

# Structure-based mechanistic analysis of the proteasome

DISSERTATION

for the award of the degree

“Doctor rerum naturalium” (Dr. rer. nat.)

of the GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

within the doctoral program

BIOMOLECULES: STRUCTURE – FUNCTION – DYNAMICS

of the Göttingen Graduate School for Neurosciences,

Biophysics, and Molecular Biosciences (GGNB)

submitted by

*Fabian Henneberg*

from München, Germany

Göttingen 2018



## **Thesis Committee**

Prof. Dr. Holger Stark  
Structural Dynamics, Max-Planck-Institute for Biophysical Chemistry, Göttingen

Prof. Dr. Kai Tittmann  
Molecular Enzymology, Albrecht-von-Haller-Institute, Göttingen

Prof. Dr. Ralf Ficner  
Molecular Structural Biology, Institute for Microbiology and Genetics, Göttingen

## **Members of the Examination Board**

1<sup>st</sup> Referee: Prof. Dr. Holger Stark  
Structural Dynamics, Max-Planck-Institute for Biophysical Chemistry, Göttingen

2<sup>nd</sup> Referee: Prof. Dr. Kai Tittmann  
Molecular Enzymology, Albrecht-von-Haller-Institute, Göttingen

Prof. Dr. Ralf Ficner  
Molecular Structural Biology, Institute for Microbiology and Genetics, Göttingen

Prof. Dr. Henning Urlaub  
Bioanalytical Mass Spectrometry, Max-Planck-Institute for Biophysical Chemistry, Göttingen

Prof. Dr. Markus Zweckstetter  
Structural Biology in Dementia, German Center for Neurodegenerative Diseases, Göttingen

Dr. Juliane Liepe  
Quantitative and Systems Biology, Max-Planck-Institute for Biophysical Chemistry, Göttingen

Date of oral examination: 05 November 2018

## AFFIDAVIT

I hereby declare that this dissertation with the title “Structure-based mechanistic analysis of the proteasome” has been written independently and with no other aids or sources than quoted. This thesis (wholly or in part) has not been submitted elsewhere for any academic award or qualification.

---

Fabian Henneberg

# Abstract

The ubiquitin proteasome pathway is one of the major protein degradation pathways in the cell. This pathway is essential for protein quality control and the regulation of the cell cycle. Proteins subjected to the ubiquitin pathway are cleaved into short peptides by the 26S proteasome. The 26S proteasome consists of the 19S regulatory particle and the 20S core proteasome. Ubiquitinated substrates are recognized through their ubiquitin moiety by the 19S regulatory particle, unfolded and translocated into the 20S core proteasome. Within the 20S proteasome hydrolysis of the protein peptide bonds takes place. The aforementioned significance of proteasome for cell function and the associated fact that the proteasome is a drug target for the treatment of cancer makes it an attractive object for structural analyses. A detailed understanding of the proteasome structure is essential to dissect the function and its modulation by proteins and small molecule effectors.

In this thesis the sample quality of purified human 20S proteasomes was improved by biochemical optimization and by the introduction of a chromatography free purification protocol. The newly developed purification protocol was shown to be transferable to the purification of human 26S proteasomes, 20S proteasomes from *Saccharomyces cerevisiae*, *Drosophila melanogaster* and *Thermoplasma acidophilum*. Crystals were grown and structures were solved for all purified 20S proteasomes. The established post-crystallization treatment resulted in human and fruit fly 20S proteasomes crystals diffracting to 2.0 Å on average. The best resolution of 1.8 Å was collected for the human 20S proteasome.

The robust purification and crystallization protocol leading to unprecedented resolutions obtained for the human 20S proteasome enabled to reanalyze inhibition mechanism of known 20S proteasome inhibitors. Furthermore, screening efforts using small molecules revealed the diazomethylketone C-3455 as new modulator of the 20S proteasome. Most of the previously described reaction mechanisms could be validated. The largest discrepancy to published reaction mechanisms was found for epoxyketone inhibitors. Instead of the proposed formation of a 1,4-morpholine ring, the newly solved structures suggest the formation of a 1,4-oxazepane ring. The observed reaction mechanism was validated by structural analyses, electronic structure calculations of the reaction pathways and kinetic analyses. Based on these results a new reaction mechanism was proposed for epoxyketone inhibition enabling the design of new 20S proteasome inhibitors.

Electron microscopy studies of the 26S proteasome and X-ray crystallographic studies of the fruit fly 20S proteasome revealed for the first time allosteric pathways of the proteasome. Inhibition of the proteolytic active sites in the 26S holocomplex influences the

dynamics of the 19S lid movements. For the fruit fly 20S proteasome a structural rearrangement in the  $\beta 5$ -subunit was observed upon inhibition of the chymotryptic-like site or upon binding of a short peptide activator to the  $\alpha$ -rings.

In summary, the newly established purification protocol led to 20S/26S proteasome samples of superior quality. This made it possible to analyze structural features of the proteasome at resolutions not described before. The established high throughput pipeline for crystallographic analyses of human 20S proteasomes can be used for screening small molecules in future.

# Acknowledgment

Preparing this thesis for the last 3.5 years in the Department of Structural Dynamics was an exciting and most of time fun adventure. The experimental possibilities we have in the Department are really remarkable. Being able to perform all experiments one wants to do is a privilege. I'm glad that Ashwin and Holger gave me the chance to go for a scientific treasure hunt in this environment.

In particular, I want to thank my supervisor Ashwin for pushing my personal ambition and believing in my results more than I did. Ashwin listened patiently to all my wild theories and was answering every question no matter how ridiculous they were. This made it easy to learn something new everyday. The "open door policy" of Holger and Ashwin allowed to get immediate answers to questions which I could not answer myself. This open and friendly environment is one of the reasons I enjoyed my time in the Lab very much.

Even more important than the equipment in the Lab are the people who are, in my opinion, the soul of the Department. They vitalize the lab and make every workday joyful, which makes it easy to get up in the morning. Besides the scientific curiosity, the people are a major reason to go to the lab. Therefore, I want to thank all former and present lab members: Ben, Uma, Sabrina, Erik, Georg, David, Felix, Cole, Zhenwei, Niels, Mario, Tobias, Thomas, Prakash, Claudia, Gaby, Kami, Uwe, Alexey, Alex, Juliane, Elham, Dietmar, Steffi, Dirk, Frank, Alycia, Jil, Michael, Jan and Bo, who joined me on the adventure to find scientific treasures. Not only were the people of our Department responsible for the enjoyable atmosphere, but also the happy faces on the hallway from the Rodnina Department. I'm still amazed how collaborative the environment on our floor is.

Many people were involved in the proteasome project. The almost unlimited sources of HeLa extract were provided by Thomas and Kami. Alexey, the Drosophila whisperer, always stocked up the fruit fly proteasome resources. Gleb and Thomas in Hamburg provided us with always optimal conditions to get the best out of our crystals. Of course I must not forget my Master student Alex, who is a big help in the lab. It is really fun working with him. Without all your support I could not have written this thesis. Thank you all very much for your help.

As typical for every PhD-thesis, the work was accompanied by ups and downs. Having a cheerful environment around was extremely important. Major contributions creating this environment were made by Karl, Kashish and Lukas, the members of office113. Smaller contributions were made by Sabrina, Georg, Alex, Cole and our regular guest Ole. Our different personalities complement just perfectly. Office113 was an infinite source for entertainment and deep conversations - Karl's talent for reading every contract out loud, Kashish's pranks, which were most of the time funny, and Lukas's life experiences and shepherding nature made every workday terrific. It is still surprising to me that four people can form such a dynamic team. Guys, I really appreciate the time we spent together in our small office. You gave me the extra boost in motivation when it was desperately needed.

I'm grateful to have a family which has supported me from the time when "Der Ernst des Lebens" started until today. Mama and Papa, your way of living is inspirational to me and I'm glad to have you around. Having a sister like you, Katharina, makes me very proud. I value your opinions and appreciate you sharing them with me.

Many thanks to all the people who have joined me on my way so far.

TR!



# Contents

<b>Abstract</b>	<b>II</b>
<b>Acknowledgment</b>	<b>IV</b>
<b>List of Tables</b>	<b>XI</b>
<b>List of Figures</b>	<b>XIII</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Role of protein degradation in the cell . . . . .	1
1.2 Protein degradation in the lysosome . . . . .	2
1.3 The Ubiquitin-Proteasome-System (UPS) . . . . .	2
1.4 The 20S proteasome . . . . .	4
1.4.1 20S proteasome structure from <i>Thermoplasma acidophilum</i> . . . . .	4
1.4.2 Eukaryotic 20S proteasome structure . . . . .	5
1.4.3 20S proteasome assembly . . . . .	7
1.4.4 Different types of proteasomes in humans . . . . .	8
1.4.4.1 Immunoproteasome . . . . .	8
1.4.4.2 Thymoproteasome . . . . .	9
1.4.4.3 Testis-specific proteasome . . . . .	9
1.4.5 Catalytic mechanism . . . . .	9
1.4.6 Substrate specificity . . . . .	10
1.5 19S regulatory particle . . . . .	11
1.5.1 19S base substructure . . . . .	11
1.5.2 19S lid substructure . . . . .	13
1.5.3 19S lid assembly . . . . .	14
1.6 ATP independent degradation . . . . .	15
1.6.1 11S activator . . . . .	15
1.6.1.1 11S activator structures . . . . .	16
1.6.1.2 11S activator 20S proteasome interaction . . . . .	16
1.6.2 PA200 activator . . . . .	18
1.6.3 Proteasome inhibition . . . . .	19
1.6.3.1 Covalent 20S proteasome inhibitors . . . . .	20
1.6.3.2 Subunit-specific active probes . . . . .	22
1.6.3.3 Non-covalent 20S proteasome inhibitors . . . . .	23
1.6.3.4 19S regulatory particle inhibitors . . . . .	24
1.6.3.5 Relevance of proteasome inhibitors as drug . . . . .	25
1.6.4 Aims of the thesis . . . . .	26
<b>2 Material and Methods</b>	<b>29</b>
2.1 Materials . . . . .	29
2.1.1 Special Equipment . . . . .	29



2.1.2	Chemicals . . . . .	30
2.1.3	Buffers and Media . . . . .	32
2.1.4	Software . . . . .	34
2.2	Methods . . . . .	35
2.2.1	Molecular Biology . . . . .	35
2.2.2	Protein purification . . . . .	36
2.2.2.1	ProteoPlex measurement . . . . .	36
2.2.2.2	<i>Homo sapiens</i> 20S proteasome purification . . . . .	36
2.2.2.3	<i>Drosophila melanogaster</i> 20S proteasome purification . . . . .	37
2.2.2.4	<i>Saccharomyces cerevisiae</i> 20S proteasome purification . . . . .	38
2.2.2.5	<i>Thermoplasma acidophilum</i> 20S proteasome purification . . . . .	39
2.2.2.6	<i>Homo sapiens</i> 26S proteasome purification . . . . .	41
2.2.2.7	Protein concentration determination . . . . .	42
2.2.3	20S proteasome crystallization . . . . .	42
2.2.3.1	<i>Homo sapiens</i> 20S proteasome . . . . .	42
2.2.3.2	<i>Drosophila melanogaster</i> 20S proteasome . . . . .	42
2.2.3.3	<i>Saccharomyces cerevisiae</i> 20S proteasome . . . . .	43
2.2.3.4	<i>Thermoplasma acidophilum</i> 20S proteasome . . . . .	43
2.2.3.5	Post-crystallization treatment . . . . .	43
2.2.4	Inhibitor soaking protocol . . . . .	44
2.2.5	Peptide soaking protocol . . . . .	44
2.2.6	Diffraction data collection . . . . .	44
2.2.7	Structure determination . . . . .	45
2.2.8	Dose series . . . . .	46
2.2.9	Negative stain electron microscopy . . . . .	46
2.2.9.1	Inhibition protocol for negative stain electron microscopy . . . . .	46
2.2.9.2	Grid preparation . . . . .	47
2.2.9.3	Image acquisition . . . . .	47
2.2.9.4	Image processing . . . . .	47
2.2.10	Electron cryo-microscopy . . . . .	47
2.2.10.1	Inhibition protocol for negative stain electron microscopy . . . . .	48
2.2.10.2	Electron cryo-microscopy sample preparation . . . . .	48
2.2.10.3	Data acquisition . . . . .	48
2.2.10.4	Image processing . . . . .	48
2.2.10.5	Model building . . . . .	49
2.2.11	Energy landscape calculation . . . . .	50
2.2.12	Electronic structure calculations . . . . .	51
2.2.13	Enzyme kinetics . . . . .	52
2.2.14	Sequence alignments . . . . .	53
<b>3</b>	<b>Results</b>	<b>55</b>
3.1	Chromatography free purification protocol . . . . .	55
3.2	Adaptation of the human 20S purification protocol to other sources . . . . .	57
3.2.1	<i>Drosophila melanogaster</i> 20S proteasome purification . . . . .	57
3.2.2	<i>Saccharomyces cerevisiae</i> 20S proteasome purification . . . . .	58
3.2.3	<i>Thermoplasma acidophilum</i> 20S proteasome purification . . . . .	59
3.3	Crystallization of 20S proteasomes . . . . .	61
3.3.1	Crystal packing . . . . .	62

3.3.2	Post-crystallization treatment . . . . .	63
3.3.3	Small molecule soaking into 20S proteasome crystals . . . . .	65
3.4	Structural analysis of the native human 20S proteasome . . . . .	66
3.4.1	Validation of modeled ions . . . . .	67
3.4.2	Structural investigation of inhibitor design . . . . .	69
3.4.3	Structural investigation of inhibition mechanisms . . . . .	71
3.4.3.1	Aldehyde inhibitors . . . . .	72
3.4.3.2	Boronic inhibitors . . . . .	74
3.4.3.3	Vinyl sulfon inhibitors . . . . .	76
3.4.3.4	Ketoaldehyde inhibitors . . . . .	76
3.4.3.5	Epoxyketone inhibitors . . . . .	79
3.4.3.5.1	Validation of the 1,4-oxazepane linkage in an acetate 20S proteasome crystal . . . . .	81
3.4.3.5.2	Analysis of radiochemistry contribution to the reaction mechanism . . . . .	81
3.4.3.5.3	Analysis of a structurally diverse epoxyketone inhibitor . . . . .	82
3.4.3.5.4	Electronic structure calculations of 6-membered and 7-membered ring formation . . . . .	83
3.4.3.5.5	Kinetic analysis of 7-membered and 6-membered ring forming inhibitors . . . . .	84
3.4.3.6	Diazomethylketone inhibitors . . . . .	85
3.4.3.6.1	Diazomethylketone inhibitor soaked crystals harvested at 4 °C . . . . .	86
3.4.3.6.2	Diazomethylketone inhibitor soaked crystals harvested at 18 °C . . . . .	86
3.4.4	Allosteric communication between proteolytic active site and 19S lid . . . . .	91
3.4.4.1	Validation of Oprozomib binding to the 26S proteasome . . . . .	91
3.4.4.2	26S proteasome dynamics upon inhibition analyzed by negative stain EM and cryo-EM . . . . .	91
3.4.5	Structural analysis of the fruit fly 20S proteasome . . . . .	93
3.4.5.1	Sequence comparison of the human and the fruit fly 20S proteasome . . . . .	94
3.4.5.2	Structure of the native fruit fly 20S proteasome . . . . .	94
3.4.5.3	Dynamics of the $\beta$ 8 sheet of the fruit fly proteasome $\beta$ 5-subunit . . . . .	96

**4 Discussion 99**

4.1	Benefits of a chromatography free purification . . . . .	100
4.2	Crystal packing . . . . .	102
4.3	Benefit of the post-crystallization procedure . . . . .	103
4.4	Determination of correct resolution cut-off . . . . .	104
4.5	Implication of the chloride position found in the proteolytic sites . . . . .	107
4.6	Catalytic water in the structure . . . . .	107
4.7	Aldehyde inhibition . . . . .	108
4.8	Epoxyketone inhibition . . . . .	109
4.9	Allostery in the proteasome . . . . .	113
4.10	Alternative conformation of the $\beta$ 5 disorder in the fruit fly 20S proteasome . . . . .	114

4.11 Active pocket definition and substrate specificity . . . . .	115
<b>5 Conclusion and Outlook</b>	<b>117</b>
<b>A Abbreviations</b>	<b>119</b>
<b>B Supplementary information</b>	<b>123</b>
B.1 Table 1 of solved yeast and archaeal 20S proteasome structures . . . . .	123
B.2 Plasmid map . . . . .	124
B.3 Occupancy list of inhibitors soaked into human 20S proteasomes . . . . .	124
B.4 Proteasome nomenclature . . . . .	126
B.4.1 20S proteasome nomenclature . . . . .	126
B.4.2 19S regulatory, activator and assembly chaperone subunits . . . . .	128
B.5 Declaration of contribution as co-author . . . . .	130
<b>C References</b>	<b>131</b>
<b>Curriculum Vitae</b>	<b>143</b>

# List of Tables

2.1	Equipment used for experiments . . . . .	29
2.2	Chemicals used for experiments . . . . .	30
2.3	Buffers and Media used for experiments . . . . .	32
2.4	Software used for data evaluation . . . . .	34
2.5	Post-crystallization procedures used for 20S proteasome crystals . . . . .	44
2.6	Refinement strategies used for 20S structure determinations . . . . .	45
3.1	20S proteasome crystallization conditions . . . . .	62
3.2	Atomic distances used for modeling ions in the structure . . . . .	67
3.3	Peptide backbone sequences of boronic inhibitors . . . . .	76
3.4	Table 1 for all solved human 20S proteasome structures part 1/2 . . . . .	89
3.5	Table 1 for all solved human 20S proteasome structures part 2/2 . . . . .	90
3.6	Emboss needle sequence alignment of human and fruit fly 20S proteasome subunits . . . . .	94
3.7	Table 1 for all solved fruit fly 20S proteasome structures . . . . .	96
B.1	Table 1 of solved yeast and archaeal 20S proteasome structures . . . . .	123
B.2	List of 20S proteasome inhibitors soaked into human 20S proteasomes . . . . .	125
B.3	Nomenclature of <i>Thermoplasma acidophilum</i> 20S proteasome . . . . .	126
B.4	Nomenclature of eukaryotic 20S proteasomes . . . . .	127
B.5	Nomenclature of 19S regulatory, activator and assembly chaperone subunits . . . . .	129

# List of Figures

1.1	Scheme of the ubiquitin proteasome pathway . . . . .	3
1.2	Representation of the <i>Thermoplasma acidophilum</i> 20S proteasome structure . . .	5
1.3	Secondary structure elements and sequences of the <i>Thermoplasma acidophilum</i> 20S proteasome . . . . .	6
1.4	General structure of the human 20S proteasome . . . . .	7
1.5	Catalytic cycle of the peptide cleavage reaction . . . . .	10
1.6	Representation of the 19S regulatory particle . . . . .	12
1.7	Representation of the activators PA28 $\alpha\beta$ and BLM10 . . . . .	17
1.8	Overview of 20S proteasome function . . . . .	19
1.9	Covalent 20S proteasome inhibitor classes . . . . .	22
1.10	Non-covalent 20S proteasome inhibitor classes . . . . .	24
1.11	19S regulatory particle inhibitors . . . . .	25
3.1	Overview of the newly developed purification strategy . . . . .	56
3.2	Activity measurement and ProteoPlex analysis of the 20S proteasome . . . . .	57
3.3	Comparison of fruit fly and human 20S proteasome purifications . . . . .	59
3.4	Yeast 20S proteasome purification . . . . .	60
3.5	Archaeal 20S proteasome purification . . . . .	61
3.6	Time-lapse of human 20S crystal growth . . . . .	62
3.7	Comparison of the crystal packing of 20S proteasomes from different species . . .	64
3.8	Diffraction pattern of 1.8 Å human 20S proteasome dataset . . . . .	65
3.9	Structural features of the human 20S proteasome . . . . .	68
3.10	Anomalous diffraction map of bromide ions used to validate chloride positions . .	69
3.11	Alternative conformations of Met45 in the chymotryptic-like site . . . . .	71
3.12	Structural features of inhibitor binding to the chymotryptic-like site. . . . .	72
3.13	Aldehyde inhibition of the chymotryptic-like site . . . . .	73
3.14	Inhibition of the chymotryptic-like site by boronic inhibitors . . . . .	75
3.15	Vinyl-sulfon inhibition of the chymotryptic-like site . . . . .	77
3.16	Ketoaldehyde inhibition of the chymotryptic-like site . . . . .	78
3.17	Comparison of ketoaldehyde binding observed in the presented structures with the published ketoaldehyde-inhibited 20S proteasome structure . . . . .	79
3.18	Epoxyketone inhibition of the chymotryptic-like site . . . . .	80
3.19	Oprozomib inhibitor soaked in Mg(OAc) <sub>2</sub> 20S proteasome crystals . . . . .	81
3.20	Analysis of dose rate on the epoxyketone reaction mechanism . . . . .	82
3.21	Chymotryptic-like site inhibition with the epoxyketone Dihydroeponepcylin . . .	83
3.22	Electronic structure calculations for the 6-membered 6R and 7-membered 7R ring formation pathways . . . . .	84
3.23	Kinetic analysis of epoxyketone and ketoaldehyde inhibition mechanisms . . . .	85
3.24	Inhibition of the chymotryptic-like by the diazomethylketone inhibitor C-3455 . .	88
3.25	Details about the 26S proteasome structure . . . . .	92
3.26	26S proteasome dynamics analyses by electron microscopy . . . . .	93
3.27	Re-ordering of $\beta$ -subunit from <i>Drosophila melanogaster</i> upon inhibitor binding .	95
3.28	Activator peptide binding induced structural rearrangements of the $\beta$ 5 subunit .	98
4.1	Staraniso analyses of best diffracting human and yeast 20S proteasome crystals .	106

4.2	Published inhibition mechanism of the 20S proteasome by epoxyketone inhibitors	109
4.3	Newly proposed inhibition mechanism for epoxyketone 20S proteasome inhibitors	110
B.1	Plasmid map for <i>Thermoplasma Acidophilum</i> expression . . . . .	124



# 1 Introduction

## 1.1 Role of protein degradation in the cell

Every living organism from complex multicellular (e.g. *Homo sapiens*) to simpler unicellular (e.g. *Escherichia coli* (*E. coli*)) organisms needs to adapt, react and interact with their environment in order to survive. A cell is able to interact with the environment by sensing mechanical forces, electrical stimuli, or chemical substances. The presence of nutrients can trigger the uptake and subsequent metabolism to make them available in the cell. Sensing of such external signals is mediated by proteins [1].

Proteins are necessary to catalyse chemical reactions at ambient conditions. The blueprint for proteins is located in the deoxyribonucleic acid (DNA) sequence. Ribonucleic acid (RNA) polymerases transcribe DNA into RNA, of which messenger RNA (mRNA) in eukaryotes is synthesized by RNA polymerase II that is then translated to proteins by ribosomes [2]. The DNA sequence in each organism's genome contains the blueprint of several hundred proteins needed for cell survival and growth. As proteins catalyse distinct reactions, most of them are not needed throughout the whole lifespan of the cell, but for limited periods of time.

Protein levels in the cell are highly controlled through regulated synthesis and degradation. Protein degradation removes proteins that are not needed anymore or are compromised in their function. The gradual accumulation of proteins in a living cell can lead to cellular dysfunction and ultimately to its apoptosis [3]. Protein degradation functions not only as a clearance mechanism for unwanted or denatured proteins, but also as a mechanism to regulate cellular processes. For example, in eukaryotic cells that are dividing through mitosis, chromosomes are bound together by ring-like proteins called Cohesins. Chromosomes must separate equally into each daughter cell and rely on a concerted cleavage of Cohesin complexes by the protease called Separase. Prior to mitosis, Separase is bound and inactivated by Securin. At the onset of anaphase, Securin is rapidly degraded, leading to the activation of Separase. Chromosomes are then able to separate and allowing the cell to progress with cell division [4].

Not only does protein synthesis gives the cell the ability to react to environmental changes, but protein degradation also is essential for keeping a balanced cell homeostasis and for the rapid modulation of cellular pathways [5]. Two major protein degradation pathways are known – the lysosomal proteolysis pathway and the ubiquitin-proteasome pathway (UPS).



## 1.2 Protein degradation in the lysosome

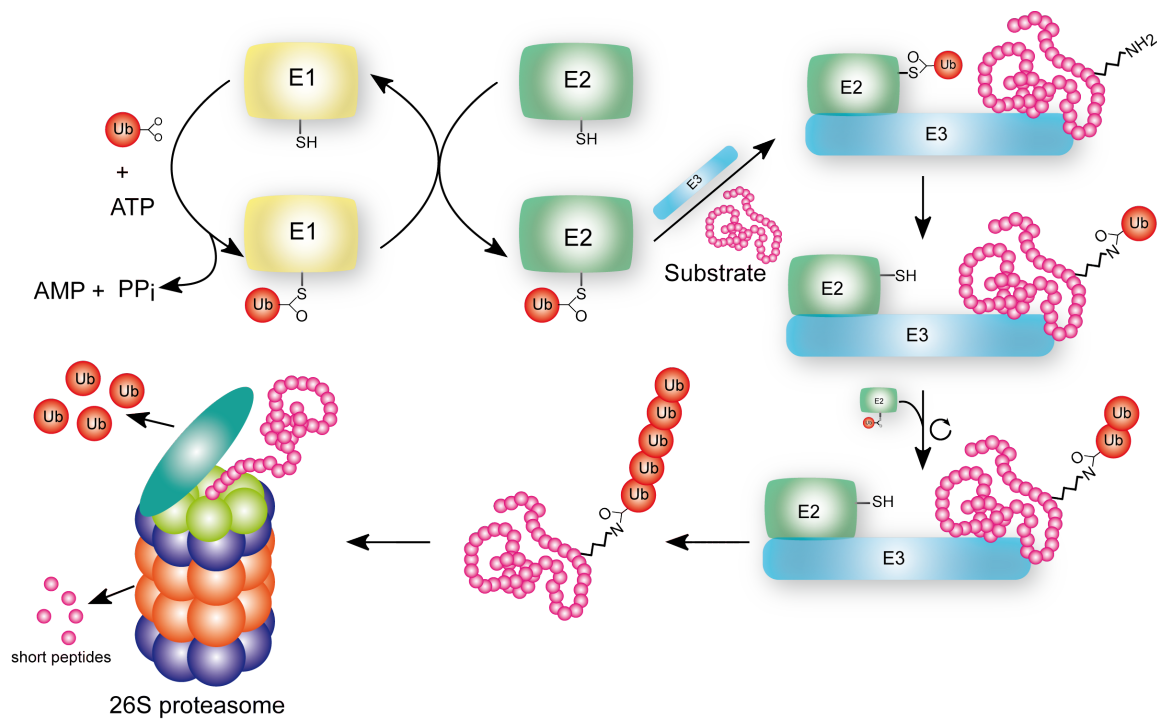
Lysosomes were first observed in 1955 [6]. Lysosomes form degradative membrane-bounded compartments in eukaryotic cells, which are filled with hydrolases. The vacuolar type H<sup>+</sup>-ATPases (v-ATPase) maintain the internal acidic pH (4.5 – 5.5) of lysosomes, which is needed for the optimal enzymatic activity of the enclosed hydrolases [7]. Misfolded or aggregated proteins, damaged cell organelles and intracellular pathogens are degraded into their constituent components by the lysosomes: amino acids, monosaccharides and free fatty acids. These components can then be recycled by the cell for metabolic processes [8]. Delivery of cargo to the lysosome is mediated by endocytic pathways, autophagy, or chaperone-mediated autophagy [9]. After the discovery of the lysosome, it was commonly believed that lysosomes were the major centre for protein degradation in the cell. However, inactivation of hydrolases by alkalinisation of lysosomes did not affect the half-life of most proteins and the discovery of an adenosine triphosphate (ATP)-dependent protein degradation pathway in reticulocyte extract by Hershko and colleagues indicated that the existence of another protein degradation pathway was active inside eukaryotic cells [10, 11].

## 1.3 The Ubiquitin-Proteasome-System (UPS)

The selective and targeted degradation of proteins in the cell is mediated by the UPS, discovered by Hershko and Ciechanover in 1978 [10]. This pathway operates by selectively labelling proteins with a general marker for degradation, ubiquitin, which recruits the target to the proteasome. Ubiquitin is a 76 kilo Dalton (kDa) signal molecule present only in eukaryotes that is attached to substrate proteins to regulate many cellular processes, like substrate degradation, transcription regulation, signal transduction, membrane-protein trafficking, endocytosis and DNA repair [12, 13]. An enzyme cascade comprised of the E1-activating enzyme, E2-like enzyme and E3 protein ligase is needed to label proteins with ubiquitin (Fig. 1.1) [14].

First, the C-terminal glycine of ubiquitin is activated in an ATP-dependent manner to form an ubiquitin adenylate intermediate. A thioester linkage between the cysteine of the E1-activating enzyme and the ubiquitin adenylate intermediate is then formed by releasing adenosine monophosphate (AMP). The activated ubiquitin is then passed from the E1 to a cysteine residue on the E2 conjugating enzyme. The E3 ligating enzyme binds to the E2 conjugating enzyme and to the substrate, which should be labeled with ubiquitin. The close proximity of an E2 activating enzyme and the substrate facilitates the formation of an iso-peptide linkage between the C-terminus of ubiquitin and the  $\epsilon$ -amino

group of a lysine encoded in the substrate sequence. The ubiquitin labeling reaction can be repeated, meaning that another ubiquitin is attached to the previously added ubiquitin, leading to a poly-ubiquitin chain [15]. As ubiquitin encodes seven lysine residues in its sequence, several different linkage types are possible [16]. Proteins that are destined for proteolysis by the 26S proteasome are marked with a Lys48-polyubiquitin chain [17].



**Fig. 1.1: Scheme of the ubiquitin proteasome pathway.** Substrates are labelled via an E1-E2-E3 enzyme cascade with ubiquitin. Polyubiquitinated substrates are recognized by the 19S regulatory particle, where deubiquitination, unfolding and translocation of the substrate into the 20S proteasome chamber takes place. Degradation of the substrate occurs at the proteolytic sites, which reside in the  $\beta$ -rings.

The ubiquitin chain is recognized by the 19S lid of the 26S proteasome. After recognition of the ubiquitin chain, the ubiquitin is removed from the substrate by deubiquitylating enzymes (DUBs). Subsequently, the substrate is unfolded and translocated into the proteolytic active chamber of the 20S proteasome, where it is enzymatically degraded [18]. The released ubiquitin is then recycled by the cell. The proteasome degrades proteins into peptides of 3 to 22 amino acids in length, although fragments of 3 to 9 amino acids form the most abundant species among the generated peptides [19–21].

## 1.4 The 20S proteasome

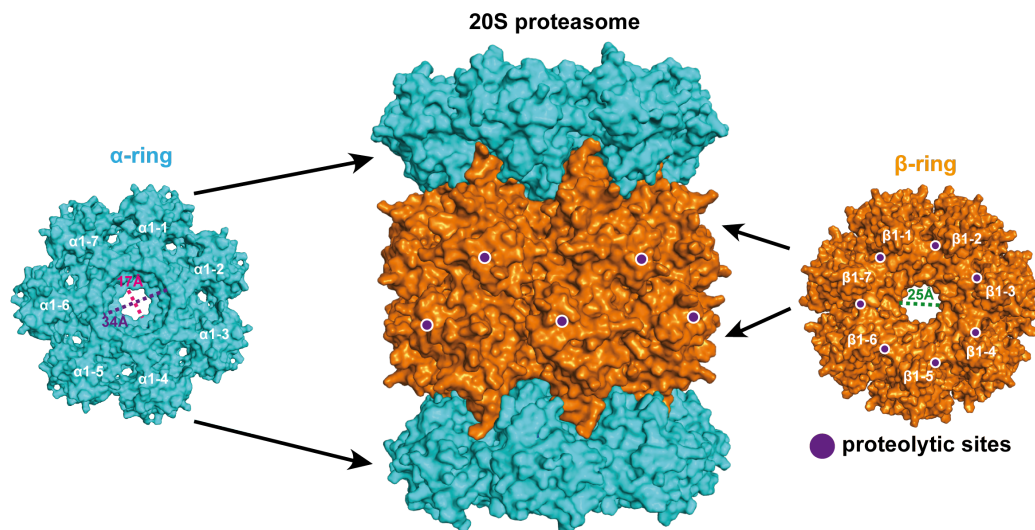
The cylindrically-shaped structure of the 20S proteasome purified from human erythrocytes was already observed by negative stain electron microscopy (EM) in 1968 [22]. Since then, similar structures were found to exist in several other organisms [23]. So far, 20S proteasomes were detected in eukaryotes, all archaea and in bacteria of the order *Actinomycetales* and *Nitrospirales*. Bacterial and archaeal 20S proteasomes are produced from two genes, in contrast to the eukaryotic 20S proteasome, which is encoded by 14 different genes [24]. Valuable insights into the structure of the proteasome were first provided by the crystal structure of the 20S proteasome from the archaeal organism *Thermoplasma acidophilum* [25].

### 1.4.1 20S proteasome structure from *Thermoplasma acidophilum*

The first crystallographic studies of the archaeal proteasome confirmed the 15 nm × 11 nm barrel shaped structure suggested by EM studies [26]. In total, the 20S proteasome is composed of 28 subunits. The barrel is formed by four coaxially stacked heptameric rings. These rings contain either seven  $\alpha$ -subunits or seven  $\beta$ -subunits. The proteolytic chamber in the center of the barrel is formed by  $\beta$ -rings. At each free site of the  $\beta$ -ring stack, an  $\alpha$ -ring is attached (Fig. 1.2). Inside of the  $\alpha$ -rings the antechambers are located. The chambers are separated by constricting annuli. The structure has a two-fold symmetry axis horizontally intersecting the  $\beta$ -rings and a seven-fold symmetry axis passing vertically through the centre of the heptameric rings [25].

The  $\alpha$ - and  $\beta$ -subunits have very similar folds, indicating that they likely emerged by gene duplication (Fig. 1.3) [27]. Two five-antiparallel  $\beta$ -sheets (bss) (bs1: S1, S2, S8, S9, S10 and bs2: S3, S4, S5, S6, S7) form a  $\beta$ -sheet sandwich in the middle of the subunits. Helices H1 and H2 are flanking the sandwich at the bs2 site and helices H3, H4 and H5 are located next to bs1. The major differences between the  $\alpha$ -subunit and the  $\beta$ -subunit are the missing H0 helix, an extended loop region between S2 and S3, a shorter H2 helix and a longer H5 helix in the  $\beta$ -subunit [25]. The interface between the  $\alpha$ - and  $\beta$ -ring is dominated by the interaction of the H1 and H2 helices in both subunits. Access to the antechamber of the 20S proteasome barrel is restricted by a 17 Å diameter annulus, formed by the loops connecting the H2 helix and the S5  $\beta$ -strand of the  $\alpha$ -subunits.

20S proteasomes belong to the class of N-terminal proteases. Mutational analysis of the  $\beta$ -subunit N-terminal threonine (Thr1) and inhibitor-soaked structures performed for the *Thermoplasma acidophilum* 20S proteasome proved the Thr1 to be the catalytic active residue [25, 28, 29].



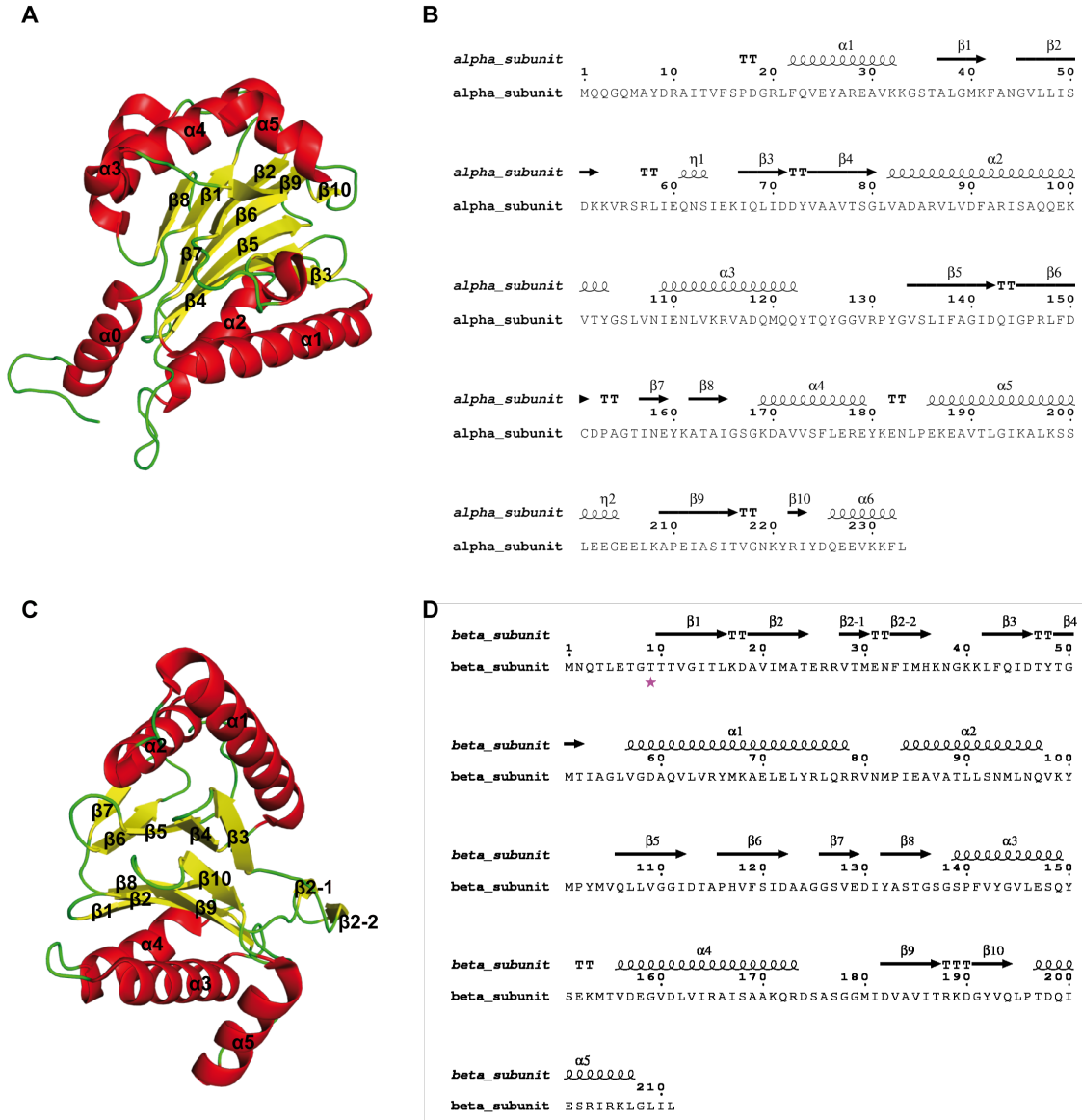
**Fig. 1.2: Volume representation of the *Thermoplasma Acidophilum* 20S proteasome structure.** The *Thermoplasma acidophilum* 20S proteasome consists of two  $\alpha$ -rings (blue) and two  $\beta$ -rings (orange). The individual subunits are numbered. The location of the proteolytic sites are marked with purple dots. The diameters of the gate and the constriction annuli are indicated with dashed lines and the respective distances in Å (PDB: 1PMA).

### 1.4.2 Eukaryotic 20S proteasome structure

The overall architecture of the eukaryotic 20S proteasome from *Saccharomyces cerevisiae* is similar to the archaeal 20S proteasome. Not only is the overall structure similar, but also the basic architecture of the  $\alpha$ - and  $\beta$ -subunits [31]. However, the structure is more complex, as the  $\alpha$ - and  $\beta$ -rings are composed by seven different subunits each. Thus, 14 different subunits form the eukaryotic 20S proteasome (Fig. 1.4). All subunits have distinct locations. Differences between the subunits can be found in turns, which vary by one to two amino acids, in N-terminal and C-terminal regions. The length of insertion loops connecting secondary structure elements is specific for each subunit, resulting in the formation of new secondary elements or the extension of existing secondary structure elements. The  $\beta$ 2-subunit has a long C-terminal extension enclosing the neighboring  $\beta$ 3-subunit and extending a  $\beta$ -sheet of the  $\beta$ 3-subunit [31].

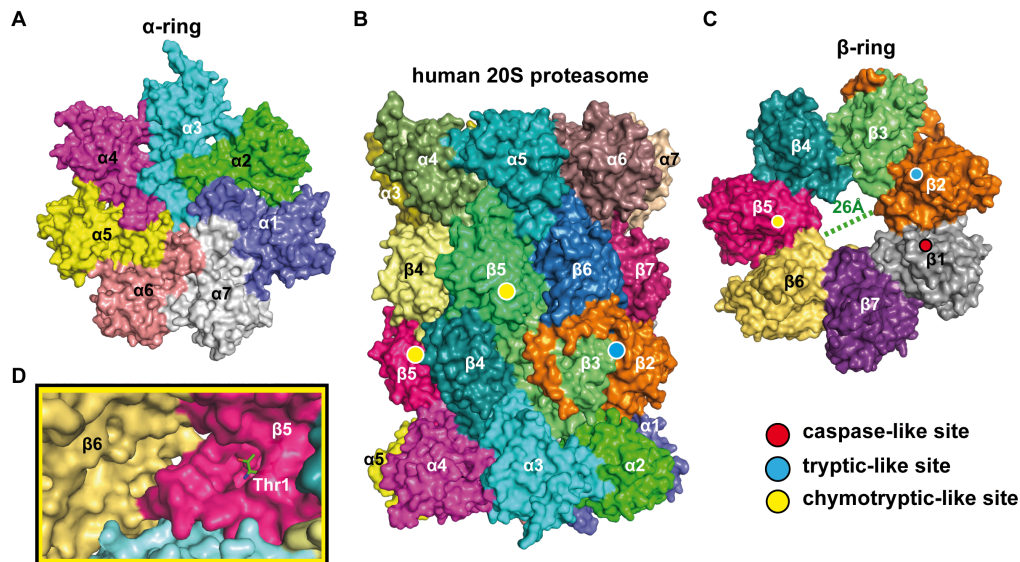
Sequence alignment studies identified three proteolytic active sites in each  $\beta$ -ring of eukaryotic 20S proteasomes. Conserved features of these subunits are the Thr1 and Lys33 [28]. Structural analysis of the inhibited 20S proteasome from *Saccharomyces cerevisiae* and mutagenetic analysis of the yeast proteasome confirmed that three different proteolytic active sites in eukaryotic 20S proteasomes exist [20, 31].

Another important functional feature of the 20S proteasome structure is the gate which restricts access to the 20S proteasome chamber. The gate is formed by N-terminal residues of the  $\alpha$ -subunits extending into the entrance pore and stacking on top of each other [31]. The comparison of *Saccharomyces cerevisiae* 20S proteasome (y20S) with mammalian 20S proteasomes reveals high structural similarity. Minor differences can be found in



**Fig. 1.3: Secondary structure elements and sequences of the *Thermoplasma acidophilum* 20S proteasome.** Ribbon representation of the  $\alpha$ -subunit (A) and  $\beta$ -subunit (C) tertiary structures of *Thermoplasma acidophilum* 20S proteasome are shown.  $\alpha$ -helices are colored in red, loops in green, and  $\beta$ -sheets in yellow. The secondary structure elements are numbered accordingly. Primary sequences of the  $\alpha$ -subunit (B) and  $\beta$ -subunit (D) are annotated with their secondary structure elements using ENDscript 2.0 [30].

the amino acid sequence, parts of connecting secondary elements and extensions of the C-termini. The observed differences do not have a decisive influence on the structure and cannot be attributed to a special functional role [32].



**Fig. 1.4: General structure of the human 20S proteasome.** Space filling representations of the human  $\alpha$ -ring (A), full 20S proteasome assembly (B) and  $\beta$ -ring (C) are shown. All subunits are shown in different colors. The caspase-like site is marked with a red dot, the tryptic-like site with a blue dot and the chymotryptic-like site with a yellow dot. D depicts a close up view of the chymotryptic-like site (PDB:5LE5).

### 1.4.3 20S proteasome assembly

28 subunits need to be expressed and assembled in a spatially regulated manner to form a functional 20S proteasome. The less complex 20S proteasomes structures from *Archaea* and *Actinomycetales* self-assemble when  $\alpha$ - and  $\beta$ -subunits are co-expressed in *E. coli* [33]. The autocatalytic cleavage of the propeptides, which protects the proteolytic active residues of the  $\beta$ -subunit from N- $\alpha$ -acetylation, is not needed for the 20S proteasome assembly when expressed in *E. coli* [34–36].

Assembly of the more complex eukaryotic 20S proteasome relies on the action of assembly chaperones POMP1 (yeast: Ump1), PAC1/PAC2 (yeast: Pba1/Pba2) and PAC3/PAC4 (yeast: Pba3/Pba4) (Tab. B.5). The assembly pathway of the yeast proteasome was studied based on structural and biochemical data [37–39]. The heterodimers Pba1/Pba2 and Pba3/Pba4 help to form a heptameric ring of  $\alpha$ -subunits and stabilize it during incorporation of the  $\beta$ -subunits. The  $\beta 2$ -subunit in complex with Ump1 is the first to associate with the  $\alpha$ -ring. The insertion of  $\beta$ -subunits 3 and 4 leads to the release of Pba3/Pba4. After  $\beta$ -subunits 5, 6 and 1 are bound, the  $\beta 7$ -subunit associates and induces dimerization of two half-proteasomes forming the pre-holo-proteasome. Finally, the Pba1/Pba2 complex dissociates and the  $\beta$ -subunits are auto-catalytically activated. After degradation of Ump1, the mature 20S proteasome is formed [40].

The autolysis is performed by a nucleophilic attack of the Thr1OH, which is deprotonated by Lys33, at the carbonyl carbon of the amide bond linking the Thr1 and Gly1. The collapse of the tetrahedral transition state frees the N-terminus of Thr1. An acyl



intermediate is formed after the nucleophilic attack by a water molecule, which is polarized by the N-terminus. Hydrolysis of the acyl intermediate frees and activates the Thr1 residues [41]. The pro-peptides of non-catalytic active  $\beta$ -subunits are cleaved off by the catalytic active  $\beta$ -subunits in the mature 20S proteasome [42].

#### 1.4.4 Different types of proteasomes in humans

Besides the constitutive 20S proteasome in mammals, which is ubiquitously present in most cells of the body, immunoproteasomes, thymoproteasomes and testis-specific proteasomes exist in cells of hematopoietic origin, cortical thymic epithelium cells (cTEC) and male germ cells entering into meiotic prophase, respectively [43]. Apart from these cell specific 20S proteasomes, several other intermediate 20S proteasomes containing different  $\alpha$ - and  $\beta$ -subunit proteoforms were also described [44].

##### 1.4.4.1 Immunoproteasome

The Immunoproteasome is involved in the immune response by generating major histocompatibility complex (MHC) class-I restricted T lymphocyte (T cell) epitopes for presentation to cytotoxic T cell. This proteasome preserves some peptide antigens that would otherwise be degraded by the constitutive 20S proteasome [45]. Via this mechanism, pathogens that have infected cells can be identified and efficiently eliminated by CD8+ T cell that scan MHC class I molecules using their T cell receptors (TCR) and CD8+ co-receptors. The three catalytic subunits  $\beta$ 1,  $\beta$ 2 and  $\beta$ 5 are exchanged with subunits LMP2 ( $\beta$ 1i), MECL-1 ( $\beta$ 2i) and LMP7 ( $\beta$ 5i). The expression of the immunoproteasome specific subunits can be induced by interferon- $\gamma$  (IFN- $\gamma$ ) or tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ). The intermediate immunoproteasome is reported to contain only  $\beta$ 5i or  $\beta$ 2i and  $\beta$ 5i [46]. The exchange of the subunits results in an altered cleavage pattern to generate peptides that can be accommodated by the MHC-I groove on the apical surface of the MHC receptor [47, 48]. Hydrophobic residues flanking the P1 pocket of  $\beta$ 1i substrate channel favor epitope production with small and nonpolar residues. The hydrophobic character of the  $\beta$ 5 is conserved in  $\beta$ 5i. The increased hydrophilicity in the active centre of  $\beta$ 5i might elevate peptide hydrolysis rates. No significant alteration can be found in the  $\beta$ 2i subunit compared to  $\beta$ 2 of the constitutive 20S proteasome [49].

#### 1.4.4.2 Thymoproteasome

cTEC expressing the thymoproteasome are essential for the positive selection of CD8+ T cells. The thymoproteasome produces unique MHC-1 binding peptides, which have a low affinity to TCRs to promote the positive selections of CD8+ T cells. Similar to the immunoproteasome, the  $\beta 1$  and  $\beta 2$  subunits of the thymoproteasome are replaced with  $\beta 1i$  and  $\beta 2i$ , whereas the  $\beta 5$  subunit is replaced with the  $\beta 5t$  subunit [50]. In the  $\beta 5t$  subunit, the hydrophobic pocket known from  $\beta 5/\beta 5i$  is assumed to be replaced by hydrophilic residues, which is supported by data showing reduced chymotryptic activity of thymoproteasomes in comparison to the constitutive- and immunoproteasome [51].

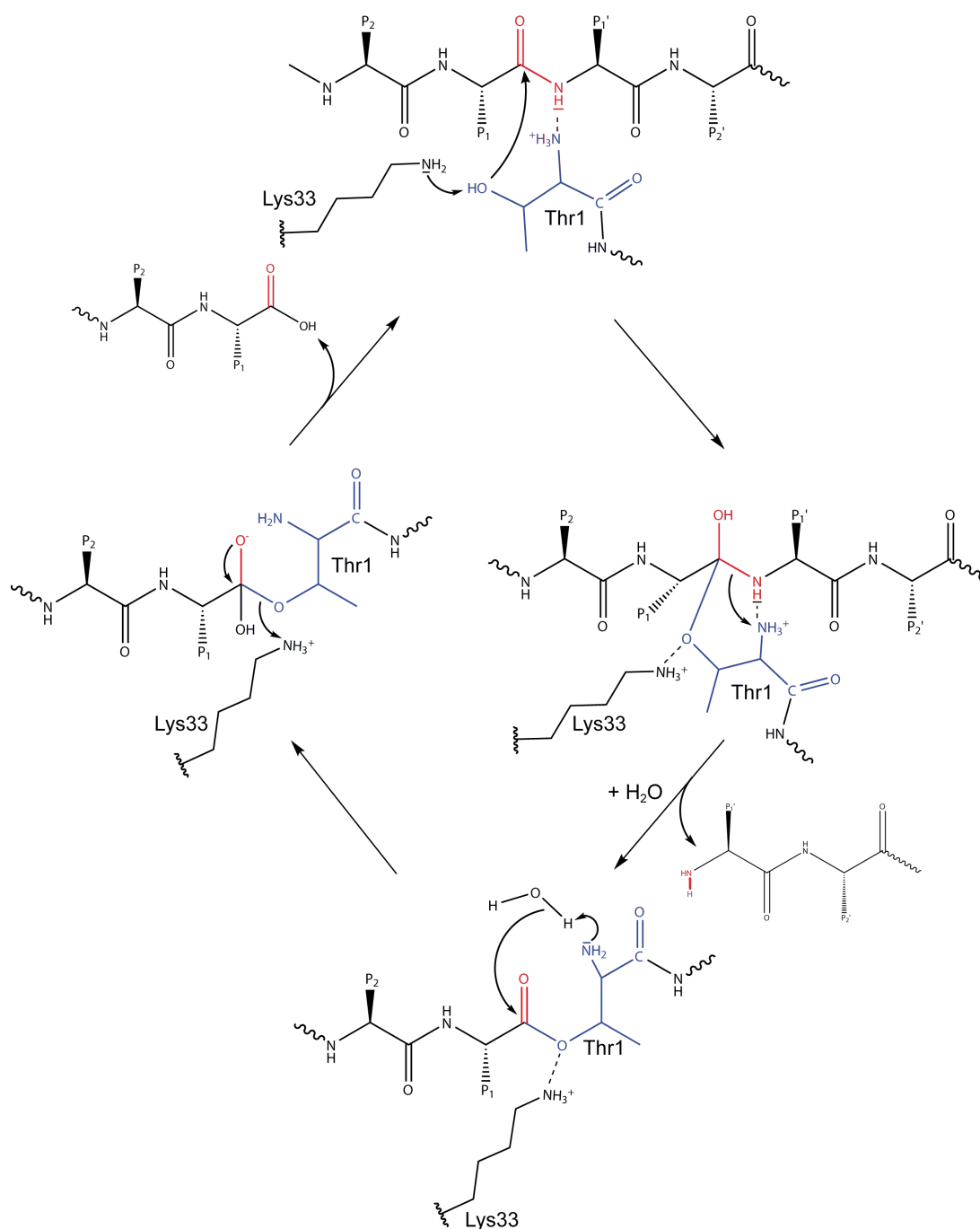
#### 1.4.4.3 Testis-specific proteasome

In testis-specific proteasomes, the  $\alpha 4s$  subunit replaces the  $\alpha 4$  subunit of the constitutive proteasome. Structural models suggest an altered charge distribution on the surface of  $\alpha 4s$  [43]. So far, the function of this subunit exchange is not understood.

