligomeric state of Anbu more precisely and in search of an explanation for the missing activity in both pro-peptide and substrate cleavage, Anbu^{wt} and Anbu^{mut} were examined structurally by X-ray crystallography. Interestingly, the mutant protein crystallised much faster, under more diverse conditions and in larger numbers than Anbu^{wt}. Preliminary X-ray analysis of crystals from *Hy*Anbu had shown before that they diffracted to about 3 Å or better, differing from all other Anbu variants studied so far, which did not diffract to resolutions better than 6 Å. These first investigations also revealed that the low

sequence identity between Anbu and the subunits of the CP and HslV would make it difficult, if not impossible to solve the structure by Patterson search methods.

The primary sequence of *Hy*Anbu contains four methionine residues, which presumably produce a strong anomalous signal, when replaced by the non-proteinogenic amino acid L-selenomethionine (SeMet). Due to the higher crystallization tendency, SeMet-labelled Anbu^{mut} protein was produced instead of Anbu^{wt}, using the naturally occurring feedback inhibition of methionine biosynthesis [125]. This method was selected, because a sufficient amount of soluble protein was obtained by heterologous expression of *anbu^{mut}* in *E. coli* BL21(DE3). Prior to induction with IPTG, an excess of the amino acids lysine, threonine, phenylalanine, leucine, isoleucine and valine as well as SeMet is added to the medium. The proliferating cells use up the remaining methionine and since its biosynthesis is blocked, the bacteria fall back to SeMet in the culture medium. Cells grew at comparable rates in LB- and M9-medium resulting in high yields of recombinant, SeMet-labelled protein, which crystallised as good as the native form of Anbu^{mut}.

The preliminary X-ray analysis had shown that the first data sets recorded of *Hy*Anbu crystals were twinned. Since this prevents structure elucidation, a multitude of precipitants as well as crystallization at 37 °C, next to 20 °C, were tested. A variety of different crystal shapes and sizes for both labelled and non-labelled mutant and wild type protein were achieved (Figure 8) and for each of the crystals, a data set was measured and analysed further as listed in Table 7.

Table 7: Overview of crystallization conditions and cryoprotectants for *Hy*Anbu crystals.

	Anbu^{mut}(SeMet) R32	**Anbu^{mut} R32**	**Anbu^{mut} $P2_12_12_1$**
Crystallization condition	0.8 M Sodium phosphate 0.8 M Potassium phosphate 0.1 M HEPES, pH 7.5	0.8 M Sodium phosphate 0.8 M Potassium phosphate 0.1 M HEPES, pH 7.5	0.2 M Di-sodium tatrate 20 % PEG 3350
Cryoprotectant	20 % Ethylenglycol	20 % Ethylenglycol	20 % PEG 400
	Anbu^{wt} R32	**Anbu^{wt} R3**	**Anbu^{wt} C2**
Crystallization condition	0.1 M Tris, pH 8.5 2.0 M Ammonium phosphate	0.05 M Ammonium sulfate 0.05 M Bis-Tris, pH 6.5 30 % Pentaerythriol-propoxylate	0.01 M Cobalt(II) chloride 0.1 M Tris, pH 8.5, 20 % PVP K15 (in presence of 1 mM BRZ)
Cryoprotectant	20 % Ethylenglycol	-	-

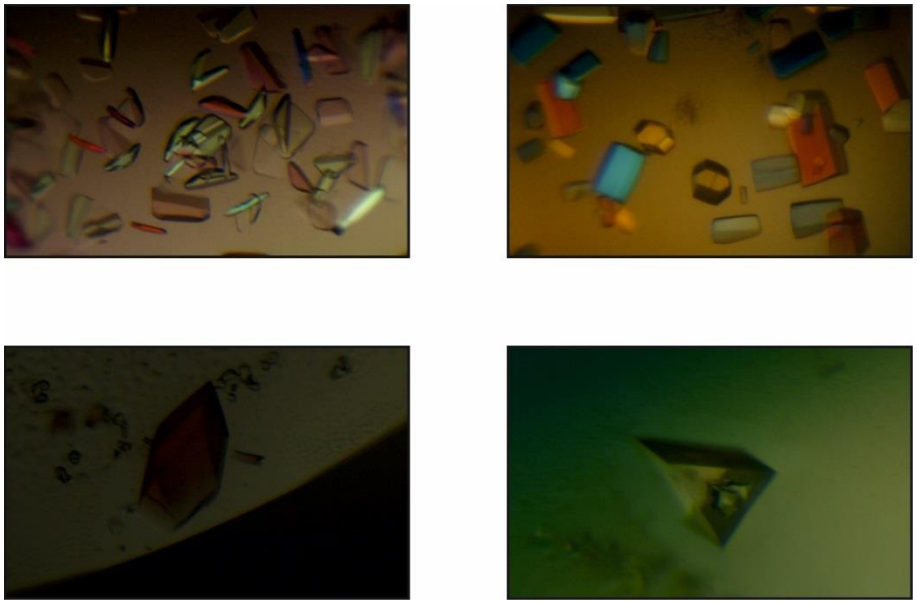


Figure 8: *Hy*Anbu crystals occur in several different shapes and sizes (viewed with a polariser).

Measurement of an X-ray fluorescence spectrum with a subsequent edge scan confirmed the presence of SeMet in the crystals of Anbu^{mut} and determined the peak wavelength to λ=0.9787 Å (f'=-8.75, f''=5.3), where a single dataset for SAD-phasing was recorded (Table 8). Data processing revealed that SeMet-labelled Anbu^{mut} crystallised in space group R32 with a resolution of 2.2 Å and unit cell parameters of a=b=176.13 Å, c=39.74 Å, α=β=90° and γ=120°. Calculation of the Matthew's coefficient predicted one molecule of *Hy*Anbu in the asymmetric unit (AU) with an estimated solvent content of about 43 %. The CRANK2 software suite found three out of four expected SeMet-sites. Automated model building placed about 90 % of all ordered residues correctly into the electron density map. Further improvement of the model resulted in final R values of R_{work} = 18.2 % and R_{free} = 20.9 %, a root mean square deviation (r.m.s.d.) for bonds lengths and angles of 0.006 Å and 1.2°, respectively and in good stereochemistry with 99.5/0.5/0 % of all residues in favoured, allowed or outlier regions (Table 8, PDB ID: 5NYF).

For both Anbu^{mut} and Anbu^{wt}, a variety of additional data sets was recorded. These structures were solved by Patterson search techniques using the above data set as a starting model. They all possess satisfying R-values and excellent stereochemistry (Table 8, PDB IDs: 5NYG/J/P/Q/R).

Table 8: X-ray data collection and refinement statistics for *Hy*Anbu structures.

	Anbu^{mut} (SeMet)	**Anbu^{mut}**	**Anbu^{mut}**	**Anbu^{wt}**	**Anbu^{wt}**	**Anbu^{wt}**
Crystal parameters						
Space group	R32	R32	$P2_12_12_1$	R32	R3	C2
Cell constants	a=176.13 Å b=176.13 Å c=39.74 Å γ=120°	a=175.55 Å b=175.55 Å c=39.70 Å γ=120°	a=103.25 Å b=116.30 Å c=197.84 Å	a=176.22 Å b=176.22 Å c=39.55 Å γ=120°	a=174.49 Å b=174.49 Å c=39.37 Å γ=120°	a=240.22 Å b=242.15 Å c=107.32 Å β=90.77°
Molecules per AU[a]	1	1	8	1	2	16
Data collection						
Beam line	X06SA, SLS	X06SA, SLS	X06SA, SLS	ID23-1, ESRF	X06SA, SLS	X06SA, SLS
Wavelength (Å)	0.9786	1.0	1.0	1.0	1.0	1.0
Resolution range (Å)[b]	40-2.2 (2.3-2.2)	40-1.95 (2.05-1.95)	40-2.4 (2.5-2.4)	40-1.9 (2.0-1.9)	40-2.0 (2.1-2.0)	40-3.2 (3.3-3.2)
No. observations	68156	176140	491003	77865	96108	373796
No. unique reflections	23158[c]	17067	92239	18422	29744	99540
Completeness (%)[b]	99.1 (99.6)	99.8 (99.7)	98.5 (99.8)	99.6 (99.7)	98.5 (99.7)	98.8 (98.6)
R_{merge} (%)[b, d]	5.5 (59.5)	4.1 (58.4)	7.9 (57.2)	4.5 (59.7)	6.0 (53.2)	8.4 (56.2)
I/σ(I)[b]	11.8 (2.1)	31.5 (4.27)	14.1 (2.7)	17.0 (2.7)	9.8 (2.5)	9.7 (2.0)
Refinement *(REFMAC5)*						
Resolution range (Å)	20-2.2	20-1.95	20-2.4	20-1.9	20-2.0	20-3.2
No. refl. working set	11356	16196	87461	17438	28233	94184
No. refl. test set	598	853	4603	921	1486	4957
No. non hydrogen atoms	1724	1751	15474	1809	3447	29884
Solvent molecules	47	62	760	59	102	0
R_{work}/R_{free} (%)[e]	18.2/20.9	17.7/21.4	20.4/24.2	19.0/21.5	18.2/21.3	23.5/27.9
r.m.s.d. bond (Å)/ angle (°)[f]	0.006/1.1	0.006/1.1	0.006/1.0	0.006/1.0	0.006/1.0	0.006/0.9
Average B-factor (Å^2)	56.5	43.4	46.0	42.8	59.1	112.2
Ramachandran Plot (%)[g]	99.5/0.5/0	100/0/0	98.3/1.7/0	99.0/1.0/0	99.2/0.8/0	97.4/2.6/0
PDB accession code	**5NYF**	**5NYQ**	**5NYG**	**5NYP**	**5NYR**	**5NYJ**

[a] Asymmetric unit.

[b] The values in parentheses of resolution range, completeness, R_{merge} and I/σ(I) correspond to the last resolution shell.

[c] Friedel pairs were treated as individual reflections.

[d] $R_{merge}(I) = \Sigma_{hkl}\Sigma_j | [I(hkl)_j - \langle I(hkl) \rangle] | / \Sigma_{hkl}\Sigma_j I(hkl)_j$, where $I(hkl)_j$ is the j^{th} measurement of the intensity of reflection hkl and $\langle I(hkl) \rangle$ is the average intensity.

[e] $R = \Sigma_{hkl} | |F_{obs}| - |F_{calc}| | / \Sigma_{hkl} |F_{obs}|$, where R_{free} is calculated for a randomly chosen 5 % of reflections, which were not used for structure refinement, and R_{work} is calculated for the remaining reflections.

[f] Deviation from ideal bond lengths/angles.

[g] Number of residues in favoured, allowed or outlier regions.

6.2.4 The structure of *Hy*Anbu

Taking all data sets into account, *Hy*Anbu crystallised in a total of four different space groups: R32 (Anbu^{wt} and Anbu^{mut}), R3 (Anbu^{wt}), $P2_12_12_1$ (Anbu^{mut}) and C2 (Anbu^{wt}) with one, two, eight or 16 copies of the molecule in the AU. Those can be divided again into two separate groups based on the following criteria: (i) crystals of space groups R32 and R3 diffracted to considerably higher resolutions (Table 8), (ii) the resulting structures, however, show a higher degree of structural disorder (up to 15 % compared to 5 %) and (iii) the three-dimensional arrangement of Anbu molecules in these crystals is clearly different to that found in the $P2_12_12_1$ structure of Anbu^{mut} and the C2 structure of Anbu^{wt} (Figure 11).

In contrast, the presence of the Gly(-1) residue in Anbu^{mut} had no clear influence on the overall structure of *Hy*Anbu, so, in the following it will only be distinguished between wild type and mutant protein if it makes a relevant difference for the general outcome.

6.2.4.1 Anbu's subunits adopt the fold of Ntn-hydrolases

A first look at the different structures reveals that an individual subunit of *Hy*Anbu shows the same structural features as seen both in α- and β-subunits of the CP and in the subunits of HslV. The order and three-dimensional arrangement of strands and helices follows the characteristic αββα-motif. Hence, by its fold, Anbu is clearly identified as a member of the family of Ntn-hydrolases (Figure 9).

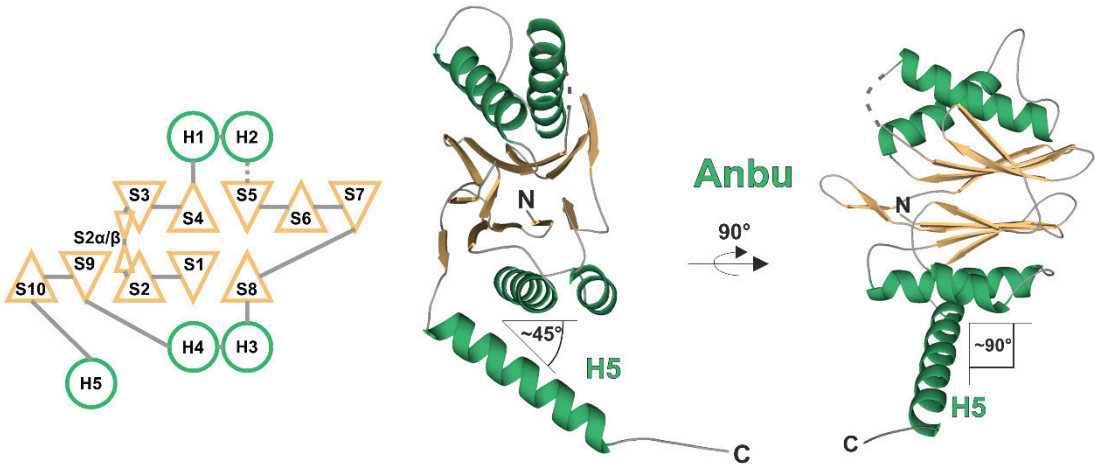


Figure 9: The fold of *Hy*Anbu's subunits is characteristic for Ntn-hydrolases.

The α-helices H1 to H4 and β-strands S1 to S10 of *Hy*Anbu are arranged in the characteristic αββα-motif found in Ntn-hydrolases, as seen in the crystal structure(s) (front and side view: middle, right) and illustrated schematically (left). Compared to the core structure, helix H5 is rotated inwards by about 90° and protrudes from there at an angle of about 45° relative to the α-layer formed by helices H3 and H4 (α-helix: yellow/triangle; β-strand: green/circle; loop region: grey; disordered residues: dashed line, PDB ID: 5NYJ, chain E, for copyright information see section 10).

Multiple sequence and cluster analyses predicted Anbu to be more closely related to the subunits of the CP than to HslV [14]. From a structural point of view, this is supported by the finding that both β-sheets in Anbu are formed by five β-strands each, just like in the proteasomal subunits but opposed to HslV, where strand S10 is missing (Figure 10a) [48, 112]. In agreement with its suggested role as a proteolytically active proteasomal homologue, Anbu has no additional helix H0 as found in the α-subunits. Instead, its encodes threonine as N-terminal amino acid, like all active β-subunits (Figure 10a). On the structural level, there is another crucial feature that distinguishes the two kinds of subunits in the CP: the loop connecting strands S2 and S3. In the α-subunits there is a small helical twist halfway of the loop, whereas in the β-subunits a short β-hairpin is formed on its tip. In

*Hy*Anbu, there is also such a hairpin located in-between the two strands, which will be referred to as S2α/β-loop hereinafter (Figure 9).

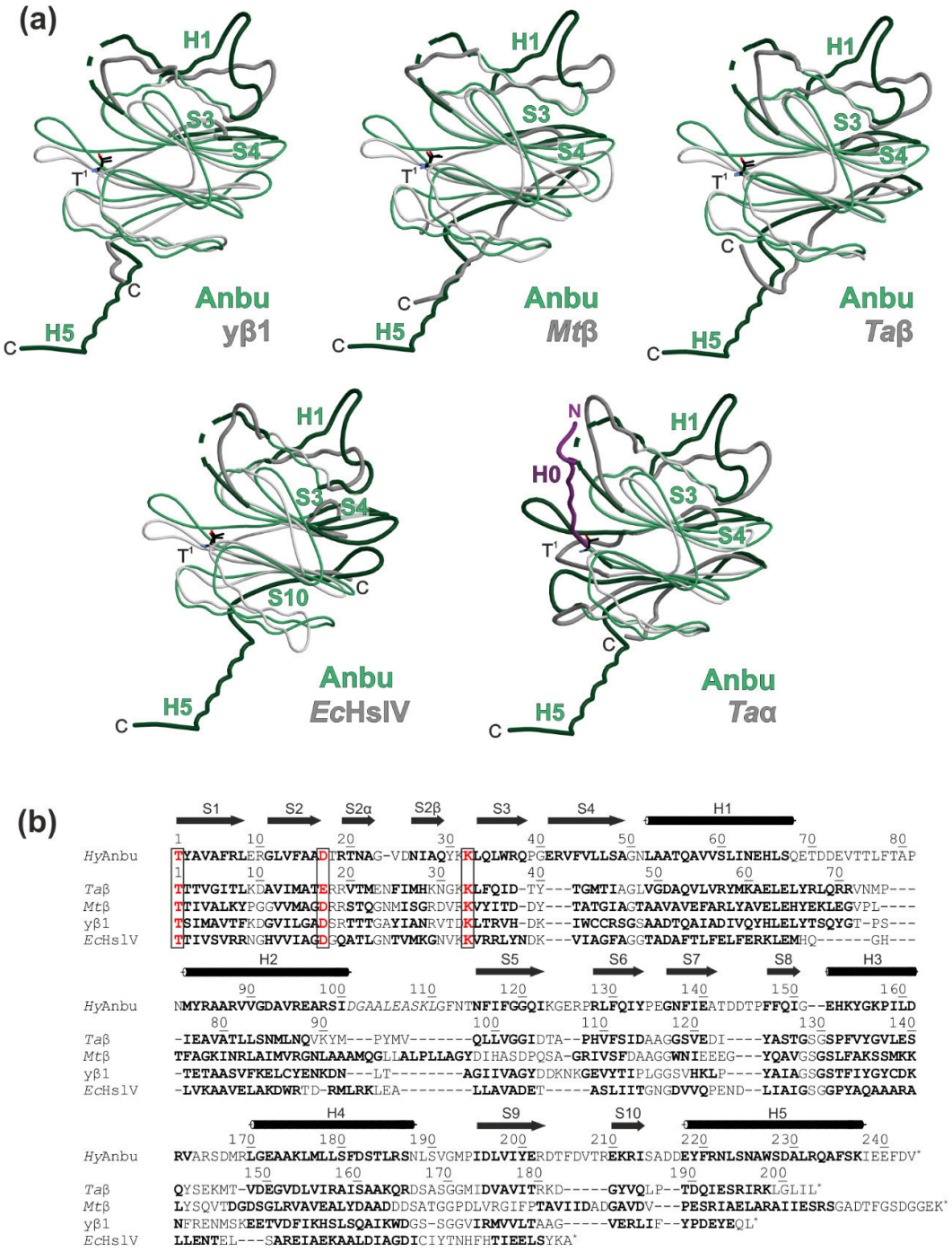


Figure 10: *Hy*Anbu shares many structural properties with the proteasomal β-subunits.

(a) Superposition of an *Hy*Anbu monomer (green, PDB ID: 5NYJ) with the β1-subunit from yeast proteasome (yβ1), the β-subunit from the CP of *M. tuberculosis* (*Mt*β, PDB ID: 2FHG), the α- and β-subunit from the CP of *T. acidophilum* (*Taα*/β, PDB ID: 1PMA), and HslV from *E. coli* (*Ec*HslV, PDB ID: 1NED) (all in grey). Structurally deviating regions are labelled and highlighted by thicker lines and darker colours. The H0 helix in *Taα* is shown in magenta, whereas Thr1 is depicted in black in all other structures. **(b)** Structure-based sequence alignment of *Hy*Anbu, proteasomal β-subunits and HslV analogous to the structural overlays shown in **(a)**. The numbering for the β-subunits and HslV is oriented on that defined for *Ta*β. The Thr1O$^{\gamma}$ activating triad is highlighted in red and framed. Secondary structure elements are emphasised in bold in the sequences and depicted as symbols above the alignment (for copyright information see section 10).

Above observations were supported by the results of a Dali search ([19], see also section 5.1.12), which identified proteasomal β-subunits as the ones most closely related to Anbu. According to its eubacterial origin, it seemed to be likely that Anbu would resemble the prokaryotic β-subunits best, but in fact the β1-subunit of yeast CP (yβ1) with a Z-score of about 22 has the highest structural similarity. The structural overlays of Anbu with yβ1 as well as with the β-subunits of *M. tuberculosis* (*Mt*β) and *T. acidophilum* (*Ta*β) show that the positions of strand S10 and helix H5 deviate the most (Figure 10a).

*Hy*Anbu's last helix H5 starts at residue 218 and includes approximately the next 20 amino acids, thereby reaching almost to the C-terminus of the protein (Figure 9). It is rotated by ~90° inwards relative to the plane of the αββα core motif and protrudes away from there at an angle of ~45°. Based on its length and orientation, helix H5 is the distinguishing feature of Anbu. *Mt*β possesses a similarly long C-terminal helix H5. The preceding strand S10, however, is much shorter than in Anbu, so, the position of the helix is clearly shifted, much closer to the structural core of the subunit. A comparable situation is found in *Ta*β, whereas in yβ1 the position of strand S10 is almost the same as in Anbu. Further structural differences with respect to all three proteasomal β-subunits include strands S3 and S4 – both are extended by a few amino acids in *Hy*Anbu – and the position of helix H1, which is tilted by about 15°, also affecting the position of the following loop (Figure 10a).

Structural variations, however, exist not only between *Hy*Anbu and its proteasomal relatives but also between the Anbu monomers of the different crystal structures per se. Both helix H5 and the S2α/β-loop seem to be comparatively flexible in their relative positions. In general, the loop regions of Anbu show a higher internal mobility than the rest of the structure, which is else well-defined in the electron density map. The best example is the loop connecting helix H2 and strand S5 that turned out to be flexible in any of the structures. However, one also needs to differentiate once more between the various space groups. In C2 (PDB ID: 5NYJ) and $P2_12_12_1$ (PDB ID: 5NYG) the other loop region with a higher degree of structural disorder lies between helix H1 and H2 (in $P2_12_12_1$ it could only be traced in chain A), whereas it is well defined in R32 (PDB ID: 5NYP/Q) and R3 (PDB ID: 5NYR). Here, the S2α/β-loop as well as the loop connecting helix H4 and strand S9 lack defined electron density.

