Assay development for monitoring intramembrane (Rhomboid) serine protease substrate cleavage and inhibition

Oliver Vosyka

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. D. R. Langosch
Prüfer der Dissertation:

1. TUM Junior Fellow Dr. St. Verhelst
2. Univ.-Prof. Dr. St. Lichtenthaler
3. Univ.-Prof. Dr. A. Kapurniotu

Die Dissertation wurde am 25.02.2013 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 26.08.2013 angenommen.

Index of contents

Abstract

Rhomboid proteases are a recently discovered member of the unusual family of intramembrane proteases. They are conserved in all kingdoms of life and fulfill highly regulatory functions like EGF-receptor signaling, parasitic host cell invasion or quorum sensing in the pathogenic bacterium *Providencia stuarti*. Although they seem to be involved in a variety of processes in health and disease, little is known about their substrates and mechanism of action, partly due to the lack of suitable specific inhibitors and activity-based probes (ABP).

We developed a mass spectrometry based assay that utilizes a natural protein substrate and used it to screen for inhibitors and activators of bacterial rhomboid proteases. This assay may also be applicable for rhomboids from other species, other intramembrane proteases and is shown to be suitable to determine rhomboid cleavage kinetics. We identified a range of inhibitors with IC_{50} values in the low micromolar range and discovered an unusual mode of inhibitor binding using protein crystallization and determine their mechanism of action in biochemical studies.

The new ABPs, identified in the screening are molecules that label active rhomboids but not their inactive counterparts and can for example be used to study regulatory mechanisms of protease activity. We used the ABPs to label endogenous *E. coli* rhomboid GlpG in cell lysates as well as expressed GlpG *in vivo*. This provides a powerful set of tools for functional cellular assays addressing the biological function of bacterial rhomboids, which is not known for many rhomboids.

1 Introduction

1.1 Proteases

In the human genome 500-600 proteases have been identified, which is approximately 2% of the genes. Proteases are enzymes that catalyze hydrolysis of peptide bonds within a polypeptide chain. Even though proteases were first thought to only have degradational function, proteolytic mechanisms that regulate a broad spectrum of cellular processes such as apoptosis, antigen presentation and blood coagulation have been discovered. Proteases are optimal tools to irreversibly regulate such important biological processes. Dysregulated proteolytic activity is known to be involved in a variety of diseases like neurodegenerative (Lichtenthaler et al, 2011) and vascular diseases (Siefert & Sarkar, 2012) as well as cancer (Puente et al, 2003). This pushed proteases into the spotlight of pharmaceutical interest as potential drug targets.

1.1.1 Protease classification

Proteases are classified into five main subgroups according to their active site architecture. (I) In the active site of threonine proteases a nucleophilic threonine is located. One prominent representative is the catalytic subunit of the proteasome, which is involved in protein quality control and antigen presentation. (II) Cysteine proteases have a catalytic triad with a cysteine - nucleophile located in the active site. Members of this protease subgroup are for example caspases, which are involved in processes like apoptosis and inflammation. Cathepsins are another family of cysteine proteases, which are for example contributing to lysosomal degradation. (III) The HIV1 protease is an important drug target in AIDS therapeutic strategies and member of the aspartate proteases, in which generally two aspartate residues form the active site. (IV) A special class of proteases are metalloproteases. The catalytic activity of these enzymes requires a metal ion – generally zinc or cobalt – coordinated in the active site by three amino acid residues. (V) The largest class of proteases are the serine proteases. They ubiquitously occur in both eukaryotes and prokaryotes and carry a nucleophilic serine in their active center. These enzymes catalyze protein degradation during digestion (e.g. trypsin), but also regulate processes like blood coagulation (thrombin) and immune response.

1.1.2 Intramembrane proteases

Most proteases are soluble and in physiological conditions are surrounded by an aqueous environment as are their substrate proteins but the newest and most unexpected class are intramembrane proteases (Wolfe, 2009). In the last decades it became clear that proteolysis does not only occur in the soluble regions of a substrate protein, but also in the transmembrane domain (TMD) of integral membrane proteins deeply buried into the hydrophobic lipid bilayer.

1.1.2.1 Intramembrane proteases classification

Several families of intramembrane proteases are known (numbering according to the Merops database): the metalloprotease family M50 (site-2-protease) the aspartic protease family A22 (signal peptide peptidase and γ-secretase) and the serine protease family S54 (rhomboid). The mammalian site-2 protease (S2P) was identified as the first intramembrane protease (Rawson et al, 1997). It cleaves inside the TMD of transcription factors and releases them from their membrane anchor but is also involved in endoplasmatic reticulum stress response (Ye et al, 2000). The signal peptide peptidase and γ-secretase hold two aspartate residues in the active site. The latter is a protease complex consisting of at least four different proteins with presenilin being the catalytic subunit. Dysfunctional presenilin is thought to be the major risk factor for the neurodegenerative Alzheimer's disease since miscleavage of the amyloid precursor protein produces $A\beta_{42}$ peptides, a major component of the senile plaques (Wolfe, 2009). The signal peptide peptidase processes signal peptides and is involved in the quality control of membrane proteins located in the ER (Schroder & Saftig, 2010) but it also regulates e.g. the reproduction rate of the Hepatitis C virus. Rhomboid proteases were identified as intramembrane serine proteases in 2001 by Urban et al. (Urban et al, 2001).

1.1.2.2 Rhomboid protease

Rhomboid proteases, which are conserved in all kingdoms of life were actually discovered in genetic studies of the fruitfly *Drosophila melanogaster*, when a mutation in rhomboid-1 gene showed a rhomboid shaped – and therefore name giving – head skeleton shape (Mayer & Nusslein-Volhard, 1988). Further studies on rhomboid-1 - the corresponding protein – revealed its function in developmental control of *Drosophila* embryos by catalyzing the proteolytic release, and therefore activation of the EGF-like growth factor Spitz from its membrane anchor TMD in the epidermal growth factor receptor (EGFR) pathway (Freeman, 1994; Sturtevant et al, 1993). It also catalyzes biological processes as diverse as mitochondrial dynamics, invasion of host cells by apicomplexan parasites (Malaria (O'Donnell et al, 2006)) and bacterial protein export (Quorum sensing (Stevenson et al, 2007)). Nevertheless the roles of many rhomboids remain to be discovered.

Rhomboids differ from the other intramembrane protease families by releasing factors to the outside of cells rather than to the cytosol and by cleaving intact membrane proteins without the requirement of pre-shedding of the substrate protein, which is e.g. the case for γ-secretase which requires an initial ectodomian shedding event by the protease BACE1 for substrate recognition (Lichtenthaler et al, 2011; Prox et al, 2012). The reason for that might be the different active site architecture of rhomboid proteases.

1.1.2.2.1 Structure of rhomboids

The structure of *E. coli* rhomboid GlpG was the first one to be solved and is the best understood so far, even though its biological function remains to be discovered. Not only did this structure confirm that the proteases active site is indeed located within the plane of the lipid bilayer and gave a first insight into the positions of the catalytic dyad residues. It also revealed the protein architecture which allows access of water to the active which appears problematic since the localization of the active site inside the hydrophobic environment of a membrane. Structures of GlpG show that rhomboid solves that paradox by forming a cup-shaped, water filled cavity (Figure 1) which allows water molecules to access the enzymes active site.
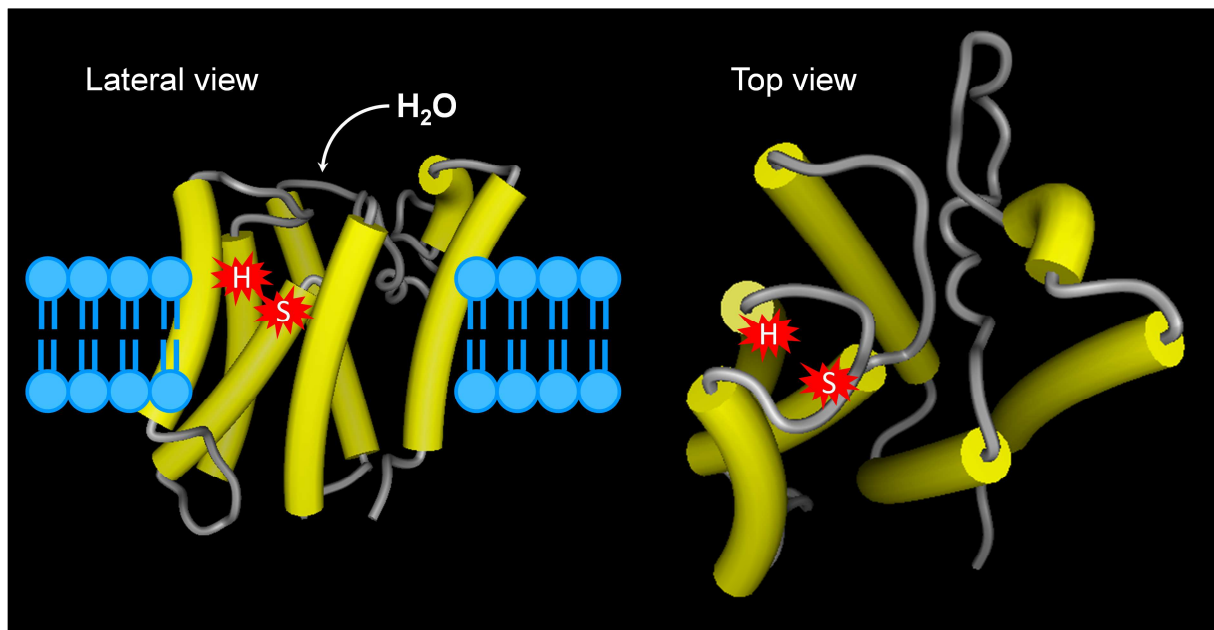
Figure 1: Structure of *E. coli* rhomboid GlpG. The active site residues (serine-histidine catalytic dyad) of GlpG are located several Å beneath the surface of the lipid bilayer (lateral view; left). The active site is shielded to the hydrophobic environment by the surrounding TMDs (top view; right) and allows water molecules, crucial for hydrolysis of the substrate protein to access the active site catalytic dyad.

Recent studies of GlpG revealed a hydrophilic cavity close-by the proteases active site suggested as water retention site. In that case water molecules do not randomly get access to the active site, but might be provided by a water molecule enriching site close to the active site serine (Zhou et al, 2012). Crystallographic analysis of both, the detergent solubilized GlpG (Ben-Shem et al, 2007; Wang et al, 2006; Wu et al, 2006) as well as reconstituted in liposomes (Vinothkumar, 2011) show that GlpG consists of 6 TMDs, a core structure which is common upon all rhomboid proteases and harbors the amino acid residues of the serine-histidine catalytic dyad (described in more detail in chapter 1.1.2.2.2). Some few microbial rhomboids differ from that by consisting of only 5 TMDs (Kateete et al, 2012) in which the active site residues are located in TMDs 3 and 5 instead of 4 and 6 which is the case for most other rhomboids (Figure 2). Members of the rhomboid family may also have a variable N-terminal domain that sticks out into the cytosol or a seventh TMD. It was shown that GlpG lacking the soluble, cytoplasmic domain displays reduced enzymatic activity in comparison to the full-length protein (Lohi et al, 2004; Sherratt et al, 2012; Wang et al, 2006). A recent study suggests this soluble domain being involved in substrate binding (Lazareno-Saez et al, 2013). Rhomboid proteases from eukaryotic organism

typically have a seventh, C-terminal TMD now making the C-terminus point to the luminal side of the membrane (Ya 2008) (Figure 2A). One exception are mitochondrial rhomboids which have an extra N-terminal TMD. But since the N-terminus remains sticking out at the luminal side, the orientation of these rhomboids appears to be upside down with the enzymes active center facing - and consequently releasing the soluble portion of their substrate to - the matrix side of the membrane (Figure 2B) (Hill & Pellegrini, 2010).
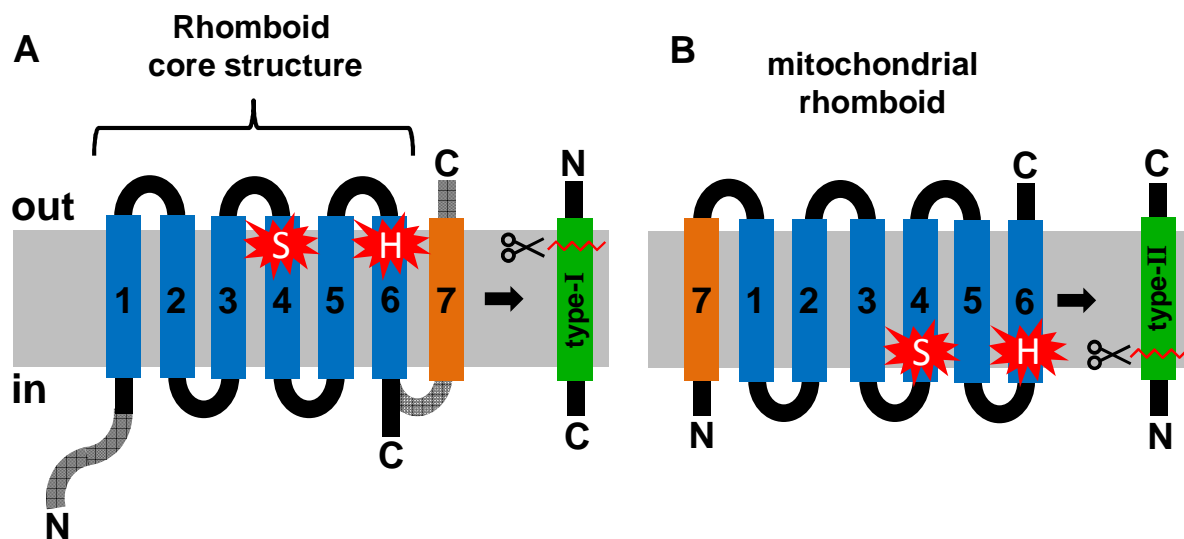
Figure 2: Schematic structure of rhomboid proteases: (A) Rhomboid consists of a general rhomboid core structure of 6 TMDs (blue; e.g. *E.coil* GlpG), with an optional 7th C-terminal TMD found in most eukaryotic rhomboids and some bacterial rhomboids (orange; e.g. *P.stuartii* AarA) and optional elongated N-terminus (grey). The typical substrates are type-I membrane proteins (N_{in} - C_{out}) (green). (B) General structure of mitochondrial rhomboids with an additional 7th TMD at the N-terminus (orange), typically cleaving type-II membrane proteins (C_{in} - N_{out}) (green).

In order for the substrate to gain access to the membrane buried active site, conformational changes in the enzyme structure must take place. The exact mechanism of this is still not known. One hypothesis is based on the observation that a surface loop (L5), which covers the catalytic dyad is flexible (Maegawa et al, 2007; Wang et al, 2007) and that the lipid bilayer appears to be narrowed around the protease. The substrate TMD may stick out the membrane and bend into the active site via the opening created by shifting the L5 cap (Wang & Ha, 2007; Xue et al, 2012). The other hypothesis focuses on the observation that TMD5 is twisted away from the enzyme creating a lateral gate inside the lipid bilayer which is proposed to be the substrates entrance to the active site (Baker et al, 2007; Wu et al, 2006).

1.1.2.2.2 Mechanism of rhomboid proteolysis

The active center of a typical, soluble serine protease like trypsin harbors the catalytic triad, which is preserved in almost all serine proteases. It consists of the amino acids histidine, serine and aspartic acid in a strictly coordinated structure. The hydroxyl group of the serine is capable of acting as a strong nucleophile, which is attacking the carbonyl carbon atom of the scissle bond. In order to do so, the nitrogen of the histidine needs to abstract a proton from the serine. The aspartic acid forms a hydrogen bond with the histidine and thereby makes the nitrogen atom more electronegative. The active site of rhomboid proteases differs from the described regular active site by having a catalytic dyad, which consists of a serine and a histidine residue.
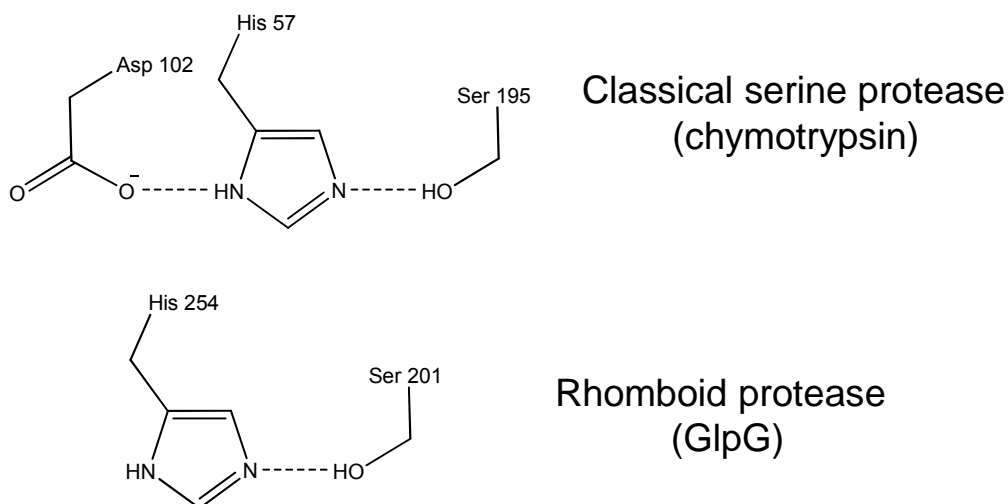
Figure 3: **Rhomboid active site**. The active site of classical serine proteases contains a triad of serine, histidine and aspartic acid (chymotrypsin amino acid numbering). The catalytic machinery of rhomboids is membrane embedded and consists of a catalytic dyad formed by S201 and H254 in TM4 and TM6, respectively (GlpG, numbering; depicted structure 2IC8).

An asparagine rather than an aspartic acid is located very close to the catalytic dyad, which appears to be important albeit not crucial for rhomboid substrate cleavage (Baker et al, 2007). The exact function of the asparagine therefore remains elusive but may be involved in stabilizing the oxyanion hole transition state. The catalytic serine is located at the end of TMD 4, several Å beneath the surface of the lipid bilayer. Crystal structures of GlpG revealed the active site histidine, located in TMD 6

is in close enough proximity to act as a general base, which is capable of taking away a proton from the serine and turning it into a reactive nucleophile. The attack onto the scissile bond of the substrate is proposed to occur at the si-face, which is the opposite site where cleavage occurs with most other serine proteases (Baker et al, 2007; Brooks et al, 2011).

1.1.2.2.3 Function and role of rhomboid proteases in disease

Rhomboid proteases are almost ubiquitous distributed among all living organisms and exerting their general role of cleaving transmembrane proteins. Subsequently cleaved, soluble parts of the substrate proteins are released from the membrane in order to trigger certain cellular responses. Whether the soluble or the membrane bound cleavage product is the activated protein depends on the organism. In the human pathogen bacterium *Providencia stuartii,* which causes inflammations of the urinal tract (Dealler et al, 1988), rhomboid is involved in intracellular communication (quorum sensing) by regulating assembly of the twin arginine transport-pore (Tat) (Fritsch et al, 2012). *P. stuartii* TatA, in contrast to other bacteria has an N-terminal extension which needs to be removed by the rhomboid protease AarA to form the Tat-pore.

Rhomboid is also involved in host cell invasion of the malaria pathogen *Plasmodium berghei*. In this apicomplexan organism, rhomboid sheds surface proteins involved in recognition of - and interaction with - host cell surface-receptors in order to invade cells (Ejigiri et al, 2012; Srinivasan et al, 2009).

Human rhomboids are involved in processes like wound healing by cleaving thrombomodulin (Cheng et al, 2011), EGFR-signalling since rhomboid RHBDL-2 was recently shown to cleave EGF (Adrain et al, 2011) and function of mitochondrial rhomboids has been linked to Parkinson's disease (Meissner et al, 2011; Shi et al, 2011).

1.2 Activity-based probes

The big challenge in proteomics is not only the identification and quantification of proteins, but also the assignment of protein function, protein localization and the mapping of regulatory pathways and networks. Unfortunately classical proteomic methods do not provide information about the activity state of enzymes. Small molecule (SM) chemical probes have become powerful tools for qualitative analysis of a proteome. Electrophilic probes with intrinsic reactivity can be used to covalently label enzymes like proteases. This allows for example mass spectrometry-based identification of the binding site or gel-based analysis of the enzyme-probe complexes. Such tools are particularly useful for the study of proteolytic activity in complex proteomes. Chemical probes can be designed in a way that they exclusively bind to an intact and functional active site of a protease but not their inactive counterparts such as zymogens or inhibitor bound enzymes. These so called activity-based probes (ABP) have recently been used for imaging of proteases in intact organisms (Bogyo et al, 2004; Ren et al, 2011), for identifying new protease inhibitors (Deu et al, 2010; Knuckley et al, 2010) and to monitor proteolytic activity and regulation thereof (Heal et al, 2011).

Hydrolysis of substrate proteins by proteases is often under control of post translational modifications, which directly regulate enzymatic activity as well as the proteases substrate specificity, generally directed by the amino acid residues flanking the scissile bond. Most proteases are expressed in an inactive (zymogen) form which needs to undergo proteolytic processing for activation. Unfortunately many proteomic methods can't distinguish between the active and the zymogen form of the protease which is important for understanding their biological function. Therefore the use of inhibitors and ABPs has led to remarkable gain in understanding the role of proteases in physiological and pathogenic processes. Until now, for rhomboid proteases no posttranslational regulation has been discovered. Except for one known example, the human rhomboid RHBDL2, which is expressed in a zymogenic form and needs to be processed by an unknown protease to gain proteolytic activity (Lei & Li, 2009), all rhomboid proteases are expressed in their catalytically active form. In fact, modulation of activity of these highly regulative enzymes remains elusive. One mechanism of regulation is the localization of rhomboid and its substrate protein in

different organelle membranes as it is the case for *Drosophila* RHOMBOID-1 which is located in the membrane of the Golgi apparatus. Its substrate protein Spitz is located in the endoplasmatic reticulum membrane and requires relocalization to the Golgi by the transmembrane protein Star (Klambt, 2002; Lee et al, 2001). Star is therefore a factor that regulates rhomboid activity by limiting the access of the substrate protein to the rhomboid protease.

Another strategy of regulating rhomboid proteolytic activity was recently discovered: it involves a subfamily of inactive rhomboid homologues named iRhoms (Lemberg & Freeman, 2007; Zettl et al, 2011). As it is the case for a large number of different enzymes (proteases, kinases, phosphatases etc.) (Pils & Schultz, 2004; Todd et al, 2002) iRhoms - inactive due to lacking the active site residues - are not only expressed but also well conserved in living organisms, indicating that they are underlying selective pressure in order to maintain their function (Adrain & Freeman, 2012). In *Drosophila*, iRhom counteracts the cleavage of EGFR-signal protein by the active rhomboid and consequently inhibits EGFR-signaling. iRhom furthermore directs rhomboid substrates to ER-associated protein degradation by the proteasome (Zettl et al, 2011). Therefore iRhoms regulate rhomboid cleavage by controlling the accessibility of its substrate proteins that are located inside the same membrane.

It was also proposed that the composition of the lipid bilayer, which hosts the rhomboid protease, might be a regulating factor as it is the case for other intramembrane proteases like γ-secretase (Osenkowski et al, 2008). This hypothesis is based on the finding that rhomboid displays – when performing substrate cleavage in the presence of different natural lipid extracts - changed enzymatic activities compared to the detergent solubilized rhomboid (Urban & Wolfe, 2005). ABPs are optimal to monitor the influence of various parameters, such as lipid or detergent environment (Vosyka et al, 2013) on enzymatic activity.