### 1.4.5 Catalytic mechanism

After the substrates reach the proteolytic chamber, they are degraded by the N-terminal threonine proteases located in the  $\beta$ -subunits. Knowledge about the catalytic mechanism was inferred from several inhibitor-20S proteasome structures and mutational studies of catalytically important residues [25, 28, 31]. The residues Thr1, Asp/Glu17 and Lys33 are essential for the catalytic activity of the 20S proteasome. These residues build up a catalytic triad. Asp17/O $\delta$  increases the pK<sub>a</sub> of Lys33/NH<sub>2</sub> and fixes it in a state, in which it can deprotonate Thr1OH. The position is further stabilized by the carbonyl oxygen of Arg19. After deprotonation, the Thr1O undergoes a nucleophilic attack at the carbonyl carbon atom of the scissile peptide bond to form a tetrahedral intermediate. The amide is then protonated by the Thr1/NH<sub>3</sub><sup>+</sup> resulting in a collapse of the intermediate state and the generation of a free N-terminus at the C-terminal cleavage product. The nucleophilic attack of a water activated by Thr1/NH<sub>2</sub> at the Thr1/O ester releases the N-terminus of the remaining peptide substrate with a C-terminal carboxylic acid and frees the Thr1/OH, regenerating the active site. After one cycle is finished, the catalytic residues are ready to perform another cleavage reaction (Fig 1.5).





**Fig. 1.5: Catalytic cycle of the peptide cleavage reaction taking place in the proteolytic active sites of the 20S proteasome.** The Lys33 increases the nucleophilicity of the Thr1/O leading to a nucleophilic attack at the ketone carbon of the substrate peptide bond. The formed intermediate transition state collapses upon protonation of the amide by the Thr1/NH<sub>3</sub><sup>+</sup> and the C-terminal cleavage product is released. A water molecule activated by Thr1/NH<sub>2</sub> nucleophilic attacks the ester bond formed between Thr1/O and the substrate then. This results in the release of the N-terminal part of the substrate and the Thr1/OH is regenerated for the next catalytic cycle.

### 1.4.6 Substrate specificity

Each of the three-distinct proteolytic active sites of the eukaryotic 20S proteasome has its own cleavage characteristic. The chymotryptic-like site ( $\beta 5$ ), tryptic-like site ( $\beta 2$ ),

caspase-like site ( $\beta$ 1) cleave after hydrophobic, basic, acid residues present in P1 position, respectively. The P1 position describes the first amino acid residue of a protease substrate located N-terminal of the cleavage site. The following N-terminal amino acid residues of the protease substrate are numbered in increasing order (P1, P2, P3, etc.). Amino residues residing C-terminally to the cleavage site are marked with increasing numbers with an additional apostrophe (P1', P2', P3', etc.) to differentiate them from the N-terminal amino acid residues. The respective residue pockets are denoted with a S and the respective number of the residue (S1 pocket for the P1 residue). This nomenclature was introduced by Schechter and Berger [52].

The substrate specificities were first observed in activity measurements performed with synthetic fluorogenic tripeptides in the presence and absence of inhibitors, which selectively inhibit a specific proteolytic site [53, 54]. Further validation of the described substrate specificities was accomplished by degradation analysis of the authentic protein substrate enolase, which is thermolabile and can be degraded by the 20S proteasome. Mutations of the proteolytic active sites, leading to their inactivation, allowed the degradation pattern analysis of each respective proteolytic site [20].

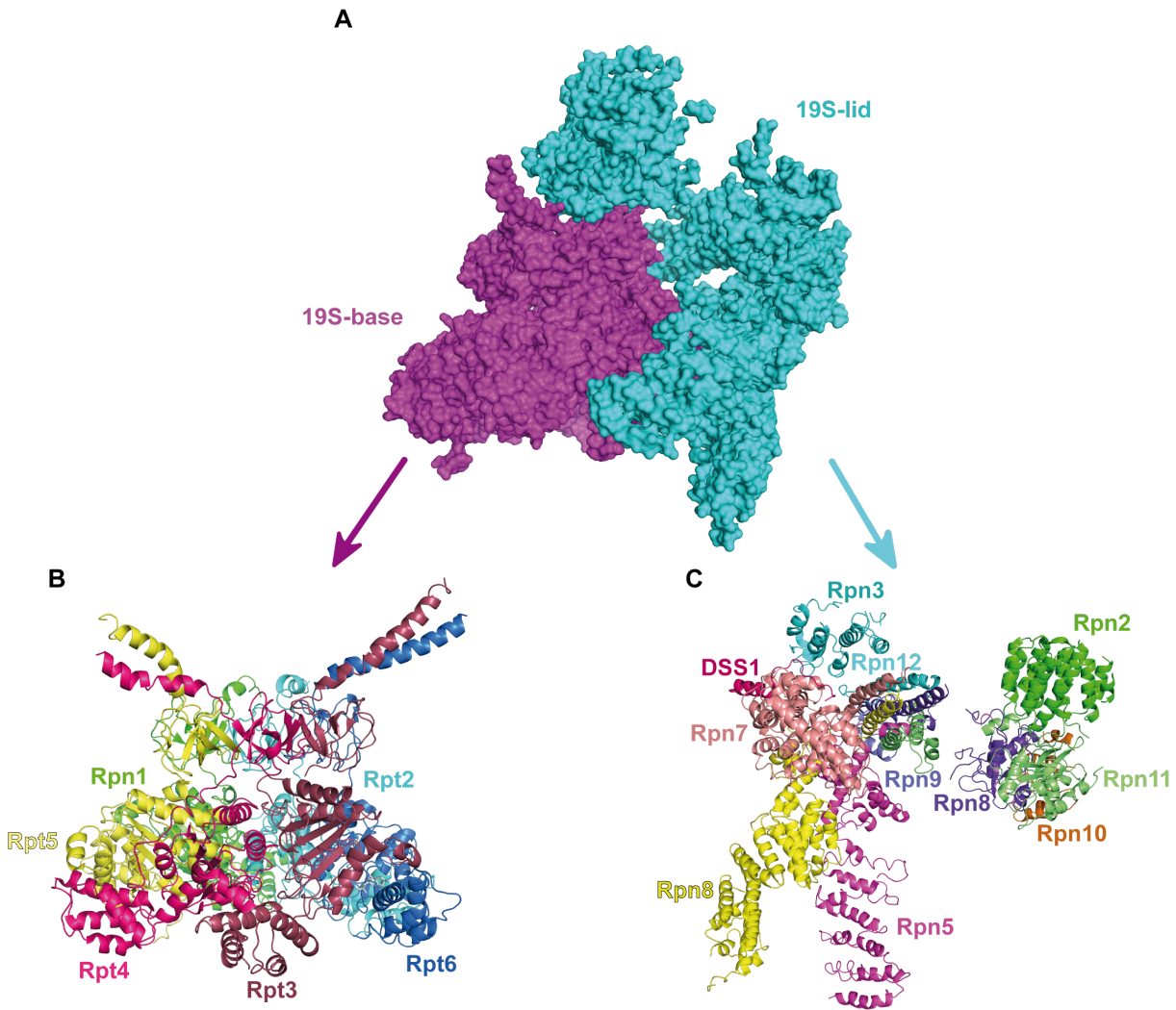
## 1.5 19S regulatory particle

The 19S regulatory particle (19S RP) functions as a gate to ensure the selective recognition and degradation of ubiquitinated substrates. The whole assembly is either called 26S or 30S proteasome based on the presence of one or two 19S RPs in the complex. Any attempts to attribute distinct functional roles to these different assemblies have not been successful [55]. The 19S RP can also be found in hybrid proteasomes, in which an 11S or a 200S activator is bound to the free  $\alpha$ -ring of the 26S proteasome [56, 57].

The 19S RP has a size of 900 kDa. The protein complex can be divided into two parts: the base and the lid (Fig. 1.6) [58]. Structural information for single proteins comprising the 19S RP was inferred from nuclear magnetic resonance (NMR) and X-ray structures [59]. Recent cryo-EM structures of the 19S RP and the human and yeast 26S proteasome have reached resolutions of as high as 4.0 Å. Those structures reveal detailed information of the 26S proteasome arrangement and functional states [60–66].

### 1.5.1 19S base substructure

The 19S base is composed of six homologous ATPases associated with diverse cellular activities (AAA+ proteins) (Rpt 1 – 6) which are assembled as an asymmetric “split



**Fig. 1.6: Representation of the 19S regulatory particle and its subunits.** A: The 19S lid is colored in blue and the 19S base in purple. B + C: Ribbon representations of the 19S base (B) and 19S lid (C) with all subunits colored differently (PDB 5M32).

washer” hexameric ring, two scaffold proteins (Rpn1 and Rpn2) and two ubiquitin receptors (Rpn10 and Rpn13) [27]. In addition to the two ubiquitin receptors, Rpn1 can also interact with ubiquitin (Tab. B.5)[62]. All ATPase subunits encode an N-terminal oligosaccharide binding (OB) domain and a C-terminal AAA+ ATPase domain. The OB fold forms a ring-like structure above the ATPase channel and connects the three coiled coils, consisting of the N-terminal helical structures of the Rpt subunits, with the ATPase domain [61]. The ATPase domain is comprised of a large and a small domain. The large domain contains a Walker A motif, Walker B motif, Arginine finger and Sensor 1. On the small domain the Sensor 2 is encoded. All these motifs are involved in nucleotide binding. The nucleotide binding pocket is enclosed by the large and small domain of one Rpt subunit and the large domain of the adjacent Rpt subunit, which provides the two arginines of the arginine finger motif to stabilize nucleotide binding [67]. Published 26S

proteasome structures verified binding of nucleotides in all binding pockets, but an unambiguous annotation of the densities to ATP, adenosine diphosphate (ADP), or ADP-P<sub>i</sub> is not possible at the current resolutions [61]. Comparison of all nucleotide binding pockets reveals structural differences in the binding pocket formed by Rpt3 and Rpt6. The ATPase subunits Rpt3 and Rpt6 are located at the end of the split washer so that Rpt3 is located closest and Rpt6 is the most distal subunit to the  $\alpha$ -ring of the 20S core particle. This asymmetry opens up a cavity between the subunits and prevents the arginine finger of Rpt3 from pointing to the nucleotide which is bound to the ATP binding site. This suggests that this nucleotide binding pocket preferentially binds another nucleoside-phosphate species than the other nucleotide binding pockets [61, 62, 68].

As a consequence of the “split washer” structure of the ATPase hexamer, the aromatic hydrophobic “pore loops” of the Rpt subunits, which point to the central axis of the ATPase channel, are arranged similarly to a spiral staircase, except for the “pore loop” of Rpt6, which is at the same level as the pore loop of Rpt3. The “pore loops” are suggested to mediate translocation of substrate through the ATPase channel [62].

Four of the six Rpt proteins (Rpt1,2,3,5) contain HbYX motifs, which appear to facilitate binding of the 19S RP to the 20S core particle and modulate 20S proteasome activity [69]. The C-terminal HbYX motifs of Rpt3 and Rpt5 extend into the  $\alpha$ -pockets of the 20S core particle and were resolved in all recently published 26S proteasome structures [61–63]. By contrast, interactions of HbYX motifs encoded in Rpt1 and Rpt2 with the 20S core proteasome could not be observed .

### 1.5.2 19S lid substructure

The 19S lid is composed of six **P**roteasome, **C**OP9 signalosome, translation **I**nitiation factor (**PCI**) containing subunits (Rpn3, 5, 6, 7, 9, and 12), two **M**pr1, **P**ad1, amino-**N**-terminal (**MPN**) subunits (Rpn8 and Rpn11) and Sem1, which seem to play a role in ubiquitin recognition (Tab. B.5) [27, 70]. The lid covers the channel of the AAA+ ATPase and is structurally related to the COP9 signalosome and eIF3 [71, 72]. The PCI containing subunits are interconnected via a horseshoe-like structure formed by their PCI domains [60]. The Rpn8/Rpn11 heterodimer links the horseshoe-like structure of the 19S lid with the helical bundle formed by C-terminal helices of the Rpt subunits [60]. Rpn11 is a Zn<sup>2+</sup>-dependent DUB [73]. The DUB activity of Rpn11 is inhibited through an interaction with Rpn5 in the isolated 19S lid structure. Upon binding of the 19S lid to the 26S holocomplex, structural rearrangement of the complex repositions Rpn11 over the substrate-translocation channel to reportedly enable deubiquitylation of the substrate [66]. Besides Rpn11, no other component of the lid shows enzymatic activity. New features observed in the high resolution human 26S proteasome structures include

the protrusion of the C-terminus of Rpn3 reaching into the substrate channel and the connection between Rpn1 and Rpn2 [61, 62].

Three different states (s1, s2, s3) for the 19S lid with respect to the 20S core proteasome were reported to exist for the yeast 26S proteasome. The s1 state represents the ground state in which the 26S proteasome is waiting for substrate. The s2 state depicts the positioning of the substrate to the mouth of the ATPase, while the s3 state exemplifies the translocation and degradation state of substrate [64]. The presence of ATP in the buffer results in state s1. Addition of ATP $\gamma$  or polyubiquitylated model substrates induce state s3 [64]. In the s1 state, the channel of the ATPase is not aligned with the channel of the 20S core proteasome and Rpn6 links the  $\alpha$ 2-subunit with the 19S RP. The channel of the ATPase coaxially aligns with the channel of the 20S core proteasome in the s3 state. Moreover, the Rpn8/11 heterodimer moves to the axis of the ATPase central ring and Rpn6 detaches from the  $\alpha$ 2-subunit. The ubiquitin receptor Rpn10 shifts closer to the mouth of the ATPase and Rpn1 turns counter-clockwise to enable binding of proteasome interacting proteins (PIP) or substrates [74]. The s2 state depicts an intermediate state between s1 and s3, in which the ATPase channel starts aligning with the 20S core channel and the Rpn8/11 dimer is rotating by 25° around the axis of the OB ring.

Structural analyses of the 26S proteasome provide first insights into the functional cycle of the proteasome. In all published structures, the subunit Rpn13 could not be resolved [61–63]. Moreover, the interpretation of the structures suffer from poorly resolved regions especially in the 19S regulatory particle, due to conformational flexibility, transiently interacting DUBs and missing substrate interactions. The low resolution and the missing subunit are two major factors limiting the understanding of this molecular machine.

### 1.5.3 19S lid assembly

The 19S base and the 19S lid structure of the 19S RP are separately assembled. Assembly of the 19S base differs among the species. The following model was proposed for the assembly of the mammalian 19S lid base: The assembly of the Rpt subunits is assisted by regulatory particle assembly chaperones (RACs), Rpn14, p28, p27 and S5b, since the Rpt subunits do not contain any information about their correct assembly in their primary sequence [75]. First, the Rpt dimer pair chaperone complexes p28-Rpt3-Rpt6-Rpn14 and S5b-Rpt1-Rpt2-Rpn1 are formed. The two modules form a complex under the release of S5b. Secondly, Rpn2, Rpn13 and the remaining Rpt dimer Rpt4-Rpt5-p27 bind to the preformed complex to build up the final 19S base hexameric structure. The chaperone p27 leaves the complex upon binding of Rpt4-Rpt5. The exact sequence of the final step

is not yet known[40].

At the 19S base, the assembly of the 19S lid takes place. The most comprehensive 19S lid assembly study was performed for the 19S lid components from *Saccharomyces cerevisiae* [76]. The 19S lid assembly functions without the help of assembly factors or chaperones. Rpn11 and Rpn8 form a dimer through their MPN domains. The formed helical bundle provides the platform for interactions of the C-termini of Rpn5 and Rpn9. Then, Rpn6 is attached to the Rpn5/8/9/11 complex via its C-terminus helix to the existing helical bundle. The 19S lid is completed by sequential addition of the heterodimer Rpn3/Rpn7 and Rpn12 with their C-termini to the helical bundle. In the assembled 19S lid, the PCI domains of Rpn3/5/6/7/9/12 form a horseshoe like motif. The fully assembled 19S lid binds to the 19S base to finalize the assembly of the 19S RP [76].

## 1.6 ATP independent degradation

Aging, mutations, or oxidation results in partial or complete unfolding of proteins. Such proteins can be degraded by the 20S proteasome in an ubiquitin independent manner [77]. Moreover, proteins carrying large unstructured segments, known as intrinsically disordered proteins (IDPs), and totally disordered sequences are substrates of the 20S proteasome core particle. These proteins include key regulatory and signalling proteins involved in cell cycle progression, cellular growth control and oncogenesis [78]. ATP independent degradation is assisted by the 11S activators and PA200, which bind to the  $\alpha$ -ring of the 20S proteasome and modulate its activity [79–81].

### 1.6.1 11S activator

Two different 11S activators, PA28 $\alpha\beta$  and PA28 $\gamma$ , are expressed in mammalian cells. Binding of 11S activators to the 20S proteasome results in an increase of its activity measured by fluorogenic substrates [80].

The 11S activator PA28 $\alpha\beta$  can be found in the cytosol. The expression of PA28 $\alpha\beta$  is synchronously upregulated alongside the immunoproteasome, MHC class I receptors and transporters associated with antigen processing. Together, this suggests that PA28 $\alpha\beta$  is involved in processing of antigen for loading into MHC-I. However, a more detailed understanding of its cellular role is still lacking [82].

The other 11S activator, PA28 $\gamma$ , resides exclusively in the nucleus. Degradation of nuclear substrates associated with regulation of cell-cycle progression and intra-nuclear dynamics

are reported to be regulated by PA28 $\gamma$ [83].

### 1.6.1.1 11S activator structures

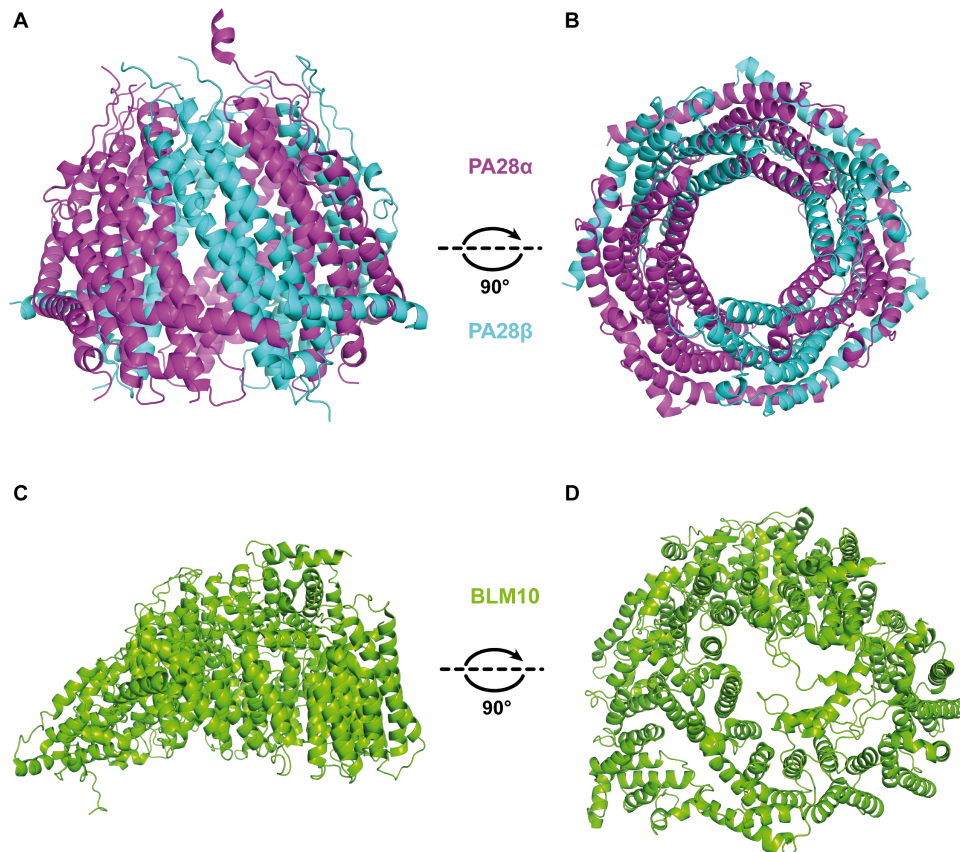
The knowledge about structures of the 11S activators is based on the PA28 $\alpha$  homoheptamer and the PA28 $\alpha\beta$  heteroheptamer assembly [82, 84]. The PA28 $\alpha$  subunit encodes four helices in its sequence that form a bundle along its entire length. The N-terminus points away from the C-terminus, as helix 1 is kinked by 45° at position Pro34 enabling interactions of helix 1 with helix 2. Seven of the  $\alpha$ -subunits build up a funnel-like structure with an opening of 20 Å at the top and 30 Å at the bottom site, where the C-termini carrying the HbYX motifs, which are essential for binding to the 20S proteasome are located. The N-terminal part of helix 1 forms a ring around the funnel by interacting with helix 4 of the adjacent subunit [84]. In the cell, the PA28 $\alpha$  homoheptamer is physiologically irrelevant since the PA28 $\alpha\beta$  heteroheptamer comprised of four PA28 $\alpha$ - subunits and three PA28 $\beta$ -subunits is the active species (Fig. 1.7). Thermal stability assays reveal a higher thermal stability of PA28 $\alpha\beta$  compared to the homoheptamers of PA28 $\alpha$  or PA28 $\beta$  [82]. Every PA28 $\beta$ -subunit is surrounded by two neighboring PA28 $\alpha$  subunits. Apart from a 13 amino acid longer insertion loop between helix 2 and 3, the fold of both subunits is almost identical. The overall structure of the heteroheptamer resembles the structure observed for the PA28 $\alpha$  heptamer.

A structurally similar 11S activator known as PA26 that is found in *Trypanosoma brucei* can increase the activity of the 20S proteasome from several species [85]. Like PA28 $\alpha\beta$  and PA28 $\gamma$ , seven PA26 subunits, composed out of four helices each, form a funnel-like heptameric structure, even though its sequence is highly divergent. Major structural differences are the presence of an internal loop in helix in PA26, the length of the loop connecting the helices, the missing kink in helix 1 and the sandwich-positioning of helix 1 between helix 2 and helix 4. So far, all structural 20S proteasome-11S activator interactions studies were performed with PA26 [86].

### 1.6.1.2 11S activator 20S proteasome interaction

Structural investigations of the interactions between the 11S activator and the 20S proteasome are based on PA26-y20S complexes and PA26-*Thermoplasma acidophilum* 20S proteasome (T20S) complex crystal studies [85–87].

The C-termini of PA26 and the activation loop located between helix 2 and 3 mediate binding of PA26 to the 20S core proteasome and modulate its activity [88, 89]. The C-termini anchor PA26 into the pockets between the  $\alpha$ -subunits of the 20S proteasome.



**Fig. 1.7: Ribbon representations of the 11S activator PA28 $\alpha\beta$  and BLM10.** A + B: Side view and top view of the PA28 $\alpha\beta$  complex is shown. The complex is composed of four  $\alpha$ -subunits (pink) and three  $\beta$ -subunits (blue)(PDB: 5MX5). C + D: BLM10 structure is illustrated as ribbon representation in a top view and side view orientation (PDB: 4V70).

There they form main-chain — main-chain hydrogen bonds. The 11S activator C-termini differ significantly in distinct species, which might be explained by main-chain — main-chain bonds mediating binding specificities [86]. A salt bridge is established between the C-terminal carboxylate with the Lys66 residue, which is highly conserved. Of the seven  $\alpha$ -subunits, six in yeast and five in humans contain a lysine at position 66. The importance of the lysine is demonstrated when it was shown that mutation of Lys66 to alanine or serine abolishes binding of PA26 [86]. Binding of PA26 to the 20S proteasome leads to repositioning of the proline at position 17 located in  $\alpha$ -subunits induced by Glu102 of the PA26 activation loop. The movement of Pro17 destabilizes hydrogen bonds of the closed gate conformation resulting in widening of the 20S proteasome entrance pore by circular alignment of the N-termini. This conformation is described as an "open gate" in the literature [86]. Structural and sequence alignments of PA26 with PA28 subunits reveals a missing Glu102 in the PA28 activation loop. The Asp144, conserved in humans, is likely to have the same functional role as the Glu102. The relocation of Pro17 enables the formation of hydrogen bonds between the conserved residues Tyr8, Asp9 and Tyr26



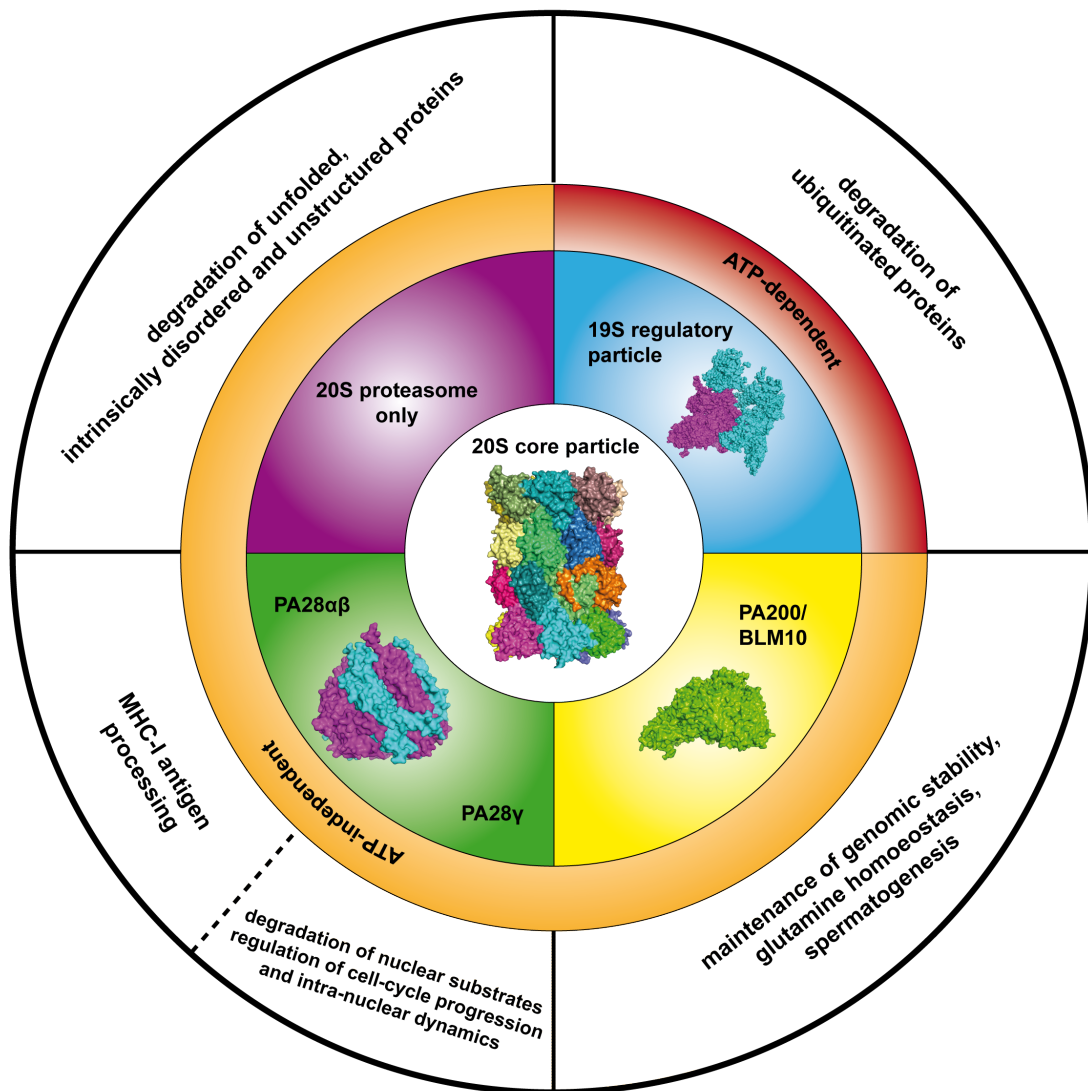
of the  $\alpha$ -subunits stabilizing the open gate conformation [86, 87].

### 1.6.2 PA200 activator

Like PA28 $\gamma$ , PA200 (BLM10: yeast homolog) is constitutively expressed and mainly resides in the nucleus [90]. PA200 functions to maintain genomic stability, glutamine homeostasis and is necessary for spermatogenesis [81]. Sequence alignments of yeast and mammalian PA200 show an identity of less than 20% indicating that this protein is not well-conserved [91]. Similar to the 11S activator, PA200 can stimulate the degradation of fluorogenic model substrates, but not of proteins [92]. How PA200 functions is not well understood. However, structural analyses of the BLM10-y20S complex provides valuable information about its structure and how it docks to the 20S core proteasome.

PA200 is a 210 kDa monomer (BLM10: 250 kDa). The N-terminus of BLM10 is disordered until residue Thr79. Before the first **H**untingtin, **e**longation factor 3, **P**R/**A** subunit of protein phosphatase 2A and the **T**OR lipid kinase (HEAT) repeat begins, three helices and loops are located between Thr79 and His133. In total, BLM10 folds into 32 HEAT repeats. One HEAT repeat is composed of two helices connected by a turn and a linker to the adjacent repeat. The length of the helices ranges from 8 to 35 residues, from 2 to 87 residues for the turns and from 1 to 88 residues for the linkers. The overall structure is reminiscent of a dome composed of a left handed 1.5 spiral turn of the HEAT repeats. An area of 11000 Å<sup>3</sup> is enclosed under the dome (Fig. 1.7). Unlike the 11S activator, BLM10 does not provide an obvious substrate channel, as its largest opening is 13 Å × 22 Å, measured from the atomic nuclei. Notably, this opening could be occluded by residues not visible in the structure.

Since PA200 is a monomer, it has only one C-terminus, which anchors the activator to the 20S proteasome core. The C-terminus binds in the pocket between  $\alpha$ 5- and  $\alpha$ 6-subunit. Hydrogen bonds between the main chains and the salt bridge with Lys66 keep the C-terminus in place. In contrast to the 11S activators, the  $\alpha$ 5/Pro17 turn movement is stabilized by the hydrogen bond interaction of the BLM10/Tyr2142 with the carbonyl group of  $\alpha$ 5/Gly19. This structural rearrangement is not sufficient to trigger the open conformation of the gate, as BLM 10 does not interact with the Pro17 turn of  $\alpha$ 2,  $\alpha$ 3, and  $\alpha$ 4, but resembles an intermediate conformation between the closed state and the open state observed for PA26 binding. Moreover, the formation of the stabilized Tyr8 and Asp9 configuration in  $\alpha$ 5 and  $\alpha$ 6 is hindered by the displacement of  $\alpha$ 5/Asp6 and  $\alpha$ 7/Tyr8 by BLM10 residues Lys2112 and Asp676 [92].



**Fig. 1.8: Overview of proteasome functions.** 20 proteasome can function in an ATP dependent and ATP independent manner. Binding of different adapters to the 20S proteasome modulates its function in the cell.

### 1.6.3 Proteasome inhibition

Beginning with the discovery of the 20S/26S proteasome complex, there was a great interest in the availability of compounds to inhibit the proteasome. Such compounds could be used to investigate the functional role of the proteasome in the cell, since the 20S proteasome has essential functions in protein degradation and antigen presentation. First, inhibitors like Leupeptin, which inhibit proteases like papain or cathepsin B, were used [54]. As these inhibitors did not only inhibit the proteasome but also other proteases, the demand for specific inhibitors provoked the discovery of different small molecules targeting the proteasome exclusively. Lactacystin, purified from *Streptomyces*, was the first reported proteasome specific inhibitor shown to covalently bind to the  $\beta 5$ -subunit of the 20S proteasome [93, 94].

Currently, several 20S proteasome specific inhibitors are available. The vast majority of inhibitors target the 20S proteasome. The 20S proteasome can be classified into covalent and non-covalent inhibitors (Fig. 1.9; 1.10). Only five inhibitors impede proteasome function by binding to the 19S RP (Fig. 1.11). Some inhibitors can be used for proteolytic-site specific inhibition while others do not show proteolytic site specificity. Besides synthetic proteasome inhibitors, some inhibitors are obtained from natural sources and were the templates for inhibitor optimization [95].

### 1.6.3.1 Covalent 20S proteasome inhibitors

Inhibitors modifying the 20S proteasome active sites all possess an electrophilic group, which interacts with the N-terminal threonine of the proteolytic sites. Covalent inhibitors are classified according to their electrophilic group. Excluding  $\beta$ -lactones and oxatiazol-2-ones, which also inhibit 20S proteasomes of bacterial origin, all 20S proteasome inhibitors have a peptide backbone consisting of two to four amino acids mimicking an authentic substrate (Fig. 1.9).

Peptide aldehydes were the first inhibitors used for functional analysis. Aldehydes form a reversible hemiacetal linkage with the hydroxyl of the threonine. The cross reactivity of the aldehyde inhibitors with other proteases and their sensitivity to oxidation restricts their clinical application [96].

Replacement of the aldehyde group with a boronate-moiety, such as Bortezomib and MLN2238, greatly enhanced the potency of inhibitors making them suitable as antineoplastic drugs used in the clinic. Boronates form a tetrahedral linkage with the hydroxyl group of the Thr1. This tetrahedral state is stabilized by a hydrogen bond between the N-terminal amino group of threonine and hydroxyl group of the boron. The higher affinity of boronic inhibitors to the 20S proteasome can be explained by additional stabilization by the amino group of the Thr1 which is lacking in serine proteases. The slow off-rate of the reversible boron-hydroxyl linkage prevents the dissociation of boronic inhibitors from the proteolytic sites during the time-span of a standard cell culture experiment or cancer treatment. The importance of this inhibitor is exemplified by development of the inhibitor CEP-18770, which is tested in clinical trials [97].

$\alpha,\beta$ -epoxyketones isolated from *Streptomyces hygroscopicus* and *Actinomyces* are the most specific proteasome inhibitor class [98]. The covalent linkage to the N-terminal threonine to proteasome catalytic sites involves both its hydroxyl and amino groups. This requirement for two reactive groups ensures high selectivity for the 20S proteasome active sites. The reaction starts with the nucleophilic attack at the carbonyl of the ketone

forming a tetrahedral intermediate. The reaction is completed with the nucleophilic attack at the  $\alpha$ -carbon of the epoxide resulting in an 1,4-morpholino ring like structure [99]. Several epoxyketone inhibitors were synthesized based on observations of structures from compounds of natural origin. Two inhibitors of this class, Carfilzomib and Oprozomib, are currently tested in clinical trials for the treatment of multiple myeloma [97].

Ketoaldehyde-peptide inhibitors share a similar reaction mechanism with the epoxyketone inhibitors because the hydroxyl group and the amino group of the threonine are involved in the linkage. Similar to epoxyketones, a nucleophilic attack of hydroxyl oxygen of the threonine at the carbonyl group of the ketone takes place. The 5,6-dihydro-2H-1,4-oxazine ring is then closed by the nucleophilic attack of the amino group at the carbonyl of the aldehyde. This two-step reaction mechanism makes the ketoaldehyde inhibitors more specific compared to their aldehyde analogues. The Schiff base residing in the 5,6-dihydro-2H-1,4-oxazine ring makes the covalent linkage reversible. Its reversibility and the high specificity make it an interesting compound for further inhibitor developments [100, 101].

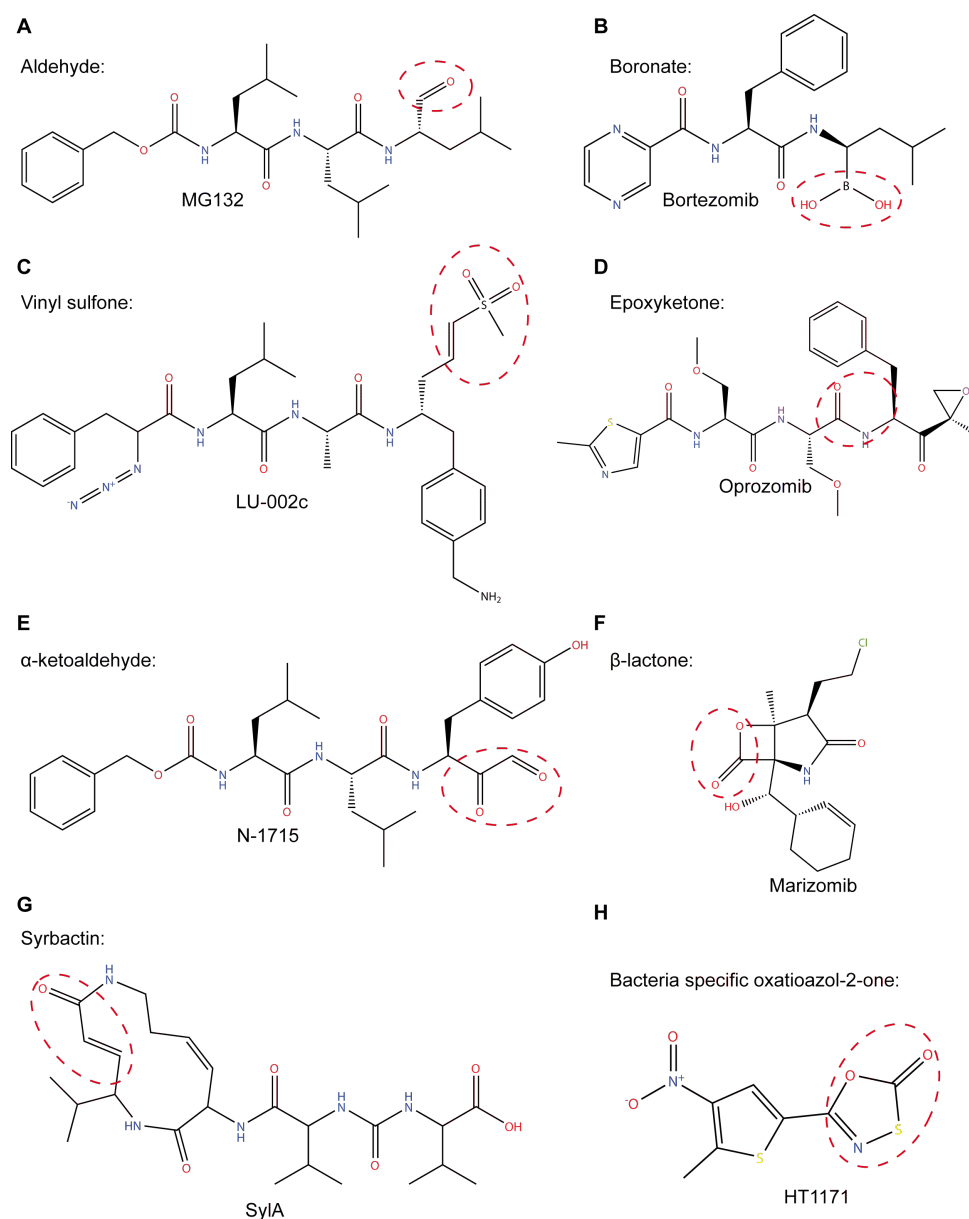
Vinyl sulfone peptide inhibitors were originally developed to inhibit cysteine proteases [102]. However, vinyl sulfones are able to block all proteolytic sites of the 20S proteasome. The hydroxyl group reacts with the double bond of vinyl sulfone via a Michael addition [103]. Vinyl sulfones are less specific and less potent compared to epoxyketones, but easier to synthesize [95].

The Syrbactin 20S proteasome inhibitor class possesses a 12-membered lactam core as an electrophilic group attached to a peptide backbone. An irreversible ether bond is formed according to a 1,4-Michael addition between the hydroxyl threonine and the double bond of the lactam ring [104]. The Syrbactin class was founded by the two compounds syringolin A and glidobactin A isolated from *Pseudomonas syringae pv syringae* and an unknown species of the order *Burkholderiales*, respectively. The chemical synthesis of syringolin A was achieved in 2009 [105].

$\beta$ -lactone inhibitors are non-peptide inhibitors. They were the first specific proteasome inhibitors reported. Since then, several compounds belonging to this class were synthesized or found in other organisms. Although  $\beta$ -lactones were initially thought to be proteasome specific inhibitors later experiments revealed that lactones also inhibit serine proteases, making them more specific than aldehyde inhibitors but less specific than epoxyketones.  $\beta$ -lactones are covalently attached to the Thr1 hydroxyl by esterification [106]. This linkage is reversible as it can slowly hydrolyze. The hydrolysis rate for the  $\beta$ -lactone Marizomib is decreased by introducing an ethylchlorid group, which reacts with the opened  $\beta$ -lactone to form a tetrahydrofuran ring. The ring formation seems to protect the ester from hydrolysis [107].

An interesting proteasome inhibitor class with potential therapeutic application are the Oxatiazol-2-ones, as this class only inhibits mycobacterium proteasomes. Binding of the

inhibitor induces structural rearrangements in the  $\beta$ -subunit. Reaction with Oxataiazol-2-ones results in an irreversible cyclocarbonylation of the reactive threonine [108].



**Fig. 1.9: Overview of different covalent 20S proteasome inhibitor classes.** For each inhibitor class an exemplary molecule is shown. The respective reactive groups of the inhibitor classes are encircled with a dashed red line.

### 1.6.3.2 Subunit-specific active probes

Great efforts have been undertaken to develop inhibitors specifically targeting the three different proteolytic active sites present in the 20S proteasome. Site-specific inhibition allows monitoring of the catalytic activity of the individual active sites and the ability to analyse the significance of each catalytic site for cell line survival [109]. As the

epoxyketones are the most specific 20S proteasome inhibitors discovered so far, most of the site-specific inhibitors are based on epoxyketones inhibitors. Interestingly, the specificity for a chymotryptic-like site could be increased by replacing the epoxyketone with a vinyl sulfone group [110]. Site specificity of the inhibitors was achieved by permutation of the amino acid in P1, P2, and P3 position with natural and unnatural amino acids in addition to different heterocycles [111].

Recently, activity-based probes were published for the different proteolytic-active sites in  $\beta$ 1,  $\beta$ 2 and  $\beta$ 5. These probes are linked to a fluorophore N-terminally to enable the visualization of proteasome activity in different cell lines by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). This approach makes it possible to visualize the specificities of distinct inhibitors to the various catalytic sites *in vivo* for several cell lines. These fluorescent labeled active probes drove the development of immunoproteasome specific inhibitors, which are of exceptional importance to better decipher its functional role in the cell [112].

### 1.6.3.3 Non-covalent 20S proteasome inhibitors

The non-covalent inhibitors can be grouped into three major classes – cyclical peptides, capped dipeptides and non-peptide inhibitors (Fig. 1.10 ).

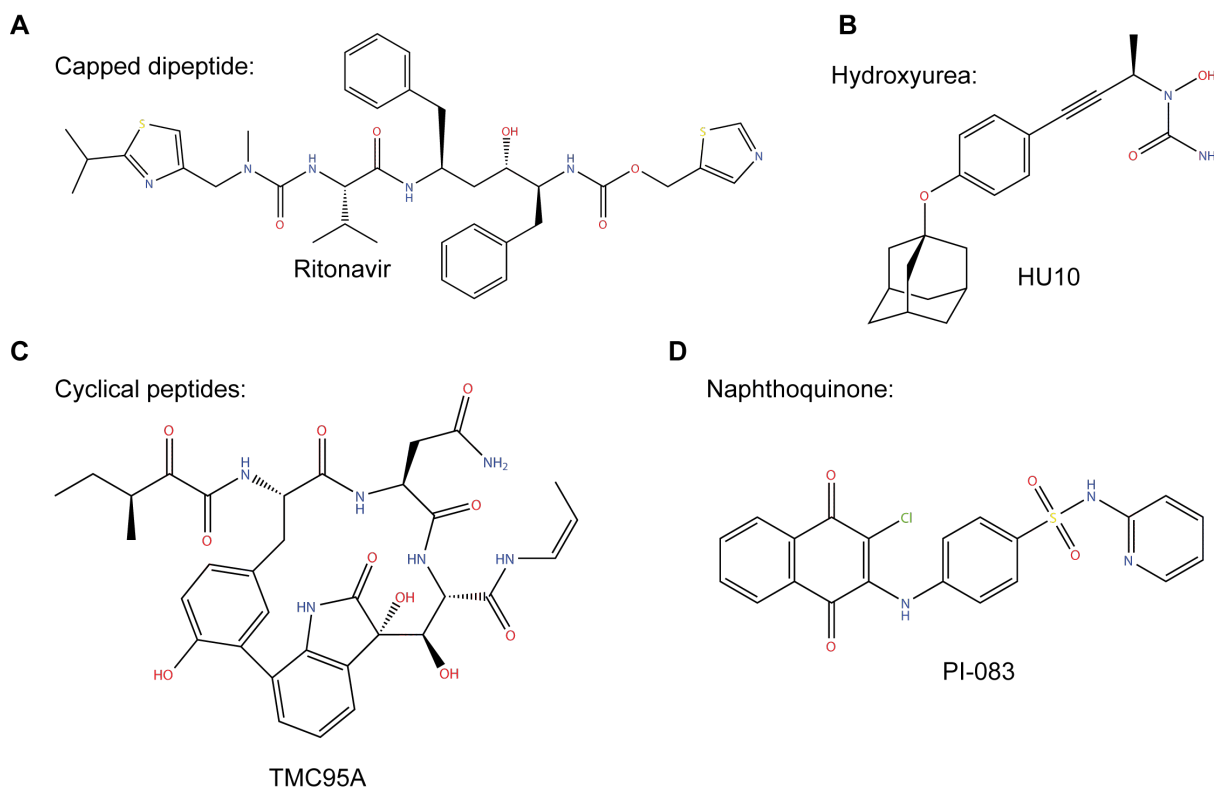
Cyclical peptides like TMC-95 are compounds of natural origin. They prevent substrate binding by occluding the entrance to the proteolytic sites. X-ray analysis of TMC-95 revealed hydrogen bonds between the backbone of the inhibitor and the main chain that result in a stabilization of the inhibitor in the active site [113]. Argyrin A is another member of this class that was shown to have antitumor activity. This demonstrates the potential for this inhibitor class to specifically inhibit the proteasome [95].

The capped dipeptide class includes two residue-peptides carrying a cap at the N- and the C-terminus. These inhibitors are quite effective since they have observed  $K_i$  concentrations for the  $\beta$ 5-subunits in the 10 nM range. The caps are involved in binding by interactions of the C-terminal cap with the S1 pocket and the N-terminal cap entering the S4 pocket of the active site. S2 and S3 pockets are occupied by the two residues of the dipeptide [114].

Non-covalent inhibitors not belonging to the other two classes are PI-083 and the hydroxyurea inhibitor. Both compounds were identified by screens of chemical libraries with purified proteasomes. PI-083 shows inhibitory activity on all active site, but only in transformed cells [115].

The hydroxyurea inhibitor binds to the chymotryptic-like site with an affinity of  $K_i = 30$  nM. The bulky parts of the inhibitor populate the S1 and S3 pocket of the  $\beta$ 5 proteolytic

site [116].



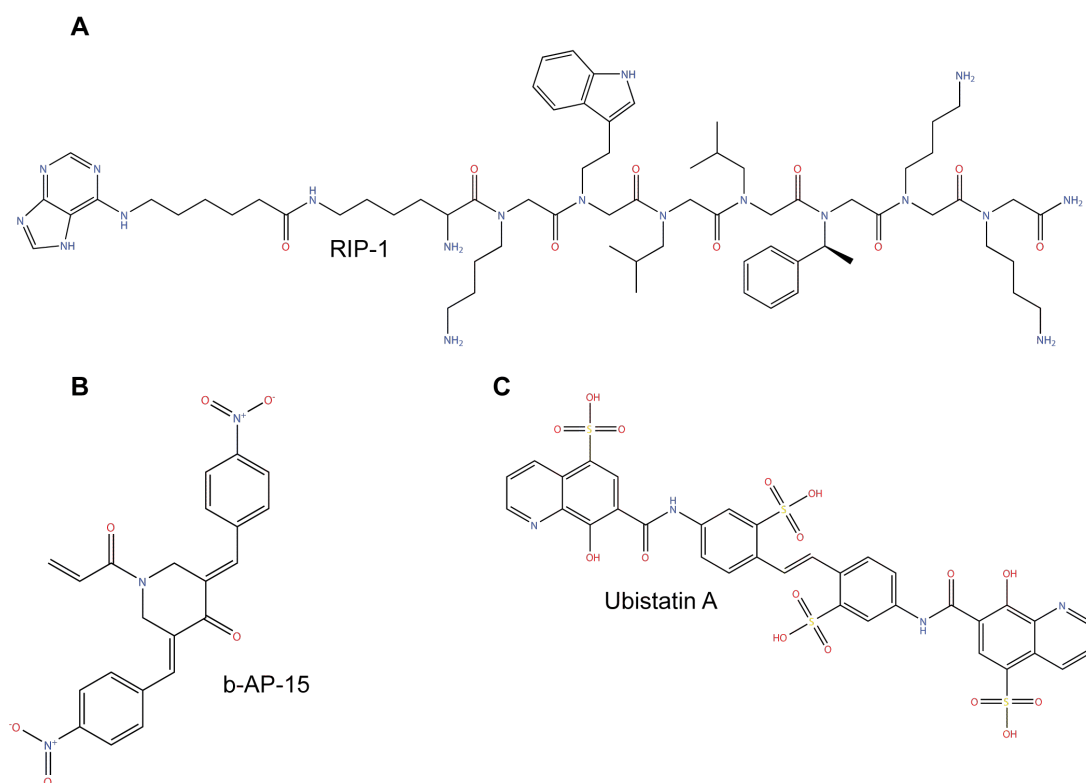
**Fig. 1.10: Overview of different non-covalent 20S proteasome inhibitor classes.** For each inhibitor class an exemplary molecule is shown.

#### 1.6.3.4 19S regulatory particle inhibitors

Inhibitors were also developed to interfere with either the ATPase activity, ubiquitin binding, or deubiquitylation of the 19S lid to compromise the 19S RP function. Two inhibitors, ubistatin A and ubistatin B, block the binding of ubiquitylated substrates to the 19S RP (Fig. 1.11).

RIP-1 inhibits the 26S proteasome by targeting Rpt4 of the 19S ATPase shown by cross-linking studies [117]. Potential off-target interaction with other ATPases limits its application.

Inhibition of the deubiquitylating enzymes Usp14 and Uch-L5 with b-AP-15 has the same effects as 20S proteasome inhibition. In contrast, the inhibition of Usp14 by IU1 stimulates the protein degradation through the proteasome-independent degradation pathway. Knowledge about inhibition chemistry and further development of new 19S RP inhibitors is hampered by missing high resolution structural data [95].



**Fig. 1.11: Overview of different 19S regulatory particle inhibitors.** Available 19S regulatory particle inhibitors are depicted.

### 1.6.3.5 Relevance of proteasome inhibitors as drug

Bortezomib was the first inhibitor showing cytotoxic activity in cell lines derived from cancer cells. Animal studies verified the efficacy of Bortezomib against xenograft tumors. Good responses of patients with multiple myeloma in clinical trials resulted a Food and Drug Administration (FDA) approval as drug against relapsed and refractory disease. Since 2008, it is used as front-line treatment for multiple myeloma patients [118].

Proteasome inhibition leads to apoptosis of multiple myeloma cells, as these cells synthesize and secrete immunoglobulin G (IgG) or immunoglobulin A (IgA) in abundant amounts and require the proteasome to cope with protein over-production. The synthesis of IgG is highly error-prone, as its structure is complex. Misfolded IgG needs to be removed through the endoplasmic reticulum degradation pathway. Consequently, proteasomes must degrade a large amount of defective protein. Therefore, proteasome inhibitor concentrations, which are tolerable by healthy cells, lead to apoptosis of transformed cells. Moreover, the proapoptotic BH3 member Phorbol-12-myristate-13-acetate-induced protein 1 (NOXA) is upregulated under endoplasmic reticulum (ER) stress [95].

So far, Bortezomib, Ixazomib and Carfilzomib are FDA-approved drugs for the treatment of multiple myeloma or mantle-cell lymphoma. Marizomib and Oprozomib are currently being tested in clinical trials [97].



### 1.6.4 Aims of the thesis

Protein degradation is essential for the cell to control protein homeostasis. Malfunctioning, mistranslated and broken proteins must be removed to keep a cell healthy. Furthermore, the cell cycle is regulated by temporal protein degradation. Besides the lysosomal protein degradation pathway, the ubiquitin-proteasome pathway is the most important one. The degradation of proteins subjected to the ubiquitin proteasome degradation pathway is catalysed in the 20S core proteasome, which is part of the 26S proteasome holocomplex. As the protein degradation by the ubiquitin proteasome pathway is of high importance for the cell, a detailed structural analysis of the 20S proteasome is of great interest. Moreover, is the human 20S proteasome known to be a drug target to treat certain types of cancer. The FDA already approved three different compounds targeting the 20S proteasome for the treatment of cancer [118]. A detailed structure of the 20S proteasome provides insight into the underlying principles of 20S proteasome function. Binding sites and reaction mechanisms of potential 20S proteasome inhibitors can be revealed by a detailed 20S proteasome structure. A robust purification and crystallization workflow for the 20S proteasome can be used to search for new inhibitors by high-throughput screening of small molecule libraries.

The first crystal structure of the 20S proteasome was published for *Thermoplasma acidophilum* at a resolution of 3.4 Å in 1995. The addition of an 11S activator PA26 mutant improved the resolution for the archaeal 20S proteasome from *Thermoplasma acidophilum* to a resolution of 1.9 Å [86]. The best diffracting eukaryotic 20S proteasome crystals are obtained from *Saccharomyces cerevisiae* 20S proteasomes. Anisotropic diffraction limits their resolution to 2.4 Å [31]. 20S proteasomes from *Saccharomyces cerevisiae* are widely used for functional and mechanistic studies. Reasons are the availability of robust purification and crystallization protocols [41]. Therefore, elucidation of 20S proteasome mechanistic and inhibition mechanisms of 20S proteasome inhibitors are mostly done by using yeast 20S proteasomes. Consequently, the major part of 20S proteasome structures deposited in the Protein Data Bank (PDB) originate from yeast. The structures published to date provided information about the general architecture and allowed the accurate modeling of most of the amino acids, composing the 20S proteasome. Moreover, binding sites and reaction mechanisms of several 20S proteasome inhibitors have been revealed.

