



Technische Universität München
Fakultät für Chemie | Lehrstuhl für Organische Chemie II

Chemical protein manipulations and their application in proteomics data mining

MATTHIAS STAHL

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

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende: Prof. Dr. Kathrin Lang
Prüfende der Dissertation: 1. Prof. Dr. Stephan A. Sieber
2. Prof. Dr. Matthias Feige
3. Prof. Dr. Steven Verhelst

Die Dissertation wurde am 18.01.2018 bei der Technischen Universität München eingereicht und
durch die Fakultät für Chemie am 19.02.2018 angenommen.



Technische Universität München
Fakultät für Chemie | Lehrstuhl für Organische Chemie II

Chemische Proteinmanipulationen und ihr Einsatz in der proteomischen Datenanalyse

MATTHIAS STAHL

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

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende: Prof. Dr. Kathrin Lang
Prüfende der Dissertation: 1. Prof. Dr. Stephan A. Sieber
2. Prof. Dr. Matthias Feige
3. Prof. Dr. Steven Verhelst

Die Dissertation wurde am 18.01.2018 bei der Technischen Universität München eingereicht und
durch die Fakultät für Chemie am 19.02.2018 angenommen.

FÜR MEINE FAMILIE.

ABSTRACT

The manipulation of protein structure by chemical compounds is a common principle in chemical biology for studying protein functions. Here, a small molecule activator of the human caseinolytic serine protease ClpP was identified and employed to unravel conformational switches that underlie its activation mode. To date, the functional role of this mitochondrial protein is largely unknown, making it an important topic of study. It is hypothesized that upon ClpP activation, the peptidic cleavage products might play an essential role in the unfolded protein response of mitochondria.

In order to clarify the nature of these product peptides, hClpP was probed for its ability to recognize specific amino acids when digesting substrates. Amongst others, whole proteins were subjected to *in vitro* degradation. Mass spectrometry-based identification of the cleavage products eventually revealed a lower cleavage specificity than for peptidic substrates and helped to shape our understanding of how such product peptide sequences look like. After establishing this method, it was further used to analyze the cleavage patterns of bacterial ClpPs from *Escherichia coli*, *Listeria monocytogenes* and *Staphylococcus aureus*.

The latter bacterial strain is known to provoke human diseases such as the toxic-shock syndrome or sepsis. With the help of small molecule inhibitors of ClpP and mass spectrometry-based proteomics, the effect of protease perturbation on virulence shut-down could be demonstrated on a systems biology level.

However, using chemical biology methods in global chemical proteomics is often a challenge due to massive background effects. For example, small molecules that are equipped with an alkyne tag and a photocrosslinker enable the identification of non-covalent protein binders, but are known to bind proteins non-specifically. Here, a set of diazirine-bearing probes was dissected in terms of non-specific protein background binding. As a result, a comprehensive negative list of common false positives of such analyses could be provided. While mass spectrometry has developed into a sophisticated method for chemical biology, mining of the emerging datasets will be the next challenge to overcome. So far, proteomics data has widely been analyzed on the protein level. To additionally access information at the peptide level, a software called ProteomeDiver was designed. It permits easy visualization of complex proteomics experiments and underlying data, and provides analytical functions, e.g. for the detection of uncorrelated peptides, which might indicate biologically important anomalies. Thus, in future, this software will help researchers to make more sense of their overall chemical proteomics data.

ZUSAMMENFASSUNG

Die Manipulation von Proteinstrukturen mit chemischen Verbindungen ist ein gängiges Prinzip der Chemischen Biologie um Proteinfunktionen zu studieren. Hier wurde ein Aktivator der humanen caseinolytischen Serinprotease ClpP identifiziert und eingesetzt, um Aktivierungs-bedingte konformationelle Änderungen am Protein aufzudecken. Dies ist wichtig, da die funktionelle Rolle dieses mitochondrialen Proteins bislang größtenteils im Dunkeln lag. Es wird jedoch vermutet, dass die Produktpeptide aus der aktiven Protease bei der Kontrolle über die mitochondriale Stressantwort essentiell sind.

Um den Charakter solcher Peptide näher zu bestimmen, wurde hClpP einer genauen Analyse hinsichtlich ihrer Schnittmuster unterzogen. Unter anderem wurden komplette Proteine zum *in vitro*-Verdau angeboten. Massenspektrometrie ermöglichte schließlich deren Identifizierung und zeigte eine deutlich abgeschwächte Schnittstellenspezifität im Vergleich zu Peptiden als Substrate. Nach Etablierung der Methode konnten ähnliche Ergebnisse für die ClpPs aus *Escherichia coli*, *Listeria monocytogenes* und *Staphylococcus aureus* erzielt werden.

Letztere Bakterien sind verantwortlich für den Ausbruch von Krankheiten wie dem Toxischen Schocksyndrom oder der Sepsis. Mit Hilfe von ClpP-Inhibitoren und Massenspektrometrie-basierter Proteomik konnte der Effekt der gestörten Protease auf die Herunterregulation von pathogenen Virulenzfaktoren demonstriert werden.

Oft ist es in der Chemischen Proteomik allerdings so, dass mit massiven Hintergrundeffekten zu kämpfen ist. Zum Beispiel ermöglichen kleine Moleküle, ausgestattet mit einem Alkin und einem Photocrosslinker, die Identifikation von nicht-kovalent bindenden Proteinen. Hier war es möglich, unspezifische Hintergrundbinder von Diazirinen zu detektieren. So konnte eine umfassende Negativliste mit gewöhnlichen Falschpositiv-Treffern aufgestellt werden.

Während die Massenspektrometrie zwischenzeitlich zu einer ausgefeilten Methode herangereift ist, wird die detaillierte Auswertung der daraus entstehenden Datensätze die nächste Herausforderung sein. Bisher wurden Proteomikdaten weitgehend auf der Proteinebene betrachtet. Um zusätzlich den dahinterliegenden Peptidlevel zu erreichen, wurde die Software ProteomeDiver entwickelt. Sie ermöglicht die einfache Visualisierung von komplexen Proteomikexperimenten und der dazugehörigen Daten. Zusätzlich sind Analysewerkzeuge enthalten, die es z.B. erleichtern nach nicht-korrelierten Peptiden zu suchen. Diese können Hinweise auf biologische Anomalien liefern. In Zukunft kann das Programm Wissenschaftlern dabei helfen, mehr Nutzen aus ihren chemischen Proteomikdaten zu ziehen.

Parts of this thesis have been published in peer-reviewed journals as listed below:

Research articles

Barrel-shaped ClpP proteases display attenuated cleavage specificities

Malte Gersch,[#] Matthias Stahl,[#] Marcin Poreba,[#] Maria Dahmen, Anna Dziedzic, Marcin Drag, Stephan A. Sieber

ACS Chem. Biol., 2016, 11(2), pp 389–399

Insights into ClpXP proteolysis: heterooligomerization and partial deactivation enhance chaperone affinity and substrate turnover in *Listeria monocytogenes*

Dóra Balogh,[#] Maria Dahmen,[#] Matthias Stahl, Marcin Poreba, Malte Gersch, Marcin Drag, Stephan A. Sieber

Chem. Sci., 2017, 8(2), pp 1592–1600

Quantitative map of β -lactone induced virulence regulation

Joanna Krysiak,[#] Matthias Stahl,[#] Jan Vomacka,[#] Christian Fetzer, Markus Lakemeyer, Anja Fux, Stephan A. Sieber

J. Proteome Res., 2017, 16(3), pp 1180–1192

A whole proteome inventory of background photocrosslinker binding

Philipp Kleiner,[#] Wolfgang Heydenreuter,[#] Matthias Stahl,[#] Vadim S. Korotkov, Stephan A. Sieber

Angew. Chem. Int. Ed., 2017, 56(5), pp 1396–1401

Review article

An amino acid domino effect orchestrates ClpP's conformational states

Matthias Stahl, Stephan A. Sieber

Curr. Opin. Chem. Biol., 2017, 40, pp 102–110

[#] these authors contributed equally

Research topics that are not covered in this thesis:

Involved persons and my individual contributions are given under respective project names.

A chemical proteomics strategy for studying pyridoxal phosphate-dependent enzymes

Project led by Annabelle Hoegl (Prof. Dr. Sieber, TU München)

Mining of chemical proteomics data to identify pyridoxal phosphate-dependent enzymes with the help of cofactor mimics.

Status: Manuscript submitted

Mass spectrometry-based identification of connexin-32 binding partners

Project led by João Coelho (Prof. Dr. Feige, TU München)

Measurement of mass spectrometry-based coimmunoprecipitations and data mining to identify connexin-32 interacting proteins in human cells.

Status: Manuscript in preparation

Software development for the identification of isotope-encoded biotin linker pairs in mass spectrometry

Project led by Dr. Hacker, TU München

Setup of a Python- and R-based data mining pipeline to identify and to quantify probe-labeled amino acids in activity-based protein profiling experiments using isotope-encoded biotin linker derivatives.

Status: Final programming

Parts of this thesis have been presented at conferences and workshops:

Doktorandenforum der Studienstiftung des deutschen Volkes

29th October – 1st November 2015, Bonn, Germany

oral presentation

Proteomics Bioinformatics Workshop at the European Bioinformatics Institute

6th – 11th December 2015, Hinxton, United Kingdom

poster presentation

Winter School of the European Bioinformatics Community

10th – 13th January 2017, Semmering, Austria

poster presentation

16th Human Proteome Organization World Congress

17th – 21st September 2017, Dublin, Ireland

poster presentation

16th Annual Swedish Proteomics Society Symposium

12th – 13th November 2017, Stockholm, Sweden

poster presentation

TALENT MEANS NOTHING, WHILE EXPERIENCE,
ACQUIRED IN HUMILITY AND WITH HARD WORK, MEANS EVERYTHING.

— Patrick Süskind

Contents

I	Scientific Background	I
o	Introduction and basic methodology	3
o.1	The single-protein level: ClpP	4
o.1.1	Structure	4
o.1.2	Function	6
o.2	The proteomics level: a global view on proteins	6
o.3	The manipulated proteomics level: chemical proteomics	8
o.3.1	Activity-based protein profiling	9
o.3.2	Affinity-based protein profiling	11
o.3.3	Mass spectrometry for proteomics	12
o.4	Retracing the proteomics journey: Contents of this thesis	13
o.5	References	15
II	Research	23
I	Selective activation of human caseinolytic protease P (ClpP)	25
I.1	Introduction	26
I.1.1	Structure of hClpP	26
I.1.2	Function of hClpP	27
I.1.3	Objective	29
I.2	Results and discussion	30
I.2.1	Discovery of D9	30
I.2.2	Structure-activity relationships	33
I.2.3	Mode of D9 action	35
I.2.4	Docking and mutational studies	36
I.2.5	X-ray crystal structure	40
I.3	Conclusion and outlook	43
I.4	Methods	45
I.4.1	Comment on compounds and their solubility	45
I.4.2	High-throughput screen reanalysis	45
I.4.3	Cloning, expression and purification of proteins	45

1.4.4	FITC-casein proteolysis assay for initial in-house compound screening	47
1.4.5	FITC-casein proteolysis assay for other purposes	48
1.4.6	Peptidolysis assay	48
1.4.7	ClpX-mediated GFP-SsrA degradation and peptidase assay	49
1.4.8	Analytical ultracentrifugation	49
1.4.9	Docking	50
1.4.10	Crystallography and X-ray analysis	51
1.4.11	Author contributions	52
1.5	References	52
2	Barrel-shaped ClpP proteases display attenuated cleavage specificities	61
3	Quantitative map of β-lactone-induced virulence regulation	77
4	A whole proteome inventory of background photocrosslinker binding	95
5	ProteomeDiver: Mining peptide intensity profiles	105
5.1	The next level of chemical proteomics	106
5.1.1	Mass spectrometry in chemical biology	106
5.1.2	From proteins back to peptides	106
5.1.3	Objective	108
5.2	Design of the ProteomeDiver software	109
5.2.1	General setup	109
5.2.2	Loading of MaxQuant result files	109
5.2.3	Data visualization and first steps	111
5.2.4	Peptide intensity profile analysis	113
5.3	Example: The quest for pyridoxal phosphate-dependent enzymes and co-factor binding sites	115
5.4	Conclusion and outlook	119
5.5	Author contributions	121
5.6	References	121
6	Research conclusion	125
III	Review	131
7	An amino acid domino effect orchestrates ClpP's conformational states	133

Part I

Scientific Background

0

Introduction and basic methodology

This thesis is a transcript of a journey from the examination of a single protein – the ClpP protease – zooming out to a global view on whole cellular proteomes and deep data mining thereof.

0.1 THE SINGLE-PROTEIN LEVEL: CLPP

Some time before mass spectrometry has been established to screen whole proteomes, a simple bacterial degradation machine was discovered in *Escherichia coli*. One of its components turned out to be the serine protease ClpP.^{1,2} The other part was later characterized as ClpA, ClpC or ClpX – ATP-dependent chaperones.^{3,4,5,6,7} After extensive research over the past decades, our picture of ClpP, its associated adaptors and common functions became much clearer.

The family of ClpPs ranges over a multitude of bacterial and eukaryotic species, but dark spots remain – especially when considering eukaryotic ClpPs.*⁸ Despite their eminent importance in many creatures, ClpPs consist of a simple and conserved setup.

0.1.1 STRUCTURE

In *Staphylococcus aureus*, for example, 14 identical ClpP monomers assemble to a hollow barrel with two axial entry pores (figure 1ab). This cylinder is composed of two heptamer rings that stack on top of each other. The monomers themselves bring their own active site with a S98-H123-D172 catalytic triad. The alignment of the charge-relay system is dependent on a surrounding network of amino acid side chains that links the active site e.g. to the apical area around the entry pores as well as to a region called the arginine sensor.^{9,10} This feature is one of the key regulators for ring-ring interactions. According to that, an R171 of one heptameric ring is able to contact D170 of the adjacent ring. This involves side chain switching, which impacts the functional alignment of the catalytic triad (figure 1c–e).¹⁰ Another outstanding characteristic of each monomer is the E-helix, which protrudes from the globular head domain. Upon the formation of ring-ring interactions, the heptamers intertwine their E-helices depending on the global complex conformation. To date,

*Here, bacterial ClpPs will be presented. The human homolog is introduced and discussed in chapter 1, which deals with the artificial activation of mitochondrial ClpP. The accompanying methods section additionally presents several procedures to probe ClpXP activity in general.

three different shades thereof have been reported: compressed, compact, and extended (figure 1a).^{9,10,11,12}

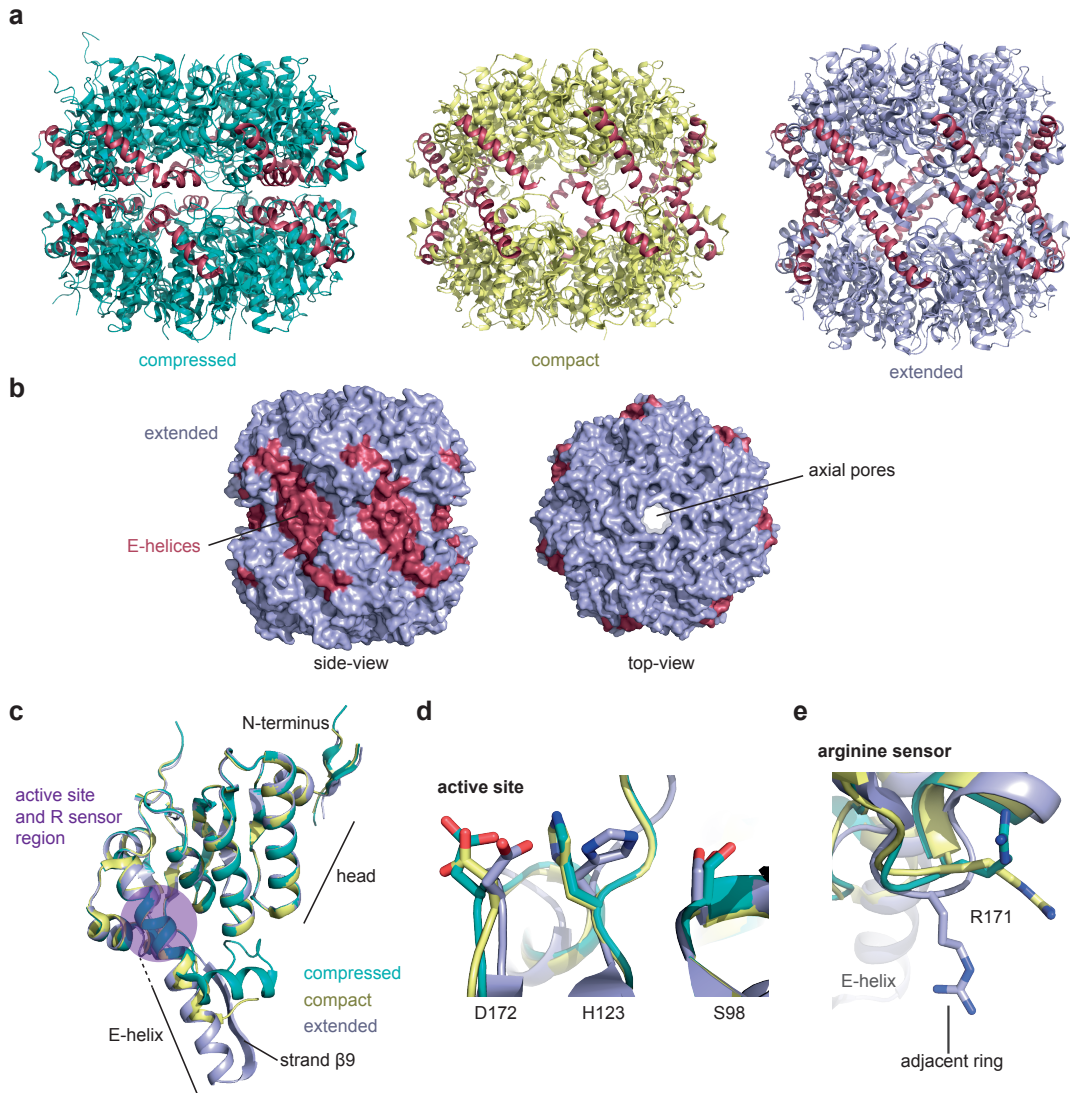


Figure 1: Different faces of ClpP from *S. aureus* (SaClpP). (a) Side-view on the compressed (PDB 3QWD), compact (PDB 4EMM) and extended (PDB 3V5E) conformations of SaClpP tetradecamers. The E-helices of each monomer are colored in red.^{9,11,10} (b) Side- and top-view of the extended conformation as surface representation. The top-view shows the axial entry pores. (c) Overlay of one monomer from each conformation of panel a. Major structural motifs are depicted: The N-terminus builds up a loop crown around the pores in the ring complex. The globular head domain houses the active site and distinct amino acid switches. The E-helix is important for ring-ring interactions in the tetradecameric complex. (d) Close-up view on the catalytic triad as overlay from all three conformations. H123 is flipped away from the catalytic serine in the compressed and compact states. (e) Example of different orientations of the arginine sensor that can communicate with the adjacent heptamer ring.

The degradation of substrates is enabled by ClpA, ClpC or ClpX. These chaperones dock to ClpP's pores and recognize, unfold and translocate proteins into the proteolytic chamber of ClpP.⁸ For ClpX, it is believed that it binds through IGF loops to hydrophobic clefts of ClpP. These pockets are located between monomers around the axial pores. A second interaction encompasses the protrusion of pore-2 loops to the N-terminal loop crown of ClpP's pores. Both modes are important for interprotein communication such as for the control of proteolytic activity of ClpP or ATP hydrolysis rates of ClpX.^{13,14,15,16,17} The digestion of substrates will be analyzed in detail for different ClpPs in chapter 2.

0.1.2 FUNCTION

The bacterial ClpP family is involved in essential cellular processes, most obviously in protein homeostasis. The complex structure is not only similar to the proteasome of eukaryotes, but it also fulfills the degradation of unneeded or unwanted proteins. Unlike the proteasome, ClpP's chaperones are able to recognize short terminal amino acid sequences or arginine phosphorylation.^{18,19,20} Furthermore, ClpP is also involved in virulence regulation, e.g. in pathogenic *S. aureus*. The secretion of alpha-hemolysin and other virulence factors is dependent on a convoluted network of signal transduction cascades. Hence, *clpP* knock-out mutants exhibit a disturbed virulence regulation. Amongst others, expression of genes from the *agr* system is reduced. This signal transduction pathway is responsible for the onset of virulence upon quorum sensing.^{21,22} However, a direct link between ClpP and the virulence regulator has been missing so far. Chapter 3 will provide some insights into ClpP's role in *S. aureus* virulence systems biology and illustrates the importance of widening the scope of research from a single isolated protein to its surrounding context in cellular systems.

0.2 THE PROTEOMICS LEVEL: A GLOBAL VIEW ON PROTEINS

In the past century, the genomics and proteomics community underwent several paradigm shifts that came along with radical advancements in the understanding of the proteome and the interplay of all proteins with each other.²³ One key development in the early days of protein research was the Edman degradation procedure.^{24,25} With this tool at hand, researchers were able to record the fingerprints of proteins: their amino acid sequence.

The simple sequence pattern of around 20 different amino acids builds the basis for life

and has been the central theme in all further breakthroughs until today. Initially, genes of single proteins could be cloned and overexpressed²⁶ and a plethora of biochemical assays allowed the broad analysis of protein characteristics, e.g. the determination of their mass by SDS-PAGE or enzymatic turnover.^{27,28} Finally, protein identity could be confirmed with western blots.²⁹

At the end of the 20th century, the situation for proteomics resembled the parable of the blind men and an elephant (figure 2).³⁰ Standing in front of the animal, one man touched the trunk and thought of a snake. Another man grasped the tusk and said it must be a spear. Just as researchers characterized single proteins, but the big picture of the proteome remained in the dark.

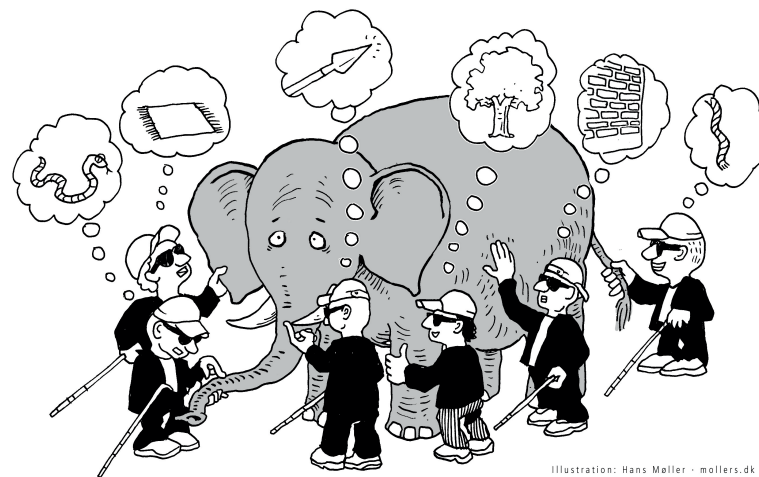


Figure 2: The proteome is an elephant. At least according to the parable of six blind men and an elephant. Originally reported from the Indian subcontinent.³⁰ Illustration reprinted with permission from Hans Møller.

In order to unravel the individual role of proteins and their communication with each other, a global and synchronous view on the full proteome is required.³¹ An important step towards this goal, was the development of 2D-gel electrophoresis – where complex protein mixtures are separated by pI and size – combined with protein identification techniques.³² The advent of soft ionization methods for mass spectrometry accompanied with ion fragmentation devices eventually leveraged high-throughput identification and quantification of complex protein mixtures.^{33,34} This achievement was followed by a myriad of methodological advancements, such as in-gel digestion of proteins^{35,36} or the multidimensional protein identification technology (MudPIT)³⁷ that uses sophisticated peptide separation steps.

Looking back on the developing field of proteomics, the focus first changed from biochemical methodology to mass spectrometry instrumentation. Nowadays, it is about to move on to data mining. The era of bioinformatics in proteomics was significantly pushed forward by SEQUEST, a cross correlation algorithm to detect amino acid sequences in peptide ion fragmentation patterns by similarity to a given peptide spectrum database.³⁸ SEQUEST and other search engines generally use scores to assess the quality of a peptide identification. Next, the nature of this measure called for statistical methods such as the target-decoy strategy to discriminate true from false positives.³⁹ It seems that the available proteomics techniques are now ready to explore the full picture of comprehensive protein networks. Through the global view of mass spectrometry-based proteomics, the six blind men not only learn to see again, but they will realize that everything they see, is interconnected and interdependent.^{40,41,42}

0.3 THE MANIPULATED PROTEOMICS LEVEL: CHEMICAL PROTEOMICS

Classical proteomics experiments are intended to examine the protein network of a cell, tissue, organ, or whole organism as it is. It is for example possible to quantify the expression status or – with the help of additional biochemistry – to assess the interactome of selected proteins.^{43,44,45} However, these are snapshots and the proteome can be viewed as a fine-tuned mobile. When applying chemical compounds to a proteomic system, the mobile changes its orientation to regain balance. This reciprocal action can again be quantified by mass spectrometry-based techniques and subsequent data mining. Results from such experiments permit insights into systems biology networks and their dynamics. An example of that is presented in chapter 3.⁴⁶

But chemical proteomics comprises more than just spoofing the steady state. Chemical biology interventions were applied throughout this thesis in many facets, often in combination with proteomics techniques. For example, two closely related methods, namely activity- and affinity-based protein profiling, were previously designed to uncover direct interactions between the proteins and chemical compounds.

0.3.1 ACTIVITY-BASED PROTEIN PROFILING

The information to build up a protein is stored in billions of DNA base pairs.⁴⁷ Static proteomics methods such as described above completely rely on this code. But when it comes e.g. to posttranslational modifications, it is more complicated. For instance, the activation state of an enzyme in a complex biological environment, i.e. a bacterial cell, is not predictable to date. At this point, chemical biology methods, such as activity-based protein profiling (ABPP), come into play.

The group of Powers pioneered the technique in the early 1990s. Biotinylated isocoumarins were used as small molecule probes to alkylate serine hydrolases.⁴⁸ The example already highlights the composition of activity-based probes: They contain (i) a reactive war head that can be trapped by active enzymes, (ii) a spacer, and (iii) a reporter group. In Powers' experiments, the biotin moiety represents the reporter, but it is also feasible to attach fluorophore tags.⁴⁹ With such probes at hand, the activity of large enzyme sets can be screened *in vitro* and in parallel. However, probes are often too bulky and electrostatically unable to cross biological membranes and are thus not favorable for *in vivo* use.

In order to tackle this problem, Cravatt and coworkers made use of Huisgen's bioorthogonal 1,3-dipolar azide-alkyne cycloaddition.⁵⁰ With the Cu(I)-catalyzed analog of this reaction from the Sharpless laboratory,⁵¹ they were able to set up a concept of reporter-free probes (figure 3).⁵² A small molecule probe containing a terminal alkyne for bioorthogonal chemistry is incubated with living cells. After the probe has bound to its target proteins, cells are lysed and a biotin-azide conjugate is coupled with Sharpless' click chemistry to the probe-labeled enzymes. Next, the labeled subproteome is captured by avidin-coated beads for enrichment. The isolated proteins are then submitted to bottom-up mass spectrometry, i.e. they are digested with an appropriate enzyme such as trypsin and subsequently identified as well as quantified with a mass spectrometer (see also figure 5).⁵³

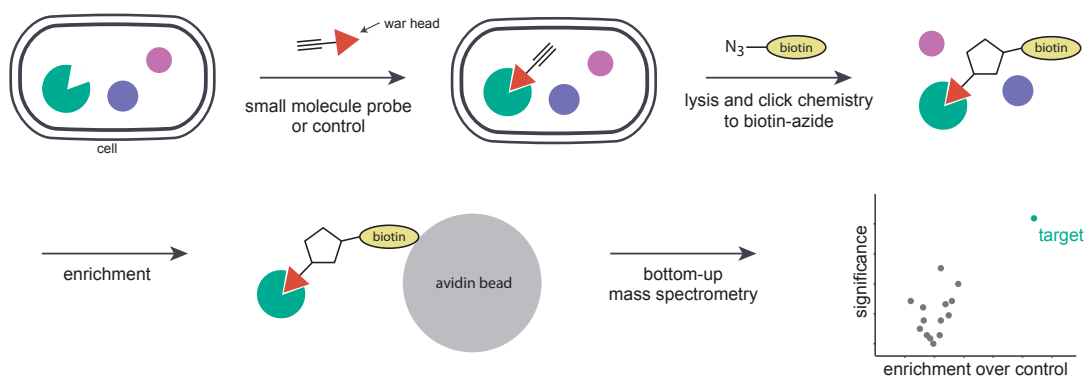


Figure 3: Principle of activity-based protein profiling (ABPP) in living cells. A small molecule probe bearing a reactive war head and an alkyne tag is incubated together with living cells. The probe labels appropriate enzymes via a stable covalent bond to their active sites. Cells are lysed and the alkyne tag of the probe is coupled to biotin-azide via a copper-catalyzed cycloaddition. The biotin tag allows for the enrichment of labeled proteins on immobilized avidin. The isolated binders are submitted to bottom-up mass spectrometry procedures for identification and quantification. If the experiment was designed with an appropriate number of replicates, results can be visualized as volcano plots.⁵³

The basic principle of ABPP can be expanded by sophisticated variations at each step (examples are referenced). For instance, the application of inactive probe derivatives helps discriminating between true and false protein targets.⁵⁴ Furthermore, using a competitor without alkyne tag gains a similar effect, but this method can also be applied to screen unmodified enzyme inhibitors for their binding potential.⁵⁵ With the introduction of cleavable linkers between biotin and azide moiety, it becomes possible to remove the probe from the beads and therefore to identify the probe binding site (tandem orthogonal proteolysis ABPP, TOP-ABPP).⁵⁶ Lastly, proteins, peptides, probe, or affinity tag can be labeled with stable isotopes such as ^{15}N , ^{13}C , or ^2H . For example, in isoTOP-ABPP (isotopic TOP-ABPP), the cleavable linker of the biotin-azide is either labeled heavy or light. This permits an early mixture of samples after click chemistry so that probe and control can be measured in one mass spectrometry run.⁵⁷

After all, ABPP requires reactive sites in the proteome that attack the probes. If only an interaction has to be shown that relies on non-covalent linkages, other methods have to be taken into account. In this context, one derivative of ABPP is widely applied: Affinity-based protein profiling.

0.3.2 AFFINITY-BASED PROTEIN PROFILING

In contrast to ABPP, affinity-based protein profiling (AfBPP) is characterized by small molecule probes that usually do not possess a reactive war head, but a photocrosslinker moiety instead. The aim is to demonstrate compound-protein interaction irrespective of reactive sites. Thus, the method is often called photoaffinity labeling.^{58,59} Mainly three different photocrosslinker classes have been spread in the scientific community: aryl azides, benzophenones, and diazirines (figure 4a).⁶⁰ Aryl azides and diazirines are preferred due to their superior target specificity.⁶¹

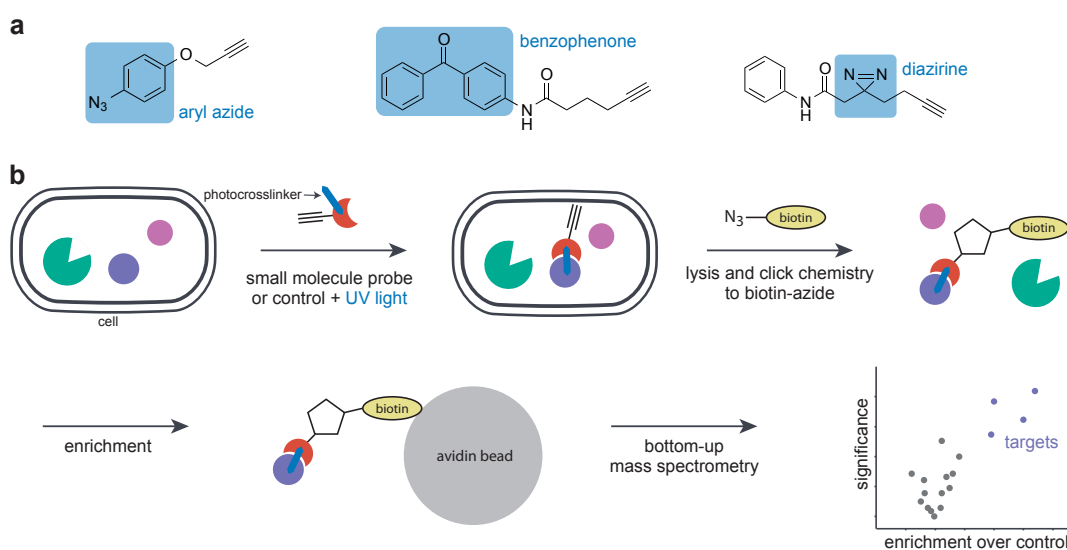


Figure 4: Photocrosslinker moieties and principle of affinity-based protein profiling (AfBPP) in living cells. The workflow is similar to ABPP (figure 3), but probes do not bear a reactive war head. Instead, they exhibit photocrosslinker groups (a). The probe is then also incubated with living cells (b). Before lysis, the whole system is irradiated with UV light to activate the radical mechanism of crosslinking. Thus, probes are getting irreversibly attached to their target protein. The subsequent procedure is the same as for ABPP.^{59,61}

A standard AfBPP experiment largely follows the workflow of an ABPP experiment as described in figure 3. The major difference consists in the composition of the photocrosslinker probe. The compound attaches to proteins via non-covalent forces, e.g. hydrophobic or electrostatic interactions. Irradiation of the sample leads to activation of the photocrosslinker part and thus the formation of a covalent bond. From there on, the sample can equally be treated as known for ABPP.⁶¹

Regarding its results, AfBPP is often accompanied with lower specificity and therefore

then leads to a larger set of identified potential targets. This disadvantage most likely traces back to both relatively high dissociation constants of the probe-target interaction and the crosslinking to neighboring proteins during irradiation.^{59,62,63} The obstacle is faced by sophisticated controls such as competitors – mostly the unmodified compound – or the use of so called minimal photocrosslinker probes.^{64,65} A bioinformatics answer to this problem is presented in chapter 4.

0.3.3 MASS SPECTROMETRY FOR PROTEOMICS

Captured proteins, whether collected by ABPP, AfBPP or other methods, are in general identified and quantified by mass spectrometry. In brief, proteins have to be digested and resulting peptides are then desalted and separated with C₁₈-based columns through high-pressure liquid chromatography (HPLC). This step aims at a reduction of sample complexity before the peptides are applied via electrospray ionization to a mass spectrometer. In this study, exclusively orbitrap-based machines have been used. After determining the full peptide m/z values (MS¹ spectra), most intense peptide ions are subjected to a collision cell, where they are fragmented (data-dependent acquisition, DDA). The fragment ions are again measured for their m/z ratio in a linear ion trap or again in the orbitrap (MS² spectra).^{40,42,66} The fragmentation pattern is dependent on the sequence of amino acids. Thus, peptides are usually identified by matching their MS² patterns to theoretical fragment patterns from a database.³⁸ Finally, in this work, quantification was mainly based on full peptide ion intensities in MS¹ spectra, either in SILAC or dimethyl labeling settings or label-free. The latter method is schematically depicted in figure 5 and chapter 5 exploits its power to quantify peptides from many experiments in parallel.^{67,68,69,70}

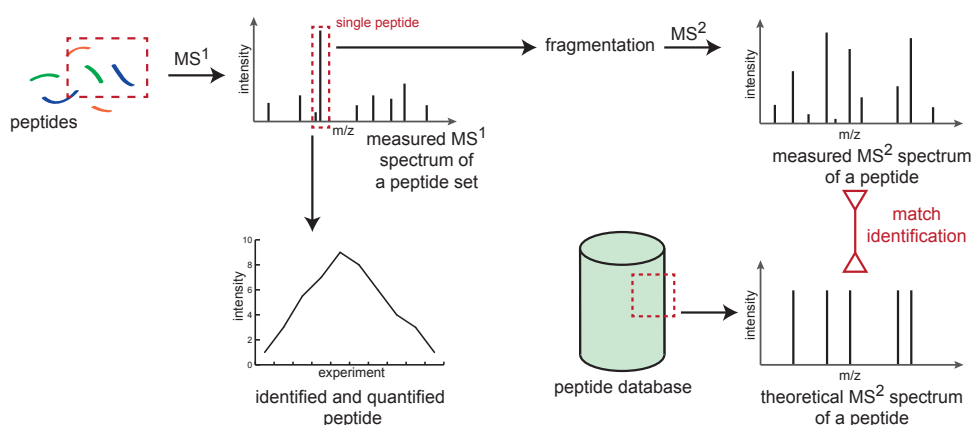


Figure 5: Methodological basis for bioinformatics to identify and quantify peptides based on MS¹ and MS² spectra in data-dependent acquisition. For identification, a peptide ion from a MS¹ spectrum is isolated and subsequently fragmented. The resulting MS² fragment spectrum is then searched against a database of *in silico* digested proteins. For quantification, peptide ion intensities from MS¹ spectra are used (here: label-free quantification).^{38,70}

0.4 RETRACING THE PROTEOMICS JOURNEY: CONTENTS OF THIS THESIS

The structure of this thesis reflects my journey through chemical biology and chemical proteomics. However, not completely in chronologic order, it starts with the conformational regulation of a single protein – ClpP –, then it continues with the dissection of ClpP’s substrate cleavage preferences. The direction then changes towards the cell biological context of ClpP. Finally, the journey temporarily ends with pure data mining of complex proteomics experiments such as photoaffinity labeling and peptide intensity profiles in mass spectrometry (figure 6).

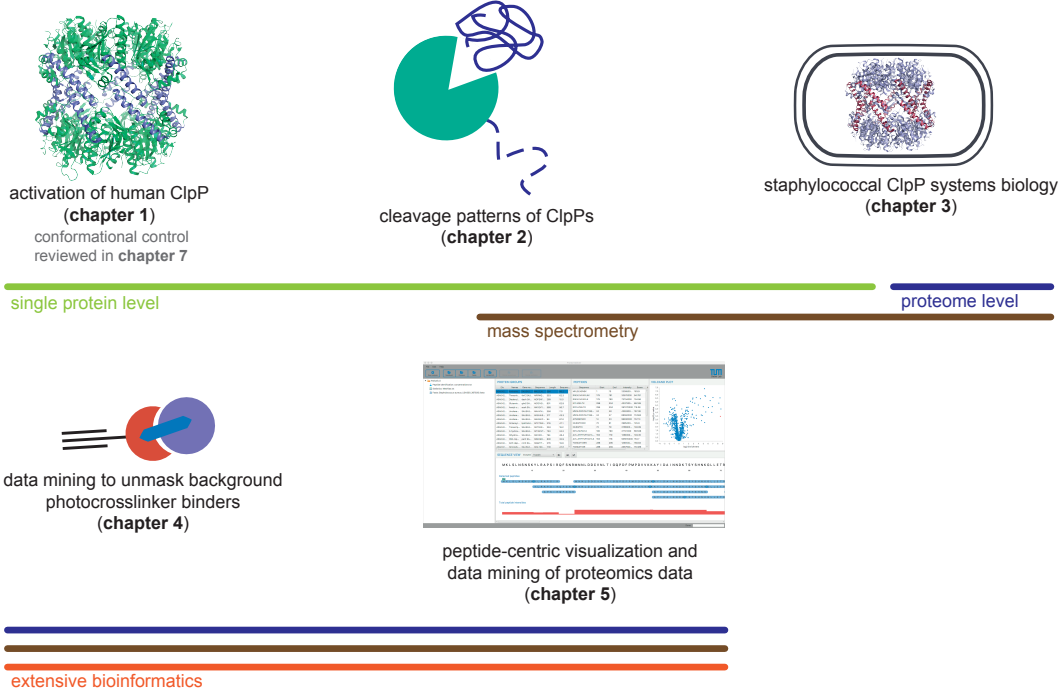


Figure 6: Schematic illustration of the contents of this thesis. The chapters are sorted to follow the increase of proteomic complexity. Furthermore, the methods used undergo a shift from classical 'wet-lab' biochemistry to 'dry-lab' bioinformatics. Both domains are linked by mass spectrometry-based proteomics.

0.5 REFERENCES

- [1] Y Katayama-Fujimura, S Gottesman, and MR Maurizi. A multiple-component, ATP-dependent protease from *Escherichia coli*. *Journal of Biological Chemistry*, 262(10):4477–4485, 1987.
- [2] Byung Joon Hwang, Woo Jin Park, Chin Ha Chung, and Alfred L Goldberg. *Escherichia coli* contains a soluble ATP-dependent protease (Ti) distinct from protease La. *Proceedings of the National Academy of Sciences*, 84(16):5550–5554, 1987.
- [3] Yoko Katayama, S Gottesman, J Pumphrey, S Rudikoff, William P Clark, and MR Maurizi. The two-component, ATP-dependent Clp protease of *Escherichia coli*. Purification, cloning, and mutational analysis of the ATP-binding component. *Journal of Biological Chemistry*, 263(29):15226–15236, 1988.
- [4] Byung Joon Hwang, Kee Min Woo, AL Goldberg, and Chin Ha Chung. Protease Ti, a new ATP-dependent protease in *Escherichia coli*, contains protein-activated ATPase and proteolytic functions in distinct subunits. *Journal of Biological Chemistry*, 263(18):8727–8734, 1988.
- [5] Diana Wojtkowiak, Costa Georgopoulos, and Maciej Zylicz. Isolation and characterization of ClpX, a new ATP-dependent specificity component of the Clp protease of *Escherichia coli*. *Journal of Biological Chemistry*, 268(30):22609–22617, 1993.
- [6] Susan Gottesman, William P Clark, Valerie de Crecy-Lagard, and Michael R Maurizi. ClpX, an alternative subunit for the ATP-dependent Clp protease of *Escherichia coli*. Sequence and in vivo activities. *Journal of Biological Chemistry*, 268(30):22618–22626, 1993.
- [7] Dorte Frees, Ulf Gerth, and Hanne Ingmer. Clp chaperones and proteases are central in stress survival, virulence and antibiotic resistance of *Staphylococcus aureus*. *International Journal of Medical Microbiology*, 304(2):142–149, 2014.
- [8] Angela Yeou Hsiung Yu and Walid A Houry. ClpP: a distinctive family of cylindrical energy-dependent serine proteases. *FEBS Letters*, 581(19):3749–3757, 2007.

- [9] Sebastian R Geiger, Thomas Böttcher, Stephan A Sieber, and Patrick Cramer. A conformational switch underlies ClpP protease function. *Angewandte Chemie International Edition*, 50(25):5749–5752, 2011.
- [10] Malte Gersch, Anja List, Michael Groll, and Stephan A Sieber. Insights into structural network responsible for oligomerization and activity of bacterial virulence regulator caseinolytic protease P (ClpP) protein. *Journal of Biological Chemistry*, 287(12):9484–9494, 2012.
- [11] Fei Ye, Jie Zhang, Hongchuan Liu, Rolf Hilgenfeld, Ruihan Zhang, Xiangqian Kong, Lianchun Li, Junyan Lu, Xinlei Zhang, Donghai Li, et al. Helix unfolding/refolding characterizes the functional dynamics of *Staphylococcus aureus* Clp protease. *Journal of Biological Chemistry*, 288(24):17643–17653, 2013.
- [12] Jie Zhang, Fei Ye, Lefu Lan, Hualiang Jiang, Cheng Luo, and Cai-Guang Yang. Structural Switching of *Staphylococcus aureus* Clp Protease a key to understanding protease dynamics. *Journal of Biological chemistry*, 286(43):37590–37601, 2011.
- [13] Andreas Martin, Tania A Baker, and Robert T Sauer. Distinct static and dynamic interactions control ATPase-peptidase communication in a AAA+ protease. *Molecular Cell*, 27(1):41–52, 2007.
- [14] Yong-In Kim, Igor Levchenko, Karolina Fraczkowska, Rachel V Woodruff, Robert T Sauer, and Tania A Baker. Molecular determinants of complex formation between Clp/Hsp100 ATPases and the ClpP peptidase. *Nature Structural & Molecular Biology*, 8(3):230–233, 2001.
- [15] Shilpa A Joshi, Greg L Hersch, Tania A Baker, and Robert T Sauer. Communication between ClpX and ClpP during substrate processing and degradation. *Nature Structural & Molecular Biology*, 11(5):404–411, 2004.
- [16] Satyendra K Singh, Jan Rozycki, Joaquin Ortega, Takashi Ishikawa, John Lo, Alasdair C Steven, and Michael R Maurizi. Functional domains of the ClpA and ClpX molecular chaperones identified by limited proteolysis and deletion analysis. *Journal of Biological Chemistry*, 276(31):29420–29429, 2001.
- [17] Matthias Stahl and Stephan A Sieber. An amino acid domino effect orchestrates ClpP’s conformational states. *Current Opinion in Chemical Biology*, 40:102–110, 2017.

- [18] Susan Gottesman, Eric Roche, YanNing Zhou, and Robert T Sauer. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes & Development*, 12(9):1338–1347, 1998.
- [19] Julia M Flynn, Saskia B Neher, Yong-In Kim, Robert T Sauer, and Tania A Baker. Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. *Molecular Cell*, 11(3):671–683, 2003.
- [20] Débora Broch Trentini, Marcin Józef Suskiewicz, Alexander Heuck, Robert Kurzbauer, Luiza Deszcz, Karl Mechtler, and Tim Clausen. Arginine phosphorylation marks proteins for degradation by a Clp protease. *Nature*, 539(7627):48–53, 2016.
- [21] Antje Michel, Franziska Agerer, Christof R Hauck, Mathias Herrmann, Joachim Ullrich, Jörg Hacker, and Knut Ohlsen. Global regulatory impact of ClpP protease of *Staphylococcus aureus* on regulons involved in virulence, oxidative stress response, autolysis, and DNA repair. *Journal of Bacteriology*, 188(16):5783–5796, 2006.
- [22] Dorte Frees, Karen Sørensen, and Hanne Ingmer. Global virulence regulation in *Staphylococcus aureus*: pinpointing the roles of ClpP and ClpX in the *sar/agr* regulatory network. *Infection and Immunity*, 73(12):8100–8108, 2005.
- [23] Lee Hood. A personal journey of discovery: developing technology and changing biology. *Annual Review of Analytical Chemistry*, 1:1–43, 2008.
- [24] Pehr Edman. Method for determination of the amino acid sequence in peptides. *Acta Chemica Scandinavica*, 4(7):283–293, 1950.
- [25] Pehr Edman and Geoffrey Begg. A protein sequenator. *The FEBS Journal*, 1(1):80–91, 1967.
- [26] Joseph Sambrook, Edward F Fritsch, Tom Maniatis, et al. *Molecular cloning: a laboratory manual*. Number Ed. 2. Cold spring harbor laboratory press, 1989.
- [27] Hermann Schägger and Gebhard Von Jagow. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Analytical Biochemistry*, 166(2):368–379, 1987.
- [28] John S Easterby. Coupled enzyme assays: a general expression for the transient. *Biochimica et Biophysica Acta (BBA)-Enzymology*, 293(2):552–558, 1973.

- [29] W Neal Burnette. “Western blotting”: electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. *Analytical Biochemistry*, 112(2):195–203, 1981.
- [30] E Bruce Goldstein. *Encyclopedia of perception*, volume 1. Sage, 2010.
- [31] Marc R Wilkins, Jean-Charles Sanchez, Andrew A Gooley, Ron D Appel, Ian Humphery-Smith, Denis F Hochstrasser, and Keith L Williams. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnology and Genetic Engineering Reviews*, 13(1):19–50, 1996.
- [32] Marc R Wilkins, Christian Pasquali, Ron D Appel, Keli Ou, Olivier Golaz, Jean-Charles Sanchez, Jun X Yan, Andrew A Gooley, Graham Hughes, Ian Humphery-Smith, et al. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Nature Biotechnology*, 14(1):61–65, 1996.
- [33] Matthias Mann, Ronald C Hendrickson, and Akhilesh Pandey. Analysis of proteins and proteomes by mass spectrometry. *Annual Review of Biochemistry*, 70(1):437–473, 2001.
- [34] Ioannis A Papayannopoulos. The interpretation of collision-induced dissociation tandem mass spectra of peptides. *Mass Spectrometry Reviews*, 14(1):49–73, 1995.
- [35] Matthias Wilm, Andrej Shevchenko, Tony Houthaeve, Stephen Breit, Lothar Schweigerer, Theodore Fotsis, and Matthias Mann. Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature*, 379(6564):466, 1996.
- [36] Anna Shevchenko, Alexander Loboda, Werner Ens, Burkhardt Schraven, Kenneth G Standing, and Andrej Shevchenko. Archived polyacrylamide gels as a resource for proteome characterization by mass spectrometry. *Electrophoresis*, 22:1194–1203, 2001.
- [37] Michael P Washburn, Dirk Wolters, and John R Yates. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology*, 19(3):242–247, 2001.

- [38] Jimmy K Eng, Ashley L McCormack, and John R Yates. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry*, 5(11):976–989, 1994.
- [39] Joshua E Elias and Steven P Gygi. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods*, 4(3):207–214, 2007.
- [40] Ruedi Aebersold and Matthias Mann. Mass spectrometry-based proteomics. *Nature*, 422(6928):198–207, 2003.
- [41] Bruno Domon and Ruedi Aebersold. Mass spectrometry and protein analysis. *Science*, 312(5771):212–217, 2006.
- [42] Ruedi Aebersold and Matthias Mann. Mass-spectrometric exploration of proteome structure and function. *Nature*, 537(7620):347–355, 2016.
- [43] Mathias Wilhelm, Judith Schlegl, Hannes Hahne, Amin Moghaddas Gholami, Marcus Lieberenz, Mikhail M Savitski, Emanuel Ziegler, Lars Butzmann, Siegfried Gessulat, Harald Marx, et al. Mass-spectrometry-based draft of the human proteome. *Nature*, 509(7502):582, 2014.
- [44] Min-Sik Kim, Sneha M Pinto, Derese Getnet, Raja Sekhar Nirujogi, Srikanth S Manda, Raghothama Chaerkady, Anil K Madugundu, Dhanashree S Kelkar, Ruth Isserlin, Shobhit Jain, et al. A draft map of the human proteome. *Nature*, 509(7502):575–581, 2014.
- [45] Eva C Keilhauer, Marco Y Hein, and Matthias Mann. Accurate protein complex retrieval by affinity enrichment mass spectrometry (AE-MS) rather than affinity purification mass spectrometry (AP-MS). *Molecular & Cellular Proteomics*, 14(1):120–135, 2015.
- [46] Norberto Pepporine Lopes and Ricardo Roberto da Silva, editors. *Mass spectrometry in chemical biology*. Chemical Biology. The Royal Society of Chemistry, 2018.
- [47] Francis Crick. Central dogma of molecular biology. *Nature*, 227(5258):561–563, 1970.

- [48] Chih Min Kam, Ahmed S Abuelyaman, Zhaozhao Li, Dorothy Hudig, and James C Powers. Biotinylated isocoumarins, new inhibitors and reagents for detection, localization, and isolation of serine proteases. *Bioconjugate Chemistry*, 4(6):560–567, 1993.
- [49] Stephan A Sieber and Benjamin F Cravatt. Analytical platforms for activity-based protein profiling—exploiting the versatility of chemistry for functional proteomics. *Chemical Communications*, (22):2311–2319, 2006.
- [50] Albert Padwa. 1, 3-Dipolar cycloaddition chemistry. *New York*, pages 1–176, 1984.
- [51] Vsevolod V Rostovtsev, Luke G Green, Valery V Fokin, and K Barry Sharpless. A stepwise Huisgen cycloaddition process: copper (I)-catalyzed regioselective “ligation” of azides and terminal alkynes. *Angewandte Chemie*, 114(14):2708–2711, 2002.
- [52] Anna E Speers, Gregory C Adam, and Benjamin F Cravatt. Activity-based protein profiling in vivo using a copper (I)-catalyzed azide-alkyne [3+2] cycloaddition. *Journal of the American Chemical Society*, 125(16):4686–4687, 2003.
- [53] Benjamin F Cravatt, Aaron T Wright, and John W Kozarich. Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. *Annu. Rev. Biochem.*, 77:383–414, 2008.
- [54] Franziska A Mandl, Volker C Kirsch, Ilke Ugur, Elena Kunold, Jan Vomacka, Christian Fetzer, Sabine Schneider, Klaus Richter, Thilo M Fuchs, Iris Antes, et al. Natural-product-inspired aminoepoxybenzoquinones kill members of the gram-negative pathogen *Salmonella* by attenuating cellular stress response. *Angewandte Chemie International Edition*, 55(47):14852–14857, 2016.
- [55] Gerjan de Bruin, Bo Tao Xin, Marianne Kraus, Mario van der Stelt, Gijsbert A van der Marel, Alexei F Kisselev, Christoph Driessen, Bogdan I Florea, and Herman S Overkleeft. A set of activity-based probes to visualize human (immuno) proteasome activities. *Angewandte Chemie International Edition*, 55(13):4199–4203, 2016.
- [56] Anna E Speers and Benjamin F Cravatt. A tandem orthogonal proteolysis strategy for high-content chemical proteomics. *Journal of the American Chemical Society*, 127(28):10018–10019, 2005.

- [57] Eranthie Weerapana, Chu Wang, Gabriel M Simon, Florian Richter, Sagar Khare, Myles BD Dillon, Daniel A Bachovchin, Kerri Mowen, David Baker, and Benjamin F Cravatt. Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature*, 468(7325):790–795, 2010.
- [58] Paul P Geurink, Laurette M Prely, Gijs A van der Marel, Rainer Bischoff, and Herman S Overkleeft. Photoaffinity labeling in activity-based protein profiling. In *Activity-Based Protein Profiling*, pages 85–113. Springer, 2011.
- [59] Ewan Smith and Ian Collins. Photoaffinity labeling in target-and binding-site identification. *Future*, 7(2):159–183, 2015.
- [60] David J Lapinsky and Douglas S Johnson. Recent developments and applications of clickable photoprobes in medicinal chemistry and chemical biology. *Future Medicinal Chemistry*, 7(16):2143–2171, 2015.
- [61] Philipp Kleiner, Wolfgang Heydenreuter, Matthias Stahl, Vadim S Korotkov, and Stephan A Sieber. A whole proteome inventory of background photocrosslinker binding. *Angewandte Chemie International Edition*, 56(5):1396–1401, 2017.
- [62] Florence Kotzyba-Hibert, Isabelle Kapfer, and Maurice Goeldner. Recent trends in photoaffinity labeling. *Angewandte Chemie International Edition*, 34(12):1296–1312, 1995.
- [63] Jongmin Park, Minseob Koh, Ja Young Koo, Sanghee Lee, and Seung Bum Park. Investigation of specific binding proteins to photoaffinity linkers for efficient deconvolution of target protein. *ACS Chemical Biology*, 11(1):44–52, 2015.
- [64] Andrew L MacKinnon and Jack Taunton. Target identification by diazirine photo-cross-linking and click chemistry. *Current Protocols in Chemical Biology*, pages 55–73, 2009.
- [65] Zhengqiu Li, Piliang Hao, Lin Li, Chelsea YJ Tan, Xiamin Cheng, Grace YJ Chen, Siu Kwan Sze, Han-Ming Shen, and Shao Q Yao. Design and synthesis of minimalist terminal alkyne-containing diazirine photo-crosslinkers and their incorporation into kinase inhibitors for cell-and tissue-based proteome profiling. *Angewandte Chemie International Edition*, 52(33):8551–8556, 2013.

- [66] Annette Michalski, Eugen Damoc, Jan-Peter Hauschild, Oliver Lange, Andreas Wiegand, Alexander Makarov, Nagarjuna Nagaraj, Juergen Cox, Matthias Mann, and Stevan Horning. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Molecular & Cellular Proteomics*, 10(9):M111–O11015, 2011.
- [67] Shao-En Ong, Blagoy Blagoev, Irina Kratchmarova, Dan Bach Kristensen, Hanno Steen, Akhilesh Pandey, and Matthias Mann. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & Cellular proteomics*, 1(5):376–386, 2002.
- [68] Jue-Liang Hsu, Sheng-Yu Huang, Nan-Haw Chow, and Shu-Hui Chen. Stable-isotope dimethyl labeling for quantitative proteomics. *Analytical Chemistry*, 75(24):6843–6852, 2003.
- [69] Jürgen Cox and Matthias Mann. MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, 26(12):1367–1372, 2008.
- [70] Jürgen Cox, Marco Y Hein, Christian A Lubner, Igor Paron, Nagarjuna Nagaraj, and Matthias Mann. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Molecular & Cellular Proteomics*, 13(9):2513–2526, 2014.