ABPs generally consist of three different elements: The warhead, a detection tag and the so called spacer. The warhead is a functional, electrophilic group that covalently binds to the proteases active site with its enhanced nucleophilicity and therefore ABPs are often derived from the large pool of mechanism-based inhibitors, which are known to irreversibly target proteases (Powers et al, 2002). For serine proteases the very general serine protease inhibitor diisopropyl fluorophosphonate (DFP) led to the development of the fluorescent FP-rhodamine ABP which is capable of labeling a

large number of serine proteases (Kidd et al, 2001). Unfortunately such probes can only be developed for cysteine-, threonine- and serine proteases since these protease classes are using the name giving amino acid residues as nucleophile, while aspartate- and metalloproteases are using an activated water molecule to perform a nucleophilic attack on the peptide scissile bond. The latter two classes are consequently not leading to covalent labeling of the enzyme but to hydrolysis of the inhibitor / ABP.
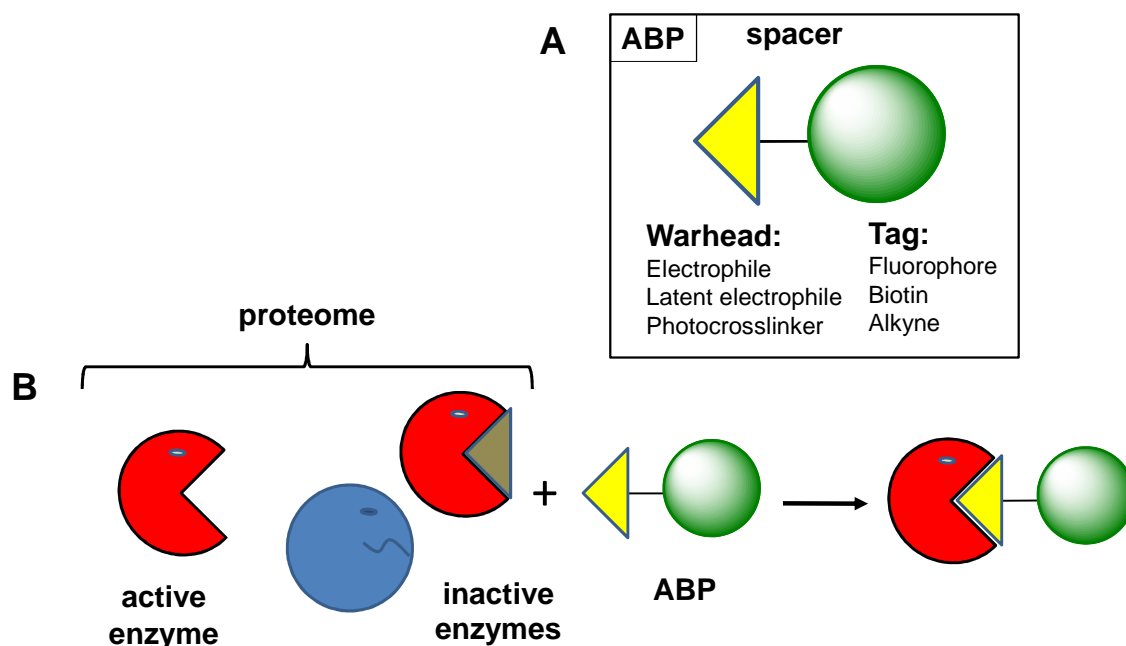
Figure 4: Activity based enzyme labeling. (A) ABPs generally consist of a warhead that covalently binds to the enzymes active site, a spacer and a detection tag. (B) ABPs only bind to catalytically active enzymes but not inactive enzymes like inhibitor bound or – in case of proteases - zymogenic forms. The detection tag enables to distinguish active from inactive enzymes.

The visualization tag can be chosen from a large variety of molecules depending on the scientific purpose. Radioisotopes such as iodine-125 are very sensitive and have been widely used for in-gel detection of protease activity but since the sensitivity of other detection tag has been improved over the last years, the use of radioisotopes has decreased. Nowadays a large variety of fluorescence-based detection tags are commonly used. Another popular detection tag is biotin, since it can not only be used to detect active enzymes but also to perform pull-down experiments (Florea et al, 2010). Unfortunately such detection tags often lead to limitation for *in vivo* and *in situ* labeling experiments due to their weak cell permeability. Additionally bulky detection groups may sterically hinder the ABP from entering the enzymes active site. This was

solved by the development of tandem labeling strategies which separates the labeling from the visualization steps by making use of bioorthogonal chemistry. In this method, the ABP is not directly linked to a reporter tag but to a small bioorthogonal reactive group that can be derivatized with a reporter tag in a second step. The reactive groups normally consist of only few atoms and therefore do not alter the structure and performance of the ABP. For proteases, the most commonly used tandem labeling strategy is the azide-alkyne 1,3-dipolar cycloaddition (Serim et al, 2012), which is often referred to as "click chemistry". Here either the warhead is linked to an alkyne functional group and the detection tag to an azide or vice versa. The alkyne reacts with the azide by forming a triazole product, covalently linking the warhead, which is in complex with the labeled enzyme to the detection tag.

Depending on the scientific question either ABP with high or low selectivity are needed. FP-rhodamine is a good choice for simultaneously labeling different serine proteases as long as they separate well during gel electrophoresis. Activity-based imaging applications require highly selective probes. The selectivity of an ABP can be tuned by the choice of warhead (Haedke et al, 2012) and by the linker. Warheads for serine proteases have mainly been based on fluorophosphonates, diphenyl phosphonates and isocoumarines (Serim et al, 2012). Depending on the active site structure a protease may prefer one of them. Introducing a peptidic linker is also a useful technique to improve selectivity towards a single protease. Most proteases have a well-defined preference for certain amino acids in the P1 and P1' position of the substrate protein, which are defining the scissile bond. In case of proteases having a regulative rather than a digestive function also the residues flanking the scissile bond may be important for substrate recognition and can consequently be used to fine-tune substrate specificity of ABPs.

1.3   Measuring enzymatic activity by mass spectrometry

Since the development of mild ionization techniques like matrix assisted laser desorption/ionization mass spectrometry (MALDI-MS) (Hillenkamp et al, 1991) and electrospray ionization mass spectrometry (ESI-MS) (Fenn et al, 1989) has become one of the most powerful methods in protein biochemistry. Even though mainly used

for protein identification and quantification, MS is an attractive method in enzymology since it can theoretically be used to analyze any enzymatic reaction that results in a mass difference between substrate and product. In the beginnings of using MS for assaying enzymatic reactions one was limited to qualitative applications like "is the expected product detected?". The development of quantitative MALDI and ESI enabled direct analysis of enzyme activity, kinetics and inhibition thereof by quantifying substrate turnover. Consequently MS based techniques have increasingly been used in biochemical and pharmaceutical studies during the last decades (Angel et al, 2012; Banerjee & Mazumdar, 2012). The advantage of using MS based methods in enzymology is the direct measurement of natural substrates and products instead of using artificial substrates since the kinetic behavior of such synthetic substrates may differ significantly from the natural and therefore resulting in falsified enzyme kinetics (Letzel, 2008; Liesener & Karst, 2005). With increasing sensitivity, throughput and miniaturization MS became a powerful tool for identifying inhibitors of enzymatic activity, a field which is so far dominated by fluorescence based methods. Such inhibitor screening assays have been reported for a variety of enzymes such as kinases, esterases and proteases (Greis et al, 2006; Liesener et al, 2005; Steinkamp et al, 2004) making use of both MALDI and ESI ionization techniques monitoring the mass shift of synthesized peptidic substrates upon enzymatic turnover. So far none of these assays made use of intact protein substrate which optimally matches the enzymes substrate selectivity. Therefore we aimed to detect rhomboid activity and inhibition using mass spectrometry in order to overcome limitations of the existing methods. The established methods may not be restricted to rhomboid but also serve as a prototype for other intramembrane proteases.

2 Results and discussion

### 2.1 Monitoring rhomboid activity by mass spectrometry

Gel based assays are the most widely used methods to monitor rhomboid substrate cleavage. These assays were used to detect rhomboid cleavage in bacteria, in cell culture and *in vitro* using purified rhomboids (Stevenson et al, 2007; Strisovsky et al, 2009; Urban & Wolfe, 2005). However, gel based protease assays are not optimal for studies that require a high sample throughput like inhibitor screenings. Substrates that lead to only a small difference in mass upon proteolytic cleavage as it is the case with TatA often cannot be resolved by gel electrophoresis. The N-terminus of TatA that is cleaved off by the rhomboid AarA consists of only 8 amino acids and leads to a 0,8 kDa mass difference, which is difficult to separate from the uncleaved form by gel electrophoresis (Stevenson et al, 2007). However, mass spectrometry is capable of monitoring these small mass differences. We therefore considered it as optimal detection method for direct measurement of rhomboid substrate cleavage.

2.2 Rhomboid inhibitor screening by MALDI-MS

Rhomboid is resistant to many of the serine protease inhibitors which might be due to the different active site architecture. In contrast to the active site of soluble serine proteases which consists of a serine, histidine aspartate catalytic triad, the rhomboid proteases hold a serine histidine catalytic dyad (Figure 3). However, some inhibitors were found to moderately inhibit rhomboid activity such as the very general serine protease inhibitors diisopropyl fluorophosphonate (DFP) (Xue & Ha, 2012) and 3,4-dichloroisocoumarin (DCI) (Harper et al, 1985; Lemberg et al, 2005; Urban et al, 2001; Urban & Wolfe, 2005). Identifying potent and selective inhibitors for rhomboids is of great scientific value since such tools can for example be used to discover the physiological role of enzymes, which is not known for many rhomboids. Protease activity or inhibition thereof is mostly monitored by making use of small fluorogenic substrates. Only one FRET-based inhibitor screening assay for the rhomboid AarA of the Gram-negative bacterium *P. stuartii* has been reported, which made use of a 16-

amino acid FRET peptide and led to the identification of monocyclic β-lactams as new mechanism based inhibitors (Pierrat et al, 2011), but many rhomboids do not cleave this substrate efficiently. The development of fluorogenic or chromogenic peptide substrates as they are available for many soluble proteases is difficult for rhomboids. Such reagents are based on the protease substrate specificity, which is generally governed by the amino acid sequence around the scissile bond. However, it is unclear how substrate recognition and binding by rhomboid proteases takes place and for most rhomboids, no peptidic substrates are available. Although helix-breaking residues within the substrate TMD seem to be an important factor for cleavage it remains largely unclear how the initial substrate recognition by rhomboid intramembrane proteases takes place. For the rhomboid AarA, a general recognition motive around the scissile bond has been identified, which successfully was used to predict AarA substrates within the *P. stuartii* proteome (Strisovsky et al, 2009). But it clearly shows that substrate recognition does not only depend on the amino acids flanking the scissile bond, but may extend over a large part of the TM. Therefore peptidic substrate for this class of proteases might miss structural conformations that are important for protease substrate recognition and cleavage. Consequently small peptide substrates may not be cleaved efficiently enough to be used as diagnostic tools in inhibitor screens. The current insights in rhomboid substrate recognition haven't resulted in the design of specific, small molecule reporters yet. Furthermore synthesizing such peptide substrates for rhomboids can be challenging because of the hydrophobic nature of the substrate proteins since peptidic substrates need to contain at least parts of the substrates TMD. One study used small molecule probes for ABPP of rhomboid GlpG (Sherratt et al, 2012) and suggests making use of a rhodamine tagged fluorophosphonate, a very general serine protease ABP (Evans & Cravatt, 2006) could potentially be used to identify new inhibitors by monitoring inhibition of ABP labeling. Such peptidic or small molecule tools have proven much value for identifying novel active site modifiers for different enzymes but are not capable of identifying inhibitors (reversible or irreversible) which interfere in enzyme-substrate complex formation by binding to exosite-positions (regions outside the active site but influencing catalytic activity) in the enzyme. Using intact substrate proteins for inhibitor screenings is suitable to identify exosite-inhibitors as well as active site inhibitors (Bannwarth et al, 2012), while conventional screening assays mainly focus on active site binders. For rhomboid proteases, various natural and

engineered protein substrates are known (Strisovsky et al, 2009). To circumvent the design of a peptide substrate, rhomboid cleavage of protein substrates and inhibition thereof was directly monitored in a label- and gel-free analysis method making use of MALDI-MS.
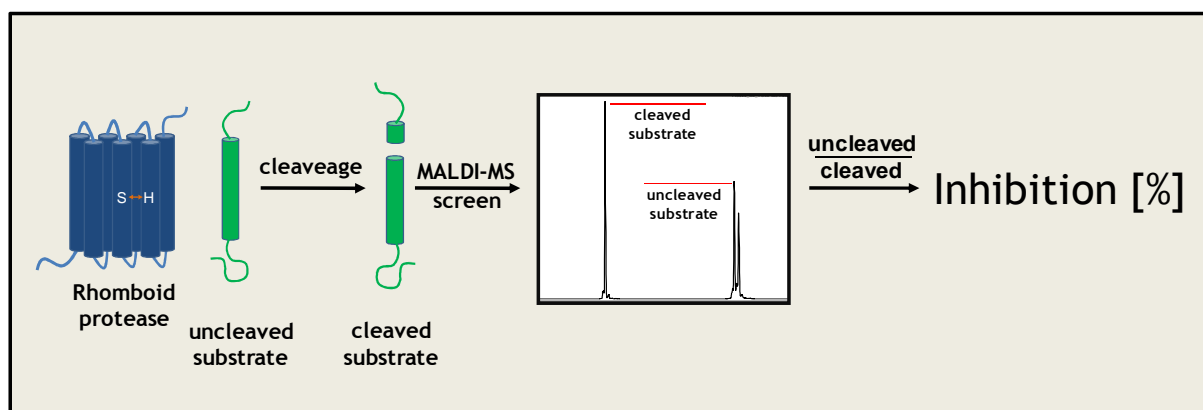
Figure 5: Schematic picture of MALDI assay. *In vitro* cleavage by rhomboid protease is monitored using MALDI mass spectrometry by detecting loss of mass of the substrate protein. The ratio of signal intensities of cleaved and uncleaved substrate is used as a measure of proteolytic activity.

To this end recombinant *E. coli* rhomboid GlpG, *P. stuartii* rhomboid AarA and its natural substrate TatA, which is also substrate of GlpG, were expressed in *E. coli* and purified in dodecylmaltoside (DDM) micelles. MALDI-MS was chosen for detecting *in vitro* rhomboid cleavage, because it is much less restricted in the use of salts and buffers compared to ESI-MS, requires only minimal sample preparation efforts and therefore enabling high throughput screenings. MALDI is therefore an ideal method to overcome the limitations of peptidic substrates.

2.2.1 Sample preparation

Inhibitor screenings by MALDI-MS requires optimal sample preparation since homogeneous co-crystallization of the MALDI-matrix and the analyte molecules is crucial to gain good quality of the mass spectra. Contaminants like salt, detergents and solvents like DMSO negatively influence the co-crystallization of the MALDI-matrix with the analyte. Unfortunately all the mentioned contaminants were present in rhomboid inhibition reactions, since detergent (DDM) is necessary for solubilization of

the membrane proteins in a HEPES buffer environment and inhibitors were stored as a DMSO sock solution. Rhomboid proteases have also been solubilized in their active form by using Triton-X-100 (Lemberg 2005) and this detergent is known to result in reduction of signal intensities and affects the quality of the spectra. The detergent DDM is more MS compatible and was therefore chosen for this study (Börensen 1997). The establishment of a high throughput inhibitor screening assay to identify new rhomboid inhibitors based on MALDI-MS technology ideally does not make use of purification steps to remove these contaminants since that would be time consuming and costly. Therefore it was pursued to optimize sample preparation in a way that enables direct transfer of the quenched sample to the MALDI target plate and to gain reproducibly robust mass spectra to monitor rhomboid activity.

2.2.1.1 Matrix selection

The first and often most important step in the optimization of sample preparation for MALDI-MS is the right choice of MALDI matrix. Today, a broad range of different MALDI matrices are available each being optimal for different MALDI-MS samples. For inhibitors screenings it is desired to repeatedly measure the same analytes in a large number of samples. Therefore it is recommendable to first choose the MALDI matrix that provides the best signal intensities for the specific sample. For the analysis of the substrate protein TatA, four different MALDI matrices were selected: α-cyano-4-hydroxycinnamic acid (CHCA), sinapinic acid (SA), 2',4',-dihydroxyacetophenone (DHAP) and 2-(4'-hydroxybenzeneazo)benzoic acid (HABA) (Figure 6). CHCA is a commonly used MALDI matrix to analyze proteins and peptides less than 10 kDa while SA is a good choice for large peptides or proteins >10 kDa. DHAP is a specialized matrix primarily used to analyze glycoproteins and complex protein mixtures. HABA can be used to analyze oligosacharides or intact proteins.

Purified TatA was diluted in rhomboid reaction buffer (50 mM HEPES / 0,05 % DDM) to concentrations of 10 µM or 20 µM, directly mixed with the different matrix solutions and analysed by MALDI-MS (Figure 6).

DHAP showed the best performance to analyze TatA, judged by the signal intensities of the unformylated TatA and was therefore used in all measurements. The signal intensity of 10 µM TatA concentration led to approximately 40 % increase signal intensity than 20 µM. This may be due to suboptimal matrix crystal formation caused by high TatA concentration. Mixing the sample 1:1 with 2% TFA to provide protons for ionization of the analyte has the positive side effect of immediately quenching proteolytic activity. DHAP was therefore chosen to measure enzymatic turnover of 10µM of TatA by rhomboid proteases.
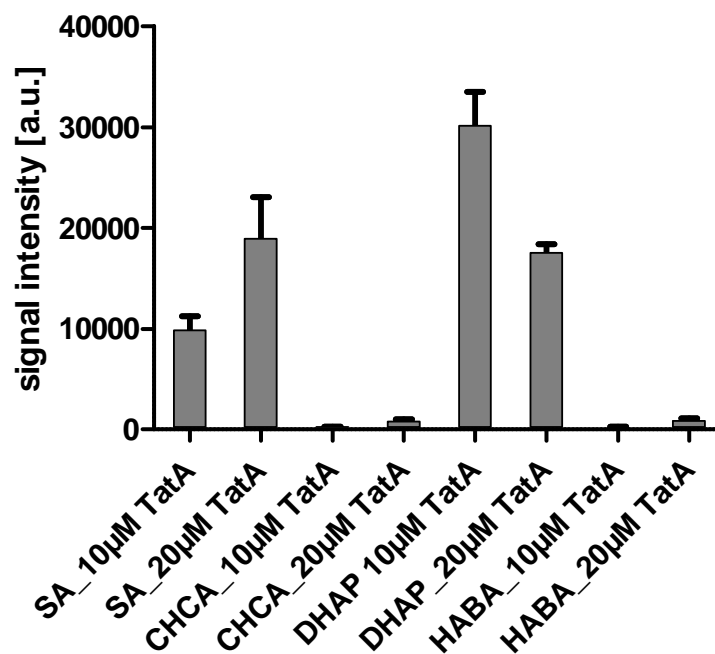
Figure 6: Different MALDI matrices to improve measurement of TatA. TatA (in HEPES reaction buffer containing 0,005% DDM; 10 or 20µM respectively) was directly analysed in five independent measurements using MALDI matrices sinapinic acid (SA), α-cyano-4-hydroxycinnamic acid (CHCA), 2',4',-dihydroxyacetophenone (DHAP) and 2-(4'-hydroxybenzeneazo)benzoic acid (HABA). Signal intensities of the unformylated TatA protein are shown and used to estimate MALDI matrix suitability.
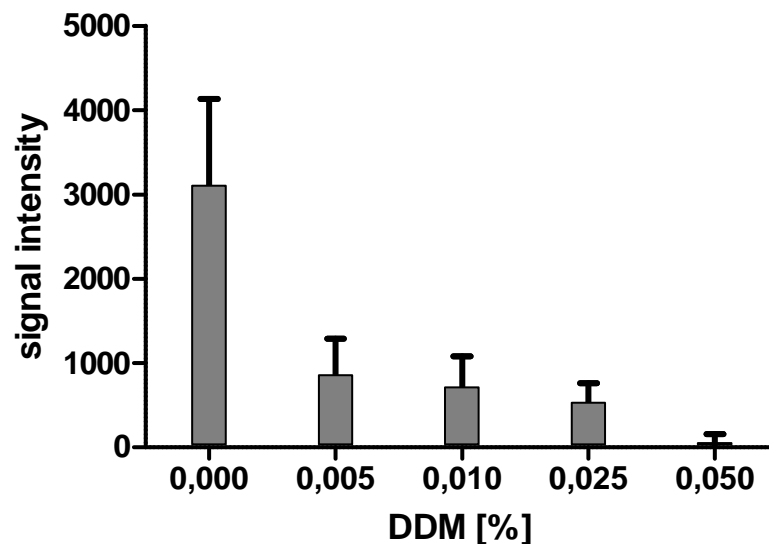
2.2.1.2 Detergent

Rhomboid *in vitro* assays were done with recombinantly expressed and detergent solubilized enzymes and substrate proteins. Detergents are surfactants consisting of a hydrophobic and a hydrophilic portion. In low concentrations detergents are forming monolayers, while at higher concentrations, the so called critical micellar

concentration (CMC), they form micelles. The choice of the right detergent depends on its physicochemical properties. The use of detergents in protein biochemistry ranges from complete denaturation to the primary structure of protein structures by using ionic detergents like sodium dodecylsulfate (SDS) (e.g. prior to gel electrophoresis) to reconstitution of native membrane proteins and even active enzymes into detergent micelles (le Maire et al, 2000) using non-ionic detergents like Triton X-100. Detergents in biochemical analytics often seem to be a necessary evil. The preparation of protein or peptide samples often requires the presence of these molecules since detergents are essential for a large number of proteomic methods (Aebersold et al, 1987; Henzel et al, 1993; Laemmli, 1970) and scientists can choose the optimal detergent from a large pool of commercially available surfactants. Unfortunately detergents can also heavily interfere with analytical methods like chromatography, protein purification or MS applications (Bornsen et al, 1997). In MS detergents leads to background signals in the lower m/z ranges often overlapping with signals of peptidic analytes and high concentrations of detergents do affect matrix-analyte crystallization during MALDI-MS negatively thus disabling successful analysis. The detergent DDM, which was used to solubilize rhomboid proteases in an active state (Urban & Wolfe, 2005) is known to be compatible with mass spectrometry (Cadene & Chait, 2000). But even though DHAP was identified as the optimal MALDI matrix to measure TatA in the presence of DDM, a detergent concentration of 0,05 % which was commonly used for rhomboid *in vitro* cleavage led to low signal intensities and often impeded matrix crystallization, thus consequently disabling analysis by MALDI-MS. Different methods to remove the detergent molecules prior to MALDI-MS analysis were tested: reversed phase desalting columns (C4 and C18 StageTipps; Proxeon), PD-10 desalting cloumns (Ge Healthcare) and Bio-Beads (BioRad).

Unfortunately all of these methods led to dramatically decreased signal intensities of TatA (>10 % compared to directly spotted control; data not shown) and no TatA could be detected after using C18 reversed phase desalting columns and Bio-Beads possibly because of strong hydrophobic interaction of the protein with the column material. Extra purification steps to monitor rhomboid cleavage in a high-throughput inhibitor screening assay would also be expensive and time consuming and therefore reduced detergent concentration during *in vitro* cleavage of TatA was considered. In

order to find the optimal DDM concentration for direct MALDI-MS analysis, RNAse was used as a model protein to test the influence of different DDM concentrations on the MALDI analysis. RNAse A (~14 kDa) has a MW that is comparable to TatA (~11 kDa), but is less hydrophobic than TatA and therefore provides good signal intensities by MALDI-MS. RNAse was solved in rhomboid reaction buffer (50 mM HEPES; 2,5 µM RNAse) containing different concentrations of DDM and analyzed by MALDI (Figure 7).

Figure 7: Influence of detergent DDM on protein detection by MALDI-MS. MALDI-MS signal intensities of RNAse A (2,5 µM) was used as a model protein to monitor effect of the detergent DDM on MALDI-MS detection. The no-detergent control shows good detection of model protein. Signal intensities decrease with increased detergent concentration.