6.2.4.2 The protomer of Anbu is a dimer

The topology of the subunit classified Anbu as a proteasomal relative and structural superposition demonstrated it to resemble eukaryotic β-subunits the most. Thus, there is still the open question, whether Anbu might form a similar proteolytic supercomplex as known for the CP and HslV [45, 79].

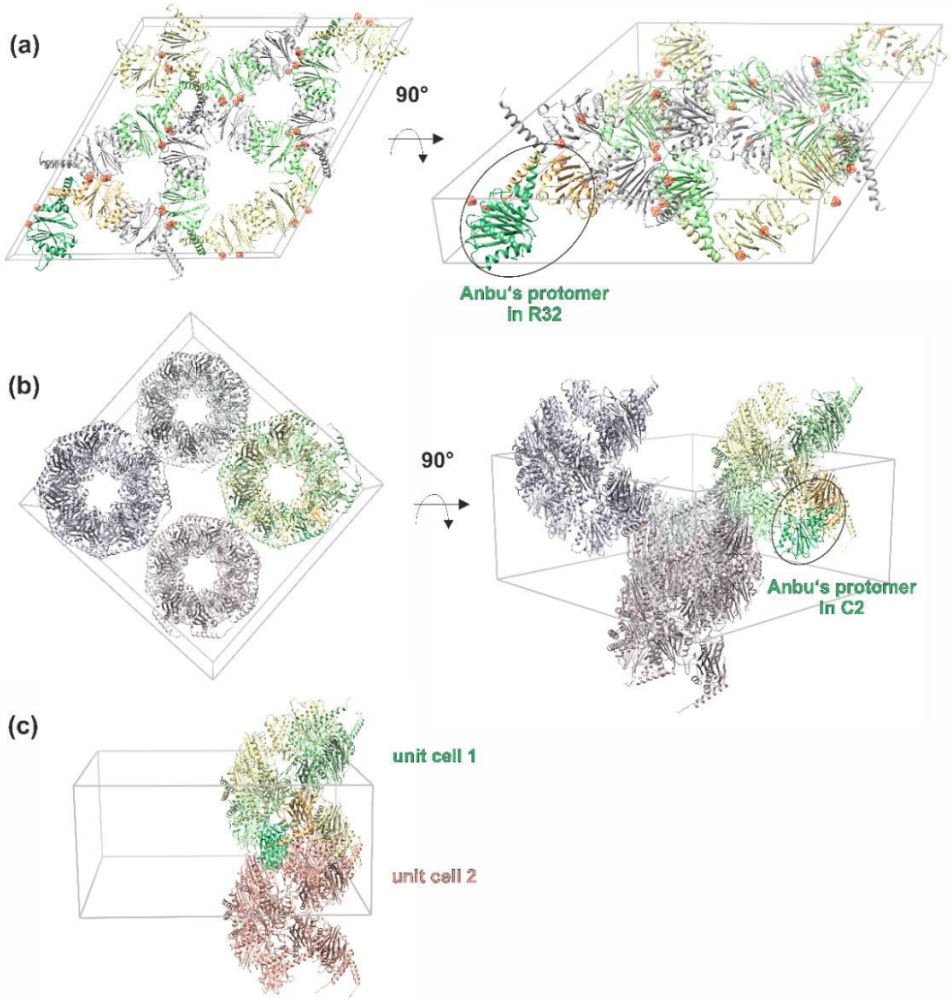


Figure 11: Oligomeric packing arrangements of *Hy*Anbu molecules in crystal lattices with space groups R32 and C2.

(a) Content of the unit cell of space group R32 in top and front view. Phosphate ions (oranges) were found to mediate important crystallographic contacts between the distinct protomers of *Hy*Anbu (encircled, subunit 1: yellow, subunit 2: green), probably impairing any further oligomerization (PDB: 5NYP). **(b)** Content of the unit cell of space group C2 in top and front view. In this space group the protomer of *Hy*Anbu (encircled, subunit 1: yellow, subunit 2: green) assembles into a left-handed helix with eight protomers per turn. One AU is depicted in colour, whereas the others are presented in different grey scales. **(c)** The stacking of unit cells in space group C2 leads to formation of a continuous helical superstructure traversing the whole crystal (the content of one AU per unit cell is shown, for copyright information see section 10).

While SEC indicated that *Hy*Anbu assembles into defined entities in solution with a molecular weight 10- to 12-fold higher than that of a single monomer (Figure 6), none of the crystal structures, however, provided an unambiguous result in that respect. The

oligomeric assembly of *Hy*Anbu molecules in space groups R32 and R3 was not even remotely reminiscent of ring-shaped particles, whereas in the other two crystal structures the subunits of Anbu formed a helical filament with resemblances to the CP or HslV as shown in Figure 11 and discussed in the next section in detail.

Nevertheless, calculation of probable quaternary structures with the PDBePISA server (see section 5.1.12) revealed that the two distinct spatial arrangements of *Hy*Anbu molecules have one thing in common: two subunits each form a tightly interacting dimer (Figure 11a/b). These dimers are organized in such way that the H5 helices of both monomers are aligned antiparallel to each other, exhibiting an internal, two-fold rotational symmetry (Figure 12a, left). The interaction is stabilized by the mutual binding of the two helices and their binding to the core of the opposite Anbu molecule. Helices H3 and H4 as well as the inward-looking parts of strands S9 and S10 are involved. Aligning several Anbu sequences illustrates that the tryptophan residue Trp227 in helix H5 is strictly conserved. In addition, there are three additional residues in this helix, for which the hydrophobic character of the amino acid seems to be obligatory. The comparison with the structural data of *Hy*Anbu discloses that the residues Leu223, Trp227, Leu231 and Ile238 form a hydrophobic network that interconnects the two H5 helices (Figure 12a, middle). The involvement of Ile159 and Leu179 in helix H3 and helix H4, respectively, extends the dimeric interaction to also include parts of the αββα core motif (Figure 12a, right). Presumably, this arrangement orchestrates the positioning of helix H5 in respect to the central αββα-motif.

The size of the interaction surface of the dimer varies between the different space groups, from ~1400 $Å^2$ in R3 to ~2100 $Å^2$ in C2 (calculated with PISA (see section 5.1.12)). Both the flexibility in helix H5 as well as the structural disorder in space groups R3 and R32 contribute to these variations. However, despite these individual differences, the dimeric contact is still the most pronounced peculiarity in all crystal structures of *Hy*Anbu, defining the dimer as the protomer of Anbu.

A comparison with the β-rings in the CP and HslV lets us conclude that there exists a pendant to the protomer formation in Anbu: the β-trans-β' contacts (Figure 12b). In the crystal structure of the CP from *T. acidophilum* they are designated as non-crystallographic dyads [48]. Notably, the amino acids in these parts of the β-subunits are conserved in HslV as well [112]. Despite the variety of β-subunits in the eukaryotic CP, with the β1β'1- and β4β'4-interactions the molecular dyads remain [53]. However, there exist some differences

in the interaction seen here compared to that described for Anbu. The subunits involved in the β-trans-β'contacts are closer together than in Anbu, where a tilt of the subunits in the protomers by about 30° prevents a similar intermolecular packing. Due to this arrangement, the H5 helices of the proteasomal β-subunits are rotated outward and solvent exposed. Instead, the contact between the H3 helices from the opposing subunits is intensified with the H4 helices flanking on both sides (Figure 12b) [126, 127]. Nonetheless, due to the missing involvement of helix H5 the surface area covered by the dyad contacts in the CP or HslV is with ~600-800 Å^2 significantly smaller than that in the protomers of Anbu.

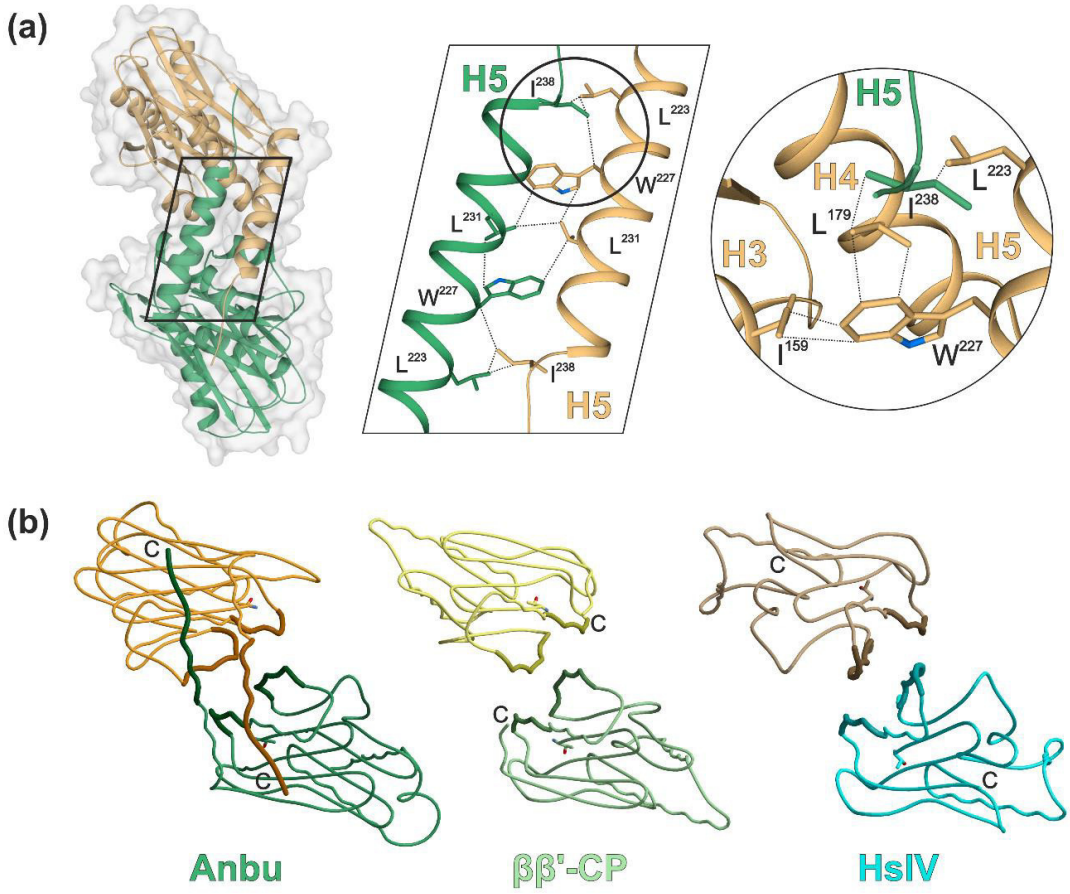


Figure 12: The protomer of Anbu is a dimer with resemblances to the molecular dyads in CPs and HslVs.

(a) Two subunits of *Hy*Anbu tightly interact with each other via their H5 helices (left). A network of hydrophilic residues around the highly conserved Trp227 mediates the mutual binding of the two helices (middle) including the helices H3 and H4 of the opposing subunit, respectively (right). Since all structures of *Hy*Anbu possess this dimer formation, we defined it as the protomer of Anbu (PDB ID: 5NYJ, chain K: green, chain M: yellow, surface representation: grey, nitrogen atoms: blue). **(b)** Despite the shift in the orientation of the subunits towards each other, the promoter contact is equivalent to the β-trans-β' contact in the CP and HslV. However, this contact is only transmitted by residues located in helices H3 and H4, weakening the strength of the network (PDB IDs: 5NYJ (Anbu), 1PMA (CP, [48]), 1KYI (HslV, [120]), subunit 1: yellow/orange, subunit 2: green/blue, interacting regions as well as N- and C-termini of each subunit are highlighted, Thr1 is shown as sticks, for copyright information see section 10).

6.2.4.3 The helical superstructure of *Hy*Anbu

In space groups C2 and $P2_12_12_1$, the molecules of *Hy*Anbu assemble in an oligomeric structure that goes beyond the dyad. Eight protomers together form the repeating unit of a

left-handed helix traversing the whole crystal. One protomer is related to the next one by a rotation of 45° and a translation of ~13 Å along the screw axis.

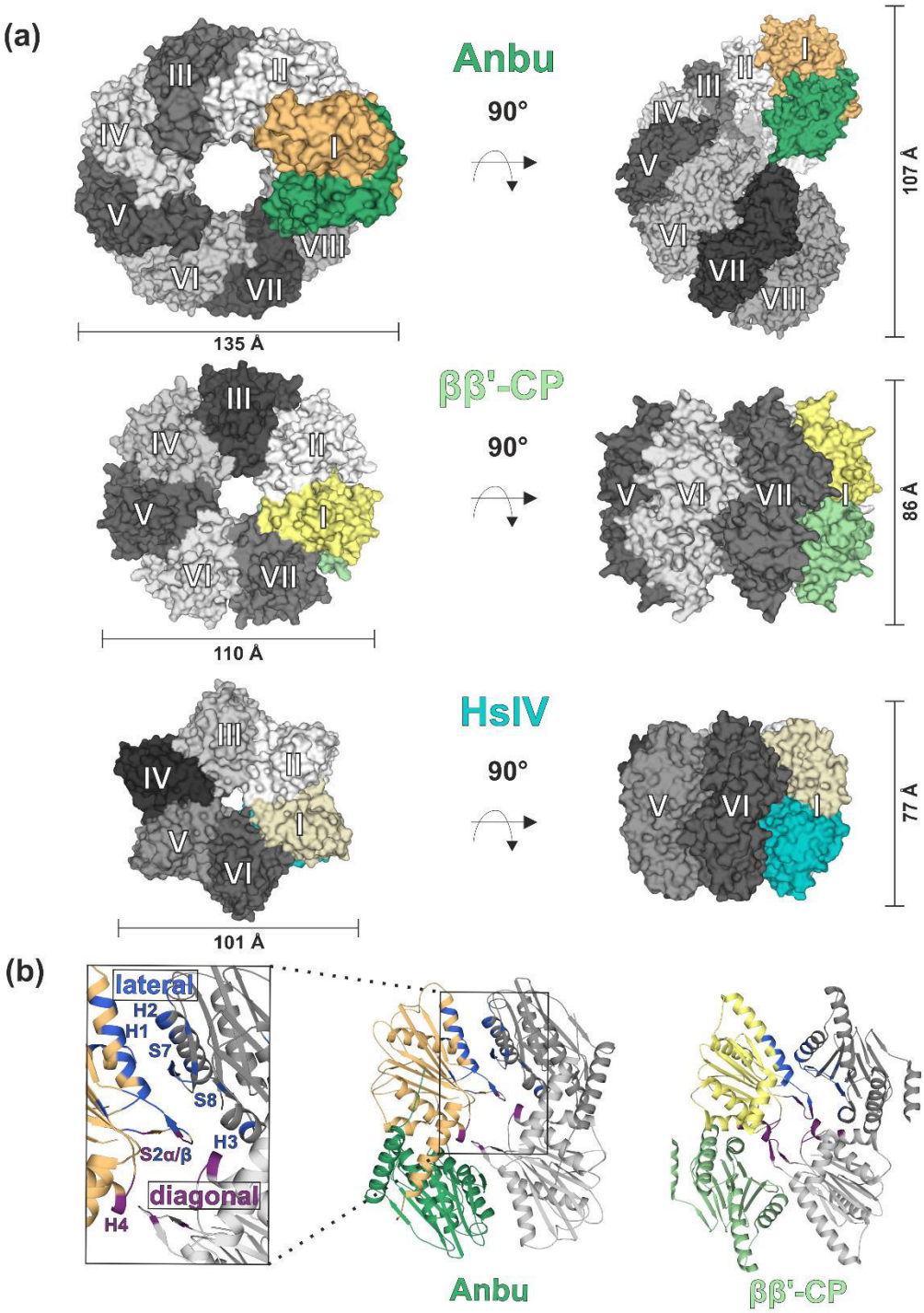


Figure 13: Protomers of *Hy*Anbu form a left-handed, helical superstructure in the crystal.

(a) Top and front view of the helical assembly seen in crystals of *Hy*Anbu. It differs both by its symmetry (8-fold vs. 7/6-fold) and shape (infinite helical superstructure vs. closed ring structure) from the β-rings of the CP and HslV (PDB IDs: 5NYJ (Anbu), 1PMA (ββ'-CP), 1NED (HslV), one protomer/molecular dyad in each structure is shown in colour, the others in various grey scales). **(b)** Anbu's 'lateral' (blue) and 'diagonal' (magenta) contacts between adjacent protomers are equivalent to cis- and trans-interactions in the β-rings of the CP and HslV. The secondary structure elements involved in the binding are conserved throughout all three homologues (for copyright information see section 10).

According to the symmetry within each protomer and the helix as a whole, only two types of interactions between neighbouring protomers exist. The sideward binding of adjacent protomers was designated as the 'lateral' contact. Parts of strands S7 and S8 as well as helices H2 and H3 of one subunit interact with helix H1 and the S2α/β-loop of the next one. The small β-turn between strands S2 and S3 is also involved in a second contact, which will be referred to as 'diagonal', because the upper subunit of one protomer binds to the lower subunit of the next one, located to its right. Here, the S2α/β-loop interacts with the end of helix H4 and the preceding loop that connects the helix with strand S9. Calculations of the size of the corresponding contact surfaces show that with about ~500 to 700 $Å^2$ they cover a much smaller area compared to what is seen in the protomer contact. Interestingly, both types of interactions are also found in the β-rings of the CP and HslV. The lateral contact is equivalent to the intramolecular binding of subunits within one ring, whereas the diagonal interaction corresponds to another β-trans-β'-contact. Remarkably, here the same secondary structure elements are involved in the subunit's binding in the distinct proteins, underlining the importance of these contacts for the complex formation in the proteasome family in total.

6.2.4.4 The putative active site of Anbu

Similarities in the sequences of Anbu, the CP and HslV are mainly found in the first 60 residues after the N-terminal nucleophile, including most of the residues involved in active site formation in the last two. Apart from Thr1 also Asp17 seems to be strictly conserved. Instead of Lys33, *Hy*Anbu encodes a lysine residue at position 32, which will likely acquire the same function within the Th1O^{γ} activating triad as its counterparts in the CP or HslV. However, the missing catalytic activity in the peptidase and protease assays suggested that, despite the conservation of these important residues, there exist other structural rearrangements in Anbu impeding it.

In structures of *Hy*Anbu that were crystallised in space groups R32 and R3, the putative catalytic centre seems to be mostly disordered (PDB IDs: 5NYR/P/Q). Thr1 is facing away from both Lys32 and Asp17. Moreover, the S2α/β-loop that precedes the lysine residue, is not defined in the electron density map at all. This also affects the side chain of Lys32, which shows an increased flexibility and thus, is probably incapable of forming stable hydrogen bonds with the other two residues in this catalytic triad.

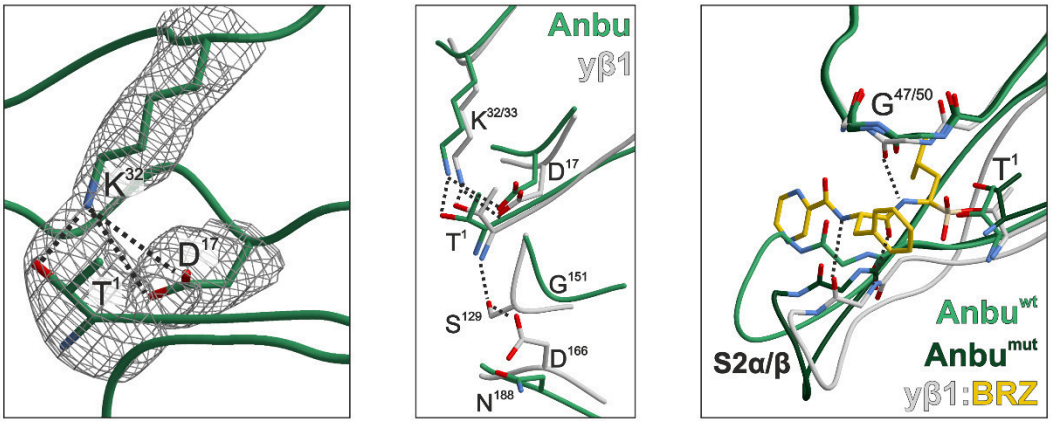


Figure 14: The putative catalytic centre of *Hy*Anbu.

In the C2 structure of *Hy*Anbu, the Thr1O$^{\gamma}$ activating triad of Thr1, Asp17 and Lys32 is able to align properly (left). Structural superposition with the β1-subunit from yeast demonstrates that Anbu lacks residues equivalent to the Thr1-NH_2 activating amino acids Ser129 and Asp166. We favour this peculiarity as a potential explanation for its missing catalytic activity (middle) and the finding that, despite co-crystallization, the high-affinity proteasome-inhibitor bortezomib (BRZ) did not react with *Hy*Anbu^{wt}. The latter might also be a result of the orientation of the S2α/β-loop in the C2 crystal structure. An overlay with BRZ-bound yβ1 illustrates that its actual position would clash severely with the inhibitor (right). However, comparison with Anbu^{mut} reveals that the S2α/β-loop is flexible and, thus, probably not the main reason for the inactivity of *Hy*Anbu (PDB IDs: 5NYJ (Anbu^{wt}, chain E/K, light green), 5NYG (Anbu^{mut}, chain A, dark green, for reasons of clarity Gly(-1) is left out here), 5CZ4 (yβ1), BRZ: yellow, oxygen and nitrogen atoms: red and blue, respectively, $2F_o$-F_c electron density map of the active site residues, contoured at 1σ: grey, for copyright information see section 10).