Part II
Research

1

Selective activation of human caseinolytic protease P (ClpP)

1.1 INTRODUCTION

During the characterization of bacterial ClpPs (see chapter 0), it was early suggested that there might be also ClpP homologs in mammalian cells.¹ Indeed, in humans, hClpP could be found in mitochondria according to the endosymbiotic theory.^{2,3} Later, the same came true for ClpX.⁴

1.1.1 STRUCTURE OF hClpP

In its simplest form, the serine hydrolase hClpP consists of seven equal monomers that are aligned to a ring (figure 1.1a).⁵ Each monomer bears an active site, which is made up of the nucleophilic S153 and its partners H178, and D227. Their alignment allows the formation of a charge-relay system in order to render the serine catalytically active (figure 1.1b). The completely different numbering of active site residues compared to *S. aureus* ClpP indicates a major difference between human and bacterial ClpPs: hClpP possesses an N-terminal extension of 56 amino acids, which is used as a mitochondrial import signal and cleaved off after the protein has reached its final destination.³ At the C-terminus, hClpP bears a 28 amino acids long tail. The fold is not resolved in the available crystal structure (PDB 1TG6), which points to a flexible arrangement. However, the function of this part remains largely unknown. Mutational studies suggest a role in the control of hClpX binding. Hence, hClpP lacking the C-terminus is found to bind more tightly to hClpX.^{6,7} The core part of a hClpP monomer is a globular head domain. From this rather compact structure, a long arm made of an α -helix (E-helix) protrudes. This region is mainly responsible for the formation of double-heptamer complexes by ring-ring stacking.⁷

To gain activity, two hClpP heptamers stack on top of each other and form a barrel-like double-ring structure, the active tetradecamer. The complex thus features two apical entrance pores and 14 stringed active sites. It was shown that tetradecamer formation is dependent on the binding of hClpX, but functional interactions with ClpX from *E. coli* (EcClpX) are possible, too.^{5,6} Additionally, acyldepsipeptides (ADEPs, cf. figure 1.7) are supposed to trigger the active conformation by mimicking the ClpX interaction.^{8,9} A similar artificial activation was observed for bacterial ClpPs, e.g. by ADEP (fragments) or Activators of Cylindrical Proteases (ACPs).^{10,11,12}

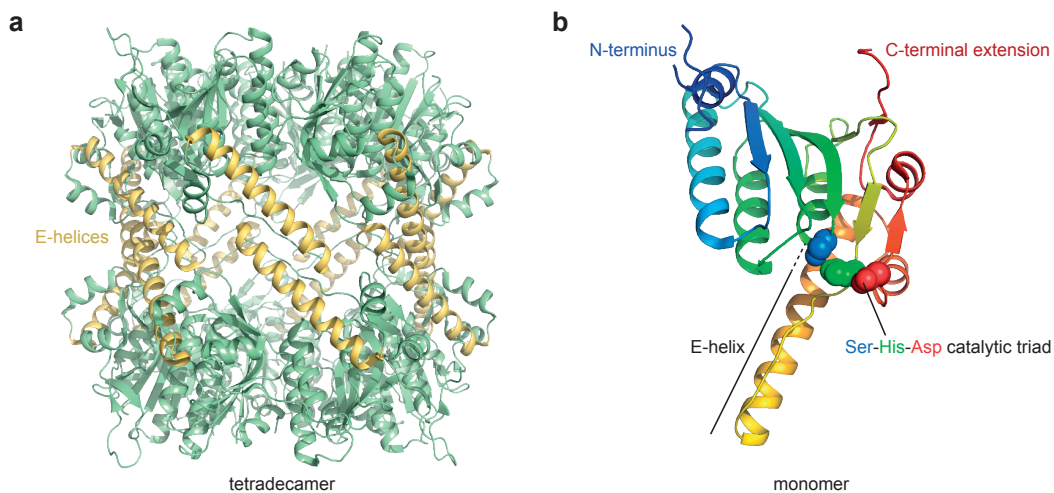


Figure 1.1: Structure of hClpP. (a) Complex of hClpP consisting of 14 monomers (PDB 1TG6), which is thought to be active. Ring-ring interactions of heptamers are partly mediated via the E-helices (highlighted in yellow). (b) Depiction of a single hClpP monomer from a. The sequence is colored from N- to C-terminus in blue to red. The C-terminal extension is not resolved. Side chains of the catalytic triad – S153, H178, and D227 – are highlighted.⁷

1.1.2 FUNCTION OF hCLpP

hClpP is embedded in the matrix of mitochondria.^{2,3} As a protease, it takes part in the mitochondrial proteostasis system. However, there is not much known about putative substrates. At least Noar, a mitochondrial matrix GTPase, was found to be degraded.¹³ Furthermore, the generation of a hClpX trap by mutating the ATP-hydrolyzing Walker B motif provided initial evidence for several other substrates or interacting proteins. Among them is the protein p32, which is thought to play a role in mitoribosome formation. However, *in vitro* degradation by hClpXP was not successful.^{9,14} Most recently, Cole *et al.* conducted a ClpP interactome study that revealed several proteins stemming from the respiratory chain and the mitochondrial metabolism.¹⁵

A ROLE OF hCLpP IN THE UNFOLDED PROTEIN RESPONSE OF MITOCHONDRIA?

As a consequence of the enrichment of unfolded or misfolded proteins in the cytoplasm, biosynthesis and activation of an armada of chaperones are induced. For example, the heat-shock proteins Hsp70 and Hsp90 are key players in a well regulated and fine-tuned mammalian stress response.¹⁶ Additionally, the endoplasmic reticulum (ER) evolved its own stress sensing systems.¹⁷ And also mitochondria harbor their proper stress response: the unfolded protein response of mitochondria (UPR^{mt}).¹⁸

As far back as in the early days of UPR^{mt} research, the name of ClpP already appeared. It was found to be upregulated, when unfolded proteins are accumulating in mitochondria.¹⁸ Indeed, in *Caenorhabditis elegans*, product peptides of ClpXP-mediated degradation find their way through the Haf1 transporter in the inner mitochondrial membrane to the cytosol.^{19,20} In the following, chaperone-related genes are upregulated.²¹

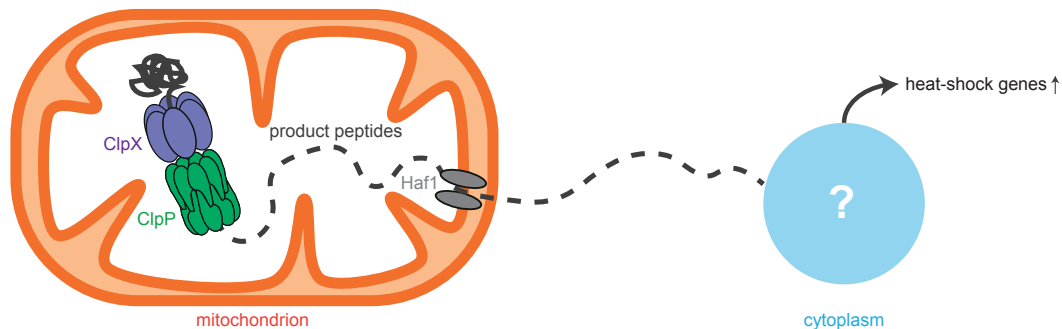


Figure 1.2: Model of the UPR^{mt} of *C. elegans*. Under stress conditions, proteins are degraded by the ClpXP machinery. Product peptides are transported through the inner mitochondrial membrane with the transporter Haf1 and translocate further to the cytosol.^{19,20} There, they trigger a transcriptional response.

If this model from UPR^{mt} for *C. elegans* is also true for mammalian and thus human cells, is still unclear to date. Seiferling *et al.* recently alerted the scientific community by showing that ClpP seems not to play a role in the UPR^{mt} of mice hearts.²² In fact, the mammalian mitochondrial stress response might more depend on ClpX as its upregulation in cultured mouse myoblasts led to increased levels of UPR^{mt} markers.²³ In future, insights into the functional role of hClpP can be mainly gained by (i) the use of tailored chemical compounds to manipulate hClpP's structure and function and (ii) by carefully characterizing disease-associated hClpP mutants.

HCLPP IN DISEASES

Disease-affiliated variants of ClpP have been continuously reported in recent years. The overwhelming number is related to cancer and collected in the COSMIC database (Catalogue of Somatic Mutations in Cancer). More than 20 point mutations, distributed over the whole sequence and out of various tissues, are documented.* Certainly, it remains elusive, if these are cause or consequence of cancer.²⁴ In a study that is presented in chapter

* As of 5th January 2018.

2, some recombinantly overexpressed hClpP mutants were subjected to an analysis of their peptidolytic and proteolytic activity. The former comprises the digestion of peptides by hClpP and the latter describes the degradation of full proteins with the help of ClpX or an artificial activator.²⁵ Nine out of ten mutants showed an aberrant peptidase behavior, which was in eight cases corrected to standard values in proteolysis. This fact demonstrates the conformational power that ClpX exerts on hClpP. However, one mutant – A131V – was completely inactive in both experiments.²⁵

In addition, hClpP is overexpressed in primary mononuclear cells of acute myeloid leukemia (AML) samples. Strikingly, a knock-down of *CLPP* in respective cell lines disturbed their viability. In accordance to that, the treatment of AML cells with a β -lactone ClpP inhibitor had cytotoxic effects.¹⁵ Vice versa, overexpression of *CLPP* in HeLa cells protects them from killing by cis-platin. Again, the opposite effect occurs, when ClpP or ClpX levels are diminished.²⁶

Besides cancer, hClpP seems to play a major role in the pathogenesis of the Perrault syndrome. This very rare and autosomal-recessive disease comes along with hearing loss and ovarian failure.²⁷ In an initial, study Jenkinson *et al.* detected two point mutations in *CLPP* – T145P and C147S – to be causative for the onset of the Perrault syndrome.²⁸ Both of them are as active as the wild-type in *in vitro* proteolysis experiments.²⁵ Later, more *CLPP* mutants were reported, but the mechanism of action to develop the pathologic state remains to be uncovered.^{29,30} A first step to solve this question was the generation of *CLPP* deficient mice, which led to increased levels of ClpX, mitochondrial DNA and inflammatory factors.³¹

1.1.3 OBJECTIVE

Despite the extensive examination of bacterial ClpP homologs in past decades, the human isoform remains poorly characterized. On a structural level, hClpP exhibits numerous similarities to its bacterial counterparts, however, its dedicated function in a cell biological context is still a riddle. One reason for that is the distinct oligomeric composition of hClpP. It usually forms inactive heptamers, whereas bacterial ClpPs in majority build tetradecamers *ab initio* that are active *per se*.⁵ This circumstance makes it challenging to develop *in vitro* and *in vivo* assays to probe the enzyme as the natural activator – hClpX – is largely unstable, hard to purify and endogenous substrates for degradation assays are rare.

This study aimed to discover an artificial small molecule activator that is easy to synthesize and facile to handle. It should be long-term stable and activate hClpP in a sufficient manner. In fact, such a molecule, named **D9**, was identified using high-throughput-screening data of a previous ClpP inhibitor study.³² It could be shown that **D9** induces the formation of tetradecamers and it seems to bind to the hydrophobic pockets, where typically loops of ClpX are interacting with. Finally, a cocrystal structure of a superactivated hClpP mutant together with **D9** could be gained, which lends insights into the fine-tuned conformational switches operating during enzyme activation.

1.2 RESULTS AND DISCUSSION

1.2.1 DISCOVERY OF **D9**

In 2015, a study was published in which the authors screened a library of about 140,000 small molecule compounds for inhibitors of *S. aureus* ClpP (SaClpP).³² The screen used the peptide *N*-succinyl-Leu-Tyr-7-amido-4-methylcoumarin (Suc-Leu-Tyr-AMC) as a substrate for ClpP-mediated peptidolysis. The protease is able to cleave off the fluorogenic dye, which thereby is released from quenching and exhibits larger fluorescence. The nature of the screen design allowed to reassess initial data and to specifically look for SaClpP activators among the library compounds (cf. Master's Thesis of M. Stahl).³³ If SaClpP significantly rose over 200% compared to the basal activity of SaClpP, the compound was considered as putative activator. After stringent filtering (for a detailed description of the selection steps, see the methods section), 70 molecules were further analyzed in a proteolysis setup, for which fluorescein-labeled casein (FITC-casein) was applied as a model substrate. SaClpP is not capable to efficiently digest this protein without an activating substance. Conversely, none of the compounds led to a stimulation of SaClpP activity. Indeed, the proteolytic potential of ClpPs from *E. coli* (EcClpP) and *L. monocytogenes* (LmClpP₂) was not amplified.³³ Eventually, an up-scaled reproduction of the peptidolytic screening conditions failed to show activation for any of the molecules (data not shown). Thus, SaClpP seems to behave differently under the high-throughput conditions.

However, when switching to human ClpP, one of the 70 compounds stood out as an activating substance. Due to its position on the well plate, it is henceforward called **D9** (figure 1.3a).³³ The maximum level of activation is comparable to the stimulation by the ADEP fragment **21** (courtesy of Pavel Kielkowski),¹¹ but only observed for hClpP and hClpPΔC

lacking the last 28 amino acids. Notably, for the tested bacterial isoforms, no significant increase in peptidase or protease activity could be manifested (figure 1.3bc). Of note, BsClpP could only be activated in peptidolysis by **2I**, not in proteolysis. This might be due to assay conditions. **D9** stimulated both hClpP peptidolysis and proteolysis in a concentration-dependent manner with an apparent dissociation constant for proteolysis in the low micromolar range (figure 1.3de). In order to check, if **D9** and **2I** are able to reach the hClpP peptidase activation level as ClpX does, peptidolysis should be measured in absence and presence of ClpX. Such an experimental setup was designed with ClpX from *E. coli* (EcClpX) because it is able to interact with hClpP and the complex can efficiently degrade the green fluorescent protein equipped with a SsrA recognition tag (GFP-SsrA, figure 1.3f).^{34,35,6} EcClpX together with hClpP increases the peptidase activity of hClpP around fivefold in comparison to **D9** or **2I** and is thus a far more potent activator than the artificial compounds. The avidity effect of multiple EcClpX binding loops and ATP-hydrolysis-driven dynamics might be possible explanations for that. In order to further explore the connection between structural features of **D9** and its activation potential, derivatives were analyzed.

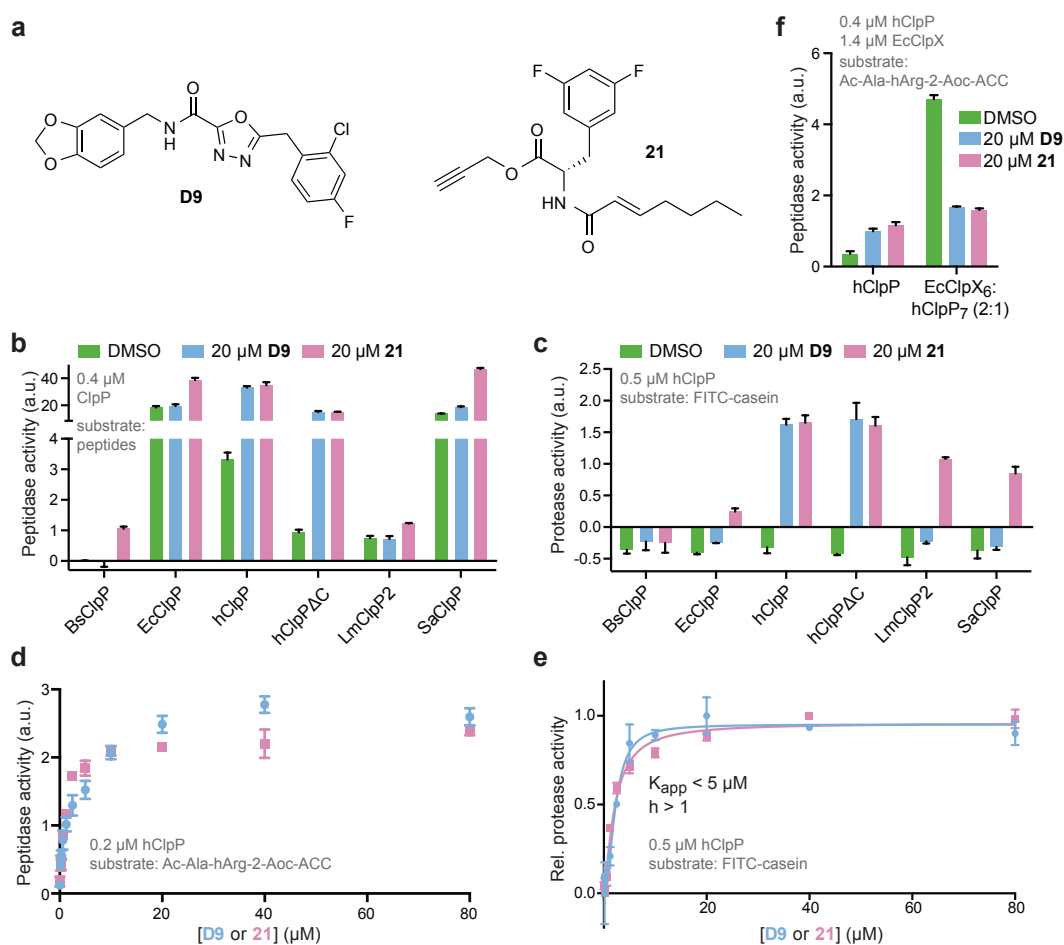


Figure 1.3: Initial characterization of D9. (a) Structures of the here discovered hClpP activator **D9** and the previously described bacterial ClpP stimulator **21**.¹¹ (b) Assessment of the activation potential of **D9** and **21** on ClpP from different species. Bs: *B. subtilis*, Ec: *E. coli*, Lm: *L. monocytogenes*, Sa: *S. aureus*, hClpPΔC is lacking 28 C-terminal amino acids compared to the wild-type, but – as all ClpPs – bears a C-terminal Strep tag. For each ClpP, an optimized peptide substrate was offered for degradation (see methods section).²⁵ Unlike as in initial screens, the same buffer (hClpP) and temperature (37 °C) conditions for all ClpPs were used. **21** enables increased peptidolysis over the DMSO control for all ClpPs, whereas **D9** only overactivates the human isoforms. (c) Respective proteolysis assay with FITC-casein as a fluorogenic substrate protein. Negative protease activities are due to fluorescent substrate quenching. BsClpP does not show any activity upon compound treatment. Human ClpPs are stimulated in equal amounts by **D9** and **21**. (d and e) Dose-dependent peptidolysis and proteolysis experiments. For peptide substrates, please refer to the methods section. Differences between **D9** and **21** are not significant. For e, data was normalized to the highest and lowest data points by Graphpad Prism and it was possible to fit an allosteric sigmoidal curve. However, apparent binding constants could not be obtained reproducibly over independent experiments, but reside in the low μM range. Hill coefficients are constantly greater than 1 indicating a cooperative binding mode. (f) Comparison of **D9**- and **21**-mediated activation of hClpP to EcClpX-stimulated peptidolysis. The EcClpX:hClpP complex reaches a fivefold higher activity compared to hClpP with **D9** or **21**. Declining activities for the mixed complex of EcClpX and hClpP together with compound might be due to EcClpX displacement effects (see figure 1.6). Data was recorded in triplicate in two independent experiments. One triplicate including standard deviations is shown (n=3).

1.2.2 STRUCTURE-ACTIVITY RELATIONSHIPS

The two outer parts of **D9** were subjected to a closer inspection of structure-activity relationships. First, the eastern moiety of **D9** was analyzed because such a substituted benzyl group also exists as a (modified) phenylalanine in ADEP molecules.³⁶ In detail, this part interacts with ClpP's hydrophobic pockets around the apical pore, which is one of the ClpX binding sites.³⁷ 13 commercially available or in-house synthesized derivatives (courtesy of Vadim S. Korotkov) were screened in proteolysis experiments with FITC-casein as substrate. Exchange of the chlorine or fluorine substitution to the corresponding halogen, in order to gain homosubstituted phenyl moieties, kept the activation potential of **D9**, albeit it is slightly reduced (compounds **1** and **2**). However, when switching two fluorine substituents from the *ortho, para* to a *meta, meta* position (**3**) – as it is found in ADEPs –, activity further decreases. A single chlorine in *ortho* or *meta* position is weakly accepted (**4** and **5**), but other constellations as well as the unsubstituted benzyl group (**8**) did only reach very poor or no stimulation of proteolysis. Thus, single substitutions are less important than the overall substitution pattern of the benzyl moiety.

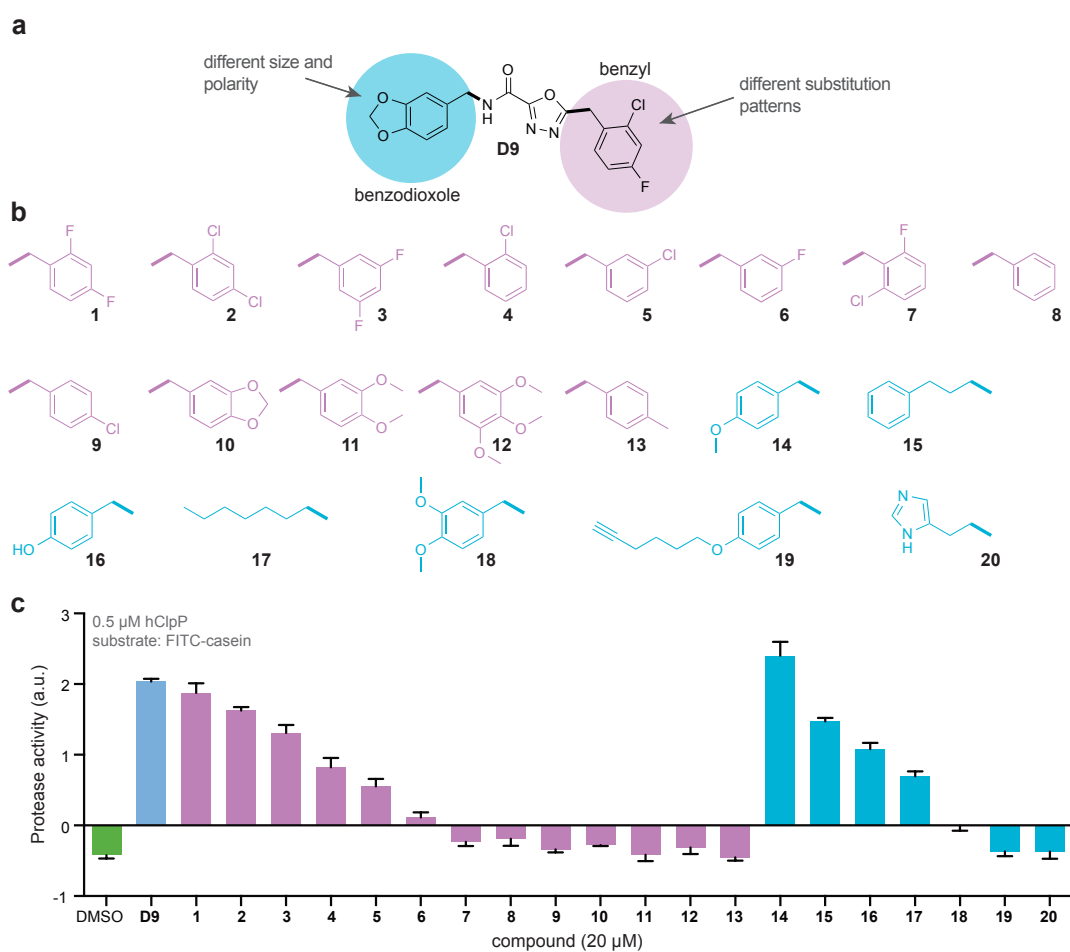


Figure 1.4: Structure-activity relationship study of **D9**. (a) The **D9** molecule can be subdivided in three parts: (i) the outer benzodioxole and (ii) substituted benzyl moieties, and (iii) the central amide bond with the oxadiazole. To probe the effect of two moieties that are located far from each other, the outer eastern and western parts were modified. The former is of special interest because of its structural similarity to ClpP activating ADEP molecules. (b) Structures of applied derivatives. (c) Derivative-dependent proteolysis screen. 20 μM compound were added to 0.5 μM hClpP. Then, digestion of FITC-casein was fluorescently monitored. Data was recorded in triplicate in two independent experiments. One triplicate including standard deviations is shown (n=3).

The western part of **D9** was also probed in respect of its structural features, although on a broader scope (figure 1.4). Tested compounds are not just substituent-exchange derivatives, but structurally diverse. For example, when changing the benzodioxole to a methoxy-substituted benzyl (**14**), no significant difference on enzyme activation compared to **D9** could be observed. But, when extending the carbon chain between amide and unsubstituted phenyl (**15**) or when replacing the methoxy moiety by an hydroxyl group (**16**), immediately a drastic decrease in activity was observed. Other changes, such as a dimethoxy

substitution pattern (**18**) or introducing an imidazole moiety (**20**), resulted in very poor or no activation of hClpP. The western part of **D9** seems therefore to be more permissive in structural changes than the eastern part. However, both parts together are important for successful activation of proteolysis. The mechanistic basis of this activation was then further examined by biophysical and biochemical methods.

1.2.3 MODE OF **D9** ACTION

It is known that hClpP *per se* forms inactive heptamers.⁵ Thus, in comparison to SaClpP, the heptamer-tetradecamer equilibrium is strongly held on the smaller complex side. The mechanism of action of **D9** might be in part based on an equilibrium shift towards higher oligomeric hClpP complexes. In order to test this hypothesis, **D9** dose-dependent analytical ultracentrifugation runs with hClpP were carried out (figure 1.5a). Indeed, a theoretical ratio of 0.3 **D9** to one hClpP tetradecamer was enough to start shifting the equilibrium to the tetradecameric state. Higher amounts of **D9** (more than nine per tetradecamer) then completely led to the formation of the tetradecameric species (figure 1.5b).

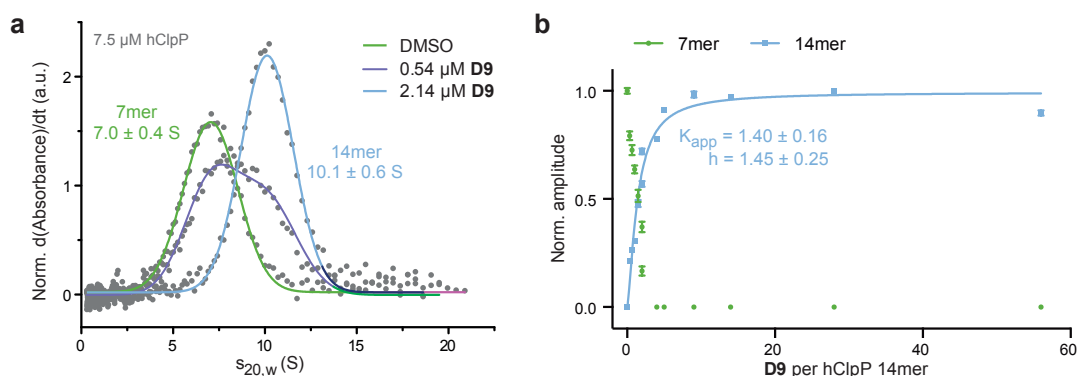


Figure 1.5: Analytical ultracentrifugation to monitor the heptamer-tetradecamer equilibrium of hClpP. (a) Example of raw data for two concentrations of **D9** and the DMSO control. Rising **D9** concentrations are reflected in a shift from a 7.0 S to a 10.1 S complex. This most likely represents the transition from heptamer to tetradecamer. (b) Plot of all theoretical **D9**:hClpP₁₄ ratios and their ability to stabilize the tetradecamer species. The blue line represents an allosteric sigmoidal fit. The apparent dissociation constant (K_{app}) and Hill coefficient (h) are given with standard errors. Data was recorded as one replicate for each concentration. Error bars represent standard errors of bigaussian fitting.

Activation of hClpP physiologically takes place through interaction with hClpX.^{5,6} Assuming that **D9** is binding in the same region as hClpX, namely around the apical entry pores, the compound should be capable of displacing the chaperone. Such a chaperone-mediated proteolysis setup was established by combining EcClpX with hClpP and GFP-SsrA as a

substrate. Indeed, rising concentrations of **D9** in an EcClpX:hClpP degradation system lowered the processing rate of GFP-SsrA, which has to be unfolded and translocated by EcClpX before digestion through hClpP (figure 1.6a). This result suggests, but does not prove, that **D9** mimics hClpP (figure 1.6b). In fact, the behavior resembles the characteristics of ADEPs that also replace ClpX in the complex.^{10,8} Similar GFP-SsrA degradation experiments were run for chaperone-free hClpP or inactive hClpP S153A and EcClpX without hClpP (figure 1.6cd). In all cases, the GFP-SsrA degradation was about tenfold lower than in EcClpX:hClpP experiments with DMSO and in neither case could a difference between DMSO and compound be observed.

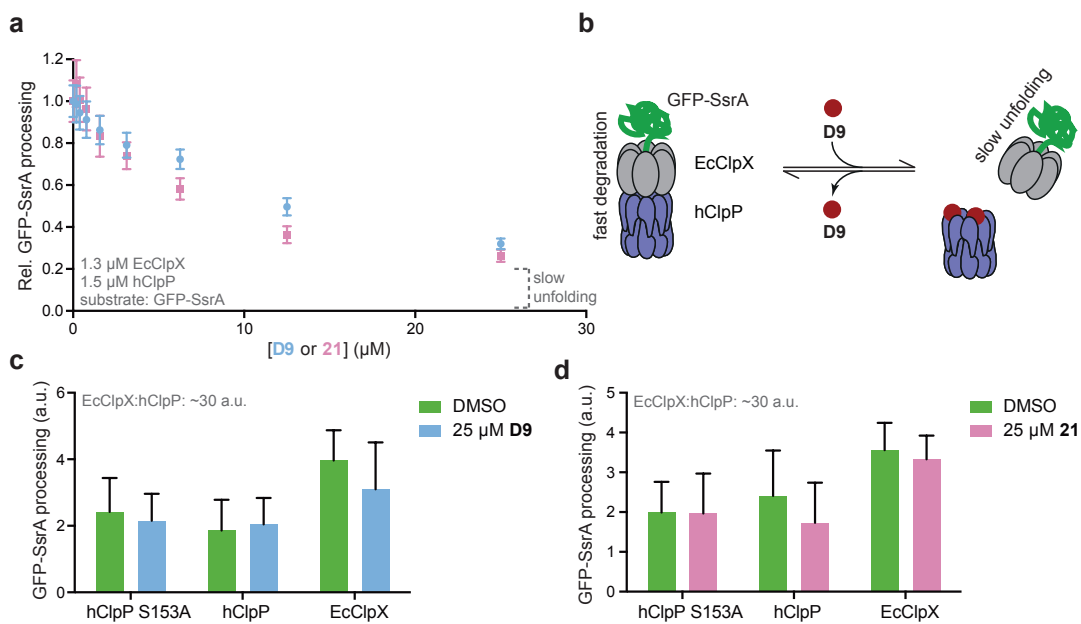


Figure 1.6: D9-dependent attenuated EcClpX-mediated GFP-SsrA processing. (a) The potential of GFP-SsrA unfolding and degradation by a complex consisting of EcClpX and hClpP was fluorescently monitored. Upon rising concentrations of **D9** or **21**, GFP-SsrA processing slowed down. The final baseline can be explained by slow GFP-SsrA unfolding through EcClpX. (b) Model of EcClpX replacement by **D9**. (c and d) Control reactions for both compounds (25 μM each) with inactive hClpP S153A only, wild-type hClpP only or EcClpX only. The compounds did not have any significant effect. Data was recorded in triplicate and in two independent experiments. Unnormalized initial slopes of all six replicates were combined and then normalized to the DMSO control. Error bars represent standard deviations (n=6).

1.2.4 DOCKING AND MUTATIONAL STUDIES

The enzymatic results obtained called for a closer view on the **D9** binding site. Therefore, *in silico* docking studies were performed. Before applying the docking procedure to **D9** and hClpP, the method was validated by a successful redocking of **ADEP2** into BsClpP (PDB

3KTJ, figure 1.7ab).³⁷ Then, the optimal binding conformation of **D9** into the hydrophobic pockets of hClpP (PDB 1TG6, figure 1.7c) was determined.⁷ The highest ranked docking solution gave evidence that the substituted benzyl moiety of **D9** is buried in an extension of the hydrophobic pockets (figure 1.7d). The entrance of the cleft is marked by a motif of three aromatic amino acids: Y₁₁₈, Y₁₃₈, and W₁₄₆ (YYW motif), which probably establish π -stacking interactions with the **D9** benzyl group (see also figure 1.8a). Additionally, docking of compound **8**, the unsubstituted benzyl derivative, was executed. A pose with the most similar conformation to **D9** revealed a deeper positioning of the benzyl moiety in the binding cavity (figure 1.7e). Likewise, compound **3** with an ADEP-like substitution pattern exhibits a perpendicular orientation of the benzyl group compared to **D9** (figure 1.7f). Most surprisingly, **21** was found to bind the other way round in the hydrophobic pocket (figure 1.7g). As these are only theoretical models, they have to be handled with care. However, an upside down binding mode of **21** would explain its activation potential towards human as well as bacterial ClpPs, where it could bind in an inverse conformation.

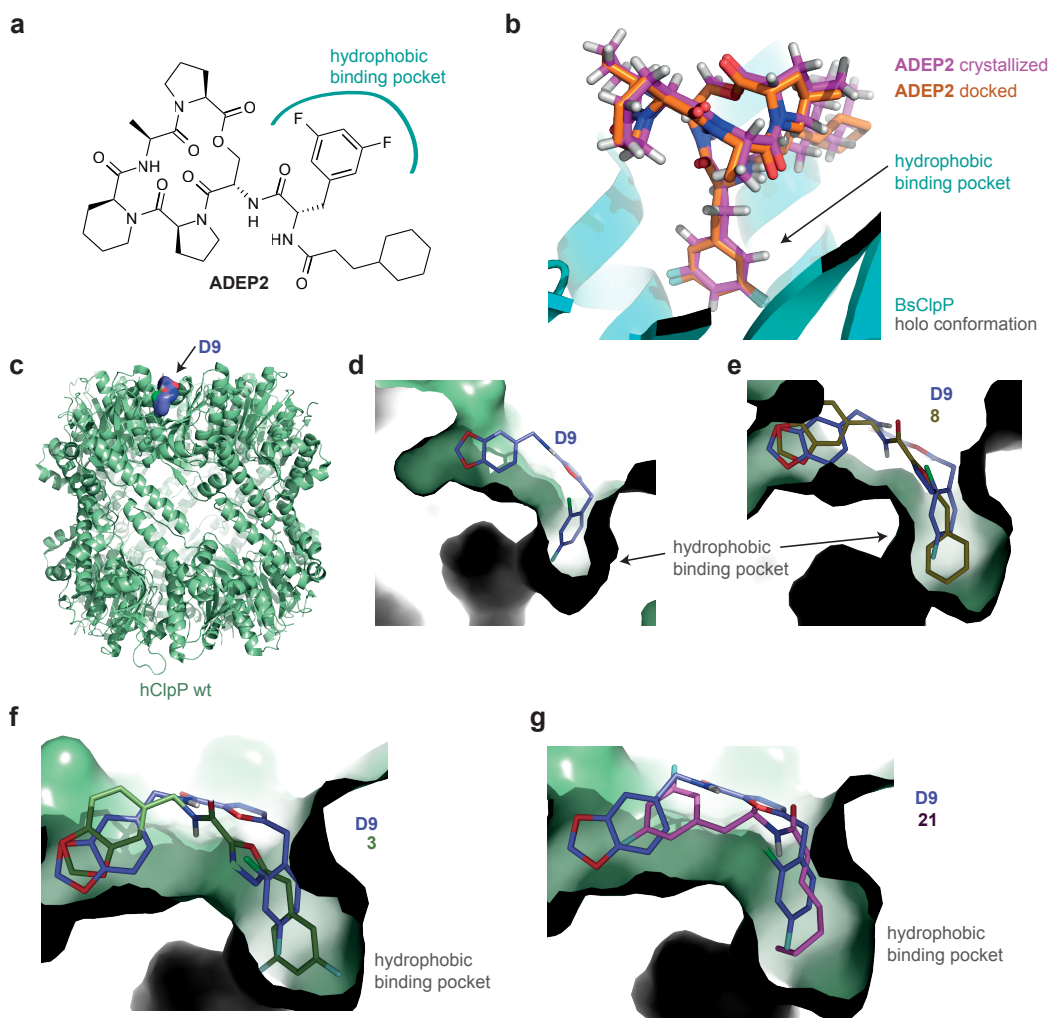


Figure 1.7: Docking of **D9** to hClpP. (a) Structure of **ADEP2**, the activating agent of BsClpP, which was used to validate the docking method. (b) Redocking of **ADEP2** to a cocrystal structure of BsClpP with activator (PDB 3KTJ).³⁷ The substituted phenyl alanine moiety fits into the hydrophobic pocket on the apical surface of BsClpP. The docked and crystallized **ADEP2** conformations show a sufficient overlay. (c) Docking of **D9** to the wild-type hClpP structure (PDB 1TG6). (d) Close-up view from the side on the hydrophobic binding pocket. The substituted benzyl group points into the cleft. The benzodioxole moiety on the other side of **D9** fits to a distant second pocket. (e) Comparison with the docking of compound **8**, which only possesses an unsubstituted benzyl group, leads to a deeper insertion of the molecule. (f and g) Docking of compounds **3** (difluoro-substitution in *meta, meta* position) and **21**; each in comparison to **D9**.

To supplement these structural findings with biochemical activation data, a set of hClpP point mutants around the presumed binding site and inspired from interspecies amino acid substitutions was generated. The mutants were subjected to a peptidolysis screen with and without **D9** present (figure 1.8). Monitoring the peptidase activity here allows to check the

general mutant wellness without compound and subsequently enables a direct comparison to **D9**-mediated activation.

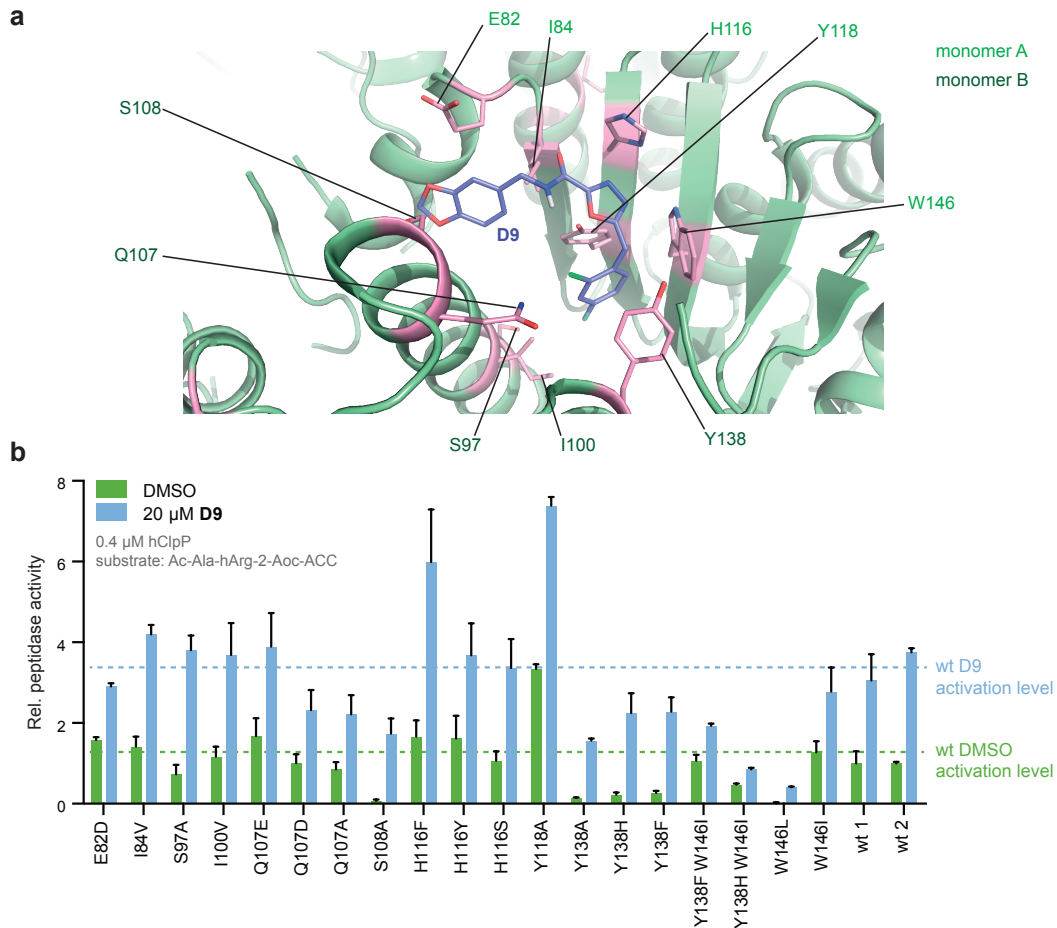


Figure 1.8: Mutant analysis of **D9**-mediated hClpP activation. (a) View into the hydrophobic binding pocket between two monomers of hClpP (PDB 1TG6).⁷ Amino acids selected for mutation are highlighted in purple. Most of them were chosen because they are altered in bacterial ClpP homologs. (b) The mutants were subjected to peptidase assays. Data was recorded in triplicate in two independent experiments from which one is shown. As the assay was conducted on two different well-plates, two wild-type measurements exist for normalization. Peptidolysis rates were normalized to hClpP wild-type with DMSO of the respective plate. Error bars represent the standard deviations (n=3).

In peptidase assays, mutants showing a normal activity level with DMSO and an altered behavior when treated with **D9** compared to the wild-type are most interesting. The peptidase activity of the point mutants E82D, Q107D, Q107A and the two double mutants is rather unaffected, however, they all show an attenuated stimulation by **D9**. Notably, the double mutants are amino acid exchanges in the YYW motif to *E. coli* (Y138F W146I) and *S. aureus* (Y138H W146I) conditions. This suggests the aromatic motif, which is in this

ClpP setting specific for hClpP, to be an important determinant for species specificity. Furthermore, two mutants were discovered that show a boosted activity with **D9** relative to the wild-type: H116F and Y118A. The latter has a basal peptidase activity on the wild-type plus **D9** level and even doubles its activity in presence of **D9**. As H116 is in its direct neighborhood, it might be possible that the two aromatic systems are part of a conformational signaling cascade. Of note, many mutants associated with the Perrault syndrome are also located around the hydrophobic binding pocket (see also chapter 2). Therefore, in future studies, **D9** will help to characterize these protein variants.

1.2.5 X-RAY CRYSTAL STRUCTURE

In order to obtain detailed insights into the binding mode of **D9** to hClpP, their cocrystallization was attempted. However, crystallization of wild-type hClpP with **D9** did not succeed. Probably, shifting the heptamer-tetradecamer equilibrium upon **D9** binding and the poor solubility of the compound in aqueous buffers ($<50 \mu\text{M}$), might disfavor (co)crystal formation. Similarly, treating hClpP crystals with **D9** did not work. However, during studies with a set of hClpP point mutants, one of them attracted special attention – hClpP Y118A. It showed exaggerated peptidase turn-over without any activator present. Thus, it was assumed that the mutant *ab initio* forms tetradecamers and would be suitable for crystallization experiments with **D9** as it could be further activated with the compound (also refer to figure 1.8b). Finally, crystals could be obtained and diffracted to a resolution of 3.1 Å.

Compared to the previously published wild-type crystal structure of hClpP (PDB 1TG6),⁷ the mutant and **D9**-bound form stands out by its approximately doubled pore diameter (ca. 30 compared to 15 Å) and reduced height (ca. 80 to 100 Å, figure 1.9a). **D9** can only be visualized as bulky electron density, because of an occupancy of only ca. 40%. However, it seems to be located between two monomers on the apical sides of the hClpP barrel (figure 1.9b). This is approximately the region, where the LGF loops of hClpX might bind to³⁸ and supports the hypothesis that **D9** is a ClpX mimic. However, it is not clear, whether the above mentioned changes in dimension occur because of **D9** binding or if they are a consequence of the Y118A mutation or a combination of both. Nevertheless, the here described widened pore is also a key feature of ADEP-mediated activation, which might suggest mutual compound-induced structural changes. Furthermore, the molecular reason for pore opening is putatively the flexible N-terminal loop crown as the first 17 amino acids are not resolved. Another segment that is not visible lies between amino acids 182-190. Strikingly,

this leads to small additional equatorial pores next to the E-helices. Conversely, these E-helices are shortened, which is usually a feature of inactive ClpP (figure 1.9c).³⁹

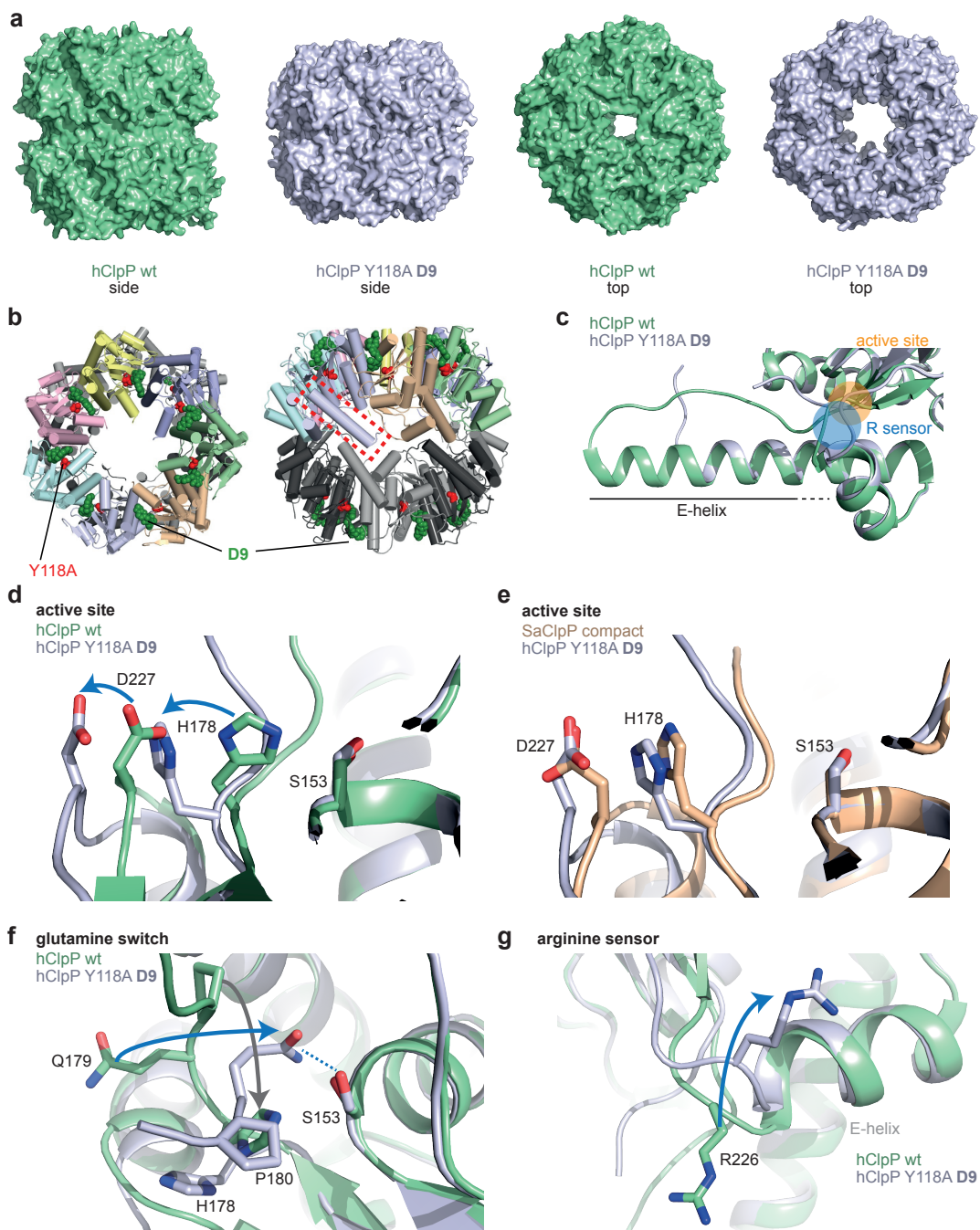


Figure 1.9: Crystal structure of hClpP Y118A partly in complex with D9. (a) The previously known wild-type structure (PDB 1TG6, green)⁷ and the novel mutant structure (not yet submitted to PDB, gray). The two proteins have different dimensions and hClpP Y118A with D9 (occupancy around 40%) exhibits a clearly enlarged pore. (b) Top- and side-view of hClpP Y118A with depiction of the localization of D9 (green). The exact conformation of the compound is not resolved due to low ligand occupancy. (c) Overlay of monomers from hClpP wild-type and the D9-bound mutant highlights a shortened E-helix (dashed red rectangle in panel b). (d) The same overlay with focus on the active site residues. In the mutant structure, H178 and D227 are turned away from S153 leading to a misaligned catalytic triad. (e) Overlay of monomers from the compact SaClpP conformation (PDB 4EMM)³⁹ and mutant hClpP shows a similar rearrangement of the active site residues. (f) Neighboring residues of H178 also undergo conformational switches in comparison to the wild-type. Both Q179 and P180 flip around ca. 180°. Subsequently, the glutamine might coordinate the hydroxyl moiety of S153 and blocks the adjacent S1 pocket. (g) Example for different conformations of R226, which is part of the arginine sensor mediating ring-ring interactions. This arginine is not resolved for all of the hClpP monomers.

A recently published study with a crystal structure of the superactive SaClpP Y63A mutant (PDB 5C90) delineates a chain of amino acids that change their conformations upon mutation.^{40,47} The exchanged tyrosine is located at the same position as Y118 in hClpP. But beyond that, the activation mode seems to be different. For example, the crucial N42 in SaClpP is replaced by S97 in hClpP, which upon mutation shows no difference to the wild-type in peptidolysis (figure 1.8b). Most remarkably, the strand β_9 is not formed in the human mutant and the overall shape of a hClpP Y118A monomer can be aligned very well to a compact and inactive structure of SaClpP (PDB 4EMM, 103 amino acid positions of 221 in hClpP are identical to SaClpP (47%)).³⁹ In both structures, the catalytic triad is not aligned (figure 1.9de). This finding is in strong contrast to the superactivity of the **D9**-stimulated hClpP mutant. The paradox continues with P180. In an inhibited SaClpP structure (PDB 5DL1) as well as in its compact conformation, this proline is found flipped 180° around compared to an active extended conformation (PDB 3V5E).^{41,39,42} This feature is a consequence of or a cause for the misalignment of the catalytic triad. In the mutant hClpP structure, this proline is also turned away from the E-helix (figure 1.9f). Simultaneously, Q179 also flips around and blocks the S1 pocket by coordinating to the active site serine. It remains in question, if such an interaction would lead to atypical activation of the serine via a Ser-Gln catalytic dyad or if it is important for cleavage product release. Also R226, a member of the arginine sensor for inter-ring contacts, is turned upside down compared to the wild-type (figure 1.9g) and would thus not be able to interact with the adjacent heptamer ring.

Taken together, hClpP Y118A exhibits many conformational characteristics that are actually known from inactive SaClpP crystal structures. Thus, it remains in question, if the present snapshot represents the same overall conformation as observed in all other biochemical assays above, where hClpP Y118A is typically more active than the wild-type. Notably, the crystals were grown under acidic conditions (pH 4.6), which might lead to the selection of unusual conformations. These, however, can be part of the catalytic proteolysis cycle. Importantly, the region of **D9** binding could be confirmed by X-ray crystallography.

1.3 CONCLUSION AND OUTLOOK

The small molecule compound **D9** was discovered out of the reanalysis of a high-throughput screen. It efficiently activates ClpX-independent peptidolysis and proteolysis, which could not be transferred to bacterial ClpP homologs. Following binding to the hydrophobic pock-

ets of hClpP, **D9** presumably acts at least via two interconnected ways: (i) by shifting the heptamer-tetradecamer equilibrium to larger complexes that are known to be functional,⁶ and (ii) by exerting conformational control over the whole hClpP complex. However, the exact mode of action at the amino acid level remains puzzling as the determined crystal structure resembles SaClpP in its compact and inactive conformation. Crucial features are a misaligned active site, a glutamine switch that closes the S_I pocket and a presumably non-functional arginine sensor. Either the crystal only captured a snapshot of an inactive hClpP conformation or the observed shape is part of the catalytic cycle, which previously was suggested for bacterial ClpPs.^{43,39} In this case, the glutamine switch is of particular interest. The residue blocks the S_I pocket, but also might establish an interaction with the catalytic serine, which could include a speed-up of the catalysis cycle. To prove such mechanistic principles, it is necessary to examine point mutants of Q₁₇₉ and other involved residues. In addition, a crystal structure of hClpP Y₁₁₈A without **D9** is needed to separate effects coming from the mutation or the compound. Finally, the docking study lent insights into a possible conformation of **D9** when binding into the hydrophobic pockets. Future studies will have to clarify the link between the allosteric binding site and conformational changes in the heart of hClpP. In addition, *in vivo* target engagement experiments are necessary to determine the specificity of **D9** in a full proteome context and to probe its physiological consequences on live cells, especially cancer cells.

1.4 METHODS

1.4.1 COMMENT ON COMPOUNDS AND THEIR SOLUBILITY

Compounds were either purchased (ChemDiv) or synthesized by Vadim S. Korotkov and Pavel Kielkowski. Due to the very poor solubility of **D9** and derivatives ($<50 \mu\text{M}$ in water, solubility in DMSO seems to be also insufficient, as tiny white powder rests could be frequently observed in tubes), a certain error must be assumed for all concentration-dependent measurements.

1.4.2 HIGH-THROUGHPUT SCREEN REANALYSIS

The screen was originally published by Hackl *et al.*³² Compounds showing 200% to 1000% activity compared to DMSO were preselected. Ten compounds were manually chosen from the $>500\%$ to 1000% region. Few conjugated double bonds were favorable. 1287 substances from the 200% to 500% region were further analyzed with RDKit.⁴⁴ Structurally similar compounds were clustered and from the clusters with the highest averaged activation potential, 240 compounds were selected. Again 100 compounds were taken from the most active compounds in the $>500\%$ to 1000% region that were not already selected via clustering. In total 350 substances were retested for autofluorescence at the COMAS facility in Dortmund, where the initial screen had been performed. 70 compounds with autofluorescence in the range of the DMSO control were further analyzed in-house.

1.4.3 CLONING, EXPRESSION AND PURIFICATION OF PROTEINS

BsClpP, EcClpP, and SaClpP were partly a kind gift of Anja Fux and Mathias Hackl.

SaClpP was basically expressed and purified as described previously.⁴² In brief, *E. coli* BL21 (DE3) carrying ClpP constructs with a C-terminal Strep-II tag in a pET301 vector were grown at 37 °C to an optical density (OD) of approximately 0.6 and induced with 0.5 mM isopropyl- β -D-thiogalactopyranoside (IPTG). Then they were usually grown at 25 °C overnight; varying incubation times and temperatures are possible. After cell lysis via ultrasound or a constant cell disruption system, the protein was first purified in 100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA by Strep-affinity chromatography and elution with additional 2.5 mM desthiobiotin. In a second step SaClpP was finally purified by size exclusion chromatography (20 mM Hepes pH 7.0, 100 mM NaCl).

BsClpP and EcClpP were expressed and purified in a similar manner with the following

changes: expression for 5 h at 37 °C, size exclusion buffer: 20 mM Hepes pH 7.4, 100 mM NaCl.

For LmClpP₂, the following changes were applied in comparison to the SaClpP protocol: induction at an OD of ca. 0.5 with 1 mM IPTG, then incubation for 5 h, Strep buffer with pH 7.0 and size exclusion in 20 mM MOPS pH 7.0, 100 mM KCl, 5% (v/v) glycerol.

hClpP and mutants thereof were expressed in *E. coli* (DE₃) Rosetta 2 or RIL CodonPlus cells in a similar way as SaClpP, but at 25 or 37 °C as well as incubated over night or for 4 h (expression and purification of proteins for crystallization is described below). The expression construct for hClpP Δ C was obtained by using a primer, which deletes 28 C-terminal amino acids, and the Gateway[®] Cloning technology. Mutant genes were gained following the QuikChange principle with primers listed in table 1.1.

Table 1.1: Primers for hCLPP construction and mutation.

Gene	Primer forward	Primer reverse
wt	GGGGACAAGTTTGTACAAAAAAGCAGGCTTTGAAGGA GATAGAACCATGCCGCTCATTCCCATCGTGG	GGGGACCACTTTGTACAAGAAAGCTGGGTGTTATT TTTCGAACCTGCGGGTGGCTCCAGGTGCTAGCTGGGACAGG
Δ C	GGGGACAAGTTTGTACAAAAAAGCAGGCTT TGAAGGAGATAGAACCATGCCGCTCATTCCCATCGTGG	GGGGACCACTTTGTACAAGAAAGCTGGGTGTTATTATT TTCGAACCTGCGGGTGGCTCCAGGGAGGGTGGACCAGAAC
E82D	ACACGATGCGATCCCGCAGCAGCGG	CGGCTGTGCGGGATCGCATCGTGT
I84V	TGACGCACACGACGGCTCCCGCAG	CTGCGGGAGCGCGTCTGTGTCGTC
S97A	CGATGACACGCTGCGCCCTTGTATCGCACAG	CTGTGCGATAACAAGGGCGGCAACGCTGTCATCG
I100V	AGCTGTGCGACAACAAGGCTGGCAACGCTG	CAGCGTTGCCAGCCTTGTGTGTCGACAGCT
Q107E	TGTTGCTCTCGGATTCCAGGAAGAGGAGCTG	CAGCTCCTCTTCTGGAATCCGAGAGCAACA
Q107D	TTCTTGTGCTCTCGGAGTCCAGGAAGAGGAGCTGTG	CACAGCTCCTCTTCTGGAATCCGAGAGCAACAAGAA
Q107A	TCTTGTGCTCTCGGATGCCAGGAAGAGGAGCTGTG	CACAGCTCCTCTTCTGGAATCCGAGAGCAACAAGAA
S108A	GCACAGCTCCTCTTCTGCAAGCCGAGAGCAAC	GTTGCTCTCGGCTTGCAGGAAGAGGAGCTGTGC
H116Y	GCTGTGATGTACATGTAGATGGGCTTCTTGTGTC	GCAACAAGAAGCCCATCTACATGTACATCAACAGC
H116F	AGGGCTGTGATGTACATGAAGATGGGCTTCTTGTGCTC	GAGCAACAAGAAGCCCATCTTATGTACATCAACAGCCCT
H116S	AGGGCTGTGATGTACATGTAGTGGGCTTCTTGTGCTC	GAGCAACAAGAAGCCCATCTACATGTACATCAACAGCCCT
Y118A	CACCAGGGCTGTTGATGGCCATGTGGATGGGCTTCT	AGAAGCCCATCCACATGGCCATCAACAGCCCTGGTG
Y138A	CAGATCGGGTTGAGGATGGCCTGCATCGTGTGCTAGAT	ATCTACGACACGATGCAGGCCATCTCAACCCGATCTG
Y138H	ATCGGGTTGAGGATGTGCTGCATCGTGTGCTAG	CTACGACACGATGCAGCACATCTCAACCCGATC
Y138F	GATCGGGTTGAGGATGAACTGCATCGTGTGCTAG	TACGACACGATGCAGTTTCACTCAACCCGATC
Y138F (W146I)	GATCGGGTTGAGGATGAACTGCATCGTGTGCTAG	TACGACACGATGCAGTTTCACTCAACCCGATC
Y138H (W146I)	ATCGGGTTGAGGATGTGCTGCATCGTGTGCTAG	CTACGACACGATGCAGCACATCTCAACCCGATC
W146A	CTGGCCACGCACGCGGTGCAGATCGGG	CCCAGTCTGCACCGCGTGGTGGGCCAG
W146L	CTGGCCACGCACAAGTGCAGATCGG	CCGATCTGCACCTTGTGCGTGGGCCAG
W146I	GGCCTGGCCACGCAGATGTGTCAGATCGGGTT	AACCCGATCTGCACCATCTGCGTGGGCCAGGCC
S153A	GGCCAGGCCCGCCCATGGGCTCCCT	AGGGAGCCATGCGGGCGGCTGGCC

EcClpX was expressed from a N-terminal His₆-TEV construct in *E. coli* (DE₃) Rosetta 2 and incubated over night at 25 °C after induction with 0.5 mM IPTG. The cell pellet was resuspended in lysis buffer (50 mM Hepes pH 7.5, 300 mM KCl, 1 mM DTT, 15% (v/v) glycerol) and lysed with a constant cell disruption system. After adding 5 mM MgCl₂ and removing the cell debris by centrifugation, the lysate was applied to a preconditioned His-affinity column (lysis buffer + 5mM MgCl₂). After a washing step with equilibration buffer + 40 mM imidazole, the protein was eluted with equilibration buffer + 500 mM imidazole with a short gradient. Protein containing fractions were pooled. After addition of 2 mM EDTA, the sample was shaken for 1 h at 4 °C and then digested by the TEV

protease (ca. 50 µg/mL) for another 3 h. Then another equal amount of TEV protease was added before incubating the sample over night at 4 °C. The next day, EcClpX was concentrated with 50 kDa cutoff tubes in a centrifuge and applied to size exclusion chromatography in lysis buffer + 2 mM MgCl₂.