The study visualized the negative influence of DDM on the signal intensity of RNAse measured by MALDI-MS. The addition of 0,005 % of DDM decreased signal intensity by more than 70 % compared to the control sample without detergent. At a DDM concentration of 0,05 %, the concentration used in the *in vitro* rhomboid cleavage of TatA, only 2 % of the signal intensity of the no-detergent control could be achieved. To improve the crystallization process and therefore increase signal intensities of TatA to gain robust data for inhibitor screenings, DDM concentration had to be reduced during the cleavage reaction. To this end TatA was expressed and purified

using 0,0125% DDM, which is slightly above the CMC (CMC of DDM = 0,01 %) (le Maire et al, 2000).
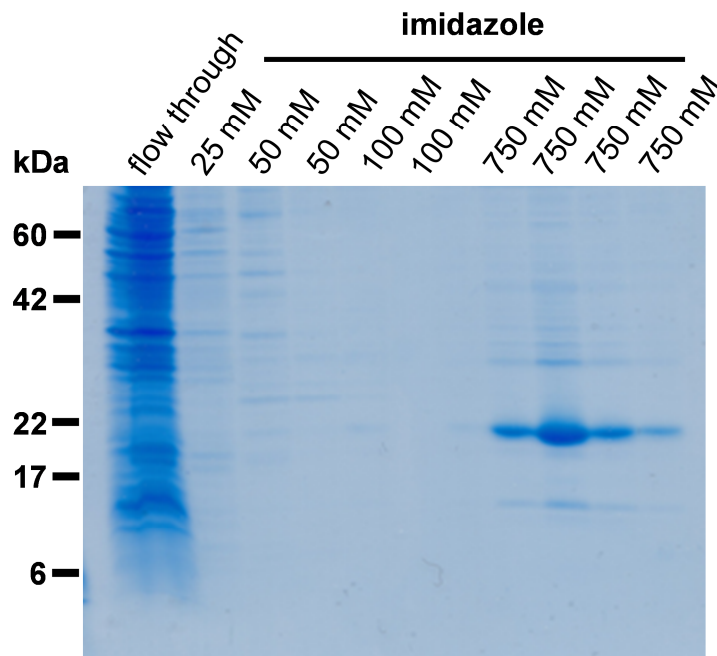
Figure 8: Expression and purification of TatA. Rhomboid substrate TatA was purified in buffer containing 0,0125% of the detergent DDM by the His6-tag using NiNTA-beads. Beads were washed using increasing imidazole concentrations (25-100 mM) and eluted with buffer containing 750 mM imidazole.

To test enzymatic activity of AarA in reduced detergent conditions, 10 µM of TatA expressed and purified with 0.0125 % DDM; Figure 8) was incubated with 0.5 µM AarA in HEPES buffer containing 0.0125 % DDM. Aliquots were taken every 20 minutes and analyzed by MALDI-MS (Figure 9).
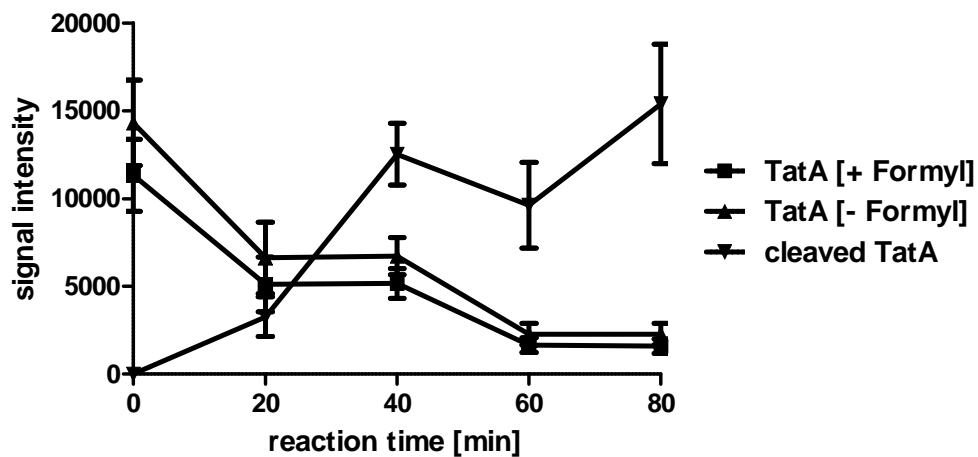
Figure 9: Cleavage of TatA by AarA in reduced detergent concentration analyzed by MALDI-MS. TatA (10 µM; expressed an purified with 0.0125 % DDM) was incubated with AarA (0.5 µM) in HEPES buffer containing 0.0125 % DDM. Every 20 min aliquots were taken from three independent reactions and substrate turnover was measured by MALDI-MS. MS Signal intensities of formylated [+Formyl] and deformylated [-Formyl] TatA as well as the N-terminal cleavage product (cleaved TatA) are shown. The cleavage product increases in a time dependent manner, while signal intensities of the substrates decrease over time.

AarA efficiently cleaves TatA in reduced detergent concentrations and TatA cleavage (decrease of substrate signal and increase of cleavage product signal intensities) was reliably measured.

This optimized sample preparation led to stable and robust signal intensities of the analytes and MALDI-MS analysis could be performed on 396 well plate format and measuring time was accelerated to 3-4 sec / sample. For each cleavage reaction 3 technical replicates were spotted and reactions were done at least as duplicate in order to rule out experimental variances.

2.2.2 Optimization of data analysis

Rhomboid cleavage can generally be monitored by using MALDI-MS as a detection method. But in order to determine rhomboid activity or inhibition thereof substrate cleavage needs to be quantified. MALDI-MS is optimal for fast and accurate determinations of molecular mass, but mass spectrometry-based quantitative

determinations are generally made by other techniques (Bantscheff et al, 2012). The reason for that is inconsistency of signal intensities and the shot-to-shot variability inherent to MALDI-MS compared to other MS methods like ESI-MS. It is generally possible to quantify peptide- or even protein samples by MALDI-MS but that often involves isotope-labeled internal standards (Bucknall et al, 2002), which are not always available or very expensive. Adding internal standards may also cover up with assay stability problems or ion suppression during MS analysis.

2.2.2.1 Usage of substrate / product ratio as measure of rhomboid activity

In order to overcome the limitation of weak shot to shot reproducibility of MALDI-MS, we aimed for a method to quantify rhomboid cleavage without making use of internal standards. Therefore we chose to use the ratio of signal intensities of substrate and product as a measure of rhomboid substrate cleavage. To do so, it is required to detect at least one of the product signals in one mass spectrum with the substrate signal in order to correctly calculate the substrate to product ratio. The calculated ratio can then be used for quantification of substrate cleavage without making use of internal standards. Unfortunately when measuring TatA incubated with AarA none of the tested matrices (see chapter 2.2.1.1) - being optimal for either peptidic or protein samples - enabled measurement of the N-terminal 8 amino acid cleavage product and the unprocessed TatA in the same MALDI-MS spectrum. We therefore used the ratio of the signal intensities from the intact substrate protein TatA and the C-terminal cleavage product as a read-out of substrate turnover. This label-free method for monitoring enzymatic activity has been reported for some enzymes including phosphatases, kinases and proteases by monitoring enzymatic turnover of small peptidic substrates (Greis et al, 2006). So far MS-based inhibition assays haven't been reported for intact and unlabeled protein substrates which optimal match the substrate specificity of their natural enzyme compared to artificial peptidic substrates.

In order to test the assay conditions TatA cleavage by AarA was monitored in time (Figure 10). The two occurring forms of TatA – with and without and N-terminal formyl group (see chapter 2.2.2.2) – were analyzed separately.
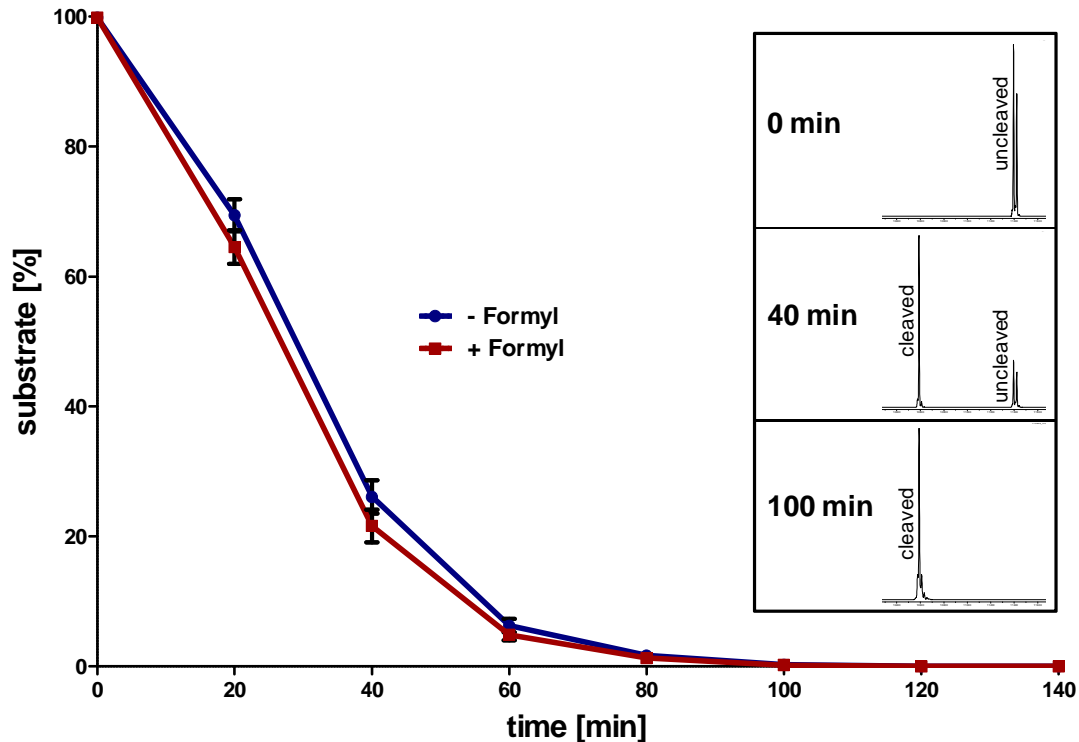
Figure 10: TatA cleacage by AarA: 10 µM TatA was incubated with 0.5 µM AarA in three independent experiments. Aliquots were taken every 20 min, immediately quenched and analyzed by MALDI-MS. Ratios of the N-terminal cleavage product to either the formylated (red) or the deformylated (blue) TatA was calculated as a measure of residual substrate.

Substrate turnover by AarA occurred in a time dependent manner (Figure 10). These results show that using the data analysis method is sufficient to monitor rhomboid substrate turnover and full substrate cleavage can be achieved within ~1.5 h.

2.2.2.2 Rhomboid cleavage of formylated TatA

Protein expression in bacterial cells and organelles (mitochondria, chloroplasts) is beginning with an initial *N*-formylmethionine, a derivate of the amino acid methionine with a formyl group linked to the amino group. The formyl group is then removed post-translationally by enzymes called deformylases. Overexpression of *P. stuartii* TatA in *E. coli* often led to incomplete deformylation of the initiator *N*-formylmethionine by endogenous *E. coli* peptide deformylase (Figure 11) and consequently two forms of this protein (formylated and deformylated TatA) were detected.
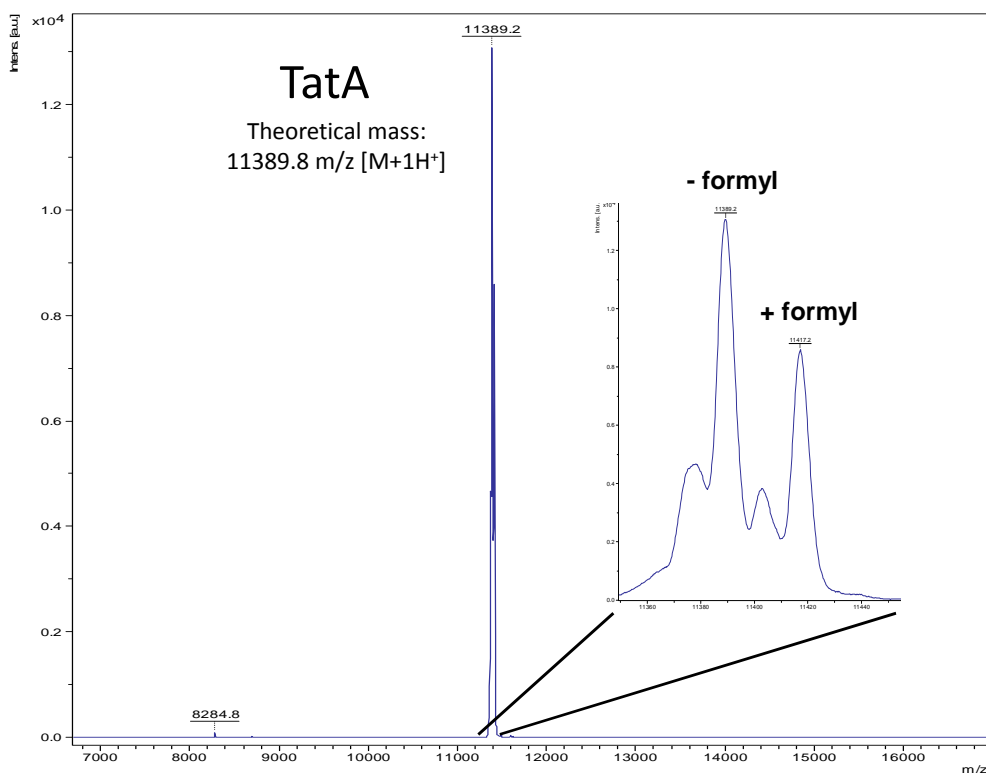
Figure 11: Incomplete deformylation of TaA during overexpression. MALDI-MS analysis shows that recombinant expression of TatA in *E. coli* leads to incomplete deformylation by endogenous *E. coli* peptide deformylase. The mass difference of 28 m/z resembles the N-terminal formyl group.

Incomplete deformylation is a known artifact for recombinantly produced proteins (Tang et al, 2004). The mass difference of these two TatA forms (~28 m/z) could easily be resolved during analysis with MALDI-MS (Figure 11). Both AarA and GlpG were able to cleave the two TatA forms at the natural cleavage site (Stevenson et al, 2007) and gave rise to a single product peak that corresponds to a loss of the first eight N-terminal amino acid residues of TatA (Figure 14). Another proof that the mass shift of the uncleaved TatA is due to N-terminal formylation is the finding that the formyl group is removed by *E. coli* peptide deformylase (see below). We intended to rule out preference of rhomboid protease for either the formylated or the unformylated substrate. Therefore signal intensities of the two substrate signals during AarA cleavage were compared (ratio of formylated TatA/deformylated TatA). Preference of rhomboid protease for one of the substrate forms due to the N-terminal formyl-group would lead to differences in the cleavage kinetics and consequently complicate data analysis.
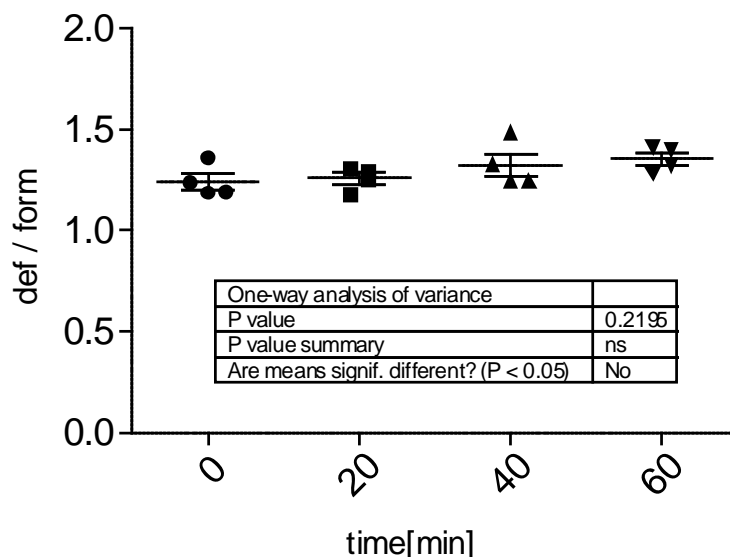
Figure 12: Comparison of rhomboid cleavage kinetics of formylated and deformylated TatA. The ratio (def/form) between the signal intensities of deformylated and formylated TatA (10 µM) over time was calculated from four independent cleavage reactions by rhomboid AarA (0,5 µM). One-way analysis of variance shows there is no significant difference in the ratio at different time points.

TatA cleavage by AarA, when followed over time, did not show significant preference for one of the substrate forms (Figure 12), which would result in significant differences in ratios calculated from MS signal intensities of the deformylated and formylated TatA. This shows that the N-terminal formyl group of the substrate does not influence rhomboid cleavage kinetics.

Even though the three proteins only marginally differ from each other (+/- formyl group; +/- 8 N-terminal amino acids), it is possible that these differences may lead to different behavior during MALDI-MS analysis. In order to correctly quantify TatA cleavage by using the ratio of signal intensities, it was necessary to determine the difference in ionizability between the two TatA forms and the C-terminal cleavage product. To this end *E. coli* peptide deformylase (PDF) was over-expressed in *E. coli*, purified via the His$_6$-tag and used to quantitatively *in vitro* deformylate a batch of fully formylated TatA. The two TatA forms – formylated and deformylated - were then

each mixed with the N-terminal cleavage product in a 1:1 molar ratio and subsequently analyzed by MALDI-MS (Figure 13).
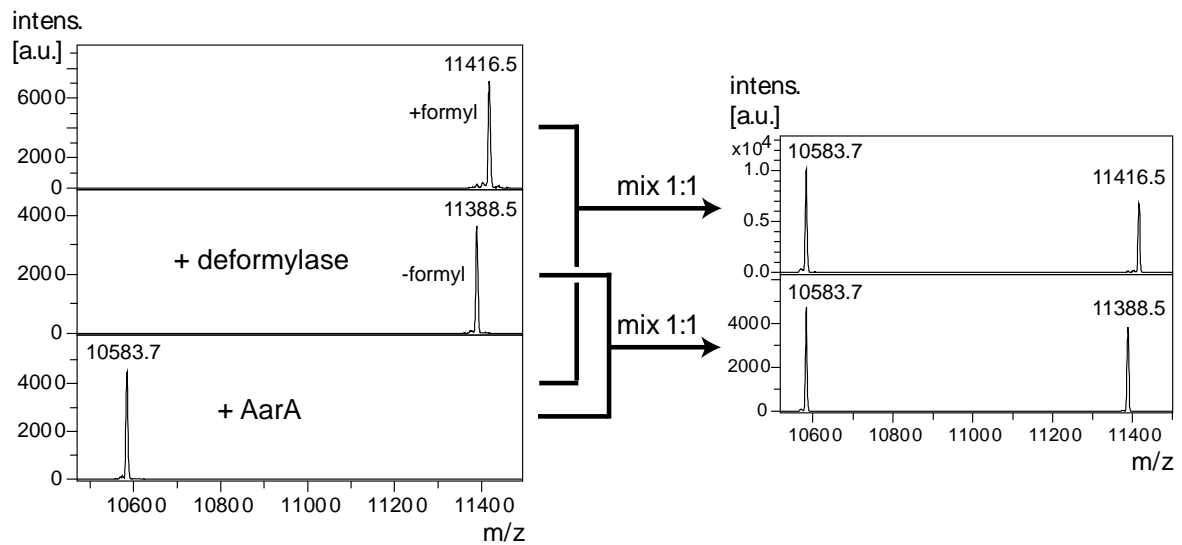
Figure 13: Determination of the ratio of ionizability of the uncleaved and cleaved TatA substrate species. Fully formylated TatA (87 µM) was *in vitro* deformylated using recombinant *E. coli* peptide deformylase (PDF; 10 µM) until full deformylation was achieved (2-3 h; tested by MALDI-MS analysis). TatA was processed further by adding AarA (5 µM) until quantitative cleavage was achieved (0.5–1 h; tested by MALDI-MS analysis). (Right) All TatA species were mixed with 2% of TFA to quench enzymatic activity of AarA and PDF. In four independent experiments, each formylated and deformylated TatA was mixed with fully cleaved TatA in a 1:1 molar ratio and analyzed directly by MALDI-MS. The ratio between signal intensities of the cleaved and the formylated or the deformylated TatA species was further used as ionization factors ($f_{formylated} = 1.56$ and $f_{deformylated} = 1.32$) to normalize all MALDI-MS measurements of TatA cleavage.

The signal intensities of the formylated and deformylated TatA in comparison to the C-terminal product were then used to determine ionization factors (Figure 13): $f_{formylated} = 1.56$ and $f_{deformylated} = 1.32$. These factors were used to normalize all measurements by using the indicated formula (Figure 14).

A

B

C Substrate cleavage $[\%] = \dfrac{I_{cleaved}}{I_{cleaved} + I_{formylated} \times f_1 + I_{cleaved} + I_{deformylated} \times f_2}$
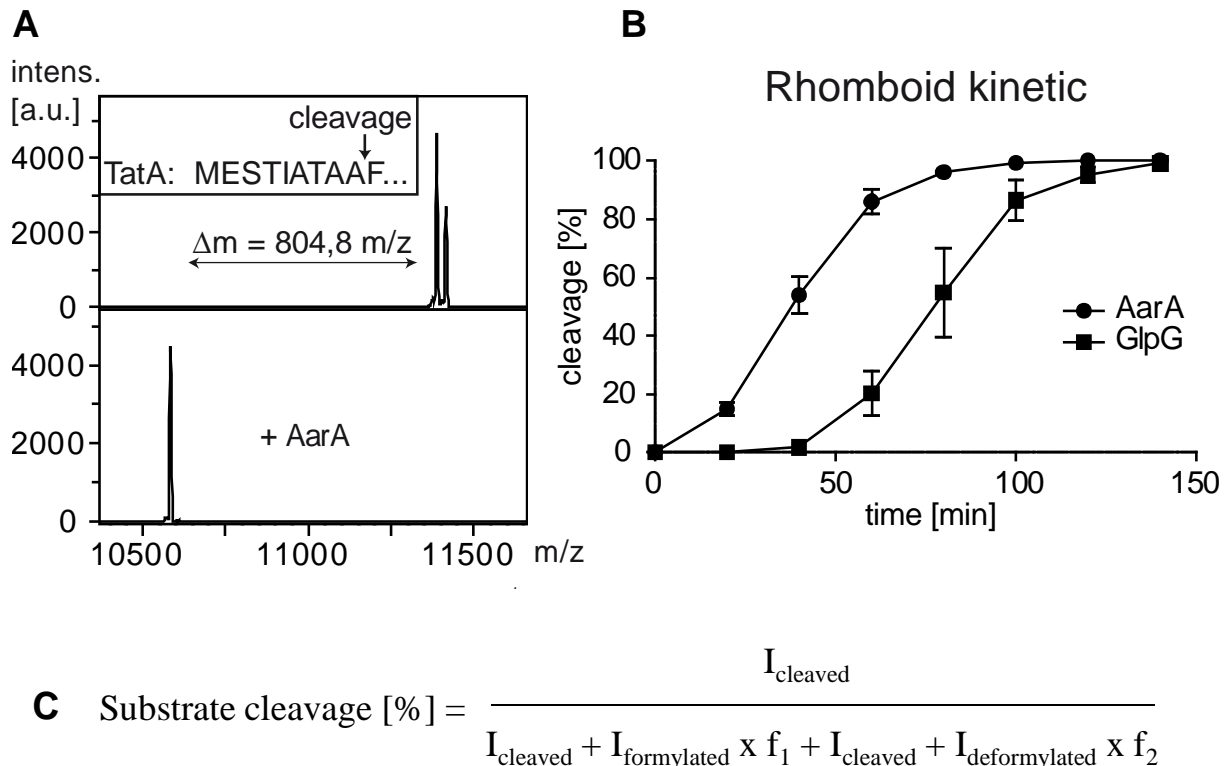
Figure 14: Rhomboid cleavage of TatA. (A) AarA cleaves the natural protein substrate TatA at the physiological cleavage site. 10 µM of TatA was incubated with 0,5 µM AarA for 2 h and monitored by MALDI-MS before and after cleavage. (B) TatA (10 µM) was incubated with rhomboid AarA (0,5 µM) and GlpG (1,5 µM) and time batches were taken every 20 minutes (0 – 140 min), quenched and directly analysed by MALDI-MS. (C) Substrate cleavage [%] was calculate by using the indicated formula with I = signal intensity and f = ionization factor (**Figure** 13).