Attempts to solve the human proteasome structure with cryo-EM resulted in a 3.5 Å structure [119]. At this resolution it is possible to detect binding sites of an inhibitor. However, the resolution is not sufficient to understand the underlying chemistry of inhibitor binding. The currently highest resolved human 20S proteasome structure was solved by X-ray crystallography at a resolution of 2.6 Å [120]. The unequal distribution of the water molecules in the structure and partly missing peptide backbone electron

densities need to be taken in account for the interpretation of this structure.

The lack of high resolution 20S proteasome structures beyond 2.0 Å are the motivation of this work. The aim was to focus on improving the resolution limit of the 20S proteasome. In this study robust purification and crystallization protocols were established for the human 20S proteasome. HeLa cells were chosen as a source for human 20S proteasomes. The sample quality of the human 20S proteasomes was significantly improved by biochemical optimization and the introduction of a new purification protocol. The newly developed chromatography-free purification protocol resulted in high yields of purified 20S proteasome with superior sample quality. The large amounts of purified 20S proteasomes and the high sample quality allowed the crystallization of the human 20S proteasome in a large scale manner.

The diffraction limit of the crystals was increased by employing post-crystallization treatments. The set-up of a purification, crystallization and inhibitor soaking pipeline to generate routinely high diffracting human 20S proteasome crystals made it possible to reinvestigate different 20S proteasome inhibitor classes in the human system.

The practicability to adapt the newly developed purification protocol of the human 20S proteasomes to other organisms was tested by successful purifications of 20S proteasomes from *Saccharomyces cerevisiae*, *Drosophila melanogaster* and *Thermoplasma acidophilum*. Crystals were obtained for all of these purified samples and structures were solved for all of them. The different structures were examined for their inherent properties, potentially resulting in even higher diffracting crystals.

Not only covalent inhibitors were examined, but also binding sites of an activator peptide constructed based on the C-terminal part of BLM10 were determined. Proposed structural rearrangements induced upon binding of the activator were analyzed.

Besides the 20S proteasome, the 26S proteasome could be purified with a modified 20S proteasome purification protocol. With the 26S proteasome sample it was possible to analyze the global conformational changes of the 26S proteasome upon inhibitor binding by cryo-EM. Valuable information about a potential feedback regulation between the proteolytic active site present in the  $\beta$ -rings and the 19S RP were deduced from the calculated energy landscapes.

In summary, the improved sample preparation of the proteasome enabled to analyze the structure and dynamics of the proteasome in more detail resulting in additional valuable knowledge for understanding its molecular functions.



## 2 Material and Methods

### 2.1 Materials

#### 2.1.1 Special Equipment

Table 2.1: Equipment used for experiments

Machine/Equipment	Manufacturer
ActiLoops	Molecular dimensions
Balances	Sartorius
Centrifuge LYNX 6000	Thermo Scientific
Copper EM grids	Plano
Crychem M Plate, 24-well	Hampton Research
Crystal Clear Sealing Tape	Jena Bioscience
Crystalgen Plate 24-well, pregreased	Jena Bioscience
EmulsiFlex-C3	Avestin, Inc.
Fiberlite rotor F14, F21, F35L	Thermo Scientific
Fluoromax-4 fluorescence spectrophotometer	Horiba Scientific
Gradient Master ip	Biocomp
Incubator Eco-Line E100	RUMED
Incubator Multitron Pro Shaker	Infors HT
Leica EM GP	Leica
LithoLoops	Molecular dimensions
Micro-Ultracentrifuge MX 150+	Thermo Scientific
MicroLoops	MiTeGen
Orbital Shaker RS-S20	Phoenix Instrument
Photometer Lambda Bio X	Perkin Elmer
Plate Sealer ALPS 3000	Thermo Scientific
Pipetting Robot Microlab Star LET	Hamilton
Quantifoil grid R3.5/1 Cu 200 mesh	Quantifoil Micro Tools GmbH

Rotor TH660	Thermo Scientific
Rotor S140AT	Thermo Scientific
Rotor SW40	Beckmann Coulter
Siliconized Glass Cover Slides	Hampton Research
Stereo microscopes M125	Leica
TEM CM200 FEG	Philips
TEM Titan Krios	FEI
Ultra Centrifugal Mill ZM 200	Retsch®
Ultracentrifuge WX Ultra 80, 90	Thermo Scientific
Vitrobot	FEI
Zebra Micro Desalt Spin Column	Thermo Scientific

## 2.1.2 Chemicals

**Table 2.2: Chemicals used for experiments**

Chemicals	Supplier
2-Methyl-2,4-pentanediol (MPD)	Sigma-Aldrich
Acetic acid	Merck
Amido Black 10 B	SERVA
	Electrophoresis
$\beta$ -Glycerophosphate disodium salt hydrate	Sigma-Aldrich
Benzamidine Hydrochloride hydrate	Sigma-Aldrich
Bis(sulfosuccinimidyl)suberate BS3	Thermo Scientific
Bis(2-hydroxyethyl)amino- tris(hydroxymethyl)methane BisTris	Sigma-Aldrich
Bradford Assay reagent	BioRad
Coomassie Brilliant Blue R-250	BioRad
Dimethyl sulfoxide DMSO	AppliChem
Dinatriumhydrogen phosphate ( $\text{Na}_2\text{HPO}_4$ )	Merck
Dipotassium phosphate ( $\text{K}_2\text{HPO}_4$ )	Merck
Dithiothreitol DTT	Roth
Ethanol	Merck
Ethylenediaminetetraacetic acid EDTA	Merck
Glucose	Merck

Glutaraldehyde	Science Services
Glycerol	Merck
Hydrochlorid acid (HCl)	Merck
Iodoacetamide	Sigma-Aldrich
Isopropyl $\beta$ -D-1-thiogalactopyranoside IPTG	Sigma-Aldrich
Isopropanol	Merck
Lauryl maltose neopentyl glycol LMNG	Anatrace
Magnesium chloride (MgCl <sub>2</sub> )	Merck
Magnesium acetate (Mg(CH <sub>3</sub> COO) <sub>2</sub> )	Merck
2-( <i>N</i> -morpholino)ethanesulfonic acid (MES)	Sigma-Aldrich
N-ethylmaleimide	Sigma-Aldrich
Octyl glucose neopentyl glycol OGNG	Anatrace
Polyethylene glycol PEG 400	Sigma-Aldrich
PEG 3350	Sigma-Aldrich
Polyethylenimine PEI	Sigma-Aldrich
Phenylmethane sulfonyl fluoride PMSF	Sigma-Aldrich
Potassium chloride (KCl)	Merck
Potassium hydroxide (KOH)	Merck
Potassiumdihydrogen phosphate (KH <sub>2</sub> PO <sub>4</sub> )	Merck
Protease inhibitor tablet , EDTA free	Roche
All Blue Protein Standard	BioRad
Sucrose	Merck
Sodium Aspartate	Sigma-Aldrich
Sodium chloride (NaCl)	Merck
Sodium creatine phosphate	Sigma-Aldrich
Sodium dodecyl sulfate polyacrylamide gel electrophoresis SDS-PAGE running buffer	Roth
Uranyl formate	Custom made

---

Substrates, inhibitors and enzymes	Supplier
Bortezomib	Selleck Chemicals
C-3455	Bachem
C-3900	Bachem
Carfilzomib	Selleck Chemicals

Creatine Kinase	Roche
Delanzomib	Selleck Chemicals
Dihydroeponeycin	APExBIO
Epoxomicin	APExBIO
Ixazomib	Selleck Chemicals
Lysozyme	Sigma-Aldrich
MG115	Selleck Chemicals
MG132	Selleck Chemicals
MG262	Selleck Chemicals
ONX-0914	AdooQ®
Oprozomib	Selleck Chemicals
Suc-LLVY-AMC (N-1715)	Bachem
LU-002c	Custom Synthesis
Z-LLY-Ketoaldehyde	Bachem

---

Cells and Plasmid Vectors	Supplier
BL21 (DE3) <i>E. coli</i>	NEB
pRSET A vector	ThermoFisher Scientific

---

### 2.1.3 Buffers and Media

If not stated otherwise all buffers and media used for the experiments were prepared according to Sambrook and Russell [121].

**Table 2.3: Buffers and Media used for experiments**

Buffer	Concentration	Compound
1x Purification Buffer	0.1 M	BisTris pH 6.5 HCl
	0.05 M	KCl
	0.01 M	MgCl <sub>2</sub>

---

1x Acetate	0.1 M	BisTris pH 6.5 CH <sub>3</sub> COOH
Purification Buffer	0.05 M	K(CH <sub>3</sub> COO)
	0.01 M	Mg(CH <sub>3</sub> COO) <sub>2</sub>
Human & Fruit Fly Crystallization Buffer	0.1 M	BisTris pH 6.5 HCl
	0.2 M	MgCl <sub>2</sub>
	10% (w/v)	PEG 3350
Human & Fruit Fly Stabilization Buffer (Stabi Buffer)	0.1 M	BisTris pH 6.5 HCl
	0.2 M	MgCl <sub>2</sub>
	20% (w/v)	PEG 3350
Human & Fruit Fly Cryo-Buffer	0.1 M	BisTris pH 6.5 HCl
	0.2 M	MgCl <sub>2</sub>
	25% (w/v)	PEG 3350
	20% (v/v)	MPD
Lysis Buffer	0.05 M	Na <sub>2</sub> HPO <sub>4</sub> pH 8.0
	0.3 M	NaCl
Yeast Crystallization Buffer 1	0.1 M	MES pH 6.5
	0.04 M	Mg(CH <sub>3</sub> COO) <sub>2</sub>
	12% (v/v)	MPD
Yeast Cryo-Buffer 1	0.1 M	BisTris pH 6.5 HCl
	0.2 M	MgCl <sub>2</sub>
	40% (v/v)	MPD
Yeast Crystallization Buffer 2	0.1 M	BisTris pH 7.5 HCl
	0.05 - 0.2 M	NaCl
	25 - 30% (w/v)	PEG 400



Yeast Cryo-Buffer 2	0.1 M	BisTris pH 7.5 HCl
	0.2 M	MgCl <sub>2</sub>
	40% (w/v)	PEG 3350
Archaeal Crystallization Buffer	0.1 M	BisTris pH 6.5 HCl
	0.1 - 0.25 M	MgCl <sub>2</sub>
	7 - 10% (w/v)	PEG 3350
Archaeal Stabilization Buffer (Stabi Buffer)	0.1 M	BisTris pH 6.5 HCl
	0.1 - 0.25 M	MgCl <sub>2</sub>
	20% (w/v)	PEG 3350
Archaeal Cryo-Buffer	0.1 M	BisTris pH 6.5 HCl
	0.1 - 0.25 M	MgCl <sub>2</sub>
	25% (w/v)	PEG 3350
	20% (v/v)	MPD
Enzyme Activity Buffer	0.05 M	BisTris pH 6.75 HCl
	0.001 M	EDTA
	0.01% (w/v)	SDS

## 2.1.4 Software

**Table 2.4: Software used for data evaluation**

Software	Source
Amira 4.1	<a href="https://www.fei.com/software">https://www.fei.com/software</a>
Adobe Creative Suite 5	<a href="http://www.adobe.com/de/creativecloud.html">http://www.adobe.com/de/creativecloud.html</a>
CCP4 [122]	<a href="http://www.ccp4.ac.uk">http://www.ccp4.ac.uk</a>
COW suite	Department of Structural Dynamics, Max-Planck-Institute for Biophysical Chemistry

---

COOT [123]	<a href="https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot">https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot</a>
ENDscript 2.0 [30]	<a href="http://escript.ibcp.fr/ESPript/ENDscript/index.php">http://escript.ibcp.fr/ESPript/ENDscript/index.php</a>
Gautomatch	<a href="http://www.mrc-lmb.cam.ac.uk/kzhang">http://www.mrc-lmb.cam.ac.uk/kzhang</a>
Gctf [124]	<a href="http://www.mrc-lmb.cam.ac.uk/kzhang">http://www.mrc-lmb.cam.ac.uk/kzhang</a>
Imagic [125]	<a href="https://www.imagescience.de/imagick_em.html">https://www.imagescience.de/imagick_em.html</a>
JLigand [126]	<a href="http://www.yesbl.york.ac.uk/mxstat/JLigand">http://www.yesbl.york.ac.uk/mxstat/JLigand</a>
KaleidaGraph 4.03	<a href="http://www.synergy.com/wordpress_650164087">http://www.synergy.com/wordpress_650164087</a>
Orca 3.0.3 [127]	<a href="https://orcaforum.cec.mpg.de/">https://orcaforum.cec.mpg.de/</a>
OriginPro 9.1	<a href="http://www.originlab.de/Origin">http://www.originlab.de/Origin</a>
Phenix [128]	<a href="https://www.phenix-online.org">https://www.phenix-online.org</a>
Python 2.7	<a href="https://www.python.org">https://www.python.org</a>
PyMOL	<a href="https://www.pymol.org">https://www.pymol.org</a>
Refmac5 [129]	<a href="https://www2.mrc-lmb.cam.ac.uk/groups/murshudov/content/refmac/refmac.html">https://www2.mrc-lmb.cam.ac.uk/groups/murshudov/content/refmac/refmac.html</a>
RELION 1.2 [130]	<a href="http://www2.mrc-lmb.cam.ac.uk/relion">http://www2.mrc-lmb.cam.ac.uk/relion</a>
ResMap [131]	<a href="http://resmap.sourceforge.net">http://resmap.sourceforge.net</a>
Robetta [132]	<a href="http://rosetta.bakerlab.org/">http://rosetta.bakerlab.org/</a>
Simple PRIME [133]	<a href="http://simplecryoem.com">http://simplecryoem.com</a>
Staraniso Server	<a href="http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi">http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi</a>
UCSF Chimera [134]	<a href="http://www.cgl.ucsf.edu/chimera">http://www.cgl.ucsf.edu/chimera</a>
Unblur [135]	<a href="http://grigoriefflab.janelia.org/unblur">http://grigoriefflab.janelia.org/unblur</a>
XDS Program Package [136]	<a href="http://xds.mpimf-heidelberg.mpg.de">http://xds.mpimf-heidelberg.mpg.de</a>

---

## 2.2 Methods

### 2.2.1 Molecular Biology

When not stated otherwise all Molecular Biology methods were performed according to Sambrook and Russell [121].

## 2.2.2 Protein purification

### 2.2.2.1 ProteoPlex measurement

The ProteoPlex analysis was done according to Chari et. al [137]. First, a protein dilution screen was performed to determine the optimal 20S proteasome concentration for the best signal readout. A concentration of  $0.3 \frac{mg}{ml}$  human 20S proteasome was used for the custom made buffer screen [137].

### 2.2.2.2 *Homo sapiens* 20S proteasome purification

HeLa cells were kindly fermented and provided by Thomas Conrad (Bioreactor Facility, Max Planck Institute for Biophysical Chemistry). HeLa cytoplasmic extract was prepared by hypotonic lysis of HeLa cells according to Dignam et al [138]. HeLa cells used for proteasome purification exceeded a viability of 85%. After hypotonic lysis and centrifugation to collect nuclei, the supernatant (the crude cytoplasmic extract) was centrifuged at  $30'000 \times g$  for 30 min at  $4 \text{ }^{\circ}\text{C}$ , flash frozen in 40 ml aliquots with liquid nitrogen and stored at  $-80 \text{ }^{\circ}\text{C}$  until further use. The S30 HeLa cytoplasmic extract was thawed in a water bath at  $37 \text{ }^{\circ}\text{C}$  and supplemented with purification buffer to 1x concentration from a 10x stock. Subsequently, sucrose powder was added to 20% (w/v), octyl glucose neopentyl glycol (OGNG) from a 10% (w/v) stock solution in water to 0.1% (w/v), Iodacetamide to 10 mM, N-Ethylmaleimide to 10 mM and Benzamidine Chloride to 10 mM. The extract was incubated at  $18 \text{ }^{\circ}\text{C}$  on a magnetic stirrer for 30 min, followed by incubation at  $30 \text{ }^{\circ}\text{C}$  in a water bath for 1 h. The treated extract was cleared by centrifugation at  $100'000 \times g$  for 2 h at  $4 \text{ }^{\circ}\text{C}$ . The supernatant was filtered through three layers each of cheese cloth and miracloth to receive a S100 HeLa cytoplasmic extract. The clarified extract was subjected to differential precipitation with polyethylene glycol (PEG) (number signifies the mean molecular weight of the PEG polymer). PEG 400 was added to a concentration of 20% (v/v) to the S100 HeLa cytoplasmic extract under stirring at  $18 \text{ }^{\circ}\text{C}$ . The mixture was incubated for 20 min. Precipitated proteins were removed by centrifugation with  $30'000 \times g$  for 30 min at  $4 \text{ }^{\circ}\text{C}$ . The supernatant was precipitated by raising the concentration of PEG 400 to 30% (v/v) as described above. The precipitate of this step contains human 20S proteasomes. The precipitate was pelleted by centrifugation at  $30'000 \times g$  for 30 min at  $4 \text{ }^{\circ}\text{C}$ . The pellet was resuspended in purification buffer containing 2% (w/v) sucrose, 10 mM dithiothreitol (DTT) and 0.01% (w/v) lauryl maltose neopentyl glycol (LMNG) on an orbital shaker at  $18 \text{ }^{\circ}\text{C}$ . The resuspended material was loaded onto 10-30% (w/v) sucrose gradients. The sucrose gradient solutions contain sucrose, 1x purification buffer and 10 mM DTT. Gradients were centrifuged with  $270'000 \times g$  for 16 h at  $4 \text{ }^{\circ}\text{C}$ . Fractionation of the gradients was done by hand in 400  $\mu\text{l}$  fractions by starting from the top

of the gradient. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was utilized to check fractions for their 20S proteasomes content [139]. Selected fractions were pooled and precipitated by the addition of 40% (v/v) PEG 400. After centrifugation with 30'000 x g for 20 min, the supernatant was removed and the pelleted precipitate was resuspended in purification buffer containing 5% (w/v) sucrose, 10 mM DTT and 0.01% (w/v) LMNG. The resuspended material was loaded onto linear 10-40% (w/v) sucrose gradients which were prepared as the previous mentioned gradients. Gradients were centrifuged with 284,000 x g for 20 h at 4 °C. Fractions containing 20S proteasomes were identified by SDS-PAGE and precipitated by the addition of 40% (v/v) PEG 400. The concentration of the final sample was determined by the volume of added resuspension buffer, which is composed of 1x purification buffer, 5% (w/v) sucrose, 5 mM DTT and 0.01% LMNG, to the 20S proteasome pellet. Concentrations of 15 mg/ml to 20 mg/ml were obtained. This procedure reproducibly yields 20 mg purified human 20S proteasomes, starting from 300 ml S100 HeLa cytoplasmic extract.

### 2.2.2.3 *Drosophila melanogaster* 20S proteasome purification

Embryonic *Drosophila melanogaster* extracts were kindly provided by Alexey Matyash (Department of Structural Dynamics, Max Planck Institute for Biophysical Chemistry). S150 *Drosophila melanogaster* embryonic extracts were prepared according to Bonte and Becker [140]. Embryonic extracts were adjusted to a protein concentration of 15 mg/mL with 1x purification buffer. Purification buffer was added to the diluted extract to 1x, sucrose powder to 20% (w/v), OGNG from a 10% (w/v) stock solution in water) to 0.1% (w/v), Iodacetamide to 10 mM, N-Ethylmaleimide to 10 mM and Benzamidine Chloride to 10 mM. The extract was incubated at 18 °C on a magnetic stirrer for 30 min. The alkylation reaction was stopped by adding 50 mM DTT and subsequent incubation for 30 min at 18 °C. The extract was then incubated for 1h at 30 °C in a water bath. The treated extract was cleared by centrifugation with 30'000 x g for 30 min at 4 °C. The supernatant was filtered through three layers each of cheese cloth and miracloth. The clarified extract was subjected to differential precipitation with PEG. PEG 400 was added to a concentration of 20% (v/v) to the embryonic extract under stirring at 18 °C and mixed for 20 min. Precipitated proteins were removed by centrifugation with 30'000 x g for 30 min at 4 °C. The supernatant was precipitated by raising the concentration of PEG 400 to 30% (v/v) as described above. The precipitate of this step contains fruit fly 20S proteasomes. The precipitate was pelleted by centrifugation with 30'000 x g for 30 min at 4 °C. The pellet was resuspended in purification buffer containing 2.5% (w/v) sucrose, 10 mM DTT and 0.01% (w/v) LMNG on an orbital shaker at 18 °C. The resuspended material was loaded onto 10-30% (w/v) sucrose gradients. The sucrose gradient solutions contained sucrose, 1x purification buffer and 10 mM DTT. Gradients were centrifuged

with 284'000 x g for 16 h at 4 °C. Fractionation of the gradients was done by hand in 400 µl fractions by starting from the top of the gradient. SDS-PAGE was utilized to check fractions for their 20S proteasome content. Selected fractions were pooled and precipitated by the addition of 40% (v/v) PEG 400. After centrifugation with 30'000 x g for 20 min, the supernatant was removed and the pelleted precipitate was resuspended in purification buffer containing 5% (w/v) sucrose, 10 mM DTT and 0.01% (w/v) LMNG. The resuspended material was loaded onto linear 10-40% (w/v) sucrose gradients which were prepared as the previous mentioned gradients. Gradients were centrifuged with 284,000 x g for 20 h at 4 °C. Fractions containing 20S proteasomes were identified by SDS-PAGE and precipitated by the addition of 40% (v/v) PEG 400. The concentration of the final sample was determined by the volume of added resuspension buffer, which is composed of 1x purification buffer, 5% (w/v) sucrose, 5 mM DTT and 0.01% LMNG, to the 20S proteasome pellet. Concentrations of 15 mg/ml to 20 mg/ml were obtained. This procedure reproducibly yields 4 mg purified fruit fly 20S proteasomes, starting from 50 ml S150 *Drosophila melanogaster* embryonic extract with a measured overall protein concentration of 15 mg/ml.

#### **2.2.2.4 *Saccharomyces cerevisiae* 20S proteasome purification**

The yeast strain *Saccharomyces cerevisiae* Meyen ex E.C. Hansen (ATTC 208277) was grown in YEP-media at 30 °C until an OD of 10 was reached. Cells were harvested with 5'000 g at 4 °C. Collected yeast cells were resuspended in yeast resuspension buffer with a volume which is two times the volume of the weight of the pelleted yeast cells. Yeast cell suspension was snap frozen as small beads in liquid nitrogen. The yeast beads were grinded with the Retsch ZM200 to disrupt the cells. Acetate purification buffer was added to the grinded yeast cell powder in a ratio 0.5 mL 1x buffer per 1 g yeast cell powder. 5 mM phenylmethane sulfonyl fluoride (PMSF), 10 mM Benzamidine Chloride, 20 mM DTT and sucrose to 20% (w/v) were added then and the mixture was incubated for 30 min at 18 °C. Cell suspension was cleared by centrifugation with 30'000 x g for 30 min at 4 °C. 0.2% OGNG from a 10% (w/v) stock solution in water was added to the supernatant before the supernatant was incubated in a water bath at 30 °C for 1 h. The supernatant was filtered through three layers each of cheese cloth and miracloth. The clarified extract was subjected to differential precipitation with PEG. PEG400 was added to a concentration of 20% (v/v) to the yeast supernatant under stirring at 18 °C and mixed for 20 min. Precipitated proteins were removed by centrifugation with 30'000 x g for 30 min at 4 °C. The supernatant was precipitated by raising the concentration of PEG 400 to 30% (v/v) as described above. The precipitate of this step contains yeast 20S proteasomes. The precipitate was pelleted by centrifugation with 30'000 x g for 30 min at 4 °C. The pellet was resuspended in 1x acetate purification buffer containing

2% (w/v) sucrose, 10 mM DTT and 0.01% (w/v) LMNG with an orbital shaker at 18 °C. The resuspended material was loaded onto 10-45% (w/v) sucrose gradients. The sucrose gradient solutions contain sucrose, 1x acetate purification buffer and 10 mM DTT. Gradients were centrifuged with 243'000 x g for 16 h at 4 °C. Fractionation of the gradients was done by hand in 400 µl fractions by starting from the top of the gradient. SDS-PAGE was utilized to check fractions for their 20S proteasomes content. Selected fractions were pooled and precipitated by the addition of 40% (v/v) PEG 400. After centrifugation with 30'000 x g for 20 min, the supernatant was removed and the pelleted precipitate was resuspended in 1x acetate purification buffer containing 5% (w/v) sucrose, 10 mM DTT and 0.01% (w/v) LMNG. The resuspended material was loaded onto linear 10-40% (w/v) sucrose gradients which were prepared as the previous mentioned gradients. Gradients were centrifuged with 217,000 x g for 16 h at 4 °C. Fractions containing 20S proteasomes were identified by SDS-PAGE, precipitated by the addition of 40% (v/v) PEG 400. After centrifugation with 30'000 x g for 20 min, the supernatant was removed and the pelleted precipitate was resuspended in 1x acetate purification buffer containing 5% (w/v) sucrose, 10 mM DTT and 0.01% (w/v) LMNG. PEG 6000 was added to a concentration of 5% (w/v) to the resuspended pellet. After incubation for 1h at 18 °C under constant stirring, precipitate was pelleted by centrifugation with 30'000 x g for 30 min at 4 °C. The supernatant was acidified with acetic acid to pH 6.0 and incubated overnight at 4 °C. The protein suspension was cleared by centrifugation with 30'000 x g for 30 min at 4 °C and alkalinized to pH 6.5 with KOH. The pH adjusted supernatant was precipitated by the addition of 40% (v/v) PEG 400. Protein pellets were collected by centrifugation with 30'000 x g for 30 min at 4 °C. The pellet was resuspended in 1 x acetate purification buffer, 5% sucrose and 10 mM DTT. The resuspended material was loaded onto linear 10-30% (w/v) sucrose gradients, which were prepared as the previous mentioned gradients. Gradients were centrifuged with 284,000 x g for 16 h at 4 °C. 400 µl fractions were collected by hand from the top of the gradient. The fractions were checked for their 20S proteasome content. The respective fractions were pooled and precipitated with 40% (v/v). Protein precipitate was collected by centrifugation with 30'000 x g for 30 min at 4 °C. The concentration of the final sample was determined by the volume of added resuspension buffer, which was composed of 1x purification buffer, 5% (w/v) sucrose, 5 mM DTT and 0.01% LMNG, to the 20S proteasome pellet. Concentrations of 15 mg/ml to 20 mg/ml were obtained. This procedure reproducibly yields 12 mg purified yeast 20S proteasomes starting from 800 g grinded yeast cell powder.

#### **2.2.2.5 *Thermoplasma acidophilum* 20S proteasome purification**

The polycistronic pST44 [141] was used to generate the bicistronic gene sequence containing the genes psmB (Uniprot: D4GYZ1) – psmA (Uniprot: P25156). The expression

vector pRSETA (ThermoFisher) was used for the expression of the assembled gene sequence (Fig. B.1). The expression vector was transformed into BL21 (DE3) *E. coli* cells. Cells were grown in Terrific Broth medium to an OD of 0.8 at 37 °C with 180 rpm before protein expression was induced with 0.1 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG). After incubation for 21 h at 37 °C, cells were harvested with 8'000 x g. Cells were resuspended in 3 mL Lysis Buffer per 1g of cells and 1mg/mL Lysozyme was added. Cells were disrupted with the EmulsiFlex-C3 and cell debris was removed by centrifugation with 22'000 x g for 20 min. 2.5 times the volume of boiling lysis buffer was added to the collected supernatant. After incubation for 15 min at 75 °C, the cells suspension was transferred to 4 °C and kept there for 30 min. All heat denatured proteins were removed by centrifugation with 24'000 x g for 30 min. Polyethylenimine PEI was added to a concentration of 0.1% (w/v) to the supernatant at 4 °C and mixed for 30 min. PEI treated supernatant was centrifuged with 24'000 x g for 30 min. The resulting supernatant was filtered through three layers each of cheese cloth and miracloth. The clarified extract was subjected to differential precipitation with PEG. PEG 400 was added to a concentration of 20% (v/v) to the filtered supernatant under stirring at 18 °C and mixed for 20 min. Precipitated proteins were removed by centrifugation with 30'000 x g for 30 min at 4 °C. The supernatant was precipitated by increasing the concentration of PEG 400 to 40% (v/v) as described above. The precipitate of this step contains archaeal 20S proteasomes. The precipitate was pelleted by centrifugation with 30'000 x g for 30 min at 4 °C. The pellet was resuspended in 1x acetate purification buffer containing 5% (w/v) sucrose, 10 mM DTT and 0.01% (w/v) LMNG with an orbital shaker at 18 °C. The resuspended material was loaded onto 10-30% (w/v) sucrose gradients. The sucrose gradient solutions contain sucrose, 1x acetate purification buffer and 10 mM DTT. Gradients were centrifuged with 154'000 x g for 16 h at 4 °C. Fractionation of the gradients was done by hand in 1000  $\mu$ l fractions by starting from the top of the gradient. SDS-PAGE was utilized to check fractions for their 20S proteasomes content. Selected fractions were pooled and precipitated by the addition of 40% (v/v) PEG 400. After centrifugation with 30'000 x g for 20 min, the supernatant was removed and the pelleted precipitate was resuspended in 1x acetate purification buffer containing 5% (w/v) sucrose, 10 mM DTT and 0.01% (w/v) LMNG. The resuspended material was loaded onto linear 10-30% (w/v) sucrose gradients, which were prepared as the previous mentioned gradients. Gradients were centrifuged at 284,000 x g for 16 h at 4 °C. Fractions containing 20S proteasomes were identified by SDS-PAGE and precipitated by the addition of 40% (v/v) PEG 400. After centrifugation with 30'000 x g for 20 min, the supernatant was removed and the pelleted precipitate was resuspended in 1x acetate purification buffer containing 5% (w/v) sucrose, 10 mM DTT and 0.01% (w/v) LMNG. 99 g harvested *E. coli* yielded 90 mg of *Thermoplasma acidophilum* 20S proteasomes.

### 2.2.2.6 *Homo sapiens* 26S proteasome purification

HeLa cells were kindly fermented and provided by Thomas Conrad (Bioreactor Facility, Max Planck Institute for Biophysical Chemistry) HeLa cytoplasmic extract was prepared by hypotonic lysis of HeLa cells according to Dignam et al.[138]. HeLa cells used for proteasome purification exceeded a viability of 85%. After hypotonic lysis and centrifugation to collect nuclei, the supernatant (the crude cytoplasmic extract) was centrifuged with 30'000 x g for 30 min at 4 °C, flash frozen in liquid nitrogen as 40 ml aliquots and stored at -80 °C until usage. The S30 HeLa cytoplasmic extract was thawed in a water bath at 37 °C and supplemented with purification buffer to 1x concentration from a 10x stock. Subsequently, sucrose powder was added to 20% (w/v), OGNG from a 10% (w/v) stock solution in water to 0.1% (w/v), Iodacetamide to 10 mM, N-Ethylmaleimide to 10 mM, Benzamidine Chloride to 10 mM and ATP to 7.5 mM. The extract was incubated at 18 °C on a magnetic stirrer for 30 min. Alkylating agents were quenched by the addition of 50 mM DTT. The mixture was incubated at 18 °C under constant stirring on a magnetic stirrer for 30 min. The quenched extract was cleared by centrifugation with 100'000 x g for 2 h at 4 °C. The supernatant was filtered through three layers each of cheese cloth and miracloth and the filtrate (S100 HeLa extract) was collected. The clarified extract was subjected to differential precipitation with PEG. PEG 400 was added to a concentration of 23% (v/v) to the S100 HeLa cytoplasmic extract at 18 °C and incubated for 20 min under constant stirring. Precipitated proteins were removed by centrifugation with 30'000 x g for 30 min at 4 °C. The supernatant was precipitated by increasing the concentration of PEG 400 to 30% (v/v) as described above. The precipitate of this step contains human 26S proteasomes. The precipitate was pelleted by centrifugation with 30'000 x g for 30 min at 4 °C. The pellet was resuspended in 1x purification buffer containing 5% (w/v) sucrose, 5 mM DTT, 0.01% (w/v) LMNG and 7.5 mM ATP on an orbital shaker at 18 °C. 10 mM sodium creatine phosphate and 5 µg/ml creatine kinase were added to the resuspended pellet and incubated for 30 min at 30 °C to regenerate ATP. The resuspended material was then loaded onto a 20%/50% sucrose cushion. The sucrose solutions contain sucrose, 1x purification buffer, 5 mM DTT and 7.5 mM ATP. Sucrose cushions were centrifuged with 260'000 x g for 14 h at 4 °C. Fractionation of the sucrose cushions was done by hand in 400 µl fractions by starting from the top of the sucrose cushion. SDS-PAGE was utilized to check the fractions for their 26S proteasome content. Selected fractions were pooled and precipitated by the addition of 40% (v/v) PEG 400. After centrifugation with 30'000 x g for 20 min, the supernatant was removed and the pelleted precipitate was resuspended in 1x purification buffer containing 5% (w/v) sucrose, 5 mM DTT, 0.01% (w/v) LMNG and 7.5 mM ATP. The resuspended material was loaded onto linear 10-40% (w/v). The sucrose gradient solutions contain sucrose, 1x purification buffer, 5 mM DTT and 7.5 mM ATP. Gradients were centrifuged with 220,000 x g for



16 h at 4 °C. Harvested 400 µl fractions containing 26S proteasomes were identified by SDS-PAGE and precipitated by the addition of 40% (v/v) PEG 400. The precipitate was pelleted by centrifugation with 30'000 x g for 20 min. The pellet was resuspended in 1x purification, 7.5% (w/v) sucrose, 5 mM DTT, 0.01% (w/v) LMNG and 7.5 mM ATP. The protein solution was loaded onto a linear 10-45% (w/v) sucrose gradient, which was prepared as the previously described sucrose gradient. Gradients were centrifuged with 260'000 x g for 16 h at 4 °C. Harvested 400 µl fractions were analyzed for their 26S proteasome content by SDS-PAGE. The concentration of the final sample was controlled by the volume of added resuspension buffer, which is composed of 2x purification buffer, 15% (w/v) sucrose, 10 mM DTT, 0.02% LMNG and 15 mM ATP to the 26S proteasome pellet. Concentrations of 30 mg/mL were obtained.

### 2.2.2.7 Protein concentration determination

Protein concentrations were determined by the Bradford assay using bovine serum albumin (BSA) as a protein standard. For protein concentration measurements 999 µl of Bradford solution was mixed with 1 µl protein and incubated for 2 min. The absorption at 595 nm was measured to determine the protein concentration according to Beer-Lambert law [142].

## 2.2.3 20S proteasome crystallization

### 2.2.3.1 *Homo sapiens* 20S proteasome

Human 20S proteasomes are crystallized at a protein concentration of 7.5 mg/ml by mixing 0.5 µl protein + 0.5 µl human crystallization buffer in Chrysem sitting drop vapor diffusion plates over a 500 µl reservoir of human Crystallization buffer. Under these conditions, human 20S proteasomes crystallize in space group  $P 2_1 2_1 2_1$  with unit cell constants of  $a = 114 \text{ \AA}$ ,  $b = 202 \text{ \AA}$ ,  $c = 302 \text{ \AA}$ ,  $\alpha = \beta = \gamma = 90^\circ$ . Typically, 3 - 6 crystals with a size of  $150 \times 150 \times 200 \text{ \mu m}^3$  grow in one well. Crystal growth is completed within 24 h of incubation at 18 °C.

### 2.2.3.2 *Drosophila melanogaster* 20S proteasome

Fruit fly 20S proteasomes are crystallized at a protein concentration of 10 mg/ml by mixing 1.0 µl protein + 1.0 µl fruit fly Crystallization buffer in Chrysem sitting drop vapor diffusion plates over a 500 µl reservoir of fruit fly Crystallization buffer. Under these conditions, fruit fly 20S proteasomes crystallize in space group  $P 1$  with unit cell

constants of  $a= 117 \text{ \AA}$ ,  $b= 171 \text{ \AA}$ ,  $c= 212 \text{ \AA}$ ,  $\alpha= 69^\circ$   $\beta= 88^\circ$   $\gamma= 70^\circ$ . Crystals grow to a size of  $150 \times 150 \times 200 \text{ \mu m}^3$  in one week at  $18^\circ \text{C}$ .

### 2.2.3.3 *Saccharomyces cerevisiae* 20S proteasome

Yeast 20S proteasomes are crystallized by mixing  $2.0 \text{ \mu l}$   $10 \text{ mg/ml}$  protein sample with  $1.0 \text{ \mu l}$  yeast Crystallization buffer on a siliconized glass cover slide. The drop was equilibrated on SuperClear hanging drop vapor diffusion plates over a  $500 \text{ \mu l}$  reservoir of yeast crystallization buffer. Under these conditions, yeast 20S proteasomes crystallize in space group  $P 1 2_1 1$  with unit cell constants of  $a= 134 \text{ \AA}$ ,  $b= 301 \text{ \AA}$ ,  $c= 143 \text{ \AA}$ ,  $\alpha= 90^\circ$   $\beta= 112^\circ$   $\gamma= 90^\circ$ . Crystals grow to a size of  $200 \times 200 \times 100 \text{ \mu m}^3$  in one week at  $18^\circ \text{C}$ .

### 2.2.3.4 *Thermoplasma acidophilum* 20S proteasome

Archaeal 20S proteasomes are crystallized by mixing  $2.0 \text{ \mu l}$   $10 \text{ mg/ml}$  protein sample with  $2.0 \text{ \mu l}$  archaeal Crystallization buffer on a siliconized glass cover slide. The drop was equilibrated on SuperClear hanging drop vapor diffusion plates over a  $500 \text{ \mu l}$  reservoir of archaeal Crystallization buffer. Under these conditions, archaeal 20S proteasomes crystallized in space group  $P 1 2_1 1$  with unit cell constants of  $a= 115 \text{ \AA}$ ,  $b= 315 \text{ \AA}$ ,  $c= 203 \text{ \AA}$ ,  $\alpha= 90^\circ$   $\beta= 94^\circ$   $\gamma= 90^\circ$ . Crystals grow to a size of  $300 \times 200 \times 100 \text{ \mu m}^3$  in four days at  $18^\circ \text{C}$ .

### 2.2.3.5 Post-crystallization treatment

The same post-crystallization treatment was used for all crystals. The added volumes of stabilization buffer and the composition of stabilization buffer and cryo buffer vary depending on the organism. The protocol described here was applied for human 20S proteasomes. Table 2.5 lists the volumes and buffer compositions, which were used for the different organisms. The steps are performed sequentially. Crystallization trays containing 20S proteasome crystals initially grown at  $18^\circ \text{C}$  were gradually cooled down to  $4^\circ \text{C}$  over a time period of 24 h in temperature controllable fridges. The plates were then transferred to the cold room and equilibrated there for at least 6 h. After equilibration, the wells were opened and  $1.0 \text{ \mu l}$  of reservoir solution was added to the drop. Then  $2.0 \text{ \mu l}$  stabilization solution was added to the crystallization drop and the reservoir was exchanged with  $500 \text{ \mu l}$  cryo buffer. The resealed drop was equilibrated for at least 48 h against the exchanged reservoir. Crystals were directly harvested from crystallization drops and flash frozen in liquid nitrogen.

**Table 2.5: Post-crystallization procedures used for 20S proteasome crystals**

Organism	Step 1	Step 2	Step 3
<i>Homo sapiens</i>	add 1 $\mu$ l reservoir to drop	add 2 $\mu$ l stabi buffer to drop	exchange reservoir with 500 $\mu$ l cryo-buffer
<i>Drosophila melanogaster</i>	-	add 2 $\mu$ l stabi buffer to drop	exchange reservoir with 500 $\mu$ l cryo-buffer
<i>Saccharomyces cerevisiae</i> PEG 400	-	add 2 $\mu$ l yeast cryo-buffer 1 to drop	exchange reservoir with 500 $\mu$ l yeast cryo-buffer 1
<i>Saccharomyces cerevisiae</i> PEG 3350	-	add 2 $\mu$ l yeast cryo-buffer 2 to drop	exchange reservoir with 500 $\mu$ l yeast cryo-buffer 2
<i>Thermoplasma acidophilum</i>	-	add 4 $\mu$ l archaeal stabi-buffer to drop	exchange reservoir with 500 $\mu$ l archaeal cryo-buffer

### 2.2.4 Inhibitor soaking protocol

100 mM stock solutions for all inhibitors were prepared in dimethyl sulfoxide (DMSO). Inhibitor stock solutions were stored at -80 °C until usage. Prior soaking, the inhibitors were diluted to 5 mM in cryo Buffer. Crystals were soaked by the addition of 5 mM inhibitor cryo buffer solution in a 1:1 volume ratio to the crystallization drop at 4 °C. Crystal were incubated in the resealed well for 48 h at least before they were harvested.

### 2.2.5 Peptide soaking protocol

The peptide (NH<sub>2</sub>-KFTQKPLERSYYS-OH) was gratefully provided by Dr. Jankowska from the University of Gdansk. A 100 mM stock solution of the peptide was prepared in DMSO. Inhibitor stock solutions were stored at 4 °C until usage. Prior soaking, the peptide was diluted to 5 mM in cryo buffer. Crystals were soaked by the addition of 5 mM inhibitor cryo buffer solution in a 1:1 volume ratio to the crystallization drop at 4 °C. Crystal were incubated for 48 h at least in the resealed well before they were harvested.

### 2.2.6 Diffraction data collection

Crystals were harvested in Litholoops or MicroLoops mounted on magnetic pins and vitrified by either plunge-freezing in liquid nitrogen or direct transfer to a nitrogen gas stream at 100 K. Data collection was performed under cryogenic conditions at 100 K. Initial data collections of human 20S proteasome crystals were done at beamline ID23-1,

ESRF, Grenoble, France. All human, fruit fly, yeast and archael high-resolution diffraction datasets reported in this thesis were collected on EMBL beamline P14 at the PETRA III storage ring (DESY, Hamburg, Germany) using a MD3 vertical-spindle diffractometer and a PILATUS2 6MF detector. The X-ray beam as produced by the undulator was defined by a white-beam compound refractive lenses (CRL) transfocator positioned 20.6 m upstream of the sample position, imaging the X-ray source at 3 - 4 m downstream of the sample and by slits 0.3 m upstream of the sample. Under these conditions, a smooth, “top-hat” beam profile with a total photon flux of up to  $2 \times 10^{13} \frac{ph}{sec}$  is generated. Using different slit settings, beam sizes between 150 - 200  $\mu\text{m}^2$  in linear dimensions – matching the crystal dimensions - were created. The best data were collected by “bathing-the-crystal” in the beam, meaning the beam size was adjusted to the size of the crystal to use all available unit cells for diffraction.

Diffraction data were collected using 40 ms exposures,  $\varphi$ -slicing of 0.05 - 0.10° and a total rotation range of 180° - 360° at 50% attenuation. Diffraction data were scaled and integrated with the XDS program package using standard protocols.

### 2.2.7 Structure determination

Initial phases for human 20S proteasomes were determined by molecular replacement using the murine 20S structure (PDB ID: 3UNE [49]). The model was built and optimized by several rounds of interactive manual model building in Coot and refinement in Refmac5. Refinement strategies for the different 20S proteasome crystals are listed in Table 2.6.

**Table 2.6: Refinement strategies used for 20S structure determinations**

<i>Homo sapiens</i>	<i>Drosophila melanogaster</i>	<i>Saccharomyces cerevisiae</i>	<i>Thermoplasma acidophilum</i>
10 cycles rigid body refinement	10 cycles rigid body refinement	10 cycles rigid body refinement	10 cycles rigid body refinement
5 cycles NCS restricted refinement	5 cycles NCS restricted refinement	5 cycles NCS restricted refinement	5 cycles NCS restricted refinement
5 cycles TLS refinement	5 cycles TLS refinement		
5 cycles NCS restricted refinement Weight Matrix 0.5	5 cycles NCS restricted refinement Weight Matrix 0.5		

For the human 20S proteasome the identity of many ions was validated by anomalous difference Fourier maps in native crystals. When necessary soaking experiments, where the ligands in question were exchanged for anomalous scatterers, were performed: e.g.,

Cl<sup>-</sup> ions were identified by their anomalous Fourier differences at 6.5 keV and by exchange through soaking crystals with Br<sup>-</sup> ions and measurements at the bromide K- absorption edge (13.47 keV); Mg<sup>2+</sup> ions were identified by exchange through soaking crystals with Mn<sup>2+</sup> ions and measurement at the manganese absorption edge (6.5 keV); K<sup>+</sup> ions were identified by their residual anomalous scattering at 6.5 keV, non-exchangeability with either Br<sup>-</sup> or Mn<sup>2+</sup> and via coordination distances to neighboring groups in the crystal structure.

For modeling the bound inhibitors, complete stereo-chemical descriptions of double- or single- covalent link attachments were created using program JLigand and applied in real-space fitting with Coot followed by refinement with Refmac5.

## 2.2.8 Dose series

A low radiation dose dataset of a 20S proteasome crystal with 0.7 MGy was measured by collecting 1800 images using a 40 ms exposure, beam transmission of 10% and a  $\varphi$ -slicing of 0.1°. After the crystal was exposed to 15 MGy, a second dataset was collected with the same settings used for the collection of the previous dataset. For both datasets, structure were calculated and the epoxyketone linkage structure was analyzed.

## 2.2.9 Negative stain electron microscopy

### 2.2.9.1 Inhibition protocol for negative stain electron microscopy

26S proteasome (12 mg/ml) was either supplemented with 2 mM DMSO as control, 2 mM Epoxomicin dissolved in DMSO (100 mM stock), or 2 mM Oprozomib dissolved in DMSO (100 mM stock). The samples were incubated for 30 min on ice. After incubation, the samples were loaded onto 10 – 30% (w/v) sucrose gradients. The sucrose gradients were composed of sucrose, 1x purification buffer, 5 mM DTT and 7.5 mM ATP. Gradients were centrifuged with 114'000 x g for 16 h at 4 °C. The gradients were harvested in 200  $\mu$ l fractions by hand from the top. Fractions were analyzed for their protein content with a dot blot stained with Amido Black.

### 2.2.9.2 Grid preparation

A continuous carbon foil was floated on the 26S proteasome containing fraction for 1 min at 4 °C. After removing excess liquid with a tissue paper, proteins were stained by floating the carbon foil on a saturated uranyl formate solution for 1 min at 4 °C. The remaining staining was removed with a tissue and the grid was dried at the air at room temperature.

### 2.2.9.3 Image acquisition

Negative stain electron microscopy images were taken with a Philips CM200 microscope. Images were acquired at a magnification of 88'000 x. The pixel size corresponds to 2.5 Å per pixel. The TVIPS charge coupled device (CCD) camera was used to record 500 micrographs per sample.

### 2.2.9.4 Image processing

Cow picker was used for particle picking. Contrast transfer function CTF correction was performed on individual particle level in the Cow. Several rounds of two-dimensional (2D) classification was employed to remove images without particles and containing contaminations like ice crystals. The cross correlation coefficient was used to determine if a particle fits better with the three-dimensional (3D) model of the rotated or with the non-rotated state. Three iterations of the particle assignment were done. After each iteration, new 3D volumes were reconstructed based on the assigned particles. The 3D volumes were low-pass filtered to the same resolution and normalized before starting the next round of particle assignment. Particles which contribute to each individual class were counted and plotted after the third iteration.

## 2.2.10 Electron cryo-microscopy

All electron microscopy work and data evaluation were done in collaboration with Dr. David Haselbach, Dr. Jil Schrader and Felix Lambrecht (Department for Structural Dynamics, Max-Planck-Institute for Biophysical Chemistry). Results of this work were published in Haselbach et al. 2017 [63].

### 2.2.10.1 Inhibition protocol for negative stain electron microscopy

Oprozomib was applied to the resuspended 26S proteasome pellet obtained from fractions of the sucrose cushion to inhibit the 26S proteasome. Oprozomib dissolved in DMSO (stock 100 mM) was added to a concentration of 0.5 mM. The mixture was incubated for 30 min at 25 °C.

### 2.2.10.2 Electron cryo-microscopy sample preparation

Human 26S proteasome (12 mg/ml) was mixed with 2.5 mM bis(sulfosuccinimidyl)suberate (BS3) and incubated for 2 h at 4 °C. The remaining BS3 was quenched with 10 mM sodium aspartate (pH 6.5). The sample was loaded then onto a linear 10-30% (w/v) sucrose gradient containing 0-0.1% (v/v) glutaraldehyde. Gradients were centrifuged with 114'000 x g for 16 h at 4 °C. 200 µl fractions were harvested by hand starting from the top of the gradients. The remaining glutaraldehyde (GA) in the fractions was quenched by adding 20 mM sodium aspartate (pH 6.5). Fractions were checked for their protein content by performing a dot blot with Amido Black staining. Buffer of fractions containing 26S proteasomes was exchanged by using Zeba spin columns. Quantifoil (3.5/1) grids with continuous carbon foil attached were floated on the sample to enable the absorption of particles to the carbon foil. Grids were then plunge frozen into liquid ethane with a Leica EM GP. The blotting chamber was adjusted to 75% humidity and 4 °C prior freezing the grids.

### 2.2.10.3 Data acquisition

Micrographs were collected with a Titan Krios equipped with a Cs-Corrector on a Falcon II detector. A nominal magnification of 110'000 x, corresponding to a pixel size of 1.27 Å per pixel, was used to record 18'707 micrographs. A total dose of 50 electrons per Å<sup>2</sup> was distributed on 17 frames. The first frame of each micrographs was discarded, as the camera shutter caused inhomogeneous illumination.

### 2.2.10.4 Image processing

Unblur was used for the alignment and weighting of the individual image frames according to the electron dose to correct for drift and charging. Gctf was used to determine the contrast transfer function (CTF) of the aligned micrograph stacks. Selection of particle was done in a template-free manner, using image statistical properties in combination with

mass centering. Individual particle coordinates were refined by alignment against twelve low resolution reference images representing different views of the proteasome. Subsequently, particles were cleaned by several image sorting steps to remove contaminations, blurred images and bad particles. Power spectra were calculated for each particle and classified using a hierarchical clustering scheme. The resulting class averages were visually inspected for their Thon ring. Particles belonging to charged or blurred classes were discarded. Following, several rounds of multi-reference alignment and 2D-classification were performed. Particles which did not classify with distinct 26S proteasome views were discarded. An initial 3D model was calculated with the best class averages using simple PRIME. 3D classification in RELION was started with the obtained 3D model as reference. The particle classified into two distinct 3D classes corresponding to two main conformational states, the rotated and non-rotated state. Correct class assignment was ensured by competitive alignment of all particles against averaged maps obtained for the two main states. The Rpn1 protein density was masked out, because it interfered with the alignments. Particles belonging to the non-rotated state of the proteasome were refined by RELION auto-refine. Further sub classes of regulatory particle conformations were revealed by hierarchical sorting. A series of 3D classification steps without alignment using smaller mask sizes for each step were performed in RELION. For the first classification step a mask for the whole proteasome holoenzyme excluding Rpn1 was applied, in the following second iteration a mask for the whole regulatory particle, in the third iteration a mask for the whole lid and lastly a mask for the Rpn2 subunit only. The remaining particles (233'000) were refined to a final resolution of 3.8 Å with RELION. The obtained map was improved by correcting the B-factor and by particle polishing of the final particle stacks with RELION.

### 2.2.10.5 Model building

The atomic coordinates from the Oprozomib-inhibited 20S proteasome crystal structure model (PDB: 5LEY) were used to model the 20S proteasome core. Models of each 19S RP subunit were predicted with Robetta. UCSF Chimera was used to dock the Robetta models as rigid bodies into the EM density map. The six nucleotides of the ATPase subunits were modeled by fitting the PAN crystal structure model (PDB: 3H4M) into the 19S base EM density map. For parts which had to be modeled partially de novo information from the psipred server about secondary elements was utilized. First, a rigid body refinement was performed using real space refinement in Phenix, followed by manual modeling in coot. Secondly, secondary structure restraints were generated using phenix.ksdssp. All secondary structure restraints were monitored and additional restraints were added manually. Several iterative rounds of real space refinement in Phenix and manual modeling



in Coot were done. The last Phenix refinements included an atomic displacement parameters refinement to determine the B-factors. According to the B-factor distributions details in the final model were interpreted. B-factors were analyzed in segments of five amino acids. Side chains were only modeled in segments with a mean atomic B-factor smaller than  $110 \text{ \AA}^2$ . Segments with mean B-factors between  $110 \text{ \AA}^2$  and  $150 \text{ \AA}^2$  were truncated to poly-alanine. Residues were not modeled and included in the final deposited PDB model for mean B-factors higher than  $150 \text{ \AA}^2$ .

### 2.2.11 Energy landscape calculation

The method for determining energy landscapes were established by Dr. David Haselbach and Felix Lambrecht (Structural Dynamics, Max Planck Institute for Biophysical Chemistry). Particles were split into two subsets based on their assignment to the native 26S proteasome or the Oprozomib-inhibited 26S proteasome. Alignment parameters of the particles belonging to one subset were determined by refinement against the initial 26S proteasome map with RELION to reduce the computational effort in the following classification steps. The aligned images were randomly split into subsets of 100,000 particles and a RELION 3D-classification was performed. The classification resulted in 40 3D classes per subset. Each resulting 3D class was further refined with the assigned particles using RELION.