GFP-SsrA was expressed and purified as described previously.⁸

All constructs were subjected to DNA sequencing. Protein concentrations are given for the monomeric species.

Casein proteins were obtained from Sigma-Aldrich/Merck (FITC-casein: C3777) and the creatine kinase was bought from Roche (10 127 566 001).

1.4.4 FITC-CASEIN PROTEOLYSIS ASSAY FOR INITIAL IN-HOUSE COMPOUND SCREENING

This part was already accomplished during the Master's Thesis of M. Stahl.³³ For EcClpP, LmClpP₂, and SaClpP (proteins were in part courtesy of M. Gersch, M. Dahmen and R. Wachtel, their expression and purification might be slightly different to the above mentioned protocols, but enzyme quality should be equal). Proteins were dissolved to 1.25 µM in an appropriate buffer (table 1.2).

Table 1.2: Buffers for different ClpPs in initial FITC-casein degradation assays.

Protein	Buffer
EcClpP	100 mM Hepes pH 7.5, 100 mM NaCl
hClpP	50 mM Hepes pH 7.5, 300 mM KCl, 1 mM DTT 15% (v/v) glycerol
LmClpP ₂	100 mM Hepes pH 7.0, 100 mM KCl, 15% (v/v) glycerol
SaClpP	100 mM Hepes pH 7.0, 100 mM NaCl

80 µL of the protein solution were added to 1 µL of 10 mM DMSO-dissolved compound on a black 96-well plate and incubated for 15 min at 37 °C. Then 20 µL of a FITC-casein mix (1.25 mg/mL casein and 0.125 mg/mL FITC-casein in PBS) were added (final ClpP concentrations after substrate addition: 1.0 µM. Fluorescence was recorded using an automatic plate reader with an excitation wavelength of 494 nm and an emission wavelength of 521 nm. Initial linear slopes after a short thermal equilibration phase were fitted according to a linear regression model.

1.4.5 FITC-CASEIN PROTEOLYSIS ASSAY FOR OTHER PURPOSES

FITC-casein degradation was measured as described above, but with 0.625 μM hClpP (final concentration after substrate addition: 0.5 μM) and an additional incubation step of 30 min at 37 °C when the protein had been dissolved in buffer and 10 min at 37 °C when the compound had been added. 20 μL casein mix (0.5 mg/mL casein and 0.75 mg/mL FITC-casein in PBS) were given and fluorescence was detected at 485/535 nm for excitation and emission, respectively. Linear slopes after a thermal equilibration phase were fitted according to a linear regression model. Further slope curve fitting was performed with Graphpad Prism and the allosteric sigmoidal equation (equation 1.1).

$$y = v_{max} * \frac{x^b}{K_{app} + x^b} \quad (1.1)$$

v_{max} is the maximum turn-over, x is the compound concentration, b the Hill coefficient and K_{app} the apparent dissociation constant.

1.4.6 PEPTIDOLYSIS ASSAY

0.2-0.4 μM ClpP was dissolved in an appropriate buffer or for better comparison in hClpP buffer (table 1.2) and prewarmed at 37 °C for 30 min. Then 98 μL of the protein solution was added to 1 μL compound in a black 96-well plate and incubated for another 10 min at 37 °C. Subsequently, 1 μL of 10 mM substrate (table 1.3) was added and fluorescence was measured at 380/430 nm for excitation and emission, respectively. The Ac-Ala-hArg-2-Aoc-ACC substrate was used for all hClpP studies except for the comparison experiment with other ClpPs (figure 1.3b). Initial linear slopes after a short thermal equilibration phase were fitted according to a linear regression model.

Table 1.3: Used substrates for different ClpPs.²⁵

Protein	Peptide substrate
BsClpP	Ac-Ala-hArg-2-Aoc-ACC
EcClpP	Ac-Ala-hArg-2-Aoc-ACC
hClpP	Ac-Phe(3,4-Cl ₂)-hArg-Leu-ACC and Ac-Ala-hArg-2-Aoc-ACC
hClpP Δ C	Ac-Phe(3,4-Cl ₂)-hArg-Leu-ACC
LmClpP ₂	Ac-Ala-hArg-Leu-ACC
SaClpP	Ac-Ala-hArg-2-Aoc-ACC

For the initial validation of high-throughput screening conditions, 1 μ M SaClpP in 100 mM Hepes pH 7.0, 100 mM NaCl was used. 100 μ L of the protein solution was then added to 1 μ L 2.5 mM compound in DMSO on a black 96-well plate. After incubation at 32 $^{\circ}$ C for 15 min, 1 μ L of 20 mM Suc-Leu-Tyr-AMC substrate was added and the fluorescence measurement performed at 32 $^{\circ}$ C (380/430 nm or 380/440 nm for excitation and emission, respectively).

1.4.7 CLPX-MEDIATED GFP-SsrA DEGRADATION AND PEPTIDASE ASSAY

For one well on a black 96-well plate, 15 μ L 4x PZ buffer (100 mM Hepes pH 7.6, 30 mM MgCl₂, 4 mM DTT, 800 mM KCl, 0.128% (v/v) NP-40, 40% (v/v) glycerol), 6 μ L 10x ATP mix (100 mM Hepes pH 7.0, 40 mM ATP, 160 mM creatine phosphate, 200 U/mL creatine kinase) and a varying amount of water (final total volume with proteins and compound: 57 μ L) were pipetted. Then EcClpX and hClpP were added to a final concentration of 1.3 μ M and 1.5 μ M, respectively. The mixture was preincubated for 10 min at 37 $^{\circ}$ C. Next, 1 μ L DMSO-dissolved compound was added and again incubated for 10 min at 37 $^{\circ}$ C. Finally, 3 μ L of a 19.8 μ M solution of GFP-SsrA were added to start the degradation reaction. The plate reader recorded fluorescence at 485/535 nm for excitation and emission, respectively.

In case of measuring the peptidase activity instead of GFP-SsrA degradation, different concentrations for both enzymes were used. EcClpX was applied in final concentrations of 0.7 μ M and 1.4 μ M depending on the complex stoichiometry and hClpP was used in a concentration of 0.4 μ M. Compounds were added prior to the proteins without an additional incubation time. Furthermore, the assay volume was adjusted to 100 μ L including 1 μ L of peptide substrate (Ac-Ala-hArg-2-Aoc-ACC). The wavelengths to record fluorescence were 380/430 nm, respectively.

1.4.8 ANALYTICAL ULTRACENTRIFUGATION

Sedimentation velocity experiments were conducted with a Beckman ProteomeLab ultracentrifuge, which was equipped with a Ti-50 rotor (Beckman Coulter) and a UV/VIS-detection system. The experiments were run with 34,000 rpm at 20 $^{\circ}$ C. hClpP was diluted to 7.5 μ M in 20 mM Hepes pH 7.0, 100 mM NaCl. **D9** was added in increasing concentrations with a maximal DMSO concentration of 0.3% (v/v). Radial scans were recorded at a wavelength of 280 nm and the first 2 h of the data were taken for the determination

of the sedimentation coefficient distribution plots. The t^*dc/dt plots were calculated with the program SedView. Sedimentation coefficient distributions were fitted to a bigaussian model with Origin. Amplitude fitting was applied with Graphpad Prism and an allosteric sigmoidal function (equation 1.1).

1.4.9 DOCKING

Docking was executed using Autodock/Vina (ADV)⁴⁵ and PyMOL⁴⁶ with the respective plugin⁴⁷. The BsClpP structure in complex with **ADEP2** (PDB 3KTJ)³⁷ and the hClpP wild-type structure (PDB 1TG6)⁷ were used for docking.

Docking in 3KTJ: PyMOL was used to add hydrogens to the receptor 3KTJ and ligand **ADEP2**. The box for docking was defined with the ADV plugin from **ADEP2**, which is bound between the two monomers A and B of BsClpP. This resulted in the following dimensions (table 1.4).

Table 1.4: Box dimensions for docking in 3KTJ.

Parameter	Value
size_x	22.50
size_y	22.50
size_z	22.50
center_x	-23.15
center_y	-31.89
center_z	4.23

No constraints were imposed on the docking apart from this box. The co-crystallized ligand was removed from the binding site and subsequently redocked with ADV for validation purposes. Before the docking was performed with the same setup as outlined for **ADEP2**, initial 3D coordinates were calculated by Open Babel⁴⁸.

Docking in 1TG6: Ligands and receptor were prepared as described for 3KTJ. The docking box now comprised different dimensions (table 1.5).

Table 1.5: Box dimensions for docking in 1TG6.

Parameter	Value
size_x	22.50
size_y	22.50
size_z	22.50
center_x	-40.48
center_y	4.63
center_z	19.83

1.4.10 CRYSTALLOGRAPHY AND X-RAY ANALYSIS

hClpP Y118A was expressed in *E. coli* (DE3) Rosetta 2 cells. A culture of 4 L was grown upon an OD of around 0.6 and then induced with 0.5 mM IPTG. Bacteria were grown for 4 h at 37 °C, harvested and subjected to a constant cell disruption system. The lysate was cleared by centrifugation and applied to a Strep-affinity column (buffers as detailed above for hClpP wild-type). Eluting protein fractions were pooled and concentrated (molecular weight cutoff: 50 kDa) to approx. 1.5 mL and snap-frozen in liquid nitrogen. Then, the protein was stored several days at -80 °C. Finally, hClpP Y118A was rethawed and applied to size exclusion chromatography (Superdex 200 pg 16/60). In comparison to the wild-type, it eluted significantly earlier (55 mL vs. 61 mL) pointing to a tetradecameric oligomerization state. Protein containing fractions were pooled and again concentrated (molecular weight cutoff: 50 kDa) to 15.1 mg/mL and snap-frozen in liquid nitrogen.

Right before crystallization, the protein was diluted in 20 mM Hepes pH 7.0, 100 mM NaCl to 10 mg/mL. Then, **D9** was added multiple times in 50 μ M steps (0.65 μ L 10 mM **D9** in DMSO to 130 μ L protein solution each) until precipitate could be observed. The final **D9** concentration was around 400 μ M, which would equal a 1:1 ratio of **D9** to hClpP Y118A monomers, although precipitate occurred. Crystals were grown by the sitting-drop vapor diffusion method in 0.15 μ L protein with 0.15 μ L precipitant solution (0.1 M sodium acetate pH 4.6 and 25% (w/v) PEG 8000) at 20 °C for approx. 5-7 days. Prior to cryo cooling in liquid nitrogen, crystals were soaked in mother liquor supplemented with 30% ethylene glycol.

Diffraction data was collected at the European Synchrotron Radiation Facility (ESRF, beamline ID23-2) in Grenoble, France at -100 °C. Subsequently, diffraction data were processed with XDS⁴⁹ and belonged to the space group $P2_1$ with cell parameters $a=115$ Å, $b=97$ Å, $c=127$ Å and beta: 93°). For structure solution and refinement diffraction data to 3.1 Å were used (1/2 CC of $\sim 50\%$, correlation coefficient of random half-data sets).^{50,51,52} The structure was determined by molecular replacement using the coordinates of human ClpP wild-type (PDB 1TG6) and PHASER.⁵³ Manual building of the model was performed with COOT⁵⁴ and the coordinates refined with REFMAC⁵⁵ and PHENIX⁵⁶, taking advantage of the non-crystallographic symmetry in the refinement protocol (R/R_{free} : $\sim 0.21/0.25$). The final steps of structure refinement are currently underway. Therefore, no complete statistics regarding data collection and structure refinement can be provided here.

1.4.II AUTHOR CONTRIBUTIONS

Matthias Stahl cloned and expressed mutants as well as performed all biochemical experiments shown and analyzed the data unless noted otherwise. Malte Gersch provided support for cloning and the initial in-house screening as well as biochemistry. Axel Pahl selected compounds and performed the docking. Vadim S. Korotkov and Pavel Kielkowski synthesized compounds. Klaus Richter recorded and analyzed ultracentrifugation data. Dóra Balogh expressed and purified some of the proteins. Leonhard Kick and Sabine Schneider crystallized mutant hClpP and measured as well as analyzed X-ray data. Matthias Stahl, Malte Gersch and Stephan A. Sieber conceived the project.

1.5 REFERENCES

- [1] MR Maurizi, William P Clark, Seung-Ho Kim, and Susan Gottesman. ClpP represents a unique family of serine proteases. *Journal of Biological Chemistry*, 265(21):12546–12552, 1990.
- [2] Peter Bross, Brage S Andresen, Inga Knudsen, Torben A Kruse, and Niels Gregersen. Human ClpP protease: cDNA sequence, tissue-specific expression and chromosomal assignment of the gene. *FEBS Letters*, 377(2):249–252, 1995.
- [3] Thomas J Corydon, Peter Bross, Henrik U Holst, Neve Søren, Karsten Kristiansen,

- Niels Gregersen, and Lars Bolund. A human homologue of *Escherichia coli* ClpP caseinolytic protease: recombinant expression, intracellular processing and subcellular localization. *Biochemical Journal*, 331(1):309–316, 1998.
- [4] Thomas J Corydon, Mette Wilsbech, Cathrine Jespersgaard, Brage S Andresen, Anders D Børglum, Søren Pedersen, Lars Bolund, Niels Gregersen, and Peter Bross. Human and mouse mitochondrial orthologs of bacterial ClpX. *Mammalian Genome*, 11(10):899–905, 2000.
- [5] Sung Gyun Kang, Mariana N Dimitrova, Joaquin Ortega, Ann Ginsburg, and Michael R Maurizi. Human mitochondrial ClpP is a stable heptamer that assembles into a tetradecamer in the presence of ClpX. *Journal of Biological Chemistry*, 280(42):35424–35432, 2005.
- [6] Sung Gyun Kang, Joaquin Ortega, Satyendra K Singh, Nan Wang, Ning-na Huang, Alasdair C Steven, and Michael R Maurizi. Functional proteolytic complexes of the human mitochondrial ATP-dependent protease, hClpXP. *Journal of Biological Chemistry*, 277(23):21095–21102, 2002.
- [7] Sung Gyun Kang, Michael R Maurizi, Mark Thompson, Timothy Mueser, and Bijan Ahvazi. Crystallography and mutagenesis point to an essential role for the N-terminus of human mitochondrial ClpP. *Journal of Structural Biology*, 148(3):338–352, 2004.
- [8] Malte Gersch, Kirsten Famulla, Maria Dahmen, Christoph Göbl, Imran Malik, Klaus Richter, Vadim S Korotkov, Peter Sass, Helga Rübsamen-Schaeff, Tobias Madl, et al. AAA+ chaperones and acyldepsipeptides activate the ClpP protease via conformational control. *Nature Communications*, 6:6320, 2015.
- [9] Bradley R Lowth, Janine Kirstein-Miles, Tamanna Saiyed, Heike Brötz-Oesterhelt, Richard I Morimoto, Kaye N Truscott, and David A Dougan. Substrate recognition and processing by a Walker B mutant of the human mitochondrial AAA+ protein ClpX. *Journal of Structural Biology*, 179(2):193–201, 2012.
- [10] Janine Kirstein, Anja Hoffmann, Hauke Lilie, Ronny Schmidt, Helga Rübsamen-Waigmann, Heike Brötz-Oesterhelt, Axel Mogk, and Kürşad Turgay. The antibiotic ADEP reprogrammes ClpP, switching it from a regulated to an uncontrolled protease. *EMBO Molecular Medicine*, 1(1):37–49, 2009.

- [11] Daniel W Carney, Corey L Compton, Karl R Schmitz, Julia P Stevens, Robert T Sauer, and Jason K Sello. A simple fragment of cyclic acyldepsipeptides is necessary and sufficient for ClpP activation and antibacterial activity. *ChemBioChem*, 15(15):2216–2220, 2014.
- [12] Elisa Leung, Alessandro Datti, Michele Cossette, Jordan Goodreid, Shannon E McCaw, Michelle Mah, Alina Nakhamchik, Koji Ogata, Majida El Bakkouri, Yi-Qiang Cheng, et al. Activators of cylindrical proteases as antimicrobials: identification and development of small molecule activators of ClpP protease. *Chemistry & Biology*, 18(9):1167–1178, 2011.
- [13] Natalie Al-Furoukh, Julia R Kardon, Marcus Krüger, Marten Szibor, Tania A Baker, and Thomas Braun. NOA1, a novel ClpXP substrate, takes an unexpected nuclear detour prior to mitochondrial import. *PLoS One*, 9(7):e103141, 2014.
- [14] Harunori Yoshikawa, Wataru Komatsu, Toshiya Hayano, Yutaka Miura, Keiichi Homma, Keiichi Izumikawa, Hideaki Ishikawa, Naoki Miyazawa, Hiroyuki Tachikawa, Yoshio Yamauchi, et al. Splicing factor 2-associated protein p32 participates in ribosome biogenesis by regulating the binding of Nop52 and fibrillarin to preribosome particles. *Molecular & Cellular Proteomics*, 10(8):M110–006148, 2011.
- [15] Alicia Cole, Zezhou Wang, Etienne Coyaud, Veronique Voisin, Marcela Gronda, Yulia Jitkova, Rachel Mattson, Rose Hurren, Sonja Babovic, Neil Maclean, et al. Inhibition of the mitochondrial protease ClpP as a therapeutic strategy for human acute myeloid leukemia. *Cancer Cell*, 27(6):864–876, 2015.
- [16] Klaus Richter, Martin Haslbeck, and Johannes Buchner. The heat shock response: life on the verge of death. *Molecular Cell*, 40(2):253–266, 2010.
- [17] David Ron and Peter Walter. Signal integration in the endoplasmic reticulum unfolded protein response. *Nature Reviews Molecular Cell Biology*, 8(7):519–529, 2007.
- [18] Quan Zhao, Jianghui Wang, Ilya V Levichkin, Stan Stasinopoulos, Michael T Ryan, and Nicholas J Hoogenraad. A mitochondrial specific stress response in mammalian cells. *The EMBO Journal*, 21(17):4411–4419, 2002.

- [19] Cole M Haynes, Kseniya Petrova, Cristina Benedetti, Yun Yang, and David Ron. ClpP mediates activation of a mitochondrial unfolded protein response in *C. elegans*. *Developmental Cell*, 13(4):467–480, 2007.
- [20] Cole M Haynes, Yun Yang, Steven P Blais, Thomas A Neubert, and David Ron. The matrix peptide exporter HAF-1 signals a mitochondrial UPR by activating the transcription factor ZC376.7 in *C. elegans*. *Molecular Cell*, 37(4):529–540, 2010.
- [21] Tomer Shpilka and Cole M Haynes. The mitochondrial UPR: mechanisms, physiological functions and implications in ageing. *Nature Reviews Molecular Cell Biology*, 2017.
- [22] Dominic Seiferling, Karolina Szczepanowska, Christina Becker, Katharina Senft, Stefan Hermans, Priyanka Maiti, Tim König, Alexandra Kukat, and Aleksandra Trifunovic. Loss of *CLPP* alleviates mitochondrial cardiomyopathy without affecting the mammalian UPR^{mt}. *EMBO Reports*, page e201642077, 2016.
- [23] Natalie Al-Furoukh, Alessandro Ianni, Hendrik Nolte, Soraya Hölper, Marcus Krüger, Sjoerd Wanrooij, and Thomas Braun. ClpX stimulates the mitochondrial unfolded protein response (UPR^{mt}) in mammalian cells. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1853(10):2580–2591, 2015.
- [24] Simon A Forbes, David Beare, Harry Boutselakis, Sally Bamford, Nidhi Bindal, John Tate, Charlotte G Cole, Sari Ward, Elisabeth Dawson, Laura Ponting, et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Research*, 45(D1):D777–D783, 2016.
- [25] Malte Gersch, Matthias Stahl, Marcin Poreba, Maria Dahmen, Anna Dziejcz, Marcin Drag, and Stephan A Sieber. Barrel-shaped ClpP proteases display attenuated cleavage specificities. *ACS Chemical Biology*, 11(2):389–399, 2015.
- [26] Yang Zhang and Michael R Maurizi. Mitochondrial ClpP activity is required for cisplatin resistance in human cells. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1862(2):252–264, 2016.
- [27] M Perrault, B Klotz, and E Housset. Two cases of Turner syndrome with deaf-mutism in two sisters. *Bulletins et mémoires de la Société médicale des hôpitaux de Paris*, 67(3-4):79–84, 1950.

- [28] Emma M Jenkinson, Atteeq U Rehman, Tom Walsh, Jill Clayton-Smith, Kwanghyuk Lee, Robert J Morell, Meghan C Drummond, Shaheen N Khan, Muhammad Asif Naeem, Bushra Rauf, et al. Perrault syndrome is caused by recessive mutations in *CLPP*, encoding a mitochondrial ATP-dependent chambered protease. *The American Journal of Human Genetics*, 92(4):605–613, 2013.
- [29] Saleem Ahmed, Musharraf Jelani, Nuha Alrayes, Hussein Sheikh Ali Mohamoud, Mona Mohammad Almrhami, Wasim Anshasi, Naushad Ali Basheer Ahmed, Jun Wang, Jamal Nasir, and Jumana Yousuf Al-Aama. Exome analysis identified a novel missense mutation in the *CLPP* gene in a consanguineous Saudi family expanding the clinical spectrum of Perrault Syndrome type-3. *Journal of the Neurological Sciences*, 353(1):149–154, 2015.
- [30] Fatma Dursun, Hussein Sheikh Ali Mohamoud, Noreen Karim, Muhammad Naeem, Musharraf Jelani, and Heves Kırmızıbekmez. A novel missense mutation in the *CLPP* gene causing Perrault syndrome type 3 in a Turkish family. *Journal of Clinical Research in Pediatric Endocrinology*, 8(4):472, 2016.
- [31] Suzana Gispert, Dajana Parganlija, Michael Klinkenberg, Stefan Dröse, Ilka Wittig, Michel Mittelbronn, Pawel Grzmil, Sebastian Koob, Andrea Hamann, Michael Walter, et al. Loss of mitochondrial peptidase ClpP leads to infertility, hearing loss plus growth retardation via accumulation of ClpX, mtDNA and inflammatory factors. *Human Molecular Genetics*, 22(24):4871–4887, 2013.
- [32] Mathias W Hackl, Markus Lakemeyer, Maria Dahmen, Manuel Glaser, Axel Pahl, Katrin Lorenz-Baath, Thomas Menzel, Sonja Sievers, Thomas Böttcher, Iris Antes, et al. Phenyl esters are potent inhibitors of caseinolytic protease p and reveal a stereogenic switch for deoligomerization. *Journal of the American Chemical Society*, 137(26):8475–8483, 2015.
- [33] Matthias Stahl. Capture and manipulation of the human protease ClpP. Master's thesis, TU München, 2013.
- [34] Guo-Fen Tu, Gavin E Reid, Jian-Guo Zhang, Robert L Moritz, and Richard J Simpson. C-terminal extension of truncated recombinant proteins in *Escherichia coli* with a 10Sa RNA decapeptide. *Journal of Biological Chemistry*, 270(16):9322–9326, 1995.

- [35] Susan Gottesman, Eric Roche, YanNing Zhou, and Robert T Sauer. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes & Development*, 12(9):1338–1347, 1998.
- [36] Heike Brötz-Oesterhelt, Dieter Beyer, Hein-Peter Kroll, Rainer Endermann, Christoph Ladel, Werner Schroeder, Berthold Hinzen, Siegfried Raddatz, Holger Paulsen, Kerstin Henninger, et al. Dysregulation of bacterial proteolytic machinery by a new class of antibiotics. *Nature Medicine*, 11(10):1082–1087, 2005.
- [37] Byung-Gil Lee, Eun Young Park, Kyung-Eun Lee, Hyesung Jeon, Kwang Hoon Sung, Holger Paulsen, Helga Rübsamen-Schaeff, Heike Brötz-Oesterhelt, and Hyun Kyu Song. Structures of ClpP in complex with acyldepsipeptide antibiotics reveal its activation mechanism. *Nature Structural & Molecular Biology*, 17(4):471–478, 2010.
- [38] Yong-In Kim, Igor Levchenko, Karolina Fraczkowska, Rachel V Woodruff, Robert T Sauer, and Tania A Baker. Molecular determinants of complex formation between Clp/Hsp100 ATPases and the ClpP peptidase. *Nature Structural & Molecular Biology*, 8(3):230–233, 2001.
- [39] Fei Ye, Jie Zhang, Hongchuan Liu, Rolf Hilgenfeld, Ruihan Zhang, Xiangqian Kong, Lianchun Li, Junyan Lu, Xinlei Zhang, Donghai Li, et al. Helix unfolding/refolding characterizes the functional dynamics of *Staphylococcus aureus* Clp protease. *Journal of Biological Chemistry*, 288(24):17643–17653, 2013.
- [40] Tengfeng Ni, Fei Ye, Xing Liu, Jie Zhang, Hongchuan Liu, Jiahui Li, Yingyi Zhang, Yinqiang Sun, Meining Wang, Cheng Luo, et al. Characterization of gain-of-function mutant provides new insights into ClpP structure. *ACS Chemical Biology*, 11(7):1964–1972, 2016.
- [41] Axel Pahl, Markus Lakemeyer, Marie-Theres Vielberg, Mathias W Hackl, Jan Vo-macka, Vadim S Korotkov, Martin L Stein, Christian Fetzer, Katrin Lorenz-Baath, Klaus Richter, et al. Reversible inhibitors arrest ClpP in a defined conformational state that can be revoked by ClpX association. *Angewandte Chemie International Edition*, 54(52):15892–15896, 2015.
- [42] Malte Gersch, Anja List, Michael Groll, and Stephan A Sieber. Insights into structural network responsible for oligomerization and activity of bacterial virulence

- regulator caseinolytic protease P (ClpP) protein. *Journal of Biological Chemistry*, 287(12):9484–9494, 2012.
- [43] Matthew S Kimber, Angela Yeou Hsiung Yu, Mikael Borg, Elisa Leung, Hue Sun Chan, and Walid A Houry. Structural and theoretical studies indicate that the cylindrical protease ClpP samples extended and compact conformations. *Structure*, 18(7):798–808, 2010.
- [44] Greg Landrum. RDKit: Open-source cheminformatics. *Online*. <http://www.rdkit.org>, 2006.
- [45] Oleg Trott and Arthur J Olson. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2):455–461, 2010.
- [46] Warren L DeLano. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography*, 40:82–92, 2002.
- [47] Daniel Seeliger and Bert L de Groot. Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *Journal of Computer-aided Molecular Design*, 24(5):417–422, 2010.
- [48] Noel M O’Boyle, Michael Banck, Craig A James, Chris Morley, Tim Vandermeersch, and Geoffrey R Hutchison. Open Babel: An open chemical toolbox. *Journal of Cheminformatics*, 3(1):33, 2011.
- [49] Wolfgang Kabsch. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D: Biological Crystallography*, 66(2):133–144, 2010.
- [50] Kay Diederichs and P Andrew Karplus. Better models by discarding data? *Acta Crystallographica Section D: Biological Crystallography*, 69(7):1215–1222, 2013.
- [51] Phil Evans. Resolving some old problems in protein crystallography. *Science*, 336(6084):986–987, 2012.
- [52] P Andrew Karplus and Kay Diederichs. Linking crystallographic model and data quality. *Science*, 336(6084):1030–1033, 2012.

- [53] Airlie J McCoy, Ralf W Grosse-Kunstleve, Paul D Adams, Martyn D Winn, Laurent C Storoni, and Randy J Read. Phaser crystallographic software. *Journal of Applied Crystallography*, 40(4):658–674, 2007.
- [54] Paul Emsley, Bernhart Lohkamp, William G Scott, and Kevin Cowtan. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, 66(4):486–501, 2010.
- [55] Garib N Murshudov, Pavol Skubák, Andrey A Lebedev, Navraj S Pannu, Roberto A Steiner, Robert A Nicholls, Martyn D Winn, Fei Long, and Alexei A Vagin. REF-MAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography*, 67(4):355–367, 2011.
- [56] Paul D Adams, Pavel V Afonine, Gábor Bunkóczi, Vincent B Chen, Ian W Davis, Nathaniel Echols, Jeffrey J Headd, L-W Hung, Gary J Kapral, Ralf W Grosse-Kunstleve, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, 66(2):213–221, 2010.

2

Barrel-shaped ClpP proteases display attenuated cleavage specificities

Published in ACS Chemical Biology, 2016, 11(2), pp 389-399

by Malte Gersch,[#] Matthias Stahl,[#] Marcin Poreba,[#] Maria Dahmen, Anna Dziejcz, Marcin Drag, and Stephan A. Sieber.

[#] equal contribution

Reprinted with permission. © 2016 American Chemical Society. DOI: 10.1021/ac-schembio.5b00757

SYNOPSIS

Barrel-shaped ClpP proteases possess an enclosed inner space for substrate degradation. The wall of this chamber exhibits 14 active sites – serine hydroxyl moieties that are in majority activated by a charge-relay system comprising histidine and aspartate. But prior to degradation, substrates are translocated into the barrel by the interacting chaperones ClpA, ClpC, or ClpX.

The present publication deals with the degradation of tripeptide substrates exhibiting varying amino acid sequences and quenched fluorogenic labels. Increasing fluorescence is therefore a measure of chaperone-independent peptidolytic ClpP activity. The peptides used can be divided in three categories, for which each amino acid position (P₁, P₂, P₃) was permuted according to a library of natural and unnatural amino acids. This peptide collection was screened for degradation activity with ClpP from *E. coli* (EcClpP), *S. aureus* (SaClpP), and human mitochondria (hClpP). For each of the ClpPs, distinct residue preferences at all three sites could be obtained.

However, these clear preferences greatly disappear when offering whole substrate proteins that are tagged with a SsrA signal, which can be recognized by ClpX. Subsequently, product peptides from ClpX-dependent proteolysis were identified via tandem mass spectrometry experiments and a downstream quantitative analysis of sequence conservation at the cleavage sites was performed. To this end, a web-based software ‘Protein|Clpper’ was developed that allows the upload of identified peptide information and enables visualization of cleavage patterns. As a result, the observed patterns for fluorogenic peptides seem to be overwritten when using substrate proteins in chaperone-dependent digestion processes.

This work supports the hypothesis that chaperone-mediated proteolysis can lead to extremely high protein concentrations in the secluded ClpP barrel. This would facilitate efficient hydrolysis of peptide bonds and would be largely independent of the type of amino acid side chain binding to ClpP’s substrate pockets.

Note: The basic methodology of analyzing product peptides of ClpP proteases was later also applied for ClpPs from *Listeria monocytogenes*.

Published in Chemical Science, 2017, 8(2), pp 1592-1600 by Dóra Balogh,[#] Maria Dahmen,[#] Matthias Stahl, Marcin Poreba, Malte Gersch, Marcin Drag, and Stephan A. Sieber.

[#] equal contribution

AUTHOR CONTRIBUTIONS

Marcin Poreba and Anna Dziezic synthesized the peptide library. Malte Gersch and Maria Dahmen performed fluorogenic peptidolysis assays. Matthias Stahl and Malte Gersch cloned and mutated *clpP/CLPP* and substrate genes. Matthias Stahl and Malte Gersch executed further biochemical experiments. Matthias Stahl determined kinetic enzyme values and screened hClpP variants. Matthias Stahl developed the *in vitro* proteolysis assay, performed mass spectrometry identification of product peptides and together with Malte Gersch wrote the analysis software 'Protein|Clpper'. Malte Gersch wrote the manuscript with input from all authors. Marcin Drag and Stephan A. Sieber conceived and supervised the project.

Barrel-shaped ClpP Proteases Display Attenuated Cleavage Specificities

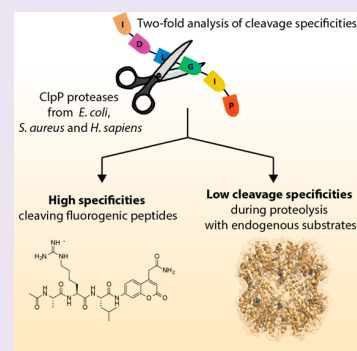
Malte Gersch,^{†,§,||} Matthias Stahl,^{†,||} Marcin Poreba,^{‡,||} Maria Dahmen,[†] Anna Dziedzic,[‡] Marcin Drag,^{*,‡} and Stephan A. Sieber^{*,†}

[†]Center for Integrated Protein Science (CIPSM), Technische Universität München, Department of Chemistry, Lichtenbergstraße 4, 85747 Garching, Germany

[‡]Department of Bioorganic Chemistry, Faculty of Chemistry, Wrocław University of Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland

S Supporting Information

ABSTRACT: ClpP is a self-compartmentalizing protease with crucial roles in bacterial and mitochondrial protein quality control. Although the ClpP homocomplex is composed of 14 equivalent active sites, it degrades a multitude of substrates to small peptides, demonstrating its capability to carry out diverse cleavage reactions. Here, we show that ClpP proteases from *E. coli*, *S. aureus*, and human mitochondria exhibit preferences for certain amino acids in the P1, P2, and P3 positions using a tailored fluorogenic substrate library. However, this high specificity is not retained during proteolysis of endogenous substrates as shown by mass spectrometric analysis of peptides produced in ClpXP-mediated degradation reactions. Our data suggest a mechanism that implicates the barrel-shaped architecture of ClpP not only in shielding the active sites to prevent uncontrolled proteolysis but also in providing high local substrate concentrations to enable efficient proteolytic processing. Furthermore, we introduce customized fluorogenic substrates with unnatural amino acids that greatly surpass the sensitivity of previously used tools. We used these to profile the activity of cancer-patient- and Perrault-syndrome-derived ClpP mutant proteins.



Proteases regulate key physiological processes in all living organisms through the cleavage of peptide bonds.¹ Since uncontrolled proteolytic activity poses a threat to cellular viability, proteases are tightly regulated by many different mechanisms, including zymogen activation, spatial coordination, cleavage specificity, allosteric activation, domain topology, and environment-dependent activity.² Many proteases have a demarcated specificity for certain amino acids around the scissile bond (Figure 1A), which is encoded through their respective structural arrangements around the active site. Virtually all types of preferences exist, and some examples include trypsin cleaving after basic residues, caspases cleaving after acidic residues, chymotrypsin cleaving after aromatic residues, and elastase cleaving after small, aliphatic amino acids.^{3–5}

While some proteases exert their function by carrying out a small set of cleavage reactions with high specificity, other proteases, such as members of the family of self-compartmentalizing proteases, are able to cleave a large number of different proteins to small peptides.⁶ Its most prominent and most evolved member, the 20S proteasome, employs three catalytic subunits with distinct cleavage specificities to facilitate efficient processing of substrates (cleavage after acidic residues by $\beta 1c$ with caspase-like activity, cleavage after basic residues by $\beta 2c$ with trypsin-like activity, and cleavage after hydrophobic residues by $\beta 5c$ with chymotrypsin-like activity).⁷

However, bacterial and mitochondrial members of this family, such as ClpXP, Lon, FtsH, and HslVU, generally feature only one type of active site which raises the question: how can “one size fit all” for proteolysis? In these machineries, proteolytic subunits (e.g., ClpP) together with AAA+ ATPase chaperones (e.g., ClpX) form barrel-shaped proteolytic complexes (e.g., ClpXP).^{8–10} In this assembly, the chaperone is responsible for substrate engagement, unfolding, and propagation,¹¹ whereas ClpP performs the peptide bond hydrolysis reaction. ClpP is composed of two heptameric rings which form a proteolytic chamber with 14 identical active sites.¹² This topology shields the active sites from the cytosol or the mitochondrial matrix which, in combination with further molecular mechanisms, protects the cell from uncontrolled degradation.^{13–15} The role of bacterial ClpXP comprises both the degradation of full-length proteins with an internal degradation tag¹⁶ for protein homeostasis purposes as well as the degradation of truncated proteins from stalled ribosomes which are targeted to ClpXP through an SsrA tag.¹⁷ Unlike other proteases whose substrate spectrum is dictated by a distinct cleavage site specificity, ClpP must thus exhibit pronounced substrate promiscuity to be able to process any

Received: September 22, 2015

Accepted: November 25, 2015

Published: November 25, 2015

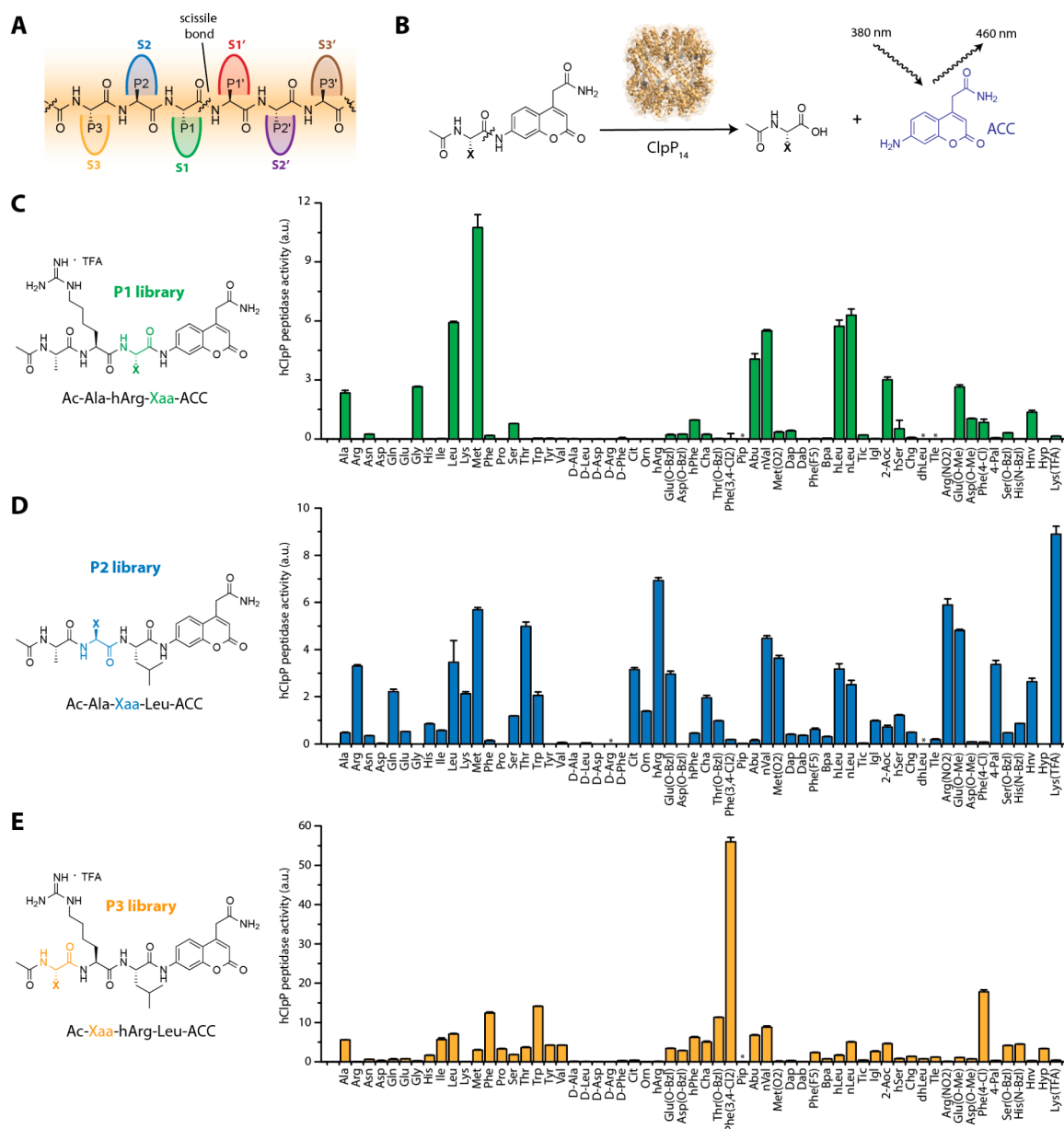


Figure 1. Cleavage specificity of human mitochondrial hClpP determined from tailored fluorogenic substrate libraries. (A) Protease nomenclature. P3, P2, P1, P1'... refer to the substrate amino acids adjacent to the cleavage site, while S3, S2, S1, S1'... refer to the respective accommodating binding pockets. (B) Assay principle. Cleavage of the highlighted amide bond leads to the liberation of 7-amino-4-carbamoylmethylcoumarin, which can be quantified through fluorescence readout. (C) P1 profiling with an Ac-Ala-hArg-Xaa-ACC library indicated a strong preference for aliphatic amino acid side chains in the P1 position. (D) P2 profiling using an Ac-Ala-Xaa-Leu-ACC library indicated a weak preference for basic amino acid side chains in the P2 position. (E) P3 profiling with an Ac-Xaa-hArg-Leu-ACC library indicated a strong preference for aromatic amino acid side chains in the P3 position. An asterisk (*) denotes that this peptide could not be obtained. Means \pm s.d. from triplicate measurements are shown.

protein that is threaded into its catalytic chamber by a cognate chaperone.^{16,18,19} It is currently unclear how ClpP is capable of processing nearly any amino acid sequence despite possessing only one type of active site.

Here, we systematically dissect the cleavage site specificities of three ClpP homologues using a tailored fluorogenic substrate library. Surprisingly, we found that the peptide products from ClpXP-mediated degradation reactions of endogenous substrates did not recapitulate the pronounced and conserved specificities we identified in the library screen. To reconcile these observations, we propose a mechanism in which the

barrel-shaped architecture of ClpP facilitates efficient cleavage of substrates even at nonpreferred amino acid sequences by bringing the substrate and active site into close proximity. Furthermore, the preferred substrates we were able to identify allowed us to profile human ClpP mutants associated with cancer and Perrault syndrome, thereby enabling the investigation of the catalytic activity of these enzyme variants.²⁰

RESULTS

Fluorogenic Substrate Library Screening Demonstrates Conserved Specificities. Three characteristic homo-

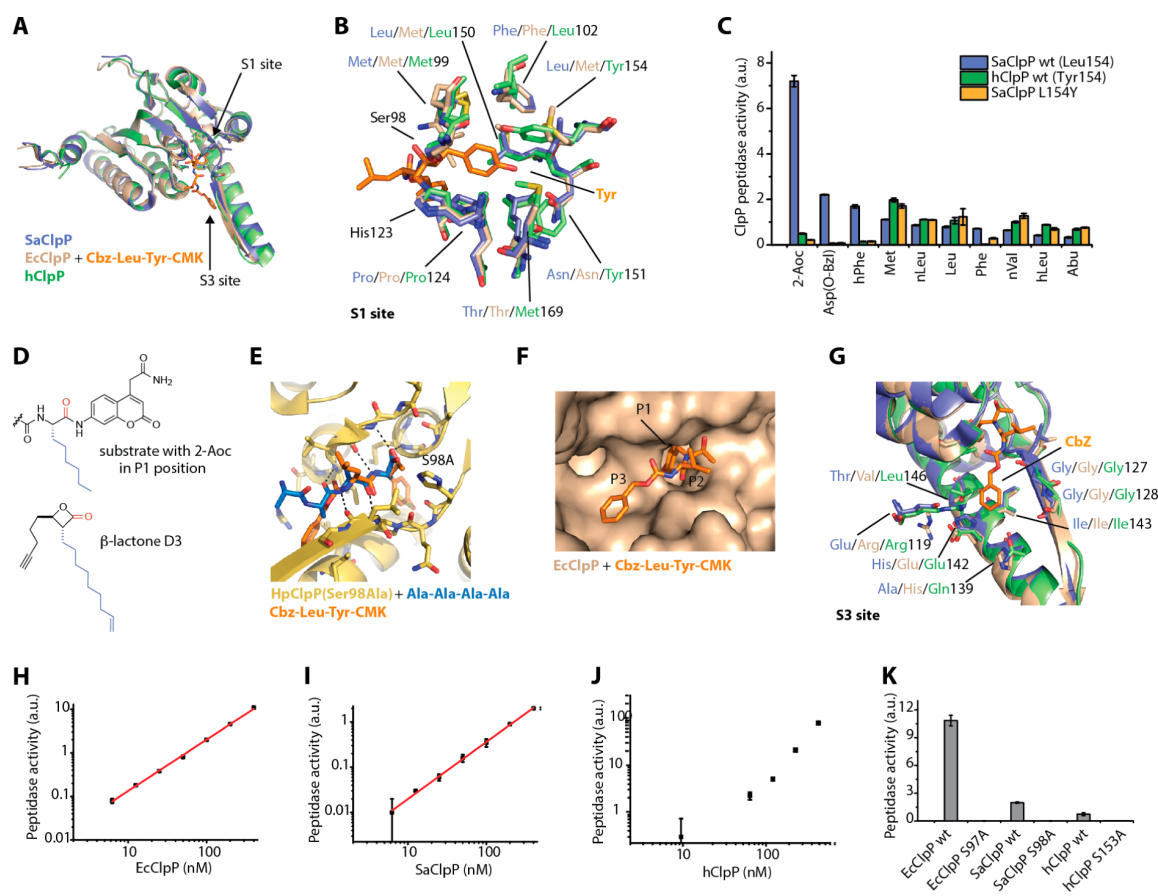


Figure 2. Improved substrates reflect S1 and S3 pocket characteristics. (A) Overlay of holo structures of SaClpP (PDB 3VSE) and hClpP (PDB 1TG6) with the structure of EcClpP in complex with Cbz-Leu-Tyr-CMK (PDB 2FZS). (B) Close-up view of the S1 binding pocket. Residues in the three enzymes are indicated according to the color code outlined in panel A. Residue numbers are given as in SaClpP for easy comparison. (C) Comparison of peptidase activities of indicated P1 library substrates for hClpP, SaClpP, as well as the hClpP-mimicking SaClpP mutant L154Y. (D) Structural similarity of the 2-Aoc P1 substrate and the previously identified β -lactone D3. The reactive carbonyl is highlighted in red; the alkyl chain occupying the S1 pocket is shown in blue. (E) Structure of *Helicobacter pylori* ClpP (Ser98Ala) in complex with a tetra-Ala-peptide (PDB 2ZL4). β -sheet interactions are indicated by black dotted lines. The position of the tetra-Ala-peptides agrees with the positioning of the CMK peptide in EcClpP. (F) Surface representation of EcClpP substrate binding cleft (PDB 2FZS). (G) Close-up view of the S3 binding pocket. (H) Enzyme-concentration-dependent activity of SaClpP with Ac-Ala-hArg-Leu-ACC. (I) hClpP showed no linear correlation of activity and protein concentration (measured with Ac-Phe(3,4-Cl₂)-hArg-Leu-ACC), in line with reduced stability of the active tetradecameric assembly.²⁵ (J) Observed activities are enzyme-dependent as shown with mutant enzymes lacking the catalytic serine. Means \pm s.d. from triplicate measurements are shown in all panels.

logues of the ClpP protease family were analyzed for individual peptide cleavage preferences: EcClpP from Gram-negative *Escherichia coli*, SaClpP from Gram-positive *Staphylococcus aureus*, and human mitochondrial ClpP (hClpP) as a eukaryotic representative. We iteratively optimized libraries of fluorogenic substrates so that individual members could be used in enzymatic assays. We started by screening a small fluorogenic (ACC-tagged) substrate library containing the general sequence Ac-Xaa-ACC against SaClpP at high enzyme concentrations. In line with previous reports on EcClpP from model peptide cleavage experiments,²¹ we found that SaClpP exhibits a preference for hydrophobic amino acids such as Leu, Met, Tyr, Ala, and Nle (norleucine, see Supporting Figure 1 for structures of all amino acids) at the P1 position of the substrate (Figure 1B). In order to systematically optimize the substrates, we next synthesized a library of ACC-labeled tripeptides with variation at the P2 position (Ac-Ala-Xaa-Leu-ACC), which contained both natural and unnatural amino acids (Xaa). We identified a preference for the basic amino acid homoarginine (hArg) at P2

of the tripeptide and therefore proceeded to synthesize a second-generation P1 library based on the scaffold Ac-Ala-hArg-Xaa-ACC. Additionally, a P3 library (Ac-Xaa-hArg-Leu-ACC) containing optimized P1 and P2 amino acids was synthesized. In each ACC-tagged tripeptide library, 19 natural amino acids (except Cys) and more than 40 unnatural amino acids were used. Each substrate was purified by HPLC, and its composition was confirmed by high-resolution mass-spectrometry (HR-MS; Supporting Tables 1–3). To unravel species-specific cleavage preferences, all 175 substrates were then tested against the panel of ClpP proteins under individually optimized assay conditions (Figure 1C–E, Supporting Figures 2–4).

From the extended P1 library screening, we identified a pronounced preference of hClpP for hydrophobic amino acids in this position (Figure 1C). The top 10 substrates featured Met, Leu, and Ala, as well as unnatural amino acids with aliphatic side chains such as Abu (ethyl moiety), nVal (propyl moiety), nLeu (butyl moiety), hLeu (isopentyl moiety), and 2-Aoc (hexyl moiety). These were also preferred by the bacterial

ClpP proteins, which additionally tolerated bulkier hydrophobic residues such as Phe, Trp, Tyr, and related unnatural amino acids such as hPhe (homophenylalanine), Phe(4-Cl) (4-chlorophenylalanine), Cha (cyclohexylalanine), and benzyl-protected amino acids (Ser, His, Asp, Glu; Supporting Figures 2B and 3B).

Most strikingly, 2-Aoc was by far the best P1 residue for bacterial ClpPs but not for hClpP. This difference can be explained by the diverging amino acid arrangements in the S1 pockets of these protein homologues, as illustrated by the overlay of holo-SaClpP and hClpP with a cocrystal structure of EcClpP covalently inhibited by a chloromethylketone (CMK; Figure 2A,B).²² hClpP has a smaller, but still largely hydrophobic, S1 pocket by virtue of three evolutionary mutations compared to bacterial ClpPs (T169M, N151Y, and L154Y for SaClpP → hClpP). We found that the L154Y mutation was sufficient to mimic the characteristics of hClpP. This mutation prevented SaClpP from cleaving more sterically encumbered P1 substrates (2-Aoc, hPhe, Asp(O-Bzl)), while those with smaller residues (Met, nLeu, Leu) were unaffected (Figure 2C).

Of note, the preference for substrates with aliphatic P1 residues is supported by previous studies with covalent inhibitors of the β -lactone-scaffold²⁴ (Figure 2D), whose aliphatic chains selectively address the hydrophobic SaClpP S1 pocket.²⁵ The binding mode of the CMK inhibitor to EcClpP as well as structures of *Helicobacter pylori* ClpP in complex with peptides²⁶ reveal a binding mode in which the substrate peptide packs against surrounding residues of the active site as an antiparallel β sheet (Figure 2E). These backbone-mediated interactions explain why the absence of a P1 side chain is well tolerated, as seen by the efficient turnover of the Gly P1 substrate.

Analysis of the P2 library screen demonstrated a consistent preference for basic amino acids such as Arg, Lys, hArg, Lys(TFA), and Arg(NO₂), although many other amino acids were also tolerated (Figure 1D, Supporting Figures 2C and 3C). Only few natural amino acids were disfavored, among them Asp, Phe, Pro, Tyr, and Val. Analogously to the proteasome, ClpP does not exhibit an S2 pocket, and so the P2 side chain extends out into the solvent-filled proteolytic chamber (Figure 2F).

Screening of the P3 library revealed an unexpectedly strong preference of hClpP for Phe(3,4-Cl₂) (3,4-dichlorophenylalanine), followed by Phe(4-Cl) (4-chlorophenylalanine), Phe, and Tyr (Figure 1E). However, many other amino acids were tolerated as well, albeit at lower efficiencies. The same holds true for the bacterial ClpPs, which were rather nondiscriminating and showed only weak preference for hydrophobic amino acids such as Ile and Phe for SaClpP and chlorinated phenylalanines for EcClpP (Supporting Figures 2D and 3D). Particularly noteworthy was the discrimination against basic amino acids with positive charges in the P3 position. This can be rationalized structurally through the hydrophobic environment in the S3 pockets (Figure 2F,G). However, a cation- π interaction involving Arg119 in hClpP, which is pointing directly onto the substrate phenyl ring at a 3.5 Å distance in the overlay, is a possible explanation for the preferred accommodation of aromatic residues in this pocket.

A small subset of D-configured amino acids was included in the library; however, these were consistently not tolerated at the P1 and P2 positions, and only weak activity was observed for EcClpP if they were placed in the P3 position. These results

build upon the previous finding that D-Met is not tolerated in the P1 position.²¹ Overall, this data can be rationalized by the binding mode of substrates in the ClpP active site since side chains of D-configured amino acids in the P1 position would clash with the backbone of Gln124/Pro125, which would in turn push the catalytic His123 out of the charge-relay system. With the substrate adhered to the inner wall of the proteolytic chamber, side chains of D-configured amino acids in the P2 and P3 positions are similarly incompatible, as they would clash with the backbones of Gly69/Ser70 and Leu126/Gly127, respectively.

Improved Substrates Lend Insights into Disease-Associated Mutant hClpP Activity. We next characterized the best substrate for each ClpP homologue respectively and confirmed enzyme-concentration-dependent turnover (Figure 2H–J). We verified that measured signals arise from catalytic activity through the use of active site mutant enzymes as controls (Figure 2K). Furthermore, the catalytic efficiency ($k_{\text{cat}}/K_{\text{M}}$) was determined for selected substrates (Table 1).

Table 1. Catalytic Efficiencies ($k_{\text{cat}}/K_{\text{M}}$) Determined for Indicated Substrates and Enzymes^a

substrate	SaClpP	EcClpP	hClpP
Ac-Ile-hArg-Leu-ACC	72.7 ± 10.6	169.6 ± 9.7	29.2 ± 5.2
Ac-Ala-hArg-Met-ACC	30.0 ± 4	237.9 ± 21	58.4 ± 4.2
Ac-Phe(3,4-Cl ₂)-hArg-Leu-ACC	37.3 ± 4.2	361.9 ± 39.4	251.9 ± 30.1
Ac-Ala-hArg-Leu-ACC	21.7 ± 3.1	177.6 ± 7.2	28.4 ± 3.9
Suc-Leu-Tyr-AMC	0.78 ± 0.2	56.9 ± 4.9	1.0 ± 0.2

^aThe first four substrates are described in this study, while Suc-Leu-Tyr-AMC is commercially available. Values are given in M⁻¹ s⁻¹ as mean ± s.d. and were derived from triplicate experiments. See the Methods section for assay conditions.

Interestingly, an up to 200-fold increase in turnover was observed when compared to the standard substrate Suc-Leu-Tyr-AMC (Table 1, Figure 3A). In addition, we carried out Hofmeister series salt experiments in order to confirm that the peptidase activity of ClpP arises from tetradecameric complexes (Figure 3B).²⁷ Consistent with this report, salts with high Hofmeister strength, which cause a “salting-out” effect and thereby stabilize multimeric complexes, led to increased activity, whereas salts with low Hofmeister strength, which cause a “salting-in” effect and thereby destabilize multimeric complexes, led to reduced activity. It is noteworthy that the addition of citrate in the reaction buffer can boost peptidase activity of hClpP and SaClpP activity further by roughly 20-fold.

We next sought to apply our optimized substrates to the investigation of disease-related hClpP mutant proteins. A recent study reported that Perrault syndrome characterized by ovarian dysgenesis in females and by sensorineural hearing impairment in males and females can be caused by recessive mutations leading to T145P and C147S in hClpP (Figure 4A).²⁰ However, the effects of these mutations at the protein level have not been studied, and the molecular disease mechanism remains elusive. Additionally, we found several cancer-patient-derived point mutations of CLPP in the COSMIC cancer genome database (Figure 4B). The mutated residues were predominantly found in the outer and upper part of the hClpP head domain, leaving the catalytic site, the substrate binding pocket, and the inter-ring contacts intact (Figure 4A,B). For Perrault syndrome, a causative relationship between mutations

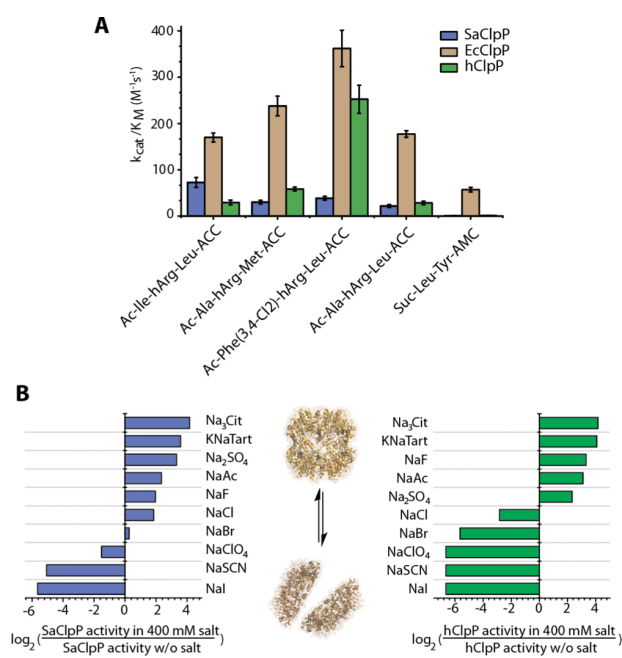


Figure 3. Improved ClpP peptidase activity. (A) Comparison of catalytic efficiencies (k_{cat}/K_M values) of four identified substrates and the commercially available ClpP substrate Suc-Leu-Tyr-AMC for the three ClpP proteins included in this study. See Table 1 for values. (B) Analysis of SaClpP and hClpP activity in buffers supplemented with indicated salts. The order of salts inferred from stimulation/reduction of catalytic activity corresponds to the Hofmeister series of anions as indicated by the cartoon. See Supporting Figure 5 for salt-concentration dependent measurements. Means \pm s.d. from triplicate measurements are shown.

in *CLPP* and the disease has been established,²⁰ whereas it is not clear whether the cancer-patient-derived *CLPP* mutations are mere passenger mutations or contribute to tumor progression. To explore the effects of these mutations, we recombinantly expressed all corresponding mutant proteins, some with much lower yields, suggestive of a destabilized fold (e.g., T145P, R167C, S173Y; see Supporting Table 4 for protein melting temperatures). With an optimal substrate for hClpP in hand, we were able to determine the peptidase activities of these mutants relative to wild-type hClpP. Most of the proteins displayed a similar or slightly reduced activity when compared to wild-type (Figure 4C), which excludes lack of catalytic activity from the list of possible Perrault disease mechanisms. The location of Thr145 and Cys147 has previously given rise to the speculation that mutation of these residues could diminish binding to ClpX.²⁰ We thus profiled all mutations in a proteolysis assay in which we monitored the fluorescence of SsrA-tagged GFP in the presence of hClpP and EcClpX,²⁸ taking into account that human ClpX does not interact with bacterial SsrA tags.²⁹ All proteins, except for the A131V mutant, showed similar turnover of GFP. The A131V mutant also exhibited completely abolished peptidase activity, likely mimicking a heterozygous *CLPP* knockout phenotype in the associated cancer patient. Importantly, both Perrault mutations showed unfolding activity indistinguishable from wild type protein, indicating a functionally competent ClpX:ClpP interaction. Our data therefore point toward a different disease mechanism such as defective transport to mitochondria, changes in protein interactions, or reduced protein levels due to decreased protein stability (Figure 4E). Notably, T213I and R228H proteins showed drastically reduced peptidase activity levels, but normal proteolysis activity. This is in agreement with a recent report in which different mutations rendered SaClpP peptidase-inactive, but still susceptible to ClpX-mediated, allosteric activation.¹⁴

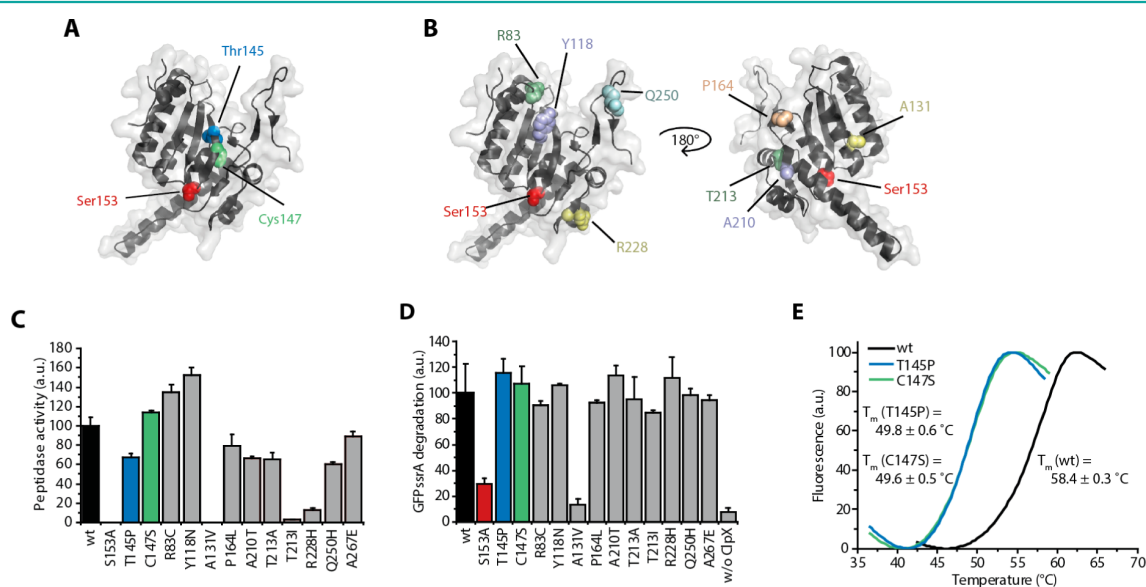


Figure 4. Profiling of disease-associated hClpP mutants. (A) Cartoon representation of an hClpP monomer with mutated residues occurring in Perrault syndrome depicted in blue and green. (B) Cartoon representations of hClpP monomers with residues found mutated in cancer genomes highlighted in color. (C) hClpP peptidase activity data with 400 nM enzyme concentration using 100 μ M of the optimized substrate Ac-Phe(3,4-Cl₂)-hArg-Leu-ACC. (D) GFP-SsrA degradation by a complex of hClpP with *E. coli* ClpX. (E) Thermal-shift assay curves of Perrault mutants in comparison to wild-type hClpP, indicating reduced protein stability. Means \pm s.d. from triplicate measurements are given in all panels.