The situation is quite different in the structures, in which *Hy*Anbu forms the helical arrangements. In the $P2_12_12_1$ structure of Anbu^{mut} all three residues are clearly resolved and linked to each other via hydrogen bonds, whereby the position of Thr1 seems to be slightly shifted compared to what is seen in the proteasomal subunits (PDB ID: 5NYG). It is possible that the additionally inserted Gly(-1) residue, likewise defined in the density and reaching into the putative substrate channel, causes this movement as no such observation was made for Anbu^{wt}. In two out of the 16 subunits of the C2 structure, the catalytic triad of Thr1O$^{\gamma}$-Lys32N$^{\varepsilon}$-Asp17O$^{\delta1}$ is properly aligned and resembles that of the proteasomal β-subunits (PDB ID: 5NYJ, chain E and K). Thus, Anbu^{wt} should be capable of substrate turnover and also inhibitor binding at least in this configuration (Figure 14). Interestingly, the C2 structure was determined from a crystal that grew in the presence of 1 mM of the highly reactive, non-specific proteasome inhibitor bortezomib (BRZ). The electron density map, however, proved that BRZ did not bind to *Hy*Anbu in the co-crystallization trials. A reasonable explanation could be that in the structure of Anbu, there are no residues equivalent to Ser129 and Asp166. Hence, the Thr1 N-terminus is not involved in any hydrogen bonding and might lack its role as a Brønsted acid (Figure 14).

6.3 Discussion

The motivation for our studies on Anbu was to analyse this mostly uncharacterised, presumed representative of an N-terminal threonine proteases in detail and to retrace the evolutionary roots of the proteasome family in bacteria. Anbu's occurrence in cyanobacteria that lack both HslV and a CP suggested it to be the progenitor of all proteasomes and proteasome-like proteins or at least closely related to it. Undoubtedly, there exists a phylogenetic relationship, but whether Anbu is actually the 'ancestral β-subunit protein' still needs further clarification.

6.3.1 Anbu proteins in various species – different or similar?

In parallel to our experiments, Fuchs and co-workers from the Martin group in Tübingen published a study on Anbu from *Pseudomonas aeruginosa* in June 2017 (*Pa*Anbu, PDB ID: 5LOX, [128]), whereas Anbu from *Yersinia bercovieri* (*Yb*Anbu, PDB ID: 5NYW) was characterised structurally by Piasecka *et al.* from the Bochtler lab in Warsaw and us (*manuscript under revision*). Considering the differences between *Hy*Anbu monomers, the structural deviations to *Pa*Anbu and *Yb*Anbu are minor (r.m.s.d. (C_α) ≤ 0.8). In *Hy*Anbu, there is a slight shift in the position of helix H1 and the following loop, which is extended by three residues compared to the other two species. Moreover, the loop between helix H2 and strand S5 could only be traced in the electron density map of the structure from *P. aeruginosa* [128]. Even there, it was found to be highly flexible with the corresponding residues having the highest B-factors of the entire structure. Interestingly, this loop plays an important role for both the CP and HslV in constricting the pore size of the different rings by stretching out into their centre of masses, and is therefore well-resolved in these crystal structures [48, 53, 112]. The divergence in Anbu might suggest a specific structural and/or functional property.

Like in *Hy*Anbu, the C-terminal residues of *Pa*- and *Yb*Anbu form an extended helix H5, confirming it to be the distinguishing attribute of Anbu molecules (Figure 9). The role it plays in the dimer formation of Anbu subunits emphasises its functional significance for the protein. The contact between two monomers is transmitted by binding of the H5 helices towards each other as well as to helices H3 and H4 of the opposing subunit via a strictly conserved network of hydrophobic contacts (Figure 12a). This holds without exception for all structures of Anbu and lets us define it as the protomer contact of Anbu.

Heterologous expression and purification of *Hy*-, *Pa*- or *Yb*Anbu likewise resulted in the finding that Anbu assembles into a multimer about 10 to 14-times larger than the respective monomer. Hence, it is likely there exists a common arrangement in the solution state for all kinds of Anbu proteins. No such agreement was observed regarding their oligomerization behaviour during crystallization. Both *Pa*- and *Hy*Anbu formed helical superstructures, but with different rises and symmetries (PDB IDs: 5LOX [128] and 5NYJ, Figure 11b/c). In addition, *Hy*Anbu also crystallised in space groups R3 and R32, where no oligomerization beyond the protomer formation could be detected (PDB IDs: 5NYP/R, Figure 11a). In contrast, *Yb*Anbu adopts a finite oligomeric complex with a lock-washer-like shape in the crystal lattice (PDB ID: 5NYW). Consisting of 14 subunits in total, this symmetric and open ring-like structure with a small helical rise seems to reflect Anbu's physiological multi-subunit state best so far, which is supported by SAXS measurements performed both by Piasecka and colleagues (unpublished results) as well as Fuchs and co-workers [128]. Fitting the experimental data with a model of six adjacent protomers of *Pa*Anbu resulted in a structure similar to that of *Yb*Anbu seen in the crystalline state – only with twelve instead of fourteen subunits included. Therefore, we propose that the oligomeric arrangement of Anbu differs from that of its homologues.

Despite several conditions being tested, none of the three Anbu variants displayed any kind of catalytic activity against peptide or protein substrates. Interestingly, Fuchs *et al.* reported binding of *Pa*Anbu to both a peptide aldehyde and an Ntn-specific peptidic epoxyketone via mass spectroscopic analysis. However, they could not confirm their observations by using Anbu protein from *Cytophaga hutchinsonii* [128]. These contradicting results could, on the one hand, indicate some so far uncharacterised, species-dependent differences. On the other hand, the varying binding mechanisms of all three peptide ligands might be of importance. We identified the missing activation of the Thr1-NH_2 terminus in *Hy*Anbu to be a suitable explanation (6.2.4.4). In principle, binding of an peptide aldehyde or epoxyketone should not require such activation, allowing an interaction with *Pa*Anbu regardless of the state of its N-terminus [129]. In any case, further experimental evidence is required to rule out one or the other possibility.

6.3.2 The relation between Anbu and its proteasomal homologues

The investigations on *Hy*Anbu elucidated a close relationship to the subunits of the CP and HslV. Unexpected similarities, but even so differences between them were discovered regarding both structural and functional properties.

Indisputable is the finding of Anbu being a member of the Ntn-hydrolase family as it adopts the typical αββα fold. The conservation of threonine as the N-terminal nucleophile indicates that Anbu might be catalytically active, thus, having more in common with the subunits of HslV and the β-subunits than with the α-subunits of the CP. A Dali-search identified the proteasomal β-subunits to be the ones most closely related to *Hy*Anbu. From a structural point of view, this is confirmed as subunits of HslV are missing both strand S10 and helix H5 (see section 6.2.4.1 and Figure 10). Nevertheless, our experiments demonstrated that in some respects *Hy*Anbu is similar to the β-subunits of the CP(s), whereas in others it adopts characteristics of HslV(s).

Both HslV and the β-subunits of the CP assemble into rings, which then associate with each other to form the proteolytic particle and/or a catalytic chamber. For this, the proteasomal β-subunits require assistance of the α-subunits as they remain monomeric in solution if expressed alone [77, 82], whereas HslV forms ring-shaped particles on its own [112, 119]. The resulting complexes can then be crystallized and characterised structurally. Heterologous expression of *Hy*Anbu in *E. coli* led to the assembly of defined multi-subunit particles, significantly larger than a monomer, which are stable at low protein concentrations or after long-time storage at -20 °C. For Anbu no interaction partner with scaffolding function, equivalent to the proteasomal α-subunits, was identified, strongly indicating that, like HslV, it has an intrinsic ability to assemble into the correct oligomeric complex by itself. The homology with the CP and HslV suggested that such a complex is also composed of various rings stacked on top of each other.

Bearing this in mind, the results of the structural elucidation of *Hy*Anbu were rather unexpected. Assembling either in a crystal packing, where no oligomeric pattern beyond a dimeric interaction is recognisable or one, where helical superstructures traverse the whole crystal, apparently contradicts the data measured in solution. A potential explanation for the observation of helical filaments of *Hy*Anbu might be crystallographic forces, which pull the presumed ring-shaped structure of Anbu slightly apart. Thereby, it switches from a

physiologically well-defined oligomeric state in solution to one facilitating formation of the infinite fibre observed in the crystal. Similar observations were reported before for the ATPases ClpA [130] and ClpX [131]. As stated above, the obvious discrepancy between Anbu's behaviour in solution and in the crystals, is not only a phenomenon observed for *Hyphomicrobium sp.* but also described for Anbu from *P. aeruginosa* [128]. Experiments with Anbu from *Yersinia bercoveri* suggest a lock-washer-shaped structure of the multimer rather than closed rings. Such a slightly open complex might easily tolerate small shifts of its subunits, switching to the state of the helical superstructures seen in the crystals.

Nonetheless, the precise analyses of the binding of *Hy*Anbu subunits in the two different crystal packings reported here showed them to resemble the three contacts between subunits in the β-rings of the CP and in HslV. The dimeric interaction of two *Hy*Anbu monomers, defined as its protomer, has an equivalent in one of the β-trans-β'-contacts of its proteasomal homologues (Figure 12b). However, due to a reorientation of the respective subunits against each other, the secondary structure elements involved are altered. In addition, such a strong participation of helix H5 as found for the interactions in *Hy*Anbu would not be possible in the CP or HslV, due to the length of this helix in the β-subunits of the former or its complete lack in the subunits of the latter [53, 112].

No such dramatic changes were observed, while comparing the other two contacts in CPs, HslVs and Anbu protein complexes. The lateral interaction in the helical superstructure of *Hy*Anbu corresponds to the inter-ring contacts of its homologues (Figure 13b). Although the orientation of neighbouring subunits in the β-rings is not the same as that of adjacent subunits in the helical superstructure, the differences regarding the size of the contact area and the parts of the subunits involved are considerably smaller than observed for the protomer contact. The best-preserved contact is the diagonal one, which is equivalent to another β-trans-β'-contact. It connects a subunit of the upper ring with the one located to its right on the lower ring (Figure 13b). In all three cases, the interacting secondary structural elements are the S2α/β-loop of one subunit that binds to the C-terminal end of helix H4 and the following loop connecting it to strand S9 (referred to as H4-S9-loop afterwards) of the other.

Notably, the loop regions described above were found to be disordered in *Hy*Anbu subunits in space groups R32 and R3, where no other oligomeric assembly than the protomer formation was identified (Figure 11, PDB: 5NYP/R). This suggests the contact with the

other subunit being indispensable for these loops to get structured. Similar observations were made for the β-subunit from *M. tuberculosis* crystallised in a monomeric state (PDB ID: 2JAY). Here again, the S2α/β- and the H4-S9-loop are not defined in the electron density map indicating a common pattern for *Hy*Anbu and the β-rings of the CP (Figure 15). There exist no data for HslV regarding this point, but as the subunit-subunit interactions were found to be the same as in the proteasomal β-rings, the situation is likely similar.

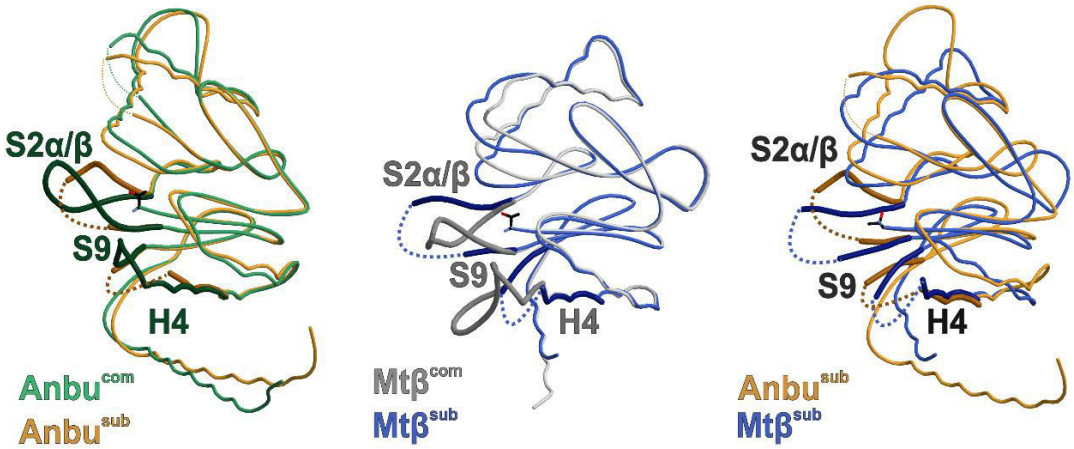


Figure 15: The S2α/β- and the H4-S9-loop are disordered in isolated β-subunits of both Anbu and the CP.

Anbu subunits crystallized in space groups R32 and R3 display the same degree of disorder as an β-subunit of the mycobacterial CP crystallised by its own. Comparison with subunits of Anbu and the CP found in a higher oligomeric complex showed that the unstructured loop regions require the presence of the adjacent subunits to become defined (PDB IDs: 5NYJ (Anbu^{com}, green), 5NYR (Anbu^{sub}, orange), 2FHG (*Mt*β^{com}, grey, [132]), 2JAY (*Mt*β^{sub}, blue), disordered residues: dotted line). The corresponding interaction was characterised as the diagonal contact in the helical superstructure of *Hy*Anbu. In the CP, it was suggested to control active site formation (for copyright information see section 10).

The strong conservation of the diagonal contact in Anbu, the CP and HslV indicates that it must be of either structural and/or functional importance for all three homologues. The residues located at the beginning and end of the S2α/β-loop are Asp17 and Lys32/33. Thus, it connects the two residues essential for the activation of Thr1O$^{\gamma}$. The crystal structures of *Hy*Anbu in space groups R32 and R3 and of isolated *Mt*β depict that a disorder in the loop is directly correlated with a disorder in one of the (putative) catalytic triads (see section 6.2.4.4). That this is a two-sided effect was revealed by the CP of *A. fulgidus*, where the T1G mutation resulted in crystallization of a proteasome-like structure with the β-rings further apart from each other than usual and the S2α/β-loop not defined in the electron density map [62]. Investigations on the *Rhodococcus* proteasome confirmed these findings revealing that mutations in the interaction of the S2α/β-loop with the N-terminal part of the H4-S9-loop do not interfere with assembly of the CP in general, but they lead to a dislocation of the S2α/β-loop, a disorder in the active site and, thus, to a lack of proteolytic

activity [127]. All these findings demonstrate that the role of the diagonal contact/S2α/β-loop-H4-S9-loop interaction has to be a regulatory one, a so-called 'activation switch' [127] and a safety mechanism as only the presence of the interacting subunit is able to align the S2α/β-loop and, thereby, the active site correctly.

Taken together, the three subunit contacts seen in the helical superstructure of *Hy*Anbu have pendants in the β-rings of the CP. From a functional point of view, the one with the most important role is the diagonal one as it has direct influence on the alignment of the active site, strongly correlating with its high degree of conservation. Considering the other two interactions, their roles seem to be switched in Anbu and the CP. The first contact formed between subunits of *Hy*Anbu is presumably the protomer contact judged by the size of covered surface area between the two monomers and the resulting strength of the interaction as well as the occurrence of this interaction in all structures solved. Sideward assembly of the protomers via the lateral contacts leads to formation of the helical superstructures in the crystals, and quite likely also to the yet uncharacterised solution state of *Hy*Anbu. In contrary, the inter-ring contacts in the CP only develop by mutual binding of two half-proteasomes to form the fully matured, proteolytic particle.

6.3.3 The missing catalytic function

Even though by the fold and structure of its subunits as well as their interaction patterns, Anbu is clearly identified as a member of the Ntn-hydrolase family, its missing enzymatic activity is puzzling, all the more since the three studies on different Anbu proteins agree in that respect.

Based on the crystal structures of *Hy*Anbu, two different potential reasons were identified: first, the position of the 'activation switch' described above and second, the missing counterparts of the proteasomal residues Ser129 and Asp166 in Anbu. In the C2 structure the S2α/β-loop occupies a position that would have interfered with the binding of bortezomib to *Hy*Anbu during co-crystallization (Figure 14, right). However, assuming that the helical superstructure of *Hy*Anbu resembles Anbu's true solution state only up to a certain degree, the orientation of the S2α/β-loop in the crystal must not necessarily reflect that since it occupies a different position in the $P2_12_12_1$ structure of $Anbu^{mut}$. Therefore, the S2α/β-loop is unlikely to be the primary cause for the lack of activity. A given fact is that the Thr1-NH_2 terminus of *Hy*Anbu is not activated via a network of hydrogen bonds.

However, a Brønsted acid is essential to complete a full catalytic reaction cycle. Hence, the CP depends on the presence of two different, catalytic triads [129], but only one of them seems to be present in *Hy*Anbu. Several conclusions may be drawn from our findings. One is that Anbu might be lacking its natural interaction/binding partner(s). For example, the catalytic turnover of HslV alone is negligible, but after complex formation with the AAA^+-ATPase HslU its degradation rate increases exponentially [117]. The allosteric regulation involves large structural rearrangements of the substrate binding channel and cooperation between HslV and HslU is reflected by forming an operon together, so that their genes are expressed simultaneously [109]. Nevertheless, the activation mechanism of *Hy*Anbu must be different than that of HslV since the structural overlay of *Hy*Anbu and HslV alone or in the HslUV complex demonstrates that these residues are already well-aligned in Anbu to form an antiparallel β-strand with a potential peptidic substrate (Figure 16).

The idea that Anbu might be activated or regulated in a different way than its proteasomal homologues was supported by a study in *Pseudomonas putida*, which showed that upon nitrogen starvation the expression of its operon was among the most upregulated ones [133]. To account for the unexpected genetic environment of Anbu and based on biocomputational studies, Iyer and co-workers proposed it to be involved in a peptide synthesis and/or tagging system, instead of being a protease in the classical sense [134]. Being part of such a signalling network, Anbu might only recognize a specific substrate spectrum explaining the absence of catalytic turnover of standard protease substrates.

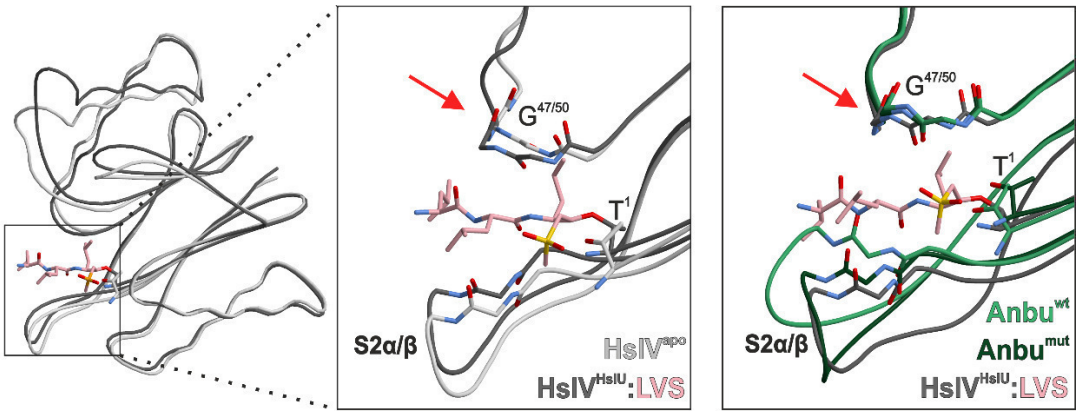


Figure 16: The activity of *Hy*Anbu seems to be regulated differently than observed for the HslUV complex.

The comparison of the structures of HslV (apo) with HslUV complexed to a peptide vinyl-sulfone inhibitor (LVS) illustrates that inhibitor binding requires structural rearrangements in the upper part of the substrate binding channel (red arrow) as well as in the S2α/β-loop. In *Hy*Anbu, the region around Gly47/50 is already structured for binding of peptidic ligands. Notably, the position of the S2α/β-loop varies considerably between different Anbu structures, which might be caused by the missing binding of Anbu's natural substrate (PDB IDs: 1JJW ($HslV^{apo}$, [119], light grey), 1KYI ($HslV^{HslU}$:LVS, [120], dark grey, inhibitor shown in rose), 5NYJ ($Anbu^{wt}$, light green), 5NYP ($Anbu^{mut}$, dark green, Gly(-1) is left out for clarity), oxygen, nitrogen and sulphur atoms shown in red, blue and yellow, respectively, for copyright information see section 10).

6.3.4 Anbu - the ancestral β-subunit protein?

The discovery of Anbu and its occurrence in cyanobacteria suggested it to be the ancestor of the family of N-terminal threonine proteases [14]. However, there are not only similarities with the proteasomal β-subunits and HslV but also some individual features of Anbu, affecting *e.g.* the subunit-subunit interactions as well as the lack of proteolytic activity. While our data could show that, from a structural point of view, Anbu is most closely related to the β-subunits of the CP, no further conclusions regarding the exact evolutionary relationship or chronological order of the origin of Anbu, the CP and HslV can be drawn from the present data. Nonetheless, the experimental proof of Anbu being another proteasomal homologue in eubacteria apart from HslV and its occurrence in cyanobacteria provide evidence for a far-back reaching root of this threonine Ntn-hydrolase family.

Alongside Anbu, another potential member of the proteasome family was discovered, called beta-proteobacteria proteasome homologue (BPH). In contrast to Anbu, the occurrence of BPH seems to be limited to a few β-proteobacteria and cluster analysis predicted it to be closer related to HslV than to the subunits of the CP or Anbu [14]. With Anbu, BPH, HslV and the 20S proteasome, the presence of members of the family of proteasome-like proteases is much more common in eubacteria than anticipated beforehand. One can imagine that there might even be more members, which have not been discovered so far because they got lost multiple times during evolution like Anbu or prevailed only in a small group of bacteria like BPH. Hence, the genomic analysis of more and more bacterial species may have a significant impact on the evolutionary perspective of this protein family.

7 Structural analysis of the ClpP1/2 heterocomplex from *Listeria monocytogenes*[4]

7.1 Introduction

7.1.1 Antibiotic resistances threaten public health care

The discovery, characterization and application of the antibiotic effect of different substances, either synthetically derived or isolated from natural sources, marked the beginning of a medical revolution in the late 19th/early 20th century [135, 136]. For the first time, a reliable instrument for the treatment of bacterial infections was available. Intensive research in the 1960s to 1980s led to the discovery of several distinct classes of antibiotics and so many, previously life-threatening diseases were now easily and effectively treatable [135].