AarA as well as GlpG mediated cleavage reaction increases slowly in the beginning, which is possibly due to the fact that the enzyme must first bind to the substrate before being able to react. Finally as the concentration of uncleaved substrate nears the concentration of the enzyme, the reaction slows over time until all substrate is cleaved. TatA cleavage by AarA was more efficient compared to GlpG (Figure 14), probably because TatA is the natural substrate of AarA and better matches its substrate selectivity.

2.2.2.3 Accelerated data analysis by using software macro

To accelerate the analysis of inhibitor screening data a Software macro was written. The used analysis software lacked a feature, which allows to search a large number

of spectra for specific analyte signals and automatically list it in a format that can be exported for further data analysis (e.g. to Excel). The developed macro listed defined MS-peaks (chosen by m/z value) with a defined threshold and – if desired, like in the label free inhibitor screening - calculates ratios thereof. The data output txt-file was then imported into GraphPad for further data analysis. This software macro dramatically accelerated data analysis and enabled analysis of hundreds of spectra in a matter of minutes.

2.2.2.4 Determination of Z' factor to estimate assay quality

To perform the rhomboid inhibitor screening, it was decided to use the MALDI-based method as an end point assay. The rhomboid was first treated with potential small molecule inhibitors and subsequently incubated with the substrate protein TatA. The reaction was quenched by adding 1 volume of 2 % TFA which immediately stopped enzymatic activity. All proteolytic reactions were stopped before 100% of cleavage was achieved, which allows to observe a decrease as well as an increase in substrate processing. The percentage of residual substrate after a certain incubation time was then used as a read-out of inhibition.

The Z'-factor is a statistical measure to judge the quality of an inhibitor screening assay ranking from Z'=1 for an ideal method to Z'=0, indicating an inapplicable assay (Zhang et al, 1999). The Z'-factor is defined by four parameters: means and standard deviations (SD) of both positive and negative controls. To judge the quality of the established rhomboid screening assay 8 replicates of positive and negative controls (AarA mutant and wild type, respectively) were used to determine the high Z' score of 0.82 which shows that this assay setup is sensitive and robust.

2.2.3 MALDI-MS based screening reveals new rhomboid inhibitors

A small, focused compound library was screened for inhibition of rhomboids AarA and GlpG. The library consisted of reactive electrophiles that are known to modify the active site of serine proteases (Powers et al, 2002): 4-chloro-isocoumarins (ICs), diphenyl phosphonates and peptido sulfonyl fluorides (Table 2). As a positive

inhibitor control DCI, known to inhibit rhomboid activity was used. Several compounds completely inhibited GlpG or AarA (Figure 15) and plotting results of a duplicate screen against each other (AarA: R^2 = 0.94; GlpG: R^2 = 0.97) showed good reproducibility of the data (Figure 15).
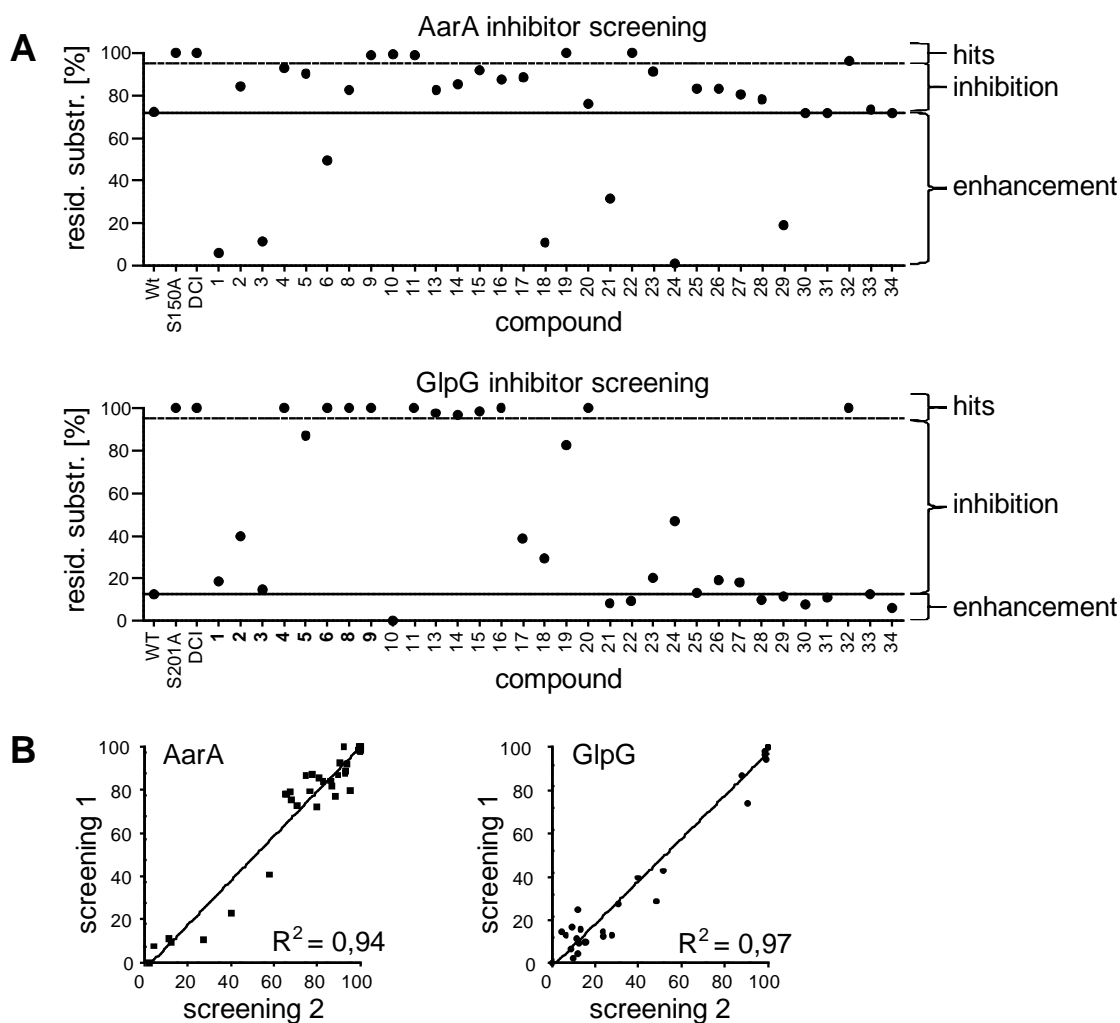
Figure 15: Inhibitor screening of rhomboids AarA and GlpG. (A) Purified recombinant rhomboid proteases AarA (0,5 µM) and GlpG (1,5 µM) were preincubated with the small molecules (200 µM) for 20 min. The reaction was started by adding the substrate protein TatA (10 µM). Samples were directly analyzed by MALDI-TOF. Compounds 7 and 12 repeatedly interfered with proper matrix crystallization (not shown). Compounds leading to increased residual substrate compared to Wt control (Wt_DMSO; solid line) were potential inhibitors, compounds leading to reduced residual substrate are potential enhancers of rhomboid activity. Molecules resulting in ≥ 95 % residual substrate (dashed line) were hit-inhibitors. (B) Screenings were done in duplicate and results were plotted against each other. The R^2 value was calculated to assess the reproducibility of the separate screenings.

For GlpG, the hits in the screening were mainly ICs and one sulfonyl fluoride. For AarA, ICs were the best inhibitors and one diphenyl phosphonate, which showed

weak inhibition of AarA. For a better quantification of inhibition, we determined the apparent IC50 values of the best hit compounds. IC **16** (structure see Table 2) displayed approximately an order of magnitude higher potency against GlpG compared to DCI (Table 1). In order to test selectivity of the inhibitors we additionally measured apparent IC_{50} values against bovine trypsin and chymotrypsin making use of fluorogenic substrates. These proteases are two representative examples of the largest family of serine proteases (S1 family). Although ICs **6** (structure see Table 2) and **16** - compared to rhomboid - showed good selectivity over trypsin, they readily inhibited chymotrypsin. This is probably because of the preference of chymotrypsin for hydrophobic residues (phenylalanine, tyrosine or tryptophan) in P1 position of the substrate. In comparison to trypsin, which prefers positively charged residues (lysine or arginine) the S1 pocket of chymotrypsin is hydrophobic and the hydrophobic substituents at position 3 of the IC inhibitors **6** and **16** can fit into that pocket.

Despite the screened library consisted of rather general serine protease inhibitors with no designed recognition element, some molecules inhibited AarA while displaying no inhibitory effect on GlpG, such as compounds **10** and **22** (structures see Table 2), while compound **20** (structure see Table 2) inhibited GlpG and had almost no effect on AarA even though they both cleave TatA. The differences in the inhibition profile could be explained by the low sequence similarities between these enzymes. Sequence alignment of AarA (P46116) and GlpG (P09391) showed only 12% sequence similarity and sequence alignments of all identified rhomboids showed only 6% sequence similarity within this class of intramembrane proteases (Urban, 2010).

```
1    ------------------------------------------------------------   0  P46116   AARA_PROST
1    MLMITSFANPRVAQAFVDYMATQGVILTIQQHNQSDVWLADESQAERVRAELARFLENPA   60  P09391   GLPG_ECOLI


1    -----------------MAEQQNPFSIKSKARFSLGAIALTLTLVLLNIAVYFYQIVFAS   43  P46116   AARA_PROST
61   DPRYLAASWQAGHTGSGLHYRRYPFFAALRE--RAGPVTWV--MMIACV------VVFIA  110  P09391   GLPG_ECOLI
                      :  :: **    :       * :: . ::: :      :** :


44   PLDSRESNLILFGANIYQLSLTGDWWRYPISMMLHSNGTHLAFNCLALFVIGICERAYG  103  P46116   AARA_PROST
111  MQILGDQEVMLWLAWPFDPTLKFEFWRYFTHALMHFSLMHILFNLLWWWYLGGAVEKRLG  170  P09391   GLPG_ECOLI
     :.:::*: *   :: :*. ::***     ::* .  *: ** *   : :* . *:  *


104  KFKLLAIYIISGIGAALFSAYWQYYEISNSDLWTDSTVYITIGVGASGAIMGIAAASVIY  163  P46116   AARA_PROST
171  SGKLIVITLISALLSG----YV---QQKFSGPWFGG-------------LSGVVYALMGY  210  P09391   GLPG_ECOLI
     .  **:.*  :**.: :.      *   :  . *  *   .        : *:.  * : *


164  LIKVVINKPNPHPVIQRRQKYQLYNLIAMIALTLINGLQSGVDNAAHIGGAIIGALISIA  223  P46116   AARA_PROST
211  VWL--RGERDPQSGIYLQRGLIIFALIWIV-AGWFDLFGMSMANGAHIAGLAVGLAMAFV  267  P09391   GLPG_ECOLI
      :      : :*:  *  ::   :: ** ::   ::  :  .: *.***.*  :*  :::.


224  YILVPHKLRVANLCITVIAASLLTMMIYLYSFSTNKHLLEEREFIYQEVYTELADANQ  281  P46116   AARA_PROST
268  DSLNARKRK------------------------------------------------  276  P09391   GLPG_ECOLI
       *   :*  :
```

Figure 16: Sequence similarity search of *P. stuartii* AarA and *E. coli* GlpG. Amino acid sequences of AarA (P46116) and GlpG (P09391) were compared using ClustalO (Uniprot) software. AarA and GlpG show only 11,7 % sequence identity. Predicted TMDs are highlighted in yellow, active site residues are highlighted in red.

Moreover GlpG structurally differs from AarA in consisting of the six core TMDs known for rhomboid proteases (Figure 2) and having an elongated N-terminus sticking out of the lipid bilayer, while AarA has an extra C-terminal TMD and a shorter N-terminus (Figure 16). This complicates the amino acid alignment of the whole proteins. Hence TMDs 4 and 6, harboring the active site residues in both, AarA and GlpG were compared. TMDs 6 (active site serine) showed 34 % sequence alignment while TMDs 4 (active site histidine) showed only 21 % sequence similarity. This suggests that there is enough structural variation around the active site in order to design selective rhomboid inhibitors even though they share similar substrate specificities like AarA and GlpG.

Besides inhibitors, also molecules were found that led to an enhancement of the TatA cleavage by GlpG or AarA and further investigations of these molecules are described in chapter 2.3.2.5.

2.2.3.1 Hit characterization

To further investigate the sulfonyl fluoride inhibitor **32** (structure see Table 2) which readily inhibited GlpG activity an analogue was synthesized containing an azide functional group (compound **35;** structure see Table 2; synthetic scheme: Vosyka et

al 2013). Compound **35** inhibited GlpG in the MALDI based assay (Figure 17 a) showing the slight changes did not influence the potency as inhibitor. Visualizing with an alkyne-TAMRA followed by SDS-PAGE showed that **35** indeed covalently reacts with purified GlpG as well as in crude lysate but also the inactive S201H mutant showing that these molecules do not act as mechanism-based inhibitors. Furthermore the inability of **32** to block labeling by ABP 11 shows it is not an active site modifier.
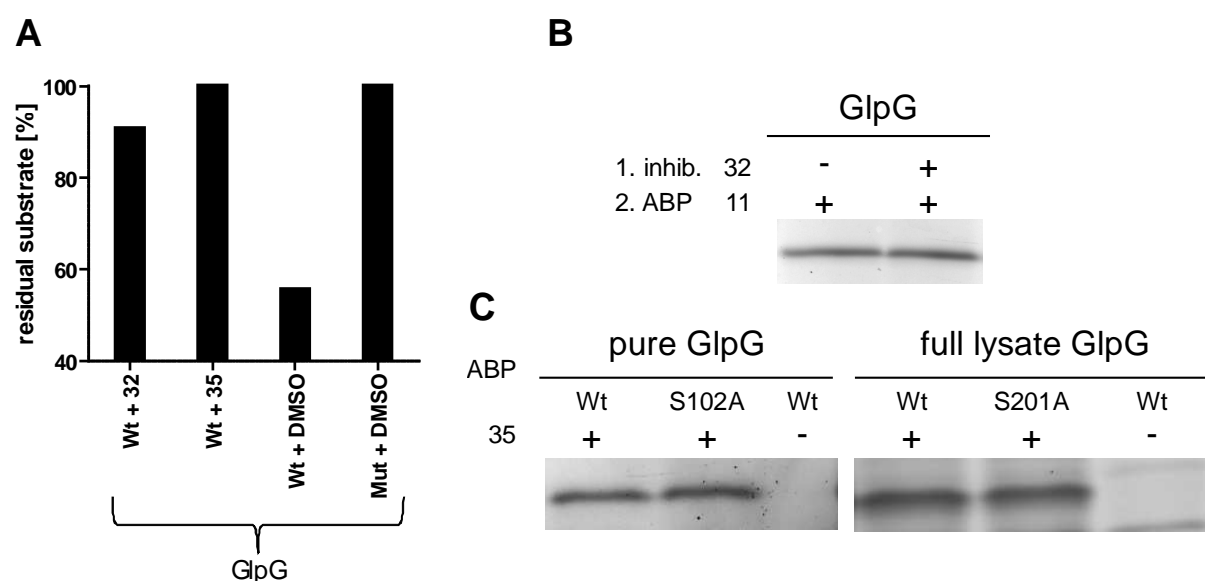
Figure 17: GlpG inhibition and labeling by sulfonyl fluorides. (A) Compound **32** and the azide coupled analogue **35** both inhibit TatA (10 µM) cleavage by rhomboid GlpG (1.5 µM) as determined by the MALDI-based screening. (B) Inhibitor **32** does not block labeling by ABP **11** and is therefore not an active site modifier. (C) Compound **35** covalently labels active GlpG (Wt) and inactive mutant (S102A) as detected by click chemistry based coupling of alkine-TAMRA in both, purified form an in *E. coli* lysates. No-probe control shows no unspecific sticking of alkine-TAMRA.

It is likely that these compounds are sulfonylating GlpG outside the active site and therefore interfering with the formation of an enzyme-substrate complex. This suggests that both, compounds **32** and **35** act as exosite inhibitors.

The diphenyl phosphonate **22** (structure see Table 2) showing weak inhibition of AarA did not act as mechanism-based inhibitor since labeling via the biotin functional group by Western-Blot analysis showed no labeling of the purified enzyme (data not shown). It therefore might act as an inhibitor that non-covalently interferes with substrate binding of AarA.

Isocoumarins are known to act as mechanism-based inhibitors of rhomboid proteases (DCI, JLK6) (Vinothkumar et al, 2010) and provided the most promising AarA and GlpG inhibitors. ICs are heterocyclic compounds that inhibit serine proteases (Powers et al, 2002) and serine hydrolases (Heynekamp et al, 2008). The first step of inhibition is a nucleophilic attack of the active site serine which results in the opening of the IC ring structure - resulting in a covalent bond – and therefore formation of an acyl-enzyme. IC inhibitors carrying an amino group at the 4 position are capable of forming a second covalent bond to the active site histidine as reported for soluble serine proteases. The chlorine (position 4) and an amino group (position 7) are important to increase the stability of the compound under physiological conditions (Powers et al, 1989). Positions 3 and 7 can be substituted in order to increase potency and selectivity towards certain serine proteases.

Table 1 Apparent IC_{50} values (µM) of rhomboid hit structures. The IC_{50} values of rhomboid (AarA, 0,5 µM; GlpG 1,5 nM) inhibition were determined using the MALDI-MS based assay. IC_{50} values of trypsin and chymotrypsin (each 5 nM) was determines using fluorogenic peptide substrates. Values arecalculated from triplicate experimentsand given ± standard error. (n.i. = no inhibition; n.d.: not determined)

Cmp	IC_{50} GlpG	IC_{50} AarA	IC_{50} trypsin	IC_{50} chymotrypsin
6	1.8 ± 0.46	n.i.	> 50	0.40±0.12
9	2.4 ± 0.70	29 ± 6.6	n.d.	n.d.
11	8.6 ±1.7	50 ± 17	6.1±2.2	0.024±0.009
16	0.74 ± 0.13	n.d.	> 50	0.11±0.02
19	n.d.	> 100	n.d.	n.d.
DCI	5.8 ± 2.8	33 ± 9.6	n.d.	n.d.

For both AarA and GlpG hydrophobic groups at position 3 of the isocoumarin structure appear to increase inhibitor potency while substituents in position 7 among the identified inhibitors seemed to be more variable. Therefore creating more hydrophobic substituents at position 3 during rational inhibitor design might result in more potent inhibitors with higher selectivity for specific rhomboids. The best GlpG inhibitor, IC **16** showed approximately tenfold increased potency compared to

previously reported DCI. In order to reveal the reason for the increased potency of IC **16** compared to other IC - inhibitors, the crystal structure of GlpG in complex with IC **16** was resolved.
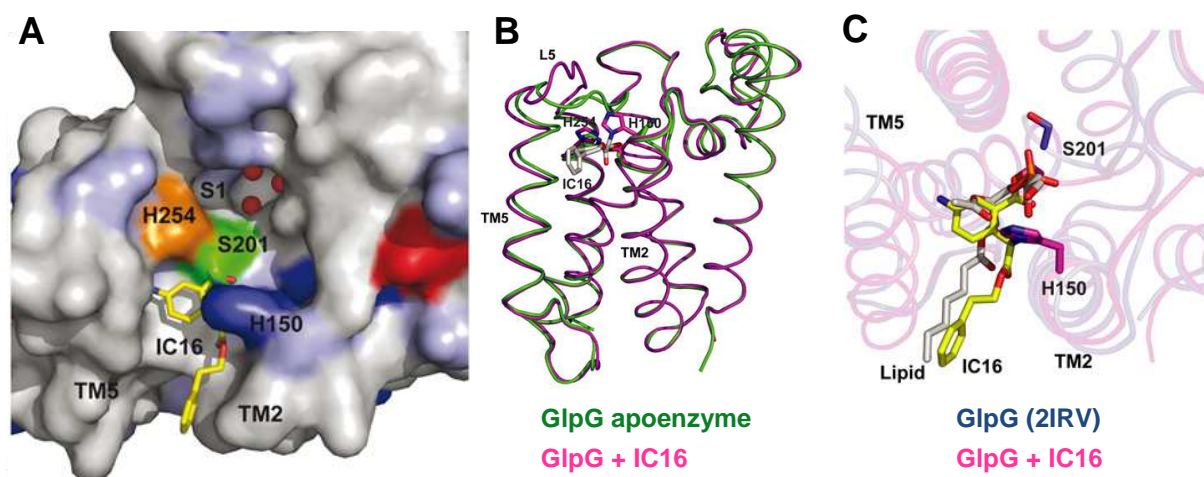

2.2.3.2 Crystal structure of GlpG in complex with IC 16

Crystal structures of enzymes in complex with inhibitors help understanding the mechanism of binding and therefore facilitate rational design of improved inhibitors. A crystal structure of GlpG in complex with the best hit inhibitor IC **16** (see chapter 2.2.3) has been resolved (Vosyka et al, 2013). As expected the open ring structure of the IC inhibitor covalently bound active site serine was observed (**Figure 18**). Unlike observed in the structure of GlpG in complex with another IC inhibitor JLK6 (Vinothkumar et al, 2010) the second nucleophilic attack does not occur at the active site histidine (H254) but at another histidine (H150) in close proximity. It is not clear yet what the function of histidine 150 is, but point mutation experiments revealed that it is crucial for GlpG proteolytic activity since it appears inactive when His150 is mutated to an alanine (Baker et al, 2007). His150 is highly conserved among rhomboids (Kateete et al, 2012) and is discussed as a possible candidate to stabilize the oxyanionhole of the active site (Baker & Urban, 2012; Baker et al, 2007). The hydrophobic group at the 3-position of IC **16** is in contact with the side chains of amino acids M149 and F153 which are located in TMD2. This hydrophobic interaction has also been observed in GlpG in complex with a phosphonate inhibitor (Xue & Ha, 2012). The hydrophobic interaction possibly forces the serine bound IC **16** more into the direction of H150 causing this difference in binding the active site of GlpG compared to IC JLK-6.

A cluster of hydrophobic amino acids are building up a hydrophobic cavity between TMD2 and TMD5. This hydrophobic cleft was postulated as the active site cavity where the lateral entrance of the substrate protein takes place (Ben-Shem et al, 2007). The recently reported crystal structures of GlpG in lipid environment (Ben-Shem et al, 2007; Bondar et al, 2009; Vinothkumar, 2011) shows a lipid molecule that takes a similar position like rhomboid inhibitor IC **16**. It was proposed that the position of the lipid molecule mimics rhomboid substrate binding. The higher potency

of inhibitors with hydrophobic side chains in the 3-position of the IC structure might be because of the need to displace a lipid (or detergent) molecule located in the active site cavity between TMD2 and TMD5 in order to efficiently bind the active site residues. Inhibitors, that are less hydrophobic might consequently fail breaking the hydrophobic interaction between the detergent molecule and the hydrophobic cleft (Figure 18) and are therefore sterically hindered to efficiently access the active site.

The structure of GlpG in complex with **16** looks very much alike the apoenzyme and as observed in other enzyme-inhibitor structures TMD5 slightly moves away from TMD2 and the L5 loop shifts upwards.

Figure 18: Structures of GlpG in complex with IC inhibitor 16. (A) Surface representation of GlpG viewed from the periplasm showing the bound IC inhibitor **16** (Vosyka et al, 2013). The protein molecule is color coded according to the biochemical nature of the amino acids: positive and negatively charged residues are shown in blue and red, polar residues in light blue, and the rest in gray. The active-site residues S201 and H254 are colored green and orange, respectively. Instead of covalently binding H254, like in other structures of GlpG in complex with an IC inhibitor (Vinothkumar et al, 2010) the second nucleophillic attack is performed by H150. Water molecules (red spheres) are found occluded in a cavity that has been postulated as the S1 substrate-binding site. The carbon atoms of the inhibitor are colored in yellow and shown in stick representation. (B) Comparison of the GlpG apoenzyme (Vinothkumar et al, 2010; PDB ID code 2XOV) and GlpG in complex with IC 16 (Vosyka et al, 2013; PDB ID code 3ZEB). The inhibitor molecule and key residues that interact with the inhibitor are shown in stick representation. The carbon atoms of IC 16 are colored in white and the amino acids in magenta. Major differences are observed in transmembrane (TM) helix 5, L5 and L1 (residues 128–135). (C) The position of the IC **16** strongly resembles the position of a lipid molecule observed previously in a structure of GlpG (Ben-Shem et al, 2007; PDB ID code 2IRV). However, in the IC 16 structure, the extent of change in TM5 and L5 is minimal compared with the structure with the lipid at the active site.