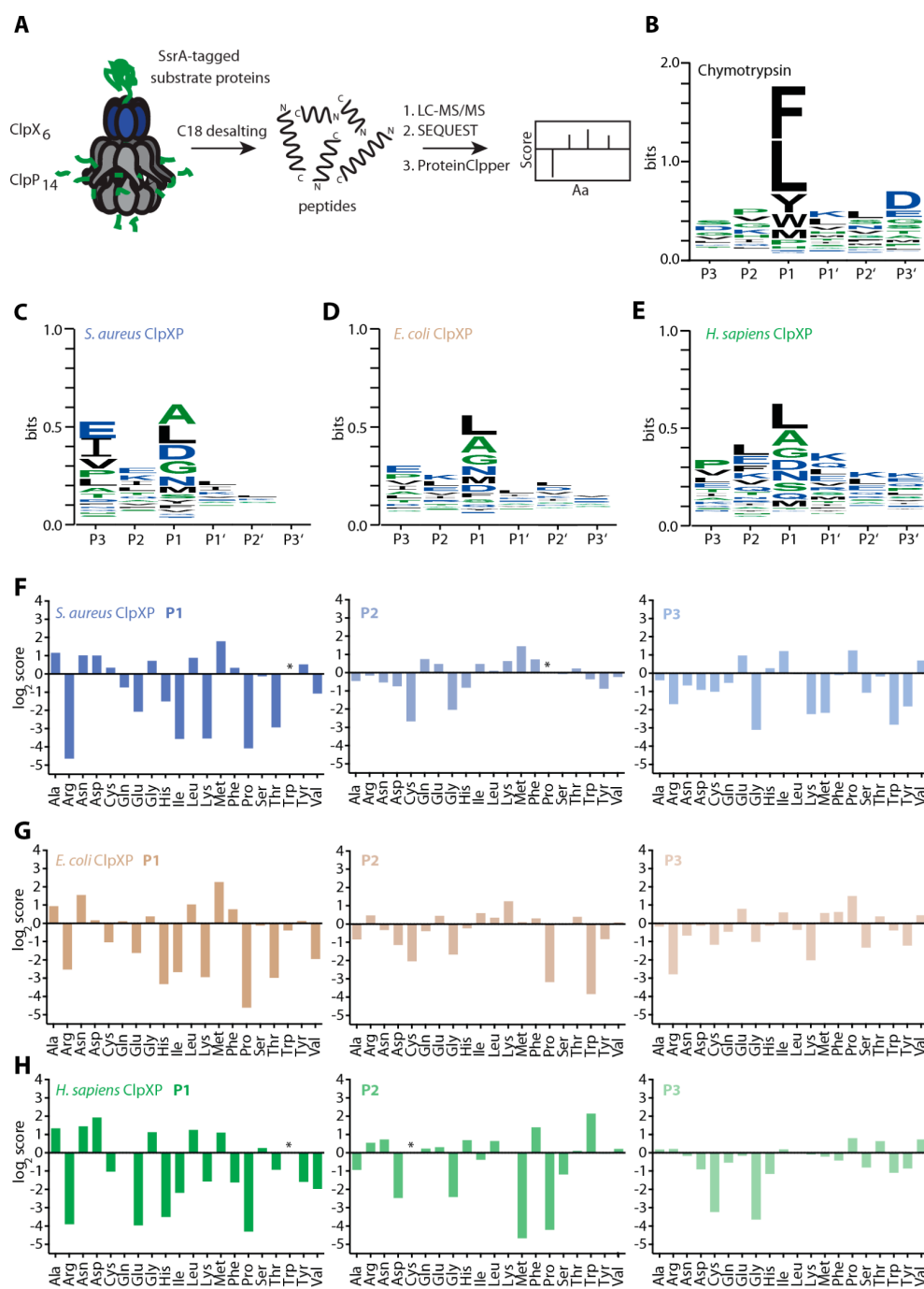


Figure 5. Analysis of ClpXP product peptides. (A) Schematic depiction of the workflow. (B, C, D, E) Weblogo representation of amino acids around the cleavage sites of all analyzed substrates for SaClpP, EcClpP, and hClpP with Chymotrypsin as a reference. Of note, the Weblogo analysis is based on the identification of a consensus sequence and thus cannot normalize the output for the amino acid abundance in the respective substrate. (F, G, H) Analysis of the occurrence of amino acids around cleavage sites with normalization to the respective amino acid abundances. A $\log_2(\text{score})$ of 0 denotes that the respective amino acid was found as expected from a random cleavage. $\log_2(\text{score}) > 0$ shows a preference for this amino acid; $\log_2(\text{score}) < 0$ shows that the amino acid at this position occurs less often than expected from a random cleavage. An asterisk (*) denotes that no peptide with the respective amino acid in this position was found; hence the score was computed to be 0. Please refer to the [Methods Section](#) for a more detailed explanation of the score. See [Supporting Figure 9](#) for an analysis of the prime site positions.

Little is known about the physiological roles of ClpP and its substrate proteins. Loss of ClpP has been linked to a longevity phenotype in the fungal model organism *P. anserina*.³⁰ Furthermore, ClpP has been implicated in the mitochondrial signaling pathway for unfolded proteins in *C. elegans*, where

ClpP-derived peptides exported from the mitochondrial matrix by HAF-1 act as signaling molecules in the cytosol to trigger a transcriptional response in the nucleus.^{31,32} However, nothing is known so far about the nature of these peptides or their receptors, and whether a similar pathway exists in mammals.

Table 2. LC-MS/MS analysis of ClpXP substrate degradation assays. Table entries specify the sequence coverage, the number of unique peptides and the number of peptide spectral matches obtained from the respective samples. Only peptides identified with high confidence were included in the analysis (see the [Methods Section](#) for details)

organismic system	proteins found in MS analysis	SaClpP + SaClpX	EcClpP + EcClpX	hClpP + hClpX
<i>E. coli</i>	EcRpoS		66%/106/191	
	EcFtsZ		83%/138/215	
	EcTnaA		60%/130/200	
	EcClpX		62%/116/434	
<i>S. aureus</i>	creatine kinase		55%/50/163	
	SaUvrA	57%/58/234		
	SaMetK	82%/184/314		
	SaClpC	81%/225/338		
	SaGudB	51%/55/72		
	SaClpX	85%/152/598		
	SaClpP	33%/11/28		
<i>H. sapiens</i>	creatine kinase	60%/58/203		
	α -casein S1			83%/75/217
	α -casein S2			74%/44/100
	β -casein			89%/82/230
	κ -casein			65%/40/64
	hClpX			65%/101/158
	hClpP			69%/24/32
	creatine kinase			27%/11/15

With the *in vitro* specificities of ClpP proteases established, we wondered whether these specificities also translated to proteolysis by the ClpXP complex, thus allowing for defined downstream signaling. We therefore carried out a thorough analysis of cleavage sites during ClpXP proteolysis.

Cleavage Site Analysis with Endogenous ClpXP Substrates. Previous proteomic studies using an inactive ClpP mutant to trap substrate proteins inside of the degradation chamber have led to the identification of endogenous substrates for *E. coli*¹⁶ and *S. aureus*¹⁸ ClpP. We selected seven of these substrate proteins from diverse classes ([Supporting Table 5](#)) including the metabolic enzymes TnaA, GudB, and MetK; the DNA damage repair protein UvrA; the transcription factor RpoS; the quality control protein ClpC; and a member of the bacterial cytoskeleton, FtsZ. We confirmed that our selection was representative in terms of its amino acid composition, both when compared to the entire sets of substrates as well as to proteomic distributions ([Supporting Figure 6](#)). Since it was unclear how well these proteins are recognized by ClpXP *in vitro* and whether accessory substrate adapters are necessary for recognition, we fused a C-terminal SsrA tag to the substrate proteins to facilitate efficient *in vitro* processing by ClpXP complexes ([Figure 5A](#)).¹⁷ We expressed and purified these proteins from the triple-protease deficient *E. coli* strain KY2266³³ and confirmed their ClpXP-dependent degradation. Peptide products in desalted ClpXP degradation reactions were then analyzed by nano-LC-MS/MS ([Table 2](#)). In order to exclude any bias in the detected cleavage patterns arising from assay conditions or sample handling, we performed a number of control experiments ([Supporting Figure 7](#)). We confirmed that the obtained size distribution ([Supporting Figure 7A,B](#)) and cleavage pattern ([Supporting Figure 7C,D](#)) are independent of the substrate concentration used in the assay. Moreover, neither dansylation of the reaction mixture prior to desalting nor a separation gradient starting at 1% organic phase instead of 4% changed the cleavage pattern, indicating only minimal loss of small hydrophilic peptides during capture on C18 material

([Supporting Figure 7E,F](#)). Finally, we demonstrated the robustness of the mass spectrometric method with replicate experiments ([Supporting Figure 7G,H](#)).

As a positive control, we analyzed peptide products from digestion reactions using chymotrypsin. We used a low-specificity chymotrypsin variant with preferred cleavage after Trp, Tyr, Phe, and, to a lesser extent, Leu and Met. Notably, this protease was chosen because of its preference for hydrophobic amino acids in the P1 position, similar to ClpP proteases (e.g., Met, Leu, Tyr, Phe, and Trp being the TOP5 preferred P1 substrates for EcClpP, and Met, Leu, Tyr, and Phe being the TOP4 preferred P1 substrates for SaClpP). In order to comprehensively visualize the results, we parsed the array of obtained peptide sequences through the Weblogo³⁴ service ([Figure 5B](#)). This algorithm is suitable for visualizing consensus sequences from multiple sequence alignments. In this representation, the overall height of the bar reflects the degree of conservation at the position, and more frequent amino acids are indicated by a larger character height. While this analysis nicely displayed the TOP5 preferred amino acids in the P1 site, it did not reflect the correct order of the amino acids (e.g., Leu seemed more preferred than Trp according to the Weblogo representation, although this is due to different abundances, see [Supporting Figure 6](#)). Moreover, this visualization lacked information regarding nonpreferred amino acids, i.e. whether they occurred randomly in the sequences or whether they were avoided around the cleavage site. We therefore developed a customized algorithm for the analysis of peptide cleavage patterns termed ProteinClpper, which is available online (www.oc2.ch.tum.de). In this analysis, data are evaluated by interpreting each observed peptide as two cleavage events (or one event if the peptide was N- or C-terminal). Next, an analysis of the occurrences of amino acids around the cleavage sites is performed. Importantly, the incidence of each amino acid at a specific position is normalized to the abundance of this amino acid in the respective protein, yielding a sequence-independent score *S* (see [Methods Section](#) for details) which can be interpreted as follows: For $\log_2(S)$ values larger than 0,

the amino acid is enriched at a given position, while $\log_2(S)$ values smaller than 0 indicate depletion. When we analyzed the chymotrypsin-derived set of peptides (Supporting Figure 8), we obtained a strong preference in the P1 position for Phe, Trp, and Tyr ($\log_2(S) > 2$) and a weak preference for Leu and Met ($\log_2(S) > 1$), which accurately reflected the reported specificities. Moreover, this representation contains information on which amino acids were observed as expected by random cleavage (e.g., His), which amino acids showed a reduced occurrence (e.g., Lys and Val), and which amino acids were completely depleted (e.g., Gly and Arg). Finally, we confirmed that the results are similar for the analysis of all peptide spectral matches, a list of only unique peptides and all peptide spectral matches weighted according to the SEQUEST cross-correlation value.

We then used this algorithm to analyze the peptide products generated in ClpXP-mediated substrate degradation reactions with all endogenous substrates and casein proteins. Substrate proteins were purified and subsequently *in vitro* digested separately by SaClpP + SaClpX, EcClpP + EcClpX, and hClpP + hClpX. Control reactions were carried out in the absence of ClpP. In total, high sequence coverage (mean of all substrates: 72%) arising from 1660 unique peptides and 3806 peptide spectral matches were observed (Table 2, see Figure 5C–E for Weblogo representations, Figure 5F–H for ProteinClpper scores for nonprime site positions, and Supporting Figure 9 for prime site positions), enabling the analysis of 3528 unique cleavage reactions (1076 for EcClpXP, 1705 for SaClpXP, and 747 for hClpXP).

Consistent with studies on other proteases,³⁵ the overall P1 position scores showed the largest discrimination between amino acids. To compare the preferences obtained from the mass spectrometric analysis to the preferences determined by fluorogenic substrate screening, we focused on those amino acids that were enriched in the P1 positions ($\log_2(S) > 0$). This subset consisted of nine amino acids for both SaClpP and EcClpP (seven amino acids for hClpP), with seven of these also present in the TOP9 P1 fluorogenic substrates for SaClpP and EcClpP (six of the seven amino acids with enrichment in P1 for hClpP were also found in the TOP7 list of P1 fluorogenic substrates of hClpP). However, apart from the conserved preference for Met, which topped both lists for SaClpP and EcClpP, the order of amino acids in these groups was largely different. It is apparent that the degree of discrimination observed in the fluorogenic substrate screening was much higher than the discrimination observed in the proteolysis reaction. As a specific example, Met and Leu are clearly preferred in the fluorogenic substrate analysis, while they appear only as one of several equally well tolerated amino acids at the P1 position in the proteolysis analysis. It is striking that multiple cleavage events with Asp and Asn in the P1 position were detected for all three ClpPs, although these amino acids showed barely above background turnover in the P1 substrate library. Moreover, several cleavage events in the SaClpXP reactions could be detected after Glu, Gln, His, and Thr, for which the corresponding P1 substrate peptides showed no detectable turnover. Similarly, sets of amino acids could be identified for EcClpP (Arg, His, Ile, Thr) and hClpP (Gln, His, Lys, Thr) that served as P1 positions in cleavage events whereas the corresponding P1 substrates showed no detectable turnover. Remarkably, cleavage events for almost all 20 natural amino acids in the P1 position were detected for each of the three ClpP proteins. Contrary to chymotrypsin, where some

amino acids at the P1 position were found to be prohibitory for cleavage, ClpP proteases thus demonstrated the capability to cleave after every natural amino acid. Moreover, preferences are less pronounced than for chymotrypsin, which can be seen from the lower Weblogo conservation scores for P1 residues (~ 1.7 bit for chymotrypsin versus ~ 0.6 bit for ClpXP complexes) as well as lower $\log_2(S)$ values for P1 residues (e.g., no amino acid apart from Met in EcClpXP displayed an equally strong preference of $\log_2(S) > 2$ as the residues preferred by chymotrypsin).

Specificities for the prime site (P1') are harder to probe with peptide libraries³⁶ but can readily be obtained from mass spectrometric analysis.³⁵ In the P1' position, Pro was consistently depleted, a phenomenon which is frequently observed for proteases.³⁷ However, apart from reduced occurrence of Cys in both P2' and P3', no other conserved specificities were found (see the Methods Section for an additional analysis of the prime site data).

Our data on the P2 and P3 positions confirmed the rather broad specificities determined in the library analysis. However, several amino acids that showed no or almost no turnover in the P2 screenings (e.g., Asp, Gly, Phe, Pro, Tyr, Val, despite being in a substrate with an optimal P1 residue) were observed or even enriched in P2 positions of cleavage events. Similarly, basic amino acids such as Lys and Arg, which were disfavored in the P3 library screenings, could now be observed in the P3 position of cleavage events (with depletion for EcClpP and SaClpP, and no effect for hClpP). A slight preference of SaClpP for Ile in the P3 position was observed in accordance with the fluorogenic P3 substrate results; however, neither of the other preferred residues (Phe, Tyr) showed enrichment at the P3 position in the product peptides.

Collectively, we found that the high degree of cleavage site specificity obtained in the fluorogenic peptide screening did not translate to the proteolysis reaction.

DISCUSSION

Cleavage specificity underlies the function of many proteolytic enzymes, and the study of their substrate cleavage motifs has greatly facilitated the investigation of associated biological roles.³⁵ Such specificity generally relies on the recognition of amino acid patterns around the scissile bond. Since many proteases have freely accessible active sites, a productive interaction that leads to cleavage is an important prerequisite for controlled substrate turnover. ClpP proteases have evolved as self-compartmentalizing enzymes that shield their active sites from the cytosol and the mitochondrial matrix by placing them inside a barrel-shaped scaffold. Substrate selection and entry into the barrel are mediated by interacting chaperones and substrate adaptor proteins and are thus structurally uncoupled from catalysis. This organization plays a crucial role in preventing uncontrolled intracellular proteolysis but requires a protease subunit that is capable of digesting many different sequences to small peptides. The cleavage specificity of the EcClpP protease was studied nearly 20 years ago through the digest of model peptides.^{21,38} These early studies uncovered a preference for Met and Leu in the P1 position with aromatic residues (Tyr, Trp, Phe) also being tolerated. In the context of these peptides, no cleavage occurred after Glu, Ser, Lys, and Ile, raising the question of how promiscuous ClpP really is and how a potentially restricted specificity affects its role in protein breakdown.

Here, we present an in-depth investigation of the cleavage site specificities of three representative members of the ClpP family of proteases. Our analysis relies on the profiling of tailored fluorogenic substrate libraries as well as on mass-spectrometric analysis of endogenous bacterial peptide products. Our data confirm a distinct preference for Met and Leu in the P1 position, which we found to be conserved. Consistent with structural and mutational analyses of the S1 pockets, bacterial ClpPs also tolerated bulkier aromatic residues in P1, such as Phe, Tyr, and Trp, while hClpP features a smaller S1 pocket that prevented efficient turnover of these substrates. In line with these insights, Tyr and Phe were found to be enriched in the P1 position of SaClpP and EcClpP-derived peptide products, while they are depleted in hClpP-derived peptides.

Our data show that some trends discovered in the library screen are also reflected in the peptide cleavage pattern of natural substrates. Moreover, the size distribution of peptide products of endogenous substrates matched those obtained in earlier studies from casein and GFP.^{39,40} Surprisingly, we found that the pronounced cleavage specificity patterns identified in the library screen were strongly attenuated in the more physiological proteolysis scenario. Strikingly, we were able to identify cleavage events after every natural amino acid. The degree of preference measured as $\log_2(S)$, where 0 denotes an unspecific cleavage, exceeded 2 for only one amino acid and was found to be between 0 and 1 for most enriched amino acids, indicating an up to 2-fold enrichment over random cleavage. This is in contrast to a low-specificity variant of chymotrypsin which displays a similar preference for aliphatic and aromatic residues in the P1, but reaches $\log_2(S) > 3$ in substrate degradation reactions (indicating >8-fold enrichment over background). Moreover, characteristics identified in the P2 and P3 screen are not retained during proteolysis.

Several factors contribute to this finding and help to reconcile these observations.

First, the recognition of the substrate peptide is also mediated by backbone interactions with the enzyme through the formation of an antiparallel β -sheet. This binding mode explains the efficient processing after Gly, an amino acid that lacks a P1 side chain. Cleavage after Gly is unusual for many proteases with the exception of glycyI endopeptidase as well as SUMO and ubiquitin-specific isopeptidases (SENPs and USPs, respectively) where terminal di-Gly motifs lie in a narrow channel next to the active site.^{41–43}

Second, the library screens were performed using only one peptide scaffold, whereas proteolysis occurred in many different sequence contexts. One could argue that different cleavage determinants are present in different sequence contexts.³⁵ However, different model peptides have revealed the same P1 specificity, which is in line with a recent report on the cleavage specificities of mycobacterial ClpP where a pooled library was used with a similar finding.⁴⁰ Moreover, structural⁴⁴ and biochemical¹⁴ evidence exists to support the notion that occupation of the regulatory hydrophobic pocket of ClpP does not change the active site geometry. This implies that the cleavage site specificity is unaffected by ClpX binding.

Third, the proteolysis reaction occurs inside a small chamber, which sets ClpP proteolysis apart from many other protease environments. If one assumes that the proteolytic chamber has an ellipsoid shape with an inner diameter of 42 Å and an inner height of 72 Å, the effective concentration of active sites in this cavity would be ~350 mM. A similar argument for efficient

processing has previously been introduced in the literature with a spherical assumption.⁴⁵ Although the exact mechanism of peptide product release is still unknown, it is generally believed that peptides remain in the chamber until they are processed to be small enough to leave the chamber by means of passive diffusion. Moreover, an electron microscopy study has shown that the chamber of an inactive ClpP can be filled extensively by substrate.⁴⁶ Taken together, these results indicate that the effective local concentrations of both the enzyme and the substrate in the chamber are in the high millimolar range. According to this model, these high concentrations would give rise to many cleavage reactions and, in turn, would contribute to the overruling of active site encoded cleavage site specificity.

In summary, our data demonstrate that ClpP proteases feature a conserved mechanism in which their active-site-encoded substrate specificity is constantly attenuated during proteolysis. This finding calls for a re-evaluation of the assumption that ClpP proteases are unspecific proteases *per se*. Instead, both the peptide binding mode and the protease scaffold together likely facilitate the efficient cleavage of diverse sequences.

Furthermore, the inclusion of unnatural amino acids⁴⁷ in the fluorogenic substrate library led to the identification of a set of substrates that greatly surpasses the capacity of the previously used commercially available substrate Suc-Leu-Tyr-AMC. This enabled us to study the catalytic activity of human ClpP localized in the mitochondria. The activity of ClpP proteins has a distinct pH dependence, and bacterial homologues from *S. aureus*¹³ and *Listeria monocytogenes*⁴⁸ were shown to exhibit optimal activity at pH 7.0. Consistent with an elevated pH in the mitochondrial matrix of around 7.8,⁴⁹ hClpP showed much reduced activity at pH 7.0 and optimal peptidase activity between pH 8 and 9 (Supporting Figure 4A), demonstrating functional adaptation. Moreover, we were able to profile a panel of disease-associated hClpP mutant proteins. Most importantly, we showed that the catalytic activity of Perrault-syndrome-associated hClpP mutants is unaffected both in peptidolytic as well as in proteolytic assays, thus arguing for a different disease mechanism than suspected previously.²⁰

CLPP expression is regulated upon mitochondrial imbalance⁵⁰ and elevated during a mitochondrial unfolded protein response,⁵¹ demonstrating a functional role for hClpP in mitochondrial proteostasis. In addition, hClpP inhibition has recently been suggested as a therapeutic strategy for human acute myeloid leukemia.⁵² Although studies on the roles of eukaryotic ClpP homologues in *C. elegans*³¹ and *P. anserina*³⁰ exist, little is known about human ClpP function and its substrates.⁵³ The insights into the molecular mechanism of substrate turnover as well as customized tools introduced herein will assist in future efforts to uncover the biological functions of ClpP.

■ ASSOCIATED CONTENT

📄 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acschembio.5b00757.

Detailed experimental procedures, supporting Figures 1–9, and supporting Tables 1–5 (PDF)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: marcin.drag@pwr.edu.pl.

*E-mail: stephan.sieber@mytum.de.

Present Address

§Medical Research Council Laboratory of Molecular Biology, Francis Crick Avenue, CB2 0QH Cambridge, United Kingdom.

Author Contributions

||These authors contributed equally to this study.

Notes

The authors declare the following competing financial interest(s): M.G., M.P., A.D., M.Dr., and S.A.S. are named inventors on a patent application covering new fluorogenic substrates for ClpP. All other authors declare no conflict of interest.

ACKNOWLEDGMENTS

We are grateful for funding from the Deutsche Forschungsgemeinschaft, SFB 749, SFB 1035, CIPS^M, and the European Research Council (start up grant 250924-antibacterials). The work was supported for M.Dr. by the Foundation for Polish Science and a statutory activity subsidy from the Polish Ministry of Science and Higher Education for the Faculty of Chemistry at Wrocław University of Technology. M.S. was supported by the German National Academic Foundation. We thank E. Weber-Ban (ETH Zürich) for the kind gift of the EcClpX expression vector, D. Ron (University of Cambridge) and C. Haynes (Memorial Sloan Kettering Cancer Center) for discussions, and S. Gottesman (NIH Bethesda) for providing strain SG1146a. We thank N. Bach and E. Kunold for help with mass spectrometry and M.W. Hackl for providing SaClpP. Furthermore, we thank K. Adatia, K. Bäuml, B. Cordes, J. Fauser, A. Kratzert, A. Fux, L. Meixner, F. Traube, and M. Wolff for help with experiments and technical assistance and A. Hoegl for critical evaluation of the manuscript.

REFERENCES

- (1) Peunte, X. S., Sanchez, L. M., Overall, C. M., and Lopez-Otin, C. (2003) Human and mouse proteases: a comparative genomic approach. *Nat. Rev. Genet.* 4, 544–558.
- (2) Drag, M., and Salvesen, G. S. (2010) Emerging principles in protease-based drug discovery. *Nat. Rev. Drug Discovery* 9, 690–701.
- (3) Harris, J. L., Backes, B. J., Leonetti, F., Mahrus, S., Ellman, J. A., and Craik, C. S. (2000) Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries. *Proc. Natl. Acad. Sci. U. S. A.* 97, 7754–7759.
- (4) Deu, E., Verdoes, M., and Bogoy, M. (2012) New approaches for dissecting protease functions to improve probe development and drug discovery. *Nat. Struct. Mol. Biol.* 19, 9–16.
- (5) Rawlings, N. D., Waller, M., Barrett, A. J., and Bateman, A. (2014) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* 42, D503–509.
- (6) Lupas, A., Flanagan, J. M., Tamura, T., and Baumeister, W. (1997) Self-compartmentalizing proteases. *Trends Biochem. Sci.* 22, 399–404.
- (7) Nussbaum, A. K., Dick, T. P., Keilholz, W., Schirle, M., Stevanovic, S., Dietz, K., Heinemeyer, W., Groll, M., Wolf, D. H., Huber, R., Rammensee, H. G., and Schild, H. (1998) Cleavage motifs of the yeast 20S proteasome beta subunits deduced from digests of enolase 1. *Proc. Natl. Acad. Sci. U. S. A.* 95, 12504–12509.
- (8) Sauer, R. T., and Baker, T. A. (2011) AAA+ Proteases: ATP-Fueled Machines of Protein Destruction. *Annu. Rev. Biochem.* 80, 587–612.
- (9) Kress, W., Maglica, Ž., and Weber-Ban, E. (2009) Clp chaperone–proteases: structure and function. *Res. Microbiol.* 160, 618–628.
- (10) Katayama-Fujimura, Y., Gottesman, S., and Maurizi, M. R. (1987) A multiple-component, ATP-dependent protease from *Escherichia coli*. *J. Biol. Chem.* 262, 4477–4485.
- (11) Ortega, J., Singh, S. K., Ishikawa, T., Maurizi, M. R., and Steven, A. C. (2000) Visualization of substrate binding and translocation by the ATP-dependent protease, ClpXP. *Mol. Cell* 6, 1515–1521.
- (12) Wang, J., Hartling, J. A., and Flanagan, J. M. (1997) The Structure of ClpP at 2.3 Å Resolution Suggests a Model for ATP-Dependent Proteolysis. *Cell* 91, 447–456.
- (13) Gersch, M., List, A., Groll, M., and Sieber, S. A. (2012) Insights into the structural network responsible for oligomerization and activity Of the bacterial virulence regulator caseinolytic protease P (ClpP). *J. Biol. Chem.* 287, 9484–9494.
- (14) Gersch, M., Famulla, K., Dahmen, M., Gobl, C., Malik, I., Richter, K., Korotkov, V. S., Sass, P., Rubsamens-Schaeff, H., Madl, T., Brotz-Oesterhelt, H., and Sieber, S. A. (2015) AAA+ chaperones and acyldepsipeptides activate the ClpP protease via conformational control. *Nat. Commun.* 6, 6320.
- (15) Gribun, A., Kimber, M. S., Ching, R., Sprangers, R., Fiebig, K. M., and Houry, W. A. (2005) The ClpP double ring tetradecameric protease exhibits plastic ring-ring interactions, and the N termini of its subunits form flexible loops that are essential for ClpXP and ClpAP complex formation. *J. Biol. Chem.* 280, 16185–16196.
- (16) Flynn, J. M., Neher, S. B., Kim, Y. I., Sauer, R. T., and Baker, T. A. (2003) Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. *Mol. Cell* 11, 671–683.
- (17) Gottesman, S., Roche, E., Zhou, Y., and Sauer, R. T. (1998) The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev.* 12, 1338–1347.
- (18) Feng, J., Michalik, S., Varming, A. N., Andersen, J. H., Albrecht, D., Jelsbak, L., Krieger, S., Ohlsen, K., Hecker, M., Gerth, U., Ingmer, H., and Frees, D. (2013) Trapping and proteomic identification of cellular substrates of the ClpP protease in *Staphylococcus aureus*. *J. Proteome Res.* 12, 547–558.
- (19) Vass, R. H., and Chien, P. (2013) Critical clamp loader processing by an essential AAA+ protease in *Caulobacter crescentus*. *Proc. Natl. Acad. Sci. U. S. A.* 110, 18138.
- (20) Jenkinson, E. M., Rehman, A. U., Walsh, T., Clayton-Smith, J., Lee, K., Morell, R. J., Drummond, M. C., Khan, S. N., Naem, M. A., Rauf, B., Billington, N., Schultz, J. M., Urquhart, J. E., Lee, M. K., Berry, A., Hanley, N. A., Mehta, S., Cilliers, D., Clayton, P. E., Kingston, H., Smith, M. J., Warner, T. T., University of Washington Center for Mendelian, G., Black, G. C., Trump, D., Davis, J. R., Ahmad, W., Leal, S. M., Riazuddin, S., King, M. C., Friedman, T. B., and Newman, W. G. (2013) Perrault Syndrome Is Caused by Recessive Mutations in CLPP, Encoding a Mitochondrial ATP-Dependent Chambered Protease. *Am. J. Hum. Genet.* 92, 605–613.
- (21) Thompson, M. W., and Maurizi, M. R. (1994) Activity and specificity of *Escherichia coli* ClpAP protease in cleaving model peptide substrates. *J. Biol. Chem.* 269, 18201–18208.
- (22) Szyk, A., and Maurizi, M. R. (2006) Crystal structure at 1.9 Å of *E. coli* ClpP with a peptide covalently bound at the active site. *J. Struct. Biol.* 156, 165–174.
- (23) Kang, S. G., Dimitrova, M. N., Ortega, J., Ginsburg, A., and Maurizi, M. R. (2005) Human Mitochondrial ClpP Is a Stable Heptamer That Assembles into a Tetradecamer in the Presence of ClpX. *J. Biol. Chem.* 280, 35424–35432.
- (24) Böttcher, T., and Sieber, S. A. (2008) Beta-lactones as privileged structures for the active-site labeling of versatile bacterial enzyme classes. *Angew. Chem., Int. Ed.* 47, 4600–4603.
- (25) Gersch, M., Gut, F., Korotkov, V. S., Lehmann, J., Böttcher, T., Rusch, M., Hedberg, C., Waldmann, H., Klebe, G., and Sieber, S. A. (2013) The mechanism of caseinolytic protease (ClpP) inhibition. *Angew. Chem., Int. Ed.* 52, 3009–3014.

- (26) Kim, D. Y., and Kim, K. K. (2008) The structural basis for the activation and peptide recognition of bacterial ClpP. *J. Mol. Biol.* 379, 760–771.
- (27) Hachmann, J., Snipas, S. J., van Raam, B. J., Cancino, E. M., Houlihan, E. J., Poreba, M., Kasperkiewicz, P., Drag, M., and Salvesen, G. S. (2012) Mechanism and specificity of the human paracaspase MALT1. *Biochem. J.* 443, 287–295.
- (28) Kang, S. G., Ortega, J., Singh, S. K., Wang, N., Huang, N. N., Steven, A. C., and Maurizi, M. R. (2002) Functional proteolytic complexes of the human mitochondrial ATP-dependent protease, hClpXP. *J. Biol. Chem.* 277, 21095–21102.
- (29) Martin, A., Baker, T. A., and Sauer, R. T. (2008) Diverse pore loops of the AAA+ ClpX machine mediate unassisted and adaptor-dependent recognition of ssrA-tagged substrates. *Mol. Cell* 29, 441–450.
- (30) Fischer, F., Weil, A., Hamann, A., and Osiewacz, H. D. (2013) Human CLPP reverts the longevity phenotype of a fungal ClpP deletion strain. *Nat. Commun.* 4, 1397.
- (31) Haynes, C. M., Petrova, K., Benedetti, C., Yang, Y., and Ron, D. (2007) ClpP mediates activation of a mitochondrial unfolded protein response in *C. elegans*. *Dev. Cell* 13, 467–480.
- (32) Haynes, C. M., Yang, Y., Blais, S. P., Neubert, T. A., and Ron, D. (2010) The matrix peptide exporter HAF-1 signals a mitochondrial UPR by activating the transcription factor ZC376.7 in *C. elegans*. *Mol. Cell* 37, 529–540.
- (33) Kanemori, M., Nishihara, K., Yanagi, H., and Yura, T. (1997) Synergistic roles of HslVU and other ATP-dependent proteases in controlling in vivo turnover of sigma32 and abnormal proteins in *Escherichia coli*. *J. Bacteriol.* 179, 7219–7225.
- (34) Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190.
- (35) Schilling, O., and Overall, C. M. (2008) Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nat. Biotechnol.* 26, 685–694.
- (36) Barrios, A. M., and Craik, C. S. (2002) Scanning the prime-site substrate specificity of proteolytic enzymes: a novel assay based on ligand-enhanced lanthanide ion fluorescence. *Bioorg. Med. Chem. Lett.* 12, 3619–3623.
- (37) Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., and Gevaert, K. (2009) Improved visualization of protein consensus sequences by iceLogo. *Nat. Methods* 6, 786–787.
- (38) Thompson, M. W., Singh, S. K., and Maurizi, M. R. (1994) Processive degradation of proteins by the ATP-dependent Clp protease from *Escherichia coli*. Requirement for the multiple array of active sites in ClpP but not ATP hydrolysis. *J. Biol. Chem.* 269, 18209–18215.
- (39) Choi, K. H., and Licht, S. (2005) Control of peptide product sizes by the energy-dependent protease ClpAP. *Biochemistry* 44, 13921–13931.
- (40) Akopian, T., Kandrór, O., Tsu, C., Lai, J. H., Wu, W., Liu, Y., Zhao, P., Park, A., Wolf, L., Dick, L. R., Rubin, E. J., Bachovchin, W., and Goldberg, A. L. (2015) Cleavage Specificity of Mycobacterium tuberculosis ClpP1P2 Protease and Identification of Novel Peptide Substrates and Boronate Inhibitors with Anti-bacterial Activity. *J. Biol. Chem.* 290, 11008–11020.
- (41) Ponder, E. L., Albrow, V. E., Leader, B. A., Bekes, M., Mikolajczyk, J., Fonovic, U. P., Shen, A., Drag, M., Xiao, J., Deu, E., Campbell, A. J., Powers, J. C., Salvesen, G. S., and Bogyo, M. (2011) Functional characterization of a SUMO deconjugating protease of *Plasmodium falciparum* using newly identified small molecule inhibitors. *Chem. Biol.* 18, 711–721.
- (42) Sato, Y., Yoshikawa, A., Yamagata, A., Mimura, H., Yamashita, M., Ookata, K., Nureki, O., Iwai, K., Komada, M., and Fukui, S. (2008) Structural basis for specific cleavage of Lys 63-linked polyubiquitin chains. *Nature* 455, 358–362.
- (43) O'Hara, B. P., Hemmings, A. M., Buttle, D. J., and Pearl, L. H. (1995) Crystal structure of glycyl endopeptidase from *Carica papaya*: a cysteine endopeptidase of unusual substrate specificity. *Biochemistry* 34, 13190–13195.
- (44) Lee, B.-G., Park, E. Y., Lee, K.-E., Jeon, H., Sung, K. H., Paulsen, H., Rübsamen-Schaeff, H., Brötz-Oesterhelt, H., and Song, H. K. (2010) Structures of ClpP in complex with acyldepsipeptide antibiotics reveal its activation mechanism. *Nat. Struct. Mol. Biol.* 17, 471–478.
- (45) Baker, T. A., and Sauer, R. T. (2012) ClpXP, an ATP-powered unfolding and protein-degradation machine. *Biochim. Biophys. Acta, Mol. Cell Res.* 1823, 15–28.
- (46) Ortega, J., Lee, H. S., Maurizi, M. R., and Steven, A. C. (2002) Alternating translocation of protein substrates from both ends of ClpXP protease. *EMBO J.* 21, 4938–4949.
- (47) Kasperkiewicz, P., Poreba, M., Snipas, S. J., Parker, H., Winterbourn, C. C., Salvesen, G. S., and Drag, M. (2014) Design of ultrasensitive probes for human neutrophil elastase through hybrid combinatorial substrate library profiling. *Proc. Natl. Acad. Sci. U. S. A.* 111, 2518–2523.
- (48) Zeiler, E., List, A., Alte, F., Gersch, M., Wachtel, R., Poreba, M., Drag, M., Groll, M., and Sieber, S. A. (2013) Structural and functional insights into caseinolytic proteases reveal an unprecedented regulation principle of their catalytic triad. *Proc. Natl. Acad. Sci. U. S. A.* 110, 11302–11307.
- (49) Porcelli, A. M., Ghelli, A., Zanna, C., Pinton, P., Rizzuto, R., and Rugolo, M. (2005) pH difference across the outer mitochondrial membrane measured with a green fluorescent protein mutant. *Biochem. Biophys. Res. Commun.* 326, 799–804.
- (50) Houtkoper, R. H., Mouchiroud, L., Ryu, D., Moullan, N., Katsyuba, E., Knott, G., Williams, R. W., and Auwerx, J. (2013) Mitonuclear protein imbalance as a conserved longevity mechanism. *Nature* 497, 451–457.
- (51) Mohrin, M., Shin, J., Liu, Y., Brown, K., Luo, H., Xi, Y., Haynes, C. M., and Chen, D. (2015) Stem cell aging. A mitochondrial UPR-mediated metabolic checkpoint regulates hematopoietic stem cell aging. *Science* 347, 1374–1377.
- (52) Cole, A., Wang, Z., Coyaud, E., Voisin, V., Gronda, M., Jitkova, Y., Mattson, R., Hurren, R., Babovic, S., Maclean, N., Restall, I., Wang, X., Jeyaraju, D. V., Sukhai, M. A., Prabha, S., Bashir, S., Ramakrishnan, A., Leung, E., Qia, Y. H., Zhang, N., Combes, K. R., Ketela, T., Lin, F., Houry, W. A., Aman, A., Al-Awar, R., Zheng, W., Wienholds, E., Xu, C. J., Dick, J., Wang, J. C., Moffat, J., Minden, M. D., Eaves, C. J., Bader, G. D., Hao, Z., Kornblau, S. M., Raught, B., and Schimmer, A. D. (2015) Inhibition of the Mitochondrial Protease ClpP as a Therapeutic Strategy for Human Acute Myeloid Leukemia. *Cancer Cell* 27, 864–876.
- (53) Al-Furoukh, N., Kardon, J. R., Kruger, M., Szibor, M., Baker, T. A., and Braun, T. (2014) NOA1, a novel ClpXP substrate, takes an unexpected nuclear detour prior to mitochondrial import. *PLoS One* 9, e103141.



RightsLink®

[Home](#) [Create Account](#) [Help](#)



Title: Barrel-shaped ClpP Proteases Display Attenuated Cleavage Specificities
Author: Malte Gersch, Matthias Stahl, Marcin Poreba, et al
Publication: ACS Chemical Biology
Publisher: American Chemical Society
Date: Feb 1, 2016
Copyright © 2016, American Chemical Society

[LOGIN](#)
If you're a **copyright.com user**, you can login to RightsLink using your copyright.com credentials. Already a **RightsLink user** or want to [learn more?](#)

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

[BACK](#)

[CLOSE WINDOW](#)

Copyright © 2017 [Copyright Clearance Center, Inc.](#) All Rights Reserved. [Privacy statement.](#) [Terms and Conditions.](#) Comments? We would like to hear from you. E-mail us at customercare@copyright.com

3

Quantitative map of β -lactone-induced virulence regulation

Published in the Journal of Proteome Research, 2017, 16(3), pp 1180-1192
by Joanna Krysiak,[#] Matthias Stahl,[#] Jan Vomacka,[#] Christian Fetzer, Markus Lakemeyer,
Anja Fux, and Stephan A. Sieber.

[#] equal contribution

Reprinted with permission. © 2017 American Chemical Society.

DOI: 10.1021/acs.jproteome.6b00705

SYNOPSIS

β -lactones are potent inhibitors of ClpP proteases from different species. This is of particular interest when considering bacterial virulence strategies. For example, *S. aureus* uses ClpP as a major switch to activate virulence, which was shown by *clpP* knock-out strains. Thus, as an example, ClpP activity is linked to the secretion of the α -toxin (hemolysin α , Hla) virulence factor. However, the mechanistic link between ClpP-mediated proteolysis and the presence of virulence factors such as Hla remained unknown.

Hla is able to penetrate the membrane of erythrocytes, which leads to immediate blood cell lysis. In this publication, four different β -lactones were first evaluated in terms of their potential to inhibit hemolysis (figure 3.1). Whereas the lactones **D3**, **U1**, and **U1p** effectively attenuate hemolysis, **E2** is a rather weak inhibiting agent. However, it is assumed that **E2** is a more reliable inhibitor of the functional ClpXP complex because it mediates inhibition by deoligomerization of tetradecameric ClpP. In contrast, the former lactones only block the active sites, while keeping the tetradecamer. Therefore, they may be outcompeted by the conformational selection that ClpX exerts on the ClpXP complex.

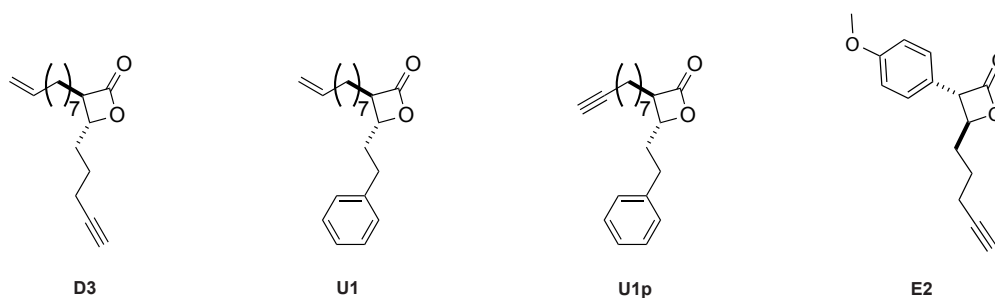


Figure 3.1: β -lactone-based ClpP inhibitors used in this study.

After an ABPP-based evaluation of the lactones' intracellular binding potential and specificity towards ClpP, mass spectrometric proteome analyses were conducted, to reveal global changes in protein expression of different *S. aureus* strains. For example, **D3** incubation for twelve hours led to a decrease of Hla levels and other known virulence effectors such as a leukocidin-like protein. Vice versa, members of known intracellular virulence signaling cascades were upregulated, e.g. the repressor of transcription Rot. A less clearer, but similar picture was gained for the deoligomerizing ClpP inhibitor **E2**.

In order to arrange all proteins involved in both virulence regulation and ClpP proteolysis, a quantitative map of changes in the bacterial proteome was created. To this end, a broad literature search for (i) molecular interactions between key proteins as well as (ii) RNAs involved in virulence mechanisms was performed to build a functional network of interdependencies. Next, all the global proteomics results were merged with this network. Thus, linkages between the ClpP subnetwork and the virulence regulation web appeared.

It turned out that ClpP inhibition through lactones is most likely linked to decreased virulence via the degradation of ClpX. ClpX in turn was previously found to control translation of *rot* mRNA. Additionally, the quantitative map highlighted a set of multi-step feedback circuits interconnecting ClpP activity, quorum sensing, and key virulence regulators, such as RNAIII and Rot.

AUTHOR CONTRIBUTIONS

Joanna Krysiak executed and analyzed proteomics experiments. Jan Vomacka performed ABPP, bacterial culture experiments and ran mass spectrometry. Matthias Stahl analyzed data, and designed as well as conducted bioinformatics analyses. Joanna Krysiak and Jan Vomacka performed biochemical experiments. Christian Fetzer executed hemolysis assays and developed as well as performed proteolysis assays. Markus Lakemeyer synthesized β -lactone **E2** and conducted analytical ABPP. Anja Fux prepared samples for mass spectrometry and together with Matthias Stahl measured **E2** data. Stephan A. Sieber conceived the project and wrote the manuscript with input from all authors.

Quantitative Map of β -Lactone-Induced Virulence Regulation

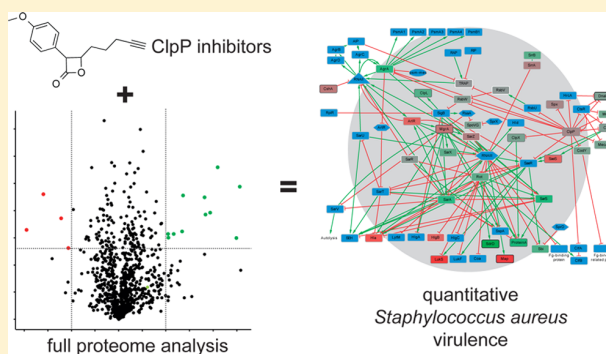
Joanna Krysiak,[†] Matthias Stahl,[†] Jan Vomacka,[†] Christian Fetzer, Markus Lakemeyer, Anja Fux, and Stephan A. Sieber*[✉]

Department of Chemistry, Chair of Organic Chemistry II, Center for Integrated Protein Science (CIPSM), Technische Universität München, Lichtenbergstrasse 4, D-85747 Garching, Germany

Supporting Information

ABSTRACT: β -Lactones have recently been introduced as the first selective ClpP inhibitors that attenuate virulence of both sensitive *Staphylococcus aureus* and multiresistant strains (MRSA). Although previous knockout studies showed that ClpP is essential for *S. aureus* alpha-toxin production, a link between β -lactone inhibition and molecular virulence mechanisms has been lacking so far. We here perform a chemical–proteomic approach to elucidate antivirulence pathways. First, we demonstrate by gel-free activity-based protein profiling that ClpP is the predominant target of β -lactones. Only a few off-targets were discovered, which, unlike ClpP, were not involved in the reduction of alpha-toxin expression. Second, in-depth mechanistic insight was provided by a full proteomic comparison between lactone treated and untreated *S. aureus* cells. Quantitative mass-spectrometric analysis revealed increased repressor of toxin (Rot) levels and a corresponding down-regulation of α -toxin, providing the first direct connection between the lactone-dependent phenotype and a corresponding cellular mechanism. By building up a quantitative virulence regulation network, we visualize the impact of ClpP inhibition in a systems biology context. Interestingly, a lack of in vitro Rot degradation by either ClpXP or ClpCP calls either for a proteolysis mechanism with yet unknown adaptor proteins or for an indirect mode of action that may involve ClpX-mediated RNA signaling and feedback circuits.

KEYWORDS: *Staphylococcus aureus* ClpP, proteomics, antivirulence, activity-based protein profiling, chemical knockout



INTRODUCTION

Current treatment of infectious diseases is largely challenged by the resistance of pathogenic bacterial strains to major antibiotics in clinical use. Among these multiresistant strains is *Staphylococcus aureus*, an opportunistic pathogen which causes a variety of infections, ranging from skin abscesses to life-threatening toxic shock syndrome and sepsis.¹ The balance between the symbiotic and pathogenic activity of *S. aureus* is tightly controlled by a broad set of colonization and virulence factors.² The production of these different cell surface-associated proteins (e.g., protein A or fibrinogen-binding protein) as well as secreted proteins (e.g., hemolysins, enterotoxins, or proteases) is spatiotemporally coordinated by a complex network of virulence regulators that orchestrate a hostile invasion of host tissue.³ This process is tightly controlled via two-component systems such as the accessory gene regulator (*agr*). In a process termed quorum sensing, a growing population of *S. aureus* produces autoinducing peptides (AIPs, encoded by *agrD*) that are secreted and sensed by the surrounding population.⁴ Upon a certain quorum of cells, AIP binding to a surface-located sensor histidine kinase (AgrC) induces phosphorylation of an intracellular response regulator (AgrA). Phosphorylation of AgrA directly promotes the

transcription of RNAPIII, a major effector of virulence.^{5–7} RNAPIII exhibits a versatile regulatory role by either inducing transcription⁸ or stabilizing as well as destabilizing specific mRNA sequences.^{7,9–12} A major target of RNAPIII is the mRNA of the DNA-binding protein repressor of toxins (Rot), the degradation of which results in reduced levels of the Rot protein in the cell.^{9,10} Rot is involved in the up- or downregulation of 146 genes¹³ and plays a crucial role in *S. aureus* virulence as a repressor of α -toxin, also termed hemolysin alpha (Hla), which is a major toxin essential for promoting pathogenesis. Thus, quorum sensing-induced *agr* signaling results in an RNAPIII-mediated reduction of Rot, representing the onset of virulence attack.

Antibiotic therapies suffer from a limited breadth of bacterial targets and are associated with rapid resistance development. Thus, methods were devised to identify new targets that extend the spectrum of available targets involved in pathogenesis. Here, recent research shows antivirulence strategies to represent a promising approach. For example, the inhibition of cholesterol biosynthesis,^{14–16} toxin production,^{17,18} or cell

Received: July 29, 2016

Published: February 10, 2017

adhesion¹⁹ has proven efficacious against *S. aureus*-induced sepsis, pneumonia, or skin infections. Antivirulence strategies are advantageous, as bacteria are not directly killed, alleviating their pressure to develop resistance pathways. However, in order to fully understand the mode of action of a compound, a detailed analysis of its involvement in virulence pathways is important. New insights into *S. aureus* virulence mechanisms may facilitate the development of sophisticated pharmaceutical strategies to impede virulence factor expression.

In previous studies, we introduced β -lactones as novel antivirulence agents that attenuate the production of major toxins and reduce skin abscesses in mice.^{18,20} A putative link between this phenotype and the cellular mode of action was established when activity-based proteome profiling (ABPP) studies with β -lactone probes revealed the bacterial serine protease caseinolytic protease P (ClpP) as a predominant target.^{21,22} ClpP has previously been shown to be essential for virulence regulation in several pathogenic strains, including *S. aureus*,^{23–27} *Listeria monocytogenes*,²⁸ and *Salmonella typhimurium*.^{29,30} ClpP associates with specific ATP-dependent chaperones (ClpX or ClpC in *S. aureus*), which are necessary for the recognition and unfolding of damaged, misfolded, or short-lived regulatory proteins.³¹ Chaperones are important for the translocation of the unfolded substrates into the barrel-shaped proteolytic chamber of ClpP.^{32,33} Accordingly, knock-outs of the chaperone ClpX³⁴ have established its essential role in virulence as well. It was assumed that β -lactones inhibit virulence through chemical inhibition of ClpP based on phenotypic observations. However, despite the clear link between ClpP and virulence, the mechanistic basis of the phenotype remains elusive. Based on the observation that transcription of RNAPIII is reduced in *clpP* and *clpX* mutants, it was hypothesized that ClpXP regulates virulence via *agr* signaling.²³ While ClpP is known to mediate the degradation of proteins associated with virulence, such as Rot,²⁵ proteomic studies with ClpP active site traps³⁵ as well as whole-proteome comparisons between wild-type and *clp* deletion strains failed to identify a candidate, virulence-associated substrate.^{26,36}

Here, we use a chemical–proteomic approach to elucidate the antivirulence mechanism of β -lactones. Lactone treated *S. aureus* cells revealed a strong upregulation of Rot and corresponding downregulation of Hla, providing the first mechanistic link to the observed antihemolytic properties.

EXPERIMENTAL SECTION

Minimal Inhibitory Concentration (MIC) and Hemolysis Assay

Bacteria were cultured as described in the Supporting Information in the presence of test compound or DMSO, respectively. After 20 h at 200 rpm and 37 °C, 100 μ L of the cultures were diluted 1:10 (v/v), transferred to a microtiter plate and measured at 600 nm with a plate reader (TECAN, Infinite M200pro). The MIC was defined as the concentration of a compound sufficient to fully inhibit bacterial growth. The remaining undiluted bacterial cultures were pelleted for 10 min at 6200g, 100 μ L of the supernatant were transferred to a microtiter plate (in triplicates per culture), incubated with 50 μ L of diluted sheep blood solution (10% (v/v) in PBS, heparinized sheep blood washed five times in PBS, Elocin lab, Germany) and measured in 1 min intervals at 600 nm, at 37 °C with a plate reader (TECAN, Infinite M200pro). Incubation of 100 μ L growth medium with 50 μ L diluted sheep blood solution (10% (v/v) in PBS) was used as a negative control. Bacterial supernatant from DMSO control samples (no inhibition of hemolysis) with 50 μ L diluted sheep blood solution was used as a positive control. Quantification was achieved via

calibration curve: a 1:2 (v/v) dilution series of bacterial supernatant from DMSO control samples with growth medium was incubated with diluted sheep blood solution. As a readout parameter from the assay, the area under the curve (absorbance at 600 nm vs time) was used.

Activity-Based Protein Profiling (ABPP)

Bacteria were grown under defined growth conditions for a desired time. Cultures were collected in a 50 mL falcon tube and centrifuged at 6000g for 10 min at 4 °C. The supernatant was disposed, bacteria were washed with warm PBS to remove remaining medium and spun down at 6000g for 10 min at 4 °C. The bacterial pellet was then resuspended in PBS to reach OD₆₀₀ = 40. 0.5 mL of the bacterial suspension and 5 μ L ABPP probe (100x stock in DMSO) or 5 μ L DMSO (as a control) were mixed and the tube was gently vortexed. Samples were incubated for 1 h at RT (ca. 25 °C) at 450 rpm. After labeling, cells were pelleted at 6000g, for 10 min at 4 °C and the supernatant was discarded. Cells were washed with 2 \times 1 mL cold PBS to remove excess of the probe and centrifuged each time at 6000g, for 10 min at 4 °C. Cells were snap frozen in liquid N₂ and stored at –80 °C until use.

Pellets were resuspended in 1 mL cold PBS and transferred to chilled 2 mL lysis tubes (Precellys Glass/Ceramic Kit SK38). Cells were lysed 2 \times 20 s with the Precellys Homogenizer (5400 rpm, run number: 1, run time: 20 s, pause: 5 s) with 5 min intermittent cooling on ice. The ball mill tubes were centrifuged (16200g, 10 min, 4 °C) and 800 μ L soluble lysate were transferred to new 1.5 mL microcentrifuge tubes and subjected to click chemistry (CC) reaction to append a reporter tag. For this, samples were treated with either 43 μ L gel-based ABPP master mix (3 μ L trifunctional linker (TFL): 5(6)-(1-[5-(4-Azido-benzoylamino)-1-carbamoyl-pentylcarbamoyl]-5-(6-biotinoylamino-hexanoylamino)-pentyl-carbamoyl)-tetramethylrhodamine; 10 mM in DMSO), 10 μ L freshly prepared TCEP (tris(2-carboxyethyl)phosphine; 50 mM in H₂O), 30 μ L TBTA (tris((1-benzyl-1H-1,2,3-triazol-5-yl)methyl)amine; 1.667 mM in 80% *t*-BuOH and 20% DMSO) or with 60 μ L gel-free ABPP master mix (20 μ L Biotin-PEG₃-N₃ (Jena Bioscience, CLK-AZ104P4-100; 10 mM in DMSO), 10 μ L freshly prepared TCEP (50 mM in H₂O), 30 μ L TBTA Ligand (1.667 mM in 80% *t*-BuOH and 20% DMSO)). Samples were then gently mixed and 10 μ L CuSO₄ solution (50 mM in H₂O) were added to initiate the 1,3-cycloaddition reaction. The final concentrations for gel-based and gel-free experiments were 35.6 μ M TFL, 593 μ M TCEP, 59.3 μ M TBTA, 593 μ M CuSO₄ and 233 μ M Biotin-PEG₃-N₃, 581 μ M TCEP, 58.2 μ M TBTA, 593 μ M CuSO₄, respectively. The lysates were mixed by gentle vortexing and incubated for 1 h at RT in the dark. After the click chemistry reaction, the lysates were transferred to 15 mL falcon tubes and 4 mL of cold acetone (–80 °C, LC-MS grade) were added. Proteins were precipitated ON at –80 °C. The precipitated proteins were thawed on ice, pelleted at 16900g for 15 min at 4 °C and the supernatant was discarded. The proteins were washed with 2 \times 1 mL cold methanol (–80 °C). Protein pellets were resuspended by sonication (15 s at 10% intensity) and proteins were pelleted again by centrifugation at 16900g for 10 min at 4 °C. The supernatant was discarded and protein pellets were air-dried at RT and subsequently resuspended in 500 μ L 0.2% SDS in PBS at RT by sonication (15 s at 10% intensity).