Originally, antibiotics are secondary metabolites produced to defend the natural habitat of a microorganism against others. As any producer of such substances is in imminent danger of getting affected itself by an antibiotic produced by a rival, bacteria also developed several strategies to disarm these compounds [137]. Therefore, it came to no surprise that shortly after the first substances were on the market, resistances against them were detected as well [138]. A massive (mis)use of antibiotics in medical treatment and factory farming over the last decades, the stagnation in the discovery of new substance classes and a strong underestimation of the possibilities of spreading antibiotic resistance genes between different genera of bacteria, led to the present situation: multiple resistances now frequently emerge in several pathogenic strains and the fatality rate of such infections increased dramatically [139].

[4] The results presented here were originally published in 'Structure and mechanism of the caseinolytic protease ClpP1/2 heterocomplex from *Listeria monocytogenes'* in *Angew. Chem. Int. Ed.* 2015, Vol. 54 ([15], Copyright © 2015 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany. Reproduced with permission (License Number: 4187681081208).). This paper is based on a cooperation with the group of Prof. Dr. Stephan A. Sieber (Chair of Organic Chemistry II, Technische Universität München). Dr. Maria Dahmen purified and characterised the *Lm*ClpP1/2 heterocomplex functionally (for experimental details: SI in [15]). Crystallographic studies have been carried out by Marie-Theres Vielberg and Prof. Dr. Michael Groll.

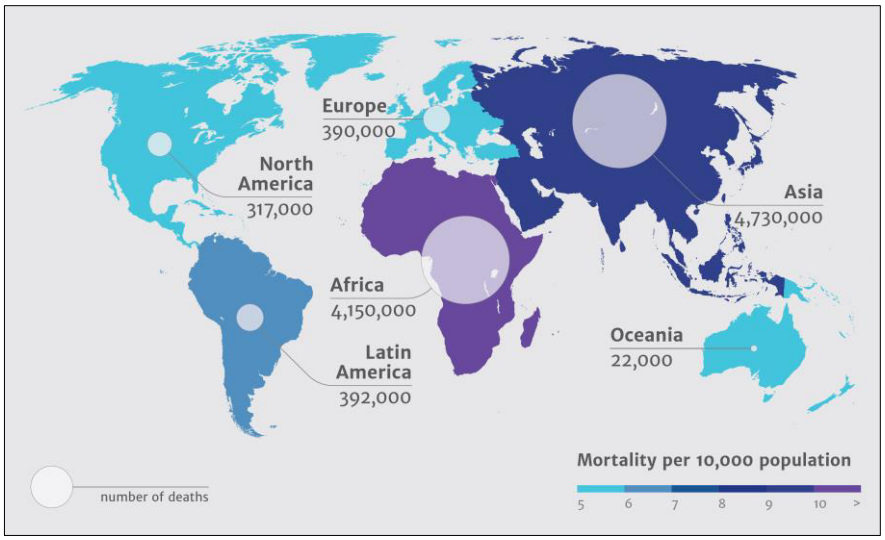


Figure 17: Estimated number of deaths per year in 2050 by antibiotic resistances.

Graph from *Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations* in 'The Review on Antimicrobial Resistance' (December 2014, chaired by Jim O'Neill, p. 13). The 'Review on Antimicrobial Resistance.' infographics have been licensed under the Creative Commons Attribution 4.0 International Public License.

In 2014, the British government launched a study to assess the dimensions of the problem (https://amr-review.org/home.html). Figure 17 shows the outcome of a possible scenario for the year 2050, in which the number of deaths per year caused by antibiotic resistances is estimated. Together with several other studies leading to a similar outcome, it illustrates the urgent need for new or altered strategies to address this threat of the achievements of modern medicine.

7.1.2 Disarming bacteria instead of killing them

The treatment with antibiotics is expected to prevent the reproduction and spread of a bacterium by causing its death. This puts a high evolutionary pressure on the targeted organism to develop a resistance and thus, escape. One strategy to interfere with that might be to solely inhibit the virulence mechanisms of a bacterium, thereby, disarming it. At the same time, its viability is not impaired. That way the pressure to mutate and evade is reduced and it should be possible to slow down the process of resistance formation. However, new antibiotic targets need to be explored for such an approach.

Infection of a host cell by a pathogen is generally associated with a dramatic change of its extracellular environment regarding *e.g.* pH, temperature and nutrient supply. This leads to enormous stress in the bacterial cell resulting, among other things, in the damage of a variety of proteins. The main task of proteases is to degrade and remove such misfolded and/or by oxidation and other mechanisms harmed proteins, thus, reducing the intra-cellular

stress. Hence, they were found to be indispensable for the virulence of various micro-organisms, which makes them, in turn, ideal candidates in the search of a new weak point within the bacterial system.

7.1.3 Caseinolytic protease P is essential for the virulence of several pathogens

The caseinolytic protease P (ClpP) was found to be a potential new antibiotic target as in many cases it regulates the virulence of the pathogen it is synthesised by. In *Staphylococcus aureus*, for example, the virulence in a Δ*clpp* mutant is strongly affected due to the significant reduction in production and secretion of the toxin alpha-haemolysin [140, 141]. Deletion of *clpp* in *Pseudomonas aeruginosa* – like *S. aureus* an opportunistic, but Gram-negative pathogen – interferes with its ability to swarm and form biofilms, both characteristics associated with increased antibiotic resistance [142]. It is often not easy to distinguish whether ClpP has solely an effect on virulence by its general function in the cellular stress regulation or whether it specifically regulates this process. So, in *Salmonella typhimurium* it was shown that ClpP has only indirect control of the expression of the pathogenicity island 1 genes, which are required for penetration of the intestinal epithelium by the pathogen [143]. It shall be emphasised that although in all analysed mutants the virulence and often also the cellular stress response was impaired, these organisms were all fully viable under normal, non-infectious growth conditions. So, the prerequisites for a new antibiotic target as defined before seem to be fulfilled.

7.1.4 The bacterial Clp-degradation machinery

The intracellular, ATP-dependent, casein-degrading activity of the caseinolytic protease P (ClpP) was described first in a Lon protease-deficient *E. coli* strain in the late 1980s [144]. Several follow-up studies elucidated many of its structural and functional characteristics, of which the most important ones are summarized in Figure 18. ClpP is a self-compartmentalized, barrel-shaped serine protease composed of 14 subunits, which are assembled in two heptameric rings stacked on top of each other [145–147]. Axial pores on the top and bottom of the particle enable access to the proteolytic chamber enclosed by the two rings [148]. The subunits of ClpP have an α/β-fold with α-helices and β-strands in alternating order. They are structurally divided into three different parts: the globular ‘head’ domain, a ‘handle’ and the N-terminal region (Figure 19) [149].

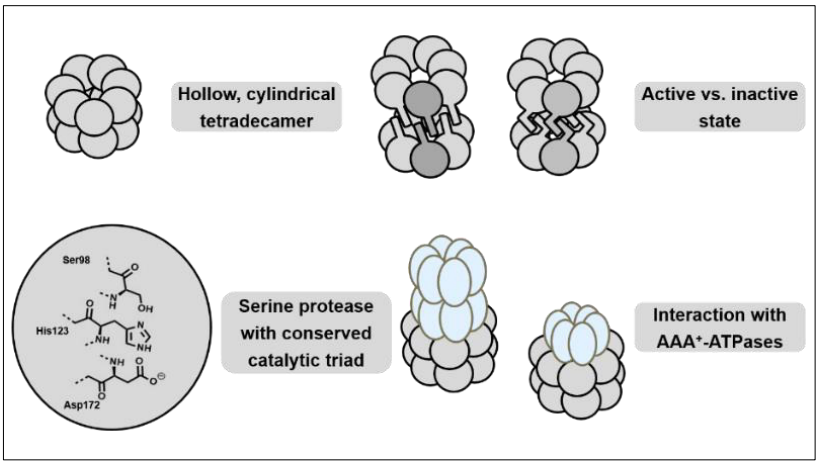


Figure 18: Schematic representation of the main characteristics of the caseinolytic protease ClpP.

Forming the inter-ring contacts, the head domain is the largest and most rigid part of each monomer. The handle that consist of α-helix E and β-strand 9 transmits the interaction between the two rings, whilst the N-terminal residues – roughly the first 25 amino acids of the (processed[5]) subunit – are responsible for the gate control at the axial pores. Both handle region and N-termini are structurally flexible to fulfil their roles in the catalytic mechanism of ClpP [151].

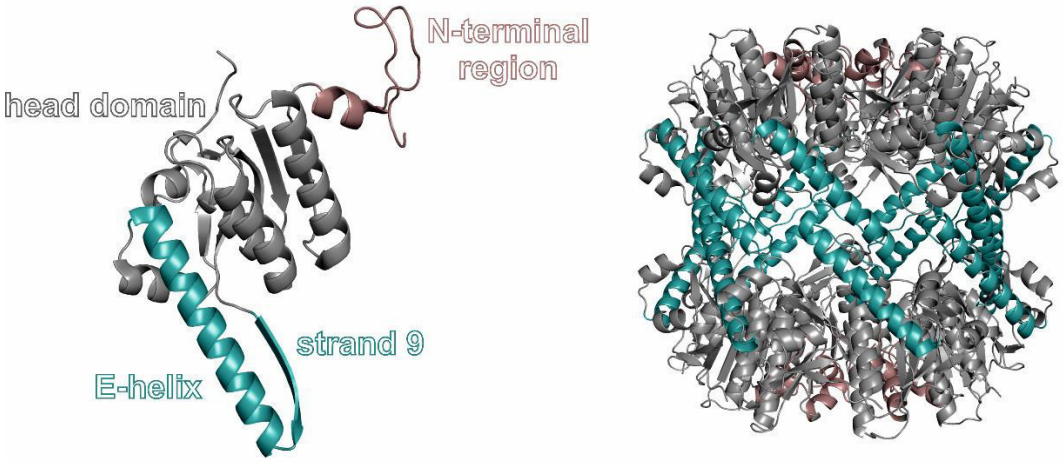


Figure 19: Subunit structure and assembly of the proteolytic complex of ClpP.

Each subunit of ClpP is divided into three different parts: the head domain, the N-terminus and the handle region consisting of the E-helix and strand 9. Fourteen of these subunits assemble into two heptameric rings stacked on top of each other to form the catalytically active complex (PDB ID: 1TYF, [149], head domain: grey, N-termini: rose, handle region: teal).

Already the initial investigations of ClpP revealed that, by itself, it is incapable of degrading most of its client proteins [152]. It has to bind to one of the AAA$^+$-ATPase of the Clp/Hsp100 family to gain its full proteolytic activity. Three different types of these

[5] Some organisms as *e.g. E. coli* produce ClpP with a precursor sequence, which is removed autocatalytically during particle maturation [150].

hexameric, ring-shaped ATPases that interact with ClpP were described so far: ClpA, ClpC and ClpX [153, 154]. All of them have other, sometimes overlapping substrate spectra and their (common) occurrence is species-dependent. Their main task is to select and prepare the proteins that shall be degraded by ClpP. For that, the ATPases bind on top and/or bottom of the proteolytic particle, forming together a bipartite degradation machinery [12]. A specific characteristic of this interaction is the apparent asymmetry between the hexameric rings of the ATPase and the heptameric rings of the protease. Its mechanistic implications are still controversial nowadays [155].

Once a particular substrate is bound by one of the Clp ATPases, it will be unfolded and, under the consumption of ATP, translocated into the proteolytic chamber of ClpP, where it gets hydrolysed subsequently [156]. The translocation of the substrate requires the opening of the axial pores of the tetradecameric protease complex, a process that is triggered by the binding of the ATPase and involves two different kinds of interactions [157]. First, highly-conserved surface loops dock into hydrophobic clefts on the surface of the proteolytic particle and second, the so-called pore-2-loops are in direct contact with the N-terminal motifs of ClpP, which, in absence of any binding partner close the axial pores [158]. To do so, the N-terminal residues must switch between an open and a closed state. It is known that they are able to form a small β-hairpin structure, but their proper characterization is difficult as no structure of the ATPase-protease complex exists so far and in most of the crystal structures of ClpP alone the corresponding regions were found to be disordered [157]. However, mutational and biochemical studies could identify some charged residue as well as a network of hydrophobic amino acids that are indispensable for this process [159, 160].

After a substrate passed the axial pores it reaches the proteolytic chamber of ClpP, which is formed between the two heptameric rings. It has a diameter of ~50 Å and harbours the 14 active sites [149]. The ‘side walls’ of the catalytic chamber are composed of the handle regions of the different subunits. Here, the interaction of the two heptameric rings is mediated via the formation of antiparallel β-sheets of the β9-strands of two oppositely located subunits [161]. An important prerequisite for β-sheet formation is the presence of a fully stretched out α-helix E as shown by crystallographic studies and molecular dynamics simulations, where three different ‘sub-forms’ of the tetradecameric particle could be identified [162]. Known as the extended, compressed and compact version of ClpP, they

differ both in their overall dimensions and, directly related to that, in the degree of disorder seen in the handle region of each subunit (Figure 20).

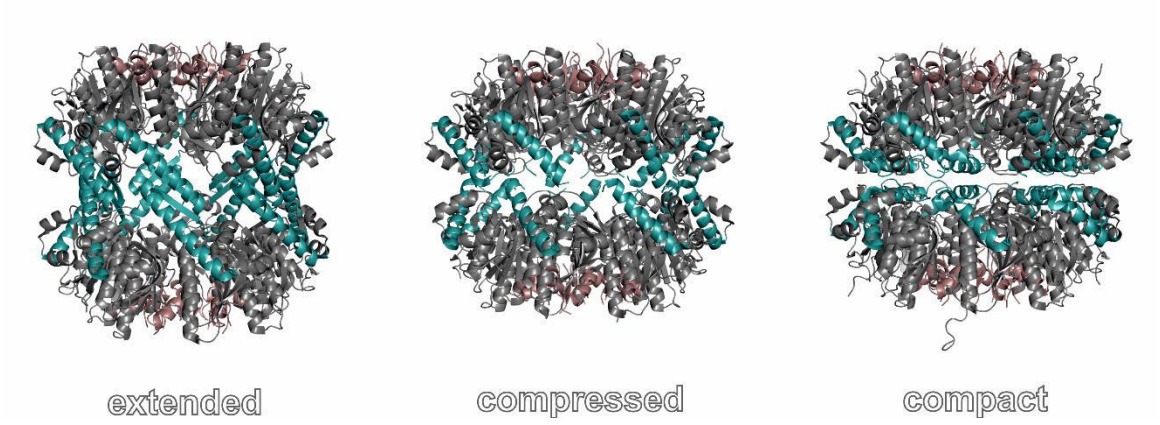


Figure 20: The extended, compressed and compact state of the ClpP tetradecamer.

ClpP crystallises in various 'sub-forms', that differ in their overall dimension as well as in the degree of structural disorder in the E-helix and strand S9 (ClpP from *Staphylococcus aureus*, PDB IDs: 3QWD (compact, [161]), 4EMM (compressed, [162]), 3V5E (extended, [163]), head domain: grey, N-termini: rose, handle region: teal). MD-simulations indicate that these states of the proteolytic particle might be present in solution and orchestrate the catalytic mechanisms of the protease.

Since the active site residues are located close by the handle region, it was suggested that the various states of the tetradecamer might be involved in regulation of the proteolytic activity of ClpP [162]. In fact, only in the extended version of the particle all three active side residues are fully resolved and hydrogen bonded to each other [163]. Inhibitor studies with diisopropyl fluorophosphate revealed for the first time that ClpP must be a serine protease [164]. Mutational as well as structural studies identified Ser97, His122 and Asp171 [165] as the catalytic triad of ClpP in *E. coli*[6] [149, 166]. For proper activity, all three residues need to be present and aligned with each other, else no proton transfer between them can take place, which is, however, an absolute requirement for formation of the active site nucleophile and, thereby, activity. Looking at the different states of the ClpP particle, it was found that the two amino acids located upstream of Asp171 play an important role in the arrangement of the catalytic site. Both the asparagine and aspartate residue engage in an inter-ring salt bonding network with the corresponding amino acid in the adjacent ring. In case of this interaction taking place, Asp171 is hydrogen bonded to His122, otherwise they are turned apart. Therefore, this interaction was defined to be a sensor for the formation of the proteolytically active particle [163].

[6] The numbering of ClpP sequences from different organisms is not defined consistently. In *E. coli*, one talks about Ser97, His122 and Asp171. The corresponding residues in *L. monocytogenes* are Ser98, His123 and Asp172. Here, the N-terminal methionine residue was considered for numbering; compare also Figure 22.

Taken together, the formation of the tetradecameric ClpP complex out of two heptameric rings is highly intertwined with the alignment of the active residues and, thus, also its proteolytic activity. This can be seen as a safety mechanism, which guarantees that the active sites only align accurately in the closed barrel-shaped assembly to prevent unspecific degradation of proteins. Other studies suggested that the flexibility in the inter-ring contacts might be involved in opening of equatorial pores in the ClpP particle for product release [151]. Thus, the exact correlation between the plasticity of the inter-ring interactions and the function of the protease is yet under debate [167].

7.1.5 The significance of ClpP for the virulence of *Listeria monocytogenes*

The human pathogen *L. monocytogenes* is a Gram-positive, facultative intracellular bacterium. It causes listeriosis, which invasive form leads to sepsis or meningoencephalitis and has a comparably high fatality rate of up to 30 % [168]. The severe manifestation of this disease is primarily found in immunocompromised, older and/or pregnant individuals as well as new-borns [169]. In general, an infection is transmitted by the consumption of contaminated food. As *L. monocytogenes* can resist harsh growth conditions with regard to pH, temperature and osmotic strength, its actual danger for food safety is high [168]. The specific threat comes from its ability to cross the three tight physiological barriers: the intestinal, the blood-brain and the mother-to-foetus barrier [170]. Various virulence genes differently expressed during the life/infectious cycle of this pathogen are key elements for the high infection rates (Figure 21) [171, 172].

What role does ClpP play for the virulence of *L. monocytogenes*? It was shown that expression of the *clpp* gene, together with *clpc* and *clpe* is negatively regulated by the presence of the CtsR regulator. However, at elevated temperatures this protein gets inactivated quickly suggesting a role for ClpP in the cellular stress response [173]. By showing that constitutive expression of *ctsr* significantly decreased virulence, it was proven that there exists a direct connection between the proper function of ClpP and the virulence of *L. monocytogenes* both *in vitro* and *in vivo* [174]. Further studies revealed ClpP being indispensable for the release of *L. monocytogenes* into the cytoplasm of bone marrow-derived macrophages [175]. Interestingly, bacteria carrying a genomic Δ*clpp* mutation were not capable of growing in macrophages during the early stages of an infection in mice [176]. Thus, ClpP is essential for the virulence of *L. monocytogenes*.

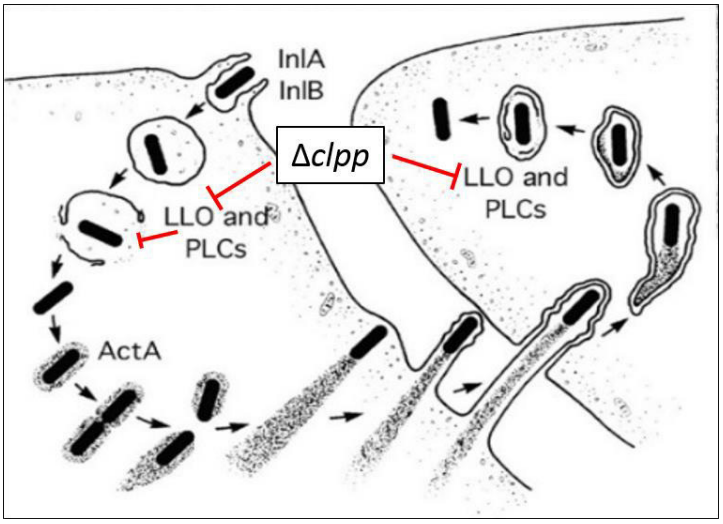


Figure 21: The effect of a Δ*clpp* mutation in *Listeria monocytogenes*.

The first step in the infectious cycle of *L. monocytogenes* is the invasion of the target cell, stimulated and mediated by many surface-exposed virulence factors as *e.g.* the internalins A and B (Inl A/B) [177]. Afterwards, *L. monocytogenes* is trapped in a vacuole which rapidly fuses with a lysosome. The change in pH activates listeriolysin O (LLO). Together with two distinct phospholipases C (PLCs) it triggers the escape into the cytoplasm [170]. Once released, the crucial tasks for *L. monocytogenes* is to multiply and to infect the neighbouring cells. The pathogen developed a specific system that provides both its intracellular motility and the possibility to move from one cell to another. The protein ActA initiates the formation of an actin filament on one side of the bacterium via actin nucleation and polymerisation. Similar to a propeller, this filament enables the movement through the cytosol and it can also literally catapult the bacterium into adjacent cells [178]. A Δ*clpp* mutant results in development of avirulent bacteria, which can be directly traced back to incorrect processing of LLO and, thereby, to reduction/absence of the haemolytic activity of the bacterium. [175]. Parts of this figure were reproduced from [171]. The original figure by Daniel A. Portnoy, Victoria Auerbuch, and Ian J. Glomski has been licensed under the Creative Commons Attribution-Share Alike 3.0 United States license.

7.1.6 The *Lm*ClpP1/2 heterocomplex

Based on the results of a Southern Blot analysis, Gaillot *et al.* claimed that *L. monocytogenes* encodes one copy of the protease (later referred to as ClpP2 or *Lm*ClpP2), in agreement with the majority of bacteria [175]. Further studies, however, revealed that this pathogen, just like *Mycobacterium tuberculosis* or *Bacillus thuringiensis*, possesses a second isoform of the ClpP protein (later referred to as ClpP1 or *Lm*ClpP1, [179, 180]).