The conformational differences between the 346 classes in total were analyzed by principle component analysis (PCA). The resulting eigenvectors  $e_j$  describe the largest covariance within the dataset. PCA sorted 3D volumes of the same molecule provide information about the movement within the molecule. The determined eigenvectors were used as conformational coordinates. All refined 3D-classes were aligned to the rigid 20S proteasome core in UCSF Chimera, before the PCA was done. This avoided calculating eigenvectors, which describe rotation and shifts between the individual 3D classes. Eigenvectors were determined by calculating the PCA for combined and normalized 3D classes from datasets of the native 26S proteasome and the Oprozomib inhibited 26S proteasome. These calculated eigenvectors resemble the movements found in both datasets and are needed to compare the result of the PCA, which were calculated for each dataset separately.

The conformer  $X_i$  is described by the eigenvector  $e_n$ , which corresponds to a certain movement in the protein complex, and the linear factor  $e_{an,I}$ , which provides information about how far the movement progressed on its trajectory.  $\bar{X}$  is the average volume and originates from the definition of the PCA.

$$X_i = a_{n,i}e_n + \bar{X} \quad (2.1)$$

The linear factors for each 3D-class with respect to each eigenvector were calculated. The linear factors enable to allocate the 3D class to a spot in N-dimensional space, which is defined by N-independent eigenvectors. The number of particles assigned to a certain 3D-class allow to calculate the free energy state of the 3D class using the Boltzmann Interpolation:

$$\Delta\Delta G = k_B T \ln \left( \frac{p_i}{p_0} \right) \quad (2.2)$$

,where  $T$  is the absolute temperature,  $p_i$  is the number of particles in state  $i$  and  $p_0$  is the number of particles in the most populated state. The number of particles belonging to each 3D class was retrieved from the data output files from RELION. Free energy differences were calculated using equation 2.2 as multiples of  $k_B T$ . Regions with high energies have a lower number of particles, whereas low energies are characterized by a high number of particles.

The energy landscape comparing the 26S proteasome in the native state and the Oprozomib inhibited state was spanned by two eigenvectors  $e_n$ ,  $e_m$  describing the major modes of motion. The 3D classes were allocated in two dimensional space based on their linear contribution of the respective eigenvectors. To this two dimensional space, a third dimension corresponding to the energy values, which are proportional to the particle numbers belonging to one 3D class as described before, are added. The space between discrete 3D classes was interpolated. The interpolation enables the description of a continuous energy landscape. 3D classes which lie in a distance of half-width of the peak around one linear factor were binned together and averaged.

### 2.2.12 Electronic structure calculations

Cluster calculations were carried out on selected model systems, considering the Dihydroeponemycin, Oprozomib and Z-LLY-Ketoaldehyde inhibitors. The starting structures were based on the crystallographic data. All stationary points and constrained optimizations were carried out at the B3LYP-D3(BJ)/def2-SVP level of theory [143–146], which include dispersion corrections as suggested by Grimme, with a Becke-Johnson type damping [147]. The nature of the stationary points was confirmed by frequency calculations. The reaction energy profiles are provided at the level of theory used for the optimizations (B3LYP-D3/def2-SVP). Free-energy corrections were computed from frequency analysis under the rigid-rotor harmonic approximation of the latter structures. The electronic

energy was determined at the B3LYP-D3/def2-TZVP level of theory, under the COSMO continuum solvation model [148]. Diethyl ether was employed as the solvent ( $\epsilon=4.3$ ), which is commonly chosen to model protein-like environments [149]. All free-energy values reported correspond to the sum of the latter electronic energies and the thermodynamic corrections computed with the smaller def2-SVP basis set. All calculations were carried out with the Orca 3.0.3 program package.

### 2.2.13 Enzyme kinetics

All kinetic measurements were performed using a FluoroMax®-4 fluorescence spectrophotometer. Succinyl-Leucine-Leucine-Valine-Tyrosine-7-Amino-4-Methylcoumarin (AMC) was used as substrate to determine the chymotryptic-like activity of the  $\beta 5$  catalytic active site of the human 20S proteasome [150]. The fluorescence emission of hydrolyzed AMC was continuously monitored at 460 nm ( $\lambda_{\text{ex}} = 380$  nm). The reaction temperature was kept at 37 °C for all measurements. The reaction buffer used for enzymatic assays is specified in Table 2.3. Suc-LLVY-AMC and inhibitors (Oprozomib, Dihydroeponemycin, Z-LLY-Ketoaldehyde) were dissolved in DMSO and stored at -80 °C until usage. The DMSO concentration did not exceed 2% (v/v) in any measurement.

For kinetic characterization of Suc-LLVY-AMC conversion, 0.035 mg/mL (50 nM) human 20S proteasome in reaction buffer was pre-incubated for 3 min at 37 °C. The reaction was started by the addition of substrate and the fluorescence signal was measured continuously.

For determination of the first-order rate inhibition constant of the respective inhibitors, the reaction mixture containing reaction buffer, 150  $\mu\text{M}$  substrate and either Oprozomib (50  $\mu\text{M}$ ), Dihydroeponemycin (50  $\mu\text{M}$ ) or Z-LLY-Ketoaldehyde (15  $\mu\text{M}$ ) were pre-incubated at 37 °C for 3 min. The reaction was then started by the addition of human 20S proteasome to a final concentration of 50 nM. The fluorescence signal was measured continuously. Data were analyzed and fitted with OriginPro 9.1 and KaleidaGraph 4.03. The equation shown in Eq. 2.3 was used to analyze the chymotryptic-like catalytic activity and catalytic activation of the 20S proteasome. For the determination of the first-order inactivation rate constants, equations were used that contained either two exponential terms in case of Z-LLY-Ketoaldehyde, or two exponential terms plus a linear term for epoxyketones (Oprozomib and Dihydroeponemycin). The first of the two exponential terms accounts for the catalytic activation, whereas the second exponential term signifies the catalytic inactivation by inhibitory action. The linear term in case of the epoxyketones was used to describe account for the residual activity of the proteasome after inactivation.

$$F(t) = F_0 + \Delta F_{ss}t - \frac{(\Delta F_{ss} - \Delta F_0)}{k_{act}}(1 - \exp(-k_{act}t)) \quad (2.3)$$

### 2.2.14 Sequence alignments

The EMBOSS needle alignment algorithm was used for sequence alignments of the human 20S proteasome subunits with the fruit fly 20S proteasome subunits [151].



## 3 Results

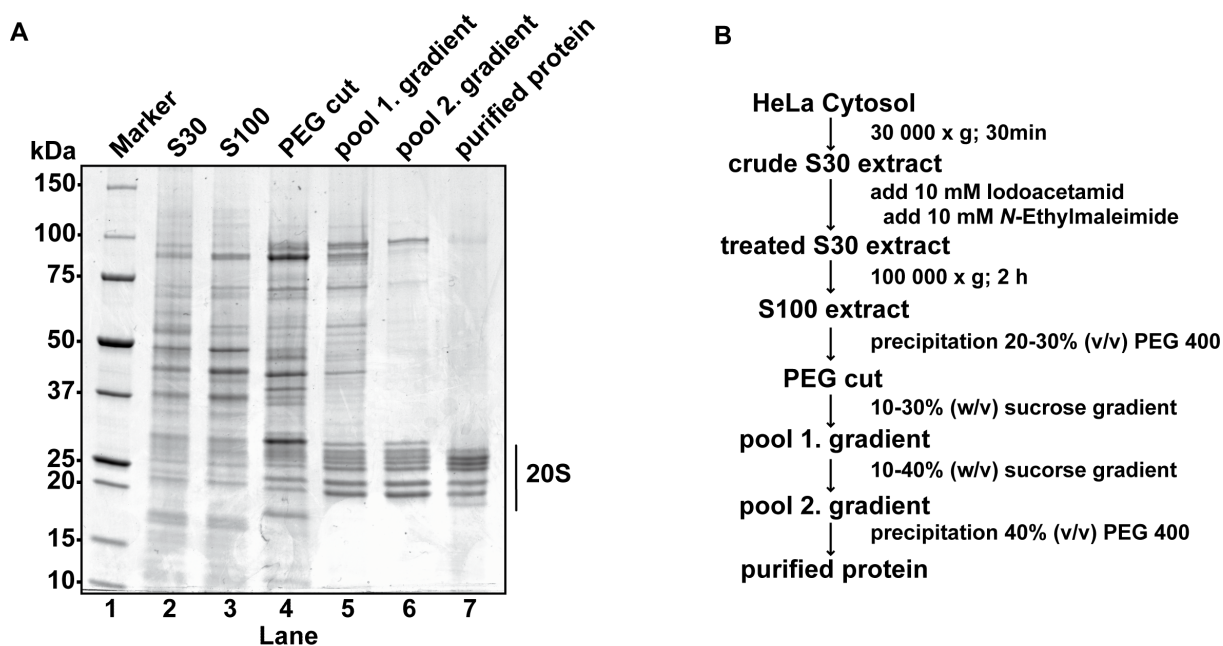
The starting point for each structural study of a protein complex is the availability of a pure and homogenous sample. Therefore, efficient and reliable purification protocols that preserve the structure of the protein complex of interest must first be established. The proteasome is a molecule of high interest because it plays a major role in protein degradation. Furthermore, the 20S proteasome is a validated anti-cancer drug target. Therefore, there is a high demand for purified 20S/26S proteasome samples for biochemical and structural analyses, which enable the development of new drugs targeting the 20S proteasome. Several protocols for the purification of the 20S proteasome exist in published scientific literature [25, 31, 32, 49, 119]. These protocols involve the application of chromatography methods to isolate the complex from lysed cells. Since the purification methods heavily influence the sample quality and the outcome of subsequent studies, we decided to establish a new purification protocol that is different from those used previously. The decisive idea was to avoid chromatography during the purification, since it is supposed that uncontrollable interactions of the protein complexes with the column material damage their structure.

### 3.1 Chromatography free purification protocol

A chromatography-free purification was established and optimized to purify 20S proteasomes from HeLa cells (Fig. 3.1). The buffer used for 20S proteasome purification was optimized with the ProteoPlex technique (Fig. 3.2; B) [137]. The optimized buffer increased the melting temperature of the protein complex, which indicates a stabilization of the 20S proteasome by the newly screened buffer condition.

First, centrifugation steps at 30'000 x g and 100'000 x g were utilized for the crude removal of cell organelles and ribosomes. After centrifugation with 30'000 x g, the supernatant containing the proteasome was alkylated with iodoacetamide and N-ethylmaleimide to reduce the complexity in protein composition in the supernatant fraction, as it causes the disintegration of contaminating protein complexes but not of the proteasome. Heat incubation after alkylation ensured the disassembly of the 26S proteasome into 19S RP and 20S proteasome. The 19S RP complex alone is not stable and disintegrates into its components. Further protein impurities were removed by precipitation of the heat-treated supernatant with 20% (v/v) PEG 400. After clearing the suspension by centrifugation, the

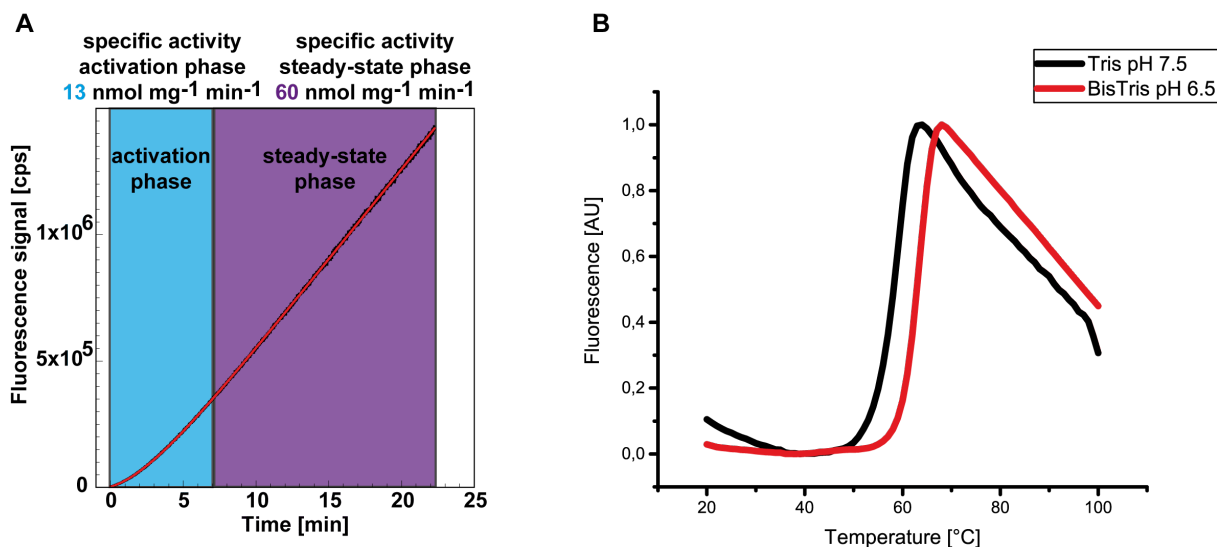
supernatant was treated with 30% (v/v) PEG 400 to precipitate the 20S proteasome. The precipitated 20S proteasomes were pelleted by centrifugation at 30'000 x g. Comparison of the SDS-PAGE analysis of the supernatant (S100) after the 100'000 x g centrifugation step with the resuspended 20S proteasome pellet demonstrates the successful separation of unwanted proteins from the 20S proteasome (Fig. 3.1; A). Remaining impurities, which have the same PEG 400 precipitation properties were removed by the utilization of two sequential sucrose gradients. The sucrose gradients enable the separation of proteins according to their molecular weight and their hydrodynamic radius. Fractions of the sucrose gradients were analyzed for their 20S proteasome content by SDS-PAGE (Fig. 3.1; A). Concentration of the 20S proteasome protein sample was achieved by precipitation of the 20S proteasome containing fractions with 40% (v/v) PEG 400 and subsequent centrifugation. The final volume used to resuspend the 20S proteasome determines the final protein concentration. The final purified 20S proteasome sample shows a purity of 95% according to SDS-PAGE and was suitable for crystallization.



**Fig. 3.1: Overview of the newly developed purification strategy.** A: SDS-PAGE analysis of intermediate purification steps and purified human 20S proteasome sample. S30 and S100 denote the supernatants after performing centrifugation with either 30'000 x g or 100'000 x g. PEG cut designates the resuspended pellet precipitated with 30% (v/v) PEG 400. The pool lanes include pooled fractions containing 20S proteasomes concentrated by PEG 400 precipitation. The last lane shows the purified sample. B: Schematic overview of the developed purification protocol is shown.

The quality of the purified sample was assessed by measuring the kinetics for the cleavage rate of a fluorogenic peptide substrate. The enzymatic assay revealed two kinetic phases, an activation phase with a specific activity of  $13 \frac{\text{nmol}}{\text{mg} \cdot \text{min}}$  and a steady state phase with a specific activity of  $60 \frac{\text{nmol}}{\text{mg} \cdot \text{min}}$  (Fig. 3.2; A). The steady state phase was reached after

7 min. Compared to published specific activities, the sample purified with the newly developed method has a ten times higher specific activity [150]. Applying the sample to a buffer exchange column significantly reduced the specific activity.



**Fig. 3.2: Activity measurement and ProteoPlex analysis of the human 20S proteasome.** A: Cleavage reaction of the fluorogenic substrate Suc-LLVY-AMC was analyzed by measuring the fluorescence signal of AMC. Two phases can be differentiated, an activation phase with a specific activity of  $13 \frac{\text{nmol}}{\text{mg} \cdot \text{min}}$  and a steady-state phase with a specific activity of  $60 \frac{\text{nmol}}{\text{mg} \cdot \text{min}}$ . B: ProteoPlex analysis of human 20S proteasome in Tris pH 7.5 and BisTris pH 6.5 is depicted. The 20S proteasome is more stable in BisTris pH 6.5 indicated by the higher melting temperature.

## 3.2 Adaptation of the human 20S purification protocol to other sources

To show that the 20S proteasome purification strategy is ubiquitously applicable, 20S proteasomes were purified from *Saccharomyces cerevisiae*, *Drosophila melanogaster* embryonic extracts and from *Thermoplasma acidophilum* exogenously expressed in *E. coli*. For all organisms, pure 20S proteasomes were obtained that were suitable for crystallization trials.

### 3.2.1 *Drosophila melanogaster* 20S proteasome purification

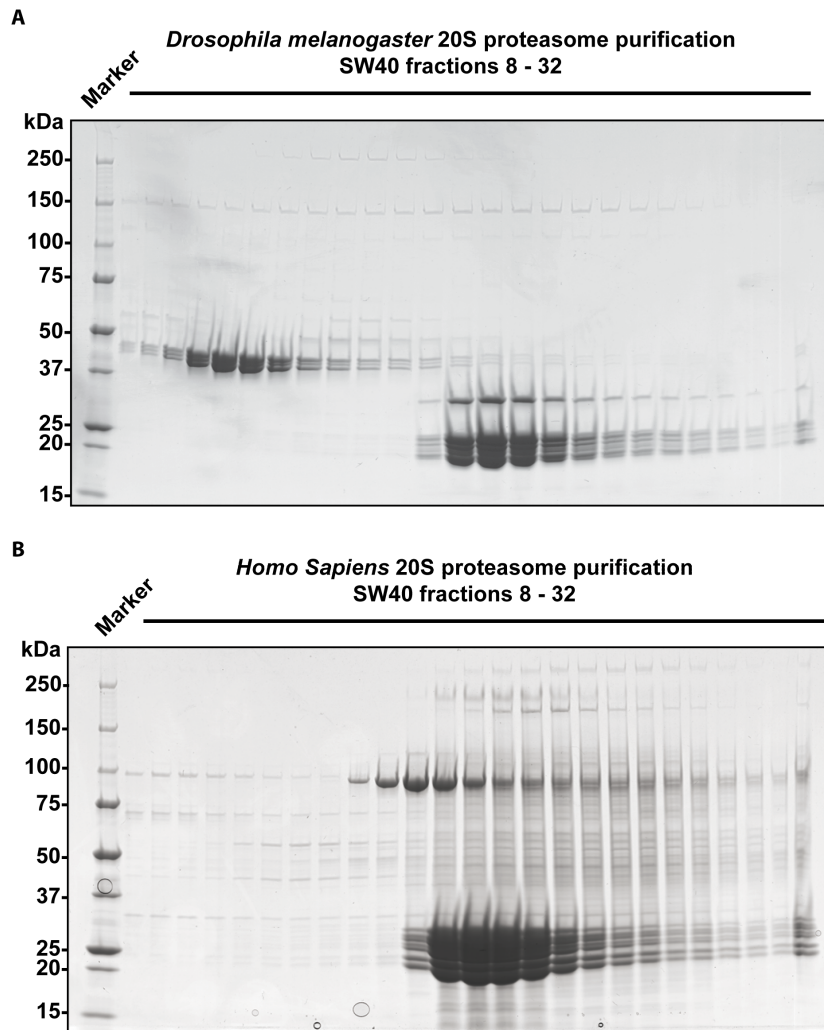
The *Drosophila melanogaster* 20S proteasome (d20S) was purified with the same protocol that was used for the purification of *Homo Sapiens* 20S proteasome (h20S). However, one major difference was the introduction of a dilution step before starting the alkylation reaction to reach an overall protein concentration of  $15 \frac{\text{mg}}{\text{ml}}$  in the embryonic



*Drosophila melanogaster* extract, which is comparable to the protein content of HeLa S10 extract preparations. The dilution step is important because the embryonic *Drosophila melanogaster* extracts have a higher starting protein concentration, which can change the PEG 400 precipitation properties. Comparison of the human and fruit fly samples after PEG 400 precipitation revealed a protein content of less complexity for the fruit fly sample (Fig. 3.3, A). The major impurity observed in the *Drosophila melanogaster* sample was the yolk protein with a size of 45 kDa, which could be efficiently separated from 20S proteasomes using sucrose density centrifugation. According to SDS-PAGE analysis the purified d20S proteasome sample had a higher purity compared to the 20S proteasome purified from HeLa cells. The SDS-PAGE band pattern of the fruit fly 20S proteasome differs from the human 20S proteasome. The fruit fly 20S proteasome showed a compact band pattern stack between 20 kDa – 30 kDa, whereas a separated high molecular band and two separated low molecular weight protein bands could be seen for the h20S (Fig. 3.1; A; Fig. 3.3; B).

### 3.2.2 *Saccharomyces cerevisiae* 20S proteasome purification

The adaption of the 20S proteasome purification protocol to the y20S purification required several modifications. The chloride ions of the purification buffer were replaced with acetate ions leading to more successful crystallization trails of the sample. In contrast to the HeLa S10 extract, the yeast lysate is more complex in its protein composition causing several protein complexes in a size range of around 700 kDa precipitating together with the 20S proteasome (Fig. 3.4; A). The higher complexity of the yeast lysate originates from harsher lysis of the cells that result in the disruption of cell organelles and subsequent release of their proteins into the cytoplasm. As it is not possible to separate protein complexes with similar sizes by sucrose density centrifugation, a precipitation step with 5% (v/v) PEG 6000 of the pooled fractions containing the 20S proteasome obtained after the second gradient was introduced. The different precipitation property of PEG 6000 was used to remove contaminating proteins, which could not be separated in previous steps. After removing the precipitate by centrifugation, the supernatant was acidified to pH 6.0 with acetic acid to further clear the 20S proteasome from impurities (Fig. 3.4; B). Alkalization to pH 6.5 with KOH restored the known 20S proteasome PEG 400 precipitation properties, enabling the use of the standard protocol for concentrating the 20S proteasome. Besides the described changes, the purification protocol could be transferred from the human purification protocol. The purification yielded 10 mg yeast 20S proteasomes per 270 g yeast cells corresponding to a wet weight of about 800 g. SDS-PAGE analysis demonstrated the high purity of the sample (Fig. 3.4; C). Two protein band pattern stacks of a size range of 20 – 30 kDa could be differentiated. The

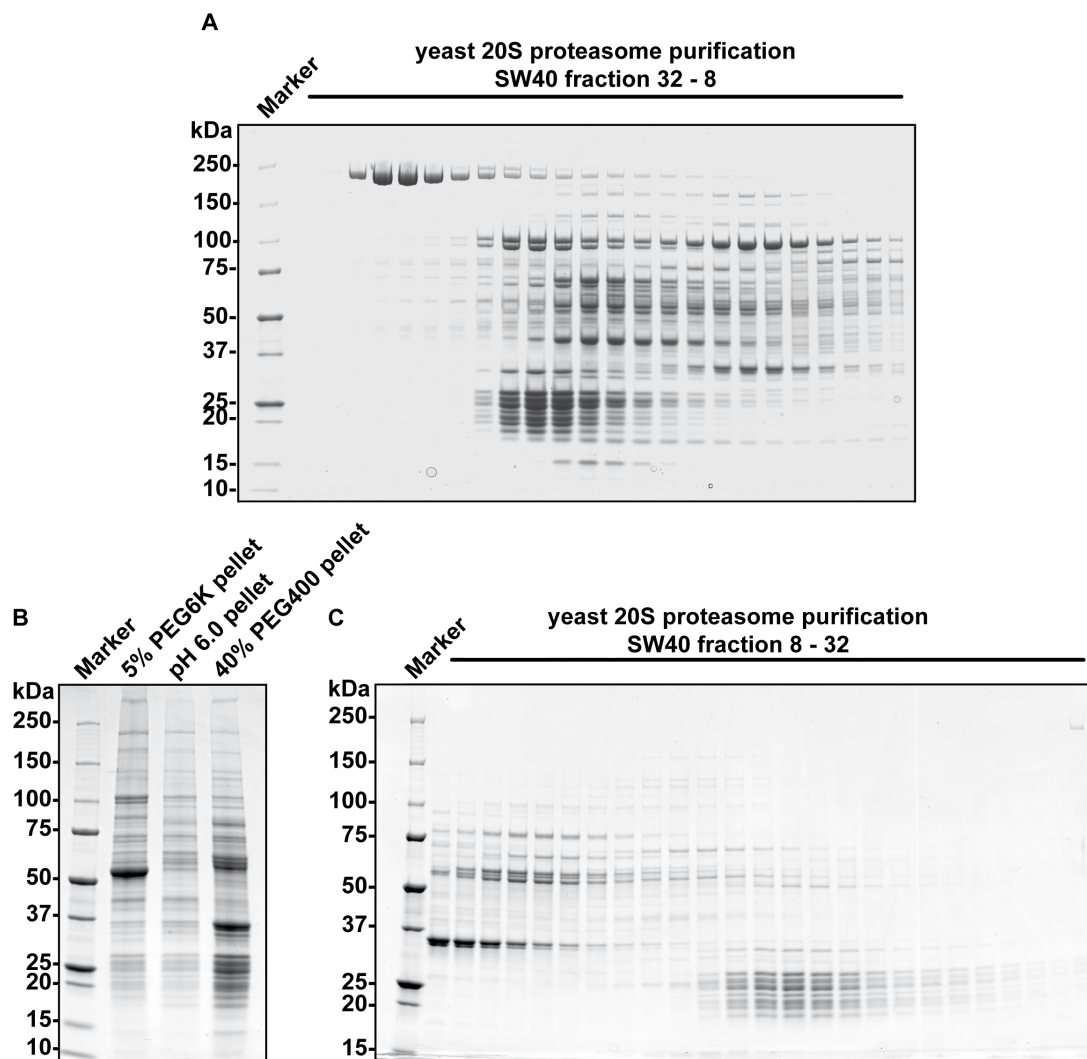


**Fig. 3.3: Comparison of fruit fly and human 20S proteasome purifications.** SDS-PAGE analyses of fruit fly (A) and human (B) fractions obtained after the second sucrose density gradient is shown. The protein composition in the fruit fly sample is less complex. The major impurity of the fruit fly sample is the yolk protein (40 kDa, fraction 11 – 14).

protein band stacks are separated by one protein band. One protein band is visible above the higher molecular weight stack and below the lower molecular weight stack.

### 3.2.3 *Thermoplasma acidophilum* 20S proteasome purification

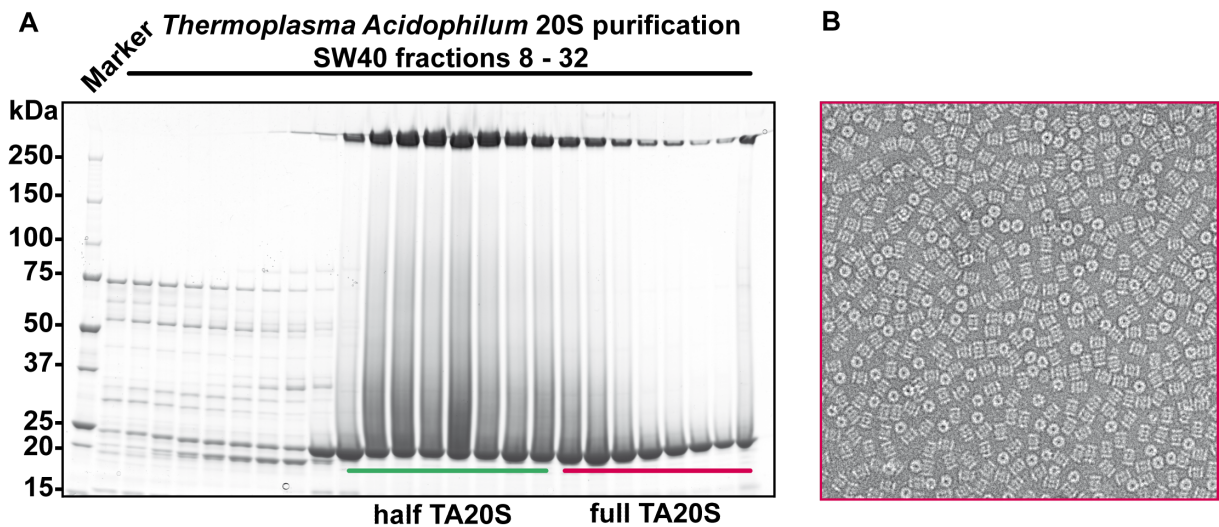
All previous described purification protocols were established for the purification of endogenously expressed human, fruit fly and yeast 20S proteasomes. To address the potential usage of the chromatography-free purification protocol to purify an exogenously expressed protein complex, *Thermoplasma acidophilum* 20S proteasomes expressed in *Escherichia coli* (*E. coli*) were purified. Since the 20S proteasome from *Thermoplasma acidophilum* is thermostable, the *E. coli* lysate was incubated at 65 °C for 15 min to



**Fig. 3.4: Intermediate purification steps of the yeast 20S proteasome.** A: SDS-PAGE analysis of the fractions harvested from the second sucrose density is displayed. The 20S proteasome containing fractions are still heavily contaminated with other protein complexes. B: Additionally introduced purification steps successfully removed impurities as shown by SDS-PAGE analysis of the pellets collected after each additional introduced precipitation step. C: SDS-PAGE analysis of fractions from the third sucrose density gradient show clean yeast 20S proteasome containing fractions.

achieve a crude clearance from impurities. A following polyethylenimine (PEI) precipitation effectively reduced contamination from nucleic acids. Without the PEI precipitation, the packing of the crystal lattice for *Thermoplasma acidophilum* 20S proteasomes was uncontrollable. After the heat treatment and PEI precipitation, PEG 400 precipitations and sucrose density gradient centrifugation, performed in accordance with the other purifications methods, yielded pure proteasome samples suitable for crystallization (Fig. 3.5; A). A critical step during sucrose density gradient centrifugation was the separation of the *Thermoplasma acidophilum* 20S half molecules, which consist of one  $\alpha$ - and  $\beta$ -ring, from the fully assembled *Thermoplasma acidophilum* 20S proteasomes. 80 mg *Thermoplasma acidophilum* 20S proteasomes were retrieved from 114 g of pelleted *E. coli* cells. The purity of the samples was verified by SDS-PAGE analysis and negative stain electron

microscopy (Fig. 3.5). The  $\alpha$ -subunit and the  $\beta$ -subunit cannot be distinguished by the used SDS-PAGE protocols, resulting in one observable protein band.

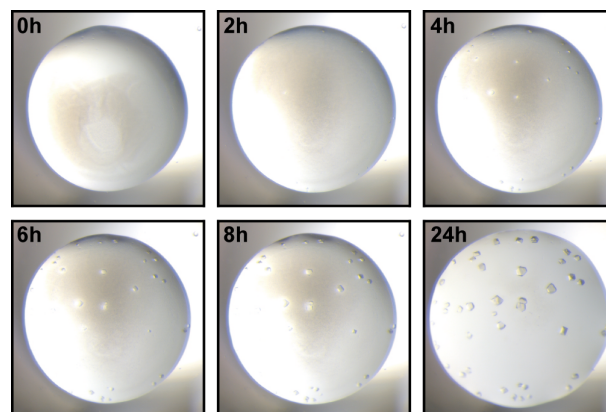


**Fig. 3.5: Analysis of *Thermoplasma acidophilum* 20S proteasome purification expressed in *E.coli*.** A: Fractions of the second sucrose density gradient performed during *Thermoplasma acidophilum* 20S proteasome purification are depicted. The two observable peaks show half *Thermoplasma acidophilum* 20S proteasomes, most likely two  $\beta$  rings, and full assembled *Thermoplasma acidophilum* 20S proteasomes. The intense bands at higher molecular weight (above 250 kDa) can be explained by the thermostability of the *Thermoplasma acidophilum* 20S proteasomes making them resistant to denaturation by sodium dodecyl sulfate polyacrylamide (SDS) and the preceding heat denaturation. B: Negative stain analysis of *Thermoplasma acidophilum* 20S proteasomes containing fractions validates the successful purification.

### 3.3 Crystallization of 20S proteasomes

20S proteasomes from *Homo sapiens* and *Drosophila melanogaster* crystallize in the same crystallization condition (Tab. 3.1). Crystals of the h20S proteasome grow overnight to a size of approximately  $200 \times 200 \times 200 \mu\text{m}^3$ , whereas crystal growth of d20S proteasomes take approximately seven days to reach a size of  $200 \times 200 \times 200 \mu\text{m}^3$  (Fig.3.6). The used crystallization condition differed from published 20S proteasome crystallization condition, which usually utilizes MPD as precipitant and not PEG 3350 [120].

The crystallization condition for T20S proteasomes included the same components as for the human proteasomes. However, the T20S proteasome crystallization tolerated a higher variation of PEG 3350 and  $\text{MgCl}_2$  concentrations as compared to h20S proteasomes or d20S proteasomes. The 20S proteasomes from *Saccharomyces cerevisiae* crystallized under the condition published by Groll [31]. Additionally, a crystallization condition utilizing PEG 400 was found by extensive screening (Tab. 3.1).



**Fig. 3.6: Crystal growth of the human 20S proteasome.** Time-lapse of crystal growth for human 20S proteasomes at 18 °C is shown. The corresponding time points are denoted in each image.

**Table 3.1: 20S proteasome crystallization conditions**

Organism	Crystallization condition	Space group	Unit cell constants	Matthews coefficient/ solvent content
<i>Homo sapiens</i>	0.1 M bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane (BisTris) pH 6.5, 0.2 M MgCl <sub>2</sub> , 10% (w/v) PEG 3350	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	a= 113 Å b= 203 Å c= 316 Å α=β=γ= 90°	2.59 / 53%
<i>Drosophila melanogaster</i>	0.1 M BisTris pH 6.5, 0.2 M MgCl <sub>2</sub> , 10% (w/v) PEG 3350	P 1	a= 117 Å b= 171 Å c= 212 Å α= 69° β= 88° γ= 70°	2.64 / 54%
<i>Thermoplasma acidophilum</i>	0.1 M BisTris pH 6.5, 0.1 - 0.25 M MgCl <sub>2</sub> , 7 - 10% (w/v) PEG 3350	P 1 2 <sub>1</sub> 1	a= 115 Å b= 198 Å c= 308 Å α= 90° β= 91° γ= 90°	2.62 / 53%
<i>Saccharomyces cerevisiae</i>	0.1 M MES pH 6.5, 0.04 M Mg(OAc) <sub>2</sub> , 12% (v/v) MPD	P 1 2 <sub>1</sub> 1	a= 134 Å b=301 Å c= 143 Å α= 90° β= 112° γ= 90°	3.82 / 68%
<i>Saccharomyces cerevisiae</i>	0.1 M BisTris pH 7.5, 0.05 - 0.2 M NaCl, 25 - 30% (v/v) PEG 400	P 1 2 <sub>1</sub> 1	a= 134 Å b=301 Å c= 143 Å α= 90° β= 112° γ= 90°	3.82 / 68%
<i>Saccharomyces cerevisiae</i>	0.1 M 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.5, 0.1 - 0.2 M NaCl, 25 - 30% (v/v) PEG 400	P 1 2 <sub>1</sub> 1	a= 134 Å b=301 Å c= 143 Å α= 90° β= 112° γ= 90°	3.82 / 68%

### 3.3.1 Crystal packing

All crystallization conditions found for the y20S proteasome resulted in the same crystal packing (Fig. 3.7, C). Like y20S proteasome crystals, T20S proteasome crystals belonged to the monoclinic space group, P 1 2<sub>1</sub> 1 (Fig. 3.7, D). Two additional two-fold screw axes could be found in the orthorhombic P 2<sub>1</sub> 2<sub>1</sub> 2<sub>1</sub> crystal packing of the h20S proteasome (Fig.3.7, A). No internal symmetries were present in the crystal packing of the d20S proteasomes (Fig.3.7, B). The asymmetric unit of the h20S proteasome and the y20S proteasome was composed of one molecule, whereas the asymmetric unit of the d20S proteasome and the T20S proteasome were comprised of two molecules.

The crystal packing of y20S proteasomes did not show any lateral interactions between the molecules but was dominated by top to bottom interactions of the α-rings, leaving large spaces between the molecules. Consequently, the Matthews Coefficient of 3.82 indicated a loose packing with a high solvent content of 68% (Tab. 3.1). The crystalline lattice

of T20S proteasomes, d20S proteasomes and h20S proteasomes showed a more dense packing structure demonstrated by decreased Matthews coefficients (2.59 – 2.75) and solvent contents compared to the y20S proteasome crystals (Tab. 3.1).

Two 20S proteasome molecules which interact laterally and are shifted vertically to each other compose the asymmetric unit of the d20S crystal (Fig. 3.7, B). The contacts between the molecules in the asymmetric unit are formed by loops connecting  $\beta$ -strands and  $\alpha$ -helices. The symmetry related asymmetric unit in z-direction form contacts between the  $\alpha$ -helices with molecule 1 and through planar top to bottom  $\alpha$ -ring interactions of molecule 2. The asymmetric units in x-direction of d20S proteasome are held in place by interactions with the layer of asymmetric units in z-direction. The lateral distance between the molecules in the asymmetric unit is shorter than the distance to the neighboring asymmetric unit, as the crystal contact between the  $\alpha$ -helices determines the distance for the asymmetric unit following in x-direction.

The h20S proteasome molecule is connected with its symmetry related molecules via lateral  $\beta$ -ring –  $\beta$ -ring and  $\alpha$ -ring –  $\alpha$ -ring interactions (Fig. 3.7, A). The next molecule layer cannot stack in a planar manner on top of the 20S proteasome, resulting in an inclination of the subsequent crystal plane. Therefore, the top to bottom  $\alpha$ -interactions are only present on one side of the  $\alpha$ -rings, whereas a solvent filled space can be found on the other side.

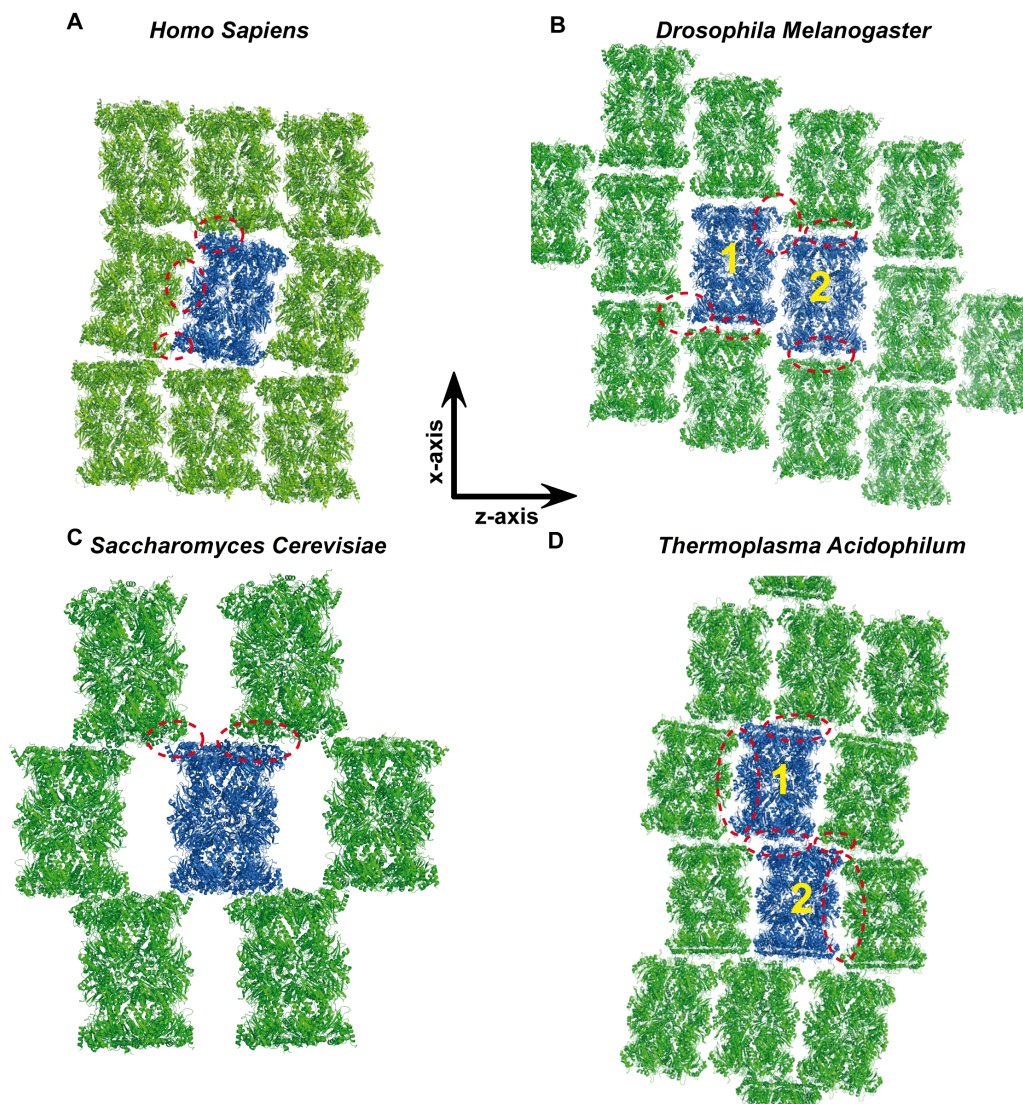
The T20S proteasome molecules form the asymmetric unit by interactions of their  $\alpha$ -rings (Fig. 3.7, D). The molecules are positioned on top of each other but are horizontally shifted. Like for the h20S, the molecule layer in x-direction is inclined. The N-terminal gate forming residues of the  $\alpha$ -rings seem to prevent the on axis planar stacking of the 20S proteasomes on top of each other.

### 3.3.2 Post-crystallization treatment

The h20S proteasome crystals were used to establish a post-crystallization treatment, as h20S proteasomes were the first species that could be purified in sufficient amounts. The highest resolved structure of the h20S proteasome in the protein data bank (PDB; [www.rcsb.org](http://www.rcsb.org); 28.08.1018) is limited to a resolution of 2.6 Å [120]. Moreover, a high-resolution structure of the h20S proteasome has the highest significance for drug development.

First, suitable cryo-protectants were screened. Cryo-protectants are essential to avoid ice formation while freezing the crystal in liquid nitrogen. A chemical compound functioning as a cryo-protectant needs to preserve the diffraction properties of the crystals. Crystals soaked in crystallization buffer supplemented with either 20% (v/v) MPD or 20% (v/v) PEG 400 at 18 °C routinely diffracted to 2.8 Å. Further dehydration by a stepwise increase of the PEG 3350 to 25% (v/w) at 18 °C resulted in poorer diffracting crystals ( $\approx$





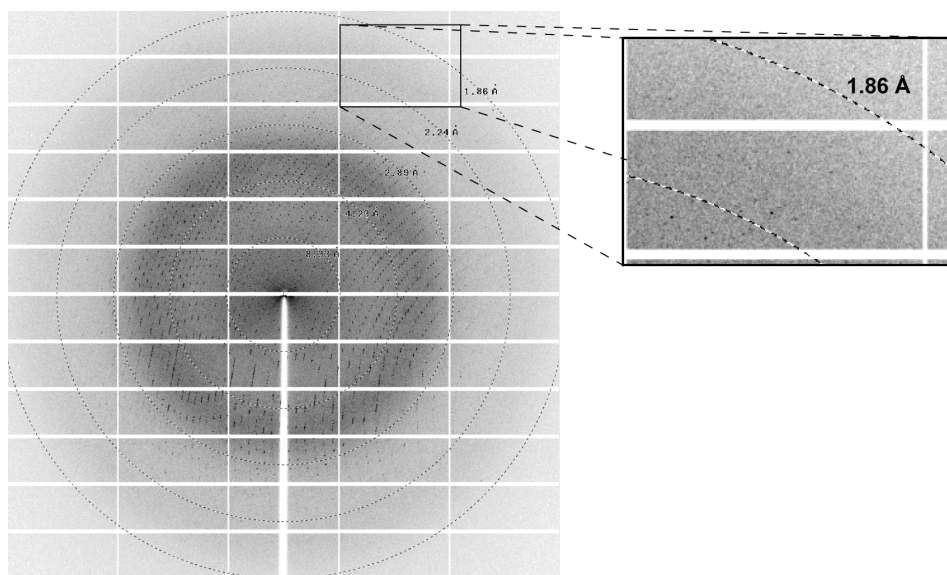
**Fig. 3.7: Comparison of the crystal packing of 20S proteasomes from different species.** Crystal contacts to adjacent asymmetric unit cell are encircled in red. Molecules for *D. melanogaster* and *T. acidophilum* were numbered as they contain two molecules in the asymmetric unit. A: human 20S proteasome. B: fruit fly 20S proteasome. C: yeast 20S proteasome D: archaeal 20S proteasome.

3.5 Å). Transfer of the crystals to 4 °C, increase of the PEG 3350 concentration to 15% in the drop and exchange of the reservoir solution with crystallization buffer containing 25% (v/w) PEG 3350 and 20% (v/v) MPD decreased the resolution limit to 2.0 Å on average, with the best crystals diffracting up to 1.8 Å. (Fig. 3.8).

The post-crystallization protocol established for h20S proteasomes was also applied to T20S proteasome and d20S proteasome crystals. Similar to the h20S proteasome crystals, the obtained resolutions for T20S proteasome and d20S proteasome crystals were on average around 2.2 Å.

A different post-crystallization protocol was used for y20S proteasome crystals because the crystallization conditions significantly differed from the crystallization condition of h20S, d20S and T20S proteasomes. Since the increase of cryo-protectant through vapor diffusion

is milder at 4 °C as observed for h20S proteasomes crystals, y20S proteasome crystals were also transferred to 4 °C, before the reservoir was exchanged with crystallization buffer containing either 40% MPD or 40% PEG 400 depending on which crystallization condition was used. Although, reflections up to 1.7 Å were observed along the  $c^*$ -unit cell axis in the diffraction patterns, the resolution of y20S proteasome crystals was limited to about 2.3 Å, due to poorer diffraction along the short axes resulting from anisotropic diffraction of the crystals (Fig. 4.1; B).



**Fig. 3.8: Exemplary diffraction pattern of 1.8 Å human 20S proteasome dataset.** Reflections up to 1.8 Å after applying the post-crystallization treatment are observed.

### 3.3.3 Small molecule soaking into 20S proteasome crystals

The soaking approach to introduce small molecules into the 20S proteasome was chosen because co-crystallization attempts with most small molecule proteasome inhibitors prevented crystallization of 20S proteasomes. Moreover, the soaking protocol can be better incorporated into a high throughput pipeline for screening small molecules. Several small molecules were successfully soaked into 20S proteasomes that were validated by solved X-ray structures. The compounds, which were dissolved in DMSO, were diluted to 4 mM with the respective cryo-buffer and added in a 1:1 ratio to the crystallization drops. Crystals were harvested after incubation for at least 48 h. Native and crystals soaked with small molecules diffracted to similar resolutions (1.9 – 2.3 Å), hence the soaking procedure did not significantly impair the diffraction properties of the crystals (Tab. 3.4, 3.5).



### 3.4 Structural analysis of the native human 20S proteasome

The h20S proteasome structure was solved with a nominal resolution of 1.8 Å. Besides the Photosystem II and the y20S proteasome, no other structure with comparable molecular weight and subunit complexity has yet been deposited into the PDB at the described resolution (24.08.2018). In total, 633'728 unique reflections with a redundancy of 13.4 were collected providing a completeness of 99.9% of the measured reflections. The highest-resolution shell used for structure determination included reflections up to 1.8 Å and showed an  $I/\sigma I$  value of 0.9. The  $CC_{1/2}$  value of 0.32 in the highest-resolution shell justified the inclusion of this resolution shell for structure refinement. The structure was refined to an overall  $R_{\text{work}}/R_{\text{free}}$  of 0.184/0.212. The refinement of the high-resolution shell resulted in a  $R_{\text{work}}/R_{\text{free}}$  of 0.383/0.388. The B-factors range from 25 in the most rigid areas, predominantly present in the  $\beta$ -subunits, to 120 in the C-terminal parts of the  $\alpha$ -subunits. The B-factors of the atoms increased with solvent accessibility and lack of stabilization by crystal contacts to residues lying in close proximity. In particular, the C-terminal residues of the  $\alpha$ -rings seem to be highly flexible indicated by the highest B-factors observed in the structure (Tab. 3.4).

The overall structure was similar to the y20S proteasome, which corresponds to the similarity in amino acid sequence. All secondary structure elements annotated in the h20S proteasome structure published by Harshberger and colleagues matched with elements observed in the newly solved h20S proteasome structure [120]. Differences of the  $\alpha$ -chain trace could be mainly found in the  $\alpha$ -rings. The density map of the solved structure was more isotropic and better defined for amino acid residues compared to the published 2.6 Å structure (Fig. 3.9; C and D) [120]. The good map quality helped to correct for several wrong conformations of amino acid residues and to model parts of the structure, such as the C-terminal parts of the  $\alpha$ -rings, with higher confidence. Alternative conformations could be modeled and refined for several amino acid residues (Fig. 3.9).

The most striking difference between the structures was the number of detected solvent molecules, metal ions and chloride ions. Notably, a total number of 3629 water molecules was observed in the structure, compared to the previously published structures, in which 185 water molecules were modeled [120]. Thus, the water molecules increased by a factor of 20. Moreover, 58 chloride ions, 10 magnesium ions and 6 potassium ions could be clearly identified. How the improved resolution sharpens the electron density can be seen in figure 3.9. The identical residues are shown in figure 3.9 C and D for the published 2.6 Å structure and the structure solved at 1.8 Å. The higher resolution helped to analyze solvent molecules and metal ions in the structure. A potassium ion coordinated by three water molecules and three backbone carbonyls could be identified in the interface between

$\beta$ 6- and  $\beta$ 2-subunit (Fig. 3.9, A). However, the coordination geometry for the potassium ion is difficult to determine, as two atoms are missing to complete the coordination sphere of eight atoms. The depicted magnesium ion is octahedrally coordinated by one water molecule, three backbone carbonyls and the carboxyl group of Asp213. However, one atom is missing for the completion of the coordination sphere. Four of the ten magnesium ions are fully coordinated octahedrally by six water molecules (Fig. 3.9; B).

Closer investigation of all different proteolytic active sites revealed a complex network of water molecules and the existence of one chloride ion, which was previously not described. The chloride ion is positioned at the same position in each active site. In previously published structures the chloride position was annotated to a catalytic water molecule [82]. The water content decreases starting from the caspase-like site, chymotryptic-like site to the tryptic-like site, which contains the fewest water molecules of all active sites (Fig. 3.9; E, F, G). For a consistent nomenclature of the water molecules present in the proteolytic active sites, the numbering of the water molecules in the caspase-like site was used as a template for the other active sites. The tryptic-like site often contains a substrate-like density, which is most likely a co-purified, partially occupied substrate. As the occupancy is very low and no sequence information about this substrate is available, modeling of this electron density was not possible.

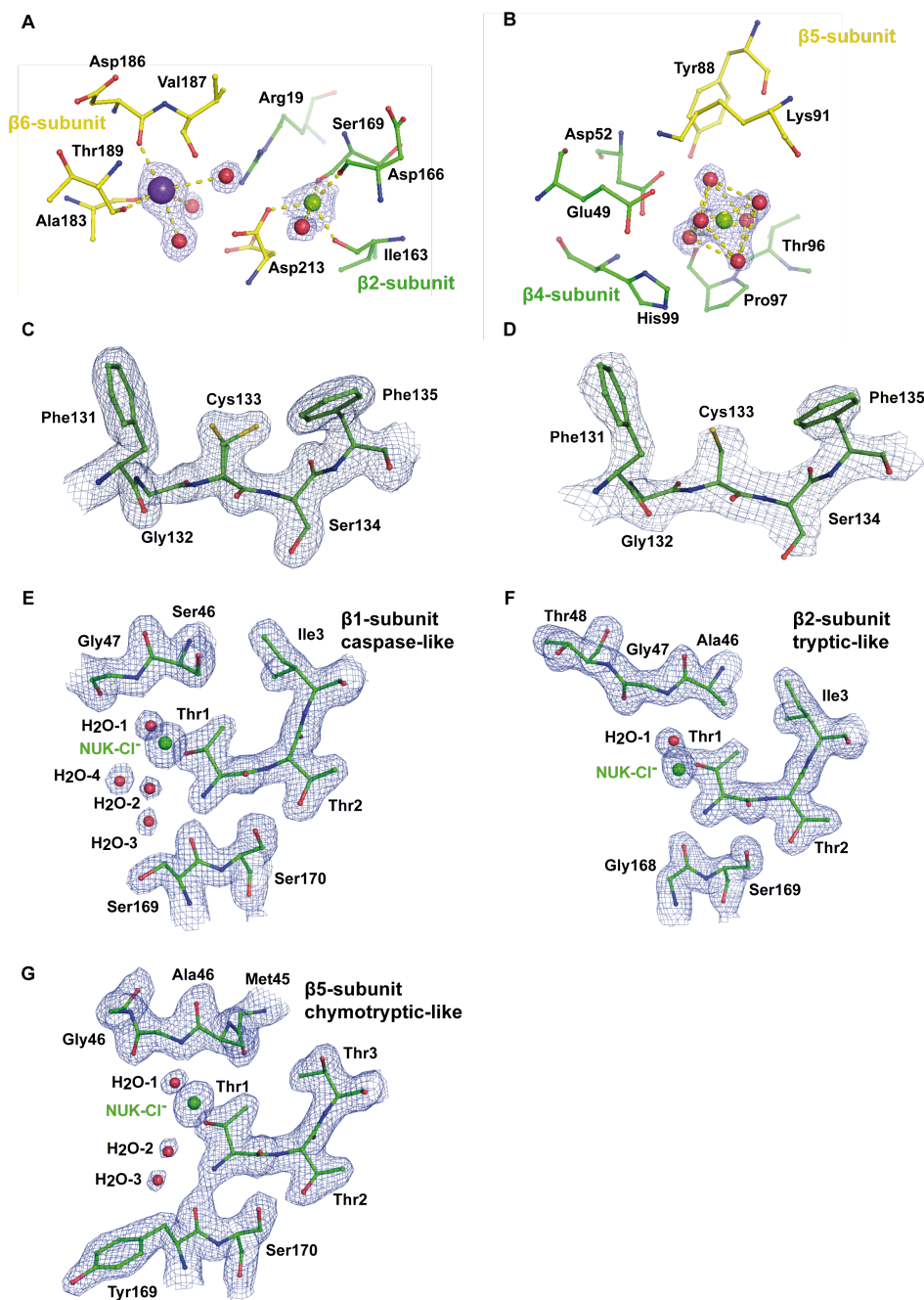
### 3.4.1 Validation of modeled ions

Ions were modeled based on detected electron density peaks. Different ions were differentiated according to their polar contact distances to their neighboring atoms (Tab. 3.2). As the distances are rough empirical estimates based on published structures, ions were further validated by anomalous scattering.