Samples were then incubated with 50 μ L of pre-equilibrated avidin-agarose beads under continuous mixing at 20 rpm for 1 h at RT (50 μ L avidin-agarose beads were resuspended by careful inverting in 1 mL 0.2% SDS in PBS in Protein LoBind Eppendorf tubes, spun down at 400g for 1 min at RT, supernatant discarded carefully, repeated 2 times). Thereafter, beads were pelleted at 400g for 1 min at RT, supernatant carefully disposed, washed 3 times with 1 mL 0.2% SDS in PBS, 2 times with 1 mL 6 M urea in water and 3 times with 1 mL PBS. After each washing step, beads were centrifuged at 400g for 1 min at RT and the supernatant was carefully discarded.

Gel-Free Preparative ABPP and Dimethyl Labeling

The ABPP procedure was followed as described above. The beads were resuspended in 200 μ L of denaturation buffer (7 M urea, 2 M thiourea in 20 mM pH 7.5 HEPES buffer). Dithiothreitol (DTT, 1 M, 0.2 μ L) was added, and the tubes were shortly mixed by vortexing and incubated in a thermomixer (450 rpm, 45 min, RT). Then, 2-

iodoacetamide (IAA, 550 mM, 2 μ L) was added, and the tubes were mixed by vortexing shortly and incubated in a thermomixer (450 rpm, 30 min, RT, in the dark). Remaining IAA was quenched by the addition of dithiothreitol (DTT, 1 M, 0.28 μ L). The tubes were shortly mixed by vortexing and incubated in a thermomixer (450 rpm, 30 min, RT). LysC (0.5 μ g/ μ L) was thawed on ice, and 1 μ L was added to each microcentrifuge tube, and the tubes were shortly mixed by vortexing and incubated in a thermomixer (450 rpm, 4 h, RT, in the dark). TEAB (triethylammonium bicarbonate buffer) (600 μ L, 50 mM in water) and then trypsin (1.5 μ L, 0.5 μ g/ μ L in 50 mM acetic acid) were added to each tube with a short vortexing step after each addition. The microcentrifuge tubes were incubated in a thermomixer (450 rpm, 13–15 h, 37 $^{\circ}$ C). The digest was stopped by adding 4 μ L formic acid (FA) and vortexing. After centrifugation (100g, 1 min, RT) the supernatant was transferred to a new Protein LoBind Eppendorf tube. FA (50 μ L, aqueous 0.1% solution) was added to the beads, and after vortexing and centrifugation (100g, 1 min, RT) the supernatant was added to the supernatant collected before. Again, FA (50 μ L, aqueous 0.1% solution) was added to the beads, and after vortexing and centrifugation (16200g, 3 min, RT) the supernatant was transferred to the combined supernatants. 50 mg SepPak C18 columns were equilibrated by gravity flow with 1 mL acetonitrile, 0.5 mL elution buffer (80% ACN, 0.5% FA), and 1 mL wash buffer 1 (aqueous 0.1% TFA solution). Subsequently, the samples were loaded by gravity flow and washed with 1 mL wash buffer 1 (aqueous 0.1% TFA solution) and 0.5 mL wash buffer 2 (aqueous 0.5% FA solution). Elution into new 2.0 mL Protein LoBind Eppendorf tubes was performed by the addition of 250 μ L elution buffer by gravity flow followed by 250 μ L elution buffer by vacuum flow until all liquid was eluted from the column. The eluates (total volume should be about 1.2 mL for each sample) were lyophilized and reconstituted in 100 μ L TEAB buffer (100 mM TEAB, 0.36% FA, pH = 6) by pipetting up and down, vortexing, and sonication (10 s) with short centrifugation steps after each reconstitution step. Then 8 μ L of fresh isotope labeling solution (light: 2% CH_2O , 0.3 M NaBH_3CN , medium: 2% CD_2O , 0.3 M NaBH_3CN , heavy: 2% $^{13}\text{CD}_2\text{O}$, 0.3 M NaBD_3CN) were added, and the tubes were mixed by vortexing, centrifuged briefly, and incubated in a thermoshaker (450 rpm, 1 h, 25 $^{\circ}$ C) (Table 1).

Table 1. Labeling Scheme for ABPP

Probe	Replicate	Heavy	Medium	Light
D3/U1p	1	10 μ M	50 μ M	DMSO
	2	50 μ M	DMSO	10 μ M
	3	DMSO	10 μ M	50 μ M
	4	10 μ M	50 μ M	DMSO
E2	1		50 μ M	DMSO
	2		DMSO	50 μ M
	3		DMSO	50 μ M

Samples were cooled on ice for 3 min and quenched with 16 μ L chilled 1% ammonia solution (NH_4OH , 25% stock solution) and 8 μ L chilled 5% FA solution with brief vortexing and centrifugation steps after each addition. Differentially labeled peptide solutions were lyophilized and stored at -20 $^{\circ}$ C.

Before MS measurement, the samples were dissolved in 30 μ L 1% FA by pipetting up and down, vortexing, and sonication for 15 min (brief centrifugation after each step). Differentially labeled samples were mixed (E2: before lyophilization (note: the method deviates due to two different sample handling workflows)). Samples were reconstituted in 1% FA, filtered, and then applied to two different MS settings for the two probe classes:

D3/U1p. Samples were analyzed via HPLC-MS/MS using an UltiMate 3000 nano HPLC system (Dionex, Sunnyvale, California, USA) equipped with an Acclaim C18 PepMap100 75 μ m ID \times 2 cm trap and Acclaim C18 PepMap RSLC, 75 μ m ID \times 15 cm separation columns coupled to Thermo Fisher LTQ Orbitrap Fusion (Thermo Fisher Scientific Inc., Waltham, Massachusetts, USA). Samples were loaded on the trap and washed for 10 min with 0.1% FA and then

transferred to the analytical column and separated using a 112 min gradient from 4% B to 35% B followed by 4 min at 80% B (at 200 nL/min flow rate) (buffer A: 95% H_2O , 5% DMSO, 0.1% FA, buffer B: 95% MeCN, 5% DMSO, 0.1% FA). LTQ Orbitrap Fusion was operated in a 3 s top speed data dependent mode. Full scan acquisition was performed in the orbitrap at a resolution of 120 000 and an ion target of 4e5 in a scan range of 300–1700 m/z . Monoisotopic precursor selection as well as dynamic exclusion for 60 s were enabled. Precursors with charge states of 2–7 and intensities greater than 5e3 were selected for fragmentation. Isolation was performed in the quadrupole using a window of 1.6 m/z . Precursors were collected to a target of 1e2 for a maximum injection time of 250 with “inject ions for all available parallelizable time” enabled. Fragments were generated using higher-energy collisional dissociation (HCD, normalized collision energy: 30%) and detected in the ion trap at a rapid scan rate. Internal calibration was performed using the ion signal of fluoranthene cations (EASY-ETD/IC source).

E2. Samples were analyzed via HPLC-MS/MS using an UltiMate 3000 nano HPLC system (Dionex, Sunnyvale, California, USA) equipped with an Acclaim C18 PepMap100 75 μ m ID \times 2 cm trap and an Acclaim Pepmap RSLC C18 separation column (75 μ m ID \times 50 cm) in an EASY-spray setting coupled to a Thermo Fischer LTQ Orbitrap Fusion (Thermo Fisher Scientific Inc., Waltham, Massachusetts, USA). Samples were loaded on the trap and washed with 0.1% TFA, then transferred to the analytical column and separated using a nonlinear 115 min gradient from 5% A to 32% B, then in 10 min to 90% B followed by another 10 min at 90% B (at 300 nL/min flow rate) (buffer A: H_2O with 0.1% FA, buffer B: MeCN with 0.1% FA). LTQ Orbitrap Fusion was operated in a 3 s top speed data dependent mode. Full scan acquisition was performed in the orbitrap at a resolution of 120000 and an AGC target of 2e5 in a scan range of 300–1500 m/z . Monoisotopic precursor selection as well as dynamic exclusion for 60 s were enabled. Precursors with charge states of 2–7 and intensities greater than 5e3 were selected for fragmentation. Isolation was performed in the quadrupole using a window of 1.6 m/z . Precursors were collected to an AGC target of 1e4 for a maximum injection time of 50 ms with “inject ions for all available parallelizable time” enabled. Fragments were generated using higher-energy collisional dissociation (HCD, normalized collision energy: 30%) and detected in the ion trap at a rapid scan rate. Internal calibration was performed using the ion signal of fluoranthene cations (EASY-IC).

Global Proteomics Analysis of D3 and U1

Bacteria were cultured under defined growth conditions for 12 h (stationary phase) or 20 h (poststationary phase). Bacteria were incubated with 20 μ L DMSO (control) or 20 μ L 100 mM compound stock in DMSO (1:1000 (v/v) dilution, final concentration 100 μ M). Compounds were added either at the beginning or 4 h after the start of the culture. After the indicated time, bacterial cells were harvested by centrifugation at 6000g for 10 min at 4 $^{\circ}$ C and the supernatant was discarded. Pellets were washed with 2 \times 1 mL chilled PBS (phosphate buffered saline), snap frozen in liquid nitrogen, and stored at -80 $^{\circ}$ C until use.

Pellets were resuspended in 1 mL lysis buffer (8 M urea, 1% (w/v) *N*-octylglucoside in 75 mM NaCl, 50 mM Tris-HCl, pH 8.2, supplemented with Protease Inhibitor Cocktail (Complete, EDTA-free, Roche)) (4 $^{\circ}$ C) and transferred to chilled 2 mL lysis tubes (Precellys Glass/Ceramic Kit SK38). Tubes were cooled on ice for 5 min, and cells were lysed with the Precellys Homogenizer using six times lysis program 3 (5400 rpm, run number: 1, run time: 20 s, pause: 5 s). After each lysis run, the tubes were cooled on ice for 2 min. The ball mill tubes were centrifuged (16200g, 20 min, 4 $^{\circ}$ C), supernatant was transferred to new Eppendorf tubes. Soluble proteome was filtered using 0.2 μ m filter. Total protein concentration was measured using a BCA protein concentration assay kit (Carl Roth Roti-Quant universal BCA Protein Assay Kit (0120.1) according to manufacturer's instructions) and adjusted to 1 mg/mL.

200 μ g protein (1 mg/mL) was transferred to LoBind Eppendorf tubes. Dithiothreitol (DTT, 100 μ M, 2 μ L) was added, and the tubes were mixed by vortexing shortly and incubated in a thermomixer (450

rpm, 45 min, 37 °C). Then 2-iodoacetamide (IAA, 550 mM, 2 μ L) was added, and the tubes were mixed by vortexing shortly and incubated in a thermomixer (450 rpm, 30 min, RT, in the dark). Remaining IAA was quenched by the addition of dithiothreitol (DTT, 1 M, 0.8 μ L). The tubes were shortly mixed by vortexing and incubated in a thermomixer (450 rpm, 30 min, RT). Endopeptidase Lys-C (0.5 μ g/ μ L in water, Wako Laboratory Chemicals) was thawed on ice and 4 μ L was added (ratio 1:100 (w/w) enzyme:protein) to each microcentrifuge tube, and the tubes were shortly mixed by vortexing and incubated in a thermomixer (450 rpm, 4 h, RT, in the dark). TEAB (triethylammonium hydrogen carbonate) solution (800 μ L, 50 mM in water) and then trypsin (4 μ L, 0.5 μ g/ μ L in 50 mM acetic acid, Promega) were added to each tube with a short vortexing step after each addition. The microcentrifuge tubes were incubated in a thermomixer (450 rpm, 13–15 h, 37 °C). The digest was stopped by adding 5 μ L formic acid (FA) to a final concentration of 0.5% and pH value ca. 2.6 and vortexing.

50 mg SepPak C18 columns (Waters) were equilibrated by gravity flow with 1 mL acetonitrile, 0.5 mL elution buffer (80% ACN, 0.5% FA), and 1 mL wash buffer 1 (aqueous 0.5% FA solution). Subsequently, the samples were loaded by gravity flow, washed with 1 mL aqueous 0.5% FA solution, and labeled with 5 mL of the respective dimethyl labeling solution. The following solutions were used: light (L): 30 mM NaBH₃CN, 0.2% CH₂O, 10 mM NaH₂PO₄, 35 mM Na₂HPO₄, pH 7.5; medium (M): 30 mM NaBH₃CN, 0.2% CD₂O, 10 mM NaH₂PO₄, 35 mM Na₂HPO₄, pH 7.5; heavy (H): 30 mM NaBH₃CN, 0.2% ¹³CD₂O, 10 mM NaH₂PO₄, 35 mM Na₂HPO₄, pH 7.5 (Table 2).

Table 2. Labeling Scheme for Global Proteomics Analyses of D3 and U1

Condition	Replicate	Heavy	Medium	Light
D3/U1, 12 h	1	U1	DMSO	D3
D3/U1, 12 h	2	D3	U1	DMSO
D3/U1, 12 h	3	DMSO	D3	U1
D3/U1, 20 h	1	U1	D3	DMSO
D3/U1, 20 h	2	D3	DMSO	U1
D3/U1, 20 h	3	DMSO	U1	D3
D3/U1, 4–16 h	1	U1	D3	DMSO
D3/U1, 4–16 h	2	DMSO	U1	D3
D3/U1, 4–16 h	3	D3	DMSO	U1

The samples were again desalted by flushing the columns with 1 mL aqueous 0.5% FA solution. Differentially labeled peptide solutions were eluted with 1 mL elution buffer (80% MeCN, 0.5% FA) and mixed in a tube, lyophilized and stored at –20 °C until use.

Samples were then reconstituted in 24 μ L 87.5% MeCN (4 μ L water and 20 μ L MeCN) and 20 μ L separated via offline UltiMate 3000 HPLC system (Dionex, Sunnyvale, California, USA) equipped with HILIC 75 μ m ID \times 15 cm separation column (buffer A: 95% MeCN, 5% H₂O, 0.1% TFA, buffer B: 5% MeCN, 95% H₂O, 0.1% TFA, flow 0.2 mL/min, gradient 0 to 30% buffer B in 57.5 min, then to 50% buffer B in 2.5 min, then to 100% buffer B in 2.5 min and hold 100% buffer B for 12.5 min, then to 0% buffer B in 2.5 min and hold 0% buffer B for 22.5 min) and a fraction collector. In total, 46 fractions were collected (each 0.25 mL), which were pooled into 10 fractions according to the UV trace of the chromatogram. All fractions were lyophilized and stored at –20 °C until use.

Before MS measurement the samples were dissolved in 30 μ L 1% FA by pipetting up and down, vortexing and sonication for 15 min (brief centrifugation after each step). 0.45 μ m centrifugal filter units (VWR) were equilibrated with two times 500 μ L water, 500 μ L 0.05 N NaOH and two times 500 μ L 1% FA (centrifugation: 16200g, 1 min, RT). Reconstituted samples were filtered through the equilibrated filters (centrifugation: 16200g, 2 min, RT).

Samples were analyzed via HPLC-MS/MS, Thermo Fisher LTQ Orbitrap XL (Thermo Fisher Scientific Inc., Waltham, Massachusetts,

USA) with a UltiMate 3000 nano HPLC system (Dionex, Sunnyvale, California, USA) using an Acclaim C18 PepMap100 75 μ m ID \times 2 cm trap and Acclaim C18 PepMap RSLC, 75 μ m ID \times 15 cm separation columns (buffer A: 95% H₂O, 5% DMSO, 0.1% FA, buffer B: 95% MeCN, 5% DMSO, 0.1% FA, flow 200 nL/min, gradient 4 to 35% buffer B in 112 min, then to 80% buffer B in 4 min and hold 80% buffer B for 4 min, then to 4% buffer B in 2 min and hold 4% buffer B for 20 min). The mass spectrometer was operated in data dependent mode. Precursors were measured in the orbitrap at a resolution of 60000 and an ion target of 1e6 in a scan range from 350 to 1400 *m/z*. Preview mode for FTMS master scans was enabled. Dynamic exclusion settings were as follows: repeat count: 1, repeat duration: 30 s, exclusion duration: 120 s, exclusion mass width relative to low mass: 10 ppm, exclusion mass width relative to high mass: 10 ppm. Charge states higher than 1 and a minimum signal threshold of 1000 counts were accepted for fragmentation. Five most intensive precursors were selected for fragmentation using collision-induced dissociation (CID) with the normalized collision energy of 35%. Fragments were measured in the ion trap with an ion target of 1e4. Measurements were done with a Thermo Xcalibur™ mass spectrometry data system version 2.1.0 SP1 build 1160.

Global Proteomics Analysis of E2

Bacteria were cultured under defined growth conditions for 20 h (poststationary phase). Bacteria were incubated with 20 μ L DMSO (control) or 20 μ L 100 mM compound stock in DMSO (1:1000 (v/v) dilution, final concentration 100 μ M). After the indicated time, bacterial cells were harvested by centrifugation at 6000g for 10 min at 4 °C and the supernatant was discarded. Pellets were washed with 2 \times 1 mL chilled PBS (phosphate buffered saline), snap frozen in liquid nitrogen and stored at –80 °C until use.

Pellets were resuspended in 1 mL PBS (4 °C) and transferred to chilled 2 mL lysis tubes (Precellys Glass/Ceramic Kit SK38). Tubes were cooled on ice for 5 min and cells were lysed with the Precellys Homogenizer using three times lysis program 3 (5400 rpm, run number: 1, run time: 20 s, pause: 5 s). After each lysis run, the tubes were cooled on ice for 5 min. The ball mill tubes were centrifuged (16200g, 10 min, 4 °C), 800 μ L of supernatant were transferred to 15 mL falcon tubes and 4 mL of cold acetone (–80 °C, MS grade) were added. Proteins were precipitated ON at –80 °C. The precipitated proteins were thawed on ice, pelletized (16900g, 15 min, 4 °C) and the supernatant was disposed of. Falcon tubes were stored on ice during the following washing procedure: protein pellets were washed two times with 1 mL cold methanol (–80 °C). Resuspension was achieved by sonication (15 s at 10% intensity) and proteins were pelletized via centrifugation (16900g, 10 min, 4 °C). Only MS grade water was used for the following procedures. After two washing steps, the supernatant was disposed of and the pellet was resuspended in 200 μ L denaturation buffer (7 M urea, 2 M thiourea in 20 mM pH 7.5 HEPES buffer). Dithiothreitol (DTT, 1 M, 0.2 μ L) was added, the tubes were mixed by vortexing shortly and incubated in a thermomixer (450 rpm, 45 min, RT). Then 2-iodoacetamide (IAA, 550 mM, 2 μ L) was added, the tubes were mixed by vortexing shortly and incubated in a thermomixer (450 rpm, 30 min, RT, in the dark). Remaining IAA was quenched by the addition of dithiothreitol (DTT, 1 M, 0.28 μ L). The tubes were shortly mixed by vortexing and incubated in a thermomixer (450 rpm, 30 min, RT). LysC (0.5 μ g/ μ L) was thawed on ice and 1 μ L was added to each microcentrifuge tube, the tubes were shortly mixed by vortexing and incubated in a thermomixer (450 rpm, 4 h, RT, in the dark). TEAB solution (600 μ L, 50 mM in water) and then trypsin (1.5 μ L, 0.5 μ g/ μ L in 50 mM acetic acid) were added to each tube with a short vortexing step after each addition. The microcentrifuge tubes were incubated in a thermomixer (450 rpm, 13–15 h, 37 °C). The digest was stopped by adding 6 μ L formic acid (FA) and vortexing.

50 mg SepPak C18 columns (Waters) were equilibrated by gravity flow with 1 mL acetonitrile, 1 mL elution buffer (80% ACN, 0.5% FA) and 3 mL aqueous 0.5% FA solution. Subsequently, the samples were loaded by gravity flow, washed with 5 mL aqueous 0.5% FA solution and labeled with 5 mL of the respective dimethyl labeling solution.

The following solutions were used: light (L): 30 mM NaBH₃CN, 0.2% CH₂O, 10 mM NaH₂PO₄, 35 mM Na₂HPO₄, pH 7.5; medium (M): 30 mM NaBH₃CN, 0.2% CD₂O, 10 mM NaH₂PO₄, 35 mM Na₂HPO₄, pH 7.5; heavy (H): 30 mM NaBD₃CN, 0.2% ¹³CD₂O, 10 mM NaH₂PO₄, 35 mM Na₂HPO₄, pH 7.5 (Table 3).

Table 3. Labeling Scheme for Global Proteomics Analysis of E2

Condition	Replicate	Heavy	Medium	Light
E2, 20 h	1		DMSO	E2
E2, 20 h	2		E2	DMSO
E2, 20 h	3		E2	DMSO

Labeled peptides were eluted into new 2.0 mL Protein LoBind Eppendorf tubes. This was performed by the addition of 250 μ L elution buffer by gravity flow followed by 250 μ L elution buffer by vacuum flow until all liquid was eluted from the column. The eluates were mixed and lyophilized.

Before MS measurement the samples were dissolved in 30 μ L 1% FA by pipetting up and down, vortexing and sonication for 15 min (brief centrifugation after each step). Differentially labeled samples were mixed and filtered. Samples were analyzed via HPLC-MS/MS using an UltiMate 3000 nano HPLC system (Dionex, Sunnyvale, California, USA) equipped with an Acclaim C18 PepMap100 75 μ m ID x 2 cm trap and an Acclaim Pepmap RSLC C18 separation column (75 μ m ID x 50 cm) in an EASY-spray setting coupled to a Thermo Fischer LTQ Orbitrap Fusion (Thermo Fisher Scientific Inc., Waltham, Massachusetts, USA). Samples were loaded on the trap and washed with 0.1% TFA, then transferred to the analytical column (buffer A: H₂O with 0.1% FA, buffer B: MeCN with 0.1% FA, flow 300 nL/min, gradient 4 to 35% buffer B in 112 min, then to 80% buffer B in 4 min and hold 80% buffer B for 4 min, then to 4% buffer B in 2 min and hold 4% buffer B for 20 min). LTQ Orbitrap Fusion was operated in a 3 s top speed data dependent mode. Full scan acquisition was performed in the orbitrap at a resolution of 120000 and an AGC target of 2e5 in a scan range of 300–1500 m/z . Monoisotopic precursor selection as well as dynamic exclusion (exclusion duration: 60 s, exclusion mass width relative to low mass: 10 ppm, exclusion

mass width relative to high mass: 10 ppm) was enabled. Precursors with charge states of 2–7 and intensities greater than 5e3 were selected for fragmentation. Isolation was performed in the quadrupole using a window of 1.6 m/z . Precursors were collected to an AGC target of 1e4 for a maximum injection time of 35 ms with “inject ions for all available parallelizable time” enabled. Fragments were generated using higher-energy collisional dissociation (HCD, normalized collision energy: 30%) and detected in the ion trap at a rapid scan rate. Internal calibration was performed using the ion signal of fluoranthene cations (EASY-IC).

Bioinformatics

MS data were searched with MaxQuant software version 1.5.3.8 with most default settings (trypsin/P set as the digest enzyme, max. two missed cleavages, oxidation (M) and protein N-term acetylation as variable modifications, carbamidomethylation (C) as fixed modification, min. peptide length 7, 4.5 ppm for precursor mass tolerance (FTMS MS/MS match tolerance) and 0.5 Da for fragment mass tolerance (ITMS MS/MS match tolerance)). For protein identification the following settings were used: PSM FDR 0.01, Protein FDR 0.01, min razor + unique peptides: 1, razor protein FDR enabled, second peptides enabled. Searches were done against the Uniprot database for *S. aureus* NCTC8325 (taxon identifier: 93061, reference reviewed and unreviewed proteome (2889 entries) downloaded on 12.07.2015) and USA300 (taxon identifier: 367830, reviewed and unreviewed proteome (2607 entries) downloaded on 12.07.2015). Quantification was performed using dimethyl labeling with the following settings: light: DimethLys0, DimethNter0; medium: DimethLys4, DimethNter4 and heavy: DimethLys8, DimethNter8, min ratio count: 2, unique + razor peptides used for quantification, variable modifications included for quantification, advanced ratio estimation enabled, requantify option enabled.

Resulting data were further analyzed using Perseus software version 1.5.2.6. The rows were filtered (identified by site, reverse, contaminants), normalized ratios (compound vs DMSO) transformed 1/x (due to label switch, if not set up in MaxQuant) and log₂. One-sample Student's *t*-test was performed (at least two valid values required), and volcano plots were generated plotting log₂(fold change (FC) compound/DMSO) vs $-\log_{10}$ (*p*-value).

The mass-spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral>).

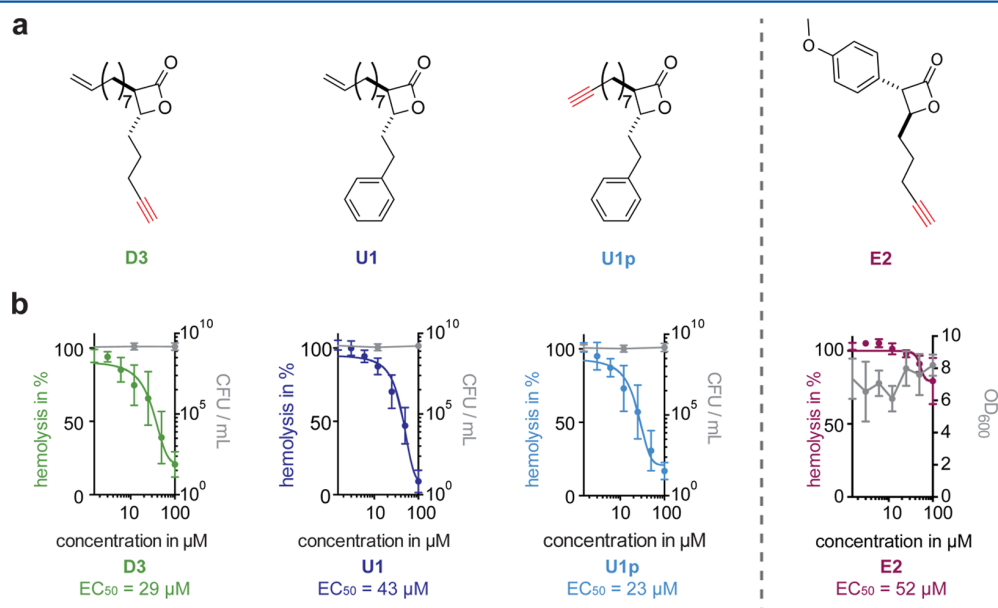


Figure 1. ClpP inhibition with β -lactones. (a) Structural depiction of β -lactones with red-colored alkyne handles for D3, U1p, and E2 which is used later in the study. (b) Hemolysis inhibition with D3, U1, U1p, and E2 in *S. aureus* NCTC8325. The mean value and standard deviation of three technical replicates are shown. Growth is indicated by colony-forming units (CFU)/mL culture or OD₆₀₀ (gray).

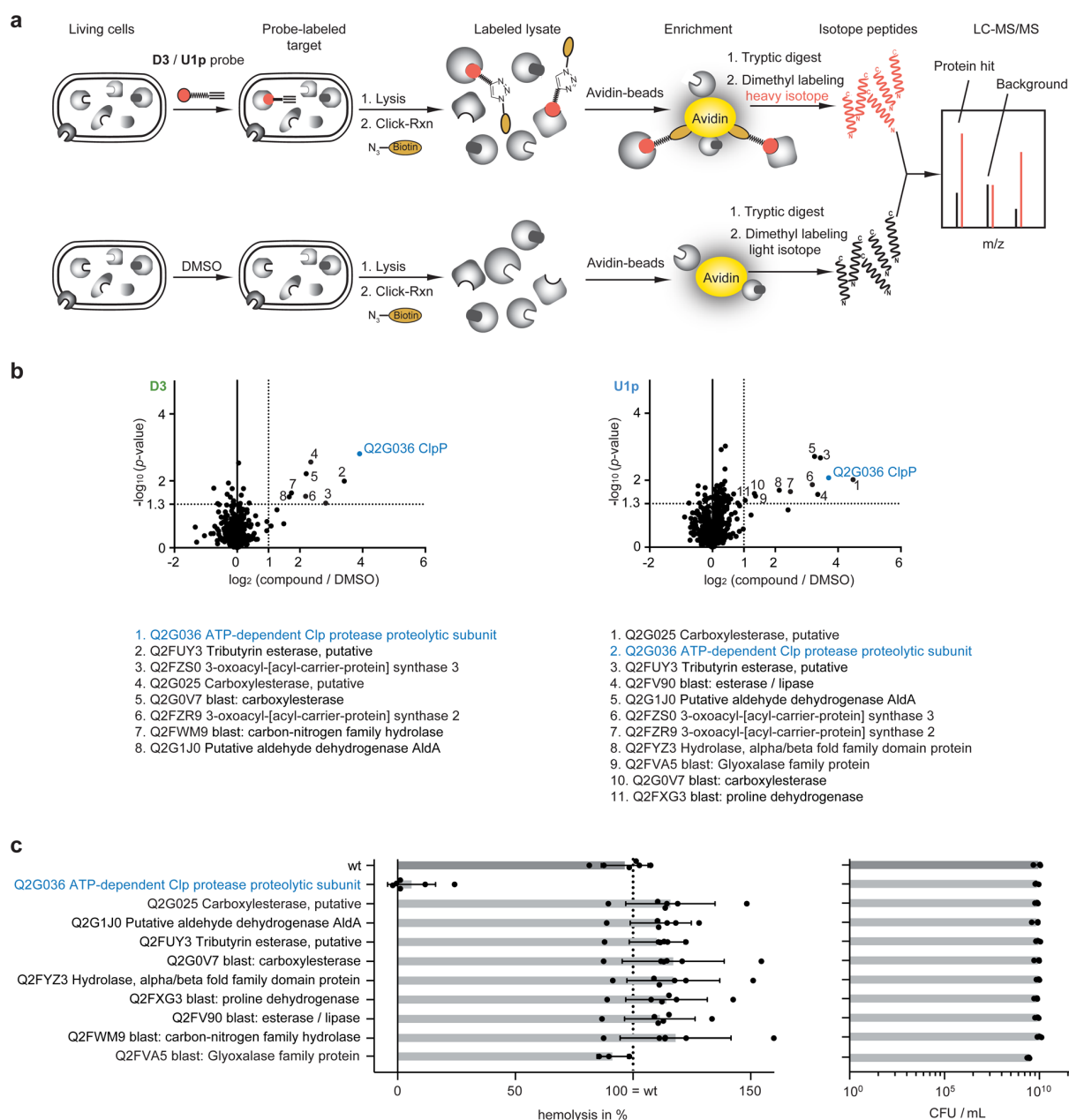


Figure 2. Activity-based protein profiling studies. (a) Workflow for the quantitative comparison of probe-treated versus DMSO-treated bacterial cells. Protein targets are labeled *in situ*, and a biotin tag is attached by click chemistry after cell lysis. Dimethyl labeling of primary amino groups following enrichment with avidin beads and tryptic digest is used to quantitatively assess protein ratios. (b) MS-based target identification in *S. aureus* NCTC8325. Volcano plots showing p -value ($-\log_{10}$ transformed) and enrichment factor (\log_2 transformed) for 50 μM compound vs DMSO. Analysis of *S. aureus* USA300 and USA300Lac strains is provided in Figure S-1. A complete list of all proteins identified with **D3** and **U1p** in *S. aureus* NCTC8325, USA300, and USA300Lac can be found in the Supporting Information. (c) Validation of ClpP as antivirulence target. Hemolytic activities of transposon mutants compared to wild-type are shown. Values are depicted as mean \pm SD ($n = 6$, $n = 3$ for Q2FVA5). Growth of the mutants was assessed via CFU assays. Here, the mean \pm SD ($n = 3$) is shown.

proteomexchange.org) via the PRIDE partner repository³⁹ with the data set identifier PXD004680.

In Vitro ClpXP/ClpCP Assay

GFP degradation assays were performed in PZ buffer (25 mM HEPES, 200 mM KCl, 5 mM MgCl₂, 1 mM DTT, 10% (v/v) glycerol, pH 7.6) with 60 μL reaction volume at 30 $^{\circ}\text{C}$. Degradation reactions contained 0.8 μM ClpX₆, 0.3 μM ClpP₁₄, 0.25 μM GFP-SsrA/0.5 μM Rot/Q2FW49/Q2FW50 and an ATP regeneration system (4 mM ATP, 16 mM creatine phosphate, 20 U/mL creatine phosphokinase) for the

ClpXP system or 1.0 μM ClpC₆, 0.3 μM ClpP₁₄, 2.5 μM MecA, 0.25 μM GFP-SsrA/0.5 μM Rot/Q2FW49/Q2FW50 and the ATP regeneration system for the ClpCP system, respectively. All reaction partners except the substrate were preincubated for 15 min at 30 $^{\circ}\text{C}$. The substrate was added afterward to start the reaction. The samples were incubated for 14 h at 30 $^{\circ}\text{C}$ and submitted to SDS-PAGE with subsequent Coomassie Brilliant Blue staining.

GO term enrichment analysis for the “biological process” branch was performed using the BiNGO plugin for Cytoscape.^{37,38} Proteins that were at least significantly 2-fold up- or down-regulated were

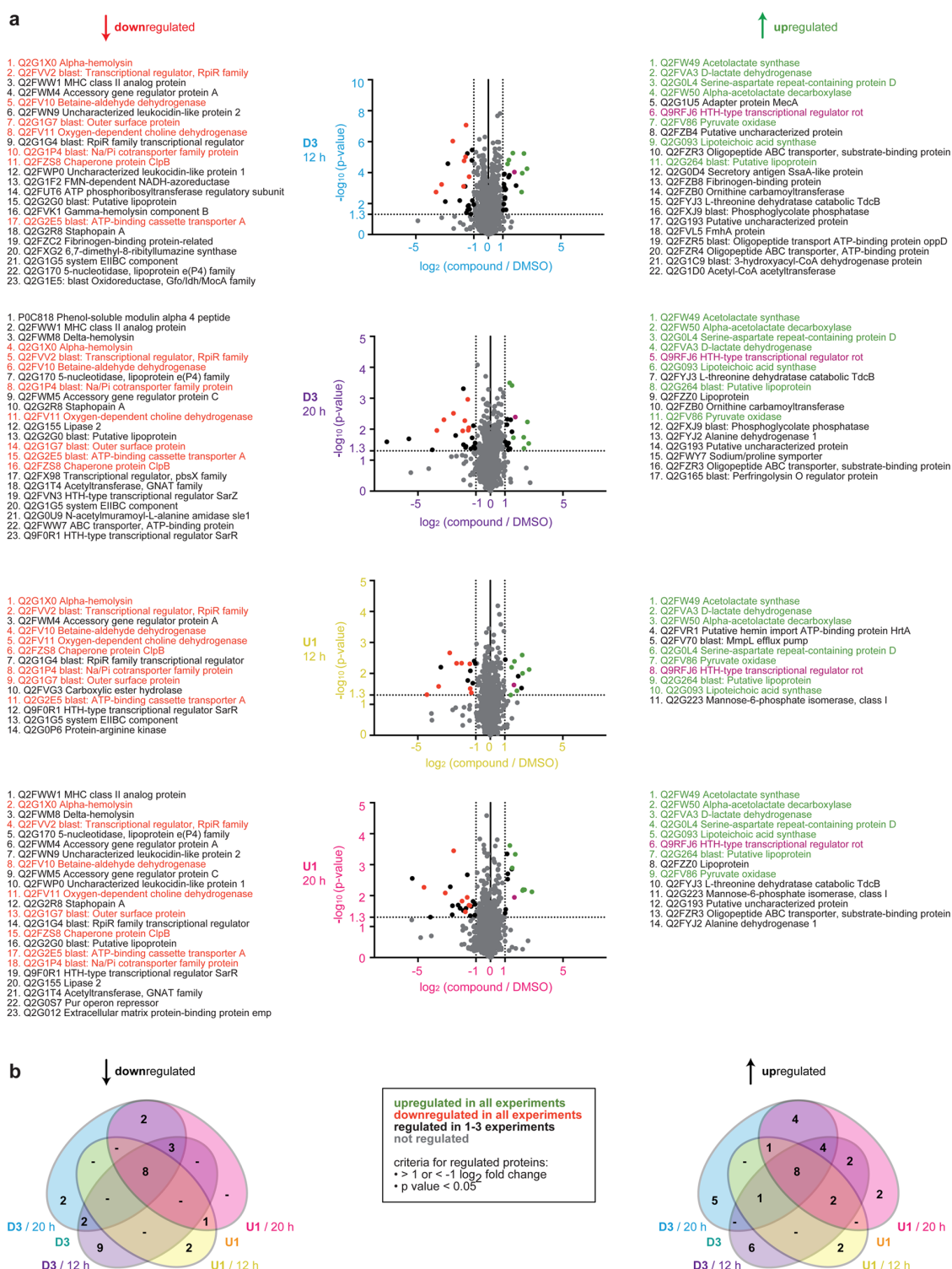


Figure 3. Whole-proteome analysis with D3 and U1. (a) *S. aureus* NCTC8325 cells were grown in the presence or absence of 100 μ M compound (D3 or U1) for 12 or 20 h; standard experiments included three biological replicates (D3/12 h: six biological replicates). The ratio of treated vs untreated cells was quantified via dimethyl labeling, and hits were considered to be up-/downregulated with a fold-change >2 or <-2 (p -value <0.05). Up- or downregulated proteins observed under all conditions by both probes are highlighted in green and red, respectively. The protein Rot is labeled in purple. (b) Venn diagrams summarize the differential regulation of all proteins. The color code matches the corresponding volcano plots. Complete identification and quantification data can be found in the [Supporting Information](#).

subjected to a hypergeometric test to decipher overrepresented GO categories. In order to construct a reference set, all GO annotations of

S. aureus NCTC8325 (taxonomy ID: 93061) were collected using QuickGO (<https://www.ebi.ac.uk/QuickGO/>, 13th January 2016).

The ontology go.obo was downloaded from <http://geneontology.org> (12th January 2016). The derived *p*-values were corrected for multiple testing by the method of Benjamini and Hochberg with a significance level of *p* = 0.05.

RESULTS AND DISCUSSION

Quantitative Proteome Profiling with Antivirulence β -Lactones

Previous ABPP studies in *S. aureus* with alkynylated β -lactone probes were performed by *in situ* labeling, followed by cell lysis, click conjugation to rhodamine/biotin azide, and separation by SDS-PAGE. Fluorescent scanning visualized putative target proteins that were isolated after avidin bead enrichment.⁴⁰ Although several targets could be identified by mass-spectrometric (MS) analysis, gel-based methodologies exhibit limited sensitivity toward low-abundant proteins. We therefore initiated our new analysis here by performing gel-free experiments via quantitative proteomics to obtain the full complement of cellular β -lactone binders. To this end, we initially selected three representative β -lactones, U1 and the alkynylated probes D3/U1p (Figure 1a). All compounds attenuated hemolysis of red blood cells without reduction of colony-forming units (CFU) or OD₆₀₀ up to concentrations of 100 μ M in *S. aureus* NCTC8325 (Figure 1b).

For gel-free ABPP, *S. aureus* strains NCTC8325, USA300, and USA300Lac were grown to poststationary phase (20 h) and incubated with either 50 μ M probe (D3 or U1p) or DMSO as a control for 1 h. Probe- and DMSO-treated cells were lysed, clicked to biotin azide, enriched on avidin beads, and digested via trypsin. Heavy or light isotopes of formaldehyde were appended to the free amino-termini of resulting peptides via reductive amination (dimethyl labeling).⁴¹ Peptides tagged with heavy and light isotopes were mixed and collectively analyzed by high-resolution MS (Figure 2a). After log₂ transformation of the normalized dimethyl labeling ratios, $-\log_{10}$ transformed *p*-values were obtained using a two-sided one sample Student's *t*-test against 0 over four biological replicates. The results are depicted in corresponding volcano plots (Figures 2b and S-1 for different strains). Proteins were regarded as hits if they exhibited more than a 2-fold enrichment over the DMSO control with a *p*-value < 0.05. Applying these criteria, we determined four hit proteins that were consistently identified with probes D3 and U1p in all three bacterial strains.

Interestingly, ClpP was among the most enriched targets, confirming previous results of gel-based studies.^{40,42} In addition, two 3-oxoacyl-[acyl-carrier-protein] synthases (observed as β -lactone targets previously)⁴² and one carboxylesterase were identified, which could contribute to the antivirulence mode of action as well. In order to dissect the individual role of major target proteins in virulence, we investigated the effect of their corresponding transposon mutants (obtained from the NARSA library) on Hla expression by hemolysis assays. Both 3-oxoacyl-[acyl-carrier-protein] synthases were not available from the library, which implies they may be essential. However, no growth inhibition upon treatment of D3 or U1p was observed (Figure 1), which points to a rather weak binding to these proteins. Previous studies have confirmed their role in fatty acid biosynthesis without any obvious connection to virulence.^{43–47} All available transposon mutants exhibited comparable CFUs, and in agreement with previous reports,²³ only the *clpP* transposon mutant revealed a drastic reduction of hemolysis compared to the wild-type (Figure 2c). This indicates that ClpP

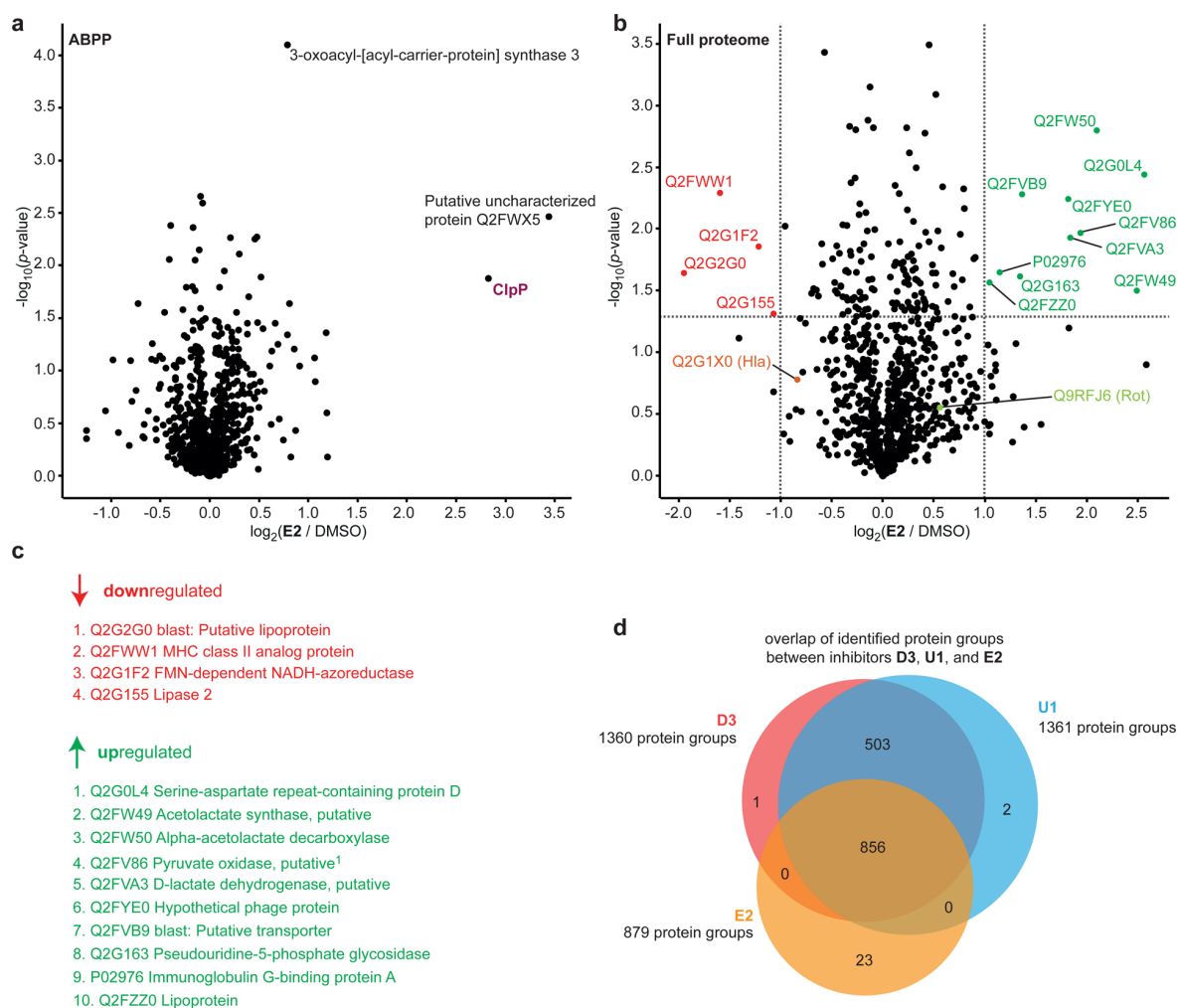
binding and inhibition represents a possible link to the molecular mode of action of the lactones.

Full Proteome Analysis Revealing of Signature Proteins with Links to Virulence

In order to analyze the molecular mechanisms of β -lactones in more detail, a full proteome analysis upon compound treatment was performed. Prior to these studies, the optimal time-point for comparative *S. aureus* proteome analysis was determined by monitoring ClpP abundance during various growth phases via Western blot analysis. Western blotting with a customized anti-SaClpP antibody (Figure S-2) revealed that the highest ClpP levels were observed during late exponential phase (6–12 h), followed by the poststationary phase (18–20 h). Similarly, previous experiments showed that several *S. aureus* virulence factors were expressed during the transition from exponential to stationary phase with maximum levels of Hla in the poststationary phase.³ Hence, for the global comparative analysis of the proteome of *S. aureus*, we selected 12 and 20 h to harvest bacterial cells, as time-points close to ClpP as well as Hla peak expression.

Prior to full proteome analysis, we ensured that D3 (50 μ M) treatment of cells led to ClpP labeling at the onset of Hla expression after 7 h of growth. Satisfyingly, an intense ClpP signal was observed by fluorescent SDS-PAGE (Figure S-3). Having confirmed target binding under the experimental conditions, we obtained protein ratios of lactone-treated versus untreated cells by quantitative proteomics experiments. *S. aureus* NCTC8325 was cultivated in medium substituted with either D3, U1, or DMSO, and cells were harvested after 12 and 20 h. Following cell lysis, protein digest, dimethyl labeling, and HILIC fractionation, peptides were analyzed via LC-MS/MS. Up- or downregulated hits are depicted in the volcano plots of Figure 3a. Interestingly, both β -lactones affected a limited number of proteins and only 66 of them showed a statistically significant >2-fold up- or downregulation.

Although differentially regulated proteins could be identified under each experimental condition (Figure 3b), most hits were independent of probe and growth phase. These results are in line with the comparable D3/U1p target spectrum and furthermore support that limited changes occur during the transition between growth phases. Importantly, a closer inspection of all shared hits revealed Rot as one of the most (3.5-fold, e.g. D3, 12 h) upregulated proteins, establishing the first direct link between β -lactones and their antivirulence effects. Accordingly, Hla, the effector of hemolysis, as well as SarR and AgrA, two transcriptional regulators of Hla expression,^{48,49} were 20.7-, 2.3-, and 10.6-fold downregulated (e.g., for U1, 12 h), respectively. In addition, a number of proteins with confirmed roles in virulence were detected. For example, D3 and U1 showed a strong downregulation of delta-hemolysin, a pore-forming toxin encoded by RNIII, as well as AgrC, a histidine kinase of the *agr* two-component signaling system, after 20 h, suggesting that expression of these virulence-associated proteins could be predominantly regulated at a later stage. Interestingly, both proteins are directly connected in a signaling pathway in which AgrC activates RNIII transcription via AgrA phosphorylation.⁵⁰ Moreover, reduction of gamma-hemolysin, an additional pore-forming toxin under the control of *agr*,⁵¹ as well as attenuation of SarZ, a transcriptional activator of *agr* signaling,⁵² highlight the larger impact on toxin regulation.



¹known substrate (Feng *et al.* 2013)

Figure 4. E2-mediated inhibition of ClpP. (a) Activity-based protein profiling (ABPP) of 50 μM E2 against DMSO. ClpP is a prominent target. (b) Full-proteome analysis of 100 μM E2 after 20 h of incubation. Proteins that were enriched or depleted more than 2-fold with a p -value smaller than 0.05 are depicted in green and red, respectively. Complete identification and quantification data can be found in the [Supporting Information S3](#) (ABPP) and [S4](#) (full proteome). (c) Overview of significantly depleted or enriched protein hits. (d) Overlap of identified protein groups with the three inhibitors D3, U1, and E2. The decrease in the number of totally identified proteins might be due to the E2-specific sample preparation and measurement workflow (see the Methods section).

In contrast to ClpP knockouts, chemical inhibitors do not require genetic manipulation and can be applied at any time to block the enzyme activity, which provide an exceptional flexibility for studying cellular systems. Therefore, we added D3 or U1 after 4 h of growth and harvested cells in the poststationary phase (20 h) to monitor time-dependent changes in protein levels. Generally, late lactone addition resulted in weaker signals, suggesting that early treatment triggers the most pronounced effects (Figure S-4).

A functional overview of the biological processes involved in this study was obtained by gene ontology (GO) enrichment analysis using the Biological Networks Gene Ontology (BiNGO) plugin for Cytoscape.^{37,38} Downregulated proteins were observed to have a prominent role in pathogenic cytolysis and amino-acid-related processes (Tables S-1 to S-7). In contrast, upregulated proteins did not show a clear and significant correlation to a certain pathway. However, some of the up-regulated proteins have previously been trapped by an

inactivated ClpP variant,³⁵ suggesting that functional inhibition results in accumulation of undigested substrates. These include pyruvate oxidase, MecA, and SsaA; however, a large fraction of previously identified ClpP substrates could not be verified here. The reason for this divergence could be manifold. On one hand, substrate trapping necessitates a ClpP deletion strain that is substituted with an inducible plasmid, which encodes the affinity-tagged ClpP active-site mutant. These strains could possess varying protein expression levels of ClpP compared to the wild-type, altering the stability of proteins that become more sensitive to unfolding and aggregation under these conditions.^{53,54} On the other hand, chemical inhibition of ClpP depends on conformational control of associated chaperones, which could generate partially active proteases under certain conditions.^{55–57} In fact, previous studies with various β -lactones showed that they either retain the tetradecameric state of ClpP upon compound binding, as observed for aliphatic D3, or dissociate into heptamers, as observed with the aromatic

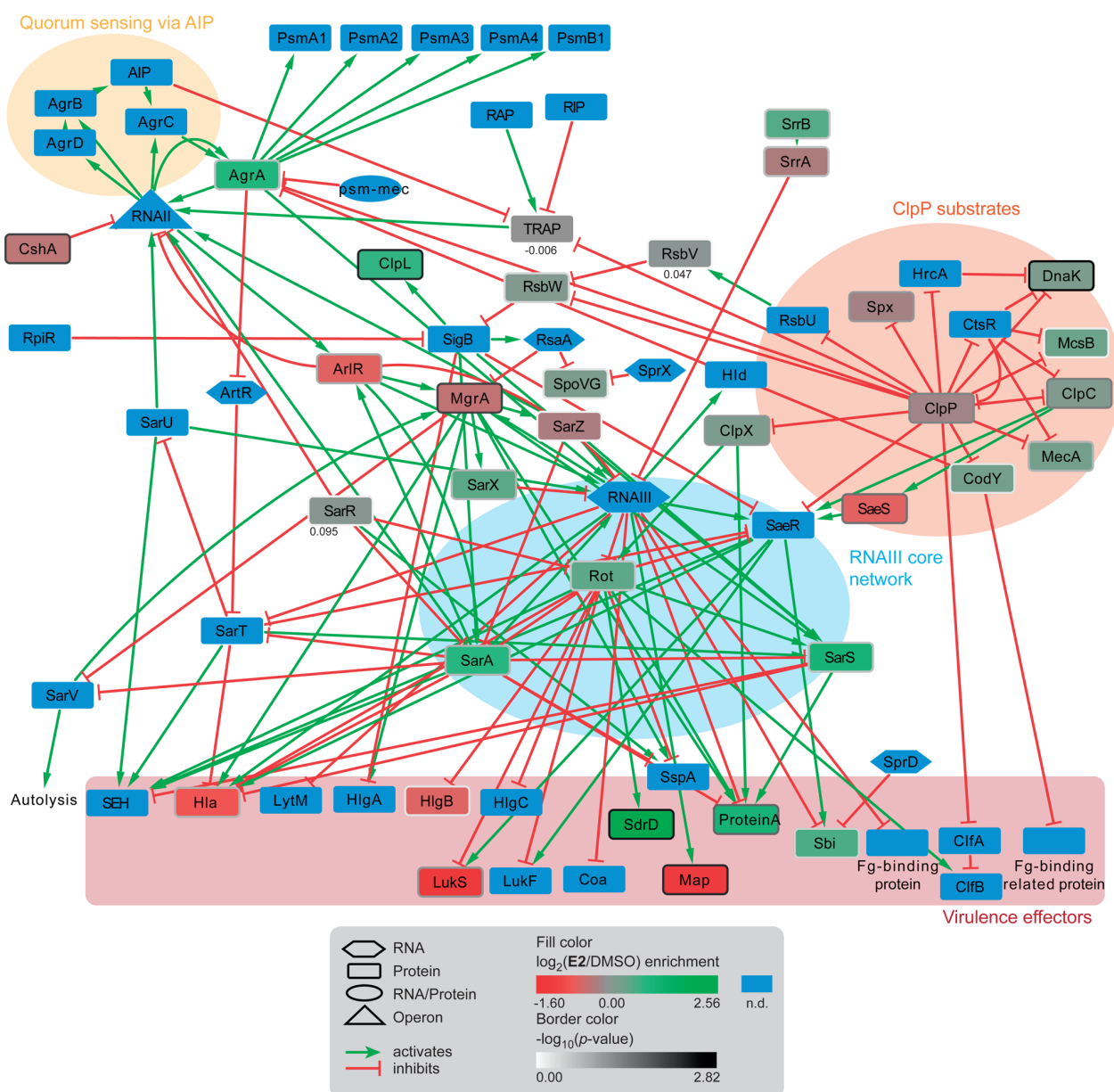


Figure 5. Quantitative map of β -lactone-induced virulence regulation (E2, 20 h). Proteins were quantified from three biological replicates. AIP: autoinducing peptide, RAP: RNAIII activating protein, RIP: RNAIII inhibiting protein, SEH: staphylococcal enterotoxin H, Fg: fibrinogen, n.d.: not determined. Proteins with \log_2 fold changes smaller than $|0.1|$ were additionally labeled with the respective level change. Maps for other β -lactones as well as a comment on the setup of the maps can be found in the [Supporting Information](#). References for each edge are listed as [Supporting References](#).

lactone E2. This deoligomerization mode also likely affects ClpXP proteolysis since D3 was a rather poor inhibitor, while deoligomerizing E2 significantly reduced proteolytic activity.^{56,58} To compare the effect of these divergent inhibition modes on proteomic profiles, E2 was subjected to a closer inspection by the above panel of experiments on a high-performance MS instrument without HILIC fractionation (Figures 1 and 4 and Figure S-5).

Although E2 reduced hemolysis less effectively (down to 80%) compared to D3 and U1 (Figure 1), *in situ* ABPP again confirmed ClpP as the predominant target as well as an uncharacterized protein as an additional hit (Figure 4a). Correspondingly, whole-proteome experiments of E2-treated

and -untreated *S. aureus* cells revealed downregulated Hla levels slightly below the enrichment cutoff (Figure 4bc). These results emphasize that while E2 disrupts the protease, it is less effective in the *in situ* reduction of virulence. Accordingly, the levels of Rot were not as elevated as observed for the nondeoligomerizing lactones. However, although the number of significantly up- or downregulated proteins was reduced, the results matched with D3 and U1 data (Figure 4d), and we therefore, conclude that reduced uptake or lower stability of the aromatic lactone E2 may contribute to a less pronounced phenotype.

Rot Upregulation, a Key Feature of *agr*-Mediated Virulence Signaling

It was previously speculated that Rot could be a substrate of ClpP-dependent degradation.²⁵ In that case, Rot degradation would be directly linked to the *agr* virulence control system. Although ClpP traps could not confirm this hypothesis,^{35,59} the accumulation of Rot in lactone-treated cells (Figure 3) called for a direct validation via a ClpXP degradation assay. We therefore overexpressed recombinant Rot, removed the strep-affinity tag using TEV cleavage, and incubated it with the reconstituted ClpXP protease according to established procedures.⁶⁰ While GFP fused to a ClpXP-specific SsrA degradation tag was efficiently degraded, no reduction in protein abundance was observed for Rot. The same results were obtained in a ClpCP proteolysis assay, which was customized for our in vitro studies based on previous reports.^{61–64} In addition, other putative protein hits, including acetolactate synthase (Q2FW49) and α -acetolactate decarboxylase (Q2FW50), could not be confirmed as substrates (Figure S-6), suggesting that lactone inhibition may trigger secondary effects that result in their upregulation or even that other chaperone adaptors are required for their targeting. Another possibility could be the direct or indirect degradation of RNAIII, a central antisense regulator of multiple genes, in response to ClpP inactivation by β -lactones. For example, RNAIII inhibits Rot mRNA translation and thereby additionally triggers regulation of Rot-dependent genes.⁹ Importantly, a connection to ClpP was already established previously by a ClpP knockout, which resulted in a strong decrease in RNAIII levels after 12 h.²⁷ However, we cannot exclude ClpP-mediated digestion of putative substrates via yet unknown mechanisms.

Quantitative Map of *S. aureus* Virulence Regulation

As RNAIII and Rot are key players in *S. aureus* virulence regulation, we inspected this regulation network in more detail. Therefore, we first built up a functional map of gene/protein signaling relations based on previous literature (see also the Supporting Procedures). RNAIII and Rot clearly stand out as central regulatory units (Figure 5). Upstream signals mainly derive from *agr*-mediated quorum sensing, and downstream signaling comprises several virulence regulators and effectors, including members of the Sar family. By incorporating our E2 quantitative proteomics data to the compiled network, we were able to highlight the effects of a chemical ClpP knockout.