Figure 22 displays the sequence alignment of *Lm*ClpP1 and 2 with the single copy of ClpP from *Escherichia coli* (*Ec*ClpP). The sequence similarity between the two *Listeria* isoforms is with 41 % much lower than the one between *Lm*ClpP2 and *Ec*ClpP, which was determined to 66 %. With about 77 % the similarity of ClpP proteins from *Bacillus subtilis* or *Staphylococcus aureus* to ClpP2 is even higher [175, 181]. Most of the differences in the primary sequences of *Lm*ClpP1 and *Lm*ClpP2 are located in their N-terminal parts. The first few residues are affected in particular, possibly indicating that these amino acids fulfil different functions in the two ClpP proteins from *L. monocytogenes*. Moreover, Asp172 in the active centre of *Lm*ClpP1 is replaced by an asparagine. These findings raised the

question, whether both isoforms have a similar structure and function and can interact with each other.

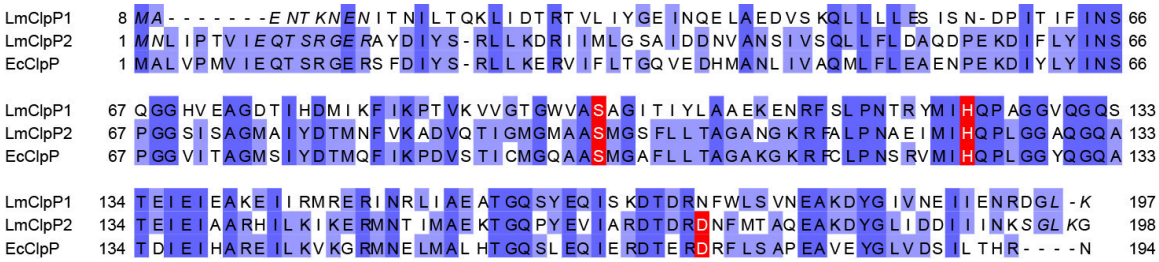


Figure 22: Alignment of the sequences of *Lm*ClpP1 and *Lm*ClpP2 with ClpP from *E. coli*.

Amino acids conserved in either two or all three sequences were labelled in light or dark blue, respectively. Active side residues were highlighted in red. The sequence for *Lm*ClpP1 was renumbered, thus the catalytic centre has the same nomenclature in both ClpP isoforms. Residues shown in italics were not resolved in the crystal structure of the *Lm*ClpP1/2 heterocomplex (Figure 24). Adapted and reproduced from Dahmen, M., Vielberg, M.-T. *et al.* [15] with permission of the Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.

Zeiler *et al.* showed that *Lm*ClpP1 on its own has no peptidiolytic activity, lacks the ability to react with β-lactones and forms heptameric rings rather than tetradecameric complexes (PDB ID: 4JCQ, [182]). *Lm*ClpP2 instead is typically found as a tetradecamer, binds to distinct inhibitors and degrades commonly used peptide substrates (PDB ID: 4JCT, [182]). Usage of a bicyclic β-lactone-probe, derived from the natural product vibralactone, gave the first indication that both ClpP isoforms together may form a heterocomplex *in vivo*. In the lysate of *L. monocytogenes* cells vibralactone also labelled ClpP1, but only in presence of ClpP2, [181]. SEC and negative stain electron-microscopy demonstrated that the heterocomplex also consists of two heptameric rings stacked on top of each other. As top and bottom of the particles are differently shaped, it was concluded that two homomeric rings were formed and no mixing between them occurred [181]. To get an idea, what the advantage of such an arrangement could be, the group around Prof. Dr. Stephan A. Sieber further compared both the peptidiolytic and, in presence of the AAA^+-ATPase ClpX, the proteolytic activity of homomeric with heteromeric complexes of ClpP proteins in *L. monocytogenes* (for experimental details see [15]).

Figure 23a illustrates that the peptidase activity of the ClpP2 homocomplex is higher than that of the heterocomplex. Abolishing the peptidolytic properties of either ClpP1 or ClpP2 by introducing a S98A mutation proved that both ClpPs contribute to the total activity of the heterocomplex [182]. Unexpectedly, in presence of *Lm*ClpX the degradation of GFP-ssrA by the heterocomplex exceeded that of the homocomplex by about ten times, regardless of the presence of seven or fourteen operable active sites (Figure 23b) [15]. Both the crystal structures of ClpP1 and ClpP2 alone as well as the electron microscopic

reconstruction of the heterocomplex did not provide a reasonable explanation for these observations. In search of one we solved the crystal structure of the *Lm*ClpP12 heterocomplex.

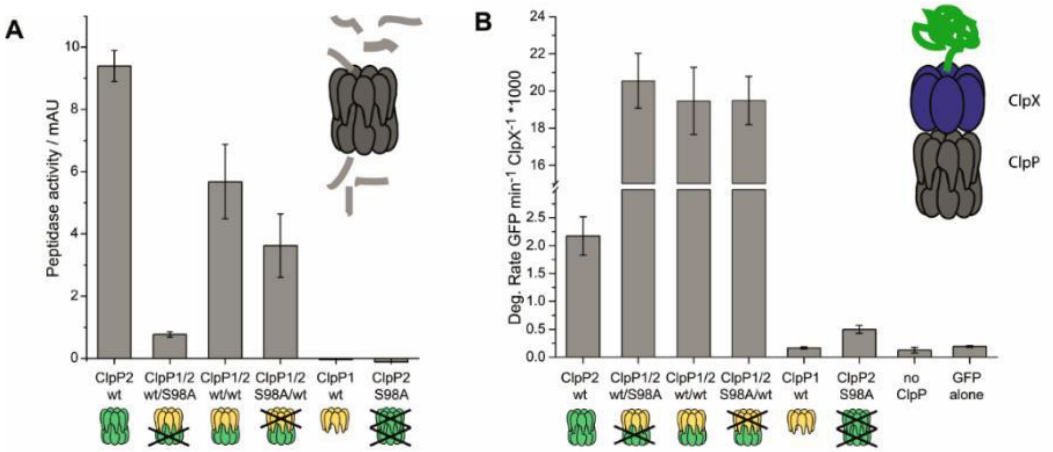


Figure 23: Comparison of the degradation rates of hetero- and homooligomeric ClpPs from *Listeria*.

(a) The peptidase activity of a ClpP2 homocomplex from *Listeria monocytogenes* deviates not significantly from that of the ClpP1/2 heterocomplex. **(b)** For the degradation of GFP the situation is quite different. In presence of *Lm*ClpX, the heterocomplex exhibits an about 10-fold higher degradation rate than observed for tetradecameric ClpP2. Reproduced from Dahmen, M., Vielberg, M.-T. *et al.* [15] with permission of the Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.

7.2 Results

7.2.1 Solving the structure of ClpP1/2 from *L. monocytogenes*

After size exclusion chromatography of the *Lm*ClpP1/2 tetradecamer in 20 mM MOPS, pH 7.0, 100 mM KCl, 5 % glycerol, fractions containing the tetradecamer were concentrated to about 10 mg/mL (for further details compare SI in [15]). The protein was used either directly or after overnight storage at 4 °C for crystallization trials at 20 °C. Initial hits were found to be mostly PEG- or MPD-based, thus NeXtal screens 8, 9 and 11 were used for fine screening (see section 5.1.11). The majority of crystals did not diffract to a reasonable resolution; very likely due to the overall size of the heterocomplex compared to the size of the crystals. Eventually, it was found that well-diffracting crystals grew in drops of 0.4 µL size containing equal amounts of protein and 0.2 M sodium malonate, 40 % MPD as reservoir solution within a month. Crystals were cryoprotected by briefly soaking them in mother-liquor.

*Lm*ClpP1/2 crystallised in space group $P2_12_12_1$ with one tetradecamer per asymmetric unit. The crystals diffracted to a final resolution of 2.8 Å (R_{free}=21.7 %, Table 9). Calculation of self-rotation functions revealed the presence of 7-fold non-crystallographic symmetry. Therefore, Patterson search techniques were performed with the coordinates of a complete heptameric ring of homomeric *Lm*ClpP1 and *Lm*ClpP2 each (PDB IDs: 4JCQ and 4JCT, [182]. As mentioned above, α-Helix E and β-strand 9 may have a high conformational flexibility and were, thus, removed from the search models. Both rings could be placed unambiguously in the asymmetric unit, confirming the presence of the heterocomplex within the crystals.

The 7-fold non-crystallographic symmetry was exploited during model building. One chain each for *Lm*ClpP1and *Lm*ClpP2 was built into the NCS-averaged electron density map and then transferred to all NCS-related copies. Side chains were allowed to refine freely to take differences between the distinct monomers into account and each chain was rebuilt individually. Final TLS-refinement resulted in good values for R_{work} and R_{free} as well as for r.m.s.d. bond and angle values. The Ramachadran plot showed that 97.9 % or 2.1 % of the residues were found to be located in favoured or allowed regions, respectively (Table 9). Numbering of the residues of ClpP1 and ClpP2 took place as indicated in Figure 22. Residues that were found to be disordered in the crystal structure are shown in italics there.

Atomic coordinates and structure factors of the *Lm*ClpP1/2 heterocomplex were deposited in the PDB under the accession code 4RYF (see section 5.1.12).

Table 9: X-ray data collection and refinement statistics.

	ClpP1/2 heterocomplex
Crystal parameters	
Space group	$P2_12_12_1$
Cell constants	a=97.2 Å b=127.2 Å c=265.5 Å
Molecules per AU[a]	1
Data collection	
Beam line	X06SA, SLS
Wavelength (Å)	1.0
Resolution range (Å)[b]	30-2.8 (2.9-2.8)
No. observations	372002
No. unique reflections [c]	78792
Completeness (%)[b]	96.2 (98.1)
R_{merge} (%)[b, d]	9.6 (48.0)
I/σ(I)[b]	14.2 (3.5)
Refinement *(REFMAC5)*	
Resolution range (Å)	15-2.8
No. refl. working set	74797
No. refl. test set	3937
No. non-hydrogen atoms	19892
Solvent molecules	175
R_{work}/R_{free} (%)[e]	19.1/21.7
r.m.s.d. bond (Å)/ angle (°)[f]	0.007/1.110
Average B-factor (Å^2)	46.3
Ramachandran Plot (%)[g]	97.9/2.1/0
PDB accession code	**4RYF**

[a] Asymmetric unit.

[b] The values in parentheses of resolution range, completeness, R_{merge} and I/σ(I) correspond to the last resolution shell.

[c] Friedel pairs were treated as identical reflections.

[d] $R_{merge}(I) = \Sigma_{hkl}\Sigma_j | [I(hkl)_j - <I(hkl)>] | / \Sigma_{hkl}\Sigma_j I(hkl)_j$, where $I(hkl)_j$ is the j^{th} measurement of the intensity of reflection hkl and $<I(hkl)>$ is the average intensity.

[e] $R = \Sigma_{hkl} | |F_{obs}| - |F_{calc}| | / \Sigma_{hkl} |F_{obs}|$, where R_{free} is calculated for a randomly chosen 5 % of reflections, which were not used for structure refinement, and R_{work} is calculated for the remaining reflections.

[f] Deviation from ideal bond lengths/angles.

[g] Number of residues in favoured, allowed or outlier regions.

7.2.2 Structural characteristics of *Lm*ClpP1/2

7.2.2.1 General features of the heterocomplex

*Lm*ClpP1/2 has the same barrel-shaped structure as other tetradecameric ClpP complexes, an overall length of 100 Å and a diameter of 95 Å (Figure 24a). Modelling of heterocomplex based on electron microscopic images already suggested that ClpP1 and ClpP2 do not form mixed heptameric rings [181]. The same was found to be true for the crystal structure (Figure 24a/b). Only the search with the two homomeric rings resulted in a PHASER solution that could be refined satisfyingly. To confirm further that the rings of ClpP1 and ClpP2 were placed correctly, the side chains of different bulky amino acids were

excluded from the model and then the whole structure was refined once again to show that there was still electron density present for the lacking atoms. In addition, superposition of one ClpP1 subunit with one subunit of the ClpP2 ring demonstrated that formation of mixed rings is prevented by steric clashes that would occur between adjacent subunits (Figure 24c). Thus, the crystal structure of the heterocomplex reveals that it possesses two different interfaces, each composed of subunits of only ClpP1 or ClpP2, respectively.

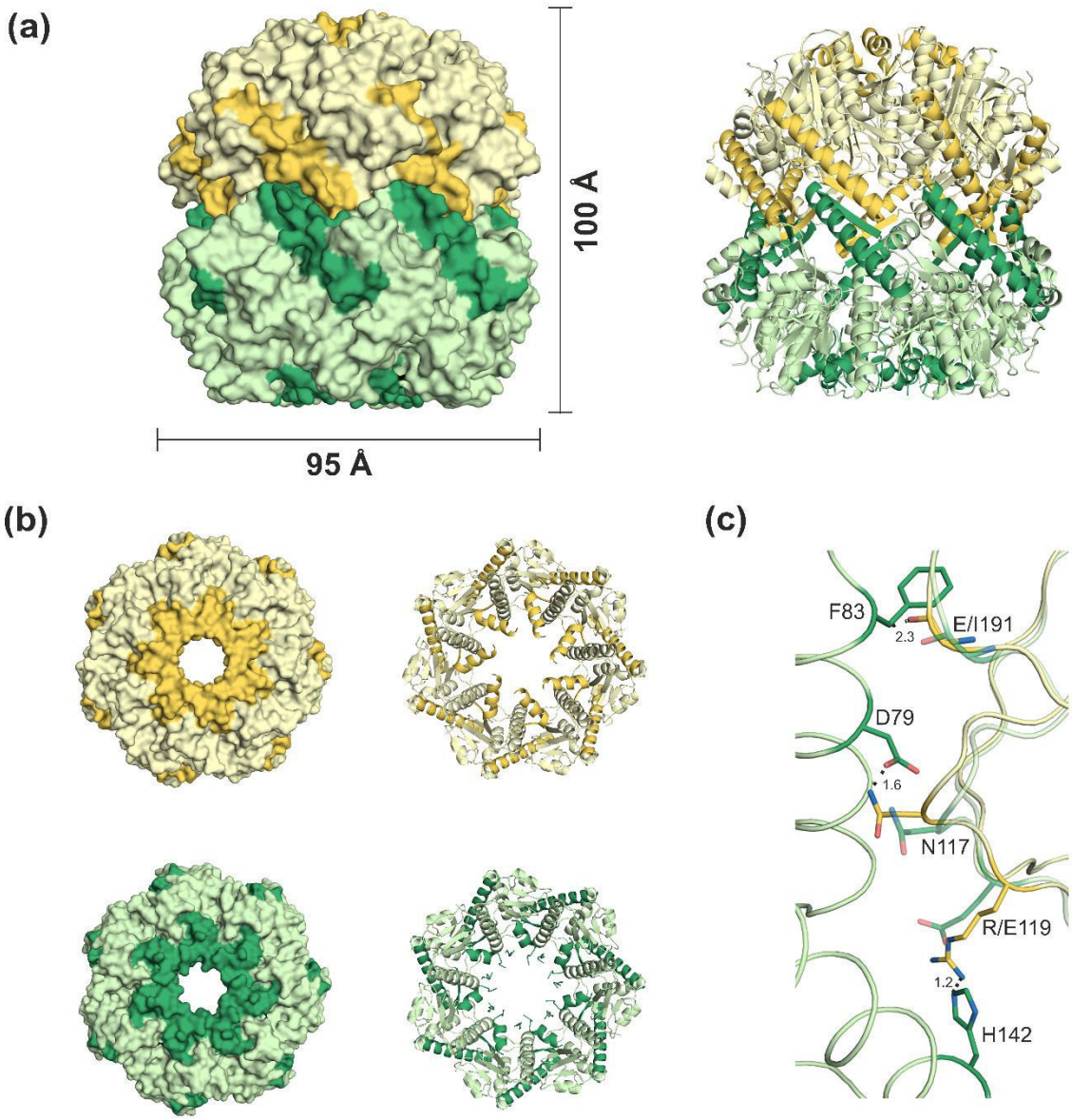


Figure 24: Overall structure of the ClpP12 heterocomplex from *Listeria monocytogenes*.

(a) Side view of the *Lm*ClpP1/2 heterocomplex, both in surface and cartoon representation (ClpP1 and ClpP2 shown in yellow or green, respectively; N-terminal regions as well as α-helix E and β-strand 9 are highlighted in dark). **(b)** Top view of the ClpP1- or the ClpP2-ring, respectively. **(c)** Formation of mixed heptameric rings is prevented by specific subunit-subunit interactions. Overlay of a single ClpP1 subunit onto the ClpP2-ring (only two adjacent subunits are shown). Clashing main or side chains are highlighted. Adapted and reproduced from Dahmen, M., Vielberg, M.-T. *et al.* [15] with permission of the Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.

7.2.2.2 The heterocomplex in its active, extended form

*Lm*ClpP1/2 has several special features in comparison to the previously published homomeric structures of ClpP1 and ClpP2 from *L. monocytogenes*. Most of the differences are located near α-helix E and β-strand 9 (Figure 25a). In contrast to the former structures, the E-helix is fully extended in the heterocomplex and a newly formed β-sheet of one strand 9 from each ClpP connects the two rings (Figure 26b). Both are characteristics of the extended form of the proteolytic particle, showing that the heterocomplex crystallised in an active state. This is also reflected in the spatial arrangement of the residues of the catalytic site, which are aligned in all fourteen subunits (Figure 25b).

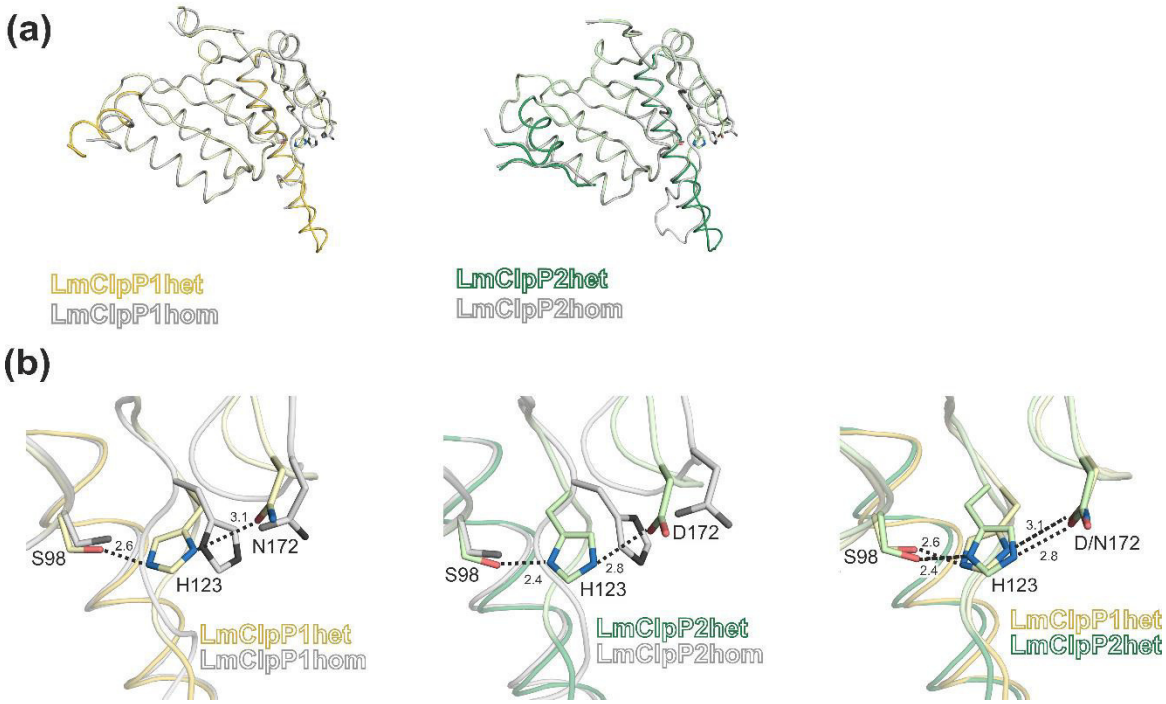


Figure 25: Differences between the homomeric and heteromeric subunit structures of *Lm*ClpP1 and *Lm*ClpP2.

(a) Overlay of one subunit of homomeric ClpP1 and ClpP2 (grey) and the corresponding subunits in the heterocomplex (in colour). Active site residues are shown as sticks. **(b)** Zoom-in to the catalytic triad (left, centre) as well as comparison of the active site in the two subunits of the heterocomplex (right) indicating the importance of Asn172 for the activity of *Lm*ClpP1. Adapted and reproduced from Dahmen, M., Vielberg, M.-T. *et al.* [15] with permission of the Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.

In ClpP2 the hydroxyl group of Ser98 points towards the basic nitrogen atom of the side chain of His123, whereas the other nitrogen is hydrogen-bonded to Asp172 (Figure 25b middle, Figure 26c). The distances between the three residues are such that a proton-transfer, prerequisite for the catalytic activity, may take place. A similar picture is drawn for ClpP1 (Figure 25b left, Figure 26a). Although Asp172 is mutated to an asparagine residue, Asn172 likewise seems to be part of the active centre as the side chains of both residues are located at exactly the same position shown in the superposition of one subunit

from each ring (Figure 25b right). The distances between Ser98, His123 and Asn172 also suggest an important role of all three residues for the catalytic mechanism.

The superposition of *Lm*ClpP1/2 with the respective homomeric ClpP1 and 2 subunits illustrates that formation of the inter-ring β-sheet directly affects the position of His123 and thereby the correct alignment of the catalytic triad (Figure 25a/b). Thus, the crystal structure of the heterocomplex is the first ClpP structure from *Listeria monocytogenes*, where the subunits, more precisely the E-helices are found in their fully extended form and the active sites are properly aligned to allow substrate catalysis.