The MALDI-based screening revealed novel mechanism-based rhomboid inhibitors which are more potent than the previously reported isocoumarin inhibitors DCI and

JLK-6. Some of the IC inhibitors carry an alkyne functional group which allows using them as activity-based probes.


2.3 Activity-based probes for rhomboid proteases

2.3.1 Evaluation of ABPs for bacterial rhomboid proteases

Activity-based probes are molecules that, equipped with a detection tag can distinguish between active and inactive enzymes by binding only to the functional active site of an enzyme. Some of the identified IC inhibitors in the screen carry an alkyne functional group amenable to functionalization with a fluorophore by the Huisgen 1,3-bipolar cycloaddition, also known as click chemistry (Kolb et al, 2001; Speers et al, 2003). Such groups enable labeling active enzymes but not of the inactive counterpart by functionalizing these active site modifiers with a fluorophore or radiolabeled tag. Another application is to perform pull down experiments of activity-based labeld enzymes by clicking on a biotin or magnetic bead. These molecules are therefore classified as ABPs. Clickable ABPs are widely used in functional proteomics due to their flexible application *in vitro* and *in vivo* (Serim et al, 2012). Azide coupled fluorophore Tetramethyl-6-Carboxyrhodamine (N_3-TAMRA) was used to visualize ABP labeled rhomboid proteases followed by analysis via SDS-PAGE.

The ICs **6** and **11** (structures see Table 2), two ICs that efficiently inhibited GlpG in the MALDI-MS based screening carry an alkyne group. Both molecules labeled GlpG, but not the GlpG S201A active site mutant, either in detergent micelles or in full *E. coli* lysates, expressing recombinant GlpG (Figure 19). In addition, pretreatment of the enzyme with DCI or IC **16** blocks labeling. The same was observed for AarA and IC **11** (Figure 19). Hence, the alkynylated isocoumarins act as true ABPs, since their labeling of rhomboids is dependent on the activity state of the protease.
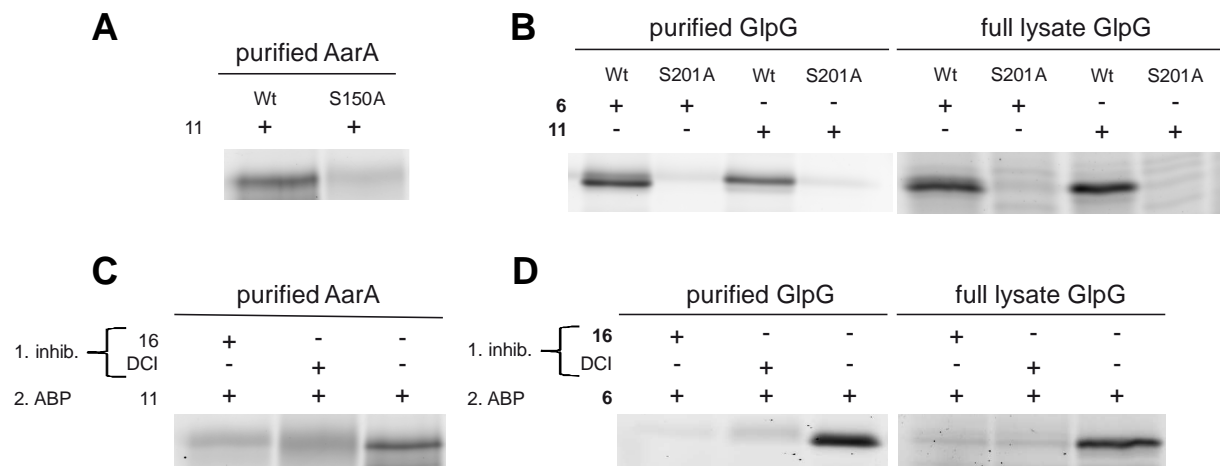
Figure 19: Activity-based labeling AarA and GlpG. (A) ABP **11** labels active (Wt) but not the active site mutant (S150A) of detergent solubilized rhomboid AarA. (B) The same is observed for ABPs **6** and **11** labeling of detergent solubilized GlpG as well as in lysates of *E. coli,* expressing GlpG. (C) Activity-based labeling of AarA is diminished, when enzymes were preblocked with active site inhibitors DCI and IC **16**. (D) The same is observed for GlpG both, as purified enzyme as well as in *E. coli* lysates expressing GlpG.

Activity-based labeling of rhomboid GlpG in crude, detergent free *E. coli* extracts is important since it monitors enzymatic activity of rhomboid in its natural lipid environment. Even though crystal structures of GlpG in detergent micelles strongly resemble structures of GlpG reconstituted in lipid bicelle environment (Vinothkumar, 2011) and therefore generally serves as a valuable tool to study rhomboid activity rhomboid activity strongly depends on the surrounding hydrophobic environment (Urban & Wolfe, 2005). This becomes obvious when labeling efficiencies of GlpG in E. coli lysates is directly compared to detergent solubilized, purified GlpG (Figure 20).
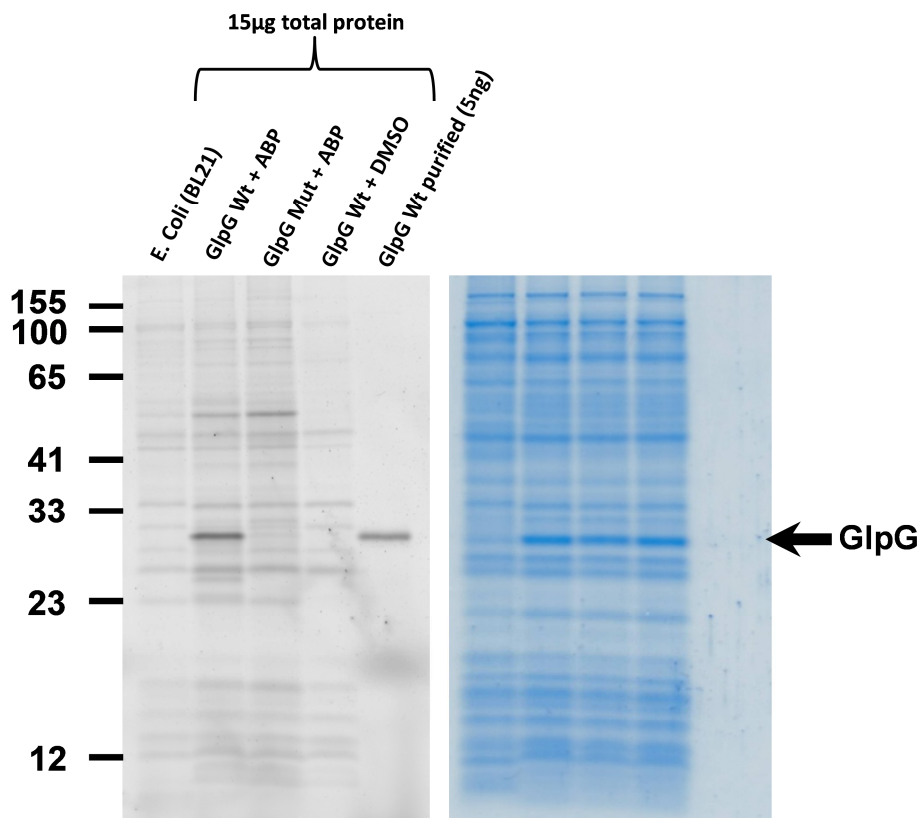
Figure 20: Labeling of GlpG with ABP 11. GlpG was detected in lysates of *E. coli* recombinantly expressing GlpG but not in inactive mutant control. DMSO control depicts unspecific click reaction and unspecific sticking of N_3-TAMRA to proteins. Labeling efficiencies of GlpG in crude *E. coli* lysate is compared to 5 ng of purified GlpG. Coomassie stain (right) shows equal protein loading.

Labeling of recombinantly expressed GlpG in lysate is several orders of magnitude less intense than labeling of 5 ng (below Coomassie detection limit; Figure 20) of detergent solubilized GlpG. This suggests that only a small fraction of GlpG in the lysate is active while it gains activity when being detergent solubilized. This is in accordance with reported results, where GlpG loses activity when reconstituted in *E. coli* lipid extracts (Urban & Wolfe, 2005). This mechanism of regulation by the surrounding lipid environment is further discussed in chapter 2.3.2.6.

In order to show the influence of the detergent DDM on rhomboid activity, GlpG in crude *E. coli* lysate was incubated with ABP **6** in the absence or presence of DDM (Figure 21). Compared to the no-detergent control, addition of 0,5 % DDM resulted in an approximately 3-fold increased enzyme labeling as estimated by band signal intensity.
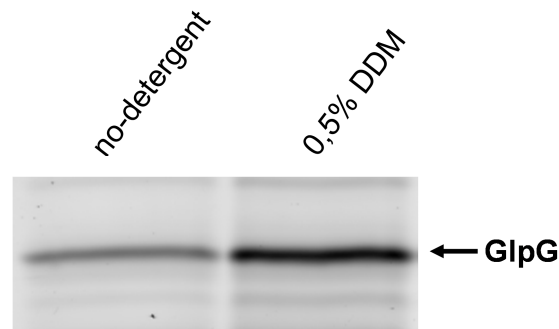
Figure 21: Detergent DDM influence on ABP labeling efficiency. Lysates of *E. coli* expressing GlpG were labeled with ABP **6** (no-detergent). Addition of 0,5% DDM increases labeling efficiency approximately threefold.

The reason for increased labeling efficiency remains unclear. One possibility is that the ABP can easier access the active site due to detergent molecules loosening up the GlpG structure. But since crystal structures of GlpG in detergent micelles strongly resembles the structure observed in lipid bicelles this is rather unlikely. Another possibility is that specific surfactant molecules (detergent or lipid), which take similar positions in both structures (Vinothkumar, 2011; Vinothkumar et al, 2010) directly influence rhomboid activity.

This suggests that GlpG might be in a rather inactive state when located in its physiological *E. coli* membrane environment. In eukaryotic cells functional membrane proteins are often located in membrane micro-domains - often referred to as lipid rafts - which are enriched in particular lipid species (Lingwood & Simons, 2010). The function of intramembrane enzymes strongly depends on the correct lipid composition in the surrounding environment. Disruption of these micro-domains can result in enzymatic dysfunction and cause diseases like Alzheimer's or Parkinson's in human (Michel & Bakovic, 2007). For intramembrane proteolysis it was shown that γ-secretase activity is directly regulated by the composition of its surrounding lipid environment (Osenkowski et al, 2008). Such micro-domains have also been discovered in bacteria (Lopez & Kolter, 2010). Since GlpG displays only weak activity, when reconstituted in lipid extracts from *E. coli* (Urban & Wolfe, 2005) the increased labeling efficiency when altering the hydrophobic environment by adding detergent suggests that GlpG might be regulated by its localization in specific lipid micro-domains as well. Another possibility of reduced rhomboid ABP-labeling might be that high concentrations of recombinantly expressed membrane proteins

disturbing the lipid micro-domains integrity and consequently leading to decreased rhomboid activities.

Although in crude lysates there is some labeling of other proteins (Figure 22), when compared to the DMSO control it becomes clear that most of the additional signals occur due to unspecific click reaction which is a known artifact when using alkyne azide cycloaddition (Speers & Cravatt, 2004). In order to reduce unspecific click reaction in the *E. coli* proteome, compound **36** was synthesized by pre-clicking a N_3-TAMRA fluorophore onto ABP **6** (Vosyka et al, 2013).
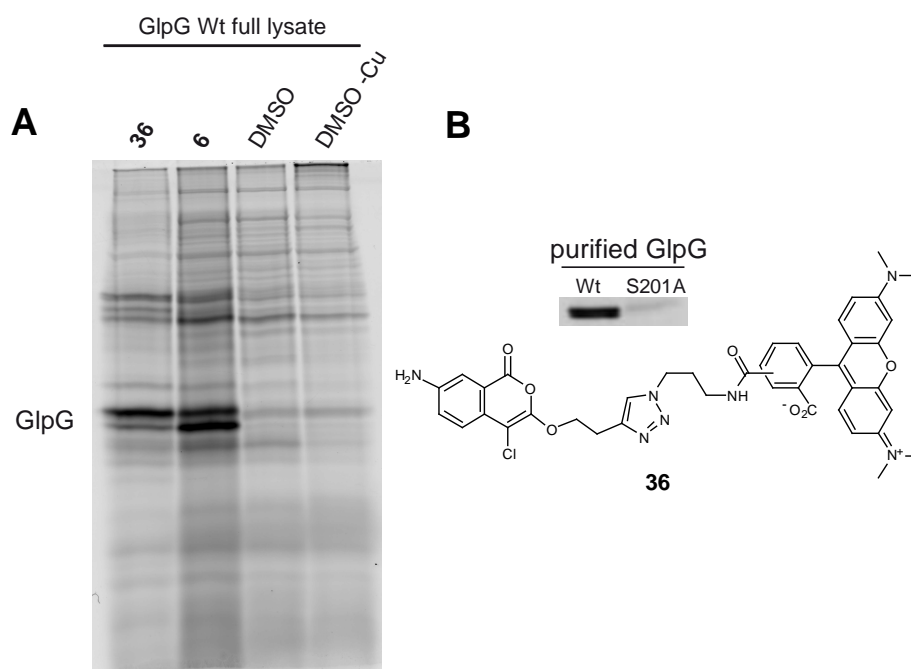

Figure 22. Comparison of ABPs 6 and 36. (A) Crude lysates from *E. coli* overexpressing GlpG were incubated with ABP **6** or **36** respectively. The DMSO control depicts the background occurring from the click reaction. Unspecific sticking of N3-TAMRA to protein is indicated in the fourth lane. Absence of catalyzing copper here disables click reaction. (B) Structure of **36**, which labels active GlpG (Wt) but not the active site mutant (S201A) and hence acts as an ABP.

This led to reduced background signals compared to labeling with ABP **6** followed by clicking on the fluorophore (Figure 22). Remaining background signals occurred mainly due to N_3-TAMRA nonselectively sticking to proteins (DMSO-Cu; no copper catalyst added). Additionally IC inhibitors are also known to inhibit other serine proteases (Table 1) (Powers et al, 2002) and hydrolases like esterases (Heynekamp et al, 2008) which probably leads to background labeling as well. Nevertheless, the

ABP-labeled rhomboid can easily be resolved on 1D gels and therefore the new rhomboid ABPs serve as a good tool for ABPP.

Interestingly, the labeling pattern of GlpG reveals the mechanism of inactivation by ICs. Theoretically IC inhibition comprises an initial attack by formation to the active site serine and subsequent forming of a covalent bond. Depending on the presence of a 7-amino group, a second attack by a histidine residue (Figure 23) occurs as reported for soluble serine proteases (Harper et al, 1985).
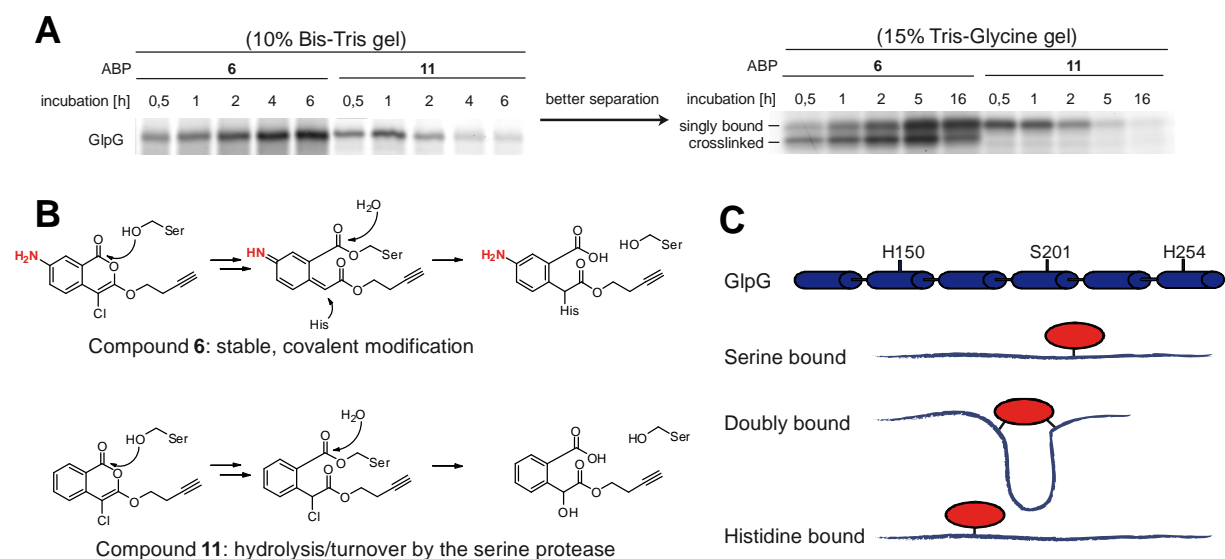
Figure 23: ABP labeling reveals binding mechanism. (A) ABP **6** shows robust labeling of GlpG in crude *E. coli* lysates while fluorescent signal of GlpG labeled with ABP **11** decreases over time (left). Better separation reveals two distinct forms of GlpG when labeled with ABP **6**, which is not observed in labeling with ABP **11**. (B) Chemical structures of ABP **6** and **11** binding to active site. (C) Schematic picture/explanation for double bond formation of GlpG labeled with ABP **6**.

Indeed, when GlpG labeling was followed over time, IC **6** showed a robust and permanent signal due to the stable carbon-nitrogen bond formed between the histidine residue and the 4-position of the IC (Figure 23A and B). In contrast, IC **11** binding showed a loss in signal over time. The latter is due to slow hydrolysis of the ester bond between the serine and the IC ring, which is in accordance with recent findings for DCI (Xue & Ha, 2012). With better gel resolution, labeling of GlpG by IC **6** shows two distinct protein species (Figure 23A). It is likely that the lower band is GlpG with IC **6** temporarily crosslinked to both serine and histidine side chains. This crosslinked species has a lower apparent MW due to a more compact protein structure (Figure 23C). The lower gel band gradually disappears in time, since the

ester bond to the active site serine is unstable. The IC **6** finally forms a single-bonded probe complex with the 4-position of the IC attached to the histidine residue. As expected, IC **11** only shows the upper (single-bonded) complex.

2.3.2 Functional analysis of rhomboid using small molecules

ABP are useful tools to analyze the biological function of the targeted enzyme. In order to use an ABP in such applications, it is necessary for the ABP to be used in a functional cellular assay that monitors the activity of the targeted enzyme at endogenous protein levels or in intact cells.

2.3.2.1 Endogenous GlpG

The natural function of GlpG still remains to be discovered and ABPs could serve as valuable tools to shed light onto that question. In order to use these new ABPs for functional studies of bacterial rhomboid it is necessary for them to not only label expressed but also endogenous levels of these enzymes. ABP **36** was therefore used to label endogenous GlpG in isolated *E. coli* membranes. The suitability of **36** to label endogenous GlpG was proven by directly comparing *E. coli* Wt and GlpG deletion mutants (Figure 24).
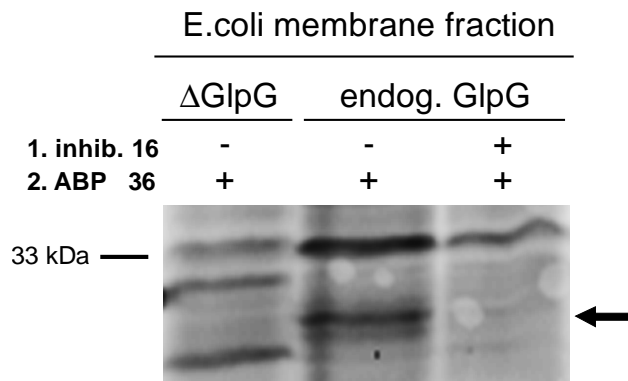
Figure 24: ABP labeling of endogenous GlpG. Membrane fractions of *E. coli* were treated with ABP **36** and analyzed by SDS-PAGE. Endogenous GlpG (MW: 31.3 kDa; black arrow) was detected in E. coli strain BL 21 (middle), but notd in a Δ*glpG* control cell strain. Labeling of endogenous GlpG was blocked upon treatment with active site inhibitor IC 16.

In *E. coli* strain BL21, expressing only endogenous GlpG a positive staining of GlpG could be observed, whereas in knock out conditions no signal was detectable. Additionally labeling of endogenous GlpG could be blocked by preincubation with inhibitor **16**. Therefore ABP **36** is sensitive enough to label endogenous protein levels of GlpG, which has the potential to monitor activity of unmodified rhomboid under physiological conditions. As described in chapter 2.3.1, surfactants like detergent or DMSO, which also increases enzymatic activity (Pierrat et al, 2011), influence rhomboid cleavage kinetics. Consequently the new ABPs could serve as a powerful tool in understanding the rhomboid enzymatic machinery and its biological function.

2.3.2.2 GlpG in vivo

Monitoring rhomboid activity in living cells could give insights into their natural function, which is not known for many rhomboids (e.g. GlpG). In order to use small molecule reporters, these tools need to get access to the enzyme. In *E. coli*, a gram negative bacterium, GlpG is located in the inner cell membrane. In order to label GlpG, ABPs consequently need to penetrate the outer cell membrane. The IC ABPs **6**, **11** and **36** were not able to penetrate the outer cell membrane when *E. coli* cells, recombinantly expressing GlpG, were directly incubated with the ABP and therefore failed labeling GlpG *in vivo*. Gram-negative bacteria possess an effective outer membrane permeability barrier that significantly reduces penetration by, and

therefore activity of antibiotics. Thus permeability of the *E. coli* outer membrane was increased by adding 1 M of the chelator EDTA to the media. EDTA destabilizes the outer membrane of gram-negative bacteria by chelating Ca^{2+} and Mg^{2+} cations that function as bridges between lipopolysaccharides of the microbial outer membrane. This results in permeabilization of the outer membrane commonly applied to gram negative bacteria (Bayer & Leive, 1977; Vaara, 1992) and used to increase susceptibility to antibacterial agents (Brown & Richards, 1965; Haque & Russell, 1974).
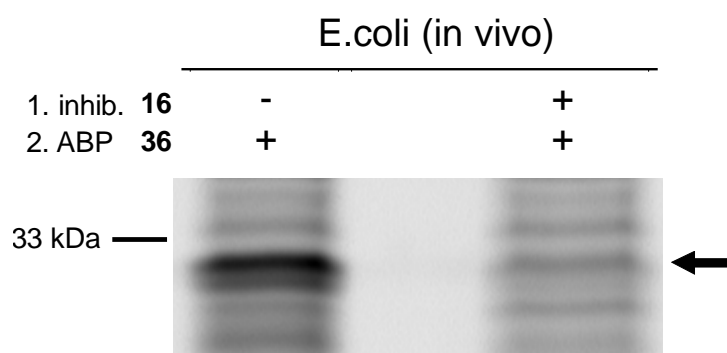
Figure 25: ABP labeling of GlpG *in vivo*. Intact *E. coli* cells recombinantly expressing GlpG were permeabilized with 1 M EDTA and incubated with 2 µM ABP **36** for 30 min. Cells were subsequently lysed by 1x sample buffer and analysed by SDS-PAGE. ABP labeling is blocked when cells were preincubated with GlpG inhibitor **16**.