In line with the literature, the observed upregulation of Rot shows an effect on Hla (down) and protein A (up) regulation.^{13,65} Concurrent SarA up-regulation seems to be a counteracting mechanism,^{66,67} possibly based on SigB-dependent signaling. However, the reason for the up-regulation of Rot upon ClpP inhibition remains unclear. A link between the protease and Rot could be, for instance, established via the RsbU-RsbV-RsbW-SigB-RsaA-MgrA-SarX axis,^{68–73} but RsbU seems to not be functional in our NCTC8325 strain.⁷⁴ Other links include ClpP substrates, such as CodY, via downregulation of AgrA, but this could not be validated with our data.³⁵ Interestingly, the ClpP-associated chaperone ClpX was found to control translation of *rot* mRNA.³⁴ As ClpX is also a ClpP substrate, we propose a ClpP-ClpX-Rot regulatory axis, which establishes a connection between proteolysis and virulence regulation. Additionally, proteins upstream of Rot signaling, such as AgrA, are also affected in lactone-treated cells, thus demonstrating the tight connection of individual components in the whole network. A putative link between the core network

and the *agr* sensors can be established via the SarTU feedback loop.⁷⁵ In this case, active SarT inhibits SarU, thereby shutting down RNAII-dependent genes such as the *agr* family or phenol-soluble modulins (PSMs). Comparing lactones D3/U1 with E2 (after 20 h of stimulation), the antivirulent phenotype remains similar. Hemolysins are downregulated, whereas protein A and upstream regulators, such as Rot, SarA, and SarS, are upregulated. However, direct effectors of quorum sensing, such as AgrA, are inversely regulated (down for D3/U1 and up for E2).

CONCLUSIONS

ClpP represents an important enzyme for mechanistic studies of bacterial cell homeostasis, which has also garnered interest as a target for antibacterial research. Acyldepsipeptides (ADEPs) have been reported as potent antibiotics that induce uncontrolled ClpP-mediated proteolysis of essential cellular proteins leading to cell death.⁷⁶ Contrary to this well-established cellular mode of action, it remains elusive how β -lactones, the first specific ClpP inhibitors discovered, promote their antivirulence properties. Off-target effects that could explain β -lactone-mediated attenuation of virulence by a ClpP-independent pathway were investigated by gel-free ABPP; however, no alternative route could be identified. Thus, this study provides a link between β -lactone, ClpP, and virulence by the observed upregulation of Rot, a major repressor protein of *S. aureus* pathogenesis. Although Rot could not be confirmed as a direct substrate of ClpXP/ClpCP in in vitro assays, a closer inspection of the RNAIII/Rot surrounding network revealed interesting systems biology insights. Therein, RNAIII is a key regulatory element whose abundance is strictly controlled over time. It acts mainly via Rot, and we hypothesize that its protein levels are dependent on ClpP activity, likely via ClpX levels. Moreover, ClpP is, in principle, able to affect virulence control factors such as SaeR in a direct manner or to affect RNAIII levels via SigB-dependent signaling. Additionally, we propose a ClpP-modulated negative feedback loop via SarTU to *agr* sensors, which are located upstream of RNAIII and explain altered expression of directly *agr*-controlled PSMs. This quantitative map of β -lactone-induced virulence is in agreement with previous findings of chemical knockout proteomics and sheds light on the multilayered role of ClpP in virulence regulation. However, many results cannot be understood without future studies to further dissect the whole picture of this network.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.6b00705.

Quantitative mass spectrometry results of D3/U1p. (XLSX)

Quantitative mass spectrometry results of D3/U1 full proteome analyses. (XLSX)

Quantitative mass spectrometry results of E2. (XLSX)

Quantitative mass spectrometry results of E2 full proteome analyses. (XLSX)

Additional figures, tables, and procedures. (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: stephan.sieber@tum.de. Tel: +49-89-28913302.

ORCID

Stephan A. Sieber: 0000-0002-9400-906X

Author Contributions

†J.K., M.S., and J.V. contributed equally to this work.

Author Contributions

J.K. executed and analyzed proteomics experiments. J.V. performed ABPP and bacterial culture experiments and ran mass spectrometry. M.S. performed mass spectrometry, analyzed data, and designed and conducted bioinformatics analyses. J.K., J.V., and M.S. performed biochemical experiments. C.F. developed and performed proteolysis assays. M.L. synthesized β -lactone E2 and conducted analytical ABPP. A.F. prepared samples for mass spectrometry. S.A.S. conceived the research and wrote the manuscript with input from all authors.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The work was funded by the Deutsche Forschungsgemeinschaft, SFB1035 and the Center for Integrated Protein Science Munich (CIPSM). Victor J. Torres (NYU School of Medicine) provided the anti-Rot antibody. We thank Maria Dahmen for cloning of MecA and ClpC expression vectors. M.S. was supported by the Wellcome Trust and the German National Merit Foundation and thanks the EuBIC for inspiring discussions. We thank the Network on Antimicrobial Resistance in *Staphylococcus aureus* (NARSA) for the supply of the Nebraska Transposon Mutant Library (NTML). Katja Bäuml, Mona Wolff, Jacqueline Kissel, and Karolina Koch technically assisted with experiments and mass spectrometry. Our group members are acknowledged for an inspiring environment, and we are grateful for critical proofreading of this manuscript by Annabelle Hoegl.

REFERENCES

- Lowy, F. D. *Staphylococcus aureus* infections. *N. Engl. J. Med.* **1998**, *339* (8), 520–32.
- Novick, R. P. Autoinduction and signal transduction in the regulation of staphylococcal virulence. *Mol. Microbiol.* **2003**, *48* (6), 1429–49.
- Cheung, A. L.; et al. Regulation of virulence determinants in vitro and in vivo in *Staphylococcus aureus*. *FEMS Immunol. Med. Microbiol.* **2004**, *40* (1), 1–9.
- George, E. A.; Muir, T. W. Molecular mechanisms of agr quorum sensing in virulent staphylococci. *ChemBioChem* **2007**, *8* (8), 847–55.
- Janzon, L.; Lofdahl, S.; Arvidson, S. Identification and nucleotide sequence of the delta-lysin gene, hld, adjacent to the accessory gene regulator (agr) of *Staphylococcus aureus*. *Mol. Gen. Genet.* **1989**, *219* (3), 480–5.
- Novick, R. P.; et al. Synthesis of staphylococcal virulence factors is controlled by a regulatory RNA molecule. *EMBO J.* **1993**, *12* (10), 3967–75.
- Benito, Y. Probing the structure of RNAIII, the *Staphylococcus aureus* agr regulatory RNA, and identification of the RNA domain involved in repression of protein A expression. *RNA* **2000**, *6* (5), 668–79.
- Gupta, R. K.; Luong, T. T.; Lee, C. Y. RNAIII of the *Staphylococcus aureus* agr system activates global regulator MgrA by stabilizing mRNA. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (45), 14036–41.

(9) Boisset, S.; et al. *Staphylococcus aureus* RNAIII coordinately represses the synthesis of virulence factors and the transcription regulator Rot by an antisense mechanism. *Genes Dev.* **2007**, *21* (11), 1353–66.

(10) Geisinger, E.; et al. Inhibition of rot translation by RNAIII, a key feature of agr function. *Mol. Microbiol.* **2006**, *61* (4), 1038–48.

(11) Huntzinger, E.; et al. *Staphylococcus aureus* RNAIII and the endoribonuclease III coordinately regulate spa gene expression. *EMBO J.* **2005**, *24* (4), 824–35.

(12) Morfeldt, E.; et al. Activation of alpha-toxin translation in *Staphylococcus aureus* by the trans-encoded antisense RNA, RNAIII. *EMBO J.* **1995**, *14* (18), 4569–77.

(13) Said-Salim, B.; et al. Global regulation of *Staphylococcus aureus* genes by Rot. *J. Bacteriol.* **2003**, *185* (2), 610–9.

(14) Liu, G. Y.; et al. *Staphylococcus aureus* golden pigment impairs neutrophil killing and promotes virulence through its antioxidant activity. *J. Exp. Med.* **2005**, *202* (2), 209–15.

(15) Liu, C. I.; et al. A cholesterol biosynthesis inhibitor blocks *Staphylococcus aureus* virulence. *Science* **2008**, *319* (5868), 1391–4.

(16) Song, Y.; et al. Inhibition of staphyloxanthin virulence factor biosynthesis in *Staphylococcus aureus*: in vitro, in vivo, and crystallographic results. *J. Med. Chem.* **2009**, *52* (13), 3869–80.

(17) Qiu, J.; et al. Baicalin protects mice from *Staphylococcus aureus* pneumonia via inhibition of the cytolytic activity of alpha-hemolysin. *J. Infect. Dis.* **2012**, *206* (2), 292–301.

(18) Weinandy, F.; et al. A beta-lactone-based antivirulence drug ameliorates *Staphylococcus aureus* skin infections in mice. *ChemMedChem* **2014**, *9* (4), 710–3.

(19) Zhang, J.; et al. Antiinfective therapy with a small molecule inhibitor of *Staphylococcus aureus* sortase. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (37), 13517–22.

(20) Vomacka, J.; Korotkov, V. S.; Bauer, B.; Weinandy, F.; Kunzmann, M. H.; Krysiak, J.; Baron, O.; Böttcher, T.; Lorenz-Baath, K.; Sieber, S. A. An aromatic hydroxyamide attenuates multidrug-resistant *Staphylococcus aureus* toxin expression. *Chem. - Eur. J.* **2016**, *22*, 1622.

(21) Evans, M. J.; Cravatt, B. F. Mechanism-based profiling of enzyme families. *Chem. Rev.* **2006**, *106* (8), 3279–301.

(22) Fonovic, M.; Bogoy, M. Activity-based probes as a tool for functional proteomic analysis of proteases. *Expert Rev. Proteomics* **2008**, *5* (5), 721–30.

(23) Frees, D.; et al. Alternative roles of ClpX and ClpP in *Staphylococcus aureus* stress tolerance and virulence. *Mol. Microbiol.* **2003**, *48* (6), 1565–78.

(24) Frees, D.; et al. Clp ATPases are required for stress tolerance, intracellular replication and biofilm formation in *Staphylococcus aureus*. *Mol. Microbiol.* **2004**, *54* (5), 1445–62.

(25) Frees, D.; Sorensen, K.; Ingmer, H. Global virulence regulation in *Staphylococcus aureus*: pinpointing the roles of ClpP and ClpX in the sar/agr regulatory network. *Infect. Immun.* **2005**, *73* (12), 8100–8.

(26) Frees, D.; et al. New insights into *Staphylococcus aureus* stress tolerance and virulence regulation from an analysis of the role of the ClpP protease in the strains Newman, COL, and SAS64. *J. Proteome Res.* **2012**, *11* (1), 95–108.

(27) Michel, A.; et al. Global regulatory impact of ClpP protease of *Staphylococcus aureus* on regulons involved in virulence, oxidative stress response, autolysis, and DNA repair. *J. Bacteriol.* **2006**, *188* (16), 5783–96.

(28) Gaillot, O.; et al. The ClpP serine protease is essential for the intracellular parasitism and virulence of *Listeria monocytogenes*. *Mol. Microbiol.* **2000**, *35* (6), 1286–94.

(29) Webb, C.; et al. Effects of DksA and ClpP protease on sigma S production and virulence in *Salmonella typhimurium*. *Mol. Microbiol.* **1999**, *34* (1), 112–23.

(30) Yamamoto, T.; et al. Disruption of the genes for ClpXP protease in *Salmonella enterica* serovar Typhimurium results in persistent infection in mice, and development of persistence requires endogenous gamma interferon and tumor necrosis factor alpha. *Infect. Immun.* **2001**, *69* (5), 3164–74.

(31) Sauer, R. T.; Baker, T. A. AAA+ proteases: ATP-fueled machines of protein destruction. *Annu. Rev. Biochem.* **2011**, *80*, 587–612.

- (32) Gersch, M.; et al. Insights into structural network responsible for oligomerization and activity of bacterial virulence regulator caseinolytic protease P (ClpP) protein. *J. Biol. Chem.* **2012**, *287* (12), 9484–94.
- (33) Gersch, M.; et al. The mechanism of caseinolytic protease (ClpP) inhibition. *Angew. Chem., Int. Ed.* **2013**, *52* (10), 3009–14.
- (34) Jelsbak, L.; et al. The chaperone ClpX stimulates expression of *Staphylococcus aureus* protein A by Rot dependent and independent pathways. *PLoS One* **2010**, *5* (9), e12752.
- (35) Feng, J.; et al. Trapping and proteomic identification of cellular substrates of the ClpP protease in *Staphylococcus aureus*. *J. Proteome Res.* **2013**, *12* (2), 547–58.
- (36) Michalik, S.; et al. Life and death of proteins: a case study of glucose-starved *Staphylococcus aureus*. *Mol. Cell. Proteomics* **2012**, *11* (9), 558–70.
- (37) Shannon, P.; et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13* (11), 2498–504.
- (38) Maere, S.; Heymans, K.; Kuiper, M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **2005**, *21* (16), 3448–9.
- (39) Vizcaino, J. A.; et al. The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* **2013**, *41* (Database issue), D1063–9.
- (40) Bottcher, T.; Sieber, S. A. Beta-lactones as specific inhibitors of ClpP attenuate the production of extracellular virulence factors of *Staphylococcus aureus*. *J. Am. Chem. Soc.* **2008**, *130* (44), 14400–1.
- (41) Hsu, J. L.; et al. Stable-isotope dimethyl labeling for quantitative proteomics. *Anal. Chem.* **2003**, *75* (24), 6843–52.
- (42) Bottcher, T.; Sieber, S. A. Beta-lactones as privileged structures for the active-site labeling of versatile bacterial enzyme classes. *Angew. Chem., Int. Ed.* **2008**, *47* (24), 4600–3.
- (43) Wang, J.; et al. Discovery of platencin, a dual FabF and FabH inhibitor with *in vivo* antibiotic properties. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (18), 7612–6.
- (44) Revill, W. P.; et al. Beta-ketoacyl acyl carrier protein synthase III (FabH) is essential for fatty acid biosynthesis in *Streptomyces coelicolor* A3(2). *J. Bacteriol.* **2001**, *183* (11), 3526–30.
- (45) Lai, C. Y.; Cronan, J. E. Beta-ketoacyl-acyl carrier protein synthase III (FabH) is essential for bacterial fatty acid synthesis. *J. Biol. Chem.* **2003**, *278* (51), 51494–503.
- (46) Tsay, J. T.; Rock, C. O.; Jackowski, S. Overproduction of beta-ketoacyl-acyl carrier protein synthase I imparts thiolactomyacin resistance to *Escherichia coli* K-12. *J. Bacteriol.* **1992**, *174* (2), 508–13.
- (47) Schujman, G. E.; et al. Response of *Bacillus subtilis* to cerulenin and acquisition of resistance. *J. Bacteriol.* **2001**, *183* (10), 3032–40.
- (48) Cheung, A. L.; et al. The SarA protein family of *Staphylococcus aureus*. *Int. J. Biochem. Cell Biol.* **2008**, *40* (3), 355–61.
- (49) Schwan, W. R.; et al. Loss of hemolysin expression in *Staphylococcus aureus* agr mutants correlates with selective survival during mixed infections in murine abscesses and wounds. *FEMS Immunol. Med. Microbiol.* **2003**, *38* (1), 23–8.
- (50) Srivastava, S. K.; et al. Influence of the AgrC-AgrA complex on the response time of *Staphylococcus aureus* quorum sensing. *J. Bacteriol.* **2014**, *196* (15), 2876–88.
- (51) Dunman, P. M.; et al. Transcription profiling-based identification of *Staphylococcus aureus* genes regulated by the agr and/or sarA loci. *J. Bacteriol.* **2001**, *183* (24), 7341–53.
- (52) Tamber, S.; Cheung, A. L. SarZ promotes the expression of virulence factors and represses biofilm formation by modulating SarA and agr in *Staphylococcus aureus*. *Infect. Immun.* **2009**, *77* (1), 419–28.
- (53) Frees, D.; et al. Clp ATPases and ClpP proteolytic complexes regulate vital biological processes in low GC, Gram-positive bacteria. *Mol. Microbiol.* **2007**, *63* (5), 1285–95.
- (54) Frees, D.; Ingmer, H. ClpP participates in the degradation of misfolded protein in *Lactococcus lactis*. *Mol. Microbiol.* **1999**, *31* (1), 79–87.
- (55) Gersch, M.; et al. AAA+ chaperones and acyldepsipeptides activate the ClpP protease via conformational control. *Nat. Commun.* **2015**, *6*, 6320.
- (56) Gersch, M.; et al. Disruption of oligomerization and dehydroalanine formation as mechanisms for ClpP protease inhibition. *J. Am. Chem. Soc.* **2014**, *136* (4), 1360–6.
- (57) Pahl, A.; et al. Reversible Inhibitors Arrest ClpP in a Defined Conformational State that Can Be Revoked by ClpX Association. *Angew. Chem., Int. Ed.* **2015**, *54* (52), 15892–6.
- (58) Hackl, M. W.; et al. Phenyl Esters Are Potent Inhibitors of Caseinolytic Protease P and Reveal a Stereogenic Switch for Deoligomerization. *J. Am. Chem. Soc.* **2015**, *137* (26), 8475–83.
- (59) Jelsbak, L.; et al. Growth phase-dependent regulation of the global virulence regulator Rot in clinical isolates of *Staphylococcus aureus*. *Int. J. Med. Microbiol.* **2010**, *300* (4), 229–36.
- (60) Pahl, A.; et al. Reversible Inhibitors Arrest ClpP in a Defined Conformational State that Can Be Revoked by ClpX Association. *Angew. Chem., Int. Ed.* **2015**, *54* (52), 15892–6.
- (61) Mei, Z.; et al. Molecular determinants of MecA as a degradation tag for the ClpCP protease. *J. Biol. Chem.* **2009**, *284* (49), 34366–75.
- (62) Turgay, K.; et al. Competence in *Bacillus subtilis* is controlled by regulated proteolysis of a transcription factor. *EMBO J.* **1998**, *17* (22), 6730–8.
- (63) Turgay, K.; et al. Biochemical characterization of a molecular switch involving the heat shock protein ClpC, which controls the activity of ComK, the competence transcription factor of *Bacillus subtilis*. *Genes Dev.* **1997**, *11* (1), 119–28.
- (64) Donegan, N. P.; Marvin, J. S.; Cheung, A. L. Role of adaptor TrfA and ClpPC in controlling levels of SsrA-tagged proteins and antitoxins in *Staphylococcus aureus*. *J. Bacteriol.* **2014**, *196* (23), 4140–51.
- (65) Tegmark, K.; Karlsson, A.; Arvidson, S. Identification and characterization of SarH1, a new global regulator of virulence gene expression in *Staphylococcus aureus*. *Mol. Microbiol.* **2000**, *37* (2), 398–409.
- (66) Manna, A.; Cheung, A. L. Characterization of sarR, a modulator of sar expression in *Staphylococcus aureus*. *Infect. Immun.* **2001**, *69* (2), 885–96.
- (67) Reyes, D.; et al. Coordinated regulation by AgrA, SarA, and SarR to control agr expression in *Staphylococcus aureus*. *J. Bacteriol.* **2011**, *193* (21), 6020–31.
- (68) Giachino, P.; Engelmann, S.; Bischoff, M. Sigma(B) activity depends on RsbU in *Staphylococcus aureus*. *J. Bacteriol.* **2001**, *183* (6), 1843–52.
- (69) Horsburgh, M. J.; et al. sigmaB modulates virulence determinant expression and stress resistance: characterization of a functional rsbU strain derived from *Staphylococcus aureus* 8325–4. *J. Bacteriol.* **2002**, *184* (19), 5457–67.
- (70) Palma, M.; Cheung, A. L. sigma(B) activity in *Staphylococcus aureus* is controlled by RsbU and an additional factor(s) during bacterial growth. *Infect. Immun.* **2001**, *69* (12), 7858–65.
- (71) Geissmann, T.; et al. A search for small noncoding RNAs in *Staphylococcus aureus* reveals a conserved sequence motif for regulation. *Nucleic Acids Res.* **2009**, *37* (21), 7239–57.
- (72) Romilly, C.; et al. A non-coding RNA promotes bacterial persistence and decreases virulence by regulating a regulator in *Staphylococcus aureus*. *PLoS Pathog.* **2014**, *10* (3), e1003979.
- (73) Ingavale, S.; et al. Rat/MgrA, a regulator of autolysis, is a regulator of virulence genes in *Staphylococcus aureus*. *Infect. Immun.* **2005**, *73* (3), 1423–31.
- (74) Kullik, I.; Giachino, P.; Fuchs, T. Deletion of the alternative sigma factor sigmaB in *Staphylococcus aureus* reveals its function as a global regulator of virulence genes. *J. Bacteriol.* **1998**, *180* (18), 4814–20.
- (75) Manna, A. C.; Cheung, A. L. sarU, a sarA homolog, is repressed by SarT and regulates virulence genes in *Staphylococcus aureus*. *Infect. Immun.* **2003**, *71* (1), 343–53.
- (76) Brotz-Oesterhelt, H.; et al. Dysregulation of bacterial proteolytic machinery by a new class of antibiotics. *Nat. Med.* **2005**, *11* (10), 1082–7.



RightsLink®

[Home](#) [Create Account](#) [Help](#)



Title: Quantitative Map of β -Lactone-Induced Virulence Regulation
Author: Joanna Krysiak, Matthias Stahl, Jan Vomacka, et al
Publication: Journal of Proteome Research
Publisher: American Chemical Society
Date: Mar 1, 2017
Copyright © 2017, American Chemical Society

[LOGIN](#)
If you're a **copyright.com user**, you can login to RightsLink using your copyright.com credentials. Already a **RightsLink user** or want to [learn more?](#)

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

[BACK](#)

[CLOSE WINDOW](#)

Copyright © 2017 [Copyright Clearance Center, Inc.](#) All Rights Reserved. [Privacy statement.](#) [Terms and Conditions.](#) Comments? We would like to hear from you. E-mail us at customercare@copyright.com

4

A whole proteome inventory of background photocrosslinker binding

Published in *Angewandte Chemie International Edition*, 2017, 56(5), pp 1396-1401
by Philipp Kleiner,[#] Wolfgang Heydenreuter,[#] Matthias Stahl,[#] Vadim S. Korotkov, and
Stephan A. Sieber.

[#] equal contribution

Reprinted with permission. © 2017 John Wiley and Sons. DOI: 10.1002/anie.201605993

SYNOPSIS

Affinity-based protein profiling (AfBPP) is able to identify non-covalent protein binding partners of small molecule compounds. In a usual case, the compound of interest is modified with an alkyne handle and a photocrosslinker moiety. The former enables subsequent click chemistry coupling to fluorophors or affinity tags and the latter allows the formation of a covalent bond upon UV irradiation. In addition to the genuine protein target, usually a myriad of unspecific background binding proteins are captured and falsely identified as a hit. The use of appropriate controls such as competition with the unmodified parent compound or the application of minimal photocrosslinker-bearing probes is inevitable.

Here, the background binders of common photocrosslinker groups are systematically analyzed. The publication especially focusses on diazirine photocrosslinkers, which are commonly used at present. The basis of this analysis consists of four different diazirine-containing compounds with a minimal structure (figure 4.1). Besides the actual photocrosslinking part, the molecules exhibit small aliphatic and aromatic structure motifs. After inserting these molecules into the standard AfBPP enrichment process, potential protein targets could be identified by mass spectrometry.

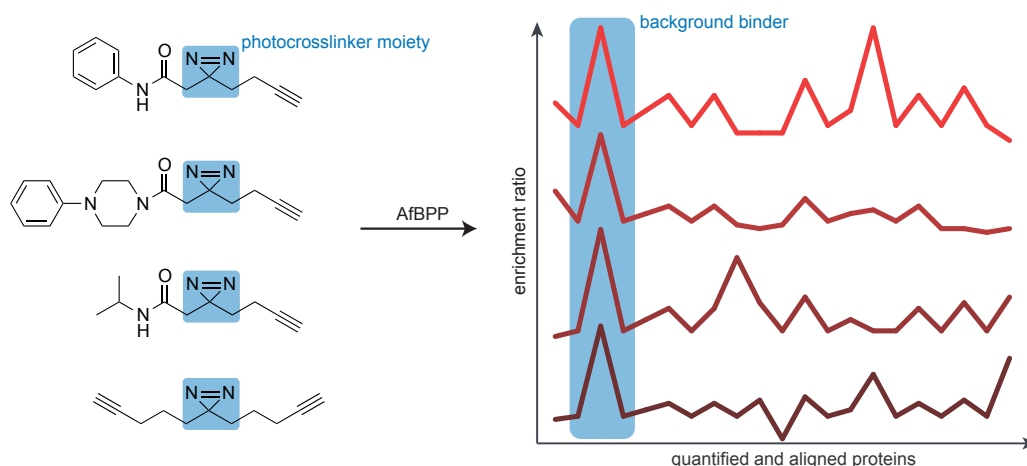


Figure 4.1: Basic idea of extracting background binders from AfBPP experiments. All proteins in each photoprobe experiment are identified and quantified and then aligned. Next, an overlay is calculated to find targets that appear regularly in a majority of the photoprobe experiments.

For example, identified proteins that appeared for all of the four compounds were considered as strong background binders. Obviously, they do not bind to the different structural

motifs, but to the photocrosslinker group, which all of the molecules have in common. In order to quantify this effect, a model from image processing was adopted, in which patterns that appear regularly are consolidated by summing them up. In contrast, patterns that only appear for some of the samples become blurred in a total overlay (figure 4.1). With this method, a negative list of common diazirine background binders in the human cell lines A549 and HeLa could be obtained, which is called the ‘photome’.

Moreover, the concept was applied for the identification of protein targets of the kinase inhibitor **H8** and it was tested using an already published study dealing with binding partners of the natural product falcarinol. In both cases, the number of putative true targets could be reduced by assigning a photome-based confidence score to the hit list.

AUTHOR CONTRIBUTIONS

Philipp Kleiner, Wolfgang Heydenreuter, and Vadim S. Korotkov synthesized compounds. Philipp Kleiner, Wolfgang Heydenreuter, and Matthias Stahl performed biochemical experiments and mass spectrometry. Philipp Kleiner and Matthias Stahl analyzed proteomics data. Matthias Stahl designed the photome assembly pipeline and wrote the analysis overlay software. Stephan A. Sieber supervised the project and wrote the manuscript with input from all authors.

Photoaffinity Labeling

International Edition: DOI: 10.1002/anie.201605993
German Edition: DOI: 10.1002/ange.201605993

A Whole Proteome Inventory of Background Photocrosslinker Binding

Philipp Kleiner⁺, Wolfgang Heydenreuter⁺, Matthias Stahl⁺, Vadim S. Korotkov, and Stephan A. Sieber*

Abstract: Affinity-based protein profiling (AfBPP) is a widely applied method for the target identification of bioactive molecules. Probes containing photocrosslinkers, such as benzophenones, diazirines, and aryl azides, irreversibly link the molecule of interest to its target protein upon irradiation with UV light. Despite their prevalent application, little is known about photocrosslinker-specific off-targets, affecting the reliability of results. Herein, we investigated background protein labeling by gel-free quantitative proteomics. Characteristic off-targets were identified for each photoreactive group and compiled in a comprehensive inventory. In a proof-of-principle study, **H8**, a protein kinase A inhibitor, was equipped with a diazidine moiety. Application of this photoprobe revealed, by alignment with the diazidine background, unprecedented insight into its *in situ* proteome targets. Taken together, our findings guide the identification of biologically relevant binders in photoprobe experiments.

Photocrosslinkers are important tools for chemical biology research and mediate irreversible linkages between proteins and ligands upon UV irradiation, for example.^[1] A prominent application is affinity-based protein profiling (AfBPP) where small molecules are equipped with a photocrosslinker to facilitate irreversible binding to a protein target.^[1b,2] Researchers have the choice between three major photoreactive groups, which differ in their size, efficacy, and possible side reactions, namely benzophenones, aryl azides, and diazirines (Figure 1A).^[1c] Alkyl diazirines,^[3] in particular, have recently experienced a renaissance in application^[4] owing to their small size and high crosslinking efficiency. Furthermore, several studies comparing the reactivity and efficiency of these three photocrosslinkers reported superior properties for diazirines, further encouraging their use.^[5] Despite this popularity, very little is known about photocrosslinker-associated off-target binding. Thus control reactions such as competition with the parent, unmodified ligand are recommended,^[6] however, these experiments are challenging owing to the irreversible nature of the photoprobe–protein interaction while the competing parent compound binds reversibly. In addition, a minimal photoprobe lacking the ligand-specific structure can be applied as a control.^[4b] A

recent study systematically identified off-targets of the three photoreactive groups described above embedded in a minimal scaffold on 2D-SDS gels.^[7] Previous work in our laboratory has also indicated a number of ligand-independent photocrosslinker-specific protein hits.^[8] However, whereas some of them correlated well with the reported gel-based off-targets, others seemed to be detected solely by quantitative gel-free procedures—the state-of-the-art technique in chemical proteomics.

To obtain a comprehensive picture focused on the needs of quantitative and gel-free target identification, we herein report a full inventory of benzophenone, aryl azide, and diazidine off-targets (photomes) in human cell lines by quantitative high-resolution proteomics. Guiding principles to understand off-target preferences and minimize non-specific labeling are also provided. Finally, we demonstrate the value added by our photome profiling study with a photoprobe based on **H8**, a protein kinase A (PKA) inhibitor. The conclusions drawn from this experiment would have been heavily biased by photome hits without our off-target exclusion list.

Our strategy for compiling an inventory of the non-specific background proteome began with the chemical design and synthesis of basic photocrosslinker probes. All three photoreactive groups, aryl azide (AA), benzophenone (BP), and diazidine (DA), were embedded in a simple molecular scaffold to minimize ligand-specific protein interactions. In the case of AA and BP, we appended the alkyne tag, which is required for target identification by click chemistry (CC; Figure 1B),^[9] to an aryl ring with a small aliphatic ether (**AA-1**) or an amide linkage (**BP-1**), respectively (Figure 1A). By definition, AA- and BP-containing photoprobes have an aryl ring in their minimal molecular scaffolds, and for comparison, we designed a minimal aromatic DA probe (**DA-1**) in a similar way. To account for photocrosslinker-independent background labeling, we selected the **DA-1** scaffold lacking the diazidine group (**NC-1**; Figure 1C). All molecules were prepared according to established procedures with slight modifications (see the Supporting Information, Schemes S1 and S2).

To evaluate the proteome binding of **AA-1**, **BP-1**, and **DA-1**, we chose two human cell lines, A549 and HeLa, as representative and commonly used test systems. The photocrosslinker probes were incubated with intact cells for 1 h and subsequently irradiated for 30 min with UV light; these conditions are similar to most common AfBPP labeling procedures. In addition, one sample was treated with the **NC-1** control under identical conditions. First, we investigated the cell permeability of all molecules by gel-based analysis of their labeling pattern in the cytosolic and membrane fractions. Interestingly, all photoprobes predominantly labeled cyto-

[*] M. Sc. P. Kleiner,^[†] Dr. W. Heydenreuter,^[†] M. Sc. M. Stahl,^[†] Dr. V. S. Korotkov, Prof. Dr. S. A. Sieber
Center for Integrated Protein Science at the Department of Chemistry
Technische Universität München
Lichtenbergstrasse 4, 85747 Garching (Germany)
E-mail: stephan.sieber@tum.de

[†] These authors contributed equally to this work.

Supporting information for this article can be found under:
<http://dx.doi.org/10.1002/anie.201605993>.

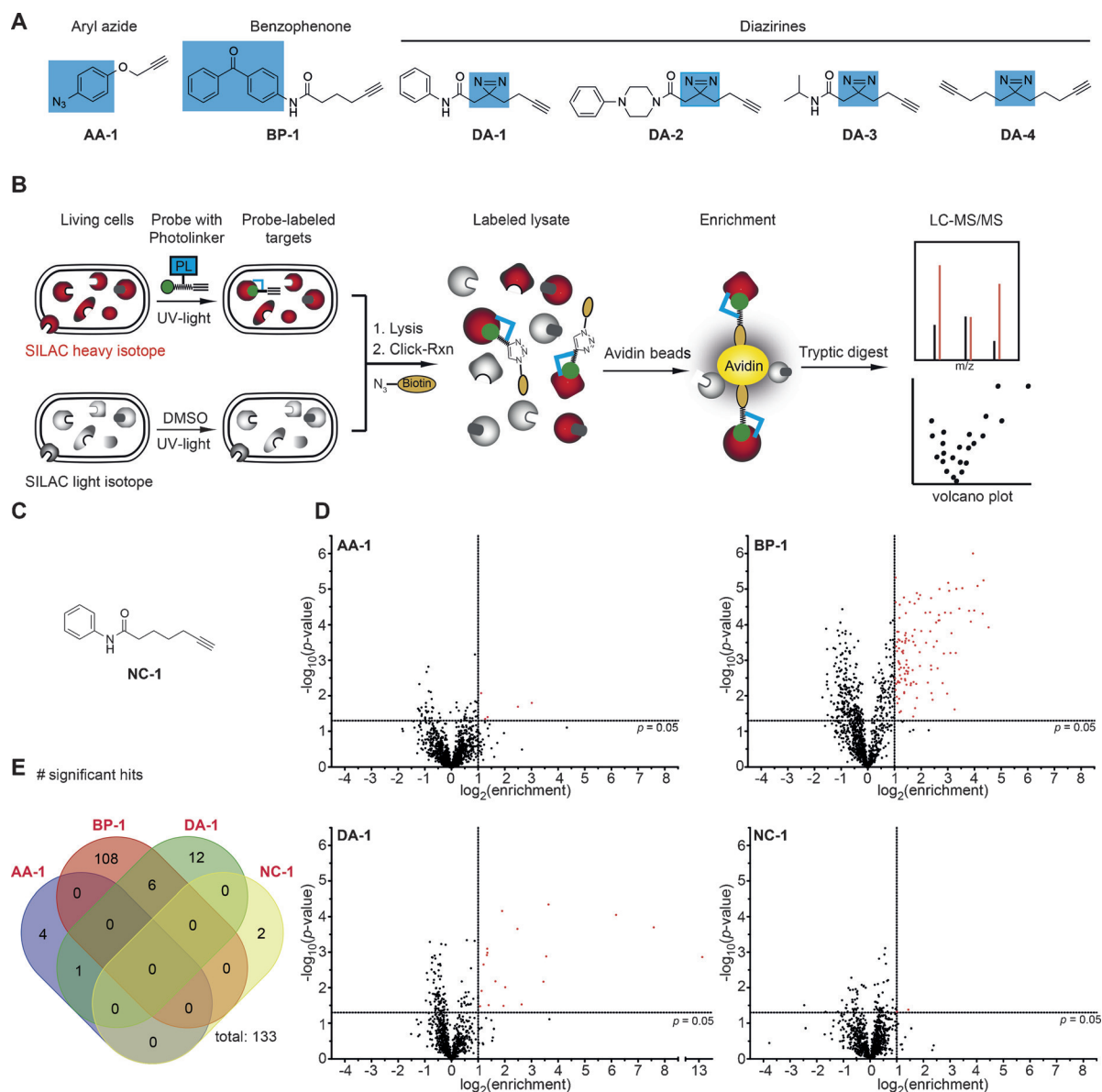


Figure 1. Analysis of photocrosslinker background binding. A) Photocrosslinker probes used in this study. The photoreactive group is highlighted in blue. B) Gel-free photolabeling strategy using SILAC to quantify probe hits over DMSO control. See the Methods Section in the Supporting Information for details. C) Structure of the negative control probe (NC-1) used for the identification of non-photospecific target proteins. D) Volcano plots of enriched proteins in SILAC experiments with the photocrosslinkers AA-1, BP-1, and DA-1 and the control NC-1. A \log_2 enrichment of > 1 implies a double excess of protein in the probe sample compared to the DMSO control. A p -value of < 0.05 was required to classify a hit as significantly enriched (highlighted in red). E) Venn representation showing proteins that were significantly enriched in all four experiments.

solic proteins, suggesting sufficient cell permeability (Figure S1). BP-1 stood out as the probe with the most intense labeling and the highest number of targets while AA-1 and DA-1 exhibited much reduced proteome labeling.

To obtain a more comprehensive inventory of background binders, we switched to gel-free, quantitative proteomics. This technique has become the gold standard in AfBPP, and its high sensitivity requires rigorous characterization of non-specific binders. We thus applied the stable isotope labeling of amino acids in cell culture (SILAC) approach and treated cells labeled with heavy or light isotopes with probe (3 μM) or DMSO, respectively.^[10] Cells were lysed, equal protein amounts from DMSO and probe-treated samples pooled,

and whole proteomes clicked to biotin azide. After enrichment and tryptic digest, the samples were analyzed by high-resolution liquid chromatography tandem mass spectrometry (LC-MS/MS; Figure 1B); \log_2 (isotope ratios) of 0 indicate no enrichment compared to the DMSO control, and \log_2 ratios of > 1 with a p -value of < 0.05 were regarded as significant hits (Figure 1D and Figure S2). Whereas NC-1 showed only marginal protein binding, all three photoprobes exhibited significant protein enrichment that was largely probe-specific. For example, BP-1, previously regarded as a selective photolinker owing to its reversible mode of photoactivation,^[11] exhibited the highest number of protein targets (114) in A549 cells followed by DA-1 (19) and AA-1 (5; Figure 1E).

Interestingly, background labeling in HeLa cells was more evenly distributed with 46, 78, and 85 hits for **AA-1**, **BP-1**, and **DA-1**, highlighting the importance of cell-line-specific off-target analysis (Figure S2). A list of all identified proteins is provided in the Supporting Information.

Considering the increasing application of DA photo-probes in chemical biology, we synthesized three additional minimal probes for in-depth analysis (Figure 1A and Scheme S3). Here, the DA moiety was embedded in different aliphatic or aromatic scaffolds with the aim of dissecting off-target binding and connecting protein hits to structural motifs. All four DAs (**DA-1–DA-4**), including a close mimic of a previously reported control compound,^[4b] were applied in gel-based cellular protein profiling. All probes labeled intracellular targets, suggesting sufficient cell permeability (Figure S3).

Subsequent gel-free, quantitative proteomics studies revealed a significant influence of the scaffold on protein binding (Figure S4). Probe **DA-2**, which contains an aryl piperazine, exhibited the highest labeling in A549 cells followed by the aromatic probe **DA-1** (Figure 2A). In HeLa cells, **DA-1** and aliphatic **DA-4** contributed to the majority of hits (Figure 2B). Interestingly, **DA-3**, with an amide bond and a short aliphatic chain, was most selective in both cell lines, suggesting that affinity probe design could benefit from such short alkyl linkers.

To gather comprehensive information on the most prominent DA-specific background targets, we overlaid the volcano plots for all probes in A549 and HeLa cells. DA-specific protein targets should thus be extensively enriched whereas scaffold-specific hits relocate around zero (Figure 2C; for a detailed explanation, see the Supporting Information). This mathematical operation revealed a set of eight and seven significant proteins for A549 and HeLa cells, respectively, after Benjamini–Hochberg correction (Figure 2D,E). Satisfyingly, 291 out of 1655 protein groups in the A549 proteome were annotated as membrane-bound. Identified hits that are predominantly localized in the lysosome and mitochondrion can be grouped into three superfamilies: 1) channels or channel-associated proteins (e.g., VDAC1 and VDAC2), 2) catabolic enzymes (e.g., CTSD), and 3) small-molecule binders (e.g., ALDH1B1). For example, VDAC1 and VDAC2 are highly abundant channels with permeability for small molecules. Similarly, ECH1 and ALDH1B1 are catabolic enzymes known to bind small molecules. A summary of all non-specific photoprobe targets is provided in Figure 2F and Table S1. To account for label-free MS quantification, a rising proteomic technique independent of SILAC, we applied all DA probes together in the labeling of intact A549 cells to obtain a comparable photome list (Figure S5 and Supporting Information). Again, a good distribution of cytosolic and membrane proteins was obtained while re-extraction of the insoluble fraction did not significantly enhance the overall coverage of membrane-bound proteins (Figure S6).

Interestingly, reported abundances (Table S1) revealed that proteins solely identified in one cell line, such as ECE1 in HeLa cells, are less expressed in the other cell line, likely explaining their cell-line-specific DA labeling. Importantly,

most of these proteins were not identified by gel-based off-target profiling,^[7] and furthermore, hits of gel studies are underrepresented here, confirming the relevance of the proteomic preparation technique for the results of AfBPP studies.

With a comprehensive photome background list in hand, we next analyzed the irradiation and concentration dependence of DA photocrosslinking reactions. For this study, we used the kinase inhibitor **SP600125** equipped with a minimal DA linker, similar as previously reported (Scheme S4).^[4b] Interestingly, while short irradiation times of 5 min and high probe concentrations of 3 μM enhanced photome labeling, longer irradiation times (20–30 min) as well as lower concentrations (125 nM) minimized background enrichment in this setting (Figures S7 and S8). Although JNK, a known target of **SP600125**, was not among the hits, concentration-dependent competition with the unmodified drug revealed several carboxylases that also exhibit ATP binding sites as putative targets (Figure S9).

For an initial evaluation of the photome list, we re-analyzed previous results obtained from A549 profiling with a falcarinol diazirine probe.^[8a] A comparison of the proteins identified with the probe and those found in the photome classified four out of seven falcarinol hits, including HMOX2 and CTSD, as photome targets (Figure S10). Importantly, ALDH2, a biochemically confirmed target of falcarinol, stood out from the photome proteins, which confirms the general utility of this approach for unbiased target identification. Finally, we sought to illustrate the applicability of the photome inventory list in a target identification study with isoquinoline sulfonamide **H8** (Figure 3A), a widely applied inhibitor of the cAMP-dependent protein kinase (PKA).^[12] **H8** and its structurally related analogue **H89** are gold standards in PKA research,^[12b] and AfBPP with this molecule should reveal PKA as a validated positive control.

Based on the lessons learned from the photocrosslinker comparisons described above, we equipped **H8** with a short aliphatic diazirine alkyne linker (Figure 3B), found to exhibit minimal background labeling, at the terminal amine. Crystallography data of PKA in complex with H-inhibitors further indicate that chain extension at this position is tolerated.^[13] However, **H8p** might alternatively directly bind to the cAMP-binding pocket of the regulatory subunit. The synthesis of this probe was conducted through the coupling of the alkyl diazirine carboxylic acid **L4** (for the structure, see Scheme S1) and *N*-(2-aminoethyl)isoquinoline-5-sulfonamide (**2**), which was prepared according to published procedures,^[14] in the final step (Figure 3B).

A549 SILAC cells were treated with **H8p** and DMSO according to established procedures. Proteomics workflow and analysis revealed the enrichment of 30 proteins matching our cut-off criteria (Figure 3C). These hits were classified into three categories according to their enrichment in the photome overlay (Figure 2D): 1) low confidence (five hits depicted in red: $\log_2(\text{enrichment})$ in the photome >1 and $p < 0.05$), 2) medium confidence (six hits depicted in orange: $0 < \log_2(\text{enrichment}) \leq 1$ and $p < 0.05$), 3) high confidence (19 remaining hits depicted in gray, which are not significantly enriched in the photome). Interestingly, medium- and high-

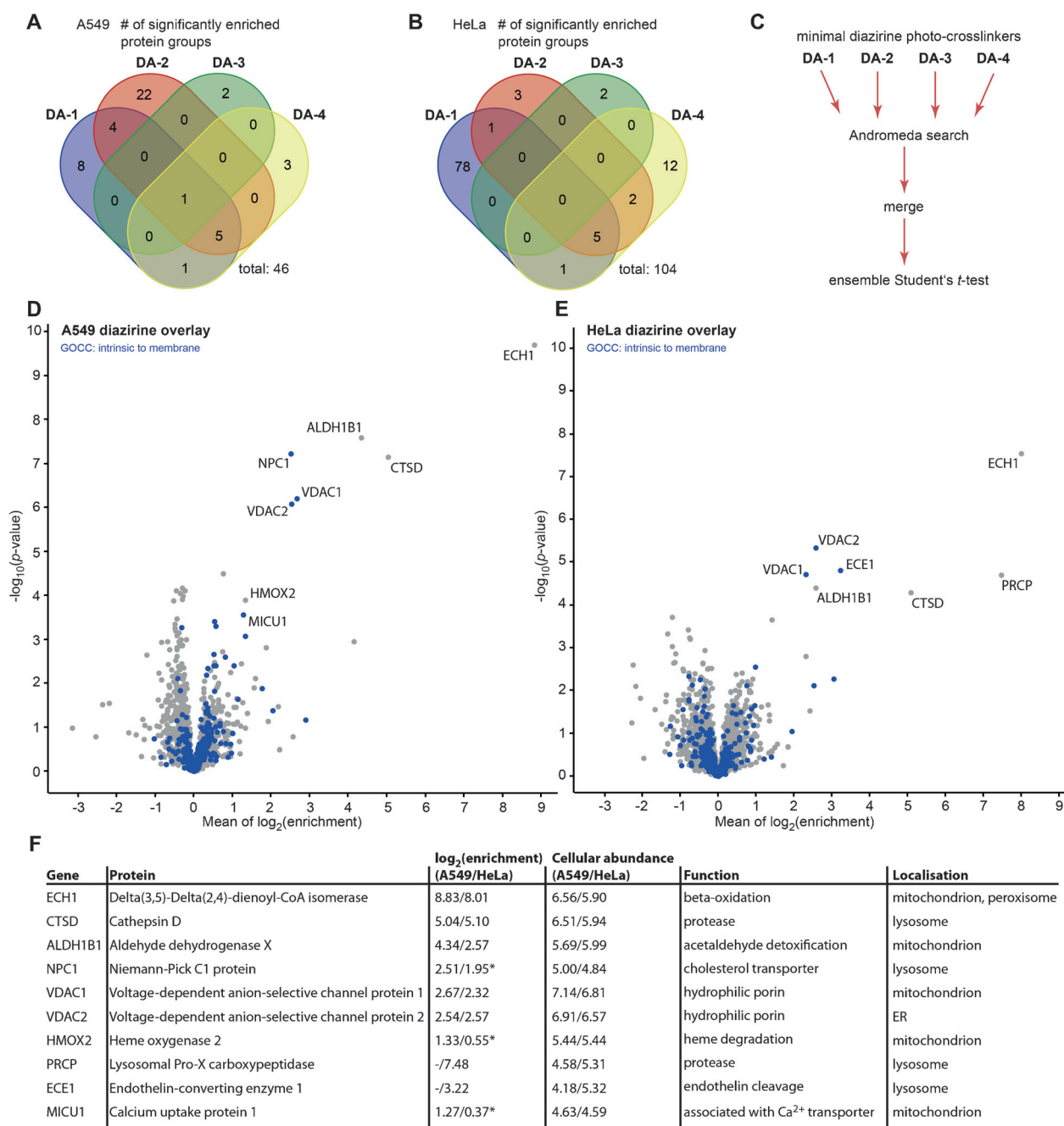


Figure 2. Background photoaffinity labeling of structurally diverse diazirines. A) Venn diagram of overlapping significantly enriched hits ($\log_2(\text{enrichment}) > 1$ and $p < 0.05$) for all diazirine probes in live A549 cells. B) Venn diagram for the analysis in live HeLa cells. C) General workflow to yield the combined diazirine photome. Four distinct diazirine proteome enrichments were analyzed individually by MaxQuant^[16] and subsequently merged (see the Methods Section in the Supporting Information). The output file was then statistically evaluated by a one-sample Student's *t*-test to gather mean enrichments and *p*-values over all valid quantifications. D) Volcano plot of the combined AfBPP results in A549 cells. To compile a master list of significant photome members, the overall significance level was adjusted by the method of Benjamini and Hochberg (FDR = 0.05). E) Volcano plot for the experiments with HeLa cells. F) List of top photome hits.

confidence proteins were predominantly linked to nucleotide binding, such as ATP/cAMP, DNA, and polyadenine RNA (Table S2). Furthermore, the regulatory subunit of PKA (PRKAR2A) was identified.

In conclusion, our study has revealed the superior performance of AA and DA compared to BP photocrosslinkers, which is in agreement with previous experiments

comparing crosslinking efficiencies.^[5] A comparison of four different DA scaffolds revealed a common set of false positives. Proteins of high abundance and with a preference for small-molecule binding belong to the most prominent hits. Small linkers with aliphatic DAs led to the least binding while aromatic substituents significantly increased the number of hits. This information complements recent findings indicating

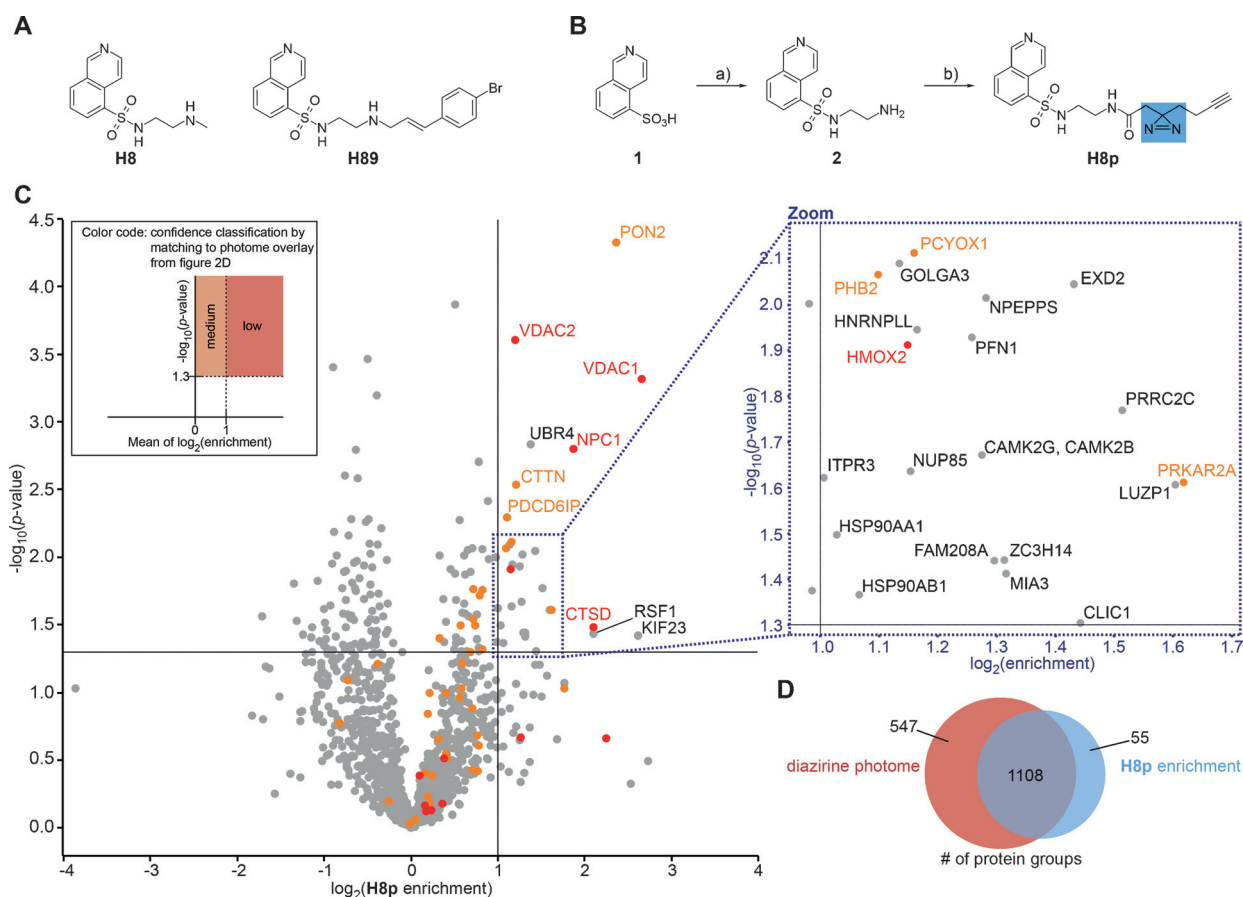


Figure 3. AFBPP of kinase inhibitor photoprobe **H8p** and photome-based confidence validation. A) Structures of two members of the H-series of protein kinase A inhibitors: **H8** and **H89**. B) Synthesis of **H8p**, a photoaffinity probe for **H8**. Reaction conditions: a) 1) SOCl_2 (35.0 equiv), DMF, reflux, 2 h; 2) NaHCO_3 (1.00 equiv), H_2O , 0°C ; 3) ethylenediamine (3.00 equiv), CH_2Cl_2 , $0^\circ\text{C} \rightarrow \text{RT}$, 2 h, 52% over two steps; b) **L4** (1.00 equiv), HOBt (1.50 equiv), EDC (1.50 equiv), DIEA (2.00 equiv), DMF, RT, 16 h, 93%. C) Volcano plot representation of **H8p** AFBPP with $3 \mu\text{M}$ probe (blow up of proteins located in the blue box shown on the right). Identified proteins were aligned with the diazirine photome (Figure 2D) and classified as low-confidence (red; $\log_2(\text{photome enrichment}) > 1$, $p < 0.05$), medium-confidence (orange; $0 < \log_2(\text{photome enrichment}) \leq 1$, $p < 0.05$), or high-confidence (gray, remaining proteins) hits. See the color code for a graphical representation of the hit positions in the photome. D) Overlap of proteins present in the photome (red) and proteins identified in the **H8p** enrichment (blue). The regulatory subunit of PKA (PRKAR2A) was thus identified. The compilation of the observed targets would have been impossible without our inventory list, once more highlighting the importance of minimal photolinker controls for the exclusion of common non-specific hits.

that branched photolinker scaffolds exhibit superior specificity.^[15] While it was possible to minimize off-target labeling within this isolated photocrosslinking study by reducing the probe size and lowering $\log P$ (Table S3), in practice, photoprobe design is limited by the molecular scaffold of interest. Here, it becomes important to attach these groups at sites so that the bioactivity and, if possible, the physicochemical properties are not affected. By applying these design criteria to a PKA inhibitor, we demonstrated successful target identification and the necessity for a photocrosslinker-specific exclusion list. Thus we highly recommend the use of minimal photocrosslinkers and their photome inventory list as an important standard control for such experiments.

Acknowledgements

We are grateful for funding from the Deutsche Forschungsgemeinschaft, CIPS^M (to S.A.S.), and the European Research

Council (ERC starting grant to S.A.S.). M.S. was supported by the Studienstiftung des deutschen Volkes. We thank Markus Lakemeyer for helpful discussions and Katja Bäuml, Mona Wolff, Daniel Nixdorf, Jacqueline Kissel, and Karolina Koch for help with experiments and mass spectrometry. We thank Megan H. Wright for critical proofreading of the manuscript.

Keywords: enzymes · photoaffinity labeling · photoreactive probes · protein profiling · proteomics

How to cite: *Angew. Chem. Int. Ed.* **2017**, *56*, 1396–1401
Angew. Chem. **2017**, *129*, 1417–1422

- [1] a) E. Smith, I. Collins, *Future Med. Chem.* **2015**, *7*, 159–183; b) P. P. Geurink, L. M. Prely, G. A. van der Marel, R. Bischoff, H. S. Overkleeft, *Top. Curr. Chem.* **2012**, *324*, 85–113; c) D. J. Lapinsky, D. S. Johnson, *Future Med. Chem.* **2015**, *7*, 2143–2171.
[2] a) A. Saghatelian, N. Jessani, A. Joseph, M. Humphrey, B. F. Cravatt, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 10000–10005; b) D. J. Lapinsky, *Bioorg. Med. Chem.* **2012**, *20*, 6237–6247.

- [3] a) L. Dubinsky, B. P. Krom, M. M. Meijler, *Bioorg. Med. Chem.* **2012**, *20*, 554–570; b) J. Das, *Chem. Rev.* **2011**, *111*, 4405–4417.
- [4] a) A. L. MacKinnon, J. L. Garrison, R. S. Hegde, J. Taunton, *J. Am. Chem. Soc.* **2007**, *129*, 14560–14561; b) Z. Li, P. Hao, L. Li, C. Y. Tan, X. Cheng, G. Y. Chen, S. K. Sze, H. M. Shen, S. Q. Yao, *Angew. Chem. Int. Ed.* **2013**, *52*, 8551–8556; *Angew. Chem.* **2013**, *125*, 8713–8718.
- [5] a) K. Sakurai, S. Ozawa, R. Yamada, T. Yasui, S. Mizuno, *ChemBioChem* **2014**, *15*, 1399–1403; b) J. T. Bush, L. J. Walport, J. F. McGouran, I. K. H. Leung, G. Berridge, S. S. Van Berkel, A. Basak, B. M. Kessler, C. J. Schofield, *Chem. Sci.* **2013**, *4*, 4115–4120.
- [6] A. L. Mackinnon, J. Taunton, *Curr. Protoc. Chem. Biol.* **2009**, *1*, 55–73.
- [7] J. Park, M. Koh, J. Y. Koo, S. Lee, S. B. Park, *ACS Chem. Biol.* **2016**, *11*, 44–52.
- [8] a) W. Heydenreuter, E. Kunold, S. A. Sieber, *Chem. Commun.* **2015**, *51*, 15784–15787; b) J. Eirich, S. Braig, L. Schyschka, P. Servatius, J. Hoffmann, S. Hecht, S. Fulda, S. Zahler, I. Antes, U. Kazmaier, S. A. Sieber, A. M. Vollmar, *Angew. Chem. Int. Ed.* **2014**, *53*, 12960–12965; *Angew. Chem.* **2014**, *126*, 13174–13179.
- [9] a) C. W. Tornøe, C. Christensen, M. Meldal, *J. Org. Chem.* **2002**, *67*, 3057–3064; b) V. V. Rostovtsev, J. G. Green, V. V. Fokin, K. B. Sharpless, *Angew. Chem. Int. Ed.* **2002**, *41*, 2596–2599; *Angew. Chem.* **2002**, *114*, 2708–2711.
- [10] S. E. Ong, B. Blagoev, I. Kratchmarova, D. B. Kristensen, H. Steen, A. Pandey, M. Mann, *Mol. Cell. Proteomics* **2002**, *1*, 376–386.
- [11] A. Kawamura, D. M. Mihai, *Methods Mol. Biol.* **2012**, *803*, 65–75.
- [12] a) M. Inagaki, S. Kawamoto, H. Itoh, M. Saitoh, M. Hagiwara, J. Takahashi, H. Hidaka, *Mol. Pharmacol.* **1986**, *29*, 577–581; b) A. Lochner, J. A. Moolman, *Cardiovasc. Drug Rev.* **2006**, *24*, 261–274.
- [13] R. A. Engh, A. Girod, V. Kinzel, R. Huber, D. Bossemeyer, *J. Biol. Chem.* **1996**, *271*, 26157–26164.
- [14] A. Morikawa, T. Sone, T. Asano, *J. Med. Chem.* **1989**, *32*, 42–46.
- [15] H. Park, J. Y. Koo, Y. V. Srikanth, S. Oh, J. Lee, J. Park, S. B. Park, *Chem. Commun.* **2016**, *52*, 5828–5831.
- [16] MaxQuant enables high peptide identification rates, individualized ppb range mass accuracies, and proteome-wide protein quantification; see: J. Cox, M. Mann, *Nat. Biotechnol.* **2008**, *26*, 1367–1372.