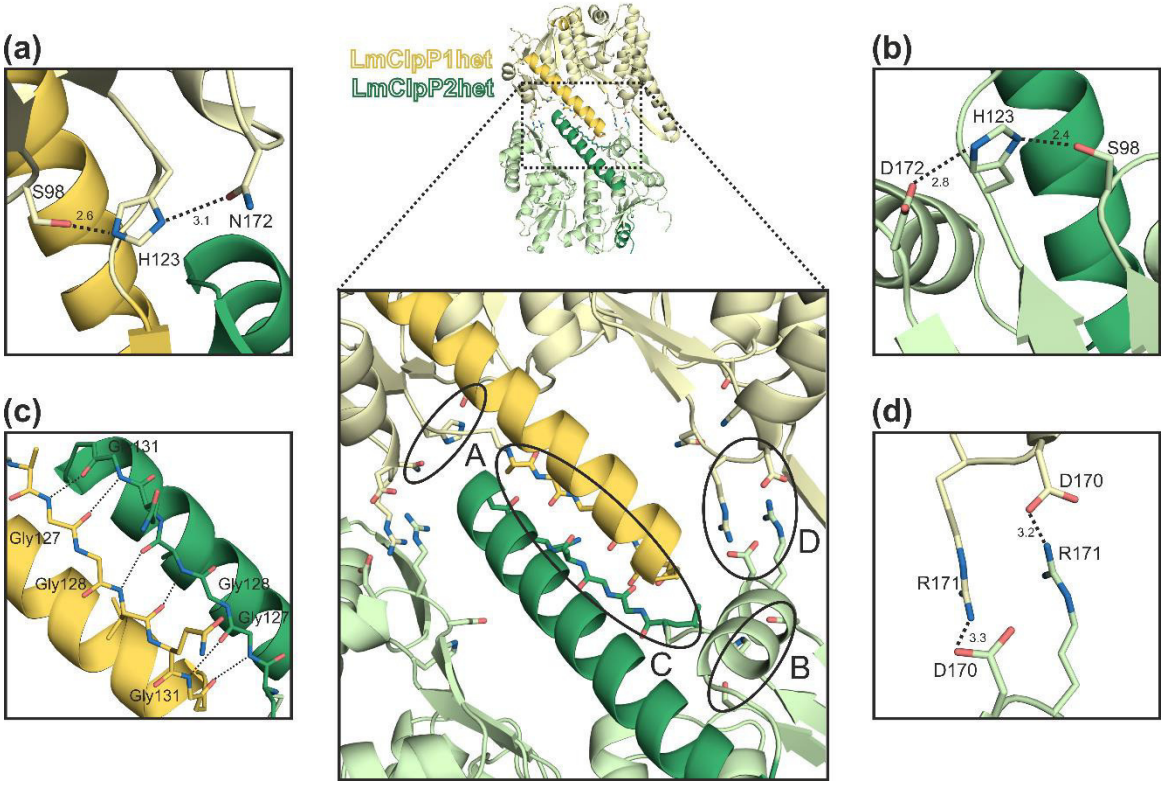


Figure 26: Important properties of the active *Lm*ClpP1/2 heterocomplex.

(a) Arrangement of the active site residues in *Lm*ClpP1. **(b)** Alignment of the catalytic triad in *Lm*ClpP2. **(c)** The inter-ring β-sheet is formed by backbone interactions of three highly-conserved glycine residues. **(d)** Hydrogen bond formation in the aspartate-arginine sensor motif. *Lm*ClpP1 and *Lm*ClpP2 are shown in yellow and green, respectively. Oxygen atoms are depicted in red, whereas nitrogen atoms are coloured blue. Dotted lines symbolize hydrogen bonds. All length specifications are given in Å. Adapted and reproduced from Dahmen, M., Vielberg, M.-T. *et al.* [15] with permission of the Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.

The Asp170-Arg171 sensor is located next to the active site(s) (Figure 26d). Former findings stated that the conformation of the E-helix is closely linked to the question whether the two rings interact with each other or not. Consistent with that, hydrogen-bonds are formed between Asp170 from one and Arg171 from the other ring also confirming that *Lm*ClpP1/2 crystallised in its active form.

The other contact site between the two rings is the inter-ring β-sheet formed by opposite β-strands 9. Remarkably, most of the contacts are mediated by the main chain atoms of three glycine residues that, in turn, are conserved between ClpP1 and ClpP2 (Figure 26b). So, the inter-ring contacts in the homomeric 14mer of ClpP2 are essentially the same as in the heterocomplex – there is no additional stabilization of the active state of the latter.

7.2.2.3 Differences between *Lm*ClpP1 and 2: N-termini and S1 pockets

Despite the structural similarities between the subunits of *Lm*ClpP1 and 2, they are unambiguously distinguishable in at least two aspects: first, the secondary structure of their N-terminal residues and second, the shape of their substrate binding pockets. The sequence alignment of the primary sequences of the two ClpP isoforms from *L. monocytogenes* together with *Ec*ClpP already indicated *Lm*ClpP1 to feature specific properties not present in other ClpPs (Figure 22). While there exists a high degree of sequence conservation between *Lm*ClpP2 and *Ec*ClpP, the N-terminal region of *Lm*ClpP1 is about 30 % shorter and its sequence deviates significantly. A superposition of the N-termini of the three proteins shows that all of them are partly α-helical (Figure 27a). In *Lm*ClpP2 and *Ec*ClpP the residues that are located in front of the α-helix, although partly disordered, form a β-hairpin structure that lines up the N-terminal pore as described for other ClpP structures [153]. No such observation is made for *Lm*ClpP1. All residues resolved in the crystal structure are located within the N-terminal helix and it is more likely that the missing residues will either prolong it or a part of a disordered loop rather than also forming a β-sheet. In addition, the superposition of the three ClpP monomers revealed that the angle, at which the N-terminal helices protrude from the rest of the protein, varies profoundly between *Lm*ClpP1 and *Lm*ClpP2, whereas *Ec*ClpP and *Lm*ClpP2 are identical in this regard (Figure 27a). Rotating the superposition at about 90° reveals that the β-sheet forming residues in *Lm*ClpP2 point in exactly the same direction as the α-helix of *Lm*ClpP1 and that the disorder in their N-termini initiates at almost the same position. So, it seems that, despite all differences, the axial pores on both sites of the ClpP1/2 heterocomplex are opened or closed to almost the same extent, respectively.

As seen in the surface representation of *Lm*ClpP1/2, the distinct N-terminal secondary structures also have an impact on the appearance of the heterocomplex in general (Figure 27b). While the side that is formed by *Lm*ClpP2 seems to be rather flat, the one by *Lm*ClpP1 is bent considerably stronger. As these are the potential interaction sites for the AAA^+-

chaperones, a variance in the interfaces might be reflected in a different binding behaviour towards them. Remarkably, the residue that form the hydrophobic binding cleft for the conserved IGF/LGF-loop of the ATPases are not conserved. The most striking substitution is T61F exchanging a small and hydrophilic residue in *Lm*ClpP1 with a much bulkier and hydrophobic one in *Lm*ClpP2 (Figure 27b).

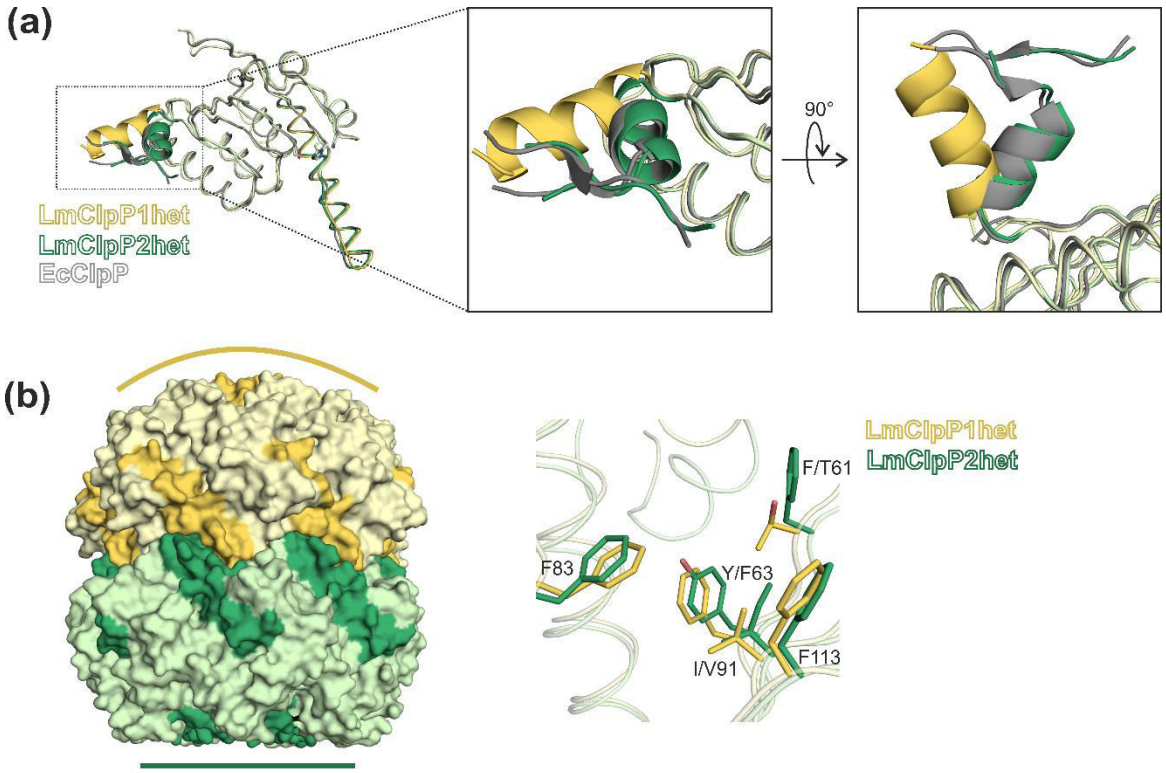


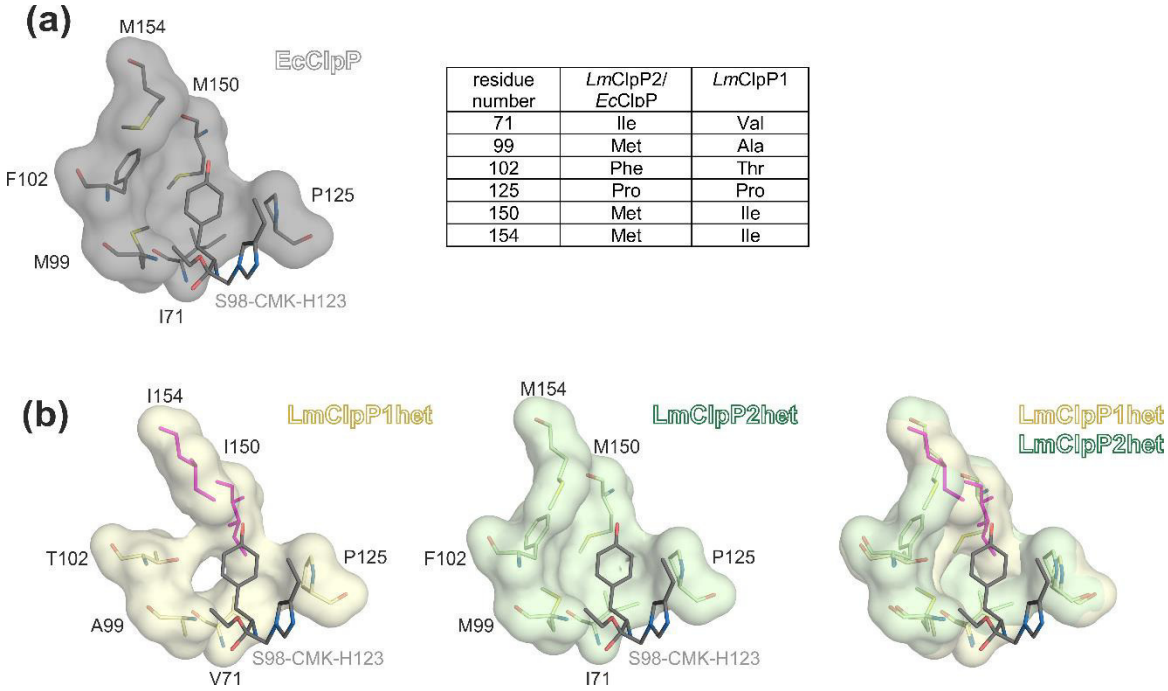
Figure 27: Asymmetry in the apical sites of the heterocomplex.

(a) Overlay of the monomeric structures of *Lm*ClpP1, 2 and *Ec*ClpP analyzing structural differences in the N-terminal regions. The axial pore in *Lm*ClpP1 is closed by an α-helix, whilst *Lm*ClpP2 forms the typical β-hairpin motif, known from other ClpP structures. **(b)** The N-terminal variations influence the overall shape of the heterocomplex: one apical site is considerably flatter than the other (left), suggesting an altered binding behaviour regarding the Clp-ATPases. This is supported by the finding that some of the residues in the hydrophobic binding clefts for the IGF/LGF-loops of the interacting ATPase are not conserved between the two isoforms (right), (*Lm*ClpP1: yellow, *Lm*ClpP2: green, *Ec*ClpP: black, oxygen atoms: red). Adapted and reproduced from Dahmen, M., Vielberg, M.-T. *et al.* [15] with permission of the Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.

*Ec*ClpP was the first representative of the ClpP proteases, whose crystal structure was solved in presence of an inhibitor covalently bound to the active site of the protease [183]. The peptide ligand with a chloro-methyl-ketone group encodes a tyrosine residue in its P1 position enabling the identification of the residues that form the S1 pocket in *Ec*ClpP (Figure 28a).

The comparison with *Lm*ClpP2 showed the similarity between the two proteins to be reflected in this aspect. All residues identified to be important for *Ec*ClpP are found in *Lm*ClpP2. This, in turn, made it possible to compare the shapes of the S1 pockets of the

two isoforms from *L. monocytogenes*, which differ both in size and hydrophobicity. In particular, the substitution of Met150 and Met154 by isoleucine residues contributes to these divergences. Superposition of the structures of *Lm*ClpP1 and *Lm*ClpP2 together with the inhibitor from *Ec*ClpP reveals that a bulky amino acid in the P1 position of an inhibitor will most likely prevent its binding to the former, whereas it should fit perfectly into the S1 pocket of the latter (Figure 28b).



residue number	*Lm*ClpP2/ *Ec*ClpP	*Lm*ClpP1
71	Ile	Val
99	Met	Ala
102	Phe	Thr
125	Pro	Pro
150	Met	Ile
154	Met	Ile

Figure 28: Identification of the S1 pockets in *Lm*ClpP1 and *Lm*ClpP2.

(a) Crystallization of *Ec*ClpP in complex with Z-LY-CMK made it possible to assign the residues forming its S1 pocket (left, PDB ID: 2FZS). The high sequence identity between *Ec*ClpP and *Lm*ClpP2 allowed to identify the respective pockets in the heterocomplex from *L. monocytogenes*. (right). **(b)** Structural comparison of the shape of the S1-pockets in *Lm*ClpP1 and 2. Substitution of Met150 and Met154 by isoleucine (pink) restrict the space in the S1 pocket of *Lm*ClpP1 and give it a more hydrophobic character (*Ec*ClpP: grey, Z-LY-CMK: grey, only the head group and the tyrosine residue are shown sticks, LmClpP1: yellow, LmClpP2: green, oxygen, nitrogen and sulphur atoms: red, blue and yellow, respectively). Adapted and reproduced from Dahmen, M., Vielberg, M.-T. *et al.* [15] with permission of the Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.

7.3 Discussion

Starting point for our crystallographic studies was the question how *Lm*ClpP1/2 differed from homomeric ClpP structures of the same or other organisms and whether these variations might provide any explanation for the accelerated protein degradation rate observed in presence of ClpX.

7.3.1 Unique properties of the *Lm*ClpP1/2 heterocomplex

In contrast to previous structures of homomeric *Lm*ClpP1 and 2, the heterocomplex crystallised in its active, extended form. This is seen by the fully extended α-helix E, the formation of the inter-ring β-sheet and the hydrogen bonds in the arginine-aspartate sensor motif as well as by alignment of the catalytic residues in all fourteen active sites (Figure 26a-d). Previous investigations on the role of *Lm*ClpP1 within the heterocomplex indicated an N172A mutation to abolish its catalytic activity completely [182]. The crystal structure underlines the importance of Asn172 as part of the active site. The residue is well-defined in the electron density map and its sidechain carbonyl oxygen atom is hydrogen bonded to His123 (Figure 25b). Independent of its orientation, an asparagine residue is not able to participate in the proton transfer reaction between the active site residues. Hence, *Lm*ClpP1 harbours a catalytic dyad of Ser98 and His123 stabilized by the polarizing activity of Asn172. Studies on other serine proteases revealed deviations from the typical Ser/His/Asp triad to be more common than expected [165]. In addition to serine and histidine, also an asparagine residue was proposed to be essential for the activity of rhomboid proteases [184, 185].

Apart from the active site architecture, there are other features, in which *Lm*ClpP1 and 2 differ significantly. Since the sequence identity between the two isoforms from *Listeria* is rather low compared to that between *Lm*ClpP2 and *Ec*ClpP, the S1 pockets in the heterocomplex vary both in their size and hydrophobicity (Figure 28). A recent follow-up study by the Sieber group confirmed that the different properties of the pockets could be used to develop subunit-specific inhibitors [186].

The finding that the apical sides of the heterocomplex are shaped differently is a direct result of the variations in the N-terminal region of the two isoforms. Unlike in other ClpP proteins, the axial pore in *Lm*ClpP1 is enclosed by N-terminal helices instead of β-sheets

(Figure 27a). This and a mutation in the residues forming the hydrophobic binding clefts on the *Lm*ClpP1 interface indicate a potentially asymmetric binding behaviour of *Lm*ClpX towards the heterocomplex (Figure 27b, [182]). The biochemical assays that accompanied the crystallographic analysis confirmed the assumption that the AAA^+-ATPase only interacts with *Lm*ClpP2 (for more details see [15]).

Except for the unilateral binding of the *Lm*ClpX chaperone, the structure of the hetero-complex does not provide any hint for the increased proteolytic activity observed in comparison to homomeric ClpP2. As stated above, *Lm*ClpP1/2 lacks any kind of special stabilization as all inter-ring contacts are formed by either main chain atoms or residues equally conserved in both isoforms. This is confirmed by an observation made in the process of structure elucidation. Micro-seeding with crystals of *Lm*ClpP1/2 resulted in growth of crystals containing *Lm*ClpP1 alone instead of the heterocomplex, indicating it to be a rather transient complex and most likely in equilibrium with tetradecameric ClpP2 particles and heptameric ClpP1-rings.

To date no high-resolution structure of any of the bipartite degradation machineries of the Clp family exist. Various reasons are imaginable: i) The symmetry mismatch between ClpP and its ATPases might prevent a regular crystal formation. ii) The Clp ATPases themselves often do not crystallise in a ring-shaped, but in a helical structure instead. The only ring-shaped structures published so far were synthetic constructs, where different subunits were covalently linked to each other. But even these did not diffract better than 3.2 Å [187].

7.3.2 ClpP heterocomplexes in *Listeria* and *Mycobacterium*

Formation of heteromeric ClpPs is not limited to *L. monocytogenes*, but was observed in the cyanobacterium *Synechococcus* and the actinobacterium *Mycobacterium tuberculosis* (*Mt*ClpP1/2) as well. Whilst the former was reported to produce tetradecameric complexes with heptameric rings consisting of two different subunits [188], ClpP1 and ClpP2 of the latter assemble into separate rings, which then bind to each other [189]. Structural studies of the resulting heterocomplex (PDB IDs: 4U0G [190] or 5E0S [191]) allow a detailed comparison with *Lm*ClpP1/2. The sequence similarity of the two isoforms of *M. tuberculosis* amounts to about 50 %. Remarkably, each of the two subunits has an even higher similarity score with *Lm*ClpP2 (~60 %), whilst they are less related to *Lm*ClpP1 (~45 %). This is also reflected in their active site composition as, according to their

sequences, both *Mt*ClpP1 and 2 encode the classical catalytic triad of serine, histidine and aspartate. In contrast to *Listeria monocytogenes*, where ClpP2 is able to form a homomeric, proteolytically active complex, *Mt*ClpP1 and 2 are only active together. Even though overproduction of each isoform alone leads to formation of tetradecameric complexes, the assembled particles lack any kind of activity. By chance, it was discovered that addition of short N-capped peptides like *e.g.* Z-Leu-Leu resulted in dissociation of the homomeric complexes into heptameric rings, which then assembled again to form active heterocomplexes [189]. These agonists were also found in the crystal structure of the *Mt*ClpP1/2 complex, sitting in the substrate binding channel of all 14 active sites [192].

The crystal structure of the heterocomplex from *Mycobacterium tuberculosis* was solved in presence of the agonist and an ADEP molecule, which only bound to *Mt*ClpP2 [191, 192]. *In vitro* studies with the AAA^+-ATPases ClpX and ClpC1 confirmed that in *Mt*ClpP1/2 the same asymmetrical binding of regulatory proteins takes places as observed in the *Listeria* heterocomplex. [193]. Remarkably, the asymmetry is reflected in the same alterations in the N-terminal region and the hydrophobic binding clefts of *Mt*ClpP1 that were described before for *Lm*ClpP1. Here as well, the axial pore is lined up by α-helical segments instead of a β-hairpin [194]. However, this pore is clearly enlarged in *Mt*ClpP1 from the heterocomplex, which might be a result of the binding of the ADEP molecule to *Mt*ClpP2. Thus, it looks like structural changes in one ring influence the state of the other, at least in *Mt*ClpP1/2.

None of the ClpP isoforms of *M. tuberculosis* was found to be active on its own. The explanation for this finding is provided by the crystal structure showing that between *Mt*ClpP1 and 2 specific inter-ring contacts are formed, which only favour heterocomplex formation [192]. The close functional interaction of the two proteins is revealed in the fact that both isoforms are expressed form a bicistronic operon [195]. This clearly differentiates them from *Lm*ClpP1/2, where the two *clpP* genes are located at distinct sites in the genome and no particular heterocomplex interactions between the heptameric rings were found (Figure 26c), what, in turn, allows the additional formation of homomeric, proteolytically active ClpP2 particles.