**Table 3.2: Atomic distances used for modeling ions in the structure [152]**

Ions/ solvent	Water	Main chain carbonyl	Monodentate carboxylate
Magnesium	2.09 Å	2.26 Å	2.07 Å
Potassium	2.82 Å	2.80 Å	2.80 Å
Chloride		3.1 - 3.7 Å	
Water	2.3 Å	-	-

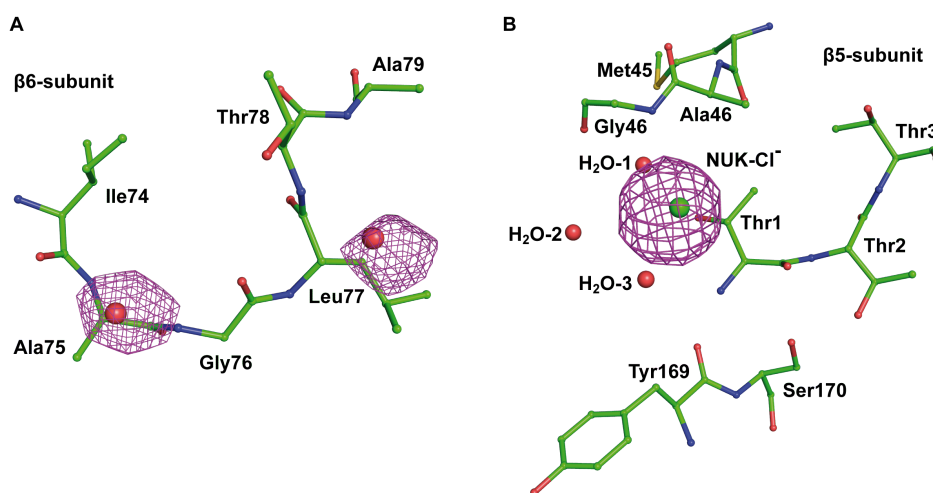
Anomalous signals from chloride ions were measured at 6.5 keV. Additionally, chloride ions were exchanged with bromide ions, as the anomalous signal of chloride ions at 6.5 keV is very weak. Bromide ions have a strong anomalous signal at 13.47 keV leading to



**Fig. 3.9: Structural features and electron densities of the 1.8 Å human 20S proteasome structure.** A: Potassium-magnesium binding site formed between  $\beta 6$  and  $\beta 2$  is selected to show the quality of the electron density map. Electron density maps for ions and water molecules are well defined. B: Electron density map for a magnesium ion octahedrally coordinated by six water molecules is depicted. C + D: Both figures show the same structural detail of the human 20S proteasome. The electron density maps of our structure (C) and of the 2.6 Å structure published by Harshberger et al [120] are compared. Electron densities are better defined and an alternative conformation of the Cys133 could be observed in our structure. E – G: All proteolytic sites - caspase-like (E), tryptic-like (F) and chymotryptic-like site (G) are shown. A previously not detected chloride ion could be observed in all proteolytic active sites. In all figures of the human 20S proteasome solved at 1.8 Å the electron densities maps for amino acids, ions and water molecules are well defined. All electron density maps are contoured at  $1.5\sigma$ .

pronounced anomalous difference map peaks (Fig. 3.10). Moreover, the shorter wavelength allowed the acquisition of higher resolved reflections compared to data acquired at 6.5 keV because the wavelength determines the distances between lattice planes that can be resolved according to Bragg's law [152]. Modeled chloride ions were verified with both approaches. In particular, the verification of the chloride ions present in the catalytic active sites was of great interest since an electron density peak at the same positions was already observed in maps of the  $\gamma$ 20S proteasome but was interpreted as catalytically important water molecule in these structures. Knowing that a chloride substituted these water molecules in the presented structures could therefore change the current understanding of the catalytic mechanism underlying the proteasome function.

To unambiguously distinguish magnesium ions from potassium ions, the magnesium ions were replaced with manganese ions to measure their anomalous signal at 6.5 keV. Anomalous differences in manganese-soaked crystals confirmed the positioning of all magnesium ions. Furthermore, the weak anomalous signal of potassium at 6.5 keV was used to prove the localization of the potassium ions in the native crystal.



**Fig. 3.10: Anomalous diffraction map of bromide ions used to validate chloride positions.**

A: Two modeled water molecules were identified as chloride ions. B: Modeled NUK-CL<sup>-</sup> in the chymotryptic-like site was validated by anomalous diffraction analysis. Electron density maps are contoured at  $4\sigma$

### 3.4.2 Structural investigation of inhibitor design

The established purification, crystallization and soaking protocols provide a framework to analyze binding of inhibitors to the 20S proteasome and their inhibition mechanism at unprecedented resolutions. In total, 20 different inhibitors from 5 inhibitor classes were successfully soaked into the human 20S proteasome. All inhibitors consist of a two to four peptide long backbone with varying sidechains, divergent electrophilic groups at the

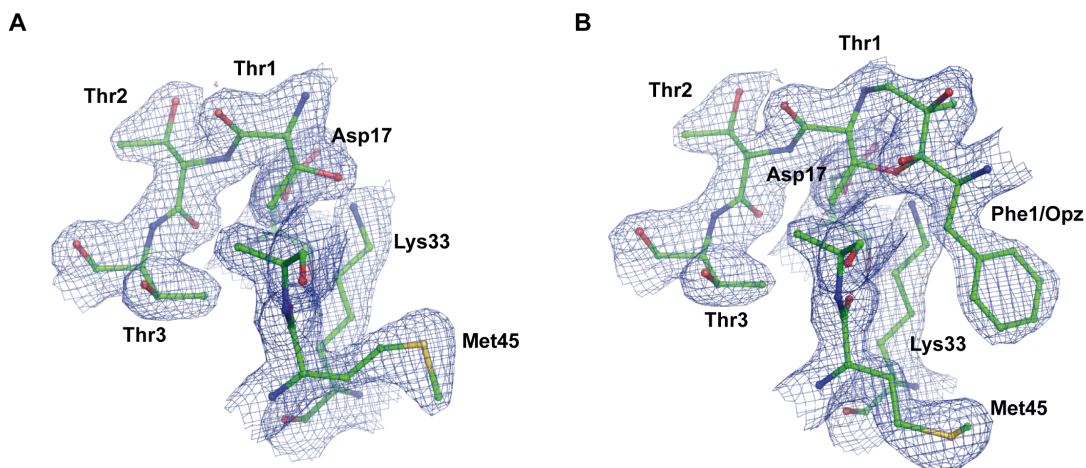
C-terminus and inhibitor specific protecting groups at the N-terminus. The resolution of the obtained proteasome-inhibitor structures ranged from 1.9 to 2.4 Å for the different 20S-inhibitor structures (Tab. B.2).

Data was collected with a minimum redundancy of five. The refinement statistics for all inhibitor structures were comparable to the refinement statistics of the native 20S proteasome structure (Tab. 3.4; 3.5). However, the observed ligand and solvent content was generally higher than for the native structure despite their worse resolution. For all solved inhibitor structures, clear electron densities for the inhibitor residing in the proteolytic active pockets were visible. The covalent linkages and the amino acid residues of the inhibitors fit accurately in the electron densities. The more flexible N-terminal parts of the inhibitors could not be modeled for all of the inhibitors. B-factors of the inhibitors, after refining the structure including the modeled inhibitor, showed values ranging from 34 to 50. These low values and the fact that they fit with the B-factors obtained for the surrounding atoms indicate a high occupancy, correct modeling and tight binding of the inhibitor.

Depending on the inhibitor the occupancy was different in each of the three unique proteolytic sites (Tab. B.2). In the Oprozomib chloride structure, the inhibitor was only present in the tryptic-like and chymotryptic-like site. Whereas, Oprozomib could be found in the caspase-like site in crystals, in which the chloride ions are exchanged with acetate ions. As for the described native proteasome structure, the tryptic-like site contained a non-interpretable impurity, which overlaid with inhibitor densities. Consequently, it was difficult to determine real occupancies of the inhibitor for the tryptic-like site.

For all inhibitors, the highest occupancy was found in the chymotryptic-like site judged by the peak height of the electron densities. In addition, the chymotryptic-like site was the only active site that was occupied by every analyzed inhibitor. Hence, binding analyses and mechanistic interrogations were restricted to this proteolytic site. Furthermore, the  $\beta$ 5-subunit, harboring the chymotryptic-like site, contains the reporter amino acid Met45 that indicates binding of an inhibitor by flipping to an alternative conformation (Fig. 3.11; A,B). The ratio between the two alternative conformations could be utilized to assess the occupancy of the inhibitor in the chymotryptic-like site.

Binding of the inhibitors to the active site is mediated through interactions between their peptide backbones with the main chains enclosing the active site. As already described in the literature, the  $\beta$ -sheet formed by strands S1, S2 and S8 is extended by inhibitor binding (Fig. 3.12; A) [99, 120]. The amino and carbonyl-groups of the peptide bonds connecting the residue in P3 position with the residue in P2 position and between the P2 and P1 set up the hydrogen bond network to extend the  $\beta$ -sheet. Most likely, the binding observed for the inhibitor resembles the interaction of an authentic substrate. The positioning of the peptide backbone in the active site was not affected by the different



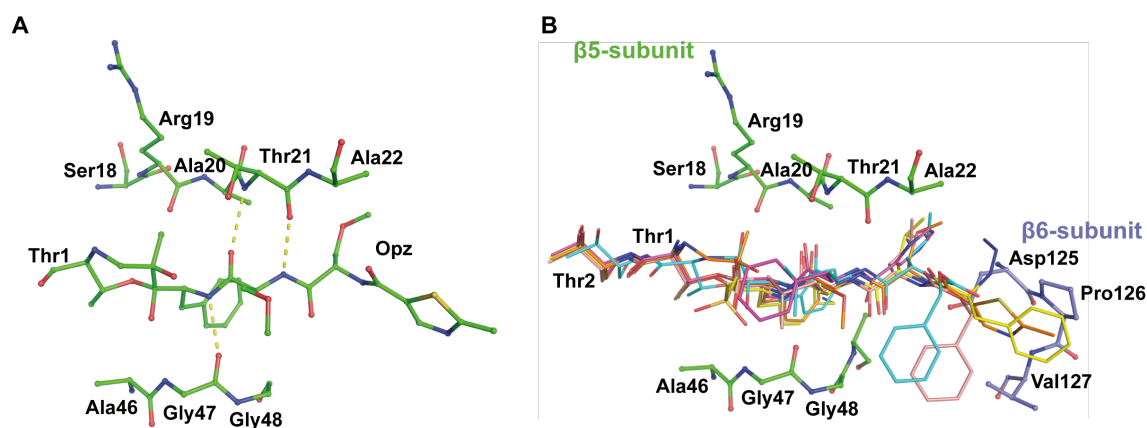
**Fig. 3.11: Alternative conformations of Met45 in the chymotryptic-like site** A: Met45 position in the native 20S proteasome structure is depicted. B: Met45 adopts a second conformation upon inhibitor binding. A 20S proteasome structure inhibited with Oprozomib was used to prepare the figure. Electron density maps are contoured at 1.5σ.

electrophilic groups shown by an overlay of inhibitor structures (Fig. 3.12; B). Differences between the inhibitors are restricted to the arrangement of the amino acid residues and the conformation of the N-terminus. Binding of the inhibitor did not lead to any obvious conformational rearrangements of the protein main chain and its residues. The only exception to this was the previously mentioned structural rearrangement of Met45 upon inhibitor binding to the chymotryptic-like site.

Most of the water molecules present in the active site were displaced after inhibitor binding. Depending on the inhibitor, certain water molecules or chloride ions remained in place. The water molecule connecting residue  $\beta 5$ /Ala50 with  $\beta 6$ /Asp125 was conserved in all structures. In some structures the chloride ion, which was detected in the native structure, was replaced by a water molecule after inhibitor binding, though its exact positions varies depending on the inhibitor. The third conserved water molecule could be found in the interspace formed between the inhibitor and residues Thr21, Ala20, Arg19 and Tyr169 of the chymotryptic-like site. The accurate spot of this water molecule changes in an inhibitor-dependent manner.

### 3.4.3 Structural investigation of inhibition mechanisms

Proteasome inhibition has already been shown to be a valid strategy to fight cancer. Understanding the chemical principles of proteasome inhibition is key for the development of more potent proteasome inhibitors and to investigate the basis of the diverse potencies of available inhibitors. For drug development specificity in targeting a certain protein



**Fig. 3.12: Structural features of inhibitor binding to the chymotryptic-like site.** A: Formed hydrogen network upon inhibitor binding is exemplified with Oprozomib. B: Inhibitors of each inhibitor class are overlaid (Aldehyde: rose, Boron: pink, Vinylsulfon: yellow, Ketoaldehyde: blue, Epoxyketone: orange) binding to the chymotryptic-like site (B).

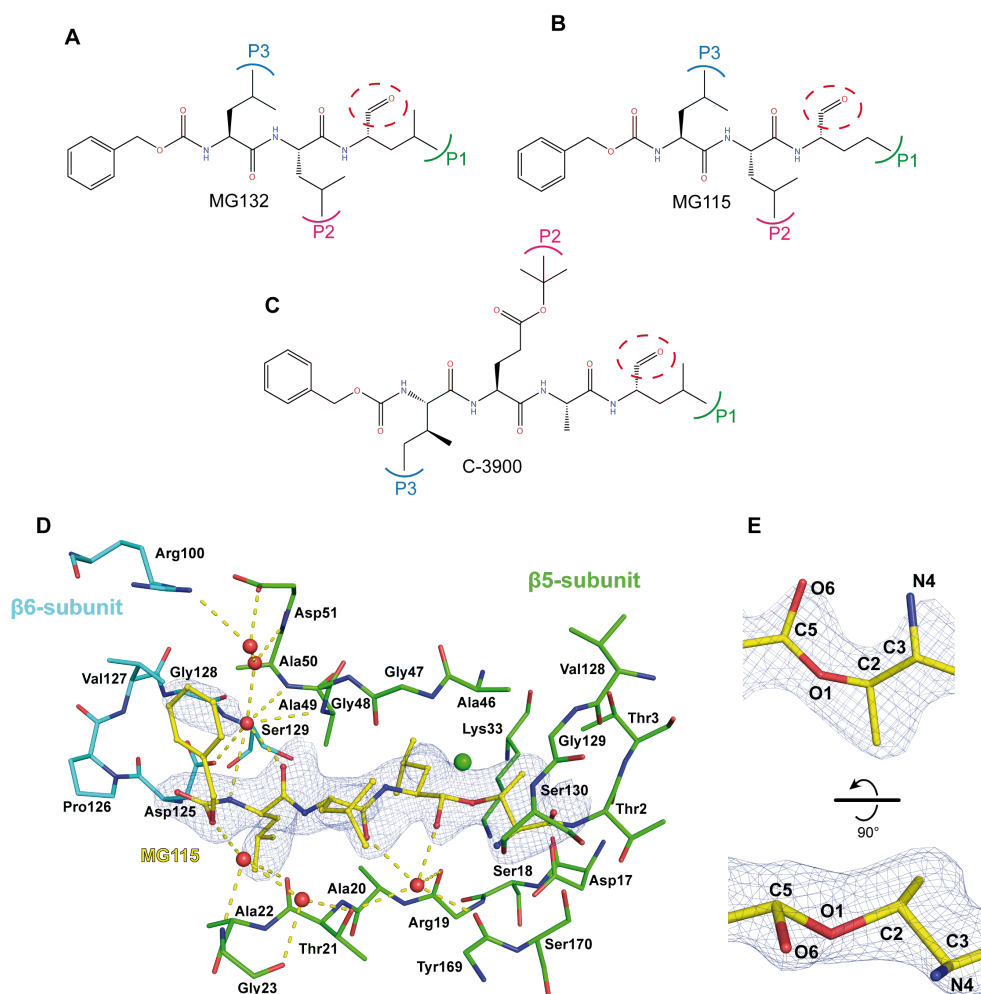
complex is essential [153]. The proteolytic active site and the catalytic active N-terminal threonine are unique features of the proteasome. Understanding the underlying principles of inhibitor binding to these pockets and the covalent modification of the threonine is the key for drug development. With this knowledge 20S proteasome specific inhibitor can be designed and the unspecific inhibition of proteases can be excluded. The newly solved human 20S proteasome-inhibitor structures reached resolutions of 1.9 – 2.2 Å. The improved resolution of the structures justified reanalyzing inhibition mechanisms in the human system, which were previously published the 20S proteasome.

### 3.4.3.1 Aldehyde inhibitors

Aldehyde inhibitors are the most ancient proteasome inhibitors. Originally, they were developed for the inhibition of other proteases [154]. In this study, three different aldehydes were used for the analyses of the aldehyde inhibition mechanism (Fig. 3.13; A - C). The occupancy of the aldehydes in the proteolytic active site was dependent on the particular inhibitor. The aldehydes MG132 and MG115 bound to all proteolytic active sites, whereas the aldehyde C-3900 only binds to the tryptic-like and chymotryptic-like sites.

The inhibitor structures could be confidently modeled into the electron density maps observed in the chymotryptic-like site for all aldehyde inhibitors (Fig. 3.13; D). The solved structures all refined to  $R_{\text{work}}/R_{\text{free}}$  values of 0.18/0.23, which are similar to the values obtained for the native chloride human proteasome structure (Tab. 3.4). All three structures of the human proteasome inhibited with an Aldehyde confirmed the same inhibition mechanism, which was already proposed by Löwe and colleagues [25]. The nucleophilic attack of the Thr1 hydroxyl group at the aldehyde carbon resulted in the formation of





**Fig. 3.13: Aldehyde inhibition of the chymotryptic-like site.** A-C: Different aldehyde inhibitors analyzed in human 20S proteasome structures are drawn. D: MG115, colored in yellow, binds to the chymotryptic-like site, which is formed between the  $\beta 6$ -subunit, colored in blue, and the  $\beta 5$ -subunit, colored in green. The chloride ion is shown in green and the water molecules in red. The hydrogen-bond network of the water molecules is visualized with yellow dashed lines. The electron density map of the inhibitor is shown as an omit map contoured at  $4\sigma$ . E: The electron density map for the aldehyde linkage to Thr1 is shown as an omit map contoured at  $6\sigma$ .

a hemiacetal group (Fig. 3.13; E). The resulting hydroxyl group coordinates a water molecule, which is preserved in the native structure. The chloride ion position does not differ from its original position in the native structure.

Comparison of the presented proteasome-aldehyde structures with y20S proteasome structures inhibited with an aldehyde solved at  $2.4 \text{ \AA}$  by Stein and colleagues (PDB: 4NNN) [155], revealed another spatial orientation of the hemiacetal hydroxyl. In contrast to the yeast structure, in which the hydroxyl faces the oxyanion hole, the hydroxyl in the human structure points in the same direction as the amine of the Thr1 (Fig. 3.13; D, E). Thus, the hydroxyl of the human structure is rotated by approximately  $120^\circ$  respective to the aldehyde-inhibited yeast proteasome structure. A small positive electron density map



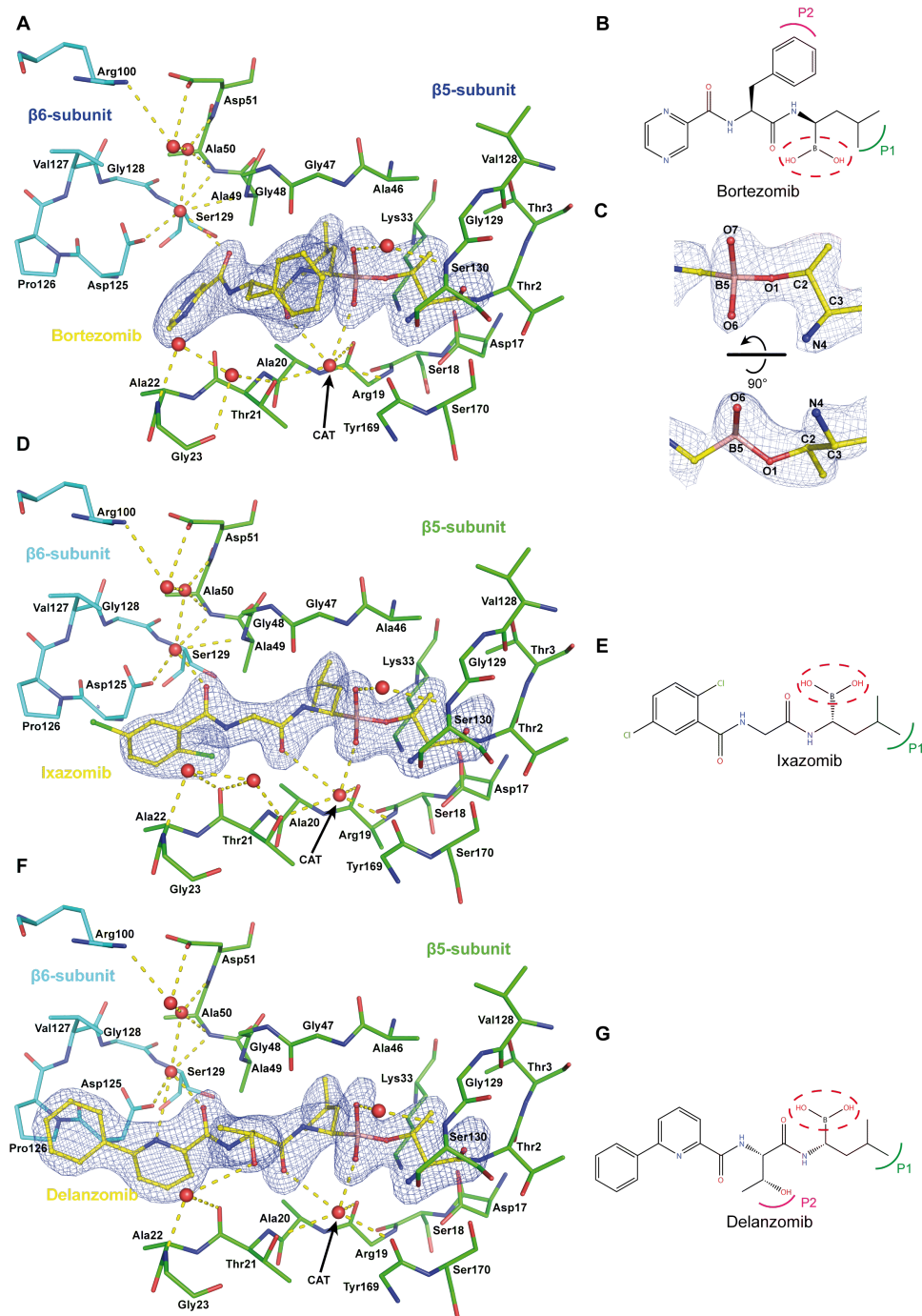
could be detected in our structure at the position where the hemiacetal hydroxyl resides in the aldehyde-inhibited yeast 20S proteasome structure. This suggested the existence of another hemiacetal conformation with low occupancy in our structure.

### 3.4.3.2 Boronic inhibitors

The first approved FDA proteasome inhibitor was Bortezomib, a member of the boronic inhibitor class. So far, no human structure inhibited with a boronic inhibitor has been published. Inhibition of three different boronic inhibitors including Bortezomib, which differ in their P2 site, were analyzed by soaks into h20S proteasome crystals. Bortezomib carries an aromatic phenylalanine at P2 (Fig. 3.14; B), Ixazomib an aliphatic and non-polar glycine (Fig. 3.14; E) and Delanzomib a polar threonine (Fig. 3.14, G; Table 3.3). Bortezomib and Ixazomib showed a high occupancy for all three proteolytic sites. In contrast, Delanzomib only bound to the caspase-like and chymotryptic-like site. All atoms of the inhibitors could be modeled into the respective electron densities. Inhibitor B-factors of the refined inhibitors-proteasome structures range from 33 to 49. Refined R-values are comparable to the R-values obtained for the native human 20S proteasome structure (Tab. 3.4).

The overlay of the human structure inhibited with Bortezomib and the corresponding yeast structure validated the published inhibition mechanism for boronic inhibitors deduced from the y20S proteasome structure [103]. The lone pair of electrons belonging to the threonine oxygen attacks the boron of the boronic acid group, which acts as a Lewis acid. The formed tetrahedral boronate complex is stabilized by hydrogen bonds to the boronic acid groups (Fig. 3.14; C). One boronic acid group is hydrogen bridged to Gly47/N and the other one to Thr1/N (Fig. 3.14; A, D, F). The hydrogen bond length between Gly47/N is in the same range as reported for the y20S proteasome structure. Whereas, the hydrogen bond length between the boronic acid group and the Thr1/N is shortened from 2.90 Å to 2.55 Å in the human structure. Two newly described water molecules present in the boronic inhibited structures have an additional stabilizing effect: one is located at the position that was occupied by the chloride ion in the native structure. This water molecule is coordinated by hydrogen bonds to one boronic acid group and to the Ser130/N. The second water molecule coordinates the second boronic acid group and is stabilized by hydrogen bonds to Thr21/O $\gamma$  and Tyr169/O.

The residues in P1 and P2 position do not show any interactions with the main chain. The main chain interactions are restricted to the peptide backbone of the inhibitor. A hydrogen bond is also formed between the pyrazine protecting group of Bortezomib and the Thr21/O $\gamma$ . This hydrogen bond is unique for the structure containing Bortezomib.



**Fig. 3.14: Inhibition of the chymotryptic-like site by boronic inhibitors** A, D, F: Bortezomib (A), Ixazomib (D) and Delanzomib (F), colored in yellow, bind to the chymotryptic-like site, which is formed between the  $\beta 6$ -subunit, colored in blue, and the  $\beta 5$ -subunit, colored in green. The water molecules are shown as red dots. The hydrogen-bond network of the water molecules is visualized with yellow dashed lines. The proposed catalytic water molecule, which stabilizes the tetrahedral intermediate state is marked with CAT. The electron density maps for the inhibitors are shown as omit map contoured at  $4\sigma$ . B, E, G: Different inhibitors belonging to the boronic class analyzed in 20S proteasome structures are depicted. D: C: The electron density map for the boronic linkage of Bortezomib to the Thr1 is shown as an omit map contoured at  $6\sigma$ .

**Table 3.3: Peptide backbone sequences of boronic inhibitors**

Inhibitor	Sequence peptide backbone
Bortezomib	Leu-Phe
Ixazomib	Leu-Gly
Delanzomib	Leu-Thr

### 3.4.3.3 Vinyl sulfon inhibitors

The main application of vinyl sulfon inhibitors is the site-specific labeling of active sites [112]. LU-002c, which was shown to selectively bind to the tryptic-like site of the constitutive human 20S proteasome, was utilized to analyze the inhibition mechanism of the vinyl sulfon electrophilic group.

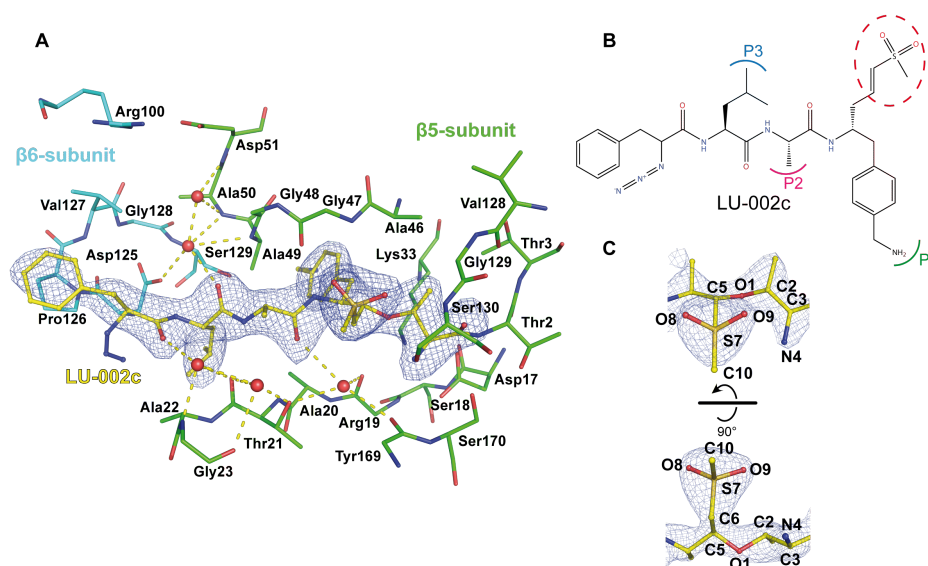
The solved structure refined to a  $R_{\text{work}}/R_{\text{free}} = 18.7/23.3$ . This value is comparable to the native structure. Electron densities for the inhibitor were found in the tryptic-like and chymotryptic-like site. The modeled inhibitor reached a B-factor of 56 in both sites after refinement (Tab. 3.5).

The reaction mechanism via a Michael addition that was previously proposed in published literature is supported by the solved structures (Fig. 3.15; C)[103]. First, the nucleophilic conjugate addition of the Thr1 hydroxyl group, the Michael donor, takes place at the vinyl, the Michael acceptor, of the vinyl sulfon to form a  $\beta$ -sulfonyl ether linkage. The reaction is completed by the protonation of the remaining unsaturated carbon. The  $\beta$ -sulfonyl ether linkage could confidently be modeled into the electron densities of the inhibitor (Fig. 3.15; B). The formed methyl-sulfon displaces the chloride ion facing the oxyanion hole.

### 3.4.3.4 Ketoaldehyde inhibitors

Ketoaldehyde inhibitors target the hydroxyl group and amino group of Thr1. The soaked Ketoaldehyde targets all proteolytic active sites. The electron density of the inhibitor visible in the chymotryptic-like site is the most distinct density compared to the other proteolytic active sites (Fig. 3.16; A). All three residues (LLY) fit into the electron density map. The structure was refined to R-values of  $R_{\text{free}}/R_{\text{work}} = 0.185/0.226$ . The N-terminal protecting group was too flexible to be modeled with high accuracy. The refined inhibitor has a B-factor of 45, which proved its correct placement (Tab. 3.5).

Comparison of the bound inhibitor in the structure solved here to the published y20S proteasome-ketoaldehyde inhibitor structure confirmed the same position of the peptide

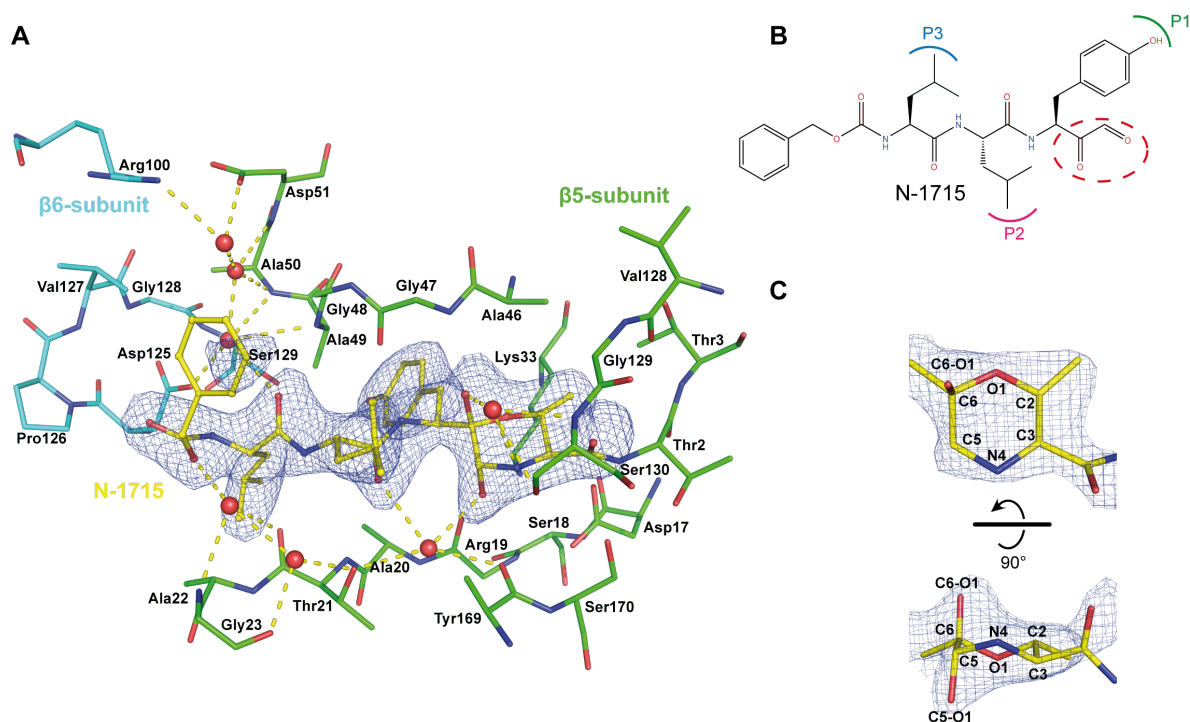


**Fig. 3.15: Vinyl-sulfon inhibition of the chymotryptic-like site.** A: LU-002c, colored in yellow, binds to the chymotryptic-like site, which is formed between the  $\beta$ 6-subunit, colored in blue and the  $\beta$ 5-subunit, colored in green. The chloride ion present in the native 20S proteasome structure is replaced by the sulfon group of the inhibitor. Water molecules are shown as red dots. The hydrogen-bond network of the water molecules is visualized with yellow dashed lines. The electron density map of the inhibitor is shown as an omit map contoured at  $4\sigma$ . B: Vinyl-sulfon inhibitor LU-002c analyzed in 20S proteasome structures is visualized. C: The electron density map for the linkage to Thr1 formed upon vinyl-sulfon inhibition is shown as an omit map contoured at  $6\sigma$ .

backbone (Fig. 3.17). The N-terminal benzyl carbonyl protecting group had a different conformation in the y20S proteasome structure [101]. The electron density for the N-terminal part was better visible in the y20S proteasome structure than in the electron density calculated for our h20S proteasome structure. In the presented structure, an occurring electron density contoured at  $1.5\sigma$  in this region, where the benzyl carbonyl was modeled for the y20S proteasome structure and the negative density of the difference map suggest an additional conformation of the benzyl carbonyl moiety present in the h20S proteasome like that observed in the y20S proteasome structure.

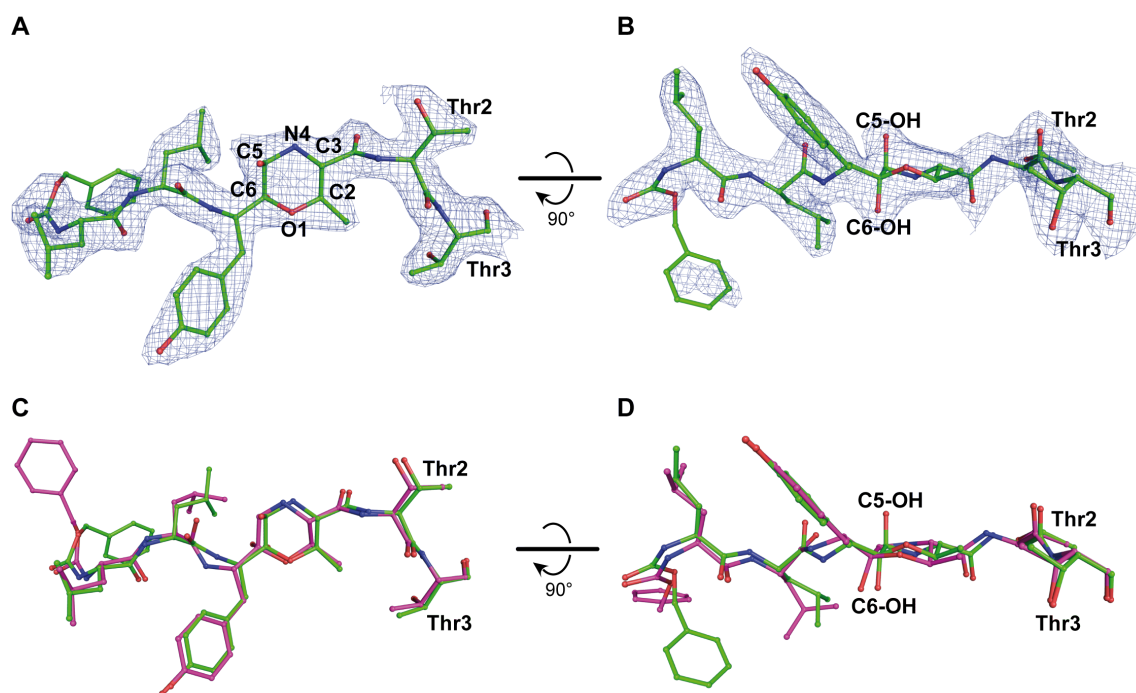
The formation of a 1,4-morpholine linkage between the ketoaldehyde group of the inhibitor and the Thr1 described for the y20S proteasome structure could be confirmed by the solved h20S proteasome structure. The calculated electron densities could be annotated to two new water molecules in close proximity to the covalent linkage. Inhibitor binding resulted in the exchange of the chloride ion, present in the native structure, with a water molecule coordinated by hydrogen bonds formed with the C6-hydroxyl, the N4 of the 1,4-morpholine ring, Ser130/N and Ser130/O $\gamma$ . The second water molecule is stabilized by a hydrogen bond network to Thr21/O $\gamma$ , Arg19/O, C4-OH and the carbonyl of P1 Leu (Fig. 3.16; A).

However, the inhibitor linkage of the ketoaldehyde in the human structure was not iden-



**Fig. 3.16: Ketoaldehyde inhibition of the chymotryptic-like site.** A: N-1715, colored in yellow, binds to the chymotryptic-like site, which is formed between the  $\beta 6$ -subunit, colored in blue and the  $\beta 5$ -subunit, colored in green. The chloride ion present in the native 20S proteasome structure is replaced by a water molecule. Water molecules are shown as red dots. The hydrogen-bond network of the water molecules is visualized with yellow dashed lines. The electron density map of the inhibitor is shown as an omit map contoured at  $4\sigma$ . B: Structure of the ketoaldehyde inhibitor N-1715 is shown. The electrophilic group is encircled with a dashed red line. C: The electron density map for the linkage to Thr1 formed upon ketoaldehyde inhibition is shown as an omit map contoured at  $6\sigma$ .

tical to the linkage described in the y20S proteasome structure (Fig. 3.17). An electron density for a hydroxyl group at the C5 carbon, coordinated by a water molecule and the carbonyl groups of Try169 and Arg19, was clearly visible in the h20S proteasome structure but was not modeled in the yeast structure. Refinement of the structure with the hydroxyl group at the C5 carbon in the 1,4-morpholine linkage, resulted in a B-factor 43 of the C5 hydroxyl oxygen atom and did not reveal any positive or negative densities. According to the higher resolved human proteasome electron density and the refinement statistics, the  $\beta$ -elimination of the C5-hydroxyl group via a condensation reaction resulting in a Schiff base does not occur as previously suggested, but instead a stable hemiaminal is formed (Fig. 3.16; C).



**Fig. 3.17: Comparison of ketoaldehyde binding in the presented structures with the published ketoaldehyde-inhibited 20S proteasome structure.** A,B: Omit map contoured at  $4\sigma$  shows the inhibitor density and the covalent linkage formed between the ketoaldehyde and the Thr1. The electron density for the C5-OH is clearly visible. C+D: Overlay of our modeled ketoaldehyde (green) and the ketoaldehyde modeled in the published yeast structure (pink) (PDB:3OKJ) [101]. The C5-OH is missing in the ketoaldehyde structure modeled for the yeast 20S proteasome inhibition.

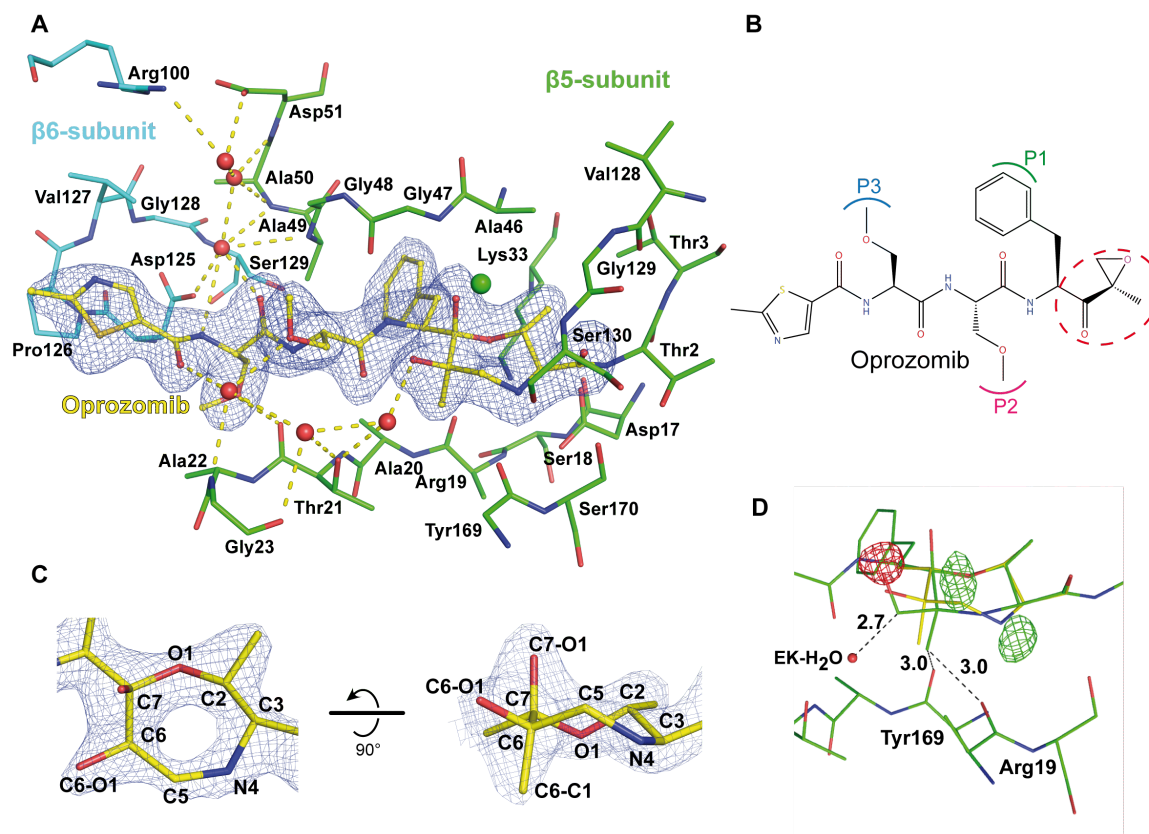
### 3.4.3.5 Epoxyketone inhibitors

The reaction mechanism of the most specific inhibitors, the epoxyketones, was analyzed by soaking several inhibitors of this class [95]. The best diffracting crystal, which reached a resolution of 1.9 Å, was obtained from an Oprozomib soak.

The visible extra density observed in the proteolytic active sites could clearly be annotated to Oprozomib (Fig. 3.18; A). The inhibitor has a B-factor of 34 after refinement. This supported the correct placement of the inhibitor (Tab. 3.5). Binding of Oprozomib caused the displacement of water molecules H<sub>2</sub>O-1 and H<sub>2</sub>O-2. The chloride ion and water molecule H<sub>2</sub>O-3 remained present in the inhibited state.

In published literature, the covalent linkage formed between the electrophilic epoxyketone and the N-terminal threonine was modeled as a 1,4-morpholine ring [99]. Refinement of the structure with the modeled inhibitor resulted in a distorted geometry of the morpholine ring. The carbon bonds to the N4 atom were elongated by 0.1 – 0.2 Å and the C5-alcohol carbon bond was shortened by 0.1 Å. Furthermore, the C5-alcohol methyl-alcohol bond angle diverged by  $-20^\circ$  from the expected value. Moreover, the van der Waals distance between the C5-methyl carbon and the backbone carbonyls of Arg19/ $\beta$ 5





**Fig. 3.18: Epoxyketone inhibition of the chymotryptic-like site.** A: Oprozomib, colored in yellow, binds to the chymotryptic-like site, which is formed between the  $\beta 6$ -subunit, colored in blue, and the  $\beta 5$ -subunit, colored in green. The chloride ion is shown as a green dot. Water molecules are shown as red dots. The hydrogen-bond network of the water molecules is visualized with yellow dashed lines. The electron density map of the inhibitor is shown as an omit map contoured at  $4\sigma$ . B: Structure of the inhibitor Oprozomib is shown. The electrophilic group is encircled with a dashed red line. C: The electron density map for the linkage to Thr1 formed upon Oprozomib inhibition is shown as an omit map contoured at  $6\sigma$ . The formed 1,4-oxazepane ring fits nicely into the electron density. D: The competitive refinement of the 1,4-oxazepane ring (yellow) and 1,4-morpholine (green) ring is illustrated. The refinement with the 1,4-morpholine ring causes van der Waals clashes with the main chain oxygens of Tyr169 and Arg19 and distorted geometry indicated by elongation of the N4-bond by 0.1 – 0.2 Å, shortening of the C5-alcohol bond by 0.1 Å and deviation of the C5-methylalcohol bond angle by  $-20^\circ$  from the expected value. Upcoming negative and positive peaks of the difference map contoured at  $4.5\sigma$  suggest the misplacement of some atoms when using the 1,4-morpholine ring during refinement. The 1,4-oxazepane ring refined without creating positive or negative density peaks and violating geometry requirements.

and Tyr169/ $\beta 5$  was too short, causing repulsive interactions. The negative difference map contoured at  $5\sigma$  suggested that the methyl-alcohol is not positioned correctly (Fig. 3.18; D). The positive difference map contoured at  $4.5\sigma$  indicated missing atoms. All these observations made it evident that the proposed reaction mechanism for the epoxyketone inhibition had to be reconsidered.

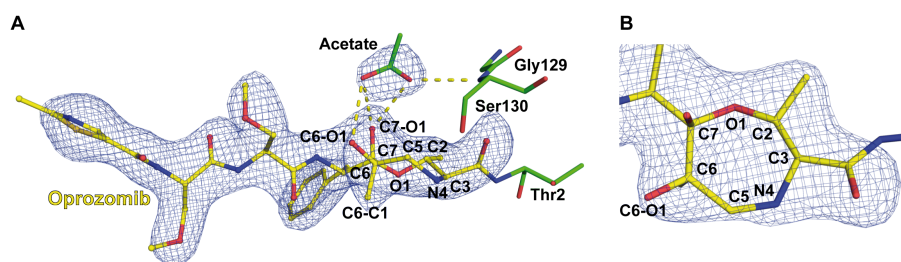
In contrast to the 1,4-morpholine linkage, modeling a 1,4-oxazepane as linkage to the Thr1 resulted in a successful refinement (Fig. 3.18; B). All atoms could be placed into the electron density without the violation of geometric requirements. No positive and neg-

active difference peaks were visible for 1,4-oxazepane linkage above a contouring level of  $2.3\sigma$ . This data suggested that not the 1,4-morpholine ring is formed, but a 1,4-oxazepane ring. To validate this result, several independent experiments were performed, which are described in the next section.

#### 3.4.3.5.1 Validation of the 1,4-oxazepane linkage in an acetate 20S proteasome crystal

The aim was to understand whether a 1,4-oxazepane ring or a 1,4-morpholine ring is formed. First, the potential influence of the chloride ion on the reaction mechanism was ruled out, which was not reported to exist in earlier publications. Slow transfer of the crystals from a chloride buffer to an acetate buffer by repetitive soaking enabled the removal of the chloride ion. The change to acetate buffer slightly decreased the obtainable resolution. A native acetate structure and an Oprozomib inhibited structure were solved at  $2.2 \text{ \AA}$ . The refinement statistics were similar to the chloride structure refinements (Tab. 3.4). No chloride ions could be found in the catalytic active sites of the structures. Instead of the chloride ion in the active sites, an acetate molecule was visible (Fig. 3.19; A).

Soaking of the acetate crystals with Oprozomib resulted in structures containing the inhibitor in all proteolytic active sites. Even the caspase-like site, which was not occupied in the native structure, was fully occupied with Oprozomib. Refinement of the Oprozomib structure revealed the 1,4-oxazepane linkage as the only possible refinement solution as described for the chloride inhibited crystal (Fig. 3.19; B).



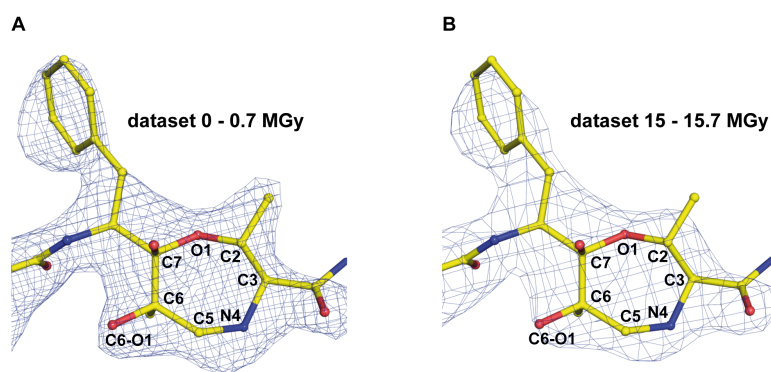
**Fig. 3.19: Oprozomib inhibitor soaked in  $\text{Mg}(\text{OAc})_2$  20S proteasome crystals.** A: Oprozomib, colored in yellow, bound to Thr1 is shown. The chloride ion is exchanged with an acetate molecule. B: Close up of Oprozomib-Thr1 linkage observed in acetate crystal is depicted. 1,4-oxazepane ring is formed upon Oprozomib inhibition in the presence of a acetate molecule. All electron density maps are contoured at  $1.5\sigma$ .

#### 3.4.3.5.2 Analysis of radiochemistry contribution to the reaction mechanism

During the collection of a dataset at the synchrotron, crystals are exposed to high energy photons. Interactions of the high energetic photons with electrons of the atoms create highly reactive free radicals that cause radiation damage in the structure [156]. This



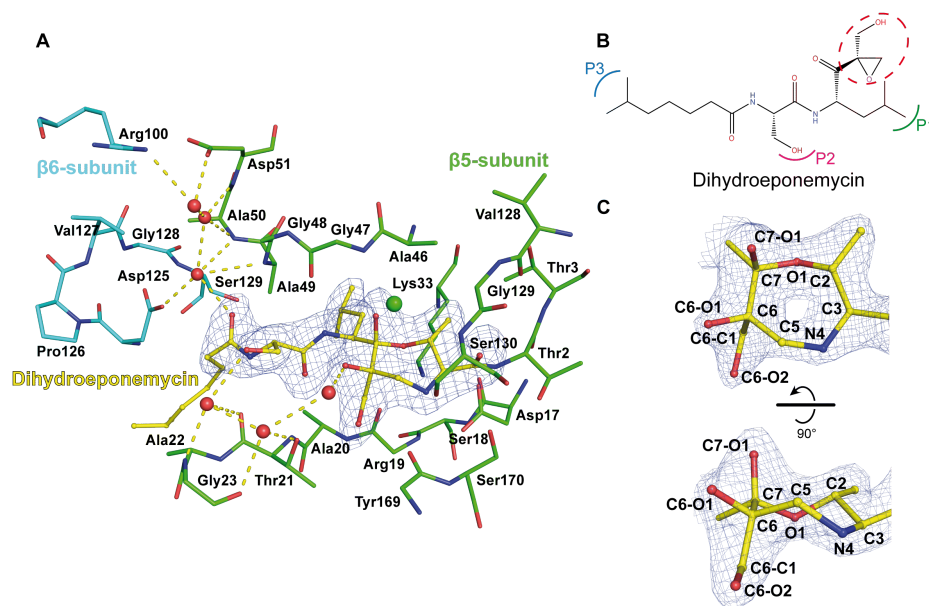
damaging radiation might have an impact on the reaction mechanism. Therefore, effects of radiation damage on the reaction mechanism had to be excluded to validate the 1,4-oxazepane ring formation. First, a structure was solved from a crystal which was exposed to a low radiation dose (0.7 MGy). The same crystal was then exposed to 15 MGy before another dataset was collected with the same settings as the first one. For both dataset structures were solved and the formed epoxyketone linkage was compared. In both structures, the formation of the 1,4-oxazepane ring upon epoxyketone inhibition could be observed. These results provided evidence that any effect of the radiation dose on the apparent reaction mechanism could be excluded (Fig. 3.20; A, B).