Manuscript received: June 20, 2016

Revised: September 15, 2016

Final Article published: December 16, 2016

JOHN WILEY AND SONS LICENSE TERMS AND CONDITIONS

Jan 16, 2018

This Agreement between Mr. Matthias Stahl ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	4271080580420
License date	Jan 16, 2018
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Angewandte Chemie International Edition
Licensed Content Title	A Whole Proteome Inventory of Background Photocrosslinker Binding
Licensed Content Author	Philipp Kleiner,Wolfgang Heydenreuter,Matthias Stahl,Vadim S. Korotkov,Stephan A. Sieber
Licensed Content Date	Dec 16, 2016
Licensed Content Pages	6
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Print and electronic
Portion	Full article
Will you be translating?	No
Title of your thesis / dissertation	Chemical protein manipulations and their application in proteomics data mining
Expected completion date	Feb 2018
Expected size (number of pages)	170
Requestor Location	Mr. Matthias Stahl Lichtenbergstraße 4 Garching, Bayern 85748 Germany Attn: Mr. Matthias Stahl
Publisher Tax ID	EU826007151
Total	0.00 EUR

[Terms and Conditions](#)

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a "Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing

5

ProteomeDiver: Mining peptide intensity profiles

5.1 THE NEXT LEVEL OF CHEMICAL PROTEOMICS

5.1.1 MASS SPECTROMETRY IN CHEMICAL BIOLOGY

Mass spectrometry is the method of choice to globally analyze cellular effects of small molecules on a proteome or to identify protein targets in activity- as well as affinity-based protein profiling. Even for non-global analyses such as binding site identification, mass spectrometry has pushed the field of chemical biology immensely forward in recent years (reviewed in¹). For example, Cravatt and coworkers have recently established methods to probe the ligandability of cysteine and lysine residues on a proteome-wide level. Therefore, they applied a combination of reactive and side-chain directed small molecule compounds with quantitative tandem mass spectrometry based on isotope-labeled enrichment linkers.² Another example from the Tate group shows the incorporation of modified metabolites, such as the myristate fatty acid, into enzymatic processes. In this case, the authors demonstrated the inhibition and chemical tracing of *N*-myristoyltransferases with the aim to explore global protein myristoylation in the Malaria pathogen *Plasmodium falciparum*.³

5.1.2 FROM PROTEINS BACK TO PEPTIDES

For the vast majority, the analysis of chemical biology-related mass spectrometry data is limited to the identification and quantification of proteins. Follow-up studies additionally comprise the localization of small molecule binding sites or the identification of posttranslational modifications, such as phosphorylation sites. However, by nature, mass spectrometry does not identify and quantify proteins but peptides. In fact, the proteins are inferred from peptides. This procedure is accompanied by a certain loss of information. For example, one peptide can be part of multiple proteins. The so called protein inference problem has been tackled in different ways.⁴ One common approach, which is used by the MaxQuant software, assigns ambiguous peptides to the protein group with the most other peptides (Occam's razor principle).^{5,6} Despite this drawback of protein inference, the peptide-centric analysis of mass spectrometric data has widely been neglected so far.

For example, focusing on peptides instead of proteins will open a powerful view on activity-based protein profiling followed by protein target identification via mass spectrometry. In a standard enrichment workflow, alkyne-tagged probes bind to their protein target and are subsequently coupled to a biotin affinity tag via click chemistry. This

enables the enrichment of probe-bound proteins on an immobilized avidin resin (cf. chapter 5). Next, proteins are digested on-bead to peptides, which are further submitted to mass spectrometry.⁷ In this scenario, the peptide including the labeled amino acid would remain on the beads (figure 5.1a). Given sufficient sequence coverage in the final proteomics experiment, the probe binding peptide can be directly identified as a missing peptide.

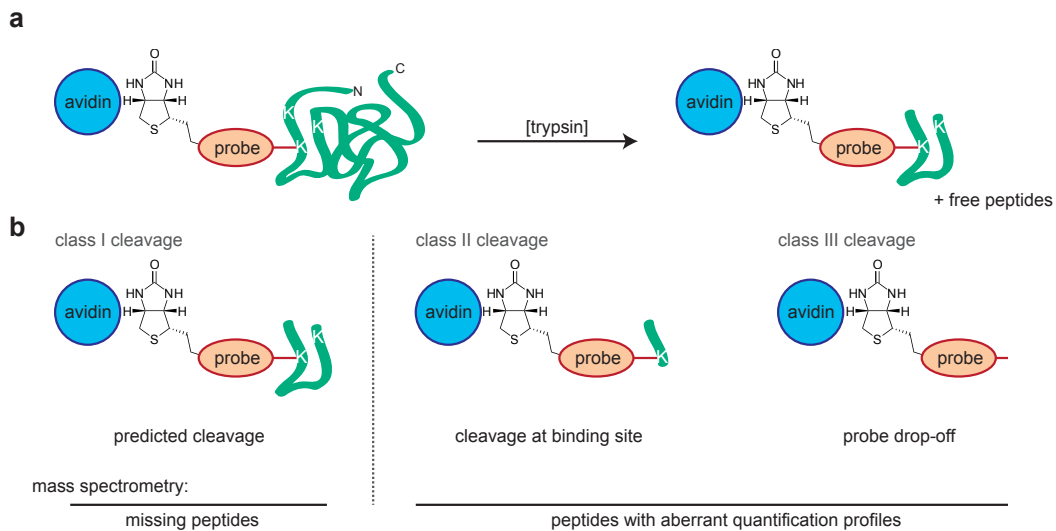


Figure 5.1: Theoretical possibilities of immobilized and enzyme-mediated protein digestion. (a) A probe-captured target protein is bound via biotin to avidin-coated beads. The probe itself is here bound to the protein via a lysine side chain. Trypsin digestion cuts off peptides, but is generally thought to skip the binding site lysine. (b) The predicted cleavages as outlined in panel a are denoted as class I cleavages. But trypsin might also be able to attack the binding site lysine (class II cleavage) or the probe drops off the protein completely (class III cleavage). Molecules remaining with the immobilization beads are shown.

Assuming that the probe binding site is also a protease cleavage site (such as the lysine in figure 5.1), the usual cleavage reaction is here denoted as class I cleavage. However, it is not fully excluded that an enzyme also recognizes the occupied binding site and cleaves at this position (figure 5.1b, class II cleavage) or the probe drops off the protein completely (class III cleavage). Importantly, the three reactions are supposed to have deviating kinetics because of individual chemical environments. Therefore, the velocities of peptide formation through class I, II, or III cleavages should be different (equation 5.1).

$$\frac{d[\text{peptide}^{\text{class I}}]}{dt} \neq \frac{d[\text{peptide}^{\text{class II}}]}{dt} \neq \frac{d[\text{peptide}^{\text{class III}}]}{dt} \quad (5.1)$$

This effect can be visualized, e.g. when using different probe concentrations in the initial ABPP experiment, while digestion times are being kept equal. As a consequence, resulting peptide intensity profiles over various probe concentrations applied, should then be distinguishable by their negative correlation to each other. Conversely, this phenomenon was described before in the context of protein quantification algorithms. There, peptides with outlier profiles ('incoherent peptides')⁸ are usually omitted from further analyses.^{8,9} In this work, exactly these peptides will be of special interest.

5.1.3 OBJECTIVE

Many research groups employ MaxQuant as software platform to analyze their mass spectrometry proteomics data.⁵ The software comprises well established algorithms for peptide identification and label-free protein quantification.¹⁰ The results are given in a set of tables and they can be searched using the MaxQuant Viewer.¹¹ However, the viewer does not allow the inspection of specific peptide profiles over different experiments. Furthermore, it also does not include statistics data, such as results of a Student's *t*-test, to identify significantly enriched proteins.

Here, an extendable software platform called ProteomeDiver was developed that allows reading in (i) MaxQuant result files, (ii) full protein sequences from databases, and (iii) statistical data, e.g. from the Perseus environment.¹² The core feature is an intuitive visualization of single protein sequences and their assigned peptides including modifications. ProteomeDiver should greatly facilitate the access to protein-specific peptide intensity profiles. As outlined above, they will provide insights into the experimental behavior of different protein segments and their cleavage in sample preparation procedures. In a proof-of-principle study with data from a pyridoxal phosphate-based pull-down, the applicability of the software could be demonstrated.

Preliminary note: The software is still under construction and in the testing phase. Thus, all results shown here, especially the ones of the peptide correlation analysis, are of preliminary nature.

5.2 DESIGN OF THE PROTEOMEDIVER SOFTWARE

5.2.1 GENERAL SETUP

ProteomeDiver was developed in Java, because this language enables the creation of easily extendible modules. Moreover, it provides a customizable and flexible XML-based visualization framework with JavaFX.

ProteomeDiver consists of three main parts:

1. an input unit to read in and deserialize MaxQuant result files,
2. a visualization unit to present the data in an intuitive manner, and
3. a peptide profile analysis unit.

Single units are interconnected via standardized interfaces and are independent of each other. For example, input units for other file types can easily be added. Vice versa, it is possible to use parts of ProteomeDiver in other software projects. In the following sections, only the very basic programming concepts used are exemplified. The full code base can be accessed on GitHub at <https://github.com/mtstahl/subsurface>.

5.2.2 LOADING OF MAXQUANT RESULT FILES

ProteomeDiver reads and processes the following MaxQuant result files: proteinGroups.txt, peptides.txt, evidence.txt, summary.txt, and *Sites.txt (the file names are expanded by the name of given chemical amino acid modifications). Additionally, it is able to read simple text file formats, such as Perseus outputs or fasta files. Each file is read in by a dedicated class that implements the *MQReader* interface.

```
public interface MQReader {
    public boolean fileExists(String txtDirectory, String prefix);
    public <E extends Enum<E> & TableHeaders> List<AnalysisComponent> read(String
        txtDirectory, Map<E, String> headerMap) throws IOException;
    public ReadOnlyDoubleProperty getProgressProperty();
    public ReadOnlyStringProperty getStatusProperty();
}
```

Listing 5.1: *MQReader* interface setting the prerequisites for all MaxQuant file readers. The *fileExists* method can be used to check the data availability before running the parser. The central method *read* takes the location of files and desired column headers and returns deserialized data objects. The last two methods return properties, which can inform the user about the status of parsing.

The core method *read* comprises all necessary functions for data deserialization. Besides the location of a folder, where the MaxQuant result files are stored, it takes standardized table headers. These are search engine-independent header names for e.g. the protein name column or the intensity columns. Only requested headers are imported from the complex result tables. The function then returns a list with elements of type *AnalysisComponent*. This is a marker interface for different data wrapper classes such as the *ProteinGroup* or *StatisticsFile* classes.

Accordingly, statistics files are parsed with readers defined by the *StatisticsReader* interface.

```
public interface StatisticsReader {
    public StatisticsFile read(String filePath, Map<StatisticsTableHeaders, String>
        headerMap) throws IOException;
    public ReadOnlyDoubleProperty getProgressProperty();
    public ReadOnlyStringProperty getStatusProperty();
}
```

Listing 5.2: *StatisticsReader* interface that sets the prerequisites for statistics file readers such as the *PerseusFileReader*.

Fasta files are imported by the *FastaFileReader* satisfying the specifications given by the *FastaReader* interface.

```
public interface StatisticsReader {
    public StatisticsFile read(String filePath, Map<StatisticsTableHeaders, String>
        headerMap) throws IOException;
    public ReadOnlyDoubleProperty getProgressProperty();
    public ReadOnlyStringProperty getStatusProperty();
}
```

Listing 5.3: *FastaReader* interface that sets the prerequisites for fasta file readers such as the *FastaFileReader*.

Both the *StatisticsReader* and *FastaReader* interfaces resign a *fileExists* method because usually, the user selects the respective files directly. Contrary, for MaxQuant analyses, the user opens a folder comprising all results and is not aware, which files are present and which are not.

After the reading process, each file type is deserialized to generic objects, e.g. stemming from the *ProteinGroup* class for proteins or the *StatisticsFile* class for files containing statistics information. The diagram in figure 5.2 gives an overview over the reader classes and the classes used to represent analysis objects.

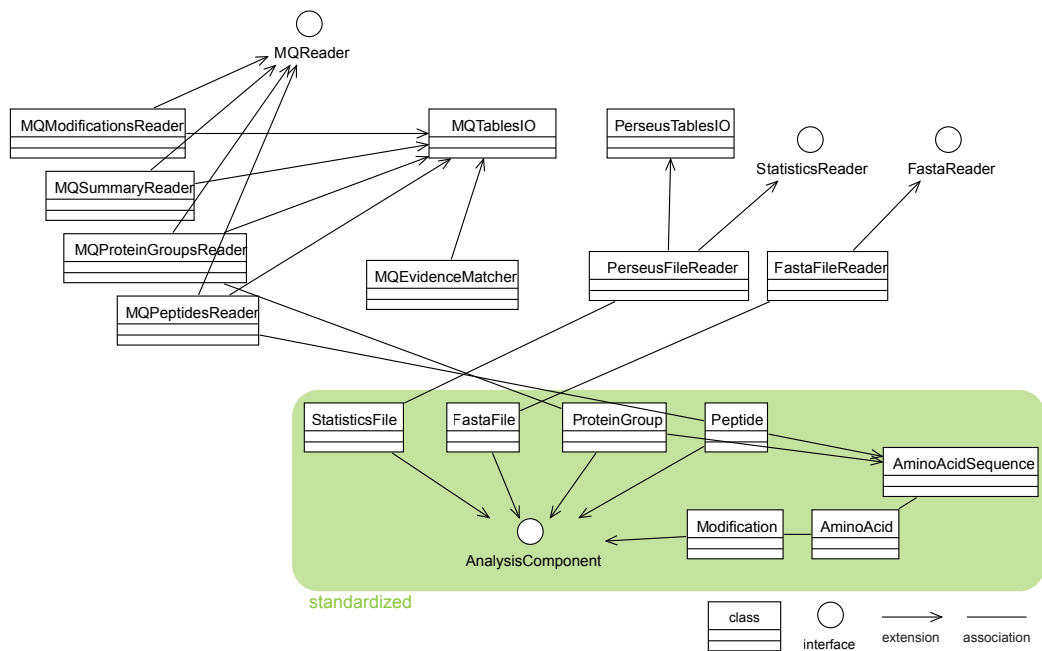


Figure 5.2: Class diagram of the input unit of ProteomeDiver for the deserialization of MaxQuant result files, Perseus-generated statistics files and fasta files. The MaxQuant readers implement the *MQReader* interface, the Perseus file reader implements the *StatisticsReader* interface, and the fasta file reader is derived from the *FastaReader* interface to ensure a high degree of standardization. The MaxQuant readers inherit table information from the abstract class *MQTablesIO*. A similar concept is applied for the *PerseusFileReader*. Some functional data wrapper classes are shown at the bottom. For example, the *Peptide* class represents a peptide and the *MQEvidenceMatcher* class is a helper class providing matching functions, e.g. to enrich the peptide objects with additional information such as intensities measured in different experiments. The green area denotes classes that are standardized meaning they are independent of the origin of input files.

Noteworthy, *MQTablesIO* is an abstract class that comes along with the standard MaxQuant result file names, standard table headers and association maps in order to convert MaxQuant table header names to standardized table headers. Thus, search engine-specific reader classes are responsible for data standardization. The data wrapper classes, i.e. *ProteinGroup* etc., are no more dependent on the origin of data.

5.2.3 DATA VISUALIZATION AND FIRST STEPS

The whole graphical user interface (GUI) of ProteomeDiver is realized with the JavaFX framework. Window layouts and styles are defined in XML (extensible markup language) files and CSS (cascading style sheets) files, respectively, and are thus separated from actual content.

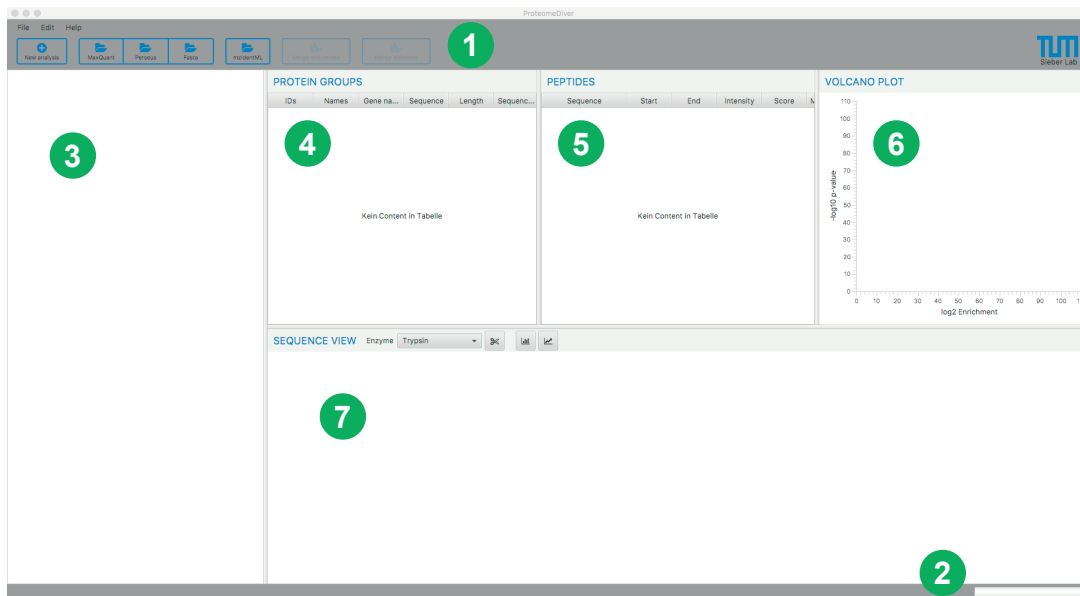


Figure 5.3: Main window of ProteomeDiver. The upper tool bar ① contains all options to load and work with data files. The progress bar on the bottom ② gives a live feedback of the import and deserialization process. The white space on the left ③ will show the collection of loaded analyses and files as tree view after having them imported. The upper panel on the right can show tables of protein groups ④ and peptides ⑤. When a statistics file with enrichment and p -values was loaded, a volcano plot can be displayed ⑥. The lower panel is the sequence view ⑦, which enables the representation of single protein sequences and their measured and assigned peptides.

After loading MaxQuant result files (peptide identification and quantification), a Perseus-generated statistics file, and an appropriate fasta database containing full protein sequences, the three file types can be merged. To this end, the respective analysis components in the tree view must be marked. By clicking the options *Merge sequences* or *Merge statistics* in the tool bar, the particular data from *StatisticsFile* or *FastaFile* objects is linked to the peptide identification analysis. In fact, respective data is transferred from *StatisticsFile* or *FastaFile* objects to the appropriate *ProteinGroup* objects.

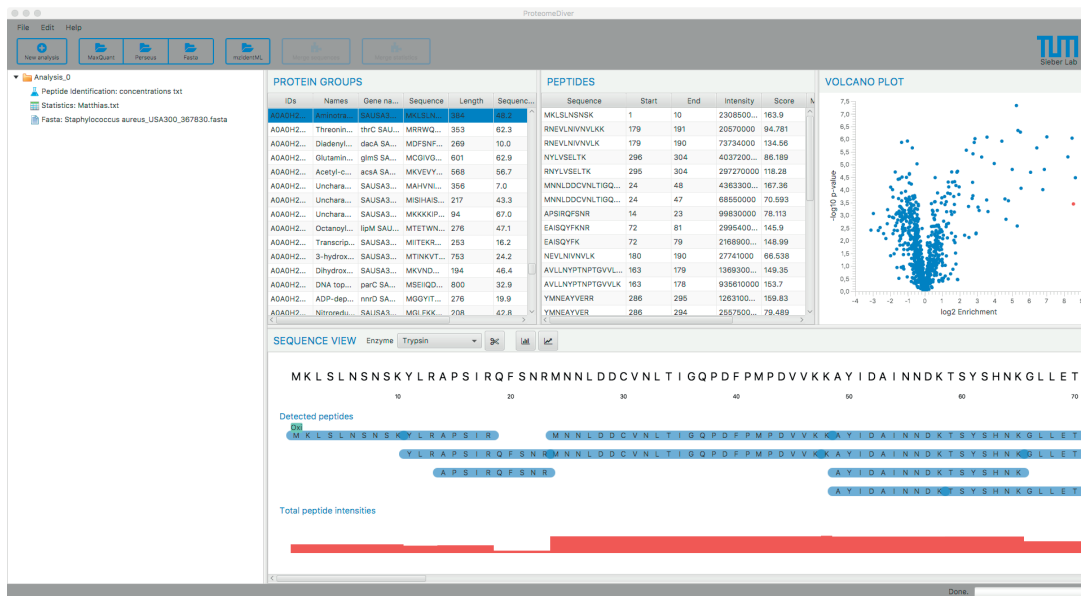



Figure 5.4: ProteomeDiver main window with one analysis loaded. The tree view on the left represents the three loaded data categories: MaxQuant results, Perseus statistics, and fasta database. Data was merged using the two different merge buttons on the tool bar. The protein groups table view depicts all identified proteins. By clicking on one entry, the peptides table view delineates all peptide identifications belonging to the selected protein. The volcano plot on the very right implements the visualization of statistics data. The selected protein is highlighted in red. Proteins can also be chosen via the volcano plot allowing easy selection of proteins of interest. Like the peptides table, the sequence view panel is active, if one specific protein is selected. It highlights the localization of identified peptides on the full protein sequence. An optional oxidative modification at the leading methionine can be observed.

Internally, ProteomeDiver handles all analysis components of one comprehensive analysis shown in the tree view in an analysis wrapper object called *Analysis*. The software is able to deal with multiple analyses, such as multiple MaxQuant runs, at once. Data merges are possible between different analyses. For example, in figure 5.4, all loaded data is organized under one *Analysis* object. The *Analysis* objects themselves are finally collected in another wrapper class named *AnalysisHandler*. This class exhibits a singleton pattern and is therefore only instantiated once and can be considered as the root of all imported data.

5.2.4 PEPTIDE INTENSITY PROFILE ANALYSIS

At this stage, ProteomeDiver supplies a solid platform for the virtualization and visualization of MaxQuant-generated peptide identifications and their quantifications. This is the basis for deeper analyses of peptide intensity profiles. ProteomeDiver is able to visualize experiment-specific intensities of peptides from selected proteins. The built-in profile analysis view is evoked by clicking the  button in the sequence view panel.

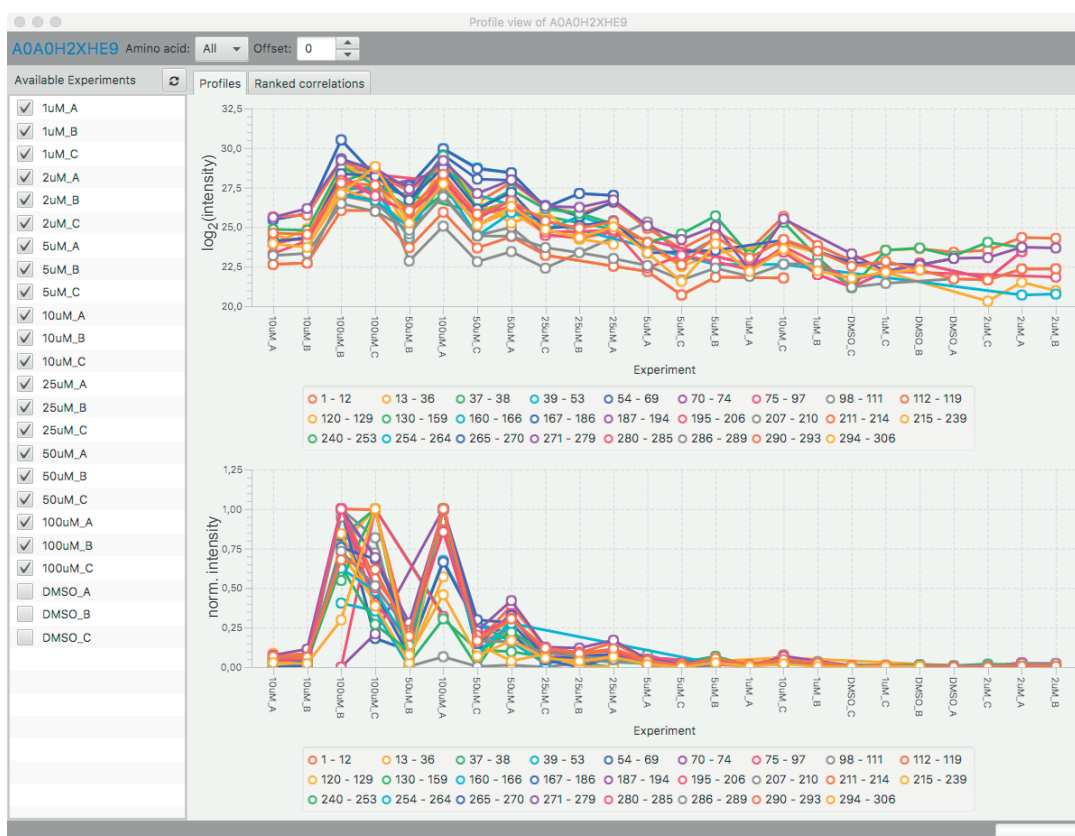


Figure 5.5: Segment intensity profile view for one specific protein. The list on the left specifies available experiments in the analysis. Here, a set of 24 experiments is shown. Only checked features will contribute to the profile correlation analysis. ProteomeDiver then subdivides the protein sequence into segments with equal summed peptide intensities (see figure 5.6). Each segment is shown as two line plot over different experiments. Here, varying probe concentrations were applied in an ABPP experiment. The upper diagram represents logarithmized segment intensities and the lower diagram shows the same data, but normalized to the maximum and minimum intensities of each segment. It is also possible to focus on specific amino acids in the segments. Therefore, an amino acid drop-down menu on the very top of the window is provided. For example, when selecting lysine, only segments that include a lysine residue are shown in the diagrams.

It is important to note that ProteomeDiver does not directly show the intensities of peptides, but of protein segments that result from the addition of overlapping peptide intensities. For example, in case of three overlapping peptides, the software generates four segments according to figure 5.6. The usage of intensity segments rather than the original peptides condenses the data and facilitates subsequent segment intensity correlation analyses.

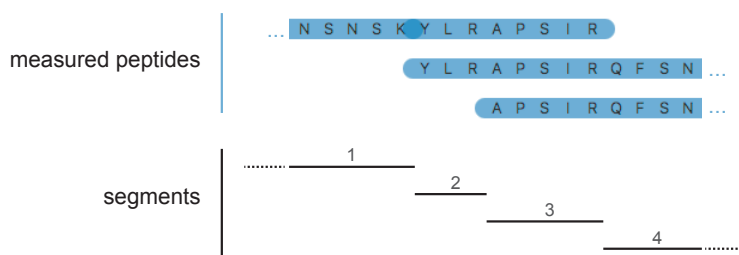


Figure 5.6: Intensity segment computation scheme. The protein sequence is scanned amino acid-wise. For each position, peptide intensity data from all experiments is evaluated. If there is no difference to the previous position, the amino acid is assigned to the same segment. If a difference occurs, a new segment is created.

The comparison of segment intensity profiles will finally deliver a measure for similar and dissimilar behaving protein regions. Again, the origin of putative outlier segments can be unconventional cleavage reactions during the sample preparation workflow as presented in 5.1. Notably, systematic errors during experiment handling, such as pipetting inaccuracies, will concern all peptides in an experiment and thus will never alter the relative segment intensity profiles studied here.

5.3 EXAMPLE: THE QUEST FOR PYRIDOXAL PHOSPHATE-DEPENDENT ENZYMES AND COFACTOR BINDING SITES

Pyridoxal phosphate (PLP, vitamin B₆) is an essential factor in certain enzyme classes such as amino transferases.¹³ Annabelle Hoegl from the research group of Prof. Dr. Sieber at the TU München developed a chemical proteomics strategy to capture PLP-dependent enzymes from living *S. aureus* bacteria (unpublished research). The generalized workflow followed the procedure of activity-based protein profiling and made use of prodrug-like PLP mimics that are first phosphorylated in the bacterial cytoplasm (figure 5.7). This trick facilitates membrane crossing and helps accumulating the active drug in the cells. However, the reversible nature of probe binding to lysine residues of the target proteins is an obstacle for standard ABPP as the probe would fall off during the enrichment steps. Therefore, the Schiff base linkage was transformed into a stable covalent bond by NaBH₄-mediated reduction.¹⁴

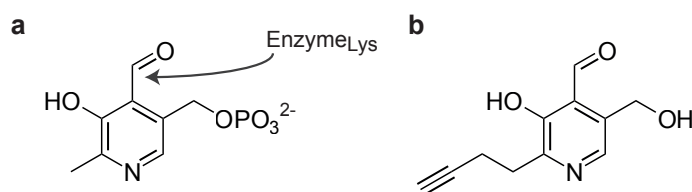


Figure 5.7: Parent compound and probe. Pyridoxal phosphate (a) and one PLP-derived probe (b), which was employed in ABPP experiments to target, enrich and identify PLP-dependent enzymes.

The aim of a ProteomeDiver-driven analysis was the prediction of probe binding sites* directly from ABPP mass spectrometry runs and subsequent MaxQuant analysis, rather than through special experiments, e.g. isoTOP-ABPP.⁴⁵ Therefore, it was assumed that the probe binds exclusively to lysines. Due to high sequence coverages, the software initially aimed to detect missing peptides, which are thought to remain on the enrichment beads (cf. figure 5.1b, class I). However, it turned out that this approach has limited applicability, either because binding site peptides are frequently measured or because of the simple fact that no peptide means no information at all. Thus, a segment intensity profile analysis of an experimental setting with different PLP probe concentrations was conducted. For example, the *S. aureus* protein AoAoH2XHH8 is known to bear a PLP binding site at lysine 32, however, in this region, peptides were measured in ABPP experiments (figure 5.8a). The same effect was observed for the protein AoAoH2XFD9 (figure 5.8).

*In fact, ProteomeDiver predicts probe binding regions and not specific amino acids. This is due to the nature of underlying peptide data: The local resolution can only be as good as the peptide resolution on a given protein sequence is. Furthermore, this section only describes the first steps of ProteomeDiver-guided segment correlation analysis and not a mature ready-to-use software.

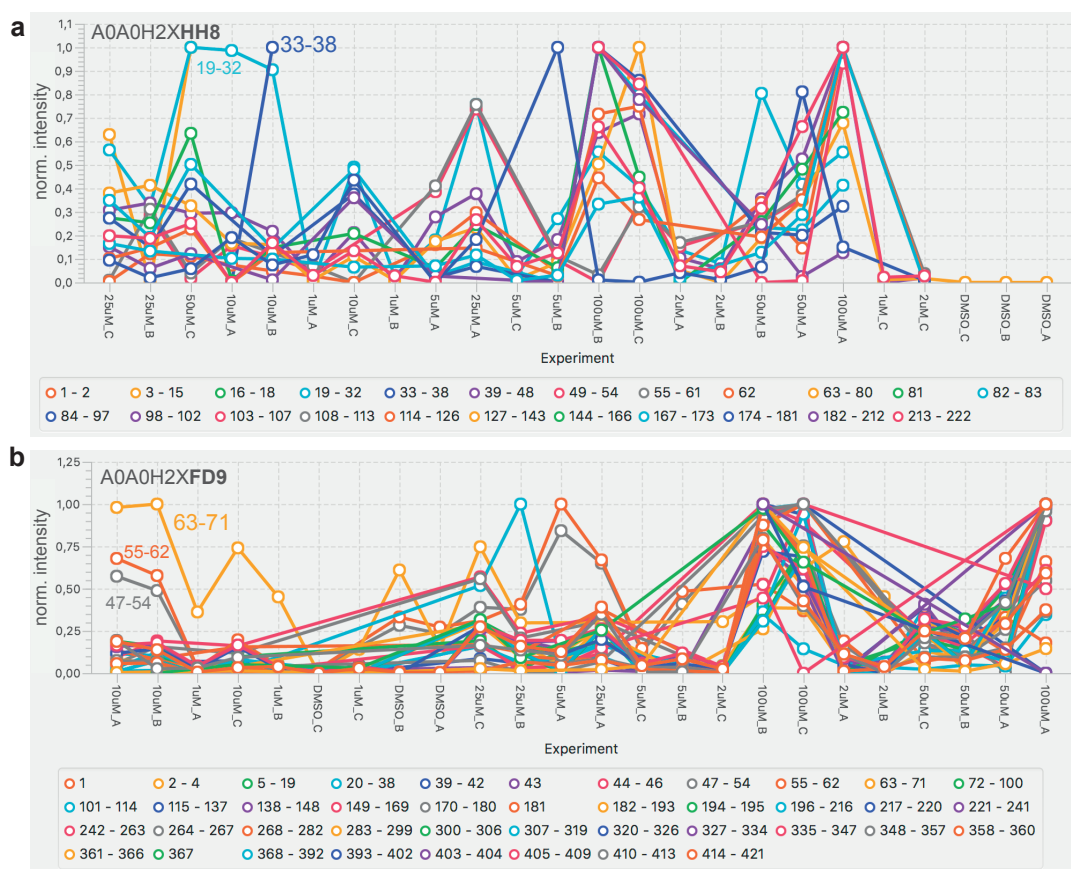


Figure 5.8: Normalized segment intensity profiles of AOA0H2XHH8 (a) and AOA0H2XFD9 (b). Uncorrelated segments are highlighted in the plot with their amino acid position in the full protein sequence. Larger numbers are further explained in the correlation analyses in figure 5.9.

In order to visualize differences in profiles, the Pearson correlation of each segment intensity profile to each other profile was determined. The resulting correlation coefficients were sorted in an increasing fashion and plotted. Thus, every profile is turned into a ranked correlation profile (figure 5.9). These profiles built the basis for further analyses such as the identification of outlier profiles. Potential PLP probe binding sites that were predicted from these outliers were partly confirmed by mass spectrometry of recombinant proteins (Annabelle Hoegl, Nina C. Bach) and UniRule annotations.¹⁶ Thus, for instance, two regions of PLP binding could be predicted for AoAoH₂XHH8 (figure 5.9a) and AoAoH₂XFD9 (figure 5.9b).

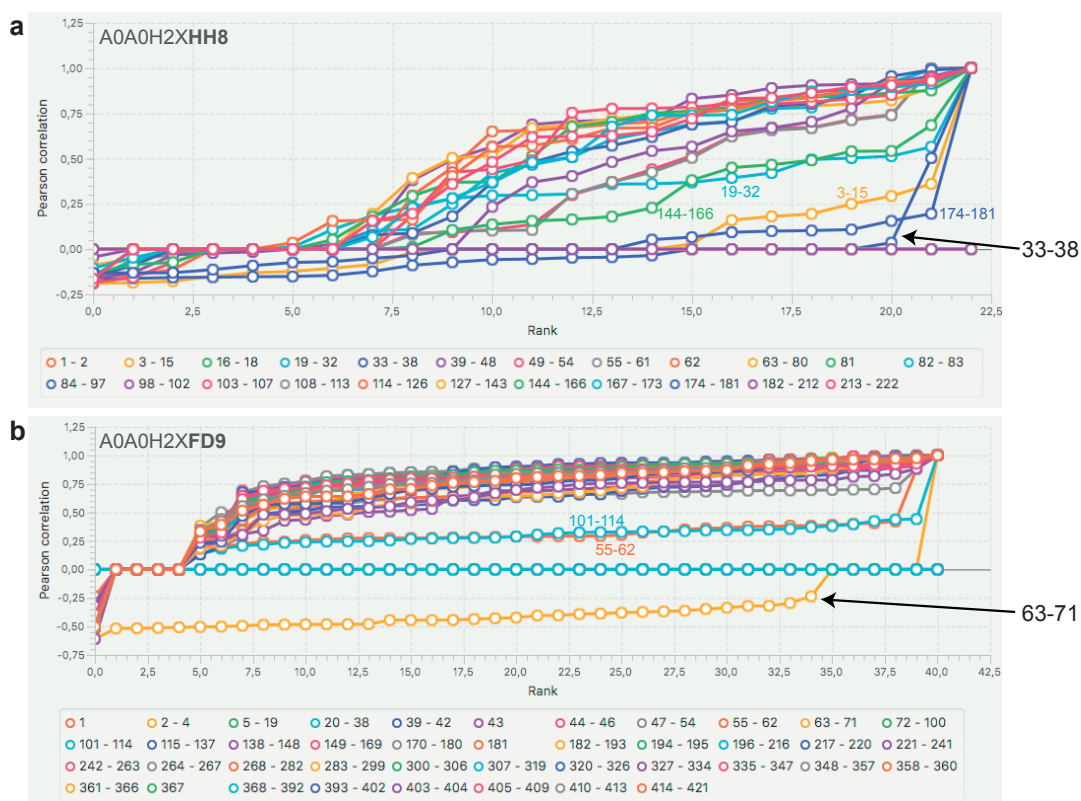


Figure 5.9: Correlation profiles of two proteins that were enriched by different concentrations of a PLP probe (for initial intensity profiles, refer to figure 5.8). Each line represents a peptide segment. The gradient of the curve is characterized by Pearson correlation-based comparisons to all other segment intensity profiles. All curves end at $y=1$ as this is the comparison of each profile to itself. (a) A0A0H2XHH8. Unpublished in-house experiments by Annabelle Hoegl and Nina C. Bach suggested lysine 32 to be the PLP binding site (data not shown). UniRule UR000497488 further supports this finding.¹⁶ Indeed, segment 33-38 is one of the outliers. It exhibits Pearson correlation coefficients of 0 or below to most of the other profiles. (b) A0A0H2XFD9. Segment 63-71 shows the most deviating and most negative correlation profile. According to UniRule UR000160031, lysine 62 should be the PLP binding site, which is supported by the profile analysis.¹⁶

The example in figure 5.9 depicts an application of the ProteomeDiver software for the prediction of probe binding sites based on the assumptions above. Outlying correlation profiles can be explained by the unforeseen cleavage behavior of trypsin, but have to be proven experimentally in future. However, there are correlation trajectories that cannot be justified in this way. For example, the correlation profile of segment 174-181 of the A0A0H2XHH8 protein exhibits a similar behavior compared to the 33-38 segment indicating the PLP probe binding site. The former segment might be partially modified with an unexpected posttranslational modification. Such peptides cannot be identified and quantified with MaxQuant. Additionally, local remaining secondary or tertiary protein folds

during trypsin digestion could lead to anomalous intensity detections. Lastly, there are enriched putative PLP-target proteins such as A0A0H2XFQ3, for which the binding site segments do not show an incoherent behavior (figure 5.10a) or no clear outlier correlation profile can be detected at all (A0A0H2XHE9, figure 5.10b). However, an estimation of how many binding sites are correctly predicted is not possible as an objective criterion to rate sufficient segment profile incoherence has to be developed in future.

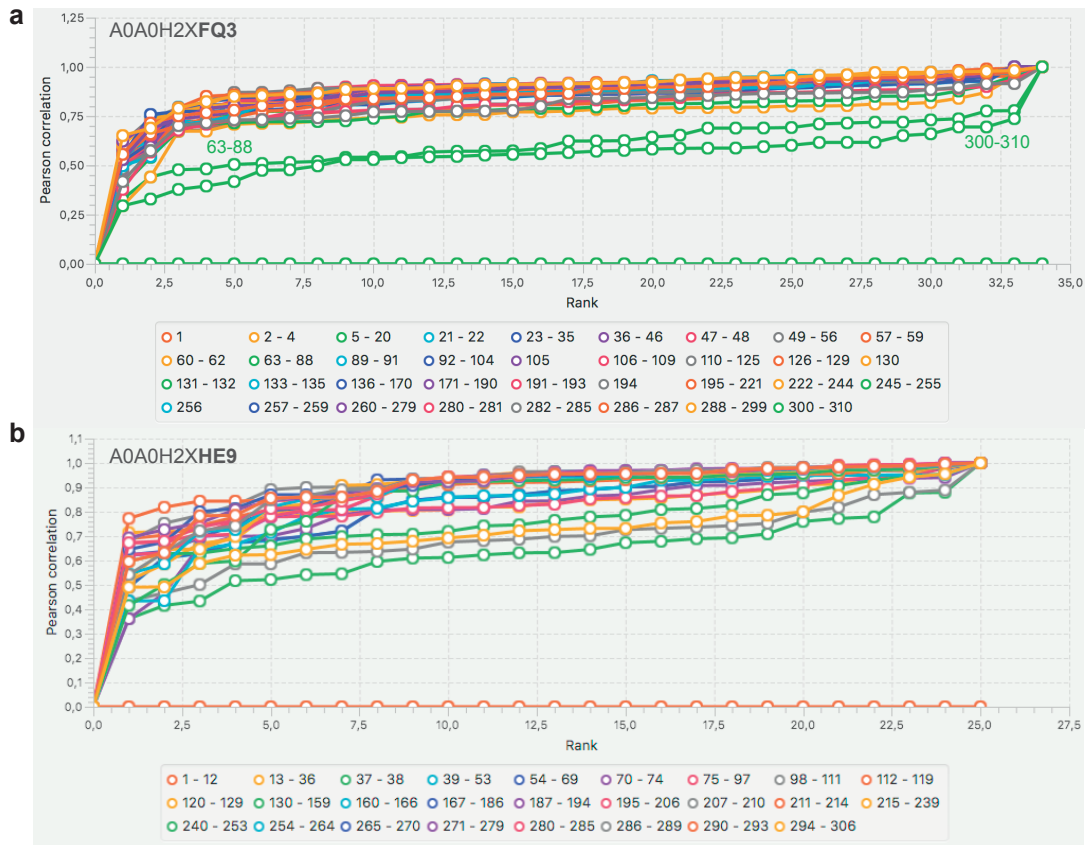


Figure 5.10: Remaining challenges of the correlation analysis method. (a) A0A0H2XFQ3. According to UniRule UR000467681, lysine 46 should be the PLP binding site.¹⁶ Here, completely different regions are highlighted by the correlation analysis. (b) A0A0H2XHE9. No correlation profile is clearly distinguishable from the bulk of profiles, despite the fact that this protein was significantly enriched in PLP pull-downs.

5.4 CONCLUSION AND OUTLOOK

ProteomeDiver allows an easy and intuitive visualization of proteomics data that has been generated by MaxQuant. The main feature comprises the sequence view panel. Identified and quantified peptides are aligned to the full protein sequence to which they

belong to. Amino acid modifications including statistical information are directly shown on the peptide representations. For each protein, a segment intensity profile analysis can be conducted. This feature allows the identification of sequence segments that behave differently compared to the majority of other peptides. This can be helpful e.g. for the prediction of probe binding sites.

The software is implemented in the Java programming language and assembled from a set of modules and classes. On the one hand, these pieces can be reused in other software projects and are freely shared via GitHub. On the other hand, ProteomeDiver's functionalities can easily be extended by other modules. One of the next steps will be the integration of a mzIdentML loader that supports the deserialization of analysis data from other search engines.¹⁷ Despite of the intriguing finding that probe-modified peptides might behave differently compared to their unmodified counterparts, this method of probe binding site prediction is lacking precision and generality.

The precision problem was observable in figure 5.9, where the correlation profile near the binding site sticks out of the other profiles, but a clear distinction of nearest neighbor segment correlation profiles that are located far away from the potential binding site is tough. At this point, further methods have to be developed. One possibility would be the introduction of a subclustering procedure. In this case, neighboring segment profiles should exhibit comparable intensity profiles because they are dependent on each other. This phenomenon is for example manifested for some correlation profile pairs of AoAoH₂XHH8 (19-32 and 33-38) and AoAoH₂XFD9 (55-62 and 63-71). Automatically combining this information should provide more evidence towards a precise prediction of probe binding sites. In addition, this approach would also be a solution to the problem of two binding sites per segment. At the moment, ProteomeDiver works segment-based and it is not possible to decide, whether a putative probe binding site is located at the N- or C-terminus of an incoherent segment.

Another issue of the current profile dissection workflow is the low degree of generality. The PLP example comprehends enriched proteins with very high sequence coverages. This is a prerequisite for a confident correlation profile analysis. Conversely, many ABPP mass spectrometry projects are lacking such high coverages. Especially for the complex protein environment of human cells, it is extremely hard to reach sufficient coverages. On

top of that, other probes than the PLP-inspired ones can have varying effects during sample preparation, which could be a starting point for future developments of ProteomeDiver. As an example, PLP probe drop-off is thought to happen via an insufficient NaBH_4 reduction of the reversible Schiff base. Thus, other probes might have different modes to promote class II or III cleavage reactions.

These efforts should be accompanied by extensive software testing. To date, only the basic functions of ProteomeDiver have been checked. For instance, a representative set of plots and related calculations should be reproduced with other software tools such as R or Perseus.

5.5 AUTHOR CONTRIBUTIONS

Annabelle Hoegl performed biochemistry and mass spectrometry. Annabelle Hoegl and Nina C. Bach performed MaxQuant searches. Matthias Stahl designed and developed ProteomeDiver and conducted all ProteomeDiver-based analyses. Stephan A. Sieber supervised the project.

5.6 REFERENCES

- [1] Norberto Peporine Lopes and Ricardo Roberto da Silva, editors. *Mass spectrometry in chemical biology*. Chemical Biology. The Royal Society of Chemistry, 2018.
- [2] Stephan M Hacker, Keriann M Backus, Michael R Lazear, Stefano Forli, Bruno E Correia, and Benjamin F Cravatt. Global profiling of lysine reactivity and ligandability in the human proteome. *Nature Chemistry*, 33, 2017.
- [3] Megan H Wright, Barbara Clough, Mark D Rackham, Kaveri Rangachari, James A Brannigan, Munira Grainger, David K Moss, Andrew R Bottrill, William P Heal, Malgorzata Broncel, et al. Validation of *N*-myristoyltransferase as an antimalarial drug target using an integrated chemical biology approach. *Nature Chemistry*, 6(2):112–121, 2014.
- [4] Enrique Audain, Julian Uszkoreit, Timo Sachsenberg, Julianus Pfeuffer, Xiao Liang, Henning Hermjakob, Aniel Sanchez, Martin Eisenacher, Knut Reinert, David L

- Tabb, et al. In-depth analysis of protein inference algorithms using multiple search engines and well-defined metrics. *Journal of Proteomics*, 150:170–182, 2017.
- [5] Jürgen Cox and Matthias Mann. MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, 26(12):1367–1372, 2008.
- [6] Stefka Tyanova, Tikira Temu, and Juergen Cox. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature Protocols*, 11(12):2301–2319, 2016.
- [7] Eliane V Wolf and Steven HL Verhelst. Activity-based protein profiling. *Chemoselective and Bioorthogonal Ligation Reactions: Concepts and Applications*, pages 305–333, 2017.
- [8] Bo Zhang, Mohammad Pirmoradian, Roman Zubarev, and Lukas Käll. Covariation of peptide abundances accurately reflects protein concentration differences. *Molecular & Cellular Proteomics*, 16(5):936–948, 2017.
- [9] Jenny Forshed, Henrik J Johansson, Maria Pernemalm, Rui MM Branca, AnnSofi Sandberg, and Janne Lehtiö. Enhanced information output from shotgun proteomics data by protein quantification and peptide quality control (PQPQ). *Molecular & Cellular Proteomics*, 10(10):M111–010264, 2011.
- [10] Jürgen Cox, Marco Y Hein, Christian A Luber, Igor Paron, Nagarjuna Nagaraj, and Matthias Mann. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Molecular & Cellular Proteomics*, 13(9):2513–2526, 2014.
- [11] Stefka Tyanova, Tikira Temu, and Jürgen Cox. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature Protocols*, 11(12):2301–2319, 2016.
- [12] Stefka Tyanova, Tikira Temu, Pavel Sinitcyn, Arthur Carlson, Marco Y Hein, Tamar Geiger, Matthias Mann, and Jürgen Cox. The Perseus computational platform for comprehensive analysis of (prote) omics data. *Nature Methods*, 2016.

- [13] Philipp Christen and David E Metzler. *Transaminases*, volume 2. John Wiley & Sons, 1985.
- [14] Eric S Simon and John Allison. Determination of pyridoxal-5'-phosphate (PLP)-bonding sites in proteins: a peptide mass fingerprinting approach based on diagnostic tandem mass spectral features of PLP-modified peptides. *Rapid Communications in Mass Spectrometry*, 23(21):3401–3408, 2009.
- [15] Eranthie Weerapana, Chu Wang, Gabriel M Simon, Florian Richter, Sagar Khare, Myles BD Dillon, Daniel A Bachovchin, Kerri Mowen, David Baker, and Benjamin F Cravatt. Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature*, 468(7325):790–795, 2010.
- [16] The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Research*, 45(D1):D158–D169, 2017.
- [17] Juan Antonio Vizcaíno, Gerhard Mayer, Simon R Perkins, Harald Barsnes, Marc Vaudel, Yasset Perez-Riverol, Tobias Ternent, Julian Uszkoreit, Martin Eisenacher, Lutz Fischer, et al. The mzIdentML data standard version 1.2, supporting advances in proteome informatics. *Molecular & Cellular Proteomics*, 2017.

6

Research conclusion

ClpP is an important protease for bacterial as well as for human cells. The first two chapters provided insights into the ClpP barrel itself. On the one hand, the mechanism of allosteric proteolytic activation was followed with the help of a small molecule, **D9**. On the other hand, the clockwork-like substrate cleavage was dissected using fluorogenic peptides and mass spectrometry. Gained results expanded our understanding of how ClpP proteases work. For example, the activator **D9** widens the scope of previously identified ClpP stimulators, such as ADEPs, ADEP fragments or ACPs.^{1,2,3,4} However, a major difference of **D9** compared to them is the species specificity for human ClpP, thus pointing to a distinct feature of allosteric proteolytic control. This will be of importance for further studies that examine hClpP in its physiological environment and its role in the unfolded protein response of mitochondria.^{5,6} Here, also the sequence characteristics of hClpP's product peptides might play an intriguing role in mitonuclear communication.^{7,8} Chapter 2 illustrates a defined, but rather unspecific cleavage behavior, which is in turn similar to tested bacterial ClpPs.

In order to obtain a global view on one of these bacterial ClpPs, staphylococcal ClpP and its cellular neighborhood was probed in chapter 3. A set of β -lactone inhibitors disturbed the protease's function *in vivo*. Subsequently, mass spectrometry-based proteomics techniques were applied to sense changes in cellular protein abundances. The study followed previous analyses of *clpP* knock-out strains and focused on a link between SaClpP inhibition and virulence shut-down.^{9,10,11,12} Strikingly, SaClpP's own chaperone, ClpX, is suggested to be crucial for SaClpP-dependent virulence regulation.¹³

The next chapter dealt with methodological advances in chemical proteomics. Chapter 4 dissected the background binding behavior of a set of diazirine photoprobes. The work was built on previous publications that uncovered the unspecificity space of photocrosslinker bearing molecules for target identification.^{14,15} As this type of experiment is broadly applied in chemical biology, the study will help researchers to better understand and interpret their results. As an example, Zhou *et al.* already made use of this dataset to classify their own proteomics results.¹⁶

Lastly, the here developed ProteomeDiver software takes already deconvoluted proteomics data as input. Whereas the classical view on proteomes is dominated by protein-centric analyses,^{17,18} the software renders the peptide level in these datasets accessible for every

researcher. Metaphorically speaking, this resembles a dive to the body of an iceberg. Peptide-centric analyses can be a treasure of data as demonstrated with the prediction of putative probe binding sites from an ABPP dataset. Thus, ProteomeDiver offers a valuable platform to make sense of thousands of megabytes of public proteomics data.¹⁹

REFERENCES

- [1] Heike Brötz-Oesterhelt, Dieter Beyer, Hein-Peter Kroll, Rainer Endermann, Christoph Ladel, Werner Schroeder, Berthold Hinzen, Siegfried Raddatz, Holger Paulsen, Kerstin Henninger, et al. Dysregulation of bacterial proteolytic machinery by a new class of antibiotics. *Nature Medicine*, 11(10):1082–1087, 2005.
- [2] Byung-Gil Lee, Eun Young Park, Kyung-Eun Lee, Hyesung Jeon, Kwang Hoon Sung, Holger Paulsen, Helga Rübsamen-Schaeff, Heike Brötz-Oesterhelt, and Hyun Kyu Song. Structures of ClpP in complex with acyldepsipeptide antibiotics reveal its activation mechanism. *Nature Structural & Molecular Biology*, 17(4):471–478, 2010.
- [3] Daniel W Carney, Corey L Compton, Karl R Schmitz, Julia P Stevens, Robert T Sauer, and Jason K Sello. A simple fragment of cyclic acyldepsipeptides is necessary and sufficient for ClpP activation and antibacterial activity. *ChemBioChem*, 15(15):2216–2220, 2014.
- [4] Elisa Leung, Alessandro Datti, Michele Cossette, Jordan Goodreid, Shannon E McCaw, Michelle Mah, Alina Nakhamchik, Koji Ogata, Majida El Bakkouri, Yi-Qiang Cheng, et al. Activators of cylindrical proteases as antimicrobials: identification and development of small molecule activators of ClpP protease. *Chemistry & Biology*, 18(9):1167–1178, 2011.
- [5] Tomer Shpilka and Cole M Haynes. The mitochondrial UPR: mechanisms, physiological functions and implications in ageing. *Nature Reviews Molecular Cell Biology*, 2017.
- [6] Cole M Haynes and David Ron. The mitochondrial UPR—protecting organelle protein homeostasis. *Journal of Cell Science*, 123(22):3849–3855, 2010.
- [7] Cole M Haynes, Kseniya Petrova, Cristina Benedetti, Yun Yang, and David Ron. ClpP

- mediates activation of a mitochondrial unfolded protein response in *C. elegans*. *Developmental Cell*, 13(4):467–480, 2007.
- [8] Cole M Haynes, Yun Yang, Steven P Blais, Thomas A Neubert, and David Ron. The matrix peptide exporter HAF-1 signals a mitochondrial UPR by activating the transcription factor ZC376.7 in *C. elegans*. *Molecular Cell*, 37(4):529–540, 2010.
- [9] Antje Michel, Franziska Agerer, Christof R Hauck, Mathias Herrmann, Joachim Ullrich, Jörg Hacker, and Knut Ohlsen. Global regulatory impact of ClpP protease of *Staphylococcus aureus* on regulons involved in virulence, oxidative stress response, autolysis, and DNA repair. *Journal of Bacteriology*, 188(16):5783–5796, 2006.
- [10] Dorte Frees, Karen Sørensen, and Hanne Ingmer. Global virulence regulation in *Staphylococcus aureus*: pinpointing the roles of ClpP and ClpX in the *sar/agr* regulatory network. *Infection and Immunity*, 73(12):8100–8108, 2005.
- [11] Dorte Frees, Julie Hove Andersen, Lene Hemmingsen, Kerttu Koskenniemi, Kristoffer T Bæk, Musemma Kedir Muhammed, Dereje Dadi Gudeta, Tuula A Nyman, Antti Sukura, Pekka Varmanen, et al. New insights into *Staphylococcus aureus* stress tolerance and virulence regulation from an analysis of the role of the ClpP protease in the strains Newman, COL, and SA564. *Journal of Proteome Research*, 11(1):95–108, 2011.
- [12] Dorte Frees, Saara NA Qazi, Philip J Hill, and Hanne Ingmer. Alternative roles of ClpX and ClpP in *Staphylococcus aureus* stress tolerance and virulence. *Molecular Microbiology*, 48(6):1565–1578, 2003.
- [13] Lotte Jelsbak, Hanne Ingmer, Lukás Valíhrach, Marianne Thorup Cohn, Mie HG Christiansen, Birgitte H Kallipolitis, and Dorte Frees. The chaperone ClpX stimulates expression of *Staphylococcus aureus* protein A by Rot dependent and independent pathways. *PLoS One*, 5(9):e12752, 2010.
- [14] Jongmin Park, Minseob Koh, Ja Young Koo, Sanghee Lee, and Seung Bum Park. Investigation of specific binding proteins to photoaffinity linkers for efficient deconvolution of target protein. *ACS Chemical Biology*, 11(1):44–52, 2015.

- [15] Ja Young Koo, VV Yellamelli, Seung Bum Park, et al. Nonspecific protein labeling of photoaffinity linkers correlates with their molecular shapes in living cells. *Chemical Communications*, 52(34):5828–5831, 2016.
- [16] Yiqing Zhou, Zhengao Di, Xiaoming Li, Yuanhong Shan, Weichao Li, Haibing Zhang, and Youli Xiao. Chemical proteomics reveal CD147 as a functional target of pseudolaric acid B in human cancer cells. *Chemical Communications*, 53(62):8671–8674, 2017.
- [17] Ruedi Aebersold and Matthias Mann. Mass spectrometry-based proteomics. *Nature*, 422(6928):198–207, 2003.
- [18] Ruedi Aebersold and Matthias Mann. Mass-spectrometric exploration of proteome structure and function. *Nature*, 537(7620):347–355, 2016.
- [19] Juan A Vizcaíno, Eric W Deutsch, Rui Wang, Attila Csordas, Florian Reisinger, Daniel Rios, Jose A Duanes, Zhi Sun, Terry Farrah, Nuno Bandeira, et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nature Biotechnology*, 32(3):223–226, 2014.

Part III

Review

7

An amino acid domino effect orchestrates ClpP's conformational states

Published in *Current Opinion of Chemical Biology*, 2017, 40, pp 102-110
by Matthias Stahl and Stephan A. Sieber.

Publishing permission not required as the author retains the right to include the
article in a dissertation. © 2017 Elsevier. DOI: [10.1016/j.cbpa.2017.08.007](https://doi.org/10.1016/j.cbpa.2017.08.007)

SYNOPSIS

Maintaining the cellular protein homeostasis means managing life on the brink of death. This balance is largely based on precise fine-tuning of enzyme activities. For instance, the ClpP protease possesses several conformational switches which are fundamental to regulating its activity. Efforts have focused on revealing the structural basis of ClpP's conformational control. In the last decade, several amino acid clusters have been identified and functionally linked to specific activation states. Researchers have now begun to couple these hotspots to one another, uncovering a global network of residues that switch in response to internal and external stimuli. For these studies, they used small molecules to mimic intermolecular interactions and point-mutational studies to shortcut regulating amino acid circuits.

The synopsis is taken from the original publication.

AUTHOR CONTRIBUTIONS

Matthias Stahl wrote the manuscript and created figures with input from Stephan A. Sieber. Stephan A. Sieber supervised the process.