This resulted in successful labeling of GlpG *in vivo* (Figure 25). Labeling could be blocked by pre-incubation of the intact cells with GlpG inhibitor **16**. This proves *in vivo* applicability of the identified ABP **36** as well as the IC inhibitor **16** in *E. coli* cells. As an alternative to EDTA permeabilized outer membranes, *E. coli* mutants with defective outer membranes might be used (Ruiz et al, 2005). Such mutant strains have already been used to prove *in vivo* applicability of rhomboid inhibitors (Pierrat et al, 2011). Additionally, the physiology of the mutant *E. coli* strain might diverge from the wild type strain and therefore lead to abnormal activities of endogenous rhomboid. Hence outer membrane permeabilization by EDTA in combination with the new ABPs is a good method for ABPP functional analysis of bacterial rhomboid under physiological conditions.

Cell permeability of ABPs is also required for applications like *in vivo* imaging studies in eukaryotic organisms. Up to date no ABPs for eukaryotic rhomboids are known.

Eukaryotic rhomboid has not been detergent-solubilized as an active enzyme, which complicates screening for selective ABPs. ABPs **6** and **11** failed to label mouse rhomboid RHBDL2 expressed in HEK-cells (data not shown). This was to be expected due to the weak sequence similarity. All together these ABPs represent a first step towards activity-based profiling of rhomboids in living organisms.

2.3.2.3 AarA inhibition in vivo

Not only ABPs but also inhibitors can provide information about enzymatic function. The rhomboid AarA enables the formation of the twin-arginine transporter, which is involved in intracellular signaling (quorum sensing) (Miller & Bassler, 2001). Quorum sensing in *P. stuartii* controls the expression of the enzyme acetyl-transferase, which modifies peptidoglycan and certain aminoglycoside antibiotics (Rather et al, 1997). Acetyl-transferase is therefore responsible for resistance of *P. stuartii* against this type of antibiotics, which is commonly used in hospitals. In order to enable quorum sensing, AarA cleaves TatA, which is expressed in a pre-form inactivated by a N-terminal extension that needs to be removed to enable assembly of the TAT-pore. Consequently AarA might serve as a target to reduce microbial growth during inflammation of the urinal tract.

AarA Knock out mutants of *P. stuartii* show a specific phenotype with (I) the inability to produce a yellow secrete, (II) absence of an putative extracellular signal that regulates cellular functions, (III) the inability to grow on MacConkey agar and (IV) they show growth in chains due to inability to separate after cell division (Kateete et al, 2012; Rather & Orosz, 1994). Despite the fact that AarA is involved in quorum sensing, the exact mechanism of these defects hat not been revealed yet.
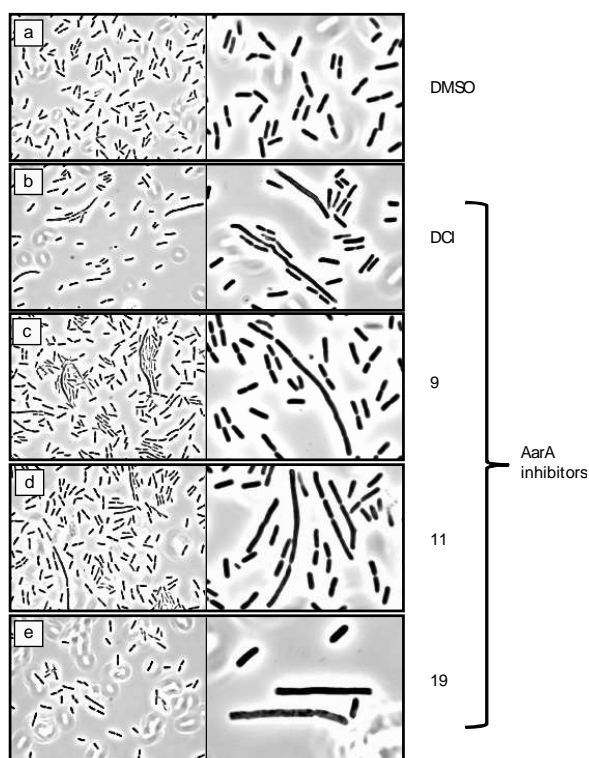
Figure 26: *In vivo* **inhibition of AarA in** ***Providencia stuartii***. *P. stuartii* cells (strain DSM4539) were incubated with the IC inhibitors DCI, **9**, **11** and **19** (b-e) or with the DMSO vehicle control (a) and subsequently analysed by microscopy (100x magnification). Treatment with the AarA inhibitors led to the same phenotype as the chain-forming Δ*aarA P. stuartii* mutant (Kateete et al, 2012).

To test *in vivo* applicability of AarA inhibitors identified in the MALDI-based screening (see chapter 2.2.3) wild type *P. stuartii* (strain DSM4539) with 100 μM inhibitor and subsequently analyzed by microscopy (100x magnification) (Figure 26). When incubated with AarA inhibitors *P. stuartii* cell showed the same cell growth phenotype as reported for the AarA knock out cell line. The cells growing in chains shows that the AarA inhibitors are capable to inhibit endogenous AarA in living cells and such molecules could possibly be used to disable quorum sensing of the human pathogen *P. stuartii*.

2.3.2.4 Evaluation of the effect of small molecules on rhomboid activity

The new ABPs can not only be used to verify that inhibitors are indeed binding the active site as it was done in chapter 2.3.1, but also monitor the effect of other molecules on rhomboid activity. Rhomboid proteases, both in crude *E. coli* lysates as

well as in detergent solubilized from are highly sensitive to external factors (e.g. detergent, DMSO) that influence their proteolytic activity.

2.3.2.5 Small molecule enhancers of rhomboid activity

The MALDI-based rhomboid inhibitor screening (see chapter 2.2.3) not only revealed novel rhomboid inhibitors but also identified molecules that enhanced rhomboid substrate cleavage. Enhancement of substrate cleavage by small molecules has been observed before (Pierrat et al, 2011) but the exact mechanism remains unclear. In order to verify enhancement of rhomboid activity, GlpG was incubated with ABP **36** in the presence of rhomboid enhancers **21** and **34**. ABP labeling kinetics were determined by SDS-PAGE followed by densitometry quantification of fluorescent band intensities and used as a measure for rhomboid enhancement.
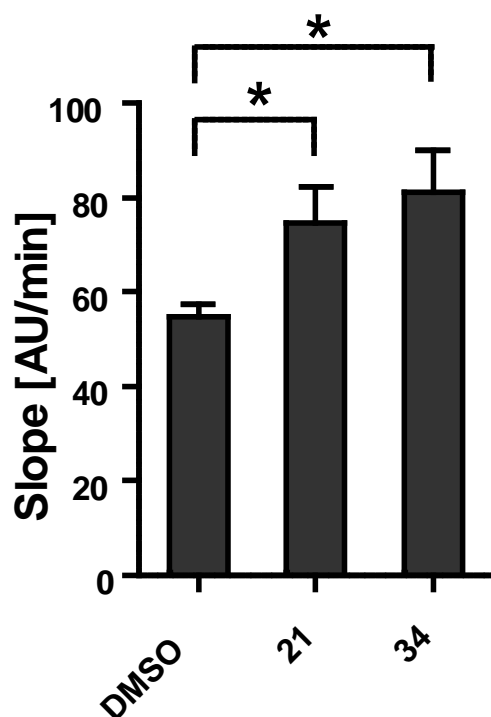

Figure 27: Rhomboid enhancement. Rhomboid GlpG (5 ng) was incubated with 2 µM of ABP **36** in the presence of rhomboid enhancers **21** and **34** (100 µM). Time batches (20, 40, 60 and 80 min) were analyzed by SDS-PAGE and labeling efficiency quantified by densitometry. Slopes of labeling kinetics were used as measure of rhomboid enhancement. Both, **21** and **34** significantly enhanced rhomboid labeling by ABP **36**.

Labeling of GlpG under the presence of these molecules enhanced labeling kinetics of 35-47% compared to a DMSO control (Figure 27). The increase of both substrate cleavage and probe labeling rules out a methodical artifact and suggests that the small molecule enhancers influence the first steps of catalysis, which are similar for both substrate cleavage and probe labeling. These steps consist of access to the active site followed by attack of the serine residue on the carbonyl group of the scissile bond or the IC. By binding the protein, the enhancers may either loosen the rhomboid structure, which increases the accessibility to the catalytic machinery which might be similar to the effect of increased detergent concentration (Figure 21), or influence the positioning of the active site residues leading to a more efficient catalytic dyad. Interestingly some of the molecules that enhanced rhomboid activity were structurally related to inhibitors, supporting the theory of creating a more efficient active site. Hence, these results give a first insight into the mechanism of how enhancers may influence the activity of rhomboids.

2.3.2.6 Detergent

Detergent solubilized rhomboid proteases are good models for studying rhomboid activity since when comparing GlpG structures in detergent and lipid environment, the detergent molecules take similar positions in binding to the protein. It was shown that lipids and detergents can have either an inhibiting or enhancing effect on rhomboid activity (see chapter 2.3.1) (Urban & Wolfe, 2005).

In order to analyze this effect under defined conditions and to evaluate the influence of the alcyl tail length of detergent on the activity state of rhomboids, GlpG was reacted with ABP **36** in the presence of maltoside detergents with increasing alkyl tail lengths, varying from eight to fourteen carbon atoms.
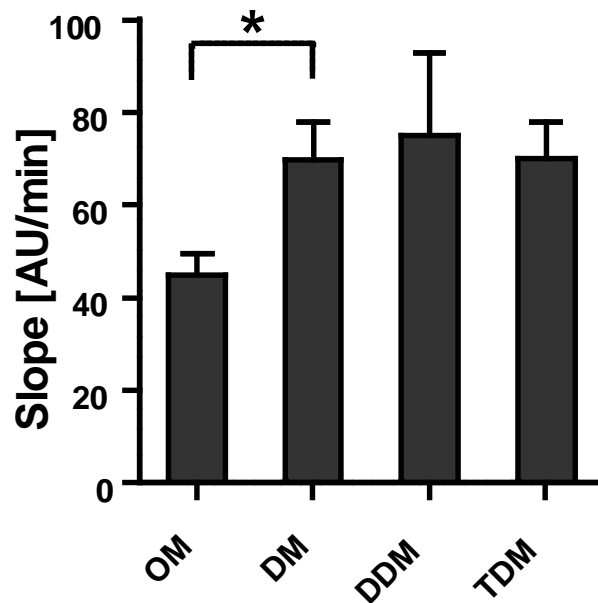
Figure 28: Labeling kinetics of GlpG in different detergent environments. Purified GlpG (5 ng) was incubated with **36** (2 µM) in the presence of maltoside detergents (5x CMC) with increasing alkyl chain length. At different reaction times, samples were analyzed by SDS-PAGE and the increase of fluorescent gel band intensity was determined. The slope of the linear part of the curve was calculated from duplicate experiments and is given in arbitrary units per minute ± standard error. Asterisk indicates a $P < 0.05$. (OM = octyl maltoside; DM = decyl maltoside; DDM = dodecyl maltoside; TDM = tetradecyl maltoside).

Determined labeling kinetics was equal in all detergent micelles except for octyl maltoside (OM), where the reaction rate was decreased by approximately 35% (Figure 28). An influence of detergent and lipid environments on rhomboid substrate cleavage has been reported (Sherratt et al, 2009; Urban & Wolfe, 2005). Here, we see that the alkyl chain length does not have a significant effect on the rhomboid activity, unless the chains are very short in comparison to the natural *E. coli* phospholipids. Overall, this supports the observation that mild, non-ionic detergents represent good environments for the reconstitution of bacterial rhomboids. These findings also contribute the hypothesis of rhomboid activity being directly regulated by its hydrophobic environment (see chapter 2.3.1) (Urban & Wolfe, 2005).

2.4 Rhomboid kinetics determined by mass spectrometry

MS-based methods are not limited to standard proteomics applications and inhibitor screenings but, can also be used to determine enzyme kinetics. Enzyme kinetics describe how fast an enzyme catalyzed reaction proceeds. The proteolytic substrate cleavage represents a first order reaction whose velocity can be described as the disappearance of the intact substrate over time or appearance of the cleavage product over time. When plotting substrate and product concentrations against time, the reaction velocity is represented by the slope. The reaction velocity is directly proportional to the concentration of the substrate and is described by the first order rate equation (v = k x [substrate]) in which k represents the first order rate constant. Since the constant does not change, the velocity begins to decrease as the concentration of substrate decreases when the reaction proceeds. For studying enzyme kinetics of proteases, fluorogenic substrates are commonly used in which the reaction velocity is defined by the slope of increasing fluorescence intensity of the generated product. In comparison to this, the advantage of using mass spectrometry as detection method to determine enzyme kinetics is that both, the substrate as well as the products can be monitored, which for example enables monitoring the activity of several enzymes simultaneously. Additionally such studies can be used to directly compare cleavage efficiencies of an enzyme towards different substrates. This can for example be used to determine the substrate specificity of a protease by altering single amino acids close to the scissile bond and monitor the effect on cleavage efficiency. In order to study enzyme-substrate interaction, cleavage kinetics of substrates with different length can be determined. This is interesting since a gel based study showed that rhomboid requires the intact substrate TMD for efficient cleavage (Pierrat et al, 2011) but the exact kinetic behind that finding remains unclear. MS-based methods to determine the cleavage kinetic e.g. of truncated rhomboid substrate proteins would help identifying recognition elements necessary for efficient cleavage.

Enzyme kinetics can also be used to characterize non covalent inhibitors of an enzyme. By determining Michaelis-Menten-kinetics a competitive inhibitor can be distinguished from a non-competitive by comparing K_M and v_{max}. MS-based enzyme kinetic studies can principally be applied to all enzymatic reactions that lead to a

mass shift upon enzymatic turnover. The enzymatic reaction could therefore either be analyzed in an on-line MS assay, which monitors substrate cleavage continuously or in time batches. MS-based enzyme kinetics are especially useful in case no artificial substrate for a particular enzyme is available, that enables monitoring of enzymatic activity e.g. by fluorescence or to figure out if artificial substrates display different K_M values than the natural substrate because of bulky and hydrophobic fluorophores that affect recognition by the enzyme. The latter is not the case when MS is used as a detection method because the untagged substrate molecules can be monitored.

Overall MS is a suitable method to detect enzymatic substrate turnover that has the advantages of simultaneously monitoring both substrate and product without the need of a labeled substrate and can, in case of proteases principally be applied to intact substrate proteins.

2.4.1 Assay setup on-line ESI-MS

In order to determine rhomboid cleavage kinetics as electrospray ionization mass spectrometry (ESI-MS) based assay that monitors rhomboid substrate cleavage in real time was considered. The use of ESI-MS enables the direct and simultaneous monitoring of both substrate and product molecules in real time. Real time analysis of enzymatic reactions has been done for a variety of enzymes such as glycosidases and chitinases (Greis, 2007; Letzel, 2008). Conventional methods to monitor protease activity like fluorogenic applications mostly make use of peptidic model substrates. Therefore the enzyme and the substrate are mixed and directly injected into the ion source and analyzed by ESI-MS. In the specific case of proteases, such studies have so far only been done using peptidic substrates (Scheerle et al, 2012; Lee et al, 1989). The goal of this study was to apply real-time ESI-MS to measure rhomboid cleavage kinetics and inhibition thereof using the recombinantly expressed and detergent solubilized rhomboid substrate TatA. The great challenge of analyzing intact membrane proteins like TatA is their hydrophobic nature. Prior to standard ESI-MS analysis membrane proteins are normally acetone precipitated in order to remove salts and detergent. Membrane proteins are then solved in solutions of methanol or acetonitrile and formic acid (Eichacker et al, 2004; Ploscher et al, 2009) for direct

injection. This is not possible in ESI-MS based real time assays, which require the direct injection of the reaction mixture into the mass analyzer. Adding solvents and acids to improve the ionization process are likely to decrease or completely quench proteolytic activity (Scheerle et al, 2012). Also the detergent molecules (DDM) used in the rhomboid cleavage assay can't be removed prior to analysis and therefore contaminate the sample.

The detergent used in this study is n-dodecyl-β-d-maltoside (DDM). Multiple studies have shown that DDM is a gentle, nonionic detergent that is often able to preserve protein activity of solubilized membrane proteins and is commonly used for rhomboid protease cleavage assays (Moller & le Maire, 1993; Seddon et al, 2004). DDM is also suitable for mass spectrometry (see chapter 2.2.1.2) and may therefore be used for real time ESI analysis of membrane proteins.

The HEPES buffer system, commonly used for rhomboid *in vitro* cleavage assays is incompatible with analysis by ESI-MS, due to the high risk of contamination of the mass spectrometric inlet with salt crystals. Therefore the activity of AarA in an ESI-MS compatible ammonium acetate buffer was tested and compared to HEPES buffer conditions using a gel-based assay. To this end a TatA variant N-terminally extended by four Serine Glycine repeats (to increase the difference in molecular weight between the processed and the uncleaved substrate) and C-terminally extended by three methionine residues was *in vitro* synthesized from mRNA in the presence of S^{35} methionine as described elsewhere (Stevenson et al, 2007). The radiolabeled TatA was incubated with different amounts of AarA in 50 mM HEPES or 10 mM ammonium acetate buffer respectively and separated on a BisTris gel (Figure 29).
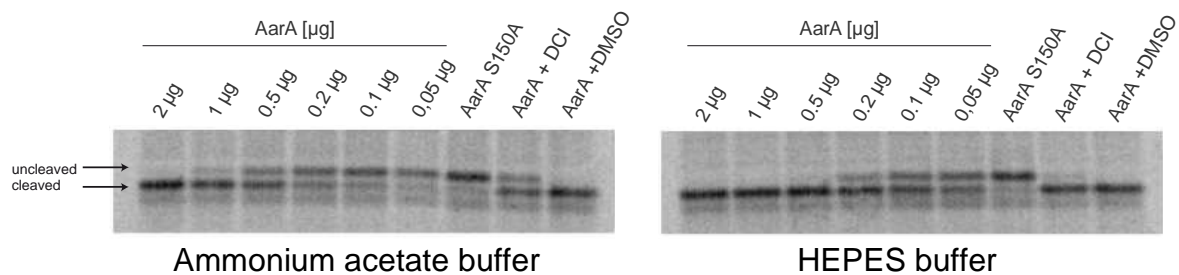
Figure 29: Cleavage of TatA by AarA under different buffer conditions. Radiolabeled (S^{35} [Met]) (SG)$_4$-TatA (Stevenson et al, 2006) variant was incubated with different amounts of AarA (2 - 0,05µg) in MS compatible 10 mM ammonium acetate buffer (AmAc) (left; pH 7,5) or 50 mM HEPES (right; pH 7,5) for 30 min at 37°C and analyzed by SDS-PAGE. No cleavage occurred when incubated with inactive AarA$_{S150A}$ mutant (AarA SA) both in AmAc or HEPES buffer condition. Preincubation with rhomboid inhibitor DCI (AarA+DCI) partly inhibited cleavage in AmAc- but had no effect on AarA in HEPES-buffer. DMSO (AarA+DMSO) had no effect on AarA substrate cleavage.

The disappearing band corresponds to the uncleaved substrate and clearly shows that AarA is cleaving the substrate in ammonium acetate buffer environment in a concentration dependent manner. Comparison of cleavage efficiency in these two buffer conditions shows that AarA in ammonium acetate is not as efficient as in HEPES buffer. Full cleavage was achieved with 0,5 µg of AarA in HEPES buffer, while 2 µg were necessary in ammonium acetate buffer to quantitatively cleave the substrate. The active site mutant AarA did not show any substrate cleavage and proteolysis could only partly be inhibited with 50 µM of the rhomboid inhibitor DCI in ammonium acetate buffer. In HEPES buffer conditions DCI had no effect on substrate cleavage probably due to increased activity compared to ammonium acetate buffer conditions. Therefore it can be concluded that AarA is active in both HEPES and ammonium acetate buffer conditions but displays reduced activity in the latter. Nevertheless ammonium acetate buffer is applicable for both enzymatic reactions as well as direct ESI-MS analysis and will be used in the following applications.

2.4.2 Proof of principle: trypsin as model enzyme for on-line ESI-MS assay

As a proof of principle and to establish the experimental setup of a real time substrate cleavage measurement by ESI-MS, trypsin was used as a model enzyme. A fluorogenic substrate, a three amino acid peptide coupled to an

amminomethylcoumarin (GGR-AMC) which becomes fluorescent upon cleavage was used to compare the two methods, the commonly used fluorescence-based assay with real-time ESI. To test detectability of the substrate and the cleavage products by mass spectrometry GGR-AMC was incubated with trypsin, quenched and analyzed by direct injection into ESI-MS.
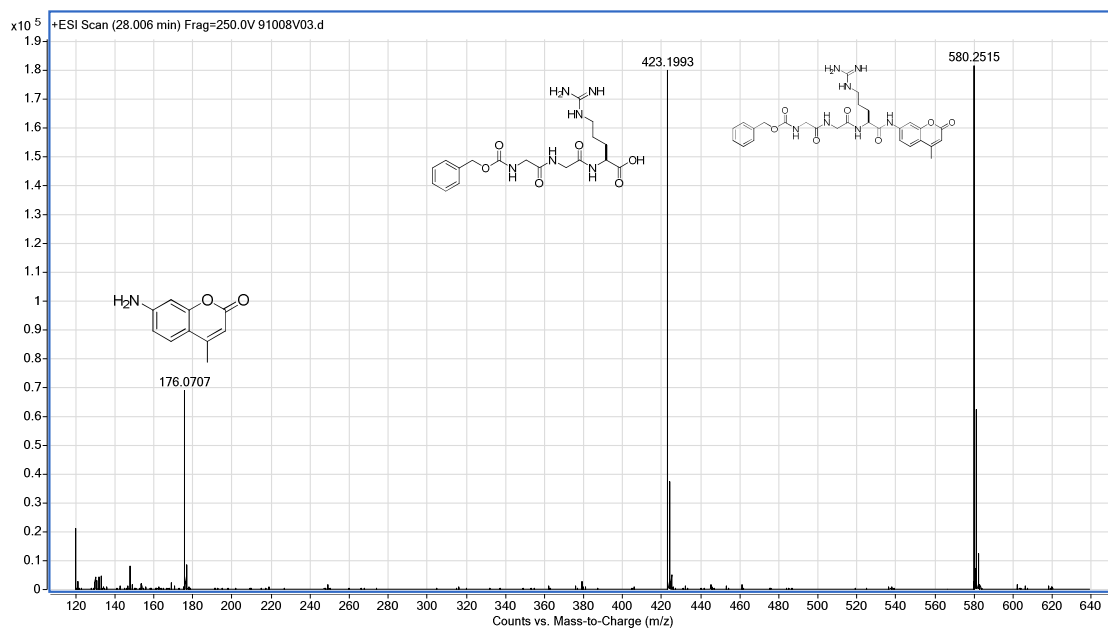

Figure 30: Tryptic cleavage of fluorogenic peptide monitored by ESI-MS. Fluorogenic trypsin substrate GGR-AMC (20 µM) was incubated with trypsin (25 nM) for 5 min, quenched by mixing 1:1 (v/v) with 50 % ACN / 0,1 %FA and analyzed by direct injection into ESI-MS. The uncleaved substrate (MW: 579,2 Da) as well as the cleavage products (MW tripeptide: 422,2 Da; MW AMC: 175,1 Da) was detected.

Both, the uncleaved substrate and the cleavage products could be readily detected by ESI-MS (Figure 30) and can consequently be used as a model substrate for establishing the real-time ESI-MS assay.