**Fig. 3.20: Analysis of dose effect on the epoxyketone reaction mechanism.** A: Oprozomib-Thr1 linkage structure solved for a dataset collected with a total dose of 0.7 MGy. B: Oprozomib-Thr1 linkage structure solved for a dataset exposed to a dose 0.7 MGy after the crystal was pre-exposed with 15 MGy. Both datasets were recorded with the same crystal. In both solved structures the 1,4-oxazepane ring can be observed. All electron density maps are contoured at  $1.5\sigma$ .

**3.4.3.5.3 Analysis of a structurally diverse epoxyketone inhibitor** The inhibitor Dihydroeponemycin, which also belongs to the class of the epoxyketones, was used to structurally validate the 1,4-oxazepane ring formation. The epoxyketone group of Oprozomib and Dihydroeponemycin slightly differ from each other. Dihydroeponemycin has a methanolic group attached to the  $\alpha$ -carbon of the epoxide instead of the methyl group present in Oprozomib. Therefore, the 1,4-oxazepane ring closure would result in the formation of a methanolic group and a hydroxyl group at the C5-carbon. In case a 1,4-morpholine ring would link the inhibitor with the Thr1, two methanolic groups would be present at the C5-carbon.

A pronounced electron density for a hydroxyl group at the C5-carbon could be seen in the solved structure validating the formation of the 1,4-oxazepane ring (Fig.3.21). The absence of any positive or negative difference map peaks and the low B-factor of 40 proved the correct modeling of Dihydroeponemycin (Tab. 3.5).

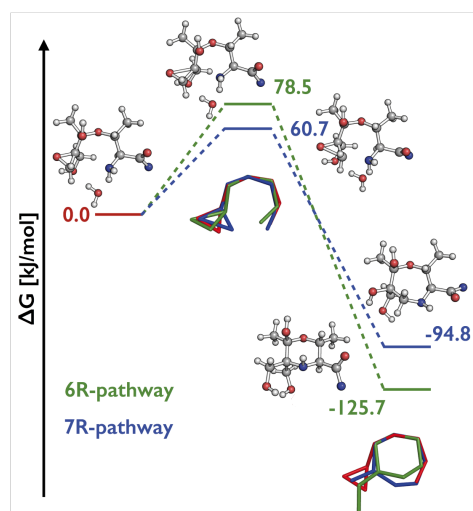


**Fig. 3.21: Inhibition with the epoxyketone Dihydroeponemycin for the chymotryptic-like.**  
 A: Dihydroeponemycin, colored in yellow, binds to the chymotryptic-like site, which is formed between the  $\beta 6$ -subunit, colored in blue, and the  $\beta 5$ -subunit, colored in green. The chloride ion is shown as a green dot. Water molecules are shown as red dots. The hydrogen-bond network of the water molecules is visualized with yellow dashed lines. The electron density map of the inhibitor is shown as an omit map contoured at  $4\sigma$ . B: Structure of the inhibitor Dihydroeponemycin is illustrated. The electrophilic group is encircled with a dashed red line. C: The electron density map for the linkage to Thr1 formed upon Oprozomib inhibition is shown as an omit map contoured at  $6\sigma$ . The formed 1,4-oxazepane ring has a chiral C6-carbon atom.

**3.4.3.5.4 Electronic structure calculations of 6-membered and 7-membered ring formation** Electronic structure calculations were performed to investigate the preferential 1,4-oxazepane formation of epoxyketones. The Dihydroeponemycin structure coordinates after the hemiacetal linkage to the Thr1/OH is formed were used as starting coordinates to calculate the potential reaction pathways resulting either in a 6-membered (6R) or a 7-membered (7R) ring (Fig. 3.22).

The calculated transition states of the 6R and 7R ring pathways revealed that the 6R ring transition state is energetically disfavored by  $18 \frac{\text{kJ}}{\text{mol}}$  compared to the energy of the 7R ring transition state. The overlay of the transition states structures from the 6R and 7R ring pathways demonstrate that the 6R ring transition state must bent more to enable the closing of a 6R ring leading to a higher strain in the peptide backbone. Hence, the observed strain increases the energy barrier for the 6R formation and accordingly the 7R ring is the preferred product state.

The calculated product energy states revealed the higher stability of the 6R ring compared to the 7R ring. The energy difference between both product states amounts to  $30 \frac{\text{kJ}}{\text{mol}}$ .



**Fig. 3.22: Electronic structure calculations for the 6R and 7R ring formation pathways.**

The starting structure and its energy state are marked in red. The 6R ring pathway and its energy states are labeled in green. The 7R ring pathway and its energy states are labeled in blue. The intermediate state for the formation of the 7R ring is lower in energy than the intermediate state of the 6R ring formation pathway. For the product state the energy of the 6R ring is lower compared to the 7R ring. The indicated energy values are relative energies in reference to the starting model.

#### 3.4.3.5.5 Kinetic analysis of 7-membered and 6-membered ring forming inhibitors

So far, the verification of the 7-membered 1,4-oxazepane ring was based on analyses of structural data and quantum chemical calculations, which explain the preferential formation of the 7R ring. As the 6R ring should be thermodynamically favored over the 7R ring, comparative inhibition kinetic measurements of epoxyketones with inhibitors forming a 6R ring should confirm the electronic structure calculations.

The Ketoaldehyde inhibitor was used as a reference for an inhibitor forming a 6R ring. Unlike the epoxyketone, ketoaldehydes only possess an electrophilic group in the  $\alpha$ -position therefore ketoaldehydes can only form 6R rings, which was demonstrated by structural analyses.

Analysis of the inhibition kinetics of the epoxyketones and the ketoaldehyde revealed a less reversible inhibition for the ketoaldehyde than for both measured epoxyketones 3.23. The fluorescence signal for the ketoaldehyde inhibition did not change significantly after 15 min, whereas the epoxyketone inhibitors exhibited a constant turnover of substrate after 10 min of observation. The remaining turnover corresponds to 3% residual 20S proteasome activity. The different inhibition capabilities of the inhibitors matched the calculated relative energy values for the 6R and 7R rings. A likely explanation is that the inhibition of the ketoaldehyde is less reversible, because the formed 6R ring is thermodynamically more stable.



**Fig. 3.23: Kinetic analysis of epoxyketone and ketoaldehyde inhibition mechanisms.** The cleavage reaction of the substrate LLVY-AMC was monitored with a fluorescence spectrophotometer. Human 20S proteasome kinetics were measured without inhibitor (black), inhibited with 50  $\mu\text{M}$  Dihydroeponemycin (grey), inhibited with 50  $\mu\text{M}$  Oprozomib (blue) and inhibited with 15  $\mu\text{M}$  Ketoaldehyde (yellow). The inhibition with the ketoaldehyde is less reversible compared to the epoxyketone inhibitor.

### 3.4.3.6 Diazomethylketone inhibitors

The information obtained from analyses of inhibited 20S proteasome structures about substrate binding is limited, as only one state in which the inhibitor is covalently attached to the Thr1 can be observed. Therefore, intermediate states are missing. Substrate binding is a more dynamic process. An inhibitor which can be analyzed at different states in the proteolytic active site could provide useful information to understand substrate binding. Screening efforts using small molecules potentially interacting with the 20S proteasome unveiled binding of a peptide to the 20S proteasome, which carry a diazomethylketone as an electrophilic group at the C-terminus. The diazomethylketone electrophilic group consists of a ketone linked via a methyl group to a diazo-group. Soaking experiments of the human 20S proteasome with a diazomethylketone peptide validated its binding to the catalytic active pockets. The diazomethylketone group was originally designed to covalently link cysteine proteases. In cysteine proteases, the thiol group of a cysteine forms a covalent linkage to a methyl carbon releasing a nitrogen product [157].

The diazomethylketone electrophile was not described before to interact with 20S proteasomes. Depending on the harvesting temperature of 20S proteasome crystals, two different states of the peptide could be observed. The chymotryptic-like site showed the highest occupancy of the diazomethylketone peptide for both states.

**3.4.3.6.1 Diazomethylketone inhibitor soaked crystals harvested at 4 °C** Structures of h20S proteasome crystals harvested at 4 °C soaked with a diazomethylketone peptide after post-crystallization treatment diffracted up to 2.0 Å. The refined structures showed  $R_{\text{work}}/R_{\text{free}}$  values of 0.18/0.215 (Tab. 3.5). Electron densities for the peptide were observed in the caspase-like site and the chymotryptic-like site. Additional electron density could be detected in the tryptic-like site. Since this site contains an impurity that was already observed in other solved structures, this density could not be attributed to the diazomethylketone peptide clearly.

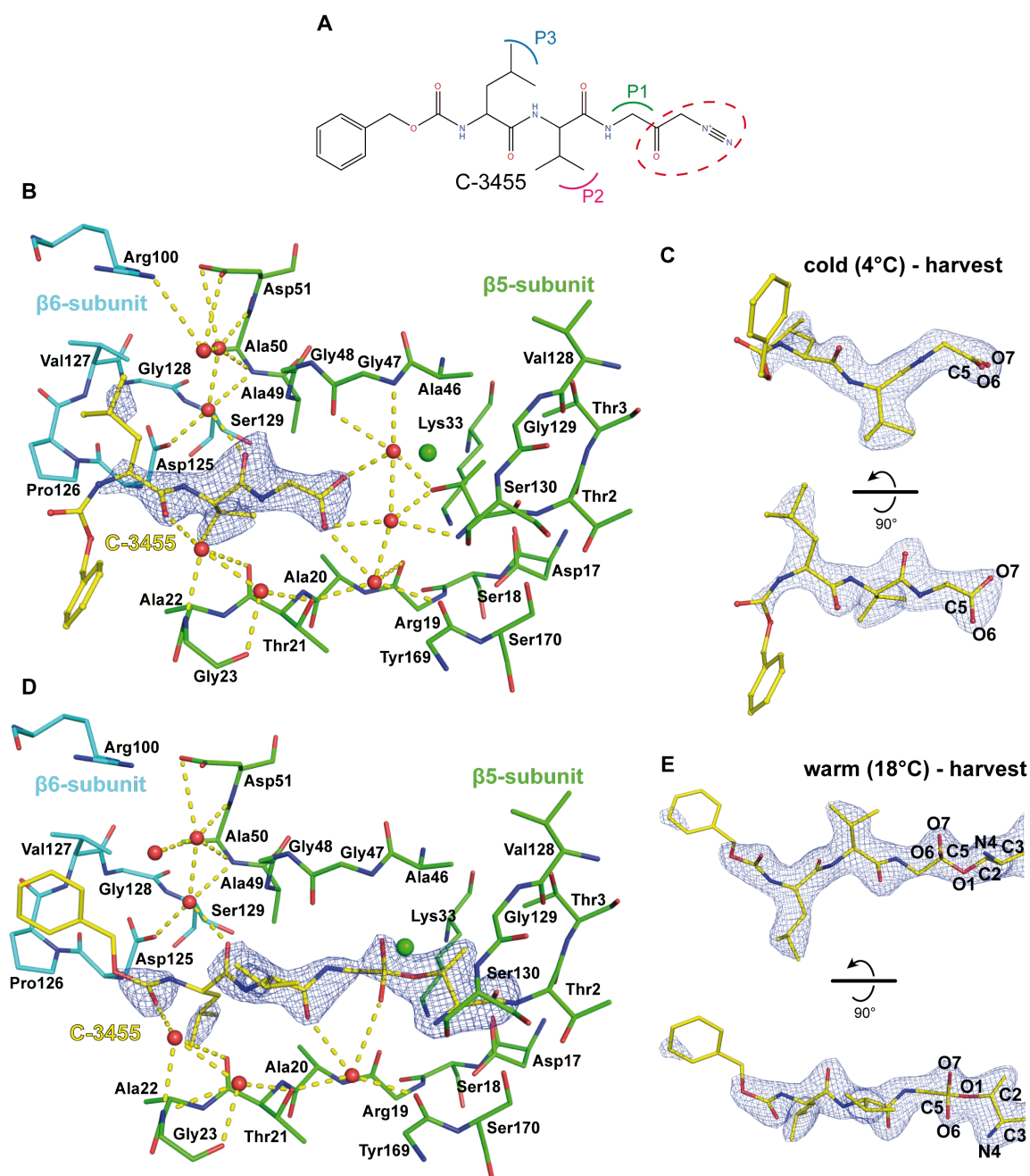
For the caspase-like site and chymotryptic-like site, the peptide could be modeled and refined to B-factors of 62 (Fig.3.24). The position of the peptide differs from the positions observed and published for the peptide backbones of 20S proteasome inhibitors. The diazomethylketone peptide is shifted by one residues out of the proteolytic pocket. Thus, the P1 position of the diazomethylketone peptide could be found at the position at which the P2 residue of all the other inhibitors is located. The  $\beta$ -sheet extension is restricted to hydrogen bond interactions of the P1 peptide bond with the neighboring main chains and to a hydrogen bond formed between the C-terminal carboxy group and the nitrogen of the Thr21 amide bond.

**3.4.3.6.2 Diazomethylketone inhibitor soaked crystals harvested at 18 °C** After soaking the h20S proteasome crystals with the diazomethylketone inhibitor at 4 °C, the crystals were transferred back to 18 °C to analyze the effect of an elevated temperature on peptide binding to the 20S proteasome. Soaked crystals harvested at 18 °C diffracted to similar resolutions obtained for crystals harvested at 4 °C. The analyzed diazomethylketone peptide-soaked crystal diffracted to 2.1 Å. The structure refined to  $R_{\text{work}}/R_{\text{free}}$  values of 18.0/22.3 (Tab. 3.5). Additional electron densities could be observed in every proteolytic active site. However, the electron densities were better defined in the chymotryptic-like and caspase-like site than in the tryptic-like site. The modeled diazomethylketone peptides refined to B-factors of 45 (Fig. 3.24).

The binding conformation of the diazomethylketone peptide changed in the crystal harvested at 18 °C compared to crystals harvested at 4 °C. The diazomethylketone peptide adopted a conformation, which has already been observed for the other 20S proteasome inhibitors [99, 103, 116, 120]. Amide bonds of P1 and P2 residues form hydrogen bond to the main chain to extend the  $\beta$ -sheet. Moreover, a covalent linkage between the diazomethylketone peptide and the Thr1 could be detected. Fitting of the peptide into the electron density revealed that the covalent linkage is different from the described inhibition mechanism detected for cysteine proteases (Fig. 3.24; E) [157]. The observed electron density can only be explained by the formation of an ether bond between the Thr1 to the carbon atom of the ketone of the diazomethylketone resulting in an ester bond.

---

Shifting of the diazomethylketone inhibitor into the proteolytic active site at 18 °C leads to the formation of a covalent linkage to Thr1 and coincides with the flipping of the residues P2 and P3 by 180° compared to their states observed in crystals harvested at 4 °C. The residues conformations of the shifted inhibitor resembled the conformations observed for all investigated 20S proteasome inhibitors.



**Fig. 3.24: Inhibition of the chymotryptic-like by the diazomethylketone inhibitor C-3455.**

A: Structure of the inhibitor C-3455 is shown. The electrophilic group is encircled with a dashed red line A, D: C-3455, colored in yellow, binds to the chymotryptic-like site, which is formed between the  $\beta 6$ -subunit, colored in blue, and the  $\beta 5$ -subunit, colored in green, in two conformations depending on the harvesting temperature. The chloride ion is shown as a green dot. Water molecules are shown as red dots. The hydrogen-bond network of the water molecules is visualized with yellow dashed lines. The electron density map of the inhibitor is shown as an omit map contoured at  $4\sigma$ . C: Position of C-3455 after solving the structure from crystals harvested at 4 °C. The omit electron density map was contoured at  $6\sigma$ . E: The electron density map for the linkage formed between C-3455 to Thr1 when the inhibitor soaked crystals are harvested at 18 °C is shown as an omit map contoured at  $6\sigma$ .

**Table 3.4: Table 1 for all solved human 20S proteasome structures part 1/2**

	<i>Homo sapiens</i>								
	native		Aldehyde			Boronic			
Ligand			MG132	MG115	C-3900	Bortezomib	Ixazomib	Delanzomib	MG262
PDB ID	5LE5	5LEX				5LF3	5LF7	5LF7	
Divalent salt	MgCl <sub>2</sub>	Mg(OAc) <sub>2</sub>	MgCl <sub>2</sub>	MgCl <sub>2</sub>	MgCl <sub>2</sub>	MgCl <sub>2</sub>	MgCl <sub>2</sub>	MgCl <sub>2</sub>	Mg(OAc) <sub>2</sub>
Space group					P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>				
Unit cell									
a (Å)	113.4	113.9	113.9	113.6	113.4	113.4	113.4	114.4	113.8
b (Å)	202.8	203.1	203.2	202.5	203.3	202.7	202.6	202.8	203.2
c (Å)	316.0	316.0	314.9	314.6	315.1	314.9	314.5	314.3	315.3
<b>Data collection</b>									
Wavelength (Å)					0.976				
Resolution range (Å)	107 - 1.80 (1.90 - 1.80)	50 - 2.20 (2.28 - 2.20)	170.71 - 2.20 (2.24 - 2.20)	170 - 1.96 (1.99 - 1.96)	203.03 - 2.12 (2.57 - 2.52)	157 - 2.10 (2.20 - 2.10)	94 - 2.00 (2.10 - 2.00)	50 - 1.99 (2.06 - 1.99)	315.31 - 2.35 (2.35 - 2.31)
No. of reflections	8923661	2519654	3124490	4574418	1709276	2017022	3047303	4210019	2159072
No. of unique reflections	633728	366728	366015	518945	249651	393306	459417	462522	319087
Completeness (%)	99.9 (99.7)	99.2 (98.1)	99.8 (97.1)	99.9 (98.3)	99.8 (96.9)	98.7 (89.1)	99.8 (99.6)	99.1 (90.0)	99.8 (97.5)
<I/σI>	18.5 (0.9)	13.0 (0.9)	12.8 (1.2)	11.8 (0.7)	12.0 (0.6)	12.4 (0.9)	14.6 (1.0)	11.3 (1.0)	12.0 (0.8)
CC <sub>(1/2)</sub>	1.00 (0.32)	1.00 (0.36)	0.998 (0.40)	1.00 (0.26)	1.00 (0.37)	1.00 (0.26)	1.00 (0.29)	0.99 (0.34)	1.00 (0.30)
R <sub>pim</sub>	0.026 (0.71)	0.039 (0.87)	0.037 (0.642)	0.035 (0.997)	0.045 (1.061)	0.052 (0.87)	0.041 (1.1)	0.045 (0.8)	0.040 (1.00)
<b>Refinement</b>									
R <sub>work</sub>	0.184	0.176	0.183	0.184	0.179	0.183	0.182	0.184	0.174
R <sub>free</sub>	0.213	0.220	0.225	0.2201	0.219	0.225	0.215	0.217	0.225
No. of. atoms									
Protein	48205	48134	47947	4777	47975	48331	48272	48272	47979
Ligand/solvent	3712	3493	4289	4286	4161	3888	3809	3794	4201
Average B values (Å <sup>2</sup> )									
Protein	46	67	56	53	54	52	53	52	66
Inhibitor	-	-	53	47	51	42	49	33	75
Metals	36	59	54	46	51	43	48	46	66
Cl ions	51	-	64	62	65	60	63	65	-
Water	48	64	60	57	59	55	58	60	70
Wilson	41	57	59	52	56	46	46	42	72
R.m.s.d. bond length (Å)	0.018	0.015	0.016	0.015	0.022	0.014	0.013	0.012	0.018
R.m.s.d bond angles (Å)	1.92	1.72	1.83	1.75	2.178	1.72	1.68	1.64	1.924
Ramachandran % most favoured/favoured	95.88/3.06	95.60/3.29	96.48/2.39	96.52/2.40	96.30/2.58	95.67/3.12	95.62/3.24	95.88/2.99	96.2/2.50



**Table 3.5: Table 1 for all solved human 20S proteasome structures part 2/2**

<i>Homo sapiens</i>									
	Diazomethylketone		Epoxyketone				Ketoaldehyde	Vinyl-sulfon	
Ligand	C-3455/ warm	C-3455/ cold	Oprozomib	Oprozomib	Epoxomicin	Dihydroeponemeyin	Carfilzomib	Z-LLY-ketoaldehyde	LU-002c
PDB ID			5LEY	5LEZ	5LF0	5LF1		5LF6	
Divalent salt	MgCl <sub>2</sub>	MgCl <sub>2</sub>	MgCl <sub>2</sub>	Mg(OAc) <sub>2</sub>	MgCl <sub>2</sub>	MgCl <sub>2</sub>	MgCl <sub>2</sub>	MgCl <sub>2</sub>	MgCl <sub>2</sub>
Space group					P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>				
Unit cell									
a (Å)	113.7	113.4	113.4	113.9	113.1	113.9	113.4	113.9	113.4
b (Å)	202.7	202.8	202.7	203.2	202.4	203.5	202.7	203.5	202.4
c (Å)	315.1	315.3	315.1	315.2	314.8	316.0	314.9	315.6	313.7
<b>Data collection</b>									
Wavelength (Å)					0.976				
Resolution range (Å)	315 – 2.10 (2.13 – 2.10)	170.58 – 2.11 (2.15 – 2.11)	170 - 1.90 (2.00 – 1.90)	50 - 2.19	107 - 2.41 (2.50 - 2.41)	50 - 2.0 (2.07 - 2.00)	314.86 – 2.11 (2.14 – 2.11)	50 - 2.07 (2.15 - 2.07)	202.4 – 2.11 (2.15 – 2.12)
No. of reflections	2877246	5603238	2653718	2510786	2519995	4381423	2836908	2792074	2791879
No. of unique reflections	422743	412672	526170	351230	262073	464668	416594	403247	410338
Completeness (%)	99.7 (95.3)	99.7 (94.1)	97.5 (99.4)	99.3 (94.3)	99.9 (99.6)	99.8 (98.3)	99.9 (98.2)	96.4 (72.3)	99.9 (99.2)
<I/σI>	8.0 (0.9)	13.7 (1.0)	15.3 (0.8)	15.2 (1.1)	13.1 (0.8)	12.2 (1.0)	12.9 (0.9)	12.9 (0.6)	11.8 (0.9)
CC <sub>(1/2)</sub>	0.995 (0.24)	1.00 (0.36)	1.00 (0.24)	1.00 (0.40)	1.00 (0.28)	1.00 (0.39)	1.00 (0.36)	0.99 (0.22)	1.00 (0.36)
R <sub>pim</sub>	0.077 (1.011)	0.039 (0.857)	0.041 (1.03)	0.036 (0.81)	0.060 (1.06)	0.038 (0.69)	0.041 (0.879)	0.043 (1.0)	0.043 (0.887)
<b>Refinement</b>									
R <sub>work</sub>	0.18	0.18	0.189	0.181	0.168	0.175	0.177	0.185	0.187
R <sub>free</sub>	0.223	0.215	0.224	0.222	0.227	0.209	0.218	0.226	0.233
No. of. atoms									
Protein	48305	48343	48202	48372	48372	48302	47961	48219	48295
Ligand/solvent	4019	3993	3756	3882	3385	3839	4292	3718	4085
Average B values (Å <sup>2</sup> )									
Protein	46	51	49	65	67	51	54	55	56
Inhibitor	45	62	34	47	48	39	51	45	56
Metals	40	50	42	57	54	42	50	45	55
Cl ions	55	60	56	-	79	56	65	68	67
Water	44	58	51	64	62	54	59	59	60
Wilson	47	51	44	56	59	46	55	49	56
R.m.s.d. bond length (Å)	0.014	0.013	0.016	0.015	0.018	0.014	0.014	0.013	0.017
R.m.s.d bond angles (Å)	1.717	1.65	1.80	1.75	1.89	1.71	1.743	1.65	1.858
Ramachandran % most favoured/favoured	96.83/2.14	96.75/ 2.27	95.58/3.30	95.24/3.45	95.91/2.88	95.68/3.16	96.52/2.34	95.49/3.35	96.62/ 2.42

### 3.4.4 Allosteric communication between proteolytic active site and 19S lid

Structural rearrangements of the 20S proteasome upon inhibitor binding are restricted to the flipping Met45 in the chymotryptic-like site. Other obvious structural rearrangements could not be detected for the h20S proteasome. To obtain a more dynamic picture of the whole 26S proteasome, the native 26S proteasome complex and the 26S proteasome complex inhibited with Oprozomib were prepared and analyzed by negative stain electron microscopy and 3D transmission electron cryo-microscopy (cryoEM).

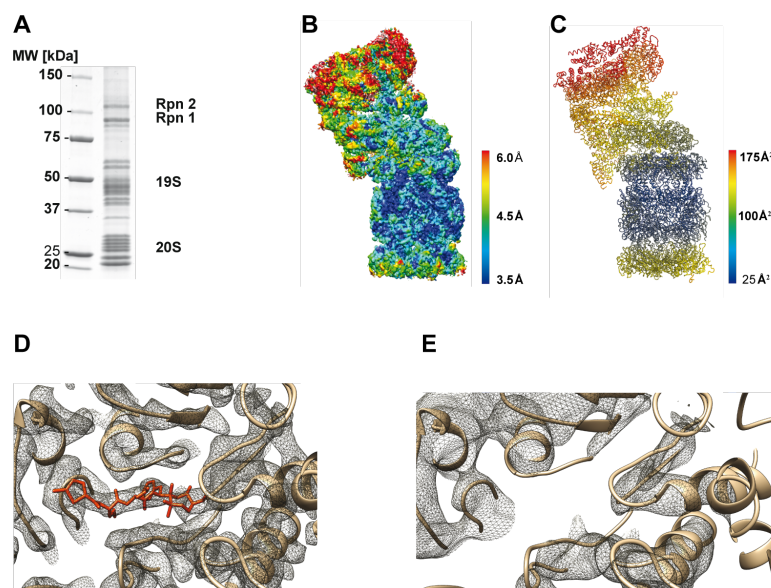
#### 3.4.4.1 Validation of Oprozomib binding to the 26S proteasome

The Oprozomib inhibited 26S proteasome structure was solved at an average resolution of 3.8 Å. The resolution was determined at FSC=0.143. The resolution ranges from 3.5 Å in the 20S proteasome core to 6 Å in the flexible regions of the 19S RP (Fig. 3.25; B). Since the best achieved resolution for the native structure of the 26S proteasome was limited to 4.8 Å, the inhibited 26S proteasome map was used for model building. B-factors of the atoms of the modeled structure vary from 25 to 175 Å<sup>2</sup> (Fig. 3.25; C). Amino acids with residues were modeled up to B-factors of 110 Å<sup>2</sup>. Between 110 Å<sup>2</sup> and 150 Å<sup>2</sup>, the structure was built as a poly-Ala trace. Atoms exceeding a B-factor of 150 Å<sup>2</sup> were discarded. The observed structural features were the same as described by Schweitzer et al. and Huang et [61, 62]. Rpn13 could not be resolved due to its conformational flexibility in agreement with the other published structures.

Comparison of the cryo-EM densities of the chymotryptic proteolytic pocket of the native and the inhibited 26S proteasome at 4.8 Å showed an extended density connected to Thr1 for the inhibited structure verifying binding of Oprozomib (Fig. 3.25; D, E). The inhibitor was also present in the symmetry related chymotryptic-like active site.

#### 3.4.4.2 26S proteasome dynamics upon inhibition analyzed by negative stain EM and cryo-EM

Two major conformational states, a non-rotated-state and a rotated state, could be differentiated for the 26S proteasome (Fig.3.26; A). The rotated state corresponds to a structure, in which the 19S lid is rotated by 25° around the 19S base. Inhibition of the 26S proteasome led to a changed population distribution of the two states compared to the native 26S proteasome (Fig. 3.26; B). The 20S proteasome inhibitors Oprozomib and Epoxomicin reduced the percentage of total particles belonging to the rotated state from

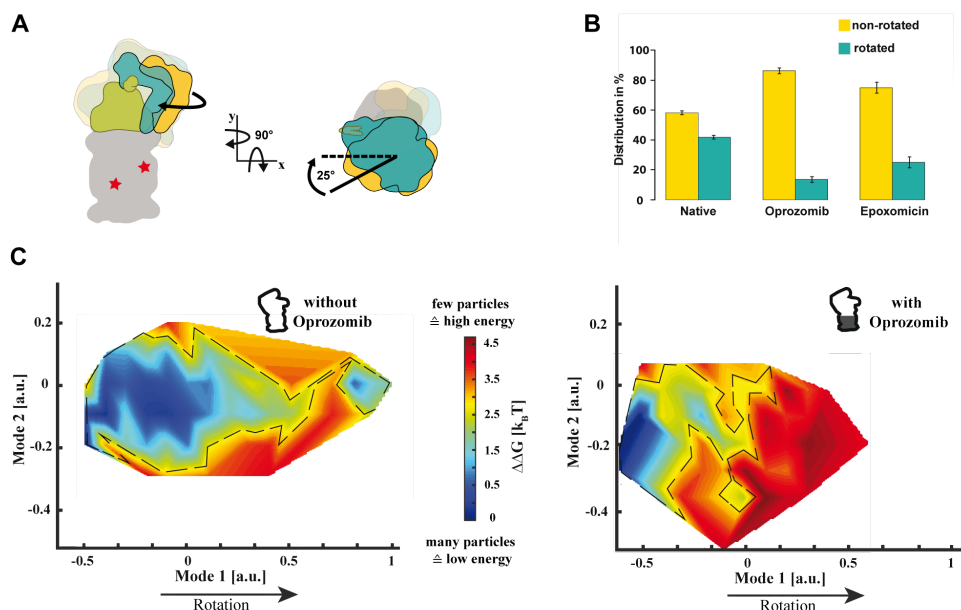


**Fig. 3.25: Details about the 26S proteasome structure.** A: SDS-PAGE analysis of the purified 26S proteasome subunits is shown. B: Local resolution map of the Oprozomib inhibited 26S proteasome calculated with ResMap is depicted. The highest resolution is obtained in the 20S proteasome core. The worst resolved features reside in the 19S lid. C: 26S proteasome model with annotated B-factors is illustrated. The lowest B-factors can be found in the 20S proteasome and the highest B-factors are located in the 19S lid. D, E: Close-up of the chymotryptic-like site of the Oprozomib-inhibited (D) and the native (E) 26S proteasome structure. Density corresponding to the inhibitor, colored in red, is not present in the native 26S proteasome structure (adapted from Haselbach et al.[63]).

41%, observed for the native structure, to 13% or 25%, respectively.

A more sophisticated conformational analysis based on cryo-EM data of the 26S proteasome was performed by PCA classification of 3D classes using the two major modes of the lid movements. The different 3D classes were calculated by several rounds of particle classification. The resulting energy landscapes of the native and inhibited 26S proteasome describe the distribution of 26S conformations and the corresponding number of particles in a two-dimensional space spanned by the two major modes (Fig. 3.26; C). The number of particles belonging to one 3D class that represent one distinct conformation is indicated by a color code and can be translated into an energy level using the Boltzmann interpolation. The red color indicates high energy levels, because these states are unlikely to be adapted signified by the low number of particles corresponding to these states. In contrast, the blue color designates highly populated states containing a large number of particles.

Inhibition reduced the conformational freedom of the 26S proteasome compared to the non-inhibited state illustrated by less possible conformations and made some conformations, observed for the native 26S proteasome, very unlikely, indicated by the low popu-



**Fig. 3.26: 26S proteasome dynamics analyses by electron microscopy.** A: Non-rotated (yellow) and rotated (green) conformations of the 19S lid are sketched. B: Distributions of particles belonging to either the non-rotated state or rotated state of the 19S lid in the native state, Oprozomib-inhibited state and Epoxomicin-inhibited state are plotted. Upon inhibition of the 26S proteasome the non-rotated is the preferred conformation. The data are based on negative stain electron microscopy. C: Energy landscapes of the native (left) and Oprozomib-inhibited (right) 26S proteasome. The conformational landscape is constrained upon inhibition with Oprozomib. Particles used for the calculation of the energy landscapes were recorded with electron cryo-microscopy (adapted from Haselbach et al.[63]).

lation number of these states indicated by the red color (Fig. 3.26, C).

The analysis of the changed 26S proteasome dynamics suggested an allosteric regulation induced by binding of an inhibitor to a proteolytic-active site mediated through the  $\alpha$ -rings to the 19S lid. The 20S proteasome structure could not provide any information how the  $\beta$ -rings can communicate with the 19S lid.

### 3.4.5 Structural analysis of the fruit fly 20S proteasome

An approach to understand the function of the 20S proteasome is the comparative analysis of 20S proteasomes from different species. Structural investigations of conserved and mutated amino acids were thought to provide additional information about the mode of function and knowledge how substrate specificity is ensured. So far, eukaryotic structures of yeast, bovine, mouse and human 20S proteasomes were solved below 3 Å [31, 32, 49, 120]. The 20S proteasome from *Drosophila melanogaster* was chosen as another model organism for structural investigations of the 20S proteasome. The d20S proteasome is abundant in embryonic extracts and can be purified with the established purification protocol.

### 3.4.5.1 Sequence comparison of the human and the fruit fly 20S proteasome

Primary sequence identities of the 20S proteasome subunits from *Drosophila melanogaster* and *Homo sapiens* describing the conservation of a specific amino acid in the sequence vary between 57% to 78% for the  $\alpha$ -subunits and between 38% to 64% for the  $\beta$ -subunits. The similarity values, which indicate the conservation of physicochemically similar amino acids at each position, ranges between 74% - 89% for the  $\alpha$ -rings and 59% - 83% for the  $\beta$ -subunits. Alignments excluding the pro-peptide regions improve the similarity values for the respective  $\beta$ -subunits by 3 – 10%. Even though, the primary sequence of the 20S proteasome subunits is not highly conserved, the secondary structural elements do not differ between the human 20S proteasome and the fruit fly 20S proteasome.

**Table 3.6: Emboss needle sequence alignment of human and fruit fly 20S proteasome subunits**

Nomenclature		Emboss needle sequence alignment			Sequence alignment without pro-peptides		
subunit	identities	positives	gaps	identities	positives	gaps	
$\alpha$ 1	164/246 (66.7%)	200/246 (81.3%)	2/246 ( 0.8%)				
$\alpha$ 2	182/234 (77.8%)	209/234 (89.3%)	0/234 ( 0.0%)				
$\alpha$ 3	186/264 (70.5%)	219/264 (83.0%)	3/264 ( 1.1%)				
$\alpha$ 4	172/251 (68.5%)	205/251 (81.7%)	5/251 ( 2.0%)				
$\alpha$ 5	172/244 (70.5%)	204/244 (83.6%)	3/244 ( 1.2%)				
$\alpha$ 6	164/246 (66.7%)	200/246 (81.3%)	2/246 ( 0.8%)				
$\alpha$ 7	145/255 (56.9%)	187/255 (73.3%)	2/255 ( 0.8%)				
$\beta$ 1	123/245 (50.2%)	161/245 (65.7%)	27/245 (11.0%)	117/209 (56.0%)	152/209 (72.7%)	4/209 ( 1.9%)	
$\beta$ 2	172/288 (59.7%)	206/288 (71.5%)	27/288 ( 9.4%)	153/245 (62.4%)	183/245 (74.7%)	23/245 ( 9.4%)	
$\beta$ 3	131/205 (63.9%)	170/205 (82.9%)	0/205 ( 0.0%)				
$\beta$ 4	100/201 (49.8%)	146/201 (72.6%)	0/201 ( 0.0%)				
$\beta$ 5	146/291 (50.2%)	203/291 (69.8%)	37/291 (12.7%)	128/209 (61.2%)	171/209 (81.8%)	5/209 ( 2.4%)	
$\beta$ 6	116/246 (47.2%)	172/246 (69.9%)	16/246 ( 6.5%)				
$\beta$ 7	103/270 (38.1%)	159/270 (58.9%)	8/270 ( 3.0%)				
$\beta$ 1i	109/229 (47.6%)	141/229 (61.6%)	15/229 ( 6.6%)	104/209 (49.8%)	134/209 (64.1%)	10/209 ( 4.8%)	
$\beta$ 2i	137/277 (49.5%)	184/277 (66.4%)	9/277 ( 3.2%)	117/236 (49.6%)	159/236 (67.4%)	5/236 ( 2.1%)	
$\beta$ 5i	139/286 (48.6%)	197/286 (68.9%)	14/286 ( 4.9%)	121/209 (57.9%)	165/209 (78.9%)	5/209 ( 2.4%)	

Identity: is the percentage of identical matches between the two sequences over the reported aligned region (including any gaps in the length).

Positives: is the percentage of matches between the two sequences over the reported aligned region (including any gaps in the length).

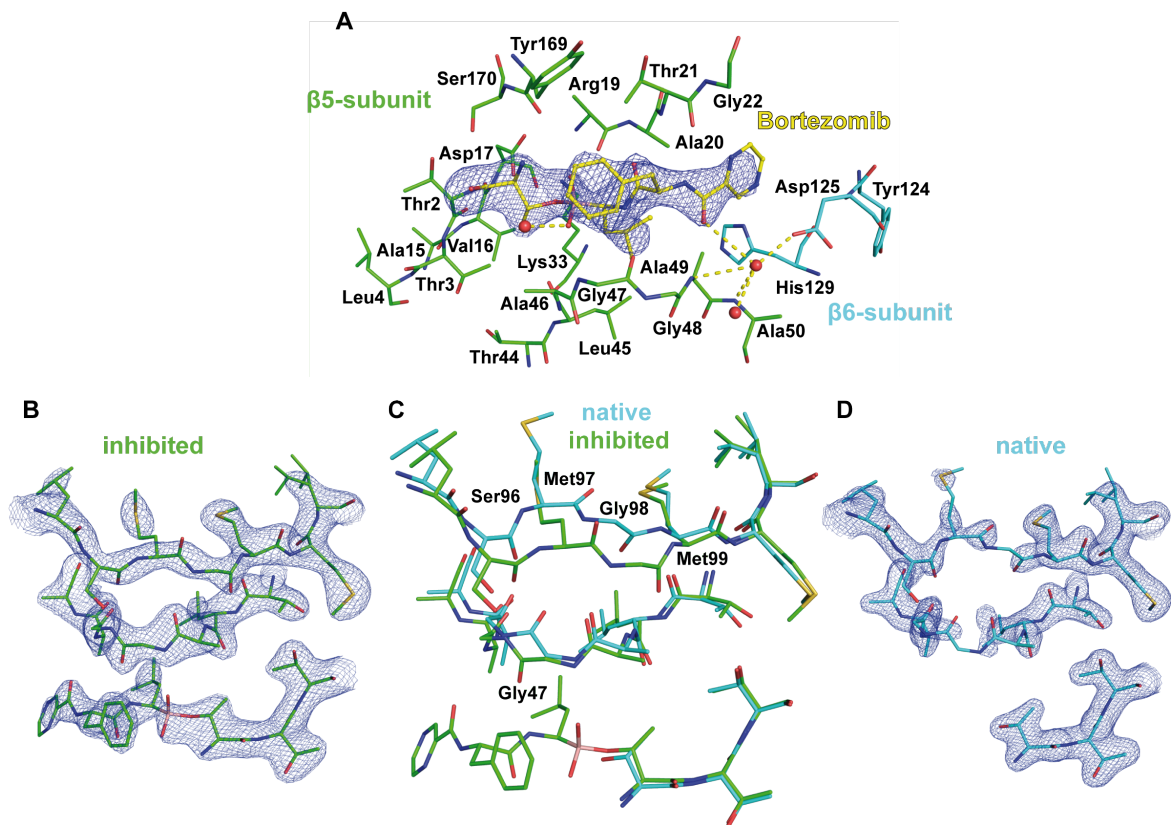
### 3.4.5.2 Structure of the native fruit fly 20S proteasome

Diffraction spots were collected up to 1.9 Å. The  $I/\sigma I$  for the highest resolution shell was 1.0 and the  $CC_{1/2}$  had a value of 0.32. In total 1131891 unique reflections were measured. The structure was solved by molecular replacement using the mutated human structure (PDB: 5LE5). The structure had two full molecules in the asymmetric resulting in 56 peptide chains which were refined. The model was refined to  $R_{\text{work}}/R_{\text{free}}$  values of 0.195/0.22. An overall B-factor of 42 was obtained, which was smaller compared to the native human

20S proteasome. The B-factors ranged from 26 in the core structures of the subunits to 90 in solvent accessible areas. The structure had colder overall B-factors than the human 20S structure. In total 3255 water molecules, 30 potassium ions and 17 magnesium ions could be modeled into structure (Tab. 3.7).

The overlay of the d20S proteasome structure with the h20S proteasome structure showed conservation of all secondary structure features. A few segments in the  $\alpha$ -rings, which are lacking electron density in the human 20S proteasome, could be confidently modeled for the fruit fly 20S proteasome structure. Most of the ion positions are conserved between the human and the fruit fly 20S proteasome.

The major structural difference between the h20S and the d20S proteasome is the partially disordered  $\beta$ -sheet  $\beta 8$  ( $\beta 5/96/\text{Ser} - \beta 5/99/\text{Met}$ ) of the  $\beta 5$  subunit. Consequently, the  $\beta$ -sheet is shorter than the corresponding  $\beta$ -sheet in the human structure (Fig. 3.27; D).



**Fig. 3.27: Re-ordering of  $\beta$ -subunit from *Drosophila melanogaster* upon inhibitor binding.**

A: Bortezomib, colored in yellow, binds to the chymotryptic-like site, which is formed between the  $\beta 6$ -subunit, colored in blue, and the  $\beta 5$ -subunit, colored in green. Water molecules are shown as red dots. The hydrogen-bond network of the water molecules is visualized with yellow dashed lines. The electron density map of the inhibitor is shown as an omit map contoured at  $4\sigma$ . B: Electron density map contoured at  $1.5\sigma$  for the inhibited  $\beta 5$  subunit is shown. The fitted model is colored in green. C: Native (blue) and Bortezomib-inhibited (green)  $\beta 5$ -subunit structures are overlaid. The overlay signifies the structural rearrangement of  $\beta$ -sheet 8 ( $\beta 8$ ) upon inhibitor binding. D: Electron density map contoured at  $1.5\sigma$  for the native  $\beta 5$  subunit is shown. The fitted model is colored in blue.

Table 3.7: Table 1 for all solved fruit fly 20S proteasome structures

<i>Drosophila melanogaster</i>			
Ligand	native	Bortezomib	Activator bound
PDB ID			
Divalent salt	MgCl <sub>2</sub>	MgCl <sub>2</sub>	MgCl <sub>2</sub>
Space group		P 1	
Unit cell			
a (Å) / $\alpha$ (°)	116.5 / 69.21	116.7 / 69.34	116.5 / 69.36
b (Å) / $\beta$ (°)	171.4 / 88.29	171.64 / 88.27	171.55 / 88.27
c (Å) / $\gamma$ (°)	212.5 / 70.35	212.02 / 70.35	212.44 / 70.31
<b>Data collection</b>			
Wavelength (Å)		0.976	
Resolution range (Å)	150.34 – 1.90 (1.93 – 1.90)	150.51 – 2.20 (2.24 – 2.20)	150.56 – 1.90 (1.93 – 1.90)
No. of reflections	8061762	5259538	5996990
No. of unique reflections	1131891	726228	1128430
Completeness (%)	99.4 (99.4)	99.0 (98.0)	99.1 (97.9)
$\langle I/\sigma I \rangle$	8.9 (1.0)	7.8 (0.7)	6.0 (0.6)
CC <sub>1/2</sub>	1.00 (0.32)	0.997 (0.254)	0.996 (0.266)
R <sub>pim</sub>	0.046 (1.33)	0.067 (1.575)	0.077 (2.202)
<b>Refinement</b>			
R <sub>work</sub>	0.190	0.192	0.1902
R <sub>free</sub>	0.220	0.2324	0.226
No. of atoms			
Protein	96629	96475	97118
Ligand/solvent	3529	3905	3489
Average B values (Å <sup>2</sup> )			
Protein	42	52	45
Inhibitor/ Activator	-	52	65
Metals	50	64	37
Cl ions	-	-	-
Water	39	49	42
Wilson	45	59	44
R.m.s.d. bond length (Å)	0.016	0.0191	0.0155
R.m.s.d bond angles (Å)	1.75	2.04	1.73
Ramachandran % most favoured/favoured	97.08/ 2.33	96.84/2.46	97.01/2.35

### 3.4.5.3 Dynamics of the $\beta 8$ sheet of the fruit fly proteasome $\beta 5$ -subunit

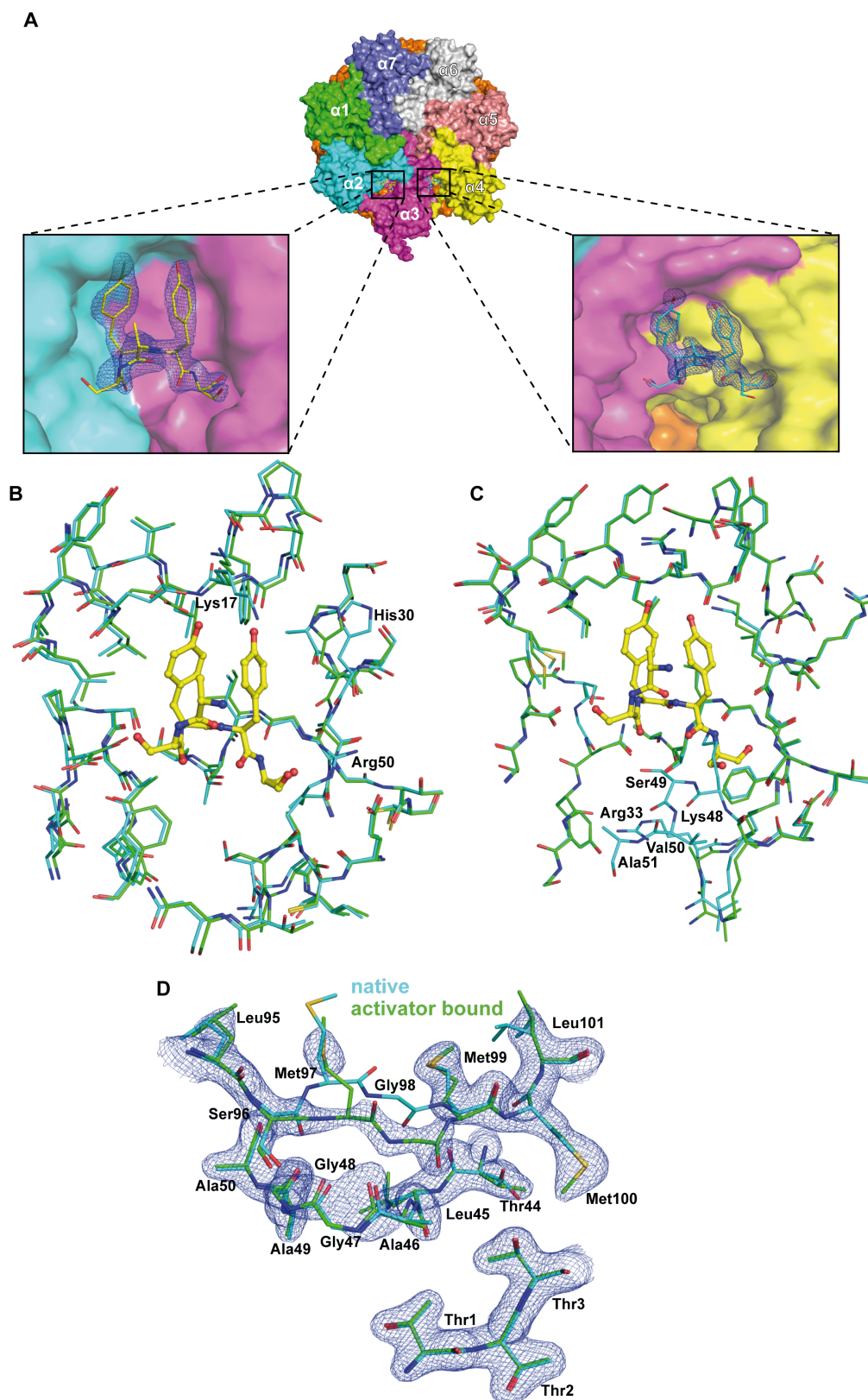
Inhibition of the *Drosophila melanogaster* 20S proteasome with an inhibitor binding to the chymotryptic-like site resulted in a reordering of the disordered  $\beta$ -sheet (Fig. 3.27). As observed for the human 20S proteasome structure, binding of the inhibitor extends the  $\beta$ -sheet,  $\beta 8$ . This enables the formation of hydrogen bonds to the main chain leading to the reorganization and stabilization of  $\beta 8$ . After stabilization, the secondary structure of  $\beta 8$  is identical with the corresponding  $\beta$ -sheet in the human 20S proteasome. All peptide-like inhibitors form the extended  $\beta$ -sheet in the proteolytic active pocket making the reorganization not inhibitor specific. The only prerequisite for the inhibitor to cause

stabilization of the  $\beta$ -sheet is the binding to the chymotryptic-like site.

The state of the disordered region can not only be regulated by inhibitor binding, but also by soaking a short peptide into the  $\alpha$ -rings. The short peptide consisted of 13 amino acid residues and contained the YY binding motif from the C-terminus of BLM10/PA200. The binding of a similar peptide to the  $\alpha$ -rings of the yeast 20S proteasome was published by Witowska et al.[158]. The peptide bound into the interface between  $\alpha$ 2 and  $\alpha$ 3 and between  $\alpha$ 3 and  $\alpha$ 4 (Fig. 3.28). The last six C-terminal residues could be modeled into the electron density for both binding sites. The overall B-factor of the modeled peptide was 65. Upon binding of the peptide, residues  $\alpha$ 4/46/Glu –  $\alpha$ 4/51/Ala and  $\alpha$ 3/30/His underwent a conformational rearrangement and the amino acid stretch  $\alpha$ 3/61/Pro -  $\alpha$ 3/64/Lys got stabilized. The C-terminal serine of the peptide is stabilized by a polar contact with  $\alpha$ 3/64/Lys in one site and with  $\alpha$ 3/61/Lys in the other site (Fig. 3.28). The proposed induced conformational rearrangement caused by an electrostatic repulsion of the tyrosine hydroxyl and the carbonyl of Gly15 shown for the yeast 20S proteasome-BLM10 structure could not be seen in the solved structures [92].

Binding of the peptide to the  $\alpha$ -rings induced a partial stabilization of the disordered region (Fig. 3.28; D). An additional electron density in the region  $\beta$ 5/96/Ser -  $\beta$ 5/99/Met became visible, which was equivalent to the electron density after soaking the crystals with an inhibitor.





**Fig. 3.28: Activator peptide binding induces structural rearrangements of the  $\beta 5$  subunit.**

A: Overview of activator binding is illustrated. Binding sites are additionally shown in a more detailed manner. Binding sites of the activator peptide are located between  $\alpha 2$  and  $\alpha 3$ , and between  $\alpha 3$  and  $\alpha 4$ . B, C: Overlays of the native binding pockets (blue) and the pockets (green) occupied with the activator peptide (yellow) are depicted. D: The  $\beta 5$  subunits from the native (blue) and the activator bound state (green) are overlaid. Binding of the activator results in a rearrangement of residues Met97, Gly98 and Met99. The electron density contoured at  $1.5\sigma$  is shown for the activator bound state.

## 4 Discussion

The underlying work of this thesis improved the biochemistry of the 20S proteasome purification and introduced a post-crystallization treatment for 20S proteasome crystals leading to high resolution structures. This allowed reevaluation of the inhibition mechanisms of the 20S proteasome proteolytic active sites in greater detail and to investigate novel mechanistic questions concerning the function of the 20S proteasome. Key points of the biochemical optimization were the stabilization of the 20S proteasome by screening different buffer compositions with the ProteoPlex technique and the chromatography-free purification. The superiority of the established purification approach was demonstrated by the increased specific activity compared to published purification strategies utilizing chromatography and the high yields of purified 20S proteasomes [150]. The universal application of the purification strategy was shown by purifying the 20S proteasome from not just one organism, but from four diverse species utilizing similar protocols. The high quality of the purified 20S proteasomes was evident as every 20S proteasome sample could be crystallized, demonstrating the homogeneity and the structural integrity of the obtained 20S proteasome complexes. X-ray diffraction for the obtained 20S proteasome crystals was limited to 3 Å.

A newly developed post-crystallization treatment procedure for the gradual dehydration of the 20S proteasome crystals led to crystals routinely diffracting to 2 Å. Essential for the post-crystallization procedure was the transfer of the crystals to 4 °C. At 4 °C the equilibration of the crystallization drop against the reservoir that contains an increased cryo-protectant concentration is milder due to slower vapor diffusion rates. Moreover, the higher humidity in the cold room prevents uncontrollable changes in relative humidity of the crystallization drop after opening the well. The improved resolution of 2 Å allowed modeling of solvent molecules and ions to higher precision, which was not described in earlier published structures. The methods presented here allowed alternative conformations of sidechains to be confidently modeled into the structure (Fig. 3.9; C). The higher resolution structures made it possible to reassess the inhibition mechanisms for several inhibitor classes of the 20S proteasome (Tab. 3.4; 3.5). Published reaction mechanisms for boronic inhibitors and vinyl-sulfon inhibitors could be validated by the newly solved structures. For the aldehyde inhibitors a slight change in the positioning of the hemiacetal hydroxyl became evident and for the  $\alpha$ -Ketoaldehyde the formation of the Schiff base proposed in earlier studies could not be supported by the solved structures [101]. The structures presented here further show a large discrepancy in the proposed reaction

mechanism for the epoxyketone inhibitors. According to the high resolution electron densities in our structures, it was determined that instead of a 1,4-morpholine ring structure, a 1,4-oxazepane ring structure is formed as a covalent linkage between the epoxyketone group and Thr1 of the proteolytic-active sites upon inhibition. Different approaches, including structural analyses, kinetic analyses and quantum chemical calculations of the reaction pathways, were used to validate the modeled 1,4-oxazepane ring. All validation experiments supported the newly proposed reaction pathway for epoxyketone inhibition.