ELSEVIER

An amino acid domino effect orchestrates ClpP's conformational states

Matthias Stahl and Stephan A Sieber



Maintaining the cellular protein homeostasis means managing life on the brink of death. This balance is largely based on precise fine-tuning of enzyme activities. For instance, the ClpP protease possesses several conformational switches which are fundamental to regulating its activity. Efforts have focused on revealing the structural basis of ClpP's conformational control. In the last decade, several amino acid clusters have been identified and functionally linked to specific activation states. Researchers have now begun to couple these hotspots to one another, uncovering a global network of residues that switch in response to internal and external stimuli. For these studies, they used small molecules to mimic intermolecular interactions and point-mutational studies to shortcut regulating amino acid circuits.

Address

Center for Integrated Protein Science Munich (CIPSM), Chair II of Organic Chemistry, Department of Chemistry, Technische Universität München, Lichtenbergstraße 4, 85747 Garching b. München, Germany

Corresponding author: Sieber, Stephan A (stephan.sieber@tum.de)

Current Opinion in Chemical Biology 2017, 40:102–110

This review comes from a themed issue on **Synthetic biomolecules**

Edited by **Peter H Seeberger** and **Beate Kocsch**

For a complete overview see the [Issue](#) and the [Editorial](#)

Available online 12th September 2017

<http://dx.doi.org/10.1016/j.cbpa.2017.08.007>

1367-5931/© 2017 Elsevier Ltd. All rights reserved.

Introduction

The human proteome comprises around 20 000 different proteins with more than 80 000 isoforms [1–3]. Both the types of proteins as well as their abundancies can give researchers a clue as to how these complex networks operate [4]. However, these data form only the tip of a huge systems biology iceberg, because the functions of proteins are significantly tuned by their environments. Specifically, proteins may (i) interact with other proteins, (ii) be modified posttranslationally, or (iii) bind small molecule ligands. Most often, a combination of these cases contribute to the overall control of different conformational states and their equilibria [5,6].

We here compile insights into the oligomeric and conformational regulation of the widespread protease

caseinolytic peptidase P (ClpP). ClpP's monomers assemble into multimeric complexes whose activity is regulated through intermolecular oligomeric rearrangements as well as subtle intramolecular conformational switches (Figure 1a).

The principles of ClpP regulation have initially been dissected using mutagenesis, small molecules or even interacting proteins. Several sites of regulation on the amino acid level were thereby revealed (Figure 1b). In the last two years, a deeper understanding of how these hotspots are interconnected has been gained. Essentially, ClpP appears to be regulated by amino acids whose conformations are guided by environmental stimuli such as protein-protein interactions. This ensures a tight control over conformational states.

The ClpXP complex: implementation of a multilayered conformational control strategy

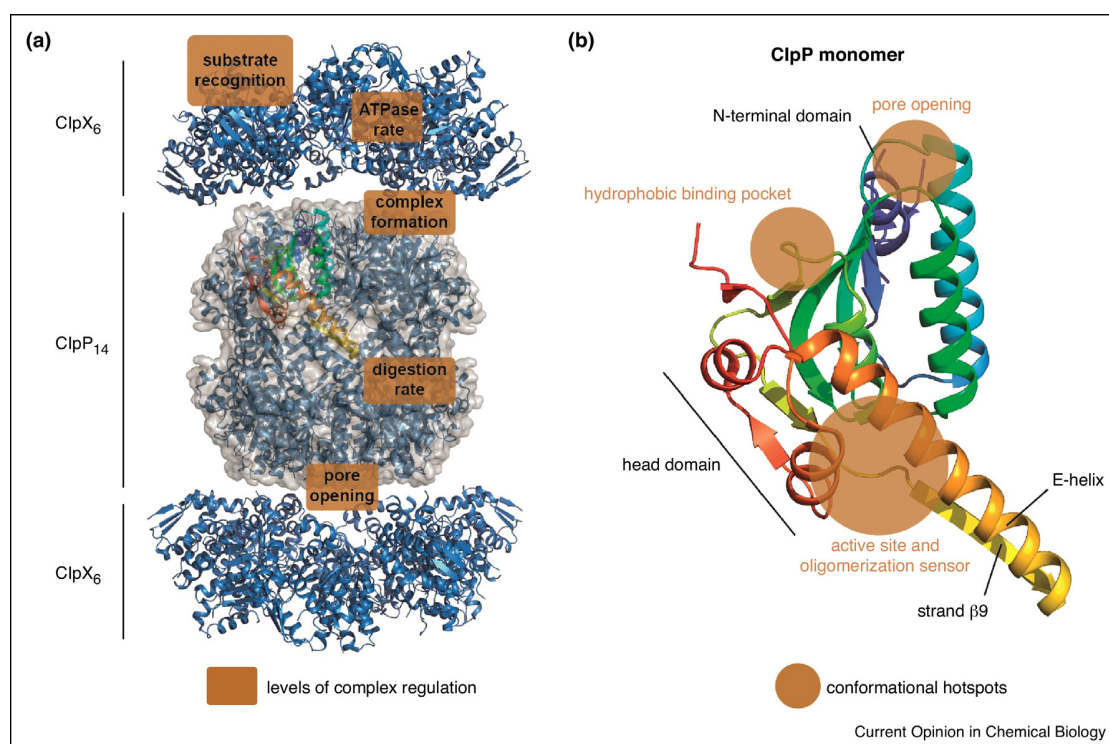
ClpP is a serine protease found in many prokaryotes and eukaryotes [7]. The functional complex consists of 14 monomers that assemble into two heptameric rings. These rings are able to stack face-to-face, thereby building up a barrel-shaped, tetradecameric structure with two axial pores. The protease therefore usually contains 14 sequestered serine-histidine-aspartate catalytic triads [8–10]. Each of these is capable of digesting substrate proteins, which are recognized, denatured and delivered into the proteolytic chamber by ClpX or a similar chaperone sitting on top of the entry pores of ClpP [11–13].

ClpX and ClpP activities need to be precisely coordinated since there is no pronounced cleavage specificity later in proteolysis [14]. For this purpose, among others, ClpX contains IGF loops,¹ which bind to specific patches on the surface of ClpP, thus enabling communication. Furthermore, crosstalk also works the other way around as the mechanical activity of ClpX is stimulated by ClpP binding [15,16^{**},17].

The proteolytic potential of ClpXP is regulated on five different levels (Figure 1a): (i) the quaternary structure: at least one ClpX hexamer and two ClpP heptamers must form a complex [18], (ii) the specificity of ClpX: substrates have to be identified by their recognition patterns [19–21], (iii) the mechanical motion of ClpX: the ATPase

¹ The consensus sequence might be different for ClpPs from other species. Only the one of *S. aureus* is mentioned here for clarity reasons.

Figure 1



The ClpXP protein degradation machine. **(a)** Artificial compilation of the full ClpXP complex consisting of two ClpX hexamers from *E. coli* (PDB: 4I81) and the active, extended variant of ClpP from *S. aureus* (PDB: 3V5E) [10,60]. The five levels of proteolytic regulation and their approximate localization are highlighted by orange rectangles. **(b)** Magnification of the rainbow-colored monomer from panel a. The sequence is visualized from the N- to C-terminus in blue to red color. Structural motifs, which serve as local hotspots of conformational control, are depicted. Among them are the N-terminal loops, which establish the substrate entry pore, the hydrophobic binding pocket – acting as a docking site for ClpX –, and the active site, which is connected to the oligomerization sensor.

rate determines the speed of substrate processing [22,23], (iv) the pore of ClpP: the protease's entry control in response to the chaperone [24*,25,26], and (v) the activity of ClpP's catalytic triad: full alignment of all involved residues is required for full catalytic activity [27]. These levels rely on precise control of protein structure and are often interconnected, *e.g.* as active site modification of ClpP influences the strength of ClpX binding [28].

We will first focus on the latter two points involving intrinsic conformational control of ClpP. Then, we will examine the influence of ClpX on ClpP's conformational integrity in order to gain an overall picture of the regulatory amino acids at play.

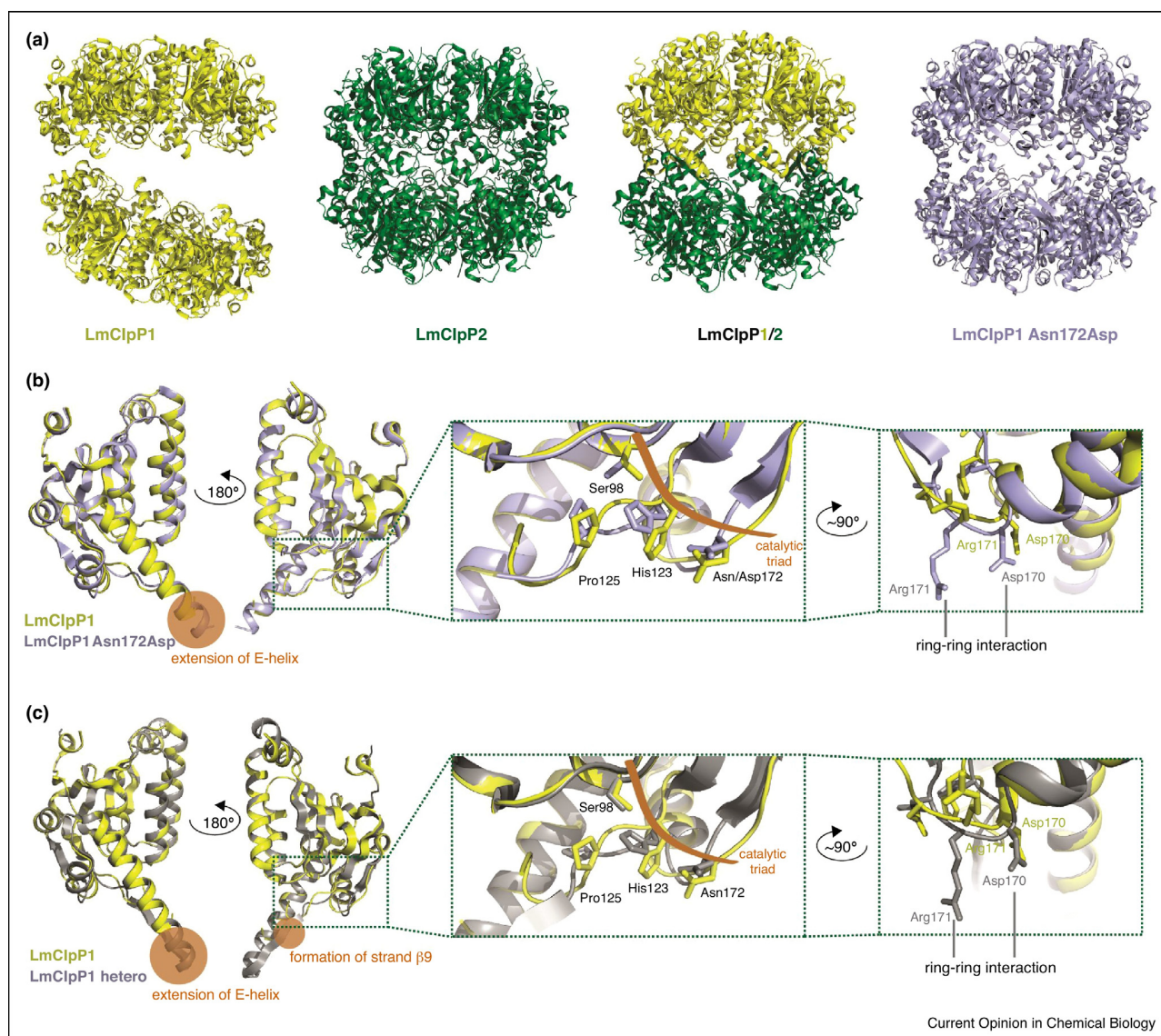
ClpP's microstructure: different faces of a tetradecamer

Our current knowledge assumes ClpP to be a tetradecamer in its active state. Indeed, a connection between the active site and the ring-ring interface has been revealed among several ClpPs including that of *Listeria monocytogenes* (LmClpP) [27]. These pathogenic bacteria encode two ClpP isoforms: LmClpP2, which is able to form an

active tetradecamer, and LmClpP1, which forms an inactive heptamer containing asparagines instead of aspartates at the third position of the catalytic triad [29,30**]. Interestingly, artificial mutation of this Asn172 back to an aspartate readily leads to tetradecamerization (Figure 2a). Comparing the LmClpP1 mutant to its wild-type, it was shown that surrounding residues Asp170 and Arg171 are significantly involved in ring-ring interactions by forming inter-ring salt bridges. In addition, Asn172 is interconnected with the catalytic triad member His123 and Pro125. In the mutant crystal structure, these residues are tilted, which may subsequently facilitate extension of the adjacent E-helix, an important feature for inter-heptamer oligomerization (Figure 2b, see Figure 1b for an overview of all secondary structure elements). Hence, formation of the tetradecamer is coupled to the arrangement of the active site [27], where alignment of the catalytic triad generates a tetradecamer-friendly ring-ring interface and *vice versa* (Figure 2b).

Studies aimed at understanding the biological function of the inactive, wild-type LmClpP1 heptamer have shown that under certain circumstances, LmClpP1 forms a

Figure 2



Prominent conformational switches in *Listeria monocytogenes* ClpP (LmClpP). **(a)** Cartoon representations of crystal structures from LmClpP1 (PDB 4JCQ), LmClpP2 homocomplex (PDB 4JCT), LmClpP1/2 heterocomplex (PDB 4RYF), and the artificially active mutant LmClpP1 Asn172Asp (PDB 4JCR) [27,30**]. **(b)** Overlay of monomers from LmClpP1 (yellow) and LmClpP1 Asn172Asp (purple). Mutation of the active site asparagine to aspartate causes turning of His123, which is also a member of the catalytic triad. This induces twisting of Pro125 that is connected to the N-terminus of the E-helix and leads to its extension (orange circle). In addition, the Asn172Asp mutation lets the oligomerization sensor (Asp170 and Arg171) swing out. Thus, LmClpP heptamer rings can stack together forming tetradecamers [27]. **(c)** Overlay of a LmClpP1 monomer (yellow) with a subunit of LmClpP1 from the LmClpP1/2 heterocomplex (gray) [30**]. Conformational changes in LmClpP1 are similar as shown in panel b and are triggered by the LmClpP2 template.

heterocomplex with LmClpP2 [29]. The LmClpP2 ring appears to serve as a template to select for an active conformation of LmClpP1. Together with ClpX, the active sites of both LmClpP1 and LmClpP2 contribute to a proteolytic rate that is significantly elevated in comparison to the LmClpP2 homocomplex. Intriguingly, the crystal structure of LmClpP1 in the LmClpP1/2 heterocomplex shows changes in amino acid orientations

within the active site and ring-ring interface that are similar to the artificially activated LmClpP1 Asn172Asp mutant (Figure 2c) [30**]. In addition, substrate occupancy of the active site seems to be crucial for the activation state. A recent study investigated the effect of chloromethyl ketone inhibitors on LmClpP activity. The inhibitors boosted proteolytic activity up to 2.5x when only one fifth of the active sites were blocked [31].

Besides LmClpP, the ClpP of *Staphylococcus aureus* (SaClpP) has been intensely researched over the past few years. In fact, similar oligomeric regulation principles as for LmClpP were found and X-ray analysis revealed three different conformational states: extended, compact and compressed (Figure 3a) [32–34]. Only the former exhibits catalytic activity, whereas the others are believed to be part of a barrel-opening cycle [35–38]. However, another explanation would be to facilitate peptide exit by the axial pores. Tetradecamer integrity is again based on a hydrogen bridge network consisting of Asp170/Arg171 (oligomerization sensor) and Gln132. The E-helix, which is involved in interheptamer contacts, is shortened in the compact state and even kinked in the compressed state, thereby pulling at Asp172 of the catalytic triad. Furthermore, formation of active ClpP requires the assembly of interacting antiparallel β -sheets near the E-helices (*i.e.* strands β 9) that are exclusively observed in the extended conformation (Figure 3b) [10].

To probe conformational coherence and to shed light on the connection of different conformational hotspots, synthetic compounds can be applied to perturb interdependencies between different residues. Therefore, phenyl esters were identified and used as a new class of active site

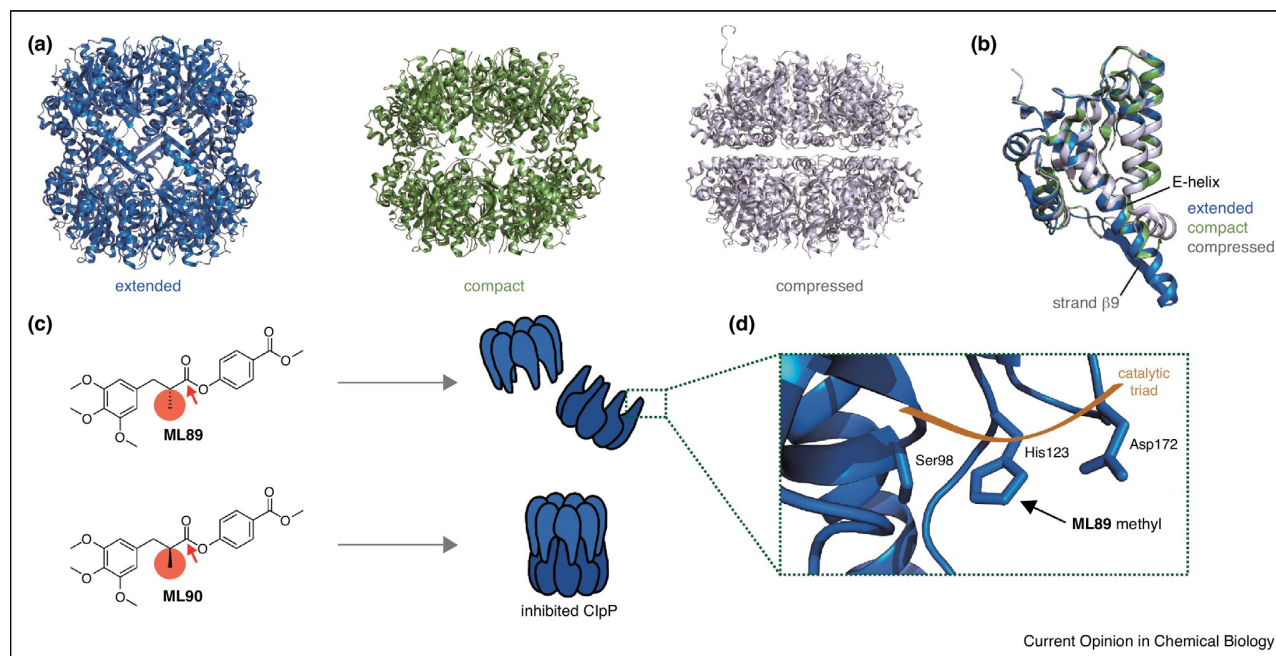
inhibitors against SaClpP [39*]. Among several derivatives, the two stereoisomers **ML89** and **ML90** attracted special attention. Depending on the orientation of a methyl substituent, the SaClpP tetradecamer was disassembled into heptamers (Figure 3c). Molecular dynamics simulations proposed the catalytic triad residue His123 to be a conformational transmitter for heptamerization (Figure 3d) [39*].

Of course, the equilibrium of different conformational states can also be affected by the environment in a cellular context. For example, the interaction of ClpP with one of its chaperones, such as ClpX, has a major influence on the protease's activation state.

ClpX guides ClpP: an analogous communication code

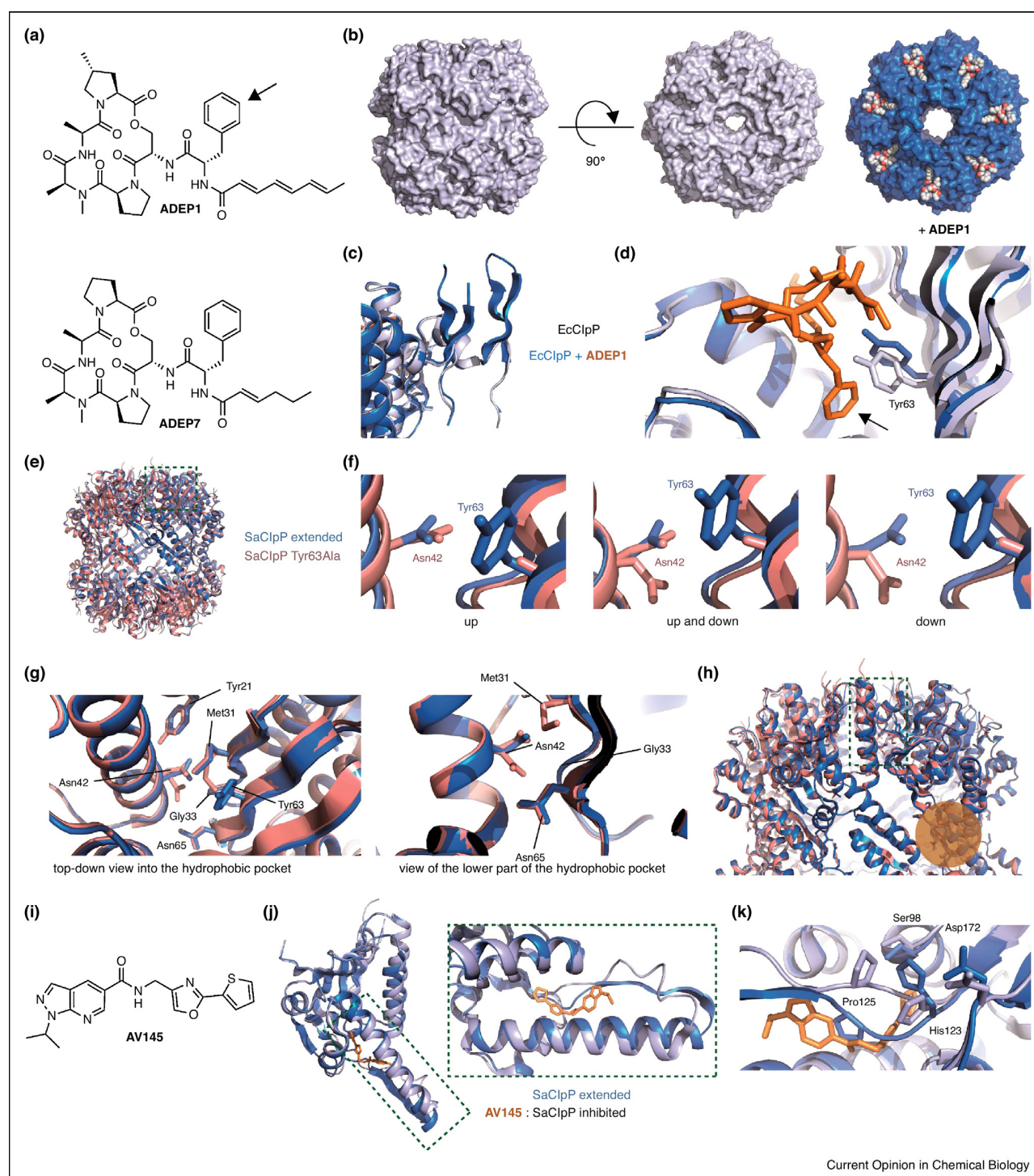
The integrity of an active ClpP complex depends on surrounding biomolecules, such as the ClpX chaperone. Therefore, ClpP contains hydrophobic clefts at the boundaries of two adjacent monomers on the apical surface. ClpX connects to these pockets using its IGF consensus loops. Additionally, the chaperone has pore-2 loops, which bind to the N-terminal loops around the pore

Figure 3



Disruption of oligomerization in *Staphylococcus aureus* ClpP (SaClpP). **(a)** X-ray crystal structures revealed three different global conformations of SaClpP: extended (PDB 3V5E), compact (PDB 4EMM), and compressed (PDB 3QWD) [10,33,34]. **(b)** Magnification and overlay of one monomer of each conformational state from panel a. Different sizes and orientations of the E-helix and the formation of strand β 9 are depicted. **(c)** Two stereoisomers of a phenyl ester: **ML89** (R)-enantiomer, and **ML90** (S)-enantiomer [39*]. A red arrow marks the electrophilic site of serine attack. Depending on the orientation of the methyl group highlighted in orange, the SaClpP tetradecamer is disassembled into heptamers or maintained through inhibition. **(d)** Illustration of the catalytic triad residues in one extended SaClpP monomer. Based on computational docking studies, the **ML89** methyl substituent is believed to cause turning of His123 and thereby provokes disruption of oligomerization.

Figure 4



Current Opinion in Chemical Biology

ClpX guides conformational switches in ClpP. **(a)** Chemical structures of **ADEP1** and **ADEP7**, which have both been used as mimics of ClpX IGF loops in conformational studies [16**,24*]. **(b)** Apo-ClpP from *Escherichia coli* (EcClpP) (blue white, PDB 1TYF) and EcClpP bound to **ADEP1** (marine blue, PDB 3MT6) [25,43]. The top view enables a comparison of the apical entry pores, which upon **ADEP1** binding, are widened by approximately 10 Å. **(c)** Zoom into the overlay of N-terminal domains of apo- and holo-EcClpP from panel b. **ADEP1** triggers a defined conformation of the N-terminal loops, which point in the direction of ClpX. **(d)** Focus on the hydrophobic binding pocket of the same overlay as in c. The side chain of Tyr63 is turned by approximately 90°. The arrow points towards the benzyl group of **ADEP1** (refer to panel a). **(e)** Overlay of extended SaClpP (PDB 3V5E) and SaClpP Tyr63Ala (PDB 5C90) [10,47**]. The region surrounded by a green rectangle is magnified in f.

of ClpP and are related to ATP turnover of ClpX [18,28,40].

The structural and conformational dissection of the ClpXP complex has proven to be challenging, presumably owing to its high flexibility, conformational heterogeneity and low *in vitro* stability. This was circumvented by the discovery of a set of acyldepsipeptides (ADEPs), which mimic the IGF loops of ClpX making ClpP activity independent of its chaperone (Figure 4ab) [41,42]. Binding of ADEPs re-organizes the N-terminal loops around the entry pore to build a defined β -hairpin crown (Figure 4c). This is accompanied by significant pore widening of around 10 Å for ClpP from *Escherichia coli* (EcClpP) (Figure 4b) [24^{*},25,43]. It is assumed that ClpX invokes similar conformational changes and that this helps substrates to enter the proteolytic chamber of ClpP after being unfolded [44].

In fact, the proteolytic rate of ClpP depends on substrate access to the active sites. But in addition, occupying the IGF pockets has a much more global effect on the overall conformation of ClpP and its substrate processing speed. In order to dissect the details of the proteolytic reaction, cleavage of the protein backbone by the active serine can be described as a two-step process. First, the carbonyl carbon is attacked by the serine, which shows enhanced nucleophilicity as a member of the catalytic triad. Second, the acyl-enzyme intermediate is hydrolyzed and product peptides are released from the active site. To probe hydrolytic activity independently of pore opening, nucleophilic attack and hydrolysis of small β -lactone inhibitors [45,46] were individually monitored in the presence or absence of ADEP. Interestingly, both catalytic turnover and hydrolysis rates were stimulated [16^{**}]. This finding suggests a direct effect of ClpX binding on the proteolytic potential of ClpP by favoring the active extended conformation of SaClpP.

The molecular connection between the hydrophobic pockets and active sites that underlies this phenomenon was still enigmatic until Ni and coworkers successfully created a gain-of-function mutant of SaClpP [47^{**}]. By removing the tyrosine, which is rotated by 90° in ADEP-bound EcClpP (Figure 4b,d), they enabled chaperone-independent proteolytic activity. Protein X-ray analysis revealed that an amino acid domino effect transmits

binding information from the apical surface into ClpP's core resulting in an active ClpP (Figure 4e–h). Taken together, ClpX binding might induce a change in the conformation of Tyr63, which seems to initiate a series of downstream events beginning at adjacent residue Met31. This causes Asn42 to flip downward, thereby switching its hydrogen bonding interactions from Tyr21 to Gly33 and Asn65 [47^{**}]. Of course, X-ray based insights into this chain reaction are only snapshots and there may be more amino acids involved. Especially the final link to the active site remains to be uncovered.

The structural control of ClpP, naturally exerted by ClpX, is fairly dominant: For example, a reversible oxazole inhibitor (Figure 4i) can paralyze proteolytic activity by binding into the space between the E-helix and strand β 9. The E-helix is shortened and, presumably *via* a complete turnaround of Pro125, the catalytic triad becomes misaligned (Figure 4jk). However, binding of ClpX or ADEP fragments [48] reverses this inactive conformational arrest and overrides inhibition [49^{*}].

In addition to this hierarchy of different allosteric effects, the ClpX-ClpP interaction is not just a binary control mechanism that switches protease activity on or off. Evidence has shown that a low ADEP:ClpP ratio mediates partial inhibition of the protease suggesting that nature implemented an analogous, multilayered code to ensure a proper communication between ClpX IGF loops and ClpP [16^{**}].

Conclusion and perspectives

Recent research has shown that the five levels of ClpXP regulation do not act independently, but are rather interconnected. In a cellular context, the ClpXP proteolytic system is an important component of the protein homeostasis machinery that is ubiquitous from bacteria to human cells. It is therefore essential to accurately control its activity in space and time to prevent accidental degradation of proteins.

Regulation of ClpP operates on two key principles including self-control through alignment of the catalytic triad only upon correct structural assembly, as well as allosteric control induced by ClpX binding, which initiates bidirectional crosstalk. ClpX thereby widens the entrance pores of ClpP and simultaneously boosts catalytic efficiency.

(Figure 4 Legend Continued) (f) Close-up view of the hydrophobic binding pocket on the apical surface of SaClpP. Wild-type and mutant structures are overlaid as in e. Directly under the Tyr63 (*cf.* panel c), an asparagine (Asn65) adopts different conformations within the same tetradecamer, either up, down or both. (g) Top-down view into the hydrophobic cleft (left) and view of the bottom part (right, *cf.* overlay from panel e). Labeled amino acids act as a chain of dominoes that are believed to react to different Tyr63 conformations, which can in turn be triggered by an activator molecule or ClpX binding. (h) The pocket in the context of a whole heptamer. The green rectangle marks the hydrophobic binding pocket. The orange circle denotes the oligomerization sensor consisting of Asp170 and Arg171. (i) Chemical structure of the reversible SaClpP inhibitor AV145 [49^{*}]. (j) Alignment of extended SaClpP with a co-crystal structure including AV145 (left, PDB 5DL1). The inhibitor binds between strand β 9 and the E-helix (right). (k) View of the catalytic triad in the overlaid structures from panel j. Pro125 undergoes a 180° rotation and the catalytic triad is misaligned as a result of a switch of His123 [49^{*}].

The important role of ClpP in a cellular system renders it an interesting target for new pharmaceuticals against bacterial infections, parasites or cancer [50–53]. As for human ClpP (hClpP), naturally occurring mutations in the hydrophobic binding pockets cause the Perrault syndrome [54–56]. To date, only few reports have addressed the allosteric events in hClpP [57,58]. A detailed knowledge of how characteristic conformational states are structurally mediated on the amino-acid level would therefore facilitate the development of tailored drug candidates.

We are convinced that subtle manipulations using small molecules in combination with detection systems such as NMR, X-ray crystallography and mass spectrometry will be a powerful strategy for dissecting conformational regulation of proteins. Furthermore, we believe that especially mass spectrometry will be valuable for future research on ClpP's conformational control, because it can be easily applied in complex cellular environments. To this end, pioneering work by Sowole and colleagues has highlighted the use of hydrogen-deuterium exchange (HDX) mass spectrometry for gaining a global view on the ClpP tetradecamer [59**].

In summary, ClpPs conformational states are well defined through *in vitro* experiments. The next step will be their elucidation in living cells. This subject is still in its infancy and only limited data is available. For example, mRNA of both LmClpP1 and LmClpP2 is overexpressed in cells undergoing heat stress, suggesting that the activated heterocomplex is needed to remove elevated levels of unfolded proteins [30**]. However, the implication of ClpP's different isoforms or conformational states *in vivo* remains a challenge for future work.

Funding

We are grateful for funding from Deutsche Forschungsgemeinschaft, SFB1035 (“Control of Protein Function by Conformational Switching”).

Acknowledgement

M.S. thanks the German National Merit Foundation for their kind support. We thank Malte Gersch, Mathias Hackl, Markus Lakemeyer, Annabelle Hoegl, and Heidi Jung for valuable comments and critical proofreading of the manuscript.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R *et al.*: **An integrated encyclopedia of DNA elements in the human genome.** *Nature* 2012, **489**:57-74.
 2. Wilhelm M, Schlegl J, Hahne H, Gholami AM, Lieberenz M, Savitski MM, Ziegler E, Butzmann L, Gessulat S, Marx H *et al.*: **Mass-spectrometry-based draft of the human proteome.** *Nature* 2014, **509**:582-587.
 3. Kim M-S, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, Kelkar DS, Isserlin R, Jain S *et al.*: **A draft map of the human proteome.** *Nature* 2014, **509**:575-581.
 4. Aebersold R, Mann M: **Mass-spectrometric exploration of proteome structure and function.** *Nature* 2016, **537**:347-355.
 5. Goh C-S, Milburn D, Gerstein M: **Conformational changes associated with protein-protein interactions.** *Curr Opin Struct Biol* 2004, **14**:104-109.
 6. Laskowski RA, Gerick F, Thornton JM: **The structural basis of allosteric regulation in proteins.** *FEBS Lett* 2009, **583**:1692-1698.
 7. Yu AYH, Houry WA: **ClpP: A distinctive family of cylindrical energy-dependent serine proteases.** *FEBS Lett* 2007, **581**:3749-3757.
 8. Kang SG, Maurizi MR, Thompson M, Mueser T, Ahvazi B: **Crystallography and mutagenesis point to an essential role for the N-terminus of human mitochondrial ClpP.** *J Struct Biol* 2004, **148**:338-352.
 9. Bewley MC, Graziano V, Griffin K, Flanagan JM: **The asymmetry in the mature amino-terminus of ClpP facilitates a local symmetry match in ClpAP and ClpXP complexes.** *J Struct Biol* 2006, **153**:113-128.
 10. Gersch M, List A, Groll M, Sieber SA: **Insights into structural network responsible for oligomerization and activity of bacterial virulence regulator caseinolytic protease P (ClpP) protein.** *J Biol Chem* 2012, **287**:9484-9494.
 11. Reid BG, Fenton WA, Horwich AL, Weber-Ban EU: **ClpA mediates directional translocation of substrate proteins into the ClpP protease.** *Proc Natl Acad Sci USA* 2001, **98**:3768-3772.
 12. Ortega J, Lee HS, Maurizi MR, Steven AC: **ClpA and ClpX ATPases bind simultaneously to opposite ends of ClpP peptidase to form active hybrid complexes.** *J Struct Biol* 2004, **146**:217-226.
 13. Baker TA, Sauer RT: **ClpXP, an ATP-powered unfolding and protein-degradation machine.** *Biochim Biophys Acta (BBA) - Mol Cell Res* 2012, **1823**:15-28.
 14. Gersch M, Stahl M, Poreba M, Dahmen M, Dzedzic A, Drag M, Sieber SA: **Barrel-shaped ClpP proteases display attenuated cleavage specificities.** *ACS Chem Biol* 2016, **11**:389-399.
 15. Martin A, Baker TA, Sauer RT: **Distinct static and dynamic interactions control ATPase-peptidase communication in a AAA+ protease.** *Mol Cell* 2007, **27**:41-52.
 16. Gersch M, Famulla K, Dahmen M, Göbl C, Malik I, Richter K, Korotkov VS, Sass P, Rübsamen-Schaeff H, Madl T *et al.*: **AAA+ chaperones and acyldepsipeptides activate the ClpP protease via conformational control.** *Nat Commun* 2015, **6**.
- This study analyzes the effects of ADEPs on SaClpP with a broad set of different biochemical methodologies including activity assays, intact-protein mass spectrometry and small-angle X-ray scattering. The authors not only show ClpP pore widening upon ADEP binding, but also deliver several lines of evidence that the chaperone mimics select for the active, extended conformation of ClpP
17. Baytshok V, Baker TA, Sauer RT: **Assaying the kinetics of protein denaturation catalyzed by AAA+ unfolding machines and proteases.** *Proc Natl Acad Sci USA* 2015, **112**:5377-5382.
 18. Kim YI, Levchenko I, Fraczkowska K, Woodruff RV, Sauer RT, Baker TA: **Molecular determinants of complex formation between Clp/Hsp100 ATPases and the ClpP peptidase.** *Nat Struct Biol* 2001, **8**:230-233.
 19. Gottesman S, Roche E, Zhou Y, Sauer RT: **The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system.** *Genes Dev* 1998, **12**:1338-1347.
 20. Flynn JM, Neher SB, Kim Y-I, Sauer RT, Baker TA: **Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals.** *Mol Cell* 2003, **11**:671-683.

21. Siddiqui SM, Sauer RT, Baker TA: **Role of the processing pore of the ClpX AAA+ ATPase in the recognition and engagement of specific protein substrates.** *Genes Dev* 2004, **18**:369-374.
22. Martin A, Baker TA, Sauer RT: **Protein unfolding by a AAA+ protease is dependent on ATP-hydrolysis rates and substrate energy landscapes.** *Nature Publishing Group* 2008, **15**:139-145.
23. Martin A, Baker TA, Sauer RT: **Pore loops of the AAA+ ClpX machine grip substrates to drive translocation and unfolding.** *Nat Struct Mol Biol* 2008, **15**:1147-1151.
24. Lee B-G, Park EY, Lee K-E, Jeon H, Sung KH, Paulsen H, Rübsamen-Schaeff H, Brötz-Oesterhelt H, Song HK: **Structures of ClpP in complex with acyldepsipeptide antibiotics reveal its activation mechanism.** *Nat Struct Mol Biol* 2010, **17**:471-478.
- This study reveals a clear widening of the *Bacillus subtilis* ClpP entrance pore upon ADEP binding. Furthermore, it lists all relevant proton- and amino acid interactions in the hydrophobic clefts of ClpP, which are useful to the development of tailored compounds to dissect the global network of residues involved in transmitting the selection of conformational states.
25. Li DHS, Chung YS, Gloyd M, Joseph E, Ghirlardo R, Wright GD, Cheng Y-Q, Maurizi MR, Guarné A, Ortega J: **Acyldepsipeptide antibiotics induce the formation of a structured axial channel in ClpP: a model for the ClpX/ClpA-Bound state of ClpP.** *Chem Biol* 2010, **17**:959-969.
26. Singh SK, Grimaud R, Hoskins JR, Wickner S, Maurizi MR: **Unfolding and internalization of proteins by the ATP-dependent proteases ClpXP and ClpAP.** *Proc Natl Acad Sci USA* 2000, **97**:8898-8903.
27. Zeiler E, List A, Alte F, Gersch M, Wachtel R, Poreba M, Drag M, Groll M, Sieber SA: **Structural and functional insights into caseinolytic proteases reveal an unprecedented regulation principle of their catalytic triad.** *Proc Natl Acad Sci USA* 2013, **110**:11302-11307.
28. Joshi SA, Hersch GL, Baker TA, Sauer RT: **Communication between ClpX and ClpP during substrate processing and degradation.** *Nat Struct Mol Biol* 2004, **11**:404-411.
29. Zeiler E, Braun N, Böttcher T, Kastenmüller A, Weinkauff S, Sieber SA: **Vibrilactone as a tool to study the activity and structure of the ClpP1/2 complex from *Listeria monocytogenes*.** *Angew Chem Int Ed Engl* 2011, **50**:11001-11004.
30. Dahmen M, Vielberg M-T, Groll M, Sieber SA: **Structure and mechanism of the caseinolytic protease ClpP1/2 heterocomplex from *Listeria monocytogenes*.** *Angew Chem Int Ed Engl* 2015, **54**:3598-3602.
- The authors dissect the functionality of the LmClpP1/2 heterocomplex from *L. monocytogenes* and discovered a boost in substrate turnover compared to the LmClpP2 homocomplex. Furthermore, they solved the LmClpP1/2 crystal structure, which provides intriguing insights into conformational complex organization.
31. Balogh D, Dahmen M, Stahl M, Poreba M, Gersch M, Drag M, Sieber SA: **Insights into ClpXP proteolysis: heterooligomerization and partial deactivation enhance chaperone affinity and substrate turnover in *Listeria monocytogenes*.** *Chem Sci* 2017 <http://dx.doi.org/10.1039/C6SC03438A>.
32. Zhang J, Ye F, Lan L, Jiang H, Luo C, Yang C-G: **Structural switching of *Staphylococcus aureus* Clp protease: a key to understanding protease dynamics.** *J Biol Chem* 2011, **286**:37590-37601.
33. Geiger SR, Böttcher T, Sieber SA, Cramer P: **A conformational switch underlies ClpP protease function.** *Angew Chem Int Ed* 2011, **50**:5749-5752.
34. Ye F, Zhang J, Liu H, Hilgenfeld R, Zhang R, Kong X, Li L, Lu J, Zhang X, Li D *et al.*: **Helix unfolding/refolding characterizes the functional dynamics of *Staphylococcus aureus* Clp protease.** *J Biol Chem* 2013, **288**:17643-17653.
35. Michael R, Maurizi, Satyendra K, Singh, Mark W, Thompson, Martin Kessel A, Ginsburg A: **Molecular properties of ClpAP protease of *Escherichia coli*: ATP-dependent association of ClpA and ClpP.** *Biochemistry* 1998 <http://dx.doi.org/10.1021/bi973093e>.
36. Gribun A, Kimber MS, Ching R, Sprangers R, Fiebig KM, Houry WA: **The ClpP double ring tetradecameric protease exhibits plastic ring-ring interactions, and the N termini of its subunits form flexible loops that are essential for ClpXP and ClpAP complex formation.** *J Biol Chem* 2005, **280**:16185-16196.
37. Sprangers R, Gribun A, Hwang PM, Houry WA, Kay LE: **Quantitative NMR spectroscopy of supramolecular complexes: dynamic side pores in ClpP are important for product release.** *Proc Natl Acad Sci USA* 2005, **102**:16678-16683.
38. Kimber MS, Yu AYH, Borg M, Leung E, Chan HS, Houry WA: **Structural and theoretical studies indicate that the cylindrical protease ClpP samples extended and compact conformations.** *Structure* 2010, **18**:798-808.
39. Hackl MW, Lakemeyer M, Dahmen M, Glaser M, Pahl A, Lorenz-Baath K, Menzel T, Sievers S, Böttcher T, Antes I *et al.*: **Phenyl esters are potent inhibitors of caseinolytic protease p and reveal a stereogenic switch for deoligomerization.** *J Am Chem Soc* 2015, **137**:8475-8483.
- Phenyl esters are covalent inhibitors of SaClpP. They bind to the active site and depending on the stereochemistry of a methyl substituent lead to disruption of tetradecamer oligomerization. The paper presents a myriad of different phenyl ester derivatives that can be used for conformational studies around ClpP's active site.
40. Singh SK, Rozycki J, Ortega J, Ishikawa T, Lo J, Steven AC, Maurizi MR: **Functional domains of the ClpA and ClpX molecular chaperones identified by limited proteolysis and deletion analysis.** *J Biol Chem* 2001, **276**:29420-29429.
41. Brötz-Oesterhelt H, Beyer D, Kroll H-P, Endermann R, Ladell C, Schroeder W, Hinzen B, Raddatz S, Paulsen H, Henninger K *et al.*: **Dysregulation of bacterial proteolytic machinery by a new class of antibiotics.** *Nat Med* 2005, **11**:1082-1087.
42. Kirstein J, Hoffmann A, Lilie H, Schmidt R, Waigmann HR, Oesterhelt HB, Mogk A, Turgay K: **The antibiotic ADEP reprogrammes ClpP, switching it from a regulated to an uncontrolled protease.** *EMBO Mol Med* 2009, **1**:37-49.
43. Wang J, Hartling JA, Flanagan JM: **The structure of ClpP at 2.3 Å resolution suggests a model for ATP-dependent proteolysis.** *Cell* 1997, **91**:447-456.
44. Alexopoulos J, Ahsan B, Homchaudhuri L, Husain N, Cheng Y-Q, Ortega J: **Structural determinants stabilizing the axial channel of ClpP for substrate translocation.** *Mol Microbiol* 2013, **90**:167-180.
45. Böttcher T, Sieber SA: **β-Lactones as specific inhibitors of ClpP attenuate the production of extracellular virulence factors of *Staphylococcus aureus*.** *J Am Chem Soc* 2008, **130**:14400-14401.
46. Gersch M, Gut F, Korotkov VS, Lehmann J, Böttcher T, Rusch M, Hedberg C, Waldmann H, Klebe G, Sieber SA: **The mechanism of caseinolytic protease (ClpP) inhibition.** *Angew Chem Int Ed Engl* 2013, **52**:3009-3014.
47. Ni T, Ye F, Liu X, Zhang J, Liu H, Li J, Zhang Y, Sun Y, Wang M, Luo C *et al.*: **Characterization of gain-of-function mutant provides new insights into ClpP structure.** *ACS Chem Biol* 2016, **11**:1964-1972.
- An artificial mutant of the hydrophobic binding clefts on the apical surface of SaClpP was found in this paper. It renders the protease active independent of its chaperone. With the help of X-ray crystallography, it could be shown that a whole series of amino acids flips around and transmits steric information deep into the core of ClpP leading to an artificial super-activation.
48. Carney DW, Compton CL, Schmitz KR, Stevens JP, Sauer RT, Sello JK: **A simple fragment of cyclic acyldepsipeptides is necessary and sufficient for clpP activation and antibacterial activity.** *ChemBioChem* 2014, **15**:2216-2220.
49. Pahl A, Lakemeyer M, Vielberg M-T, Hackl MW, Vomacka J, Korotkov VS, Stein ML, Fetzter C, Lorenz-Baath K, Richter K *et al.*: **Reversible inhibitors arrest ClpP in a defined conformational state that can be revoked by ClpX association.** *Angew Chem Int Ed Engl* 2015, **54**:15892-15896.
- The paper introduces the first reversible inhibitor of ClpP. Based on a co-crystal structure, the authors found a 180° turnaround of Pro125 and a

switch of the His123 side chain resulting in an erroneous alignment of the catalytic triad. The inhibition mechanism is unique, because it can be reversed by ClpX binding and thus reveals a link between apical ClpX binding site and equatorial active site.

50. Rathore S, Sinha D, Asad M, Böttcher T, Afrin F, Chauhan VS, Gupta D, Sieber SA, Mohammed A: **A cyanobacterial serine protease of *Plasmodium falciparum* is targeted to the apicoplast and plays an important role in its growth and development.** *Mol Microbiol* 2010, **77**:873-890.
 51. Brötz-Oesterhelt H, Sass P: **Bacterial caseinolytic proteases as novel targets for antibacterial treatment.** *Int J Med Microbiol* 2014, **304**:23-30.
 52. Cole A, Wang Z, Coyaud E, Voisin V, Gronda M, Jitkova Y, Mattson R, Hurren R, Babovic S, Maclean N *et al.*: **Inhibition of the mitochondrial protease ClpP as a therapeutic strategy for human acute myeloid leukemia.** *Cancer Cell* 2015, **27**:864-876.
 53. Zhang Y, Maurizi MR: **Mitochondrial ClpP activity is required for cisplatin resistance in human cells.** *Biochim Biophys Acta (BBA) - Mol Basis Disease* 2016, **1862**:252-264.
 54. Jenkinson EM, Rehman AU, Walsh T, Clayton-Smith J, Lee K, Morell RJ, Drummond MC, Khan SN, Naeem MA, Rauf B *et al.*: **Perrault syndrome is caused by recessive mutations in CLPP, encoding a mitochondrial ATP-dependent chambered protease.** *Am J Hum Genet* 2013, **92**:605-613.
 55. Ahmed S, Jelani M, Alrayes N, Mohamoud HSA, Almramhi MM, Anshasi W, Ahmed NAB, Wang J, Nasir J, Al-Aama JY: **Exome analysis identified a novel missense mutation in the CLPP gene in a consanguineous Saudi family expanding the clinical spectrum of Perrault Syndrome type-3.** *J Neurol Sci* 2015 <http://dx.doi.org/10.1016/j.jns.2015.04.038>.
 56. Demain LAM, Urquhart JE, O'Sullivan J, Williams SG, Bhaskar SS, Jenkinson EM, Lourenco CM, Heiberg A, Pearce SH, Shalev SA *et al.*: **Expanding the genotypic spectrum of Perrault syndrome.** *Clin Genet* 2016 <http://dx.doi.org/10.1111/cge.12776>.
 57. Kang SG, Ortega J, Singh SK, Wang N, Huang NN, Steven AC, Maurizi MR: **Functional proteolytic complexes of the human mitochondrial ATP-dependent protease, hClpXP.** *J Biol Chem* 2002, **277**:21095-21102.
 58. Kang SG, Dimitrova MN, Ortega J, Ginsburg A, Maurizi MR: **Human mitochondrial ClpP is a stable heptamer that assembles into a tetradecamer in the presence of ClpX.** *J Biol Chem* 2005, **280**:35424-35432.
 59. Sowole MA, Alexopoulos JA, Cheng Y-Q, Ortega J, Konermann L: **Activation of ClpP protease by ADEP antibiotics: insights from hydrogen exchange mass spectrometry.** *J Mol Biol* 2013, **425**:4508-4519.
- This is an example of using hydrogen-deuterium exchange (HDX) mass spectrometry to visualize global conformational changes in *E. coli* ClpP upon ADEP binding.
60. Stinson BM, Nager AR, Glynn SE, Schmitz KR, Baker TA, Sauer RT: **Nucleotide binding and conformational switching in the hexameric ring of a AAA+ machine.** *Cell* 2013, **153**:628-639.

Danke!

Vier aufregende Jahre dieser Doktorarbeit liegen hinter mir. Ein Weg, den ich weiß Gott nicht allein gegangen bin.

Mein erster Dank gilt meinem Doktorvater Stephan Sieber. Als Student hat er mich in seiner Vorlesung mit der Aussicht auf die Welt der Chemischen Biologie begeistert. An seinem Lehrstuhl konnte ich dann selbst daran teilhaben. Danke für die vielen spannenden Projekte und die stete Begleitung bei meiner Wandlung vom Biochemiker zum Bioinformatiker! Nicht zuletzt habe ich auch als Familienvater große Unterstützung und viel Verständnis von ihm erfahren.

Des Weiteren möchte ich mich bei meinen Prüfern Matthias Feige und Steven Verhelst bedanken. Ersterer beschäftigte mich nebenbei noch mit interessanten Massenspektrometrie-Fragestellungen und letzterer brachte mir einst das Präsentieren von Wissenschaft auf Englisch bei. Ferner gilt mein Dank Kathrin Lang, die nicht nur den Prüfungsvorsitz übernahm, sondern – zusammen mit Marie-Kristin von Wrisberg – mich auch immer wieder mit interessanten Crosslink-Projekten versorgte.

Diese Arbeit sähe ganz anders aus, wenn meine vielen Kolleginnen und Kollegen nicht gewesen wären. Malte Gersch führte mich in die Welt der Proteasen ein, Nina Bach, Katja Bäuml und Lena Kunold erklärten mir so manche Schraube an einem Massenspektrometer und Mona Wolff, sowie die zahlreichen Azubis, erleichterten mir das tägliche Suchen von Chemikalien oder deren Bestellung und noch vieles andere. Christina Brumer half mir zuverlässig mit allerlei Papierkram weiter. Als sehr fruchtbar erwiesen sich die regelmäßigen Kaffeepausen, insbesondere mit dem ClpP-Team aus Mathias Hackl, Maria Dahmen, Dóra Balogh, Patrick Allihn, Anja Fux, Christian Fetzer und Markus Lakemeyer. Letzterer war jederzeit zur Stelle, wenn ein Glutamin in einer Kristallstruktur nicht da war, wo man es vermutete und Diskussionsbedarf bestand. Ich glaube, durch unsere fast täglichen Erörterungsrunden haben wir unsere jeweiligen Projekte enorm vorangebracht. Dafür konnten wir uns anfangs nicht an meinem Schreibtisch treffen, da mein geschätzter

Nachbar Wolfgang Heydenreuter leider unter ClpP-Phobie litt. Dies tat den sonstigen Konversationen mit ihm allerdings keinen Abbruch. Später saß dann Caro Gleißner neben mir, der ich für viele Aufheiterungen im Laboralltag danken möchte. Auch allen anderen früheren und derzeitigen Mitgliedern des Lehrstuhls sei an dieser Stelle herzlichst für die überragende Atmosphäre gedankt.

Die einzelnen Kapitel dieser Dissertation sind nie durch mich alleine, sondern mit vielen begabten Kooperationspartnern entstanden. Besonders hervorheben möchte ich Vadim Korotkov, Pavel Kielkowski, Leonhard Kick, Sabine Schneider, Marie-Theres Vielberg, Michael Groll und Klaus Richter, die das D9-Projekt mit ihren Verbindungen, der Kristallstruktur und Daten aus der Ultrazentrifuge maßgeblich vorangebracht haben. Marcin Poreba und Marcin Drag haben zig verschiedene Peptidderivate hergestellt, damit wir dem Schnittmuster von ClpP auf die Schliche kommen konnten. Joanna Krysiak und Jan Vomacka vermaßen *S. aureus*-Proteome, ohne die ich nie ein Virulenznetzwerk hätte generieren können. Philipp Kleiner und Wolfgang Heydenreuter brachten Stunden am Abzug zu, um die Photocrosslinker für die Photom-Bestimmung zu synthetisieren. Der ProteomeDiver war ursprünglich inspiriert aus einem Projekt von Annabelle Hoegl, die eine hervorragende Biochemikerin ist und daneben unzählige Englischtexte für mich korrigiert hat. Die Software selbst basiert auf Massenspektrometrie-Daten und der *label-free quantification* Methode. Zusammen mit Megan Wright war es eine große Freude, diese Technik an unserem Lehrstuhl einzuführen. Und schließlich Marc Vaudel, der mir immer hilfsbereit zur Seite stand, wenn es in die Tiefen der Bioinformatik hinab ging.

Bei der täglichen Laborarbeit halfen mir meine Praktikantinnen und Praktikanten weiter und es war eine Freude mein Wissen mit ihnen zu teilen. Die eine oder der andere blieb sogar für Bachelor-, Master- oder Doktorarbeit am Lehrstuhl: Susann Badmann, Sabine Helmraht, Joel Fauser, Anja Fux, Katharina Lamm, Caro Gleißner, Elisabeth Kleber, Kristina Schüller, Anna Folz, Lisa Bauerschmidt, David Frieser und Ludwig Bauer. Alles Gute für euren weiteren Weg!

Während der Promotionszeit hatte ich Mentoren, die mit Eifer meine Projekte unterstützt und mir das ein oder andere Mal den richtigen Weg gewiesen haben. Das ist zum Beispiel Aymelt Itzen, der sich mit großem Interesse die Fortschritte meiner Arbeit angehört hat. Das ist auch Stephan Hacker, der mir mit seinen Fragestellungen aus der

Chemischen Biologie immer wieder Anlass zu spannenden Programmier-Nächten gab. Und das ist Juan Vizcaíno, den ich als Vorbild sehe, was meinen Weg von der Biochemie zur Bioinformatik betrifft.

Diese Arbeit wurde maßgeblich von der Studienstiftung des deutschen Volkes finanziell wie auch ideell unterstützt. Nicht nur dafür bin ich dankbar, sondern auch für die vielen Menschen, die ich über dieses Netzwerk kennenlernen durfte.

Überhaupt in die Felder der Biochemie und Informatik gerutscht, bin ich schon von Kindesbeinen an. Nicht zuletzt deshalb möchte ich meinen Eltern Birgit und Gereon herzlich für ihre Begleitung während all der Jahre danken.

Mein größter Dank gilt meiner Familie: Madita und Linnea haben mir immer wieder gezeigt, was wirklich wichtig ist im Leben und mir jeden Abend beim Nachhausekommen ein großes Lächeln ins Gesicht gezaubert. Und Esther, wie du während all der Jahre zu mir standest, ist warmherzig, lieb und unbeschreiblich herzerfüllend.

Ich liebe dich! Danke!

Matthias Stahl

Lebenslauf

Geburtsdatum und -ort: 21. August 1987 in Speyer

Ausbildung

- 01-2014 – vsl. 03-2018 **Promotion, Lehrstuhl für Organische Chemie II, Technische Universität München**
bei Prof. Dr. Stephan A. Sieber in Chemischer Biologie und Proteomik:
Chemische Proteinmanipulationen zur Analyse der ClpP-Protease-Funktion und ihre Anwendung in der Datenauswertung chemischer Proteomik
- 10-2011 – 12-2013 **Master of Science Biochemie, Technische Universität München**
mit Spezialisierung auf Chemische Biologie, Proteomik und Onkologie
Master-Arbeit: *Isolation und Manipulation der humanen Protease ClpP*
bei Prof. Dr. Stephan A. Sieber
- 10-2008 – 08-2011 **Bachelor of Science Biochemie, Technische Universität München**
Bachelor-Arbeit: *Analyse der bei Morbus Parkinson auftretenden Aggregation von α -Synuclein mit Einzelmolekülmethoden*
bei Prof. Dr. Thorsten Hugel
- 08-1998 – 03-2007 **Gymnasium im Paul-von-Denis-Schulzentrum, Schifferstadt**
Abitur
- 08-1994 – 07-1998 **Grundschule Süd, Schifferstadt**

Arbeit und Lehrerfahrung

- seit 2014 **Lehrstuhl für Organische Chemie II, Technische Universität München**
Betreuung mehrerer Master- und Bachelor-Arbeiten sowie
Forschungspraktika in Biochemie und Bioinformatik
Dozent im Seminar *Frontiers in Chemical Biology*
- 10-2011 – 12-2011 **Forschungsgruppe Prof. Dr. Thorsten Hugel, Technische Universität München**
Wissenschaftliche Hilfskraft
Einzelmolekülmethoden
- 02-2008 – 09-2008 **Eisenmann AG, Holzgerlingen**
Betriebspraktikum Informatik
Programmierung eines Visualisierungseditors für die industrielle Leittechnik

07-2007 – 12-2013 **Site-Makers GbR, Schifferstadt**
Mitgründer und geschäftsführender Gesellschafter
Webseitenprogrammierung und Services für Vereine

04-2007 – 12-2007 **Deutsches Rotes Kreuz, Mutterstadt**
Zivildienst

Sprachkenntnisse

Deutsch	Muttersprache
Englisch	fließend in Wort und Schrift
Französisch	fließend in Wort und Schrift

Andere Verdienste

seit 05-2015 **Studienstiftung des deutschen Volkes**
Doktorandenstipendium

seit 07-2013 **Junge Akademie, Technische Universität München**
Mitglied

03-2010 – 09-2013 **Studienstiftung des deutschen Volkes**
Studienstipendium

10-2010 – 09-2013 **Max Weber-Programm im Elitenetzwerk Bayern**

2014 **Science Idol Award von Elsevier**
für einen Science Slam meiner Promotionsarbeit