To obtain kinetic data of proteolytic cleavage by trypsin it was necessary to determine the stability of trypsin in the used buffer system. Trypsin is known to loose activity relatively fast by self-digestion and non-optimal buffer conditions and therefore estimating the enzymatic stability was necessary. To this end trypsin was incubated in reaction buffer at 4 °C or 37 °C and proteolytic activity was tested at different

timepoints by measuring enzymatic activity (V_{max}) by using the GGR-AMC fluorogenic substrate.
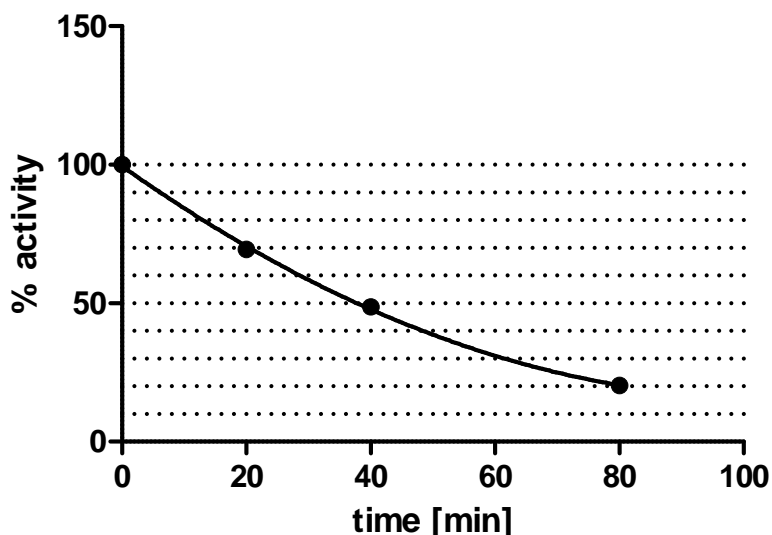
Figure 31: Stability of trypsin. 25 nM of trypsin was incubated in 10 mM ammonium acetate buffer for 0-80 min at room temerature. Aliquots were taken every 20 min and activity was determined by calculating the slope of fluorogenic substrate (GGR-AMC) cleavage over time. Slopes were plotted against incubation time to show loss of trypsin activity over time.

The results show that trypsin is losing about 25 % of activity within 20 min at 37 °C while trypsin incubated at 4 °C lost less than 5 % of activity. This suggests that trypsin is inactivated due to self-digestion. The ESI-MS measurement normally takes 20-30 min in order to gain a measurable increase in product signals, which might lead to errors in obtained kinetic data due to trypsin self-digestion and therefore inactivation of enzymatic activity. Fluorescence-based methods normally need only 3-5 mins and consequently are better suited for kinetic measurement of self-inactivating enzymes. But this effect might be overrated due to the fact that trypsin was incubated in the absence of substrate molecules. The presence of "alternative" substrate – other than trypsin itself - might decrease the rate of losing enzymatic activity due to self-digestion.

A first try to analyze protease substrate cleavage by ESI-MS in real time was done by incubating trypsin and the GGR-AMC in a 100 µl Hamilton syringe. The mixture was

subsequently injected into the mass spectrometer. This approach showed weak ionization of the analyte molecules and weak reproducibility (data not shown). In order to improve the MS analysis, a second syringe containing an acetotirile-formic acid mixture was placed into the syringe pump. Both streams were connected to each other via a T-piece and directly injected into the mass analyzer (Figure 32).
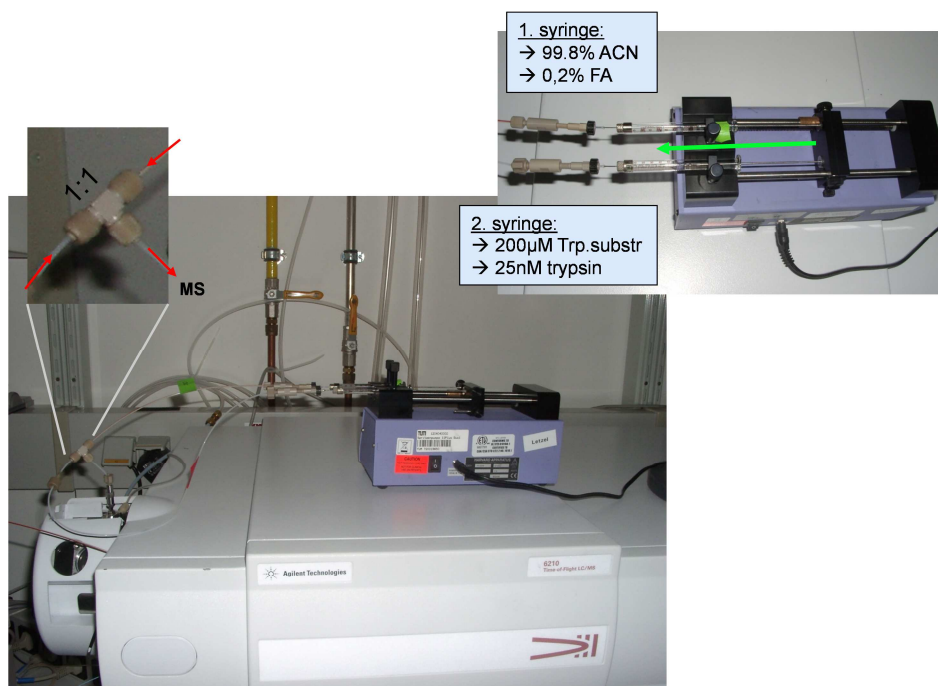
Figure 32: Experimental setup of continuous flow ESI-MS. (upper right) Two syringes were filled with (1; 500 µl Hamilton syringe) 0,2% formic acid (FA) in acetonitrile (ACN) or (1; 100 µl Hamilton syringe) reaction mixture. Syringe pump flowrate was 5 µl / min (for 100 µl syringe). The two syringes were connected by a T-piece (upper left) and mixture was directly injected into an ESI-MS mass analyzer.

This had several advantages: (I) The actetonitrile improved the electron spray process while the formic acid delivered protons for ionization of analytes and therefore increased signal intensities. (II) The higher flow rate (syringe containing enzyme-substrate mixture + syringe containing ACN/FA) entering the mass analyzer led to a more stable signal and (III) the formic acid quenched the proteolytic activity and therefore led to a precise stop of the reaction. The measurement was started immediately after mixing the enzyme and the substrate. This led to a loss of data during the first 5-8 minutes due to the time needed to set up the experiment (Figure 33).
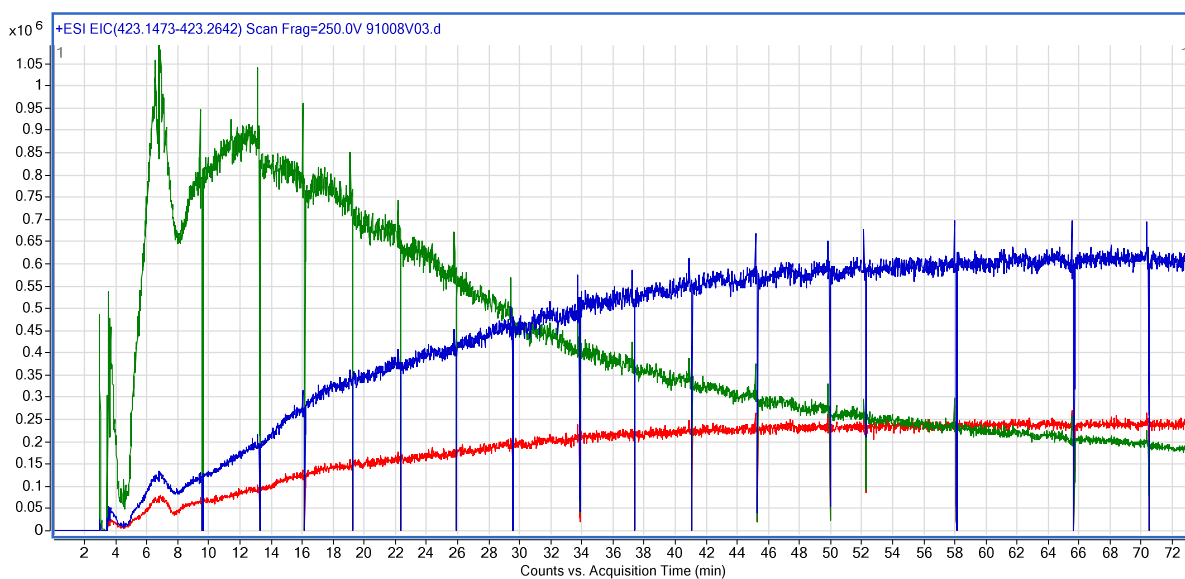
Figure 33: Tryptic hydrolysis of fluorogenic peptide substrate monitored by continuous flow ESI-MS. 20 µM of fluorogenic substrate GGR-AMC was incubated with 25 nM of trypsin at RT and monitored by real-time ESI-MS. Decrease of substrate (green) and increase of both cleavage products (blue = tripeptide; red = AMC) could be monitored over time.

The loss of data in the first minutes was disadvantageous since the slopes of increasing product as well as the slope of decreasing substrate is ideally calculated within the first minutes. Here the self-inactivation of trypsin would be at its lowest level. The velocity of the reaction begins to decrease after 10-15 min as the concentration of the substrate diminishes and possibly because of self-inactivation of trypsin. The velocity of the reaction eventually slows to zero before 100 % of substrate turnover was achieved. This is probably due to complete inactivation of trypsin. Nevertheless slopes of increasing product signals could be calculated, as soon as a stable injection flow was achieved and could therefore be used to determine enzyme kinetics of tryptic digest by ESI-MS detection.

In order to directly compare real-time ESI-MS assay with the commonly used fluorescence based assay, trypsin was incubated with increasing concentrations of the GGR-AMC substrate. Slopes of signal intensities (both, MS and fluorescence) of the fluorescent AMC cleavage product were determined and used to calculate K_M values.

A Mass Spectrometry   **B** Fluorescence
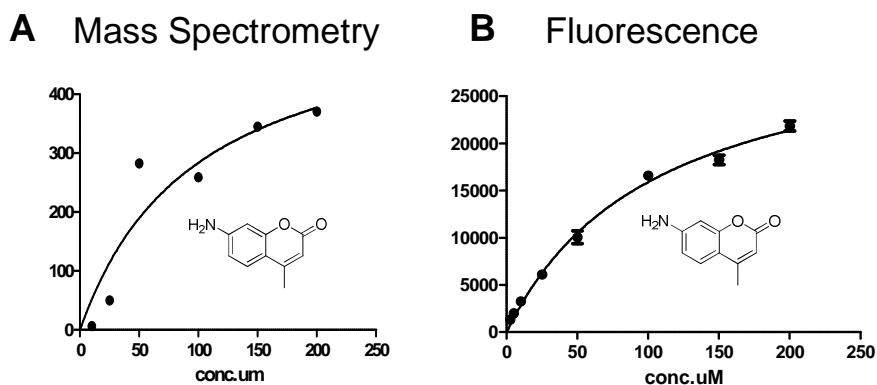
Figure 34: Trypsin cleavage kinetics of peptidic substrate monitored by online ESI-MS and fluorescence in comparison. (A) Fluorogenic trypsin substrate (5 - 200 µM) was incubated with 25 nM trypsin and cleavage was monitored by continuous flow ESI-MS analysis. Reaction velocity was determined by calculated by increase of florescent cleavage product aminomethyl coumarin (AMC). (B) Same experimental setup as in (A) but proteolysis was detected by fluorescence of cleavage product AMC). A K_M = 93,6 +/- 77,3 obtained by ESI-MS resembles the K_M = 104,8 +/- 11,52 obtained by fluorescence.

In order to determine kinetic data of tryptic substrate cleavage and directly compare the two methods Trypsin was incubated with different concentrations of substrate (0 µM – 200 µM) and the Michaelis Menten constant (K_M) was determined. Similar K_M values were determined by the fluorescent assay and the mass spectrometry based assay (Figure 34). This proves that kinetic data of the model enzyme trypsin, achieved with the commonly used fluorogenic assay can be reproduced by real-time ESI-MS measurements.

Because of relatively long measuring times and the loss of data in the first minutes, the established MS-online assay is not optimal for instable enzymes or for proteases that tend to self-digestion. Nevertheless, trypsin was a good model enzyme to optimize assay setup and conditions. Fortunately self-cleavage is not known for the rhomboid proteases used in this study and therefore is not a criterion for exclusion of this method to measure rhomboid cleavage kinetics.

2.4.3 Rhomboid kinetic determined by ESI-MS

Enzyme kinetics of rhomboid proteases have recently been determined by using a gel-based assay (Lazareno-Saez et al, 2013). The FRET-based substrate used in an

inhibitor screen (Pierrat et al, 2011) has not yet been used to study rhomboid enzyme kinetics. Additionally this substrate contained only a part of the natural TatA substrate TMD and might therefore result in different cleavage kinetics than substrates containing the intact substrate TMD. It was therefore the goal of the study to perform online ESI-MS enzyme kinetics of rhomboid making use of the intact substrate protein TatA.

To this end, recombinantly expressed and detergent purified TatA was incubated with wild type AarA or inactive AarA S150A mutant respectively, quenched and analysed by ESI-MS (Figure 35).
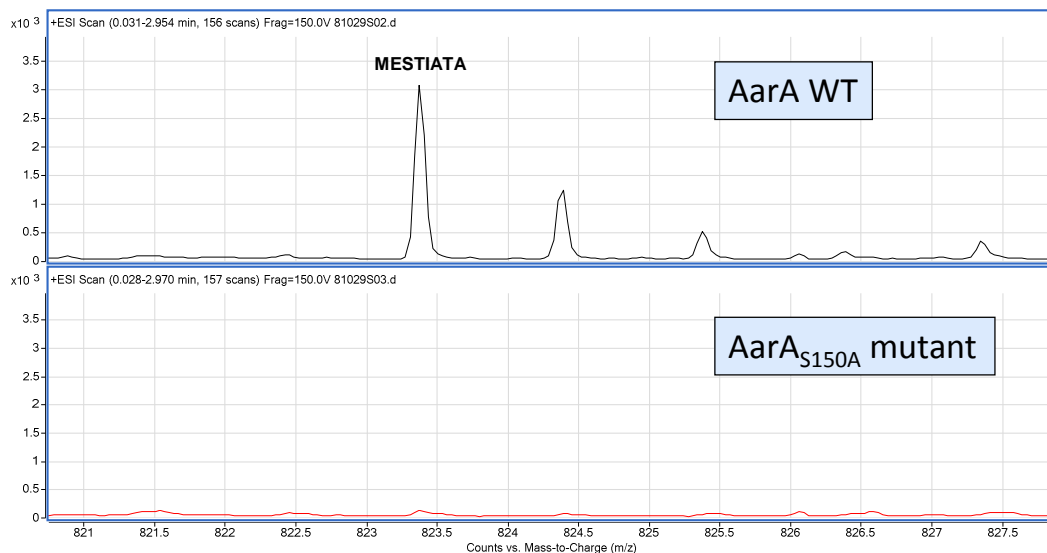

Figure 35: TatA cleavage by AarA monitored by ESI-MS. TatA (20 µM) was incubated with AarA (1 µM) for 1 h. The N-terminal peptide cleavage product (8 amino acids; 823,38 m/z) proves AarA cleavage at the physiological cleavage site. No cleavage product was detected when TatA was incubated with catalytically inactive AarA$_{S150A}$ rhomboid.

TatA cleavage at the physiological cleavage site could be detected with the wild type AarA but not the inactive mutant. Unfortunately only the N-terminal cleavage product (8 amino acid peptide MESTIATA; MW: 822.4 Da) could be detected. No MS-signals were obtained for the uncleaved substrate and the C-terminal cleavage product, which may be due to hydrophobicity. When monitoring TatA turnover by AarA in real-time using the optimized experimental setup (see chapter 2.4.2) only weak increase of the peptide product could be detected and consequently the slope could not be determined correctly in order to use this technique for rhomboid enzyme kinetics. To

determine K_M of rhomboid substrate cleavage it would probably be necessary to monitor cleavage of substrate concentrations lower than 20 µM which would not be possible with this substrate. On the other hand, increasing substrate concentration might contaminate the ESI-MS ion source.

It was therefore decided to design a peptidic rhomboid substrate based on TatA. In contrast to other peptidic rhomboid substrates (Pierrat et al, 2011) we chose to use the intact TatA TMD since interaction between rhomboid protease and the substrates TMD might be important for substrate recognition and cleavage. Enzyme kinetics determined by using intact TMD substrates are likely reflect rhomboid cleavage kinetics better than peptide not containing the intact TMD., flanked by three lysine residues (N- and C-terminal respectively) in order to increase solubility and ionizability (Figure 36).
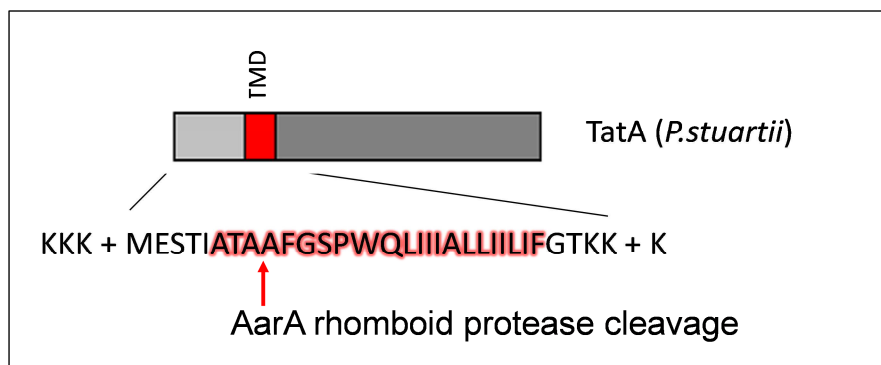
Figure 36: TatA-based peptidic rhomboid substrate. Rhomboid substrate peptide design was based on natural *P. stuartii* rhomboid AarA substrate protein TatA. Additional lysine residues were added to the N- and C-terminus (TMD is flanked by each three lysine residues) to provide solubility and ionizability.

In order to test cleavage of the synthetic substrate by rhomboid protease, lysine residues were fluorescently labeled using an amino-reactive fluorophore (TAMRA succinimidyl ester). AarA cleavage of the fluorescently labeled substrate was monitored by SDS-PAGE (Figure 37) and fluorescence scan of the gel.
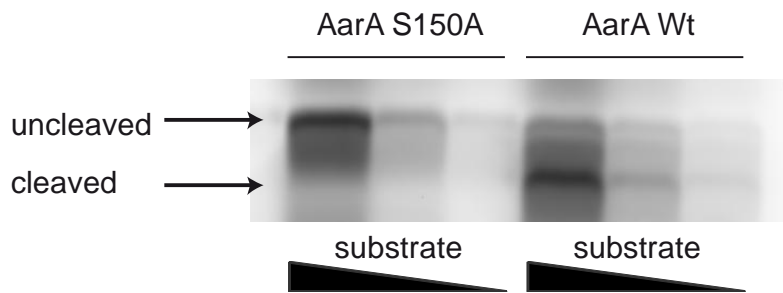
Figure 37: Hydrolysis of fluorescently labeled peptide by rhomboid AarA. N- and C-terminal lysine residues of rhomboid peptide substrate were labeled by amino-reactive TAMRA. Different amounts of the substrate were incubated with 0,5 µM of wild type AarA or inactive S150A mutant AarA for 30 min. Wild type AarA quantitatively cleaves fluorescent peptide.

AarA quantitatively cleaved the fluorescent peptide. The smeary occurrence of the fluorescent band is probably due to incomplete labeling of the six lysine residues by amino-reactive TAMRA which leads to a difference in MW of ~0.5 kDa per fluorophore. The fluorescent peptide could not be analyzed by MS to monitor the efficiency of fluorescent labeling probably due to lysine residues covalently bound to fluorophore do not serve as charge carriers any more. Additionally, bound fluorophore increases hydrophobicity of the peptide and probably complicates MS analysis as well.

The finding that AarA cleaves the peptidic substrates despite the covalently bound fluorophore molecules might facilitate the design of new FRET based substrates that better resemble physiological enzyme-substrate interaction. This could be done by exchanging lysine by arginine residues as charge carriers and introducing a lysine and a cysteine residue on each the N-and the C-terminal side of the peptide. These residues could then be specifically functionalized by fluorophores that enable FRET analysis.

ESI-MS analysis of the peptide substrate (20 µM) revealed good detectability (Figure 38) which was not achieved by using the same concentration of intact protein substrate TatA.
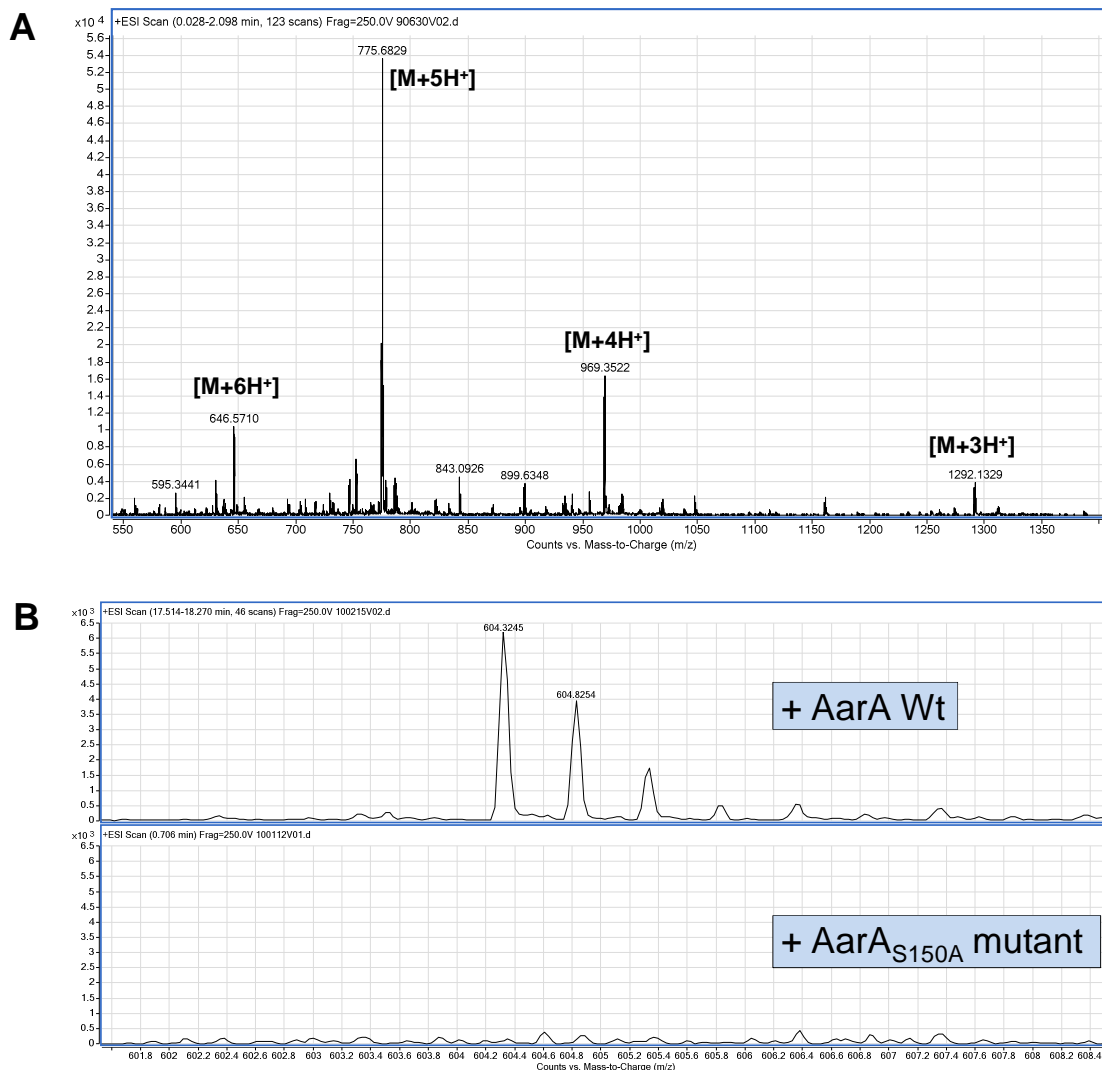
Figure 38: Peptidic rhomboid substrate analyzed by ESI-MS. (A) 20 µM of peptide substrate was prepared in 10 µM ammonium actetate reaction buffer, mixed 1:5 (vol/vol) with 50% ACN/0,1%FA (vol/vol) and analyzed by ESI-MS by direct injection. Multiple charging states were detected (indicated). (B) Rhomboid peptide substrate (20 µM) was incubated with wild type AarA or the inactive mutant $AarA_{S150A}$ respectively (each 1 µM) for 30 min, enzymatic activity was quenched by mixing with 1:5 (vol/vol) with 50% ACN/0,1%FA (vol/vol) and sample was analyzed yby ESI-MS. The N-terminal cleavage product (peptide 604,4 Da [M+2H]), corresponding to AarA cleavage at the physiological TatA cleavage site was detected when incubated with AarA Wt but not with the active site mutant $AarA_{S150A}$.