The effect of proteasome inhibition was not only analyzed locally for the respective proteolytic-active sites, but also globally for the 26S proteasome holoenzyme. Comparison between the native and inhibited 26S proteasome structures revealed structural rearrangements in the 19S lid when the 20S proteasome is inhibited. Sorting of native and inhibited 26S proteasome conformations using the major modes of 19S lid dynamics revealed a constricted variation of observable conformations for the inhibited 26S proteasome structure. These results suggest a long-range allosteric communication, which spans over a distance of 190 Å, between the proteolytic active-sites and the 19S lid. This allosteric modulation is not uni-directional, but bi-directional. This was demonstrated by structures of the *Drosophila melanogaster* 20S proteasome, in which a short peptide chain was located in the  $\alpha$ -ring pockets. Binding of this peptide induces a structural rearrangement of the  $\beta$ -sheet in the  $\beta$ 5-subunit, where the chymotryptic-like proteolytic site is located. The solved structures provided first insights into the functional regulation through allosteric interactions between the  $\beta$ -rings of the 20S proteasome core and the 19S lid. So far, the resolution of the structures is not sufficiently high to enable a comprehensive analysis on how the allostery is mediated throughout the 20S proteasome. However, the optimized workflow described in this thesis can be used as a basis to achieve even higher resolution proteasome structures in the future.

## 4.1 Benefits of a chromatography free purification

A prerequisite to study the biochemical or structural properties of a protein complex is to obtain a pure and functional active sample. A variety of approaches exist to purify proteins. Initially, proteins were purified using a fractionation approach. This technique is based on using the solubility properties of proteins determined by the charge distribution on the surface of proteins. Manipulation of solvent compounds and concentrations were used for the selective precipitation of the desired protein. The next development in protein purification methods was the introduction of chromatography in the 1960s. Chromatography is based on the spatial or temporal immobilization of the desired protein to

the column material using the surface charge, or size and shape of the protein of interest. The latest development in protein purification was made possible due to advances in gene sequencing and editing technologies. Sequence editing made it possible to generate proteins fused to an affinity tag, which is translated with the protein of interest [159]. The tag is used for immobilization of the protein to a column material that has affinity to the tag. The tag strategy protocol can be easily transferred to diverse proteins, making it a popular method for protein purification. Tagging proteins in a mammalian system is technically challenging and expensive. Therefore, genetic modified mammalian genes like tagged mammalian protein complexes are typically expressed in insect cells. As insect cells provide a suitable environment for the successful protein complex assembly, several mammalian proteins can be expressed simultaneously allowing the reconstitution of large protein complexes [160].

The established purification protocol developed for the purification of endogenously expressed 20S proteasomes solely uses the fractionation procedure with PEG, representing the most antique method to purify proteins. The advantage of this method is the preservation of protein complex structure during precipitation [161]. The precipitation is based on manipulation of the hydration shell enclosing the protein, which is reversible and does not lead to irreversible denaturation. Currently, the exact physical mechanisms of PEG precipitation are not yet fully understood, making it necessary to determine the precipitation behavior of protein complexes empirically [162]. Tagging of a subunit of the 20S proteasome for purification purposes is not needed because the existing chemical properties of the 20S proteasome are utilized to separate it from other protein complexes. As endogenously expressed 20S proteasomes are used for the purification, reconstitution of the protein complex is not necessary.

The fractionation method is beneficial in particular for large protein complexes with loosely bound components. High salt concentrations are applied during elution of labile complexes immobilized to a column material used in chromatography. The high ionic strength could possibly cause the partial disassembly of the protein complex. Furthermore, shear forces between the protein complexes and the column material might occur throughout the purification procedure harming the structural integrity of the protein complex. If a complex is immobilized via an affinity-tagged subunit, this subunit is exposed to strong forces because it carries the weight of the whole complex. The externally applied forces are likely to lead to the partial disintegration of the protein complex or small structural changes hampering its function and structure. Most of the commonly used protein purification protocols involve size exclusion chromatography. During size exclusion chromatography, the protein complex is separated by its molecular size allowing to enter pores of different diameters formed by the column material. The retention time of a protein complex to the column material correlates with its size and the pore diameter. The different retention times make it possible to separate protein complexes of various

sizes. If the pore diameter just barely accommodates the size of the protein complex, shear forces between the protein complex and the column material might be expected to impair the integrity of the protein complex.

Comparison of specific activities measured for the human 20S proteasome either purified with chromatography or with a chromatography-free protocol demonstrated a ten-fold higher activity for the 20S proteasome purified without chromatography [150]. The higher specific activity of proteasome complexes purified by different methods confirms the hypothesis that the integrity of the 20S proteasome suffers under the use of chromatography methods during purification. In addition, the crystallizability of the 20S proteasomes from different organisms purified by a chromatography-free method shows the high quality of the obtained samples.

Crystallization of the 20S proteasome is possible using chromatography purification methods, as demonstrated by the several 20S proteasome structures already deposited in the Protein Data Bank (PDB; [www.rcsb.org](http://www.rcsb.org), 24.08.2018). Interestingly, protein concentrations of 20 mg/ml to 30 mg/ml are used for the crystallization of these proteasomes. The chromatography-free purified 20S proteasome already crystallized at protein concentration of 7.5 mg/ml, suggesting a higher homogeneity of the particles due to the avoidance of chromatography as a purification strategy.

Compared to other multi component protein complexes, the 20S proteasome is a highly stable protein complex. However, improvements of the sample quality, signified by the higher specific activity and good crystallizability, by avoiding chromatography are notable. More labile complexes like the 26S proteasome may benefit greatly by using a fractionation technique for purification, such as the one developed in this thesis work.

## 4.2 Crystal packing

Crystal packing heavily influences the diffraction properties of a crystal. Misaligned protein complexes in the crystal lattice cause the loss of diffraction. In a tight crystal packing protein complexes are better ordered due to more intermolecular interactions. The Matthews coefficient is used as a dimension to describe how dense proteins are packed in a crystal as it estimates the solvent content of a unit cell [163]. The tighter molecules are packed in a crystal the less solvent can be found between the molecules.

The crystal packing of the yeast proteasome is the least dense compared to the crystal packing of the human, fruit fly and archaeal 20S proteasomes, indicated by the increased Matthews coefficient (Tab. 3.1). Analysis of the yeast 20S proteasome crystal packing shows a high-water content along the z-axis. The loose crystal packing in the z-axis is most probably the reason for the observed severe anisotropic diffraction of the yeast 20S

proteasome (Fig. 4.1). 20S proteasome crystals have a relatively large unit cell compared to the average unit cell dimension of proteins analyzed by X-ray crystallography. Consequently, the 20S proteasome crystals contain significantly fewer unit cells resulting in weaker diffraction, as the integrated intensity of a diffraction spot correlates with the number of unit cells present in a crystal [164]. Loose packing of a crystal reduces the diffraction spot intensity because solvent atoms do not contribute to diffraction, as disordered solvent molecules do not contribute to the intensity of the diffraction spots.

In comparison to the  $\gamma$ 20S proteasome, the crystal packings of human, fruit fly and archaean 20S proteasomes are more densely packed in each direction. Diffraction of these three 20S proteasome crystals was more isotropically than the diffraction observed for  $\gamma$ 20S proteasome crystals. This supports the idea that the diffraction power of the crystals is dependent on the number of unit cells in the crystal. The strong dependency between the number of unit cells and the diffraction power of a crystal could be also shown by the comparison of the diffraction spots obtained from crystals of different sizes. To be able to record high resolution diffraction data, the crystal needed a minimum size of  $150 \times 150 \times 200 \mu\text{m}^3$ .

### 4.3 Benefit of the post-crystallization procedure

Crystals which were cryo protected by short soaks in crystallization condition supplemented with 30 % MPD diffracted up to 2.5 Å. The lack of higher-resolution reflections is likely to be caused by imperfections in the crystal lattice. Crystal lattices which are composed of large unit cells are more sensitive to imperfections and the resulting loss of high resolution diffraction because a misalignment of the long unit cell constant causes a larger lateral shift of the symmetry related atom compared to a lateral shift observed for a short unit cell constant. The imperfections in the crystal lattice could be corrected with a post-crystallization treatment. Crystals measured after applying the post-crystallization treatment diffract to 2.0 Å on average.

During the post-crystallization treatment, a mild dehydration of the crystal takes place by increasing the concentration of the precipitant and by the addition of the cryo-protectant. Dehydration leads to a rearrangement of the unit cells resulting in a repair of the crystal lattice imperfections which leads to better diffraction of treated crystals.

The rate of dehydration must be carefully adjusted as to not destroy the crystal lattice. The dehydration rate is dependent on the difference in concentrations in the drop and in the reservoir, which are equilibrated via vapour diffusion. The ambient temperature is another parameter to influence the dehydration rate since the temperature changes the vapour pressure of the solutions. For the human 20S proteasome it was essential to use

precise concentrations of PEG 3350 and MPD as described in the post-crystallization protocol, which were determined empirically, and to let the different concentration equilibrate at 4 °C (Tab. 3.4; 3.5).

Crystals of a size larger than 250 x 250 x 250  $\mu\text{m}^3$  showed worse resolution of 2.5 Å. Very large crystals seem to contain too many imperfections in the crystals lattice, which cannot be repaired by the post-crystallization treatment explaining their low-resolution diffraction. Therefore, the optimal crystal size was deemed to be about 250 x 250 x 250  $\mu\text{m}^3$  for 20S proteasome crystals.

## 4.4 Determination of correct resolution cut-off

An X-ray crystallography experiment yields diffraction patterns showing the measured reflections. After integration of the reflections obtained from one diffraction experiment to generate the structure factors, the highest resolution shell containing significant data must be determined. Data exceeding this resolution shell, which are too weak, caused by cosmic rays or by background noise of the detector, are insignificant and need to be removed from the data before starting the structure refinement.

For several decades, crystallographers judged the significance of measured reflections based on the  $I/\sigma I$  and  $R_{\text{merge}}$ .  $I/\sigma I$  is a measure for the signal to noise ratio and is suggested to be  $\geq 2$  for reflections, which should be included in the highest resolution shell for structure refinement.  $R_{\text{merge}}$  gives information about how well the intensities of symmetry-related reflections correlate. If the variations of the intensities belonging to one reflection are too large, they are apparent in increased values for  $R_{\text{merge}}$ . Values above 0.8 hint at insignificant data and are recommended to be excluded from structural refinement [165]. Reflections of the highest resolution shell used to solve the human 20S proteasome structure at 1.8 Å had an average  $I/\sigma I$  of 0.92 and a  $R_{\text{merge}}$  value of 2.21. These numbers would suggest an overestimation of the resolution for the solved structure.

One reason to include these reflections, which would otherwise be rejected for refinement according to older standards, are the latest developments in detector technology. Currently, hybrid photon-counting (HPC) pixel array detectors are used for recording diffraction data. The great advantage of this generation of detectors is their noise-free readout, allowing the detection of weak reflections, which were lost in the background noise of the previous generation CCD detectors [166]. The better signal to noise ratio of the new detectors makes it possible to include reflections with an  $I/\sigma I$  lower than 2.0 for structural refinement.

In 2012 Karplus and Diederichs, showed that cutting data based on the general accepted  $I/\sigma I$  and  $R_{\text{merge}}$  values results in the loss of useful data [165]. Consequently, they intro-

duced an additional measure, called  $CC_{1/2}$  to determine the significance of the measured reflection intensities. The  $CC_{1/2}$  value is calculated by splitting the unmerged diffraction data into two parts and cross-correlating the averaged intensities of the two half data sets.  $CC_{1/2}$  values of 0.09 indicate significant data according to the Student's t-test [165]. The highest resolution shell used for the refinement of the 1.8 Å 20S proteasome structure presented in this thesis has a  $CC_{1/2}$  value of 0.32, implying that useful data is present in this resolution shell, which should therefore be used for structure refinement.

The  $I/\sigma I$  value for one resolution shell is specified as the average  $I/\sigma I$  for all reflections belonging to one resolution shell. As the diffraction of most of the crystals is anisotropic, useful reflections with a reasonable high  $I/\sigma I$  may have a lower over-all mean value due to weak reflections present in the same resolution shell. The Staraniso server was used to check for anisotropy of the diffraction data. The analysis revealed anisotropy evident by a suggested resolution limit of 2.023 Å along the  $a^*$ -, 1.853 Å along the  $b^*$ - and 1.820 Å along the  $c^*$ - axis. An anisotropic resolution cut-off improves the  $I/\sigma I$  to 1.5 and the  $CC_{1/2}$  to 0.53 for the highest resolution shell. This analysis demonstrates the presence of valuable high-resolution reflections up to 1.82 Å along the  $b^*$  and  $c^*$  axis (Fig. 4.1).

The points discussed here demonstrate the rationale of choosing a high resolution cut-off should not solely be based on fixed standard values of  $I/\sigma I$  and  $R_{\text{merge}}$  but should also take into account the  $CC_{1/2}$  measure and a possible anisotropy in the diffraction data. Applying a fixed standard threshold, would result in a loss of useful information that could help to improve the data quality of the solved structure. To validate the applied resolution cut-offs used in this thesis work the  $R_{\text{free}}/R_{\text{work}}$  of the highest resolution shell, which is 0.383/0.388 for the solved human 20S proteasome 1.8 Å structure, and the quality of the electron density were used.



A

```

Using MTZ column labels:  IMEAN SIGIMEAN I(+) SIGI(+) I(-) SIGI(-) N(+) N(-)
Unit cell and space group: 113.444 202.769 316.004 90.00 90.00 90.00 'P 21 21 21'
Nominal diffraction range: 316.004 1.456
Input reflection count: 874857
Unique reflection count: 874716

Eigenvalues & eigenvectors of redundancy anisotropy tensor:

1.773 1.0000 0.0000 0.0000 a*
1.757 0.0000 1.0000 0.0000 b*
2.039 0.0000 0.0000 1.0000 c*

Diffraction limits & eigenvectors of ellipsoid fitted to diffraction cut-off surface:

2.023 1.0000 0.0000 0.0000 a*
1.853 0.0000 1.0000 0.0000 b*
1.820 0.0000 0.0000 1.0000 c*

```

• Merging statistics table for observed data extracted from the final MRFANA log file:

Resolution	#uniq	#Rfac	Rmerge	Rmeas	Rpm	#Isig	I/sigI	Compl. Spher.		Multiplicity				Compl. Ellip.				
								all	ano	all	ano	CC(1/2)	#CCano	CC(ano)	SigAno	all	ano	
106.772	5.254	27850	180760	0.020	0.022	0.009	27850	69.410	0.9960	0.9942	6.50	3.47	0.9996	23408	-0.1005	0.758	0.9960	0.9942
5.254	4.153	27851	188435	0.022	0.024	0.009	27851	69.255	0.9964	0.9925	6.77	3.53	0.9996	24703	-0.1061	0.731	0.9964	0.9925
4.153	3.621	27850	192379	0.026	0.029	0.011	27850	58.664	0.9990	0.9975	6.91	3.57	0.9995	25227	-0.1269	0.739	0.9990	0.9975
3.621	3.286	27852	193145	0.034	0.037	0.014	27852	46.267	0.9986	0.9971	6.94	3.57	0.9993	25716	-0.1053	0.742	0.9986	0.9971
3.286	3.047	27847	193094	0.049	0.053	0.020	27847	32.808	0.9965	0.9901	6.94	3.58	0.9986	25499	-0.0570	0.762	0.9965	0.9901
3.047	2.865	27853	185710	0.067	0.073	0.028	27853	24.255	0.9971	0.9916	6.67	3.43	0.9973	25039	-0.0424	0.764	0.9971	0.9916
2.865	2.721	27850	190116	0.090	0.098	0.037	27850	18.708	0.9997	0.9986	6.83	3.49	0.9956	25922	-0.0235	0.772	0.9997	0.9986
2.721	2.601	27851	192163	0.116	0.126	0.048	27851	15.041	0.9991	0.9971	6.90	3.53	0.9932	25919	-0.0154	0.776	0.9991	0.9971
2.601	2.500	27852	190285	0.148	0.160	0.061	27852	12.030	0.9969	0.9913	6.84	3.51	0.9893	25736	-0.0144	0.779	0.9969	0.9913
2.500	2.413	27850	191182	0.181	0.196	0.074	27850	10.116	0.9975	0.9920	6.87	3.52	0.9853	25768	-0.0117	0.783	0.9975	0.9920
2.413	2.337	27849	191892	0.224	0.242	0.091	27849	8.409	0.9978	0.9917	6.89	3.53	0.9776	25673	-0.0085	0.789	0.9978	0.9917
2.337	2.269	27850	193323	0.276	0.298	0.112	27850	7.006	0.9987	0.9943	6.94	3.55	0.9664	25824	-0.0089	0.798	0.9987	0.9943
2.269	2.209	27849	193304	0.337	0.364	0.137	27849	5.985	0.9982	0.9962	6.94	3.54	0.9533	25918	-0.0061	0.795	0.9982	0.9962
2.209	2.155	27855	193572	0.410	0.443	0.167	27855	4.871	0.9980	0.9965	6.95	3.55	0.9311	26008	-0.0027	0.802	0.9980	0.9965
2.155	2.105	27849	191313	0.508	0.550	0.207	27849	3.941	0.9953	0.9891	6.87	3.52	0.8963	25870	-0.0010	0.802	0.9953	0.9891
2.105	2.060	27849	191730	0.621	0.671	0.253	27849	3.258	0.9922	0.9867	6.89	3.52	0.8542	25894	-0.0014	0.806	0.9922	0.9867
2.060	2.017	27856	192344	0.752	0.812	0.305	27856	2.704	0.9550	0.9508	6.91	3.53	0.8039	25890	-0.0069	0.800	0.9572	0.9530
2.017	1.971	27850	191236	0.864	0.934	0.351	27850	2.330	0.8367	0.8331	6.87	3.51	0.7517	25669	-0.0122	0.798	0.9160	0.9108
1.971	1.921	27852	189631	1.011	1.093	0.413	27852	1.986	0.6892	0.6837	6.81	3.48	0.6726	25392	-0.0142	0.802	0.9082	0.9024
1.921	1.822	27786	185449	1.286	1.394	0.531	27786	1.535	0.2967	0.2894	6.68	3.45	0.5299	24115	-0.0048	0.806	0.7488	0.7324
Total: 106.772	1.822	556951	3811063	0.062	0.067	0.025	556951	19.926	0.8657	0.8595	6.85	3.52	0.9995	509190	-0.0588	0.780	0.9707	0.9662

B

```

Using MTZ column labels:  IMEAN SIGIMEAN N(+) N(-)
Unit cell and space group: 135.660 299.046 143.840 90.00 113.02 90.00 'P 1 21 1'
Nominal diffraction range: 121.054 1.428
Input reflection count: 1452364
Unique reflection count: 1452357

Eigenvalues & eigenvectors of redundancy anisotropy tensor:

1.204 0.9583 0.0000 -0.2856 0.817 a* - 0.576 c*
1.125 0.0000 1.0000 0.0000 b*
1.173 0.2856 0.0000 0.9583 0.330 a* + 0.944 c*

Diffraction limits & eigenvectors of ellipsoid fitted to diffraction cut-off surface:

2.838 0.7998 0.0000 -0.6002 0.657 a* - 0.754 c*
1.725 0.0000 1.0000 0.0000 b*
2.439 0.6002 0.0000 0.7998 0.749 a* + 0.663 c*

```

• Merging statistics table for observed data extracted from the final MRFANA log file:

Resolution	#uniq	#Rfac	Rmerge	Rmeas	Rpm	#Isig	I/sigI	Compl. Spher.		Multiplicity				Compl. Ellip.				
								all	ano	all	ano	CC(1/2)	#CCano	CC(ano)	SigAno	all	ano	
121.054	6.348	21689	75986	0.030	0.035	0.018	21689	26.955	0.9709	0.9214	3.56	1.88	0.9984	15723	-0.2859	0.537	0.9709	0.9214
6.348	5.031	21698	68745	0.040	0.048	0.026	21698	22.929	0.9606	0.8799	3.22	1.71	0.9966	10672	-0.1188	0.679	0.9606	0.8799
5.031	4.397	21695	73397	0.041	0.049	0.026	21695	24.127	0.9783	0.9254	3.41	1.78	0.9967	12988	-0.1761	0.687	0.9783	0.9254
4.397	3.996	21698	74124	0.046	0.055	0.029	21698	21.970	0.9829	0.9447	3.43	1.76	0.9966	12643	-0.1564	0.733	0.9829	0.9447
3.996	3.711	21696	74431	0.054	0.063	0.034	21696	19.266	0.9790	0.9419	3.45	1.77	0.9968	12816	-0.1536	0.767	0.9790	0.9419
3.711	3.489	21697	73260	0.060	0.071	0.038	21697	16.581	0.9692	0.9147	3.40	1.77	0.9954	12671	-0.1700	0.772	0.9692	0.9147
3.489	3.314	21698	73761	0.069	0.083	0.044	21698	14.131	0.9718	0.9163	3.43	1.78	0.9949	13108	-0.1905	0.765	0.9718	0.9163
3.314	3.170	21695	75899	0.081	0.096	0.051	21695	11.613	0.9803	0.9365	3.52	1.81	0.9946	14308	-0.1963	0.715	0.9803	0.9365
3.170	3.048	21699	76183	0.098	0.116	0.061	21699	9.515	0.9813	0.9378	3.53	1.82	0.9942	14341	-0.1569	0.715	0.9813	0.9378
3.048	2.943	21695	76692	0.119	0.141	0.074	21695	8.039	0.9832	0.9443	3.55	1.82	0.9920	14435	-0.1070	0.712	0.9832	0.9443
2.943	2.850	21697	77159	0.139	0.164	0.087	21697	7.042	0.9661	0.9342	3.57	1.82	0.9909	14525	-0.1012	0.724	0.9661	0.9342
2.850	2.761	21698	77333	0.155	0.183	0.096	21698	6.524	0.8893	0.8612	3.57	1.82	0.9882	14535	-0.1031	0.730	0.9173	0.8871
2.761	2.674	21697	78045	0.179	0.211	0.111	21697	5.929	0.7978	0.7771	3.60	1.83	0.9837	14857	-0.0776	0.727	0.8806	0.8729
2.674	2.589	21697	77724	0.202	0.238	0.125	21697	5.471	0.7241	0.7034	3.59	1.83	0.9812	14527	-0.0423	0.757	0.8995	0.8705
2.589	2.502	21696	77313	0.248	0.292	0.153	21696	4.713	0.6197	0.5959	3.58	1.84	0.9681	14808	-0.0548	0.742	0.8665	0.8290
2.502	2.407	21696	75299	0.254	0.300	0.158	21696	4.650	0.4871	0.4513	3.50	1.83	0.9659	13982	-0.0460	0.747	0.8378	0.7712
2.407	2.304	21698	75705	0.267	0.315	0.166	21698	4.474	0.3842	0.3575	3.51	1.82	0.9590	13847	-0.0508	0.743	0.8812	0.8125
2.304	2.195	21696	74351	0.314	0.372	0.198	21696	3.966	0.2979	0.2743	3.44	1.79	0.9317	12516	-0.0383	0.768	0.8995	0.8199
2.195	2.074	21697	76090	0.392	0.463	0.245	21697	3.294	0.2202	0.2046	3.51	1.82	0.8753	13175	-0.0342	0.762	0.8890	0.8183
2.074	1.840	21704	69947	0.575	0.684	0.366	21704	2.236	0.0795	0.0647	3.26	1.80	0.7294	11100	-0.0161	0.765	0.6466	0.5124
Total: 121.054	1.840	433936	1501444	0.057	0.068	0.036	433936	11.171	0.4787	0.4591	3.48	1.81	0.9981	271577	-0.1902	0.727	0.9169	0.8612

Fig. 4.1: Staraniso analysis of the best diffracting human and yeast 20S proteasome crystals.

A: The diffraction is anisotropic along the different cell axes for human 20S proteasome crystals. Applying an anisotropic resolution cut-off increases the  $I/\sigma$  value to 1.53. B: The difference of the best and worst diffracting axes amounts to 1 Å for yeast 20S proteasome crystals. Applying an anisotropic resolution cut-off increases  $I/\sigma I$  to 2.2 in the highest resolution shell. The anisotropic diffraction manifests in a low completeness in the highest resolution shell when applying a anisotropic resolution cut-off. Both anisotropic analyses were performed with the Staraniso server <http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi>; 28.08.2018.

## 4.5 Implication of the chloride position found in the proteolytic sites

The high resolution data obtained for the human 20S proteasome allowed to observe electron densities for many ions and solvent molecules. Initially, solvent molecules and ions were modeled according to their interatomic distances. With this approach it is not possible to unambiguously annotate all electron densities to solvent molecules or specific ions.

In particular, the chloride ions located in the proteolytic active sites were questionable, as this specific position was previously believed to accommodate the nucleophilic water molecule that drives the cleavage reaction [41]. Two independent anomalous diffraction experiments validated the chloride positions. First, the strong anomalous signal of bromide ions at 13.47 keV was measured for crystals, in which chloride ions were replaced with bromide ions by soaking stabilized crystals with  $\text{MgBr}_2$ . Secondly, the weak anomalous signal for chloride ions at 6.5 keV was determined. The anomalous diffraction signal measured for chloride ions helped to identify chloride sites, where no replacement with bromide ions was possible.

As a chloride ion cannot function as nucleophile for the cleavage mechanism, the water molecule that was observed in previous published structures is very likely not the catalytic water molecule needed during the cleavage reaction [41].

The anomalous diffraction data helped to validate some more chloride positions that were previously annotated as water molecules and to model potassium and magnesium ions within the proteasome structure. For most cases, the interatomic and coordination distances of an ion allowed its correct placement. However, the analysis of anomalous diffraction data is a very helpful tool to discriminate ions in cases of doubt.

## 4.6 Catalytic water in the structure

For understanding the underlying principle of peptide hydrolysis taking place in the 20S proteasome it is important to determine the exact positions of all solvent molecules and ions in the proteolytic active sites. The anomalous diffraction data of the bromide and chloride ions revealed that the location of the catalytic water molecule suggested in previous structures is most likely wrong [41]. Depending on the proteolytic active site, two to four water molecules are apparent, each of which could potentially be involved in the reaction mechanism.

Inhibition of the proteasome with the Ketoaldehyde inhibitor displaces the water molecules

H2O-1, H2O-2 in the chymotryptic-like site and causes the exchange of the NUK-chloride with a water molecule. The C5-OH group is stabilized by a complex hydrogen bond network formed by the Arg19-carbonyl, the Tyr169-carbonyl, the Lys33-εN and the H2O-3. In contrast to the water molecule present at the NUK position, the location of the H2O-3 is at the correct stereo face providing the possibility that it might play a role in the cyclization step during inhibition.

Like for the Ketoaldehyde inhibitor, in all boronic structures the NUK chloride is exchanged with a water molecule and the inhibitors replace the waters H2O-1 and H2O-2. The covalent bond between the boron inhibitors and the Thr1 structurally resembles the tetrahedral intermediate state proposed for the peptide cleavage reaction and the transition states of ketoaldehyde and epoxyketone inhibition. In all boronic structures coordination of the conserved water molecule H2O-3 by Tyr169, Thr21, and one hydroxyl group of the boronic moiety can be observed. The boronic tetrahedral intermediate state seems to be stabilized by H2O-3.

The anomalous diffraction data measured for bromide and chloride ions and the wrong positioning NUK water molecule at the opposite stereo face of the ring closure reaction contradicts its function as catalytic water molecule. Instead of the NUK water, the presented structures suggest that the H2O-3 water molecule might be the catalytic active water molecule. H2O-3 can be found in all proteolytic active sites. In addition, it is positioned at the similar location in all inhibitor structures. The H2O-3 position at the right stereo face allows to facilitate the ring closure by the electrophilic attack of the threonyl amine for ketoaldehydes and epoxyketones. In conclusion, the water molecule H2O-3 might be the catalytic water molecule driving the cleavage reaction of the 20S proteasome.

## 4.7 Aldehyde inhibition

Several inhibition mechanisms of 20S proteasome inhibitors were reanalyzed at high resolutions. Compared to the previously published yeast proteasome structure, the higher resolution achieved for the 20S proteasome structure inhibited with an Aldehyde inhibitor presented in this thesis reveals a different conformation of the hydroxyl which is part of the formed hemiacetal linkage [155]. However, positive electron difference peaks suggest that besides the main observed conformation, the conformation described for the yeast proteasome additionally exists, but with lower occupancy.

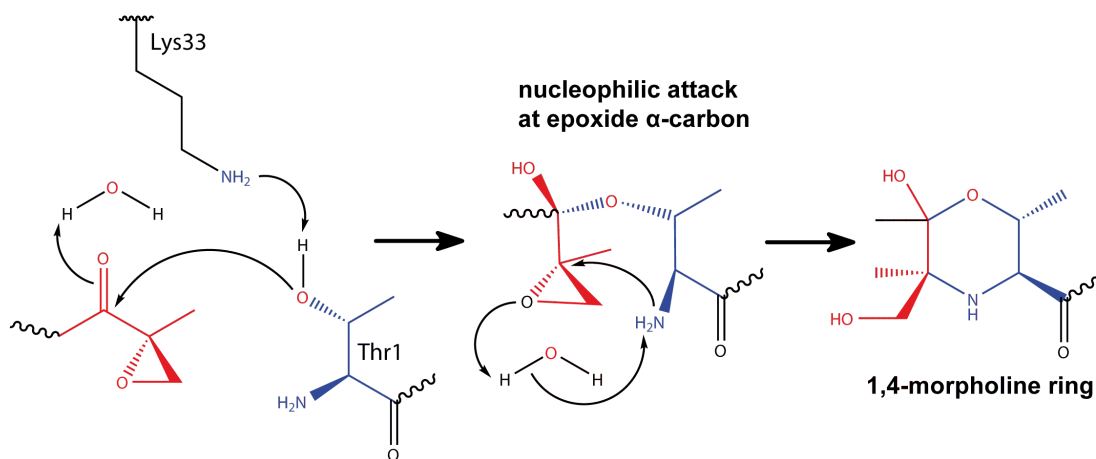
The hemiacetal conformation resembles the last step in peptide hydrolysis by the 20S proteasome. From the different observed conformations seen for the aldehyde-inhibition proteasome, one can conclude that the hydroxyl of the hemiacetal is mobile to some extent during the cleavage reaction. The rotation of the hemiacetal bond might influence

the kinetics and the processivity of peptide hydrolysis. So far, the interpretation of the described phenomena is very limited. Further experiments are needed to obtain a better understanding of the two aldehyde inhibitor conformations, in which one needs to reconstitute the hemiacetal state during peptide hydrolysis.

## 4.8 Epoxyketone inhibition

Epoxyketones belong to the most specific 20S proteasome inhibitors. Two epoxyketone have already been approved as a drug by the FDA. Therefore, the reaction mechanism of this 20S inhibitor class was analyzed in more detail.

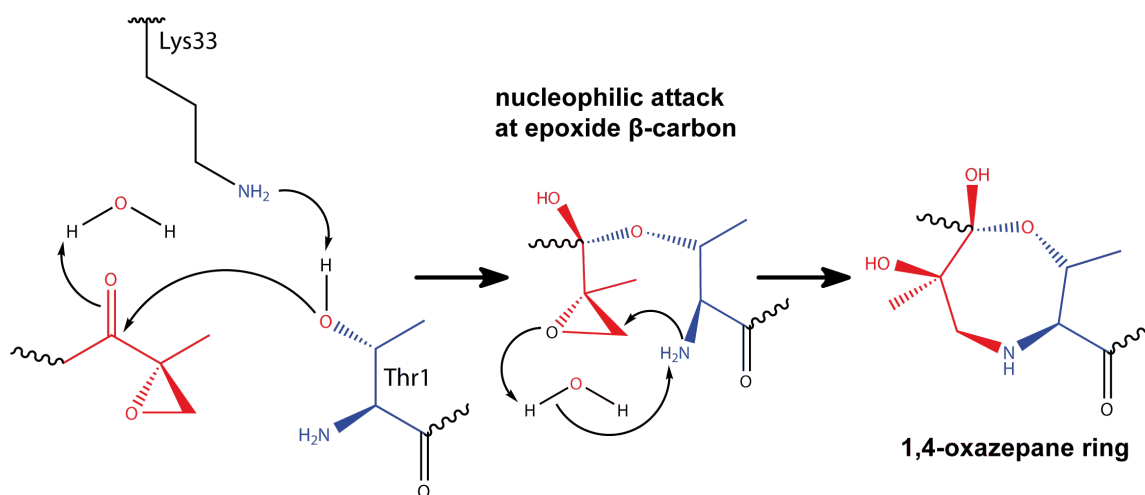
The following reaction mechanism was proposed by Groll et al. for epoxyketone inhibitors [99]. The hydroxyl of the Thr1 is deprotonated by the adjacent Lys33. The increased nucleophilicity results in a nucleophilic attack of the Thr1 hydroxyl on the carbon of the ketone group. In a second reaction step the amine group of the Thr1 electrophilic attacks the  $\alpha$ -carbon of the epoxide to form a 1,4-morpholine ring (Fig. 4.2). Groll et al. postulated the 1,4-morpholine ring structure based on their calculated electron density and justified it with the application of the Baldwin's rules for ring closures [167].



**Fig. 4.2: Published inhibition mechanism of the 20S proteasome by epoxyketone inhibitors [99]** First, the increased nucleophilicity of the Thr1/O, caused by the Lys33, enables the nucleophilic attack at the ketone carbon of the epoxyketone group. The following cyclization step is mediated via a nucleophilic attack by the Thr1/N at the  $\alpha$ -carbon of the epoxide. A 6-membered 1,4-morpholine-ring is formed via the cyclization step.

According to Baldwin's rules, the 6 Exo-tet closure leads to the 1,4-morpholine ring structure as it is favored over the 7 Endo-tet ring closure. However, the first publication describing the epoxyketone inhibition mechanism suggested that the 7 Endo-tet ring closure is the mechanism which results in an 1,4-oxazepane ring [99]. The ring closure

is designated to be Exo or Endo depending on the location of the bond that is broken during the ring closure to be either outside (exo) or inside (endo) of the newly formed ring. Analysis of the potential reaction pathway forming the 1,4-oxazepane ring suggests that the ring closure is not a 7 Endo-tet closure, but a 7 Exo-tet closure, as the broken bond between the epoxide oxygen and the  $\beta$ -carbon is not part of the 1,4-oxazepane ring. Baldwin's rules do not favor a 6 Exo-tet closure over a 7 Exo-tet closure [167]. Therefore, an electrophilic attack at the  $\beta$ -carbon based on the Baldwin's rules cannot be ruled out. Competitive refinement for the calculated electron density with a 1,4-morpholine and an 1,4-oxazepane ring resulting from an electrophilic attack at the  $\beta$ -carbon suggest a 7-membered ring is formed. Several different validation experiments were performed to rule out a reaction mechanism biased by the experimental set up and to support the newly postulated reaction pathway (Fig. 4.3).



**Fig. 4.3: Newly proposed inhibition mechanism for epoxyketone 20S proteasome inhibitors.** The cyclization step does not occur via a nucleophilic attack at the  $\alpha$ -carbon of the epoxide but at the  $\beta$ -carbon. This leads to the formation of a 7-membered 1,4-oxazepane ring instead of a 1,4-morpholine ring described previously.

The epoxyketone Dihydroeponemycin was used to structurally validate the formation of the 1,4-oxazepane ring structure. The epoxyketone group of Dihydroeponemycin carries a methanolic group at the  $\alpha$ -carbon of the epoxide. Therefore, the formation of the six membered 1,4-morpholine-ring would result in a chiral C5 carbon carrying two methanolic groups. In contrast, the formation of the 7-membered 1,4-oxazepane ring would lead to a non-chiral C6 carbon attached to a hydroxyl and methanolic group. The electron density located at the putative C5 carbon or C6 carbon, respectively, are different in their size and shape indicating the presence of a hydroxyl and methanolic group and not of two methanolic groups. These results further support the newly proposed reaction mechanism.

The solved structures presented in this thesis carry all a chloride ion at the NUK posi-

tion in the proteolytic active sites, as previously discussed. The chloride might not be present in a 20S proteasome in its native environment, but is instead caused by the high concentration of  $\text{MgCl}_2$  in the crystallization buffer. Furthermore, this chloride ion was not previously described in other structures. In this work, the chloride ion was successfully exchanged with an acetate molecule by slow transfer of the crystal to a 200 mM  $\text{Mg}(\text{OAc})_2$  containing cryo-buffer solution. Epoxyketone inhibitor soaks with  $\text{Mg}(\text{OAc})_2$  20S proteasome crystals still show the formation of an 1,4-oxazepane ring. With this result, it can be excluded that the chloride present in the native structures does change the reaction mechanisms of epoxyketones.

Reaction mechanisms deduced from X-ray crystallographic data need to be carefully interpreted. The high energy photons of X-rays might interfere with the reaction mechanism by locally changing the chemical environment of the protein forming the crystal. These phenomena are collectively attributable to radiation damage. To test for any effect of radiation damage on the reaction mechanism leading to the 1,4-oxazepane ring, one crystal was first exposed to low dosage radiation and then to a high dose of radiation before another dataset was collected to compare the influence of radiation dose on the epoxyketone inhibition mechanism. In both structures the 1,4-oxazepane ring was observed, suggesting that the reaction mechanism leading to the 1,4-oxazepane ring is not affected by radiation damage.

Electronic chemical calculations were performed to investigate the reason for the preferential formation of the 7-membered ring for epoxyketones. The transition state of 6-membered ring reaction pathway has a high energy because the N-terminal amine needs to rearrange to be able to complete the cyclization step at the  $\alpha$ -carbon. This rearrangement leads to an increased bond strain in the transition state. In contrast, the N-terminal amine is already positioned in a favorable position to attack the  $\beta$ -carbon of the epoxide. As the energy of the 7-membered ring formation pathway is lower, the 7-ring is kinetically favored over the 6-ring. Matching with known stabilities of heterocycles, the electronic chemically calculated energy states of the product states of the 6-ring and 7-ring proved the higher stability of the 6-ring compared to the 7-ring. The quantum chemical analysis provides insights into the reaction mechanisms of the 6-ring and 7-ring pathway and revealed that the epoxyketone inhibition is kinetically and not thermodynamically driven as the higher energy product state is formed.

The difference of the 6-ring and 7-ring product state was supported by a fluorometric assay. The different inhibition kinetics of the epoxyketones, a 7-ring forming inhibitor, and a ketoaldehyde, a 6-ring forming inhibitor, were analyzed. The ketoaldehyde inhibition is less reversible than the inhibition by the epoxyketones Oprozomib and Dihydroepone-mycin. The difference in reversibility can be explained by the different stabilities of the product states revealed by the quantum chemical calculations. Therefore, the kinetic analysis further validates the formation of the 1,4-oxazepane ring.

The formation of an 1,4-oxazepane ring upon inhibition with epoxyketones was demonstrated by several experiments. In conclusion, the reaction mechanism does not follow an electrophilic attack at the  $\alpha$ -carbon of the epoxide but at the  $\beta$ -carbon of the epoxide during the cyclization step (Fig. 4.3).

## 4.9 Allostery in the proteasome

Allosteric regulation is a common principle to control enzyme activity by binding of a modulator to a site remotely distant from the catalytic active site. For the proteasome no data showing the allosteric communication in the protein complex existed.

Two independent experiments showed the allosteric communication in the proteasome. cryoEM analysis of the native and Oprozomib-inhibited 26S proteasome revealed restricted movements of the 19S lid upon inhibition. The crystallographic analysis of the fruit fly 20S proteasome soaked with a peptide sequence, which is similar to the C-terminal sequence of BLM10, showed the ordering of the tertiary structure in the  $\beta$ 5-subunit in the presence of the peptide bound to the  $\alpha$ -ring pockets. The results indicate a bi-directional communication in the molecule between the proteolytic sites and the gate region which has not been described previously. The understanding of how the allostery is mediated throughout the proteasome molecule and what functional role it plays is limited and remains to be fully understood. The resolution of structures used for both experiments was too low to elucidate how the allostery is conveyed through the  $\alpha$ -rings. The restricted conformations of the inhibited 26S proteasome suggest that the proteasome can communicate its inhibition to the 19S lid to prevent translocation of more substrate into the proteasome. The peptides designed based on the C-terminal sequence of BLM10 can increase the proteolytic activity of the 20S proteasome. Interestingly, the mode how BLM10 activates the proteasome is not yet understood, as it forms a dome-like structure on top of the  $\alpha$ -rings with no obvious opening for substrates [92]. The published structural rearrangements induced by BLM10 were not observed in the solved activator-peptide structures suggesting that other residues within the BLM10 20S proteasome interface are essential to cause the described structural rearrangements [92]. However, they are not needed to enable the binding of the C-terminal residues shown by the solved structures.

For the 11S activators, which bind via their C-terminal residues to the  $\alpha$ -ring pocket similar to BLM10, the increased proteolytic activity is explained by gate opening of the 20S proteasome. This makes the proteolytic sites easier accessible for substrate [86]. The activator-20S proteasome structure provides evidence that the increase of proteolytic activity is not only related to the gate opening, but also to structural rearrangements of the  $\beta$ -subunit. One can hypothesize that the conformational change might resemble a priming step to make an efficient substrate binding possible before the substrate can be cleaved. In these experiments, first structural evidence for an allosteric communication between the proteolytic sites and the  $\alpha$ -rings are shown. They rely the signal further to the 19S lid in case of the 26S proteasome through an unknown mechanism. So far, this process can only be vaguely described, and the data only allow a rough functional interpretation of our observations. Understanding the details of the allosteric interaction between the 19S regulatory particle and the 20S core particle will provide valuable information about



the function of the proteasome.

## 4.10 Alternative conformation of the $\beta 5$ disorder in the fruit fly 20S proteasome

20S proteasome structures from four different organism were solved. However, structural analysis of the 20S proteasomes showed the disordered region of the  $\beta 5$ -subunit for the fruit fly 20S proteasome only but not for the 20S proteasome of the other organisms. The disordered state might resemble the activation state observed in the kinetic measurements for the human 20S proteasome.

The fruit fly 20S proteasome was purified from embryonic fruit fly extracts harvested from eggs collected up to 90 min after they were laid. The 20S proteasomes, which are maternally inherited, undergo low degradation turnover because *de novo* translation begins 90 minutes after an egg is laid [168]. Therefore, 20S proteasome purified from this source degraded only a small number of substrates or even none. In contrast, purified 20S proteasome are exposed to a steady-state protein turnover in yeast and high protein turnover in HeLa cells. The differing environmental conditions could explain the missing disordered region for yeast and human 20S proteasomes. Additionally, the impurity observed in the tryptic-like site of the human 20S proteasomes might prevent the formation of the disordered state due to an allosteric effect on the  $\beta 5$ -subunit.

Crystal packing of the 20S proteasomes could also explain the absence of the disordered region in the 20S proteasomes, with the exception of the fruit fly 20S proteasome. Crystal contacts can induce and stabilize conformations of the 20S proteasome. The crystals of the fruit fly 20S proteasome belong to the space group P 1, meaning they do not have any symmetry. One 2-fold screw axis can be found for the yeast and the archaeal 20S proteasome crystals and three 2-fold screw axes are present in the human 20S proteasome crystal. The higher symmetries might be caused by stabilization of the ordered conformation of the 20S proteasomes in the crystal lattice.

Further experiments are needed to test the hypothesis that the newly observed conformation of the  $\beta 5$ -subunit is connected to the protein turnover or the crystal packing. To test this hypothesis, the fruit fly 20S proteasomes could be subjected to artificial protein substrates before crystallization to cause re-ordering of the disordered region. New crystallization condition resulting in a different crystal packing might favor the ordered conformation.

## 4.11 Active pocket definition and substrate specificity

According to published literature, S1 – S4 pockets are defined by the residues surrounding the side chains P1 – P4 of an bound inhibitor [120]. On a tertiary structural level all different proteolytic active sites present in eukaryotic 20S proteasomes are identical. Therefore, sequence differences in the distinct proteolytic sites seem to mediate the substrate specificity. Analysis of cleavage patterns from yeast 20S proteasome, in which a distinct proteolytic site is inactivated by mutation, supported the idea of different substrate specificities for the respective proteolytic active site [20].

The most noticeable differences of the S1 pocket can be found at residue position 45. In each proteolytic active-site S1 pocket, another amino acid can be found at position 45. This residue can also differ between the different organisms. The chymotryptic-like site Met45 of the human 20S proteasome is replaced with a Leucine in the fruit 20S proteasome, which is also a hydrophobic residue. Moreover, the fruit fly chymotryptic-like site S1 pocket has a cysteine located at position 52 and not a threonine, which is present in the other S1 pockets. The different sizes of the residue side chains present at the 45 position might constrict the available space for the accommodation of the P1 side chain of the substrate. Inhibitors with a large residue in position P1 bind poorest to the caspase-like site (Tab. B.2). A likely reason is that the Arginine at position 45 does not allow the binding of large P1 residues. The dependency on the size of side chain in the P1 position can be seen by analysis of the sequence of active site-specific inhibitors. Inhibitors that selectively inhibit the tryptic-like site have a (4-ethylphenyl)methanamine in P1 position, chymotryptic-specific inhibitors carry a hydrophobic Isoleucine and caspase-like inhibitors an Aspartate residue in P1 position [112]. Based on this observation, the tryptic-like site should be the most active proteolytic active site, since Gly45 does not have any side chain which would restrict the access to the P1 pocket. From cleavage activities published in the literature and our own measurements, it is evident that the chymotryptic-like site is the most active proteolytic site. Therefore, the residue 45 is unlikely to be the only decisive residue to mediate substrate specificity.

From the crystallographic structures presented in this thesis, it is not possible to determine the amino acids that form the S2 and S4 pockets. There are no obvious amino acids that can be found in close proximity that could interfere with binding of the P2 and P4 residues. The S3 pockets show sequence and structural differences for each active site. Therefore, the orientation of the N-terminal cap of the inhibitor can differ for the different proteolytic sites. As the exact positioning of a potential substrate is not known, it is difficult to correlate the sequence differences of the S3 pocket to a functional role in substrate recognition. One can only hypothesize why some conformations of the N-terminal cap are unlikely to happen due to steric clashes with specific amino acids.

Overall, Inhibitor-20S proteasome structures can be used to get a general idea about sub-

strate specificity of the proteolytic active sites. However, they can provide only a static picture of substrate-binding as they only depict the final state of a binding event. Moreover, *in vivo* substrate-binding to the proteolytic sites within the proteasome is a dynamic process. For example, it could be shown that the residues of the diazomethylketone inhibitor adopt different conformations depending on their positioning relative to the active sites. Substrate specificity is not only mediated through the interaction of the P residues with the S pockets, but also through the pathways that guide residues to the respective pockets. Snapshots at intermediate steps of a substrate progressing in the proteolytic site to reach the catalytic active Thr1 are needed for to provide a better understanding of the substrate specificity of each active site.

## 5 Conclusion and Outlook

In this thesis, we obtained high-resolution crystallographic structures of the 20S proteasomes from a diverse number of species. They provide unprecedented mechanistic information about the proteolytic active sites of the proteasome. This knowledge helps to better describe the catalytic mechanisms of protein degradation that take place in the 20S proteasome. Furthermore, the observed structural dynamics in the 20S proteasome and 26S proteasome provided first insights into the complex allosteric interactions in the proteasome. Still, the exact mode of action and the regulation of the proteasome remains elusive.

We provided a new purification methodology based on an established yet neglected technique that could provide the foundation for future studies. We demonstrated that the purification protocols are likely to enable the purification of proteasomes from even more species than those presented in this study. Besides the constitutive proteasome, the immune proteasome and the thymus proteasome can likely be purified with a similar protocol. The difficulty lies in generating a cell line that exclusively expresses immune proteasomes or thymus proteasomes. Our method for proteasome purification is not selective for a specific type of proteasome species, which requires cell lines expressing only one of the three occurring human 20S proteasomes. High resolution structures of the three different 20S proteasomes occurring in humans are essential to structurally examine the functional necessity for the different kinds of 20S proteasomes. Such structures could also enable the design of selective proteasome inhibitors.

The optimized crystallization procedures, which allow for the collection of 20S proteasome data sets at a resolution of 2 Å on average, offer a framework to determine the accurate binding site and mechanisms of known and potential ligands binding to the 20S proteasome in a high-throughput manner. The routinely high-resolution structures obtained and the robust crystallization of purified proteasomes lends feasibility to using fragment-based lead discovery. The technique relies on protein structures resolved at a maximal resolution of 2.2 Å to identify potential binding fragments in the structure. New inhibition mechanisms and binding sites are likely to be found, which could result in a better understanding of the 20S proteasome function. There is not solely a scientific interest to find modulators of the 20S proteasome, but also a great interest for the development of new pharmaceutical reagents, as the 20S proteasome is already a validated drug target in cancer therapy.

To make the fragment screening more effective and reliable, crystals should be available in

different cryo-solutions containing various salt concentrations. Differences in inhibitor occupancy could already be observed comparing inhibitor soaks with crystals in the  $\text{MgCl}_2$  or  $\text{Mg}(\text{OAc})_2$  cryo-buffers. Therefore, protocols for the transfer of crystals to other cryo-solutions without losing the diffraction power of the crystals need to be developed or new crystallization conditions need to be screened.

Further work should focus on the optimization of the post-crystallization procedure, as this may allow even higher resolved structures. An even higher resolution structure would give important clues about the solvent molecules present in the 20S proteasome and reveal additional side chain conformation to better understand the allosteric communication occurring inside the protein complex.

The established protein purification protocols are the basic requirement to further address questions concerning the function of the proteasome, as the 20S and the 26S proteasome of unprecedented sample quality can now be obtained in sufficient amounts. An interesting question is how an authentic substrate can be accommodated in the 20S proteasome. Another unanswered question is how substrates are translocated via the 19S regulatory particle into the 20S proteasome and where the cleavage products exit the proteasome. Furthermore, it is still not clear if the 26S or the 30S proteasome is the functional state of the proteasome in the cell. The interplay between all available 20S proteasome adapters, such as the 19S regulatory particle, the 11S activator, and the PA200, is only vaguely understood and needs further investigation. The newly developed purification protocol yields high amounts of 20S and 26S proteasomes enabling the screening for suitable substrates and the investigation of the proteasome kinetics in the presence of different 20S proteasome adapters to address these questions.