Taken together, the comparison between the ClpP heterocomplexes in *Listeria* and *Mycobacterium* is suitable to only a limited extent to draw any conclusion for understanding the role of the two different isoforms in the respective organism. More

insight might be expected from a recently published study on *Pseudomonas aeruginosa*, which also encodes two ClpP isoforms. Both are expressed differently during growth of the bacterium and seem to have specific impacts on its virulence [196]. Here again, one of the two ClpPs is able to form a tetradecamer by itself, whereas the other one stays heptameric. The infectious cycle of *Listeria monocytogenes* requires the switch between highly deviating environmental conditions. A specialised function of the heterocomplex in some of these stages, as it appears to be the case for *P. aeruginosa*, would be an elegant solution to deal with the different quantities or types of degradable proteins. Further experimental evidence, however, is required to prove such hypothesis.

8 Bibliography

1. Sontag EM, Samant RS, Frydman J (2017). Mechanisms and functions of spatial protein quality control. *Annu. Rev. Biochem.* 86, 97–122.
2. Vendruscolo M (2012). Proteome folding and aggregation. *Curr. Opin. Struct. Biol.* 22, 138–143.
3. Radzicka A, Wolfenden R (1996). Rates of uncatalyzed peptide bond hydrolysis in neutral solution and the transition state affinities of proteases. *J. Am. Chem. Soc.* 118, 6105–6109.
4. Hooper NM (2002). Proteases: a primer. *Essays Biochem.* 38, 1–8.
5. Rawlings ND, Barrett AJ, Finn R (2016). Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* 44, D343-50.
6. Drag M, Salvesen GS (2010). Emerging principles in protease-based drug discovery. *Nat. Rev. Drug Discov.* 9, 690–701.
7. Rawlings ND, Barrett AJ, Bateman A (2011). Asparagine peptide lyases: a seventh catalytic type of proteolytic enzymes. *J. Biol. Chem.* 286, 38321–38328.
8. Powers JC, Asgian JL, Ekici ÖD, James KE (2002). Irreversible inhibitors of serine, cysteine, and threonine proteases. *Chem. Rev.* 102, 4639–4750.
9. Schechter I, Berger A (1967). On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* 425, 497–502.
10. Lupas A, Flanagan JM, Tamura T, Baumeister W (1997). Self-compartmentalizing proteases. *Trends Biochem. Sci.* 22, 399–404.
11. Striebel F, Kress W, Weber-Ban E (2009). Controlled destruction: AAA+ ATPases in protein degradation from bacteria to eukaryotes. *Curr. Opin. Struct. Biol.* 19, 209–217.
12. Sauer RT, Baker TA (2011). AAA+ proteases: ATP-fueled machines of protein destruction. *Annu. Rev. Biochem.* 80, 587–612.
13. Dahlmann B (2005). Proteasomes. *Essays Biochem.* 41, 31–48.
14. Valas RE, Bourne PE (2008). Rethinking proteasome evolution: two novel bacterial proteasomes. *J. Mol. Evol.* 66, 494–504.
15. Dahmen M, Vielberg M-T, Groll M, Sieber SA (2015). Structure and mechanism of the caseinolytic protease ClpP1/2 heterocomplex from *Listeria monocytogenes*. *Angew. Chem. Int. Ed.* 54, 3598–3602.
16. Pahl A *et al.* (2015). Reversible inhibitors arrest ClpP in a defined conformational state that can be revoked by ClpX association. *Angew. Chem. Int. Ed.* 54, 15892–15896.
17. Winn MD *et al.* (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D* 67, 235–242.

18. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010). Features and development of Coot. *Acta Crystallogr. Sect. D* 66, 486–501.

19. Holm L, Rosenstrom P (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Res.* 38, W545-9.

20. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009). Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191.

21. Drozdetskiy A, Cole C, Procter J, Barton GJ (2015). JPred4: a protein secondary structure prediction server. *Nucleic acids research* 43, W389-94.

22. Chen VB *et al.* (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D* 66, 12–21.

23. Kraulis PJ (1991). MOLSCRIPT. A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24, 946–950.

24. Krissinel E, Henrick K (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372, 774–797.

25. Gasteiger E *et al.* (2005) in *The Proteomics Protocols Handbook*, ed Walker JM. (Humana Press Inc, Totowa, NJ), pp 571–608.

26. Schrödinger LL (2015) *The PyMOL Molecular Graphics System, Version 1.3*.

27. Schultz J, Milpetz F, Bork P, Ponting CP (1998). SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl. Acad. Sci. U.S.A.* 95, 5857–5864.

28. Notredame C, Higgins DG, Heringa J (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 302, 205–217.

29. Pettersen EF *et al.* (2004). UCSF Chimera - a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612.

30. (2017). UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 45, D158-D169.

31. Kabsch W (2010). XDS. *Acta Crystallogr. Sect. D* 66, 125–132.

32. Kabsch W (2010). Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. Sect. D* 66, 133–144.

33. van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J (1993). Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* 229, 105–124.

34. Mueller M, Wang M, Schulze-Briese C (2012). Optimal fine phi-slicing for single-photon-counting pixel detectors. *Acta Crystallogr. Sect. D* 68, 42–56.

35. Nurizzo D *et al.* (2006). The ID23-1 structural biology beamline at the ESRF. *J. Synchrotron. Radiat.* 13, 227–238.

36. Potterton E, Briggs P, Turkenburg M, Dodson E (2003). A graphical user interface to the CCP4 program suite. *Acta Crystallogr. Sect. D* 59, 1131–1137.

37. Evans PR, Murshudov GN (2013). How good are my data and what is the resolution? *Acta Crystallogr. Sect. D* 69, 1204–1214.

38. Tong L, Rossmann MG (1997). Rotation function calculations with GLRF program. *Meth. Enzymol.* 276, 594–611.

39. Pannu NS *et al.* (2011). Recent advances in the CRANK software suite for experimental phasing. *Acta Crystallogr. Sect. D* 67, 331–337.

40. McCoy AJ *et al.* (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658–674.

41. Murshudov GN, Vagin AA, Dodson EJ (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D* 53, 240–255.

42. Lamzin VS, Wilson KS (1993). Automated refinement of protein models. *Acta Crystallogr. Sect. D* 49, 129–147.

43. Vaguine AA, Richelle J, Wodak SJ (1999). SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr. Sect. D* 55, 191–205.

44. Dahlmann B *et al.* (1989). The multicatalytic proteinase (prosome) is ubiquitous from eukaryotes to archaebacteria. *FEBS Lett.* 251, 125–131.

45. Groll M, Clausen T (2003). Molecular shredders: how proteasomes fulfill their role. *Curr. Opin. Struct. Biol.* 13, 665–673.

46. Baumeister W *et al.* (1988). Electron microscopy and image analysis of the multicatalytic proteinase. *FEBS Lett.* 241, 239–245.

47. Harris JR (1969). Some negative contrast staining features of a protein from erythrocyte ghosts. *J. Mol. Biol.* 46, 329–335.

48. Löwe J *et al.* (1995). Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science* 268, 533–539.

49. Pühler G *et al.* (1992). Subunit stoichiometry and three-dimensional arrangement in proteasomes from *Thermoplasma acidophilum*. *EMBO J.* 11, 1607–1616.

50. Grziwa A, Baumeister W, Dahlmann B, Kopp F (1991). Localization of subunits in proteasomes from *Thermoplasma acidophilum* by immunoelectron microscopy. *FEBS Lett.* 290, 186–190.

51. Zwickl P *et al.* (1992). Primary structure of the *Thermoplasma* proteasome and its implications for the structure, function, and evolution of the multicatalytic proteinase. *Biochemistry* 31, 964–972.

52. Heinemeyer W, Kleinschmidt JA, Saidowsky J, Escher C, Wolf DH (1991). Proteinase yscE, the yeast proteasome/multicatalytic-multifunctional proteinase: mutants unravel its function in stress induced proteolysis and uncover its necessity for cell survival. *EMBO J.* 10, 555–562.

53. Groll M *et al.* (1997). Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* 386, 463–471.

54. Dahlmann B, Kuehn L, Grziwa A, Zwickl P, Baumeister W (1992). Biochemical properties of the proteasome from *Thermoplasma acidophilum*. *Eur. J. Biochem.* 208, 789–797.

55. Kaczowka SJ, Maupin-Furlow JA (2003). Subunit topology of two 20S proteasomes from *Haloferax volcanii*. *J. Bacteriol.* 185, 165–174.

56. Kocabiyik S, Ozdemir I, Zwickl P, Ozdoğan S (2010). Molecular cloning and co-expression of *Thermoplasma volcanium* proteasome subunit genes. *Protein Expr. Purif.* 73, 223–230.

57. Madding LS *et al.* (2007). Role of the beta1 subunit in the function and stability of the 20S proteasome in the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Bacteriol.* 189, 583–590.

58. Brannigan JA *et al.* (1995). A protein catalytic framework with an N-terminal nucleophile is capable of self-activation. *Nature* 378, 416–419.

59. Hughes AL (1997). Evolution of the proteasome components. *Immunogenetics* 46, 82–92.

60. Köhler A *et al.* (2001). The substrate translocation channel of the proteasome. *Biochimie* 83, 325–332.

61. Groll M *et al.* (2000). A gated channel into the proteasome core particle. *Nat. Struct. Mol. Biol.* 7, 1062–1067.

62. Groll M, Brandstetter H, Bartunik H, Bourenkow G, Huber R (2003). Investigations on the maturation and regulation of archaebacterial proteasomes. *J. Mol. Biol.* 327, 75–83.

63. Benaroudj N, Zwickl P, Seemüller E, Baumeister W, Goldberg AL (2003). ATP hydrolysis by the proteasome regulatory complex PAN serves multiple functions in protein degradation. *Mol. Cell* 11, 69–78.

64. Religa TL, Sprangers R, Kay LE (2010). Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR. *Science* 328, 98–102.

65. Seemüller E *et al.* (1995). Proteasome from *Thermoplasma acidophilum*: a threonine protease. *Science* 268, 579–582.

66. Arendt CS, Hochstrasser M (1997). Identification of the yeast 20S proteasome catalytic centers and subunit interactions required for active-site formation. *Proc. Natl. Acad. Sci. U.S.A.* 94, 7156–7161.

67. Harris JL, Alper PB, Li J, Rechsteiner M, Backes BJ (2001). Substrate specificity of the human proteasome. *Chem. Biol.* 8, 1131–1141.

68. Kniepert A, Groettrup M (2014). The unique functions of tissue-specific proteasomes. *Trends Biochem. Sci.* 39, 17–24.

69. Orlowski M, Wilk S (2000). Catalytic activities of the 20S proteasome, a multicatalytic proteinase complex. *Arch. Biochem. Biophys.* 383, 1–16.

70. Kunjappu MJ, Hochstrasser M (2014). Assembly of the 20S proteasome. *Biochim. Biophys. Acta* 1843, 2–12.

71. Jäger S, Groll M, Huber R, Wolf DH, Heinemeyer W (1999). Proteasome beta-type subunits: unequal roles of propeptides in core particle maturation and a hierarchy of active site function. *J. Mol. Biol.* 291, 997–1013.

72. Maupin-Furlow JA (2013). Archaeal proteasomes and sampylation. *Subcell. Biochem.* 66, 297–327.

73. Chen P, Hochstrasser M (1995). Biogenesis, structure and function of the yeast 20S proteasome. *EMBO J.* 14, 2620–2630.

74. Schmidtke G *et al.* (1996). Analysis of mammalian 20S proteasome biogenesis: the maturation of beta-subunits is an ordered two-step mechanism involving autocatalysis. *EMBO J.* 15, 6887–6898.

75. Ditzel L *et al.* (1998). Conformational constraints for protein self-cleavage in the proteasome. *J. Mol. Biol.* 279, 1187–1191.

76. Huber EM *et al.* (2016). A unified mechanism for proteolysis and autocatalytic activation in the 20S proteasome. *Nat. Commun.* 7, 10900.

77. Gallastegui N, Groll M (2010). The 26S proteasome: assembly and function of a destructive machine. *Trends Biochem. Sci.* 35, 634–642.

78. Gerards WL *et al.* (1997). The human alpha-type proteasomal subunit HsC8 forms a double ringlike structure, but does not assemble into proteasome-like particles with the beta-type subunits HsDelta or HsBPROS26. *J. Biol. Chem.* 272, 10080–10086.

79. Groll M, Bochtler M, Brandstetter H, Clausen T, Huber R (2005). Molecular machines for protein degradation. *ChemBioChem* 6, 222–256.

80. Murata S, Yashiroda H, Tanaka K (2009). Molecular mechanisms of proteasome assembly. *Nat. Rev. Mol. Cell Biol.* 10, 104–115.

81. Zwickl P, Lottspeich F, Baumeister W (1992). Expression of functional *Thermoplasma acidophilum* proteasomes in *Escherichia coli*. *FEBS Lett.* 312, 157–160.

82. Zwickl P, Kleinz J, Baumeister W (1994). Critical elements in proteasome assembly. *Nat. Struct. Mol. Biol.* 1, 765–770.

83. Glickman MH *et al.* (1998). A subcomplex of the proteasome regulatory particle required for ubiquitin-conjugate degradation and related to the COP9-signalosome and eIF3. *Cell* 94, 615–623.

84. Bar-Nun S, Glickman MH (2012). Proteasomal AAA-ATPases: structure and function. *Biochim. Biophys. Acta* 1823, 67–82.

85. Ciechanover A, Stanhill A (2014). The complexity of recognition of ubiquitinated substrates by the 26S proteasome. *Biochim. Biophys. Acta* 1843, 86–96.

86. Zwickl P, Ng D, Woo KM, Klenk HP, Goldberg AL (1999). An archaebacterial ATPase, homologous to ATPases in the eukaryotic 26S proteasome, activates protein breakdown by 20S proteasomes. *J. Biol. Chem.* 274, 26008–26014.

87. Rabl J *et al.* (2008). Mechanism of gate opening in the 20S proteasome by the proteasomal ATPases. *Mol. Cell* 30, 360–368.

88. Barthelme D, Sauer RT (2012). Identification of the Cdc48•20S proteasome as an ancient AAA+ proteolytic machine. *Science* 337, 843–846.

89. Forouzan D *et al.* (2012). The archaeal proteasome is regulated by a network of AAA ATPases. *J. Biol. Chem.* 287, 39254–39262.

90. Fu X *et al.* (2016). Ubiquitin-like proteasome system represents a eukaryotic-like pathway for targeted proteolysis in archaea. *MBio* 7, e00379-16.

91. Pouch MN, Cournoyer B, Baumeister W (2000). Characterization of the 20S proteasome from the actinomycete *Frankia*. *Mol. Microbiol.* 35, 368–377.

92. Nagy I, Tamura T, Vanderleyden J, Baumeister W, Mot R de (1998). The 20S proteasome of *Streptomyces coelicolor*. *J. Bacteriol.* 180, 5448–5453.

93. Tamura T *et al.* (1995). The first characterization of a eubacterial proteasome: the 20S complex of *Rhodococcus*. *Curr. Biol.* 5, 766–774.

94. Lupas A, Zwickl P, Baumeister W (1994). Proteasome sequences in eubacteria. *Trends Biochem. Sci.* 19, 533–534.

95. Jastrab JB, Darwin KH (2015). Bacterial Proteasomes. *Annu. Rev. Microbiol.* 69, 109–127.

96. Knipfer N, Shrader TE (1997). Inactivation of the 20S proteasome in *Mycobacterium smegmatis*. *Mol. Microbiol.* 25, 375–383.

97. Darwin KH, Ehrt S, Gutierrez-Ramos J-C, Weich N, Nathan CF (2003). The proteasome of *Mycobacterium tuberculosis* is required for resistance to nitric oxide. *Science* 302, 1963–1966.

98. Lupas A *et al.* (1997). Eubacterial proteasomes. *Mol. Biol. Rep.* 24, 125–131.

99. Zühl F *et al.* (1997). Subunit topology of the *Rhodococcus* proteasome. *FEBS Lett.* 400, 83–90.

100. Zühl F, Seemüller E, Golbik R, Baumeister W (1997). Dissecting the assembly pathway of the 20S proteasome. *FEBS Lett.* 418, 189–194.

101. Lin G *et al.* (2006). *Mycobacterium tuberculosis prcBA* genes encode a gated proteasome with broad oligopeptide specificity. *Mol. Microbiol.* 59, 1405–1416.

102. Kwon YD, Nagy I, Adams PD, Baumeister W, Jap BK (2004). Crystal structures of the *Rhodococcus* proteasome with and without its pro-peptides: implications for the role of the pro-peptide in proteasome assembly. *J. Mol. Biol.* 335, 233–245.

103. Li D *et al.* (2010). Structural basis for the assembly and gate closure mechanisms of the *Mycobacterium tuberculosis* 20S proteasome. *EMBO J.* 29, 2037–2047.

104. Darwin KH, Lin G, Chen Z, Li H, Nathan CF (2005). Characterization of a *Mycobacterium tuberculosis* proteasomal ATPase homologue. *Mol. Microbiol.* 55, 561–571.

105. Wolf S *et al.* (1998). Characterization of ARC, a divergent member of the AAA ATPase family from *Rhodococcus erythropolis*. *J. Mol. Biol.* 277, 13–25.

106. Wu Y *et al.* (2017). *Mycobacterium tuberculosis* proteasomal ATPase Mpa has a β-grasp domain that hinders docking with the proteasome core protease. *Mol. Microbiol.* 105, 227–241.

107. Pearce MJ, Mintseris J, Ferreyra J, Gygi SP, Darwin KH (2008). Ubiquitin-like protein involved in the proteasome pathway of *Mycobacterium tuberculosis*. *Science* 322, 1104–1107.

108. Chuang SE, Blattner FR (1993). Characterization of twenty-six new heat shock genes of *Escherichia coli*. *J. Bacteriol.* 175, 5242–5252.

109. Chuang SE, Burland V, Plunkett G3, Daniels DL, Blattner FR (1993). Sequence analysis of four new heat-shock genes constituting the hslTS/ibpAB and hslVU operons in *Escherichia coli*. *Gene* 134, 1–6.

110. Rohrwild M *et al.* (1996). HslV-HslU: A novel ATP-dependent protease complex in *Escherichia coli* related to the eukaryotic proteasome. *Proc. Natl. Acad. Sci. U.S.A.* 93, 5808–5813.

111. Kessel M *et al.* (1996). Six-fold rotational symmetry of ClpQ, the *E. coli* homolog of the 20S proteasome, and its ATP-dependent activator, ClpY. *FEBS Lett.* 398, 274–278.

112. Bochtler M, Ditzel L, Groll M, Huber R (1997). Crystal structure of heat shock locus V (HslV) from *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 94, 6070–6074.

113. Rohrwild M *et al.* (1997). The ATP-dependent HslVU protease from *Escherichia coli* is a four-ring structure resembling the proteasome. *Nat. Struct. Mol. Biol.* 4, 133–139.

114. Yoo SJ *et al.* (1997). Mutagenesis of two N-terminal Thr and five Ser residues in HslV, the proteolytic component of the ATP-dependent HslVU protease. *FEBS Lett.* 412, 57–60.

115. Missiakas D, Schwager F, Betton JM, Georgopoulos C, Raina S (1996). Identification and characterization of HsIV HsIU (ClpQ ClpY) proteins involved in overall proteolysis of misfolded proteins in *Escherichia coli*. *EMBO J.* 15, 6899–6909.

116. Yoo SJ *et al.* (1996). Purification and characterization of the heat shock proteins HslV and HslU that form a new ATP-dependent protease in *Escherichia coli*. *J. Biol. Chem.* 271, 14035–14040.

117. Seol JH *et al.* (1997). The heat-shock protein HslVU from *Escherichia coli* is a protein-activated ATPase as well as an ATP-dependent proteinase. *Eur. J. Biochem.* 247, 1143–1150.

118. Iyer LM, Leipe DD, Koonin EV, Aravind L (2004). Evolutionary history and higher order classification of AAA+ ATPases. *J. Struct. Biol.* 146, 11–31.

119. Sousa MC, McKay DB (2001). Structure of *Haemophilus influenzae* HslV protein at 1.9 Å resolution, revealing a cation-binding site near the catalytic site. *Acta Crystallogr. Sect. D* 57, 1950–1954.

120. Sousa MC, Kessler BM, Overkleeft HS, McKay DB (2002). Crystal structure of HslUV complexed with a vinyl sulfone inhibitor: corroboration of a proposed mechanism of allosteric activation of HslV by HslU. *J. Mol. Biol.* 318, 779–785.

121. Sousa MC *et al.* (2000). Crystal and solution structures of an HslUV protease-chaperone complex. *Cell* 103, 633–643.

122. Gille C *et al.* (2003). A comprehensive view on proteasomal sequences: implications for the evolution of the proteasome. *J. Mol. Biol.* 326, 1437–1448.

123. Golubic S, Seong-Joo L (1999). Early cyanobacterial fossil record: preservation, palaeoenvironments and identification. *Eur. J. Phycol.* 34, 339–348.

124. Slabinski L *et al.* (2007). XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics* 23, 3403–3405.

125. Bright SW, Lea PJ, Miflin BJ (1979). The regulation of methionine biosynthesis and metabolism in plants and bacteria. *Ciba Foundation symposium*, 101–117.

126. Chen P, Hochstrasser M (1996). Autocatalytic subunit processing couples active site formation in the 20S proteasome to completion of assembly. *Cell* 86, 961–972.

127. Witt S *et al.* (2006). Proteasome assembly triggers a switch required for active-site maturation. *Structure* 14, 1179–1188.

128. Fuchs ACD *et al.* (2017). The architecture of the Anbu complex reflects an evolutionary intermediate at the origin of the proteasome system. *Structure* 25, 834–845.

129. Huber EM, Groll M (2012). Inhibitors for the immuno- and constitutive proteasome: current and future trends in drug development. *Angew. Chem. Int. Ed.* 51, 8708–8720.