The introduction of six lysine residues led to multiple charged states of the peptide detected by ESI-MS with the 5-fold charged peptide being the most prominent MS signal. Incubating this substrate peptide with AarA gave rise to a double charged at 604,4 m/z which matches the MW of the N-terminal cleavage product after rhomboid cleavage at the physiological cleavage site (Figure 38). In contrast to the N-terminal cleavage product of the protein substrate TatA, signal intensity of N- terminal product

was intense enough to monitor its increase and calculate the slope for kinetic studies (Figure 39).
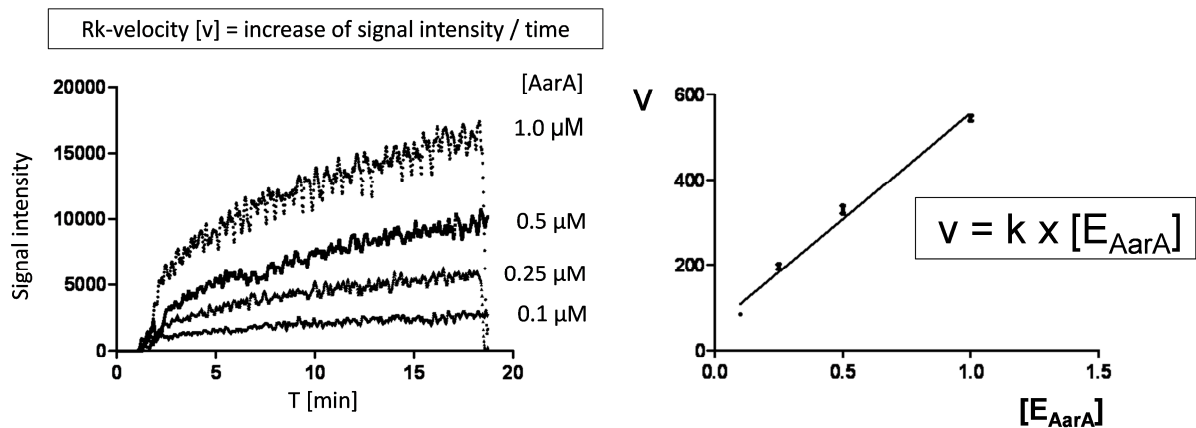
Figure 39: Rhomboid cleavage kinetics by online ESI-MS. (A) Substrate peptide (20 μM) was incubated with different concentrations of AarA (ranging from 0,1 to 1 μM) in duplicate reactions (only one example set of reactions shown). Reaction velocity was determined by the slope of increase of signal intensity of the N-terminal cleavage product over time. (B) Linear correlation between Reaction velocity and AarA concentration shows applicability of the assay.
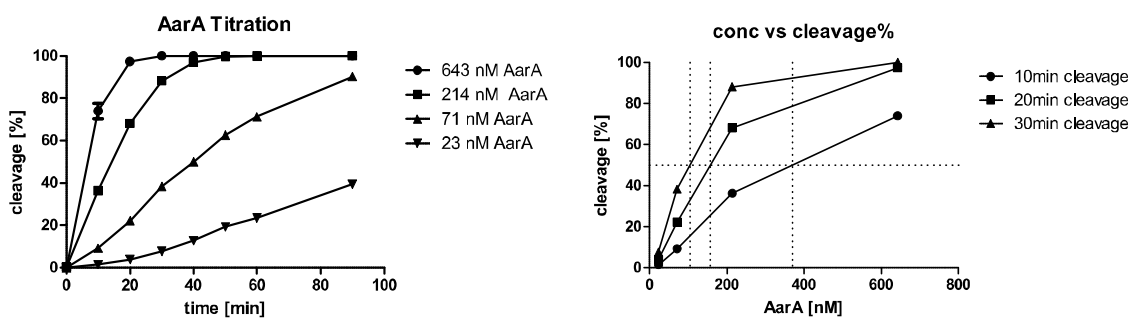
Linear correlation of reaction velocity and enzyme concentration shows that the peptide substrate gets cleaved in an enzyme-concentration dependent manner. Unfortunately the assay performed weak day to day reproducibility and was very time intensive (30-45 min / measurement). The weak reproducibility was probably due to interaction of rhomboid with surfaces of the tubing and the syringe. AarA cleavage of TatA was strongly influenced by the plastic surface when reaction was performed in different plastic tubes as determined by MALDI-MS (data not shown). Even though tubings and the syringes were flushed, hence pre-coated with detergent containing buffer, enzymes might stick to the surfaces and therefore lead to reduced substrate cleavage. Furthermore, even though the new synthetic peptide substrate contains the intact substrate TMD and probably matches the substrate specificity better than reported peptide substrates lysine residues might influence cleavage kinetics compared to the protein substrate.

To overcome this limitation, a MALDI-based method was considered to measure rhomboid cleavage kinetics of intact protein substrates.

2.4.4 Rhomboid kinetics determined by MALDI-MS

Even though the ESI-online assay developed in chapter 2.4.3 principally works, the artificial nature of the used peptide substrate might influence cleavage kinetics and obtained data might not resemble rhomboid cleavage kinetics of the natural substrate protein. As described in chapter 2.2, MALDI is a good method to monitor rhomboid cleavage of the intact protein substrate TatA. With this assay not only inhibition of rhomboid proteases can be detected but may also be used in future studies of rhomboid cleavage kinetics. Comparison of parameters such as enzyme efficiency k_{cat}/K_M of different rhomboid substrates and rhomboid mutants could give better insight into rhomboid substrate recognition. A gel-based assay has been used to study caspase kinetics for calculating k_{cat}/K_M by incubating protein substrates with different protease concentrations (Timmer et al, 2009). In that study, the enzyme efficiency (k_{cat}/K_M) of caspase was calculated by using formula indicated in Figure 40. $[E_{1/2}]$, the enzyme concentration that led to turnover of 50 % substrate protein after a certain incubation time was determined by the disappearance of the full length protein as estimated by densiometry.

In order to calculate k_{cat}/K_M, rhomboid substrate TatA was incubated with increasing AarA concentrations and cleavage was followed over time. The k_{cat}/K_M values calculated for three different incubation times (t = 10, 20, 30 min; Figure 40) showed good reproducibility of the obtained kinetic data. While in the inhibitor screening partly formylated TatA was used in order to prevent cross reactions of potential inhibitors with PDF, TatA was quantitatively deformylated for measuring enzyme kinetics. Plotting the slopes of substrate cleavage against the enzyme concentration showed a linear correlation between AarA concentration and substrate conversion which is crucial for an enzyme kinetic assay. For correctly determining the AarA efficiency, it is important to choose a substrate concentration which is bigger than the K_M. At high substrate concentrations, the reaction velocity appears to be independent of concentration of the substrate concentrations so differences in enzyme are the rate limiting factor. In this experiment TatA was used at a 20 μM concentration, with an unknown K_M. In case the actual K_M value for TatA is approaching or lower than 20 μM, the k_{cat}/K_M will be underestimated.

AarA Titration — 643 nM AarA, 214 nM AarA, 71 nM AarA, 23 nM AarA

conc vs cleavage% — 10min cleavage, 20min cleavage, 30min cleavage

$$\frac{k_{cat}}{K_m} = \frac{\ln 2}{t \cdot E_{1/2}}$$

10min: $E_{1/2}$ = 370nM k_{cat}/K_M: $3.1 * 10^3$ $M^{-1}s^{-1}$
20min: $E_{1/2}$ = 158nM k_{cat}/K_M: $3.7 * 10^3$ $M^{-1}s^{-1}$
30min: $E_{1/2}$ = 104nM k_{cat}/K_M: $3.7 * 10^3$ $M^{-1}s^{-1}$

Figure 40: TatA cleavage kinetics. (a) Measurement of TatA cleavage by different concentrations of AarA followed over time. TatA was quantitatively deformylated using PDF. Cleavage percentages were calculated after correction for the difference in ionization of cleaved and uncleaved TatA (see Figure 13) (b) The cleavage percentage plotted against the AarA concentration. (c) Calculation of apparent k_{cat} / K_M can be determined by the indicated formula[2], as long as [S] << K_M. In this experiment, [TatA] was used at a 20 μM concentration, with an unknown K_M. At a K_M value approaching or lower than 20 μM, the k_{cat}/K_M will be underestimated.

The K_M of TatA cleavage by AarA could possibly be determined by this method as well by measuring substrate cleavage of constant rhomboid concentration with increasing TatA concentrations. In case of the substrate concentration being so high, that it interferes with proper matrix crystal formation, cleavage reaction could be diluted after quenching. K_M is inversely proportional to the fraction of enzyme in the reaction that is in complex with a substrate molecule and con therefore be used to estimate the affinity of a certain substrate to the enzyme.

This new method enables enzyme kinetics by using intact substrate proteins. This circumvents the synthesis of synthetic substrate peptides and is therefore less artificial. Additionally it provides the possibility to investigate into the proteases substrate specificity. It is relatively easy to obtain a large number of substrate variants by simple point mutations (Quick change mutagenesis). Using MALDI technology to monitor substrate protein hydrolysis allows a much higher sample throughput compared to gel electrophoresis.

Therefore the established assay represents a suitable method for future rhomboid kinetic studies. This will help to get a better understanding of rhomboid enzymology.

3 Methods

3.1 Competent *E. coli* cells

5 ml of sterile LB-medium (Lennox; Carl Roth; if necessary containing antibiotics) were inoculated with *E. coli* (picked colony or 5 µl 10 % Glycerol stock) and incubated for 8 h at 37 °C. 100 µl of this culture was then diluted in fresh 5 ml of LB-medium and incubated over night at 37 °C. The culture was diluted 1:100 in fresh LB and incubated at 37 °C to $OD_{600} = 0,4$ and subsequently pelleted for 5 min; 5000 rpm at 4 °C. Cells were resuspended in 200 ml cold (4 °C), sterile $MgCl_2$ (0,1 M) and incubated on ice for 30 min. Cells were centrifuged (5 min: 5000 rpm; 4°C), resuspended in 4 ml cold $CaCl_2$ and incubated for 30 min on ice. Cells were then aliquoted to 50 µl in sterile 1,5 ml reaction tubes, snap frozen in liquid nitrogen and stored till usage at -80 °C.

3.2 Transformation of *E. coli*

50 µl of competent *E. coli* cells were mixed with 200 ng plasmid DNA and incubated on ice for 10-30 min. For heat shock cells were incubated at 42 °C for 2 min, immediately diluted 1:10 with fresh LB (without antibiotics) and incubated for 30-60 min at 37 °C under permanent shaking. 100-200 µl of transformed cells were streaked out on a LB-agar-plate containing appropriate antibiotics and incubated for 16 h at 37 °C. Plasmids used in this study were kindly provided by Matthew Freeman (AarA Wt and SA mutant as well as GlpG Wt and SA mutant; pET-25b(+) vector; and TatA; pET-21a(+) vector; Lemberg et al, 2005) or received from NAIST, Japan (PDF; ASKA clone JW3248; plasmid vector pCA24N)

3.3 Protein purification

Transformed bacteria (*E. coli* strain BL21 (DE3) for AarA, GlpG and deformylase overexpression; *E. coli* ΔglpEGR::kan derived from *E.coli* MG1655 for TatA overexpression) were induced at an $OD_{600} = 0.8$ (TatA and deformylase) and $OD_{600} = 0.6$ (GlpG and AarA). In case of TatA Kanamycin was disclaimed after induction due to faster cell growth. Protein expression was performed for 4 h at 37°C

(TatA and deformylase) and o/n at 16°C (GlpG and AarA). The pelleted cells were lysed in a buffer (20 mM HEPES pH 7.4, 10 mM NaCl, 10 % Glycerol, Roche Complete inhibitor mix) by a French pressure cell press (SLM Instruments *Simo Aminco*) and sequentially centrifuged 3000 x g, 15 min, 4°C to remove unlysed cells, supernatant was centrifuged at 100000 x g, 30 min, 4°C to pellet membrane fraction. For deformylase purification supernatant was directly subjected to Ni-NTA beads (see below). For purification of membrane proteins the pellet was resuspended in buffer (20mM HEPES-NaOH pH 7.4, 10% glycerol, 300mM NaCl, 10mM imidazole) and solubilized by adding 1,5 % beta-D-dodecyl-maltoside (DDM) and shaking for 3 h at 4°C. Centrifugation at 100000 g, 30 min, 4°C removes unsolubilized membrane debris and supernatant is subjected to Ni-NTA agarose beads (Qiagen; 0,5 ml beads per 2 l bacterial expression culture) for several hours and eluted by standard imidazole washing steps (25 mM, 50 mM, 100 mM, 750 mM final elution). Dialysis was performed against buffer (20mM HEPES-NaOH pH 7.4, 10% glycerol, 300mM NaCl, 10mM imidazole, 0,05% DDM for rhomboids AarA and GlpG, 0,0125% DDM for TatA and deformylase).

3.4 SDS-PAGE

Polyacrylamid gelelectrophoresis was done by using a two-gel electrophoresis chamber (PerfectBlue Dual Gel System Twin S; Peqlab).

Proteins were mixed with 4x sample buffer (40 % Glycerol (v/v), 10 % 1 M Tris pH 6,8 (v/v), 10 % β-mercaptoethanol (v/v), 6 % SDS (w/v), 0,02 % (w/v) Bromphenolblau) and separated using 15 % Tris-Glycine gels and running buffer (1,5 % (w/v) Tris basic, 7,2 % (w/v) glycine, 0,5 % (w/v) SDS). Gel electrophoresis was done at 120 V. Gels were prepared unsing 30 % acrylamide (acrylamide:bisacrylamide 37,5:1), 4x Tris buffer for Stacking gel (0,5 M Tris, pH 6,6 with HCl, 0,002 % bromphenol blue) and 4x Tris buffer for Running gel (3 M Tris, pH 8,8 with HCl).

Ingredient	Separating gel (15%)	Stacking gel (4%)
4x Tris buffer	20 ml	10 ml
30% acrylamide	40 ml	6.66 ml
ddH$_2$O	20 ml	23.34 ml
Total	80 ml	40 ml
Mix well before adding APS and TEMED.		
10% APS	600 µl	400 µl
TEMED	60 µl	40 µl
Then pour immediately		

Peptides were mixed with 4x sample buffer and separated using 15 % Bis-Tris gels and MES running puffer (MES SDS running buffer 20x; Invitrogen) at 100 V voltage. Gels were prepared by using 30% acrylamide (19:1 acryl:bisacryl) and 3.5X gel buffer (1.25 M Bistris, pH 6.5-6.8 with HCl).

Ingredient	Running gel (10%)	Stacking gel (4%)
Water	19.1 mL	8.7 mL
3.5 x buffer	14.3 mL	4.3 mL
30% acrylamide	16.6 mL	2 mL
Total	50 mL	15 mL
Mix well before adding APS and TEMED.		
10% APS	250 µL	100 µL
TEMED	25 µL	10 µL
Then pour immediately		

3.5 Coomassie staining of SDS-PAGE gels

Gels were incubated with Coomassie staining solution (Roti®-Blue, kolloidal Coomassie; Carl Roth) for 16 h and washed three times 20 min with ddH$_2$O.

3.6 Western Blot / Immunodetection

Proteins, separated by SDS-PAGE were transferred on nitrocellulose (NC) membrane using a semi-dry blotter (VWR). The NC membrane was then incubated in 3 % nonfat dry *milk* in PBS buffer containing 0,1 % Tween 20 detergent (PBST) to block to prevent unspecific binding for at least 1-2 h. NC membrane was washed 3 times 5 min with PBST buffer and incubated with Anti His$_6$ horse radish peroxidase coupled antibody (Roche Applied Science) diluted 1:3000 in 3 % milk-PBST for 1-2 h (RT) or 16 h (4 °C). After washing the NC membrane 3 times 15 min with PBST it was incubated with enhanced chemiluminescence substrate (ECL Plus, GE healthcare) for 2 min. Luminescence was detected by film (X-Omat LS, 180x240 mm; Kodak) using a film developer (Kodak).

3.7 MALDI matrix preparation / spotting

Sinapic acid and CHCA were solved in 30 % acteonitrile / 0,1 % trifluoric acid - HABA was soved in 50 % acteonitrile / 0,1 % trifluoric acid (10 mg in 1 ml) by ultrasonication for 5 min and centrifuged for 5 min at 12000 rpm. DHAP was solved in a 3:1 mixture of ethanol and 80 mM diammonium-hydrogen citrate (10 mg im 658 µl) by ultrasonication (5 min) and centrifuged for 5 min at 12000 rpm. For SA, CHCA and HABA sample was diluted 1:1 (v/v) with 30 % acteonitrile / 0,1 % trifluoric acid, mixed 1:1 (v/v) with the MALDI matrix solution and 2 µl thereof were spotted on the MALDI-target plate. For DHAP sample was mixed 1:1 (v/v) with 2 % TFA and 1 vol. equivalent of MALDI matrix was added prior to spotting 2 µl onto the MALDI target plate.

3.8 Deformylation of TatA and ionization factor

For deformylation of TatA enzyme was added in a 1:5 molar ratio and proceeded until full deformylation of TatA was achieved (~2h, tested by MALDI-MS). Ionization factor

was determined by mixing formylated or deformylated TatA and the proteolytic product (100% cleavage by AarA) in a 1:1 molar ratio. Differences in signal intensities due to different ionization of these molecules during MALDI-MS analysis were used to calculate an ionization factor. This factor was used to normalize screening and titration data.

3.9   MALDI-MS inhibitor Screening / inhibitor titration

Purified recombinant rhomboid proteases were preincubated with the small molecules (200 μM for screening; 200-0,02 μM for titration) for 20 min. All molecules were solved in DMSO and dilutions of these stocks were prepared to end up with 2% DMSO (v/v) in cleavage reactions. The reaction was started by adding the substrate proteins. The quenching of the reaction was done by 1:1 dilution with 2% TFA. Samples were directly analyzed by MALDI-TOF.

3.10 MALDI-MS Substrate cleavage assay

TatA cleavage by rhomboids GlpG and AarA was performed in buffer (50 mM HEPES pH 7.4, 10 % Glycerol, 0,0125 % DDM) at 37°C. Reaction was stopped by adding 2 % TFA and subjected to the 2,5-Dihydroxyacetophenone-MALDI Matrix (*Bruker* Daltonics, Bremen, Germany) (sample:TFA:matrix 1:1:1). Previously the matrix crystals were prepared by dissolving 0,1M in 25 % 80mM ammonium citrate bibasic / 75 % ethanol abs. by sonication. 1,5 μl of sample-TFA-matrix solution was spotted on the MALDI target plate (Bruker Daltonics, MTP 384 target plate ground steel T F). Samples were measured using a MALDI TOF/TOF mass spectrometer (*Bruker* Daltonics, ultrafleXtreme). Mass spectrometer settings were positive reflectron ion mode, 65% laser intensity, 1500 shots / sample, random walk, 100 shots / position on spot, frequency of laser 1000 shots / sec. Mass spectra were corrected using protein standard (*Bruker* Daltonics*, Protein Calibration Standard I*). The signal intensities of the protein substrate and the rhomboid cleavage product were analyzed using a custom made (*Bruker* Daltonics, Bremen, Germany) macro for flexAnalysis (Version 3.3.) software and imported into a Microsoft Excel worksheet for further analysis.

3.11 Z'-factor

Z'-factor was calculated (Zhang 1999) by measuring the substrate to product ratio of 8 independent sets of positive (0,5µM AarA S201A inactive mutant + 10µM TatA; no cleavage = full inhibition) and negative control (0,5µM AarA + 10µM TatA; full cleavage = no inhibition). The reaction was stopped after 30 minutes and analyzed by MALDI-MS.

$$Z' = 1 - \frac{3 \times (SD_{positive} - SD_{negative})}{|\, mean_{positive} - mean_{negative} \,|}$$

3.12 MALDI-MS based screening

Purified AarA (0,5µM) or GlpG (1,5µM) was incubated with the small molecules (200µM) for 20 min. All molecules were stored as 10mM DMSO stocks. The cleavage reaction was started by adding 10µM TatA and reacted at 37°C for 30 minutes (AarA) or 60 minutes (GlpG). All reactions were done in 10µl reaction volume, immediately quenched by adding 2% TFA and stored at -20°C until MALDI-MS analysis.

3.13 Fluorescence-based protease assay using fluorogenic peptide substrates

Proteolytic activity of trypsin (sequencing grade modified Trypsin, Invitrogen) and chymotrypsin (alpha-chymotrypsin grade I; Applichem) was determined using fluorogenic peptide-aminomethyl coumarine substrates (trypsin substrate: Z-Gly-Gly-Arg-AMC; Bachem and chymotrypsin substrate: Suc-Ala-Ala-Phe-AMC; Bachem). The fluorescence intensity of the free AMC of each substrate was determined at 25°C using a multimode plate reader (FLUOstar OPTIMA, BMG Labtech GmbH, Offenburg, Germany). The operating conditions for this fluorescence assay were as follows: Excitation filter 355 nm, emission filter 460 nm for the measurement of fluorescence intensity of free AMC. The samples were shaken 3 seconds before the measurement started. The cycle time and the total time of measurement varied according to the number of wells, which were measured. The reaction was started by mixing 50 µL of substrate (200 µM from 20 mM DMSO stock solution in 20 mM HEPES buffer, pH 7,4, 10 % glycerol) were placed in one well of a 96-well plate (COSTAR 96) and 50 µL of enzyme (2x end concentration in HEPES buffer, pH 7,4,

10 % glycerol) were added. Data were analyzed using Excel (Version 2007, Microsoft, USA) and GraphPad Prism 5.

3.14 ESI-MS real-time assay

Enzymes and their substrates were filtered (0,2 µm pore size), mixed and subsequently transferred into a 100 µl Hamilton syringe and placed in a syringe nano pump (HARVARD Aparatus). In the second position of the pump, a 500 µl syringe containing 50% ACN / 0,1% FA was placed. The latter was in order to improve mass spectrometric analysis and to stop enzymatic reaction immediately. Both syringes were connected to via a T-piece with teflon tubings and directly connected to the ESI-MS ion source. Mass detection was done by an ESI-TOF mass spectrometer (G210a time of flight mass spectrometrer; Agilent Technologies). The measurement was started immediately when mixing enzyme and substrate which led to a loss of data for the first 2-3 minutes due to experiment setup. Mass spectrometer settings were: Positive ion mode, capillary voltage 4000 V, nebulizer gas 15 p.s.i.g., drying gas 3,0 liter/min, gas temperature 300°C. The injection flow rate was set for the 100 µl syringe to 5 µL/min. The obtained data was exported to Microsoft Excel (version 20003) and was further analyzed using GraphPad Prism 5.

3.15 ABP labeling of rhomboids.

Purified enzymes and bacterial lysates were incubated with ABPs at 37°C and directly analysed by SDS-PAGE (in case of fluorophore bound ABP **36**) or visualized using click chemistry prior to gel electrophoresis. Samples were separated using 15% Tris-Glycine or 10% Bis-Tris SDS polyacrylamide gels. Fluorescent gel bands were detected using TRIO+ fluorescent scanner (settings for TAMRA detection: excitation at 532 nm, emission at 580 nm).

3.16 Activity-based labeling of endogenous GlpG in *E. coli* lysates

E. coli (BL21-Gold(DE3) pLysS strain; stratagene) cells were grown in 50 ml LB to OD = 0.6 and harvested by centrifugation (45min, 4°C, 3000 x g). Cells were lysed in buffer (20 mM HEPES pH 7.4, 10 mM NaCl, 10 % Glycerol, Roche Complete inhibitor mix) by using French pressure cell press (SLM Instruments *Simo Aminco*) and

sequentially centrifuged 3000 x g, 15 min, 4°C to remove unlysed cells. Supernatant was centrifuged at 100000 x g, 30 min, 4°C to pellet membrane fraction and pellet was resuspended in buffer (20mM HEPES-NaOH pH 7.4, 10% glycerol) to total protein concentration of 