# A Abbreviations

<b>19S RP</b>	19S regulatory particle
<b>2D</b>	two-dimensional
<b>3D</b>	three-dimensional
<b>6R</b>	6-membered
<b>7R</b>	7-membered
<b>AAA+ proteins</b>	ATPases associated with diverse cellular activities
<b>AMC</b>	7-Amino-4-Methylcoumarin
<b>AMP</b>	adenosine monophosphate
<b>ADP</b>	adenosine diphosphate
<b>ATP</b>	adenosine triphosphate
<b>bs</b>	$\beta$ -sheet
<b>BisTris</b>	bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane
<b>BS3</b>	bis(sulfosuccinimidyl)suberate
<b>BSA</b>	bovine serum albumin
<b>CCD</b>	charge coupled device
<b>cTEC</b>	cortical thymic epithelium cells
<b>CRL</b>	compound refractive lenses
<b>cryoEM</b>	3D transmission electron cryo-microscopy
<b>CTF</b>	contrast transfer function
<b>d20S</b>	<i>Drosophila melanogaster</i> 20S proteasome
<b>DMSO</b>	dimethyl sulfoxide
<b>DNA</b>	deoxyribonucleic acid
<b>DTT</b>	dithiothreitol
<b>DUB</b>	deubiquitylating enzyme
<b><i>E. coli</i></b>	<i>Escherichia coli</i>
<b>EDTA</b>	ethylenediaminetetraacetic acid
<b>EM</b>	electron microscopy
<b>ER</b>	endoplasmatic reticulum
<b>FDA</b>	Food and Drug Administration

---

<b>GA</b>	glutaraldehyde
<b>h20S</b>	<i>Homo Sapiens</i> 20S proteasome
<b>HEAT</b>	<b>H</b> untingtin, <b>e</b> longation factor 3, <b>P</b> R/ <b>A</b> subunit of protein phosphatase 2A and the <b>T</b> OR lipid kinase
<b>HEPES</b>	4-(2- <b>h</b> ydroxy <b>e</b> thyl)-1- <b>p</b> iperazine <b>e</b> thanesulfonic acid
<b>HPC</b>	hybrid photon-counting
<b>IFN-<math>\gamma</math></b>	interferon- $\gamma$
<b>IgG</b>	immunglobulin G
<b>IgA</b>	immunglobulin A
<b>IDPs</b>	intrinsically disordered proteins
<b>IPTG</b>	isopropyl $\beta$ -D-1-thiogalactopyranoside
<b>kDa</b>	kilo Dalton
<b>LMNG</b>	lauryl maltose neopentyl glycol
<b>MES</b>	2-( <i>N</i> -morpholino)ethanesulfonic acid
<b>MHC</b>	major histocompatibility complex
<b>MPD</b>	2-Methyl-2,4-pentanediol
<b>MPN</b>	<b>M</b> pr1, <b>P</b> ad1, amino- <b>N</b> -terminal
<b>mRNA</b>	messenger RNA
<b>NMR</b>	nuclear magnetic resonance
<b>NOXA</b>	Phorbol-12-myristate-13-acetate-induced protein 1
<b>OGNG</b>	octyl glucose neopentyl glycol
<b>PCA</b>	principle component analysis
<b>PCI</b>	<b>P</b> roteasome, <b>C</b> OP9 signalosome, translation <b>I</b> nitiation factor
<b>PDB</b>	Protein Data Bank
<b>PEG</b>	polyethylene glycol
<b>PEI</b>	polyethylenimine
<b>PIP</b>	proteasome interacting proteins
<b>PMSF</b>	phenylmethane sulfonyl fluoride
<b>RACs</b>	regulatory particle assembly chaperones
<b>RNA</b>	Ribonucleic acid
<b>SDS-PAGE</b>	sodium dodecyl sulfate <b>p</b> oly <b>a</b> crylamide <b>g</b> el <b>e</b> lectrophoresis
<b>SDS</b>	sodium dodecyl sulfate <b>p</b> oly <b>a</b> crylamide
<b>OB</b>	oligosaccharide binding
<b>UPS</b>	ubiquitin-proteasome pathway

<b>T cell</b>	T lymphocyte
<b>T20S</b>	<i>Thermoplasma acidophilum</i> 20S proteasome
<b>TCR</b>	T cell receptors
<b>TNF-<math>\alpha</math></b>	tumor necrosis factor- $\alpha$
<b>UPS</b>	ubiquitin-proteasome pathway
<b>v-ATPase</b>	vacuolar type H <sup>+</sup> -ATPases
<b>y20S</b>	<i>Saccharomyces cerevisiae</i> 20S proteasome





## B Supplementary information

### B.1 Table 1 of solved yeast and archaeal 20S proteasome structures

Table B.1: Table 1 of solved yeast and archaeal 20S proteasome structures

	<i>Saccharomyces cerevisiae</i>	<i>Thermoplasma acidophilum</i>
Ligand	Epoxomicin	ONX-0914
PDB ID		
Divalent salt	Mg(CH <sub>3</sub> COO) <sub>2</sub>	MgCl <sub>2</sub>
Space group		P 1 21 1
Unit cell		
a (Å) / α(°)	135.77 / 90.00	114.81 / 90.00
b (Å) / β(°)	299.37 / 113.00	198.15 / 91.32
c (Å) / γ(°)	143.99 / 90.00	308.4 / 90.00
<b>Data collection</b>		
Wavelength (Å)		0.976
Resolution range (Å)	146.69 – 2.01 (2.01 – 2.04)	198.15 – 1.95 (1.98 – 1.95)
No. of reflections	2400077	67644227
No. of unique reflections	682324	990981
Completeness (%)	97.5 (92.0)	99.3 (99.0)
<I/σI>	7.8 (0.7)	10.3 (1.0)
CC <sub>(1/2)</sub>	0.998 (0.268)	0.998 (0.270)
R <sub>pim</sub>	0.048 (1.391)	0.051 (0.844)
<b>Refinement</b>		
R <sub>work</sub>	0.201	0.252
R <sub>free</sub>	0.234	0.270
No. of atoms		
Protein	49307	92119
Ligand/solvent	3718	2581
Average B values (Å <sup>2</sup> )		
Protein	60	41
Inhibitor	51	27
Metals	60	-
Cl ions	-	-
Water	65	31
Wilson	55	44
R.m.s.d. bond length (Å)	0.016	0.011
R.m.s.d bond angles (Å)	1.832	1.541
Ramachandran % most favoured/favoured	96.42/2.75	97.62/1.90

## B.2 Plasmid map

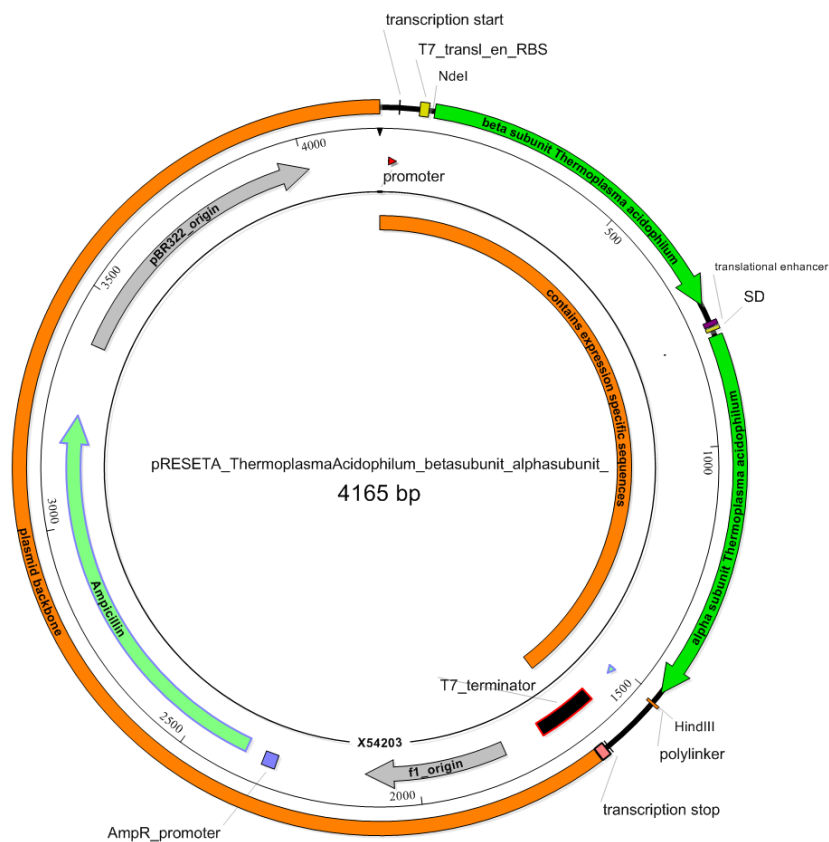


Fig. B.1: Plasmid map of vector used for *Thermoplasma Acidophilum* 20S proteasome expression.

## B.3 Occupancy list of inhibitors soaked into human 20S proteasomes

Table B.2: List of 20S proteasome inhibitors soaked into human 20S proteasomes

Inhibitor	Caspase-like site	Tryptic-like site	Chymotryptic-like site	Functional group	Condition Cl <sup>-</sup> /(OAc) <sup>-</sup>	Resolution Å
Bortezomib	X	X	X	Boron	Cl	2.1
C-3455 cold	X	impurity	X	Diazomethylketone	Cl	2.1
C-3455 warm	X	impurity	X	Diazomethylketone	Cl	2.1
C-3900 (IE(OtBu)-COH)	0	X	X	Aldehyde	Cl	2.1
Carflizomib	0	X	X	Epoxyketone	Cl	2.1
Delanzomib	X	impurity	X	Boron	Cl	2.0
Dihydroeponemycin	X	X	X	Epoxyketone	Cl	2.0
Epoxomicin	X	X	X	Epoxyketone	Cl	2.4
Ixazomib	X	X	X	Boron	Cl	2.0
LU-002c	0	X	X	Vinyl-sulfon	Cl	2.1
MG115	X	X	X	Aldehyde	Cl	2.0
MG132	X	X	X	Aldehyde	Cl	2.0
MG262	0	X	X	Boron	OAc	2.3
Oprozomib	0	X	X	Epoxyketone	Cl	1.9
Oprozomib	X	X	X	Epoxyketone	OAc	2.2
Z-LLY-Ketoaldehyde	X	X	X	Ketoaldehyde	Cl	2.1

X denotes the occupancy by an inhibitor of the respective proteolytic active site. An empty proteolytic active site is marked by a 0 . In some crystals an endogenous impurity is detected in the tryptic-like site.

## B.4 Proteasome nomenclature

### B.4.1 20S proteasome nomenclature

Table B.3: Nomenclature of *Thermoplasma acidophilum* 20S proteasome

systematic name	<i>Thermoplasma acidophilum</i>	archaeal coot chains	Function
$\alpha$ -subunit	PsmA	A - G a - g V - Z 1, 2, o -u	
$\beta$ -subunit	PsmB	H - U k - n 3 - 9	Proteolytic-active site

Table B.4: Nomenclature of eukaryotic 20S proteasomes

systematic nomenclature	HUGO	<i>Homo sapiens</i>	human coot_chains	<i>Drosophila melanogaster</i>	fruit fly coot_chains	<i>Saccharomyces cerevisiae</i>	yeast coot_chains	Function
$\alpha 1$	PSMA6	iota	G, U	Prosalpha1	G, U, 8, m	SCL1, YC7	G, U	
$\alpha 2$	PSMA2	C3	A, O	Prosalpha2	A, O, 2, g	PRE8, Y7	A, O	
$\alpha 3$	PSMA4	C9	B, P	Prosalpha3	B, P, 3, h	PRE9, Y13	B, P	
$\alpha 4$	PSMA7	C6	C, Q	Prosalpha4	C, Q, 4, i	PRE6	C, Q	
$\alpha 5$	PSMA5	zeta	D, R	Prosalpha5	D, R, 5, j	PUP2, DOA5	D, R	
$\alpha 6$	PSMA1	C2	E,S	Prosalpha6	E,S, 6, k	PRE5	E,S	
$\alpha 7$	PSMA3	C8	F, T	Prosalpha7	F, T, 7, l	PRE10, YC1	F, T	
$\beta 1$	PSMB6	Y, delta	N, b	Prosbeta1	N, 1, f, t	PRE3	N, b	Caspase-like activity
$\beta 2$	PSMB7	Z	H, V	Prosbeta2	H, V, 9, n	PUP1	H, V	Tryptic-like activity
$\beta 3$	PSMB3	C10	I, W	Prosbeta3	I, W, a, o	PUP3	I, W	
$\beta 4$	PSMB2	C7	J, X	Prosbeta4	J, X, b, p	PRE1	J, X	
$\beta 5$	PSMB5	X	K, Y	Prosbeta5	K, Y, c, q	PRE2, DOA3	K, Y	Chymotryptic-like activity
$\beta 6$	PSMB1	C5	L, Z	Prosbeta6	L, Z, d, r	PRE7	L, Z	
$\beta 7$	PSMB4	N3, beta	M, a	Prosbeta7	M, 0, e, s	PRE4	M, a	
$\beta 1i$	PSMB9	LMP2						
$\beta 2i$	PSMB10	MECL1						
$\beta 5i$	PSMB8	LMP7						
$\beta 5t$	PSMB11	PSMB11						

HUGO: Human Genome Organization

**B.4.2 19S regulatory, activator and assembly chaperone subunits**

**Table B.5: Nomenclature of 19S regulatory, proteasome activators and assembly chaperones subunits [18]**

Category	Subclassification	Systematic nomenclature	HUGO	Human	Yeast	Function	human_coot_chain
19S RP	ATPase subunits	Rpt1	PSMC2	S7, Mss1	YTA3, CIM5	ATPase	c
		Rpt2	PSMC1	S4, p56	YTA5/mts2	ATPase, Gate-opening	d
		Rpt3	PSMC4	S6, Tbp7, P48	YTA2	ATPase, Gate-opening	e
		Rpt4	PSMC6	S10b, p42	SUG2, PCS1, CRL13	ATPase,	f
		Rpt5	PSMC3	S6', Tbp1	YTA1	ATPase, Gate-opening	g
		Rpt6	PSMC5	S8, p45, Trip1	SUG1, CRL3, CIM3/let1	ATPase	h
	non-ATPase subunits	Rpn1	PSMD2	S2, p97	HRD2, NAS1/mts4	PIPs scaffold	
		Rpn2	PSMD1	S1, p112	SEN3	PIPs scaffold	i
		Rpn3	PSMD3	S23, p58	SUN2		j
		Rpn5	PSMD12	p55	NAS5		k
		Rpn6	PSMD11	S9, p44.5	NAS4		l
		Rpn7	PSMD6	S10a, P44			m
		Rpn8	PSMD7	S12, p40, MOV34	NAS3		n
		Rpn9	PSMD13	S11, p40.5	NAS7/mts1		o
		Rpn10	PSMD4	S5a, Mbp1	SUN1, MCB1/pus1	Ub receptor	p
Rpn11	PSMD14	S13, Poh1	MPR1/pad1, mts5	DUB	q		
Rpn12	PSMD8	S14, p31	NIN/mts3		r		
Rpn13	ADRM1	ADRM1	DAQ1	UB receptor, Uch37 recruit	-		
Rpn15	SHFM1	DSS1, SHFM11	SEM1		s		
PA28		PSME1	PA28alpha, REGalpha	-	activator		
		PSME2	PA28beta, REGbeta	-	activator		
		PMSE3	PA28gamma, REGgamma, Ki	-	activator		
PA200		PSME4	PA200, TEMO	BLM10	activator		
Assembling chaperones		hUmp1	POMP	Proteasemblin	UMP1		
		PAC1	PSMG1		Pba1, Poc1		
		PAC2	PSMG2		Pba2, ADD66, Poc2		
		PAC3	PSMG3		Pba3, Dmp2, Poc3		
		PAC4	PSMG4		Pba4, Dmp1, Poc4		



## B.5 Declaration of contribution as co-author

**The inhibition mechanism of human 20S proteasomes enables next-generation inhibitor design**

Schrader J, Henneberg F, Mata RA, Tittmann K, Schneider TR, Stark H, Bourenkov G, Chari A: Science, 353 (6299): 594-598

### **Thesis author's contributions**

Conception: 20 %

Experimental contribution: 40 %

Formulation of results: 30 %

**Long-range allosteric regulation of the human 26S proteasome by 20S proteasome-targeting cancer drugs**

Haselbach D, Schrader J, Lambrecht F, Henneberg F, Chari A, Stark H: Nat. Commun., 8: 15578

### **Thesis author's contributions**

Conception: 10 %

Experimental contribution: 10 %

Formulation of results: 10 %

## C References

- [1] Efeyan, A., Comb, W. C. & Sabatini, D. M. Nutrient-sensing mechanisms and pathways. *Nature* 517, 302–310 (2015). <http://doi.org/10.1038/nature14190>.
- [2] Crick, F. Central dogma of molecular biology. *Nature* 227, 561–3 (1970). <http://doi.org/10.1038/227561a0>.
- [3] Meister, S. *et al.* Extensive immunoglobulin production sensitizes myeloma cells for proteasome inhibition. *Cancer research* 67, 1783–92 (2007). <http://doi.org/10.1158/0008-5472.CAN-06-2258>.
- [4] Hornig, N. C., Knowles, P. P., McDonald, N. Q. & Uhlmann, F. The Dual Mechanism of Separase Regulation by Securin. *Current Biology* 12, 973–982 (2002). [http://doi.org/10.1016/S0960-9822\(02\)00847-3](http://doi.org/10.1016/S0960-9822(02)00847-3).
- [5] Hartl, F. U. Cellular Homeostasis and Aging. *Annual Review of Biochemistry* 85, 1–4 (2016). <http://doi.org/10.1146/annurev-biochem-011116-110806>.
- [6] Appelmans, F. & de Duve, C. Tissue fractionation studies. 3. Further observations on the binding of acid phosphatase by rat-liver particles. *The Biochemical journal* 59, 426–33 (1955). <http://doi.org/10.1042/bj0590426>.
- [7] Forgac, M. Vacuolar ATPases: Rotary proton pumps in physiology and pathophysiology. *Nature Reviews Molecular Cell Biology* 8, 917–929 (2007). <http://doi.org/10.1038/nrm2272>.
- [8] Glick, D., Barth, S. & Macleod, K. F. Autophagy : cellular and molecular mechanisms. *Journal of Pathology The* 221, 3–12 (2010). <http://doi.org/10.1002/path.2697>. Autophagy.
- [9] Perera, R. M. & Zoncu, R. The Lysosome as a Regulatory Hub. *Annual Review of Cell and Developmental Biology* 32, 223–253 (2016). <http://doi.org/10.1146/annurev-cellbio-111315-125125>.
- [10] Ciechanover, A., Hod, Y. & Hershko, A. A heat-stable polypeptide component of an ATP-dependent proteolytic system from reticulocytes. *Biochemical and biophysical research communications* 81, 1100–5 (1978). [http://doi.org/10.1016/0006-291X\(78\)91249-4](http://doi.org/10.1016/0006-291X(78)91249-4).
- [11] Clague, M. J. & Urbé, S. Ubiquitin: Same molecule, different degradation pathways. *Cell* 143, 682–685 (2010). <http://doi.org/10.1016/j.cell.2010.11.012>.
- [12] Hochstrasser, M. Origin and function of ubiquitin-like proteins. *Nature* 458, 422–429 (2009). <http://doi.org/10.1038/nature07958>.
- [13] Kerscher, O., Felberbaum, R. & Hochstrasser, M. Modification of Proteins by Ubiquitin and Ubiquitin-Like Proteins. *Annual Review of Cell and Developmental Biology* 22, 159–180 (2006). <http://doi.org/10.1146/annurev.cellbio.22.010605.093503>.
- [14] Hershko, A. Ubiquitin-mediated protein degradation. *The Journal of biological chemistry* 263, 15237–40 (1988).

- [15] Pickart, C. M. & Eddins, M. J. Ubiquitin: Structures, functions, mechanisms. *Biochimica et Biophysica Acta - Molecular Cell Research* 1695, 55–72 (2004). <http://doi.org/10.1016/j.bbamcr.2004.09.019>.
- [16] Hershko, A. & Ciechanover, A. The ubiquitin system. *Annual review of biochemistry* 67, 425–79 (1998). <http://doi.org/10.1146/annurev.biochem.67.1.425>.
- [17] Chau, V. *et al.* A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science (New York, N.Y.)* 243, 1576–83 (1989). <http://doi.org/10.1126/science.2538923>.
- [18] Tanaka, K. The proteasome: overview of structure and functions. *Proceedings of the Japan Academy. Series B, Physical and biological sciences* 85, 12–36 (2009). <http://doi.org/10.2183/pjab.85.12>.
- [19] Kisselev, A. F., Akopian, T. N., Castillo, V. & Goldberg, A. L. Proteasome active sites allosterically regulate each other, suggesting a cyclical bite-chew mechanism for protein breakdown. *Molecular Cell* 4, 395–402 (1999). [http://doi.org/10.1016/S1097-2765\(00\)80341-X](http://doi.org/10.1016/S1097-2765(00)80341-X).
- [20] Nussbaum, A. K. *et al.* Cleavage motifs of the yeast 20S proteasome beta subunits deduced from digests of enolase 1. *Proceedings of the National Academy of Sciences of the United States of America* 95, 12504–9 (1998). <http://doi.org/10.1073/PNAS.95.21.12504>.
- [21] Wenzel, T., Eckerskorn, C., Lottspeich, F. & Baumeister, W. Existence of a molecular ruler in proteasomes suggested by analysis of degradation products. *FEBS Letters* 349, 205–209 (1994). [http://doi.org/10.1016/0014-5793\(94\)00665-2](http://doi.org/10.1016/0014-5793(94)00665-2).
- [22] Harris, J. Release of a macromolecular protein component from human erythrocyte ghosts. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 150, 534–537 (1968). [http://doi.org/10.1016/0005-2736\(68\)90157-0](http://doi.org/10.1016/0005-2736(68)90157-0).
- [23] Pühler, G., Pitzer, F., Zwickl, P. & Baumeister, W. Proteasomes: Multisubunit Proteinases common to Thermoplasma and Eukaryotes. *Systematic and Applied Microbiology* 16, 734–741 (1993). [http://doi.org/10.1016/S0723-2020\(11\)80347-4](http://doi.org/10.1016/S0723-2020(11)80347-4).
- [24] Jastrab, J. B. & Darwin, K. H. Bacterial Proteasomes. *Annual Review of Microbiology* 69, 109–127 (2015). <http://doi.org/10.1146/annurev-micro-091014-104201>.
- [25] Löwe, J. *et al.* Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science (New York, N.Y.)* 268, 533–9 (1995). <http://doi.org/10.1126/science.7725097>.
- [26] Dahlmann, B. *et al.* The multicatalytic proteinase (prosome) is ubiquitous from eukaryotes to archaeobacteria. *FEBS letters* 251, 125–31 (1989). [http://doi.org/10.1016/0014-5793\(89\)81441-3](http://doi.org/10.1016/0014-5793(89)81441-3).
- [27] Voges, D., Zwickl, P. & Baumeister, W. the 26S Proteasome : a Molecular Machine Designed for Controlled. *Annual review of biochemistry* 68, 1015–1068 (1999). <http://doi.org/10.1146/annurev.biochem.68.1.1015>.
- [28] Seemüller, E. *et al.* Proteasome from *Thermoplasma acidophilum*: a threonine protease. *Science (New York, N.Y.)* 268, 579–82 (1995). <http://doi.org/10.1126/science.7725107>.
- [29] Ekici, Ö. D., Paetzel, M. & Dalbey, R. E. Unconventional serine proteases: Variations on the catalytic Ser/His/Asp triad configuration. *Protein Science* 17, 2023–2037 (2008). <http://doi.org/10.1110/ps.035436.108>.

- [30] Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDScript server. *Nucleic Acids Research* 42, 320–324 (2014). <http://doi.org/10.1093/nar/gku316>.
- [31] Groll, M. *et al.* Structure of 20S proteasome from yeast at 2.4 Å resolution (1997). <http://doi.org/10.1038/386463a0>.
- [32] Unno, M. *et al.* The structure of the mammalian 20S proteasome at 2.75 Å resolution. *Structure (London, England : 1993)* 10, 609–18 (2002). [http://doi.org/10.1016/S0969-2126\(02\)00748-7](http://doi.org/10.1016/S0969-2126(02)00748-7).
- [33] Zwickl, P., Lottspeich, F. & Baumeister, W. Expression of functional Thermoplasma acidophilum proteasomes in Escherichia coli. *FEBS Letters* 312, 157–160 (1992). [http://doi.org/10.1016/0014-5793\(92\)80925-7](http://doi.org/10.1016/0014-5793(92)80925-7).
- [34] Seemuller, E., Lupas, A. & Baumeister, W. Autocatalytic processing of the 20S proteasome. *Nature* 382, 468–71 (1996). <http://doi.org/10.1038/382468a0>.
- [35] Lin, G. *et al.* Mycobacterium tuberculosis prcBA genes encode a gated proteasome with broad oligopeptide specificity. *Molecular microbiology* 59, 1405–16 (2006). <http://doi.org/10.1111/j.1365-2958.2005.05035.x>.
- [36] Zwickl, P., Kleinz, J. & Baumeister, W. Critical elements in proteasome assembly. *Nature Structural Biology* 1, 765–770 (1994). <http://doi.org/10.1038/nsb1194-765>.
- [37] Le Tallec, B. *et al.* 20S proteasome assembly is orchestrated by two distinct pairs of chaperones in yeast and in mammals. *Molecular cell* 27, 660–74 (2007). <http://doi.org/10.1016/j.molcel.2007.06.025>.
- [38] Kurimoto, E. *et al.* Crystal structure of human proteasome assembly chaperone PAC4 involved in proteasome formation. *Protein science : a publication of the Protein Society* 26, 1080–1085 (2017). <http://doi.org/10.1002/pro.3153>.
- [39] Sikdar, A., Satoh, T., Kawasaki, M. & Kato, K. Crystal structure of archaeal homolog of proteasome-assembly chaperone PbaA. *Biochemical and Biophysical Research Communications* 453, 493–497 (2014). <http://doi.org/10.1016/j.bbrc.2014.09.114>.
- [40] Budenholzer, L., Cheng, C. L., Li, Y. & Hochstrasser, M. Proteasome Structure and Assembly. *Journal of molecular biology* 429, 3500–3524 (2017). <http://doi.org/10.1016/j.jmb.2017.05.027>.
- [41] Huber, E. M. *et al.* A unified mechanism for proteolysis and autocatalytic activation in the 20S proteasome. *Nature communications* 7, 10900 (2016). <http://doi.org/10.1038/ncomms10900>.
- [42] Groll, M. *et al.* The catalytic sites of 20S proteasomes and their role in subunit maturation: a mutational and crystallographic study. *Proceedings of the National Academy of Sciences of the United States of America* 96, 10976–83 (1999). <http://doi.org/10.1073/pnas.96.20.10976>.
- [43] Uechi, H., Hamazaki, J. & Murata, S. Characterization of the Testis-specific Proteasome Subunit  $\alpha$ 4s in Mammals. *Journal of Biological Chemistry* 289, 12365–12374 (2014). <http://doi.org/10.1074/jbc.M114.558866>.
- [44] Dahlmann, B. Mammalian proteasome subtypes: Their diversity in structure and function. *Archives of Biochemistry and Biophysics* 591, 132–140 (2016). <http://doi.org/10.1016/j.abb.2015.12.012>.

- [45] Basler, M., Kirk, C. J. & Groettrup, M. The immunoproteasome in antigen processing and other immunological functions. *Current Opinion in Immunology* 25, 74–80 (2013). <http://doi.org/10.1016/j.coi.2012.11.004>.
- [46] Guillaume, B. *et al.* Two abundant proteasome subtypes that uniquely process some antigens presented by HLA class I molecules. *Proceedings of the National Academy of Sciences of the United States of America* 107, 18599–604 (2010). <http://doi.org/10.1073/pnas.1009778107>.
- [47] Bjorkman, P. J. *et al.* Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* 329, 506–12 (1987). <http://doi.org/10.1038/329506a0>.
- [48] Gaczynska, M., Rock, K. L. & Goldberg, A. L. Role of proteasomes in antigen presentation. *Enzyme & protein* 47, 354–69 (1993).
- [49] Huber, E. M. *et al.* Immuno- and constitutive proteasome crystal structures reveal differences in substrate and inhibitor specificity. *Cell* 148, 727–38 (2012). <http://doi.org/10.1016/j.cell.2011.12.030>.
- [50] Sasaki, K. *et al.* Thymoproteasomes produce unique peptide motifs for positive selection of CD8(+) T cells. *Nature communications* 6, 7484 (2015). <http://doi.org/10.1038/ncomms8484>.
- [51] Murata, S. *et al.* Regulation of CD8+ T cell development by thymus-specific proteasomes. *Science (New York, N.Y.)* 316, 1349–53 (2007). <http://doi.org/10.1126/science.1141915>.
- [52] Schechter, I. & Berger, A. On the size of the active site in proteases. I. Papain. 1967. *Biochemical and biophysical research communications* 425, 497–502 (2012). <http://doi.org/10.1016/j.bbrc.2012.08.015>.
- [53] Orlowski, M. & Wilk, S. A multicatalytical protease complex from pituitary that forms enkephalin and enkephalin containing peptides. *Biochemical and Biophysical Research Communications* 101, 814–822 (1981). [http://doi.org/10.1016/0006-291X\(81\)91823-4](http://doi.org/10.1016/0006-291X(81)91823-4).
- [54] Wilk, S. & Orlowski, M. Evidence that Pituitary Cation-Sensitive Neutral Endopeptidase Is a Multicatalytic Protease Complex. *Journal of Neurochemistry* 40, 842–849 (1983). <http://doi.org/10.1111/j.1471-4159.1983.tb08056.x>.
- [55] Collins, G. A. & Goldberg, A. L. The Logic of the 26S Proteasome. *Cell* 169, 792–806 (2017). <http://doi.org/10.1016/j.cell.2017.04.023>.
- [56] Cascio, P., Call, M., Petre, B. M., Walz, T. & Goldberg, A. L. Properties of the hybrid form of the 26S proteasome containing both 19S and PA28 complexes. *The EMBO Journal* 21, 2636–2645 (2002). <http://doi.org/10.1093/emboj/21.11.2636>.
- [57] Blickwedehl, J. *et al.* Role for proteasome activator PA200 and postglutamyl proteasome activity in genomic stability. *Proceedings of the National Academy of Sciences of the United States of America* 105, 16165–70 (2008). <http://doi.org/10.1073/pnas.0803145105>.
- [58] Glickman, M. H. *et al.* A subcomplex of the proteasome regulatory particle required for ubiquitin-conjugate degradation and related to the COP9-signalosome and eIF3. *Cell* 94, 615–23 (1998). [http://doi.org/10.1016/S0092-8674\(00\)81603-7](http://doi.org/10.1016/S0092-8674(00)81603-7).
- [59] Kish-Trier, E. & Hill, C. P. Structural Biology of the Proteasome. *Annual Review of Biophysics* 42, 29–49 (2013). <http://doi.org/10.1146/annurev-biophys-083012-130417>.

- 
- [60] Lander, G. C., Martin, A. & Nogales, E. The proteasome under the microscope: The regulatory particle in focus. *Current Opinion in Structural Biology* 23, 243–251 (2013). <http://doi.org/10.1016/j.sbi.2013.02.004>.
- [61] Schweitzer, A. *et al.* Structure of the human 26S proteasome at a resolution of 3.9 Å. *Proceedings of the National Academy of Sciences of the United States of America* 113, 7816–21 (2016). <http://doi.org/10.1073/pnas.1608050113>.
- [62] Huang, X., Luan, B., Wu, J. & Shi, Y. An atomic structure of the human 26S proteasome. *Nature Structural & Molecular Biology* 23, 778–785 (2016). <http://doi.org/10.1038/nsmb.3273>.
- [63] Haselbach, D. *et al.* Long-range allosteric regulation of the human 26S proteasome by 20S proteasome-targeting cancer drugs. *Nature communications* 8, 15578 (2017). <http://doi.org/10.1038/ncomms15578>.
- [64] Unverdorben, P. *et al.* Deep classification of a large cryo-EM dataset defines the conformational landscape of the 26S proteasome. *Proceedings of the National Academy of Sciences of the United States of America* 111, 5544–9 (2014). <http://doi.org/10.1073/pnas.1403409111>.
- [65] Śledź, P. & Baumeister, W. Structure-Driven Developments of 26S Proteasome Inhibitors. *Annual review of pharmacology and toxicology* 56, 191–209 (2016). <http://doi.org/10.1146/annurev-pharmtox-010814-124727>.
- [66] Dambacher, C. M., Worden, E. J., Herzik, M. A., Martin, A. & Lander, G. C. Atomic structure of the 26S proteasome lid reveals the mechanism of deubiquitinase inhibition. *eLife* 5, e13027 (2016). <http://doi.org/10.7554/eLife.13027>.
- [67] Nyquist, K. & Martin, A. Marching to the beat of the ring: Polypeptide translocation by AAA+ proteases. *Trends in Biochemical Sciences* 39, 53–60 (2014). <http://doi.org/10.1016/j.tibs.2013.11.003>.
- [68] Sledz, P. *et al.* Structure of the 26S proteasome with ATP- S bound provides insights into the mechanism of nucleotide-dependent substrate translocation. *Proceedings of the National Academy of Sciences* 110, 7264–7269 (2013). <http://doi.org/10.1073/pnas.1305782110>.
- [69] Smith, D. M. *et al.* Docking of the proteasomal ATPases' carboxyl termini in the 20S proteasome's alpha ring opens the gate for substrate entry. *Molecular cell* 27, 731–44 (2007). <http://doi.org/10.1016/j.molcel.2007.06.033>.
- [70] Paraskevopoulos, K. *et al.* Dss1 is a 26S proteasome ubiquitin receptor. *Molecular cell* 56, 453–61 (2014). <http://doi.org/10.1016/j.molcel.2014.09.008>.
- [71] Querol-Audi, J. *et al.* Architecture of human translation initiation factor 3. *Structure (London, England : 1993)* 21, 920–8 (2013). <http://doi.org/10.1016/j.str.2013.04.002>.
- [72] Lingaraju, G. M. *et al.* Crystal structure of the human COP9 signalosome. *Nature* 512, 161–5 (2014). <http://doi.org/10.1038/nature13566>.
- [73] Verma, R. *et al.* Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. *Science (New York, N.Y.)* 298, 611–5 (2002). <http://doi.org/10.1126/science.1075898>.

- [74] Wehmer, M. & Sakata, E. Recent advances in the structural biology of the 26S proteasome. *International Journal of Biochemistry and Cell Biology* 79, 437–442 (2016). <http://doi.org/10.1016/j.biocel.2016.08.008>.
- [75] Tomko, R. J. *et al.* A Single  $\alpha$  Helix Drives Extensive Remodeling of the Proteasome Lid and Completion of Regulatory Particle Assembly. *Cell* 163, 432–444 (2015). <http://doi.org/10.1016/j.cell.2015.09.022>.
- [76] Estrin, E., Lopez-Blanco, J. R., Chacón, P. & Martin, A. Formation of an intricate helical bundle dictates the assembly of the 26S proteasome Lid. *Structure* 21, 1624–1635 (2013). <http://doi.org/10.1016/j.str.2013.06.023>.
- [77] Ben-Nissan, G. & Sharon, M. Regulating the 20S Proteasome Ubiquitin-Independent Degradation Pathway. *Biomolecules* 4, 862–884 (2014). <http://doi.org/10.3390/biom4030862>.
- [78] Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology* 6, 197–208 (2005). <http://doi.org/10.1038/nrm1589>.
- [79] Pickering, A. M. & Davies, K. J. A. Differential roles of proteasome and immunoproteasome regulators Pa28 $\alpha\beta$ , Pa28 $\gamma$  and Pa200 in the degradation of oxidized proteins. *Archives of Biochemistry and Biophysics* 523, 181–190 (2012). <http://doi.org/10.1016/j.abb.2012.04.018>.
- [80] Stadtmueller, B. M. & Hill, C. P. Proteasome Activators. *Molecular Cell* 41, 8–19 (2011). <http://doi.org/10.1016/j.molcel.2010.12.020>.
- [81] Welk, V. *et al.* Inhibition of proteasome activity induces formation of alternative proteasome complexes. *Journal of Biological Chemistry* 291, 13147–13159 (2016). <http://doi.org/10.1074/jbc.M116.717652>.
- [82] Huber, E. M. & Groll, M. The Mammalian Proteasome Activator PA28 Forms an Asymmetric  $\alpha 4\beta 3$  Complex. *Structure* 25, 1473–1480.e3 (2017). <http://doi.org/10.1016/j.str.2017.07.013>.
- [83] Yu, G. *et al.* Comparative analysis of REG $\{\gamma\}$  expression in mouse and human tissues. *Journal of molecular cell biology* 2, 192–198 (2010). <http://doi.org/10.1093/jmcb/mjq009>.
- [84] Knowlton, J. R. *et al.* Structure of the proteasome activator REG $\alpha$  (PA28 $\alpha$ ). *Nature* 390, 639–643 (1997). <http://doi.org/10.1038/37670>.
- [85] Whitby, F. G. *et al.* Structural basis for the activation of 20S proteasomes by 11S regulators. *Nature* 408, 115–120 (2000). <http://doi.org/10.1038/35040607>.
- [86] Förster, A., Masters, E. I., Whitby, F. G., Robinson, H. & Hill, C. P. The 1.9 Å structure of a proteasome-11S activator complex and implications for proteasome-PAN/PA700 interactions. *Molecular cell* 18, 589–99 (2005). <http://doi.org/10.1016/j.molcel.2005.04.016>.
- [87] Förster, A., Whitby, F. G. & Hill, C. P. The pore of activated 20S proteasomes has an ordered 7-fold symmetric conformation. *The EMBO journal* 22, 4356–64 (2003). <http://doi.org/10.1093/emboj/cdg436>.
- [88] Chu-Ping, M., Willy, P. J., Slaughter, C. A. & DeMartino, G. N. PA28, an activator of the 20 S proteasome, is inactivated by proteolytic modification at its carboxyl terminus. *Journal of Biological Chemistry* 268, 22514–22519 (1993).

- 
- [89] Zhang, Z. *et al.* Identification of an activation region in the proteasome activator REGal-pha. *Proceedings of the National Academy of Sciences of the United States of America* 95, 2807–2811 (1998). <http://doi.org/10.1073/pnas.95.6.2807>.
- [90] Ustrell, V., Hoffman, L., Pratt, G. & Rechsteiner, M. PA200, a nuclear proteasome activator involved in DNA repair. *The EMBO Journal* 21, 3516–3525 (2002). <http://doi.org/10.1093/emboj/cdf333>.
- [91] Savulescu, A. F. & Glickman, M. H. Proteasome Activator 200: The HEAT is on... *Molecular & Cellular Proteomics* 10, R110.006890 (2011). <http://doi.org/10.1074/mcp.R110.006890>.
- [92] Sadre-bazzaz, K., Whitby, F. G., Robinson, H., Formosa, T. & Hill, C. P. Structure of a Blm10 Complex Reveals Common Mechanisms for Proteasome Binding and Gate Opening. *Molecular Cell* 37, 728–735 (2010). <http://doi.org/10.1016/j.molcel.2010.02.002>.
- [93] Omura, S. *et al.* Lactacystin, a novel microbial metabolite, induces neurogenesis of neuroblastoma cells. *The Journal of antibiotics* 44, 113–6 (1991). <http://doi.org/10.7164/antibiotics.44.113>.
- [94] Fenteany, G. *et al.* Inhibition of proteasome activities and subunit-specific amino-terminal threonine modification by lactacystin. *Science* 268, 726–731 (1995). <http://doi.org/10.1126/science.7732382>.
- [95] Kisselev, A., van der Linden, W. A. & Overkleeft, H. Proteasome Inhibitors: An Expanding Army Attacking a Unique Target. *Chemistry & Biology* 19, 99–115 (2012). <http://doi.org/10.1016/j.chembiol.2012.01.003>.
- [96] Lindsten, K. *et al.* A transgenic mouse model of the ubiquitin/proteasome system. *Nature Biotechnology* 21, 897–902 (2003). <http://doi.org/10.1038/nbt851>.
- [97] Manasanch, E. E. & Orlowski, R. Z. Proteasome inhibitors in cancer therapy. *Nature Reviews Clinical Oncology* 14, 417–433 (2017). <http://doi.org/10.1038/nrclinonc.2016.206>.
- [98] Meng, L. *et al.* Epoxomicin, a potent and selective proteasome inhibitor, exhibits in vivo antiinflammatory activity. *Proceedings of the National Academy of Sciences of the United States of America* 96, 10403–8 (1999). <http://doi.org/10.1073/pnas.96.18.10403>.
- [99] Groll, M., Kim, K. B., Kairies, N., Huber, R. & Crews, C. M. Crystal Structure of Epoxomicin:20S Proteasome Reveals a Molecular Basis for Selectivity of  $\alpha'$ , $\beta'$ -Epoxyketone Proteasome Inhibitors. *Journal of the American Chemical Society* 122, 1237–1238 (2000). <http://doi.org/10.1021/ja993588m>.
- [100] Lynas, J. F., Harriott, P., Healy, A., McKervey, M. A. & Walker, B. Inhibitors of the chymotrypsin-like activity of proteasome based on di- and tri-peptidyl  $\alpha$ -keto aldehyde (glyoxals). *Bioorganic and Medicinal Chemistry Letters* 8, 373–378 (1998). [http://doi.org/10.1016/S0960-894X\(98\)00030-4](http://doi.org/10.1016/S0960-894X(98)00030-4).
- [101] Gräwert, M. A. *et al.* Elucidation of the  $\alpha$ -keto-aldehyde binding mechanism: a lead structure motif for proteasome inhibition. *Angewandte Chemie (International ed. in English)* 50, 542–4 (2011). <http://doi.org/10.1002/anie.201005488>.
- [102] Bogyo, M. *et al.* Covalent modification of the active site threonine of proteasomal  $\beta$  subunits and the Escherichia coli homolog HsIV by a new class of inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* 94, 6629–6634 (1997). <http://doi.org/10.1073/pnas.94.13.6629>.



- [103] Groll, M., Nazif, T., Huber, R. & Bogoy, M. Probing Structural Determinants Distal to the Site of Hydrolysis that Control Substrate Specificity of the 20S Proteasome. *Chemistry & Biology* 9, 655–662 (2002). [http://doi.org/10.1016/S1074-5521\(02\)00144-8](http://doi.org/10.1016/S1074-5521(02)00144-8).
- [104] Groll, M. *et al.* A plant pathogen virulence factor inhibits the eukaryotic proteasome by a novel mechanism. *Nature* 452, 755–758 (2008). <http://doi.org/10.1038/nature06782>.
- [105] Archer, C. R. *et al.* Activity Enhancement of the Synthetic Syrbactin Proteasome Inhibitor Hybrid and Biological Evaluation in Tumor Cells. *Biochemistry* 51, 6880–6888 (2012). <http://doi.org/10.1021/bi300841r>.
- [106] Groll, M., Larionov, O. V., Huber, R. & de Meijere, A. Inhibitor-binding mode of homobactosin C to proteasomes: new insights into class I MHC ligand generation. *Proceedings of the National Academy of Sciences of the United States of America* 103, 4576–9 (2006). <http://doi.org/10.1073/pnas.0600647103>.
- [107] Manam, R. R. *et al.* Leaving Groups Prolong the Duration of 20S Proteasome Inhibition and Enhance the Potency of Salinosporamides. *Journal of Medicinal Chemistry* 51, 6711–6724 (2008). <http://doi.org/10.1021/jm800548b>.
- [108] Lin, G. *et al.* Inhibitors selective for mycobacterial versus human proteasomes. *Nature* 461, 621–6 (2009). <http://doi.org/10.1038/nature08357>.
- [109] Kisselev, A. F. & Groettrup, M. Subunit specific inhibitors of proteasomes and their potential for immunomodulation. *Current Opinion in Chemical Biology* 23, 16–22 (2014). <http://doi.org/10.1016/j.cbpa.2014.08.012>.
- [110] Screen, M. *et al.* Nature of Pharmacophore Influences Active Site Specificity of Proteasome Inhibitors. *Journal of Biological Chemistry* 285, 40125–40134 (2010). <http://doi.org/10.1074/jbc.M110.160606>.
- [111] Geurink, P. P. *et al.* Incorporation of Non-natural Amino Acids Improves Cell Permeability and Potency of Specific Inhibitors of Proteasome Trypsin-like Sites. *Journal of Medicinal Chemistry* 56, 1262–1275 (2013). <http://doi.org/10.1021/jm3016987>.
- [112] de Bruin, G. *et al.* A Set of Activity-Based Probes to Visualize Human (Immuno)proteasome Activities. *Angewandte Chemie (International ed. in English)* 55, 4199–203 (2016). <http://doi.org/10.1002/anie.201509092>.
- [113] Groll, M., Koguchi, Y., Huber, R. & Kohno, J. Crystal structure of the 20 S proteasome:TMC-95A complex: a non-covalent proteasome inhibitor 1 Edited by I. A. Wilson. *Journal of Molecular Biology* 311, 543–548 (2001). <http://doi.org/10.1006/jmbi.2001.4869>.
- [114] Blackburn, C. *et al.* Optimization of a series of dipeptides with a P3 threonine residue as non-covalent inhibitors of the chymotrypsin-like activity of the human 20S proteasome. *Bioorganic and Medicinal Chemistry Letters* 20, 6581–6586 (2010). <http://doi.org/10.1016/j.bmcl.2010.09.032>.
- [115] Kazi, A. *et al.* Discovery of a novel proteasome inhibitor selective for cancer cells over non-transformed cells. *Cell Cycle* 8, 1940–1951 (2009). <http://doi.org/10.4161/cc.8.12.8798>.
- [116] Besche, H. C. & Goldberg, A. L. Affinity purification of mammalian 26S proteasomes using an ubiquitin-like domain. *Methods in molecular biology (Clifton, N.J.)* 832, 423–32 (2012). [http://doi.org/10.1007/978-1-61779-474-2\\_29](http://doi.org/10.1007/978-1-61779-474-2_29).

- [117] Lim, H. S., Cai, D., Archer, C. T. & Kodadek, T. Periodate-triggered cross-linking reveals Sug2/Rpt4 as the molecular target of a peptoid inhibitor of the 19S proteasome regulatory particle. *Journal of the American Chemical Society* 129, 12936–12937 (2007). <http://doi.org/10.1021/ja075469+>.
- [118] Park, J. E., Miller, Z., Jun, Y., Lee, W. & Kim, K. B. Next-generation proteasome inhibitors for cancer therapy. *Translational Research* (2018). <http://doi.org/10.1016/j.trsl.2018.03.002>.
- [119] da Fonseca, P. C. & Morris, E. P. Cryo-EM reveals the conformation of a substrate analogue in the human 20S proteasome core. *Nature Communications* 6, 7573 (2015). <http://doi.org/10.1038/ncomms8573>.
- [120] Harshbarger, W., Miller, C., Diedrich, C. & Sacchettini, J. Crystal Structure of the Human 20S Proteasome in Complex with Carfilzomib. *Structure* 23, 418–424 (2015). <http://doi.org/10.1016/j.str.2014.11.017>.
- [121] Sambrook, J. & Russell, D. W. *Molecular Cloning - A Laboratory Manual* (CSHL Press, Cold Spring Harbour, New York, 2001), 3 edn.
- [122] Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta crystallographica. Section D, Biological crystallography* 67, 235–42 (2011). <http://doi.org/10.1107/S0907444910045749>.
- [123] Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* 66, 486–501 (2010). <http://doi.org/10.1107/S0907444910007493>.
- [124] Zhang, K. Gctf: Real-time CTF determination and correction. *Journal of Structural Biology* 193, 1–12 (2016). <http://doi.org/https://doi.org/10.1016/j.jsb.2015.11.003>.
- [125] van Heel, M. *et al.* *Four-dimensional cryo-electron microscopy at quasi-atomic resolution: IMAGIC 4D*, chap. 19.9, 624–628 (American Cancer Society, 2012). <http://doi.org/10.1107/97809553602060000875>.
- [126] Lebedev, A. A. *et al.* JLigand : a graphical tool for the CCP 4 template-restraint library. *Acta Crystallographica Section D Biological Crystallography* 68, 431–440 (2012). <http://doi.org/10.1107/S090744491200251X>.
- [127] Neese, F. The ORCA program system. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 2, 73–78 (2012). <http://doi.org/10.1002/wcms.81>.
- [128] Adams, P. D. *et al.* The Phenix software for automated determination of macromolecular structures. *Methods* 55, 94–106 (2011). <http://doi.org/https://doi.org/10.1016/j.ymeth.2011.07.005>.
- [129] Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography* 67, 355–367 (2011). <http://doi.org/10.1107/S0907444911001314>.
- [130] Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology* 180, 519–530 (2012). <http://doi.org/https://doi.org/10.1016/j.jsb.2012.09.006>.
- [131] Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nature methods* 11, 63–5 (2014). <http://doi.org/10.1038/nmeth.2727>.

- [132] Kim, D. E., Chivian, D. & Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research* 32, 526–531 (2004). <http://doi.org/10.1093/nar/gkh468>.
- [133] Elmlund, H., Elmlund, D. & Bengio, S. PRIME: Probabilistic Initial 3D Model Generation for Single-Particle Cryo-Electron Microscopy. *Structure* 21, 1299–1306 (2013). <http://doi.org/https://doi.org/10.1016/j.str.2013.07.002>.
- [134] Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry* 25, 1605–12 (2004). <http://doi.org/10.1002/jcc.20084>.
- [135] Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* 4, e06980 (2015). <http://doi.org/10.7554/eLife.06980>.
- [136] Kabsch, W. XDS. *Acta Crystallographica Section D Biological Crystallography* 66, 125–132 (2010). <http://doi.org/10.1107/S0907444909047337>.
- [137] Chari, A. *et al.* ProteoPlex: stability optimization of macromolecular complexes by sparse-matrix screening of chemical space. *Nature methods* 12, 859–65 (2015). <http://doi.org/10.1038/nmeth.3493>.
- [138] Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Research* 11, 1475–1489 (1983). <http://doi.org/10.1093/nar/11.5.1475>.
- [139] Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680–5 (1970). <http://doi.org/10.1038/227680a0>.
- [140] Bonte, E. & Becker, P. B. Preparation of Chromatin Assembly Extracts from Preblastoderm Drosophila Embryos. In Chellappan, S. P. (ed.) *Chromatin Protocols*, vol. 523 of *Methods in Molecular Biology*, chap. 1, 1–10 (Humana Press, Totowa, NJ, 2009). [http://doi.org/10.1007/978-1-59745-190-1\\_1](http://doi.org/10.1007/978-1-59745-190-1_1).
- [141] Tan, S., Kern, R. C. & Selleck, W. The pST44 polycistronic expression system for producing protein complexes in Escherichia coli. *Protein Expression and Purification* 40, 385–395 (2005). <http://doi.org/10.1016/j.pep.2004.12.002>.
- [142] Swinehart, D. F. The Beer-Lambert Law. *Journal of Chemical Education* 39, 333 (1962). <http://doi.org/10.1021/ed039p333>.
- [143] Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics* 98, 5648–5652 (1993). <http://doi.org/10.1063/1.464913>.
- [144] Grimme, S., Ehrlich, S. & Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *Journal of Computational Chemistry* 32, 1456–1465 (2011). <http://doi.org/10.1002/jcc.21759>.
- [145] Lee, C., Yang, W. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B* 37, 785–789 (1988). <http://doi.org/10.1103/PhysRevB.37.785>.
- [146] Weigend, F. & Ahlrichs, R. {B}alanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for {H} to {R}n: {D}esign and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* 7, 3297–3305 (2005). <http://doi.org/10.1039/b508541a>.

- 
- [147] Johnson, E. R. & Becke, A. D. A post-Hartree-Fock model of intermolecular interactions: Inclusion of higher-order corrections. *The Journal of Chemical Physics* 124, 174104 (2006). <http://doi.org/10.1063/1.2190220>.
- [148] Klamt, A. & Schüürmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2* 799–805 (1993). <http://doi.org/10.1039/P29930000799>.
- [149] Blomberg, M. R. A., Borowski, T., Himo, F., Liao, R.-Z. & Siegbahn, P. E. M. Quantum Chemical Studies of Mechanisms for Metalloenzymes. *Chemical Reviews* 114, 3601–3658 (2014). <http://doi.org/10.1021/cr400388t>.
- [150] Stein, R. L., Melandri, F. & Dick, L. Kinetic Characterization of the Chymotryptic Activity of the 20S Proteasome. *Biochemistry* 35, 3899–3908 (1996). <http://doi.org/10.1021/bi952262x>.
- [151] Li, W. *et al.* The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic acids research* 43, W580–4 (2015). <http://doi.org/10.1093/nar/gkv279>.
- [152] Rupp, B. & Kantardjieff, K. *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology* (Garland Science, 2010).
- [153] Grover, A. *Drug Design: Principles and Applications* (Springer Singapore, Singapore, 2017). <http://doi.org/10.1007/978-981-10-5187-6>.
- [154] Fehrentz, J. A. *et al.* Solid Phase Synthesis of C-Terminal Peptide Aldehydes. *Journal of Organic Chemistry* 62, 6792–6796 (1997). <http://doi.org/10.1021/jo962408d>.
- [155] Stein, M. L. *et al.* Systematic Comparison of Peptidic Proteasome Inhibitors Highlights the  $\alpha$ -Ketoamide Electrophile as an Auspicious Reversible Lead Motif. *Angewandte Chemie International Edition* 53, 1679–1683 (2014). <http://doi.org/10.1002/anie.201308984>.
- [156] Garman, E. F. Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta crystallographica. Section D, Biological crystallography* 66, 339–51 (2010). <http://doi.org/10.1107/S0907444910008656>.
- [157] Leary, R., Larsen, D., Watanabe, H. & Shaw, E. Diazomethyl Ketone Substrate Derivatives as Active-Site-Directed Inhibitors of Thiol Proteases. Papain. *Biochemistry* 16, 5857–5861 (1977). <http://doi.org/10.1021/bi00645a033>.
- [158] Witkowska, J. *et al.* Crystal structure of a low molecular weight activator Blm-pep with yeast 20S proteasome – insights into the enzyme activation mechanism. *Scientific Reports* 7, 6177 (2017). <http://doi.org/10.1038/s41598-017-05997-4>.
- [159] Scopes, R. K. Overview of protein purification and characterization. *Current protocols in protein science* Chapter 1, Unit 1.1 (2001). <http://doi.org/10.1002/0471140864.ps0101s00>.
- [160] Sari, D. *et al.* The MultiBac Baculovirus/Insect Cell Expression Vector System for Producing Complex Protein Biologics. *Advances in experimental medicine and biology* 896, 199–215 (2016). [http://doi.org/10.1007/978-3-319-27216-0\\_13](http://doi.org/10.1007/978-3-319-27216-0_13).
- [161] Atha, D. H. & Ingham, K. C. Mechanism of precipitation of proteins by polyethylene glycols. Analysis in terms of excluded volume. *The Journal of biological chemistry* 256, 12108–17 (1981).

- [162] Sim, S.-L. *et al.* Protein precipitation by polyethylene glycol: a generalized model based on hydrodynamic radius. *Journal of biotechnology* 157, 315–9 (2012). <http://doi.org/10.1016/j.jbiotec.2011.09.028>.
- [163] Kantardjieff, K. A. & Rupp, B. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Science* 12, 1865–1871 (2003). <http://doi.org/10.1110/ps.0350503>.
- [164] Holton, J. M. & Frankel, K. A. The minimum crystal size needed for a complete diffraction data set. *Acta Crystallographica Section D Biological Crystallography* 66, 393–408 (2010). <http://doi.org/10.1107/s0907444910007262>.
- [165] Karplus, P. A. & Diederichs, K. Linking Crystallographic Model and Data Quality. *Science* 336, 1030–1033 (2012). <http://doi.org/10.1126/science.1218231>.
- [166] Casanas, A. *et al.* EIGER detector: application in macromolecular crystallography. *Acta Crystallographica Section D Structural Biology* 72, 1036–1048 (2016). <http://doi.org/10.1107/S2059798316012304>.
- [167] Baldwin, J. E. Rules for ring closure. *Journal of the Chemical Society, Chemical Communications* 734 (1976). <http://doi.org/10.1039/c39760000734>.
- [168] Tadros, W. & Lipshitz, H. D. The maternal-to-zygotic transition: a play in two acts. *Development* 136, 3033–3042 (2009). <http://doi.org/10.1242/dev.033183>.