130. Guo F, Maurizi MR, Esser L, Di Xia (2002). Crystal structure of ClpA, an Hsp100 chaperone and regulator of ClpAP protease. *J. Biol. Chem.* 277, 46743–46752.

131. Kim DY, Kim KK (2003). Crystal structure of ClpX molecular chaperone from *Helicobacter pylori*. *J. Biol. Chem.* 278, 50664–50670.

132. Hu G *et al.* (2006). Structure of the *Mycobacterium tuberculosis* proteasome and mechanism of inhibition by a peptidyl boronate. *Mol. Microbiol.* 59, 1417–1428.

133. Hervás AB, Canosa I, Santero E (2008). Transcriptome analysis of *Pseudomonas putida* in response to nitrogen availability. *J. Bacteriol.* 190, 416–420.

134. Iyer LM, Abhiman S, Maxwell Burroughs A, Aravind L (2009). Amidoligases with ATP-grasp, glutamine synthetase-like and acetyltransferase-like domains: synthesis of novel metabolites and peptide modifications of proteins. *Mol. Biosyst.* 5, 1636–1660.

135. Wright GD (2007). The antibiotic resistome: the nexus of chemical and genetic diversity. *Nat. Rev. Microbiol.* 5, 175–186.

136. Fleming A (1929). On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of *B. influenzae*. *Br. J. Exp. Pathol.* 10, 226–236.

137. D'Costa VM *et al.* (2011). Antibiotic resistance is ancient. *Nature* 477, 457–461.

138. Sugarman B, Pesanti E (1980). Treatment failures secondary to in vivo development of drug resistance by microorganisms. *Rev. Infect. Dis.* 2, 153–168.

139. McGowan JE (2006). Resistance in nonfermenting gram-negative bacteria: multidrug resistance to the maximum. *Am. J. Med.* 119, S29-36.

140. Frees D, Qazi SNA, Hill PJ, Ingmer H (2003). Alternative roles of ClpX and ClpP in *Staphylococcus aureus* stress tolerance and virulence. *Mol. Microbiol.* 48, 1565–1578.

141. Frees D, Sorensen K, Ingmer H (2005). Global virulence regulation in *Staphylococcus aureus*: pinpointing the roles of ClpP and ClpX in the sar/agr regulatory network. *Infect. Immun.* 73, 8100–8108.

142. Fernandez L, Breidenstein EBM, Song D, Hancock REW (2012). Role of intracellular proteases in the antibiotic resistance, motility, and biofilm formation of *Pseudomonas aeruginosa*. *Antimicrob. Agents Chemother.* 56, 1128–1132.

143. Knudsen GM *et al.* (2013). ClpP deletion causes attenuation of *Salmonella typhimurium* virulence through mis-regulation of RpoS and indirect control of CsrA and the SPI genes. *Microbiology* 159, 1497–1509.

144. Katayama-Fujimura Y, Gottesman S, Maurizi MR (1987). A multiple-component, ATP-dependent protease from *Escherichia coli*. *J. Biol. Chem.* 262, 4477–4485.

145. Kessel M *et al.* (1995). Homology in structural organization between *E. coli* ClpAP protease and the eukaryotic 26 S proteasome. *J. Mol. Biol.* 250, 587–594.

146. Flanagan JM, Wall JS, Capel MS, Schneider DK, Shanklin J (1995). Scanning transmission electron microscopy and small-angle scattering provide evidence that native *Escherichia coli* ClpP is a tetradecamer with an axial pore. *Biochemistry* 34, 10910–10917.

147. Shin DH, Lee CS, Chung CH, Suh SW (1996). Molecular symmetry of the ClpP component of the ATP-dependent Clp protease, an *Escherichia coli* homolog of 20S proteasome. *J. Mol. Biol.* 262, 71–76.

148. Ortega J, Lee HS, Maurizi MR, Steven AC (2004). ClpA and ClpX ATPases bind simultaneously to opposite ends of ClpP peptidase to form active hybrid complexes. *J. Struct. Biol.* 146, 217–226.

149. Wang J, Hartling JA, Flanagan JM (1997). The structure of ClpP at 2.3 Å resolution suggests a model for ATP-dependent proteolysis. *Cell* 91, 447–456.

150. Maurizi MR *et al.* (1990). Sequence and structure of Clp P, the proteolytic component of the ATP-dependent Clp protease of *Escherichia coli*. *J. Biol. Chem.* 265, 12536–12545.

151. Sprangers R, Gribun A, Hwang PM, Houry WA, Kay LE (2005). Quantitative NMR spectroscopy of supramolecular complexes: dynamic side pores in ClpP are important for product release. *Proc. Natl. Acad. Sci. U.S.A.* 102, 16678–16683.

152. Katayama Y *et al.* (1988). The two-component, ATP-dependent Clp protease of *Escherichia coli*. Purification, cloning, and mutational analysis of the ATP-binding component. *J. Biol. Chem.* 263, 15226–15236.

153. Alexopoulos J *et al.* (2013). Structural determinants stabilizing the axial channel of ClpP for substrate translocation. *Mol. Microbiol.* 90, 167–180.

154. Kirstein J *et al.* (2006). Adaptor protein controlled oligomerization activates the AAA+ protein ClpC. *EMBO J.* 25, 1481–1491.

155. Bewley MC, Graziano V, Griffin K, Flanagan JM (2006). The asymmetry in the mature amino-terminus of ClpP facilitates a local symmetry match in ClpAP and ClpXP complexes. *J. Struct. Biol.* 153, 113–128.

156. Sen M *et al.* (2013). The ClpXP protease unfolds substrates using a constant rate of pulling but different gears. *Cell* 155, 636–646.

157. Effantin G, Maurizi MR, Steven AC (2010). Binding of the ClpA unfoldase opens the axial gate of ClpP peptidase. *J. Biol. Chem.* 285, 14834–14840.

158. Baker TA, Sauer RT (2012). ClpXP, an ATP-powered unfolding and protein-degradation machine. *Biochim. Biophys. Acta* 1823, 15–28.

159. Gribun A *et al.* (2005). The ClpP double ring tetradecameric protease exhibits plastic ring-ring interactions, and the N termini of its subunits form flexible loops that are essential for ClpXP and ClpAP complex formation. *J. Biol. Chem.* 280, 16185–16196.

160. Lee ME, Baker TA, Sauer RT (2010). Control of substrate gating and translocation into ClpP by channel residues and ClpX binding. *J. Mol. Biol.* 399, 707–718.

161. Geiger SR, Böttcher T, Sieber SA, Cramer P (2011). A conformational switch underlies ClpP protease function. *Angew. Chem. Int. Ed.* 50, 5749–5752.

162. Ye F *et al.* (2013). Helix unfolding/refolding characterizes the functional dynamics of *Staphylococcus aureus* Clp protease. *J. Biol. Chem.* 288, 17643–17653.

163. Gersch M, List A, Groll M, Sieber SA (2012). Insights into structural network responsible for oligomerization and activity of bacterial virulence regulator caseinolytic protease P (ClpP) protein. *J. Biol. Chem.* 287, 9484–9494.

164. Hwang BJ, Woo KM, Goldberg AL, Chung CH (1988). Protease Ti, a new ATP-dependent protease in *Escherichia coli*, contains protein-activated ATPase and proteolytic functions in distinct subunits. *J. Biol. Chem.* 263, 8727–8734.

165. Ekici ÖD, Paetzel M, Dalbey RE (2008). Unconventional serine proteases: variations on the catalytic Ser/His/Asp triad configuration. *Protein Sci.* 17, 2023–2037.

166. Maurizi MR, Clark WP, Kim SH, Gottesman S (1990). Clp P represents a unique family of serine proteases. *J. Biol. Chem.* 265, 12546–12552.

167. Stahl M, Sieber SA (2017). An amino acid domino effect orchestrates ClpP's conformational states. *Current opinion in chemical biology* 40, 102–110.

168. Hernandez-Milian A, Payeras-Cifre A (2014). What is new in listeriosis? *BioMed Res. Int.* 2014, 358051.

169. Allerberger F, Wagner M (2010). Listeriosis: a resurgent foodborne infection. *Clin. Microbiol. Infect.* 16, 16–23.

170. Hamon MA, Ribet D, Stavru F, Cossart P (2012). Listeriolysin O: the Swiss army knife of *Listeria*. *Trends Microbiol.* 20, 360–368.

171. Portnoy DA, Auerbuch V, Glomski IJ (2002). The cell biology of *Listeria monocytogenes* infection: the intersection of bacterial pathogenesis and cell-mediated immunity. *J. Cell Biol.* 158, 409–414.

172. Tilney LG, Portnoy DA (1989). Actin filaments and the growth, movement, and spread of the intracellular bacterial parasite, *Listeria monocytogenes*. *J. Cell Biol.* 109, 1597–1608.

173. Derre I, Rapoport G, Msadek T (1999). CtsR, a novel regulator of stress and heat shock response, controls clp and molecular chaperone gene expression in gram-positive bacteria. *Mol. Microbiol.* 31, 117–131.

174. Nair S, Derre I, Msadek T, Gaillot O, Berche P (2000). CtsR controls class III heat shock gene expression in the human pathogen *Listeria monocytogenes*. *Mol. Microbiol.* 35, 800–811.

175. Gaillot O, Pellegrini E, Bregenholt S, Nair S, Berche P (2000). The ClpP serine protease is essential for the intracellular parasitism and virulence of *Listeria monocytogenes*. *Mol. Microbiol.* 35, 1286–1294.

176. Gaillot O, Bregenholt S, Jaubert F, Di Santo JP, Berche P (2001). Stress-induced ClpP serine protease of *Listeria monocytogenes* is essential for induction of listeriolysin O-dependent protective immunity. *Infect. Immun.* 69, 4938–4943.

177. Camejo A *et al.* (2011). The arsenal of virulence factors deployed by *Listeria monocytogenes* to promote its cell infection cycle. *Virulence* 2, 379–394.

178. Travier L, Lecuit M (2014). *Listeria monocytogenes* ActA: a new function for a 'classic' virulence factor. *Curr. Opin. Microbiol.* 17, 53–60.

179. Ollinger J, O'Malley T, Kesicki EA, Odingo J, Parish T (2012). Validation of the essential ClpP protease in *Mycobacterium tuberculosis* as a novel drug target. *J. Bacteriol.* 194, 663–668.

180. Fedhila S, Msadek T, Nel P, Lereclus D (2002). Distinct clpP genes control specific adaptive responses in *Bacillus thuringiensis*. *J. Bacteriol.* 184, 5554–5562.

181. Zeiler E *et al.* (2011). Vibralactone as a tool to study the activity and structure of the ClpP1P2 complex from *Listeria monocytogenes*. *Angew. Chem. Int. Ed.* 50, 11001–11004.

182. Zeiler E *et al.* (2013). Structural and functional insights into caseinolytic proteases reveal an unprecedented regulation principle of their catalytic triad. *Proc. Natl. Acad. Sci. U.S.A.* 110, 11302–11307.

183. Szyk A, Maurizi MR (2006). Crystal structure at 1.9 Å of *E. coli* ClpP with a peptide covalently bound at the active site. *J. Struct. Biol.* 156, 165–174.

184. Urban S, Lee JR, Freeman M (2001). *Drosophila* rhomboid-1 defines a family of putative intramembrane serine proteases. *Cell* 107, 173–182.

185. Maegawa S, Ito K, Akiyama Y (2005). Proteolytic action of GlpG, a rhomboid protease in the *Escherichia coli* cytoplasmic membrane. *Biochemistry* 44, 13543–13552.

186. Balogh D *et al.* (2017). Insights into ClpXP proteolysis: heterooligomerization and partial deactivation enhance chaperone affinity and substrate turnover in *Listeria monocytogenes*. *Chem Sci.* 8, 1592–1600.

187. Glynn SE, Martin A, Nager AR, Baker TA, Sauer RT (2009). Structures of asymmetric ClpX hexamers reveal nucleotide-dependent motions in a AAA+ protein-unfolding machine. *Cell* 139, 744–756.

188. Andersson FI *et al.* (2009). Structure and function of a novel type of ATP-dependent Clp protease. *J. Biol. Chem.* 284, 13519–13532.

189. Akopian T *et al.* (2012). The active ClpP protease from *M. tuberculosis* is a complex composed of a heptameric ClpP1 and a ClpP2 ring. *EMBO J.* 31, 1529–1541.

190. Schmitz KR, Sauer RT (2014). Substrate delivery by the AAA+ ClpX and ClpC1 unfoldases activates the mycobacterial ClpP1P2 peptidase. *Mol. Microbiol.* 93, 617–628.

191. Li M *et al.* (2016). Structure and functional properties of the active form of the proteolytic complex, ClpP1P2, from *Mycobacterium tuberculosis*. *J. Biol. Chem.* 291, 7465–7476.

192. Schmitz KR, Carney DW, Sello JK, Sauer RT (2014). Crystal structure of *Mycobacterium tuberculosis* ClpP1P2 suggests a model for peptidase activation by AAA+ partner binding and substrate delivery. *Proc. Natl. Acad. Sci. U.S.A.* 111, E4587-95.

193. Leodolter J, Warweg J, Weber-Ban E (2015). The *Mycobacterium tuberculosis* ClpP1P2 protease interacts asymmetrically with its ATPase partners ClpX and ClpC1. *PLOS ONE* 10, e0125345.

194. Ingvarsson H *et al.* (2007). Insights into the inter-ring plasticity of caseinolytic proteases from the X-ray structure of *Mycobacterium tuberculosis* ClpP1. *Acta Crystallogr. Sect. D* 63, 249–259.

195. Raju RM *et al.* (2012). *Mycobacterium tuberculosis* ClpP1 and ClpP2 function together in protein degradation and are required for viability in vitro and during infection. *PLOS Pathog.* 8, e1002511.

196. Hall BM *et al.* (2016). Two isoforms of Clp peptidase in *Pseudomonas aeruginosa* control distinct aspects of cellular physiology. *J. Bacteriol.* 199, e00568-16.

9 Abbreviations

aa	amino acid
AAA^+	ATPases associated with a variety of cellular activities
Ac	acetic acid
Anbu	ancestral β-subunit protein
AMC	7-amino-4-methylcoumarin
Amp	ampicillin
APS	ammonium persulfate
ATP	adenosine triphosphate
AU	asymmetric unit
Å	Ångström
Boc-	*tert*-butyloxycarbonyl-
bp	base pair
BRZ	bortezomib
Clp	caseinolytic proteinase
CMK	chlorormethylketone
CP	20S proteasome core particle
CV	column volume
(k)Da	(kilo)Dalton(s)
ddH_2O	double distilled water
DMSO	dimethylsulfoxid
DNA	deoxyribonucleic acid
dNTP	deoxyriboncleotide triphosphate
DTT	dithiothreitol
Ec	*Escherichia coli*
EDTA	ethylendiaminetetraacetic acidl
ESRF	European synchrotron radiation facility
EtOH	ethanol
g	g-force
FPLC	fast protein liquid chromatography

HCl	hydrochloric acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPSF	high purity salt free
HPLC	high performance liquid chromatography
HslV	heat shock locus V
Hy	*Hyphomicrobium sp.* strain MC1
ID	identification
IPTG	isopropyl-β-D-thiogalactopyranoside
K	Kelvin
Kan	kanamycin
kb	kilo base pairs
lac	*lactose*
LB	lysogeny broth
Lm	*Listeria monocytogenes*
mA	milliampere
MCS	multiple cloning site
MPD	2-methyl-2,4-pentaendiol
Mt	*Mycobacterium tuberculosis*
mut	mutant
MWCO	molecular weight cut off
NaN_3	sodium azide
NaOH	sodium hydroxide
NCS	non-crystallographic symmetry
Ntn	N-terminal nucleophile
OD	optical density
Pa	*Pseudomonas aeruginosa*
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEG	polyethylene glycol

PVP K15	polyvinylpyrrolidone K15
R_{free}	free R-factor
r.m.s.d.	root mean square deviation
rpm	rounds per minute
R_{work}	crystallographic R-factor
S	Svedberg
SAD	single wave-length anomalous dispersion
SDS	sodium dodecyl sulphate
SEC	size exclusion chromatography
SeMet	L-selenomethionine
SLS	Swiss Light Source
SI	supplementary information
SOC	super optimal broth with catabolite repression
Suc-	succinyl-
TAE	tris-acetate-EDTA
TEMED	*N, N, N', N'*-tetramethylethylendiamine
TLS	translation, libration, screw
Tris	tris(hydroxylmethyl-)aminomethane
UV	ultraviolet
V	volt
Yb	*Yersinia bercovieri*
Z-	carboxybenzyl-
v/v	volume per volume
wt	wild type
w/v	weight per volume
°	degree
°C	degree Celsius
%	percent

10 Publications

Parts of this thesis will be or have been published in peer-reviewed scientific journals:

Vielberg, M.-T., Bauer V.C., Groll, M., On the trails of the proteasome fold: Structural and functional analysis of the ancestral β-subunit protein Anbu., *submitted for publication*[7]

Piasecka, A., Czapińska, H., Vielberg, M.-T., Szczepanowski, R., Kiefersauer, R., Reed, S., Groll, M., Bochtler, M., The *Y. bercovieri* Anbu crystal structure sheds light on the evolution of highly (pseudo)symmetric multimers., *submitted for publication*[7]

Dahmen, M.*, Vielberg, M.-T.*, Groll, M., Sieber, S.A. (2015), Structure and mechanism of the caseinolytic protease ClpP1/2 heterocomplex from *Listeria monocytogenes*., *Angew. Chem. Int. Ed.*, 54, 3598-602.

Dahmen, M.*, Vielberg, M.-T.*, Groll, M., Sieber, S.A. (2015), Struktur und Mechanismus des Heterokomplexes der caseinolytischen Protease ClpP1/2 aus *Listeria monocytogenes., Angew. Chem.*, 127, 3668-73.

Publications not highlighted in this work:

Pahl, A.*, Lakemeyer, M.*, Vielberg, M.-T.*, Hackl, M.W., Vomacka, J., Korotkov, V.S., Stein, M.L., Fetzer, C., Lorenz-Baath, K., Richter, K. Waldmann, H., Groll, M., Sieber S.A. (2015), Reversible inhibitors arrest ClpP in a defined conformational state that can be revoked by ClpX association., *Angew. Chem. Int. Ed.*, 54, 15892-6.

Pahl, A.*, Lakemeyer, M.*, Vielberg, M.-T.*, Hackl, M.W., Vomacka, J., Korotkov, V.S., Stein, M.L., Fetzer, C., Lorenz-Baath, K., Richter, K. Waldmann, H., Groll, M., Sieber S.A.

[7] *Note added only for publication of this thesis:* Shortly after this thesis had been defended, both articles on Anbu were accepted for publication by the Journal of Molecular Biology. Their full citation information is now as follows: i) Vielberg, M.-T., Bauer, V.C., Groll, M. (2018), On the trails of the proteasome fold: structural and functional analysis of the ancestral β-subunit protein Anbu., *J. Mol. Biol.*, 430, 628-40 (doi: 10.1016/j.jmb.2018.01.004.) and ii) Piasecka, A., Czapińska, H., Vielberg, M.-T., Szczepanowski, R., Kiefersauer, R., Reed, S., Groll, M., Bochtler, M. (2018), The *Y. bercovieri* Anbu crystal structure sheds light on the evolution of highly (pseudo)symmetric multimers., *J. Mol. Biol.*, 430, 611-27 (doi: 10.1016/j.jmb.2017.11.016.). The respective figures and parts in chapters 6.2 and 6.3 are now licensed under the © 2018 CC-BY-NC-ND 4.0 license.

(2015), Reversible Inhibitoren arretieren ClpP in einer definierten Konformation, die durch Bindung von ClpX aufgehoben wird., *Angew. Chem.*, 127, 16121–6.

Annamalai, K., Liberta, F., Vielberg, M.-T., Close, W., Lilie, H., Gührs, K.-H., Schierhorn, A., Koehler, R., Schmidt, A., Haupt, C., Hegenbart, U., Schönland, S., Schmidt, M., Groll, M., Fändrich, M. (2017), Common fibril structures imply systematically conserved protein misfolding pathways in vivo., *Angew. Chem. Int. Ed.*, 56, 7510-14.

* These authors contributed equally to this work.

11 Acknowledgement

First and foremost, I would like to thank my supervisor, Prof. Michael Groll, for granting me the opportunity of writing my PhD thesis at the chair of biochemistry and pursuing my interests in structural biology. His enthusiasm for X-ray crystallography and structure elucidation of proteins inspired me and I am especially grateful for the variety of projects that allowed me to acquire the knowledge and skills of solving my 'own' crystal structures. His advice and ideas were of valuable support to my work.

My special thanks go to Astrid König, Ute Kashoa, Kathrin Gärtner, Christoph Graßberger, Richard Feicht and Winfried Meining for dealing with all the big and small challenges apart from the actual research and providing so best working conditions.

Most of my projects were done in close interaction with other research groups, so I would like to thank my collaborators for the successful cooperations, above all Dr. Ania Piasecka, Dr. Honorata Czapińska and Prof. Dr. Matthias Bochtler from the International Institute of Molecular and Cell Biology in Warsaw, Dr. Maria Dahmen, Markus Lakemeyer and Prof. Dr. Stephan Sieber from the Technische Universität München and Dr. Karthikeyan Annamalai and Prof. Dr. Marcus Fändrich from the Universität Ulm.

Furthermore, I would like to acknowledge all my students as well as our apprentices for supporting me in the lab, especially Verena Bauer, who did the first experiments for characterization of *Hy*Anbu.

I am very grateful to all members of the Groll lab, in particular my PhD and post-doc colleagues for the good time we had together.

Finally, I would like to thank Michael Sammon and my family for their constant love and support.

12 Declaration

I, Marie-Theres Vielberg, hereby declare that I independently prepared the present thesis, using only the references and resources stated. This work has not been submitted to any examination board yet. Parts of this work have been or will be published in scientific journals.

Garching, November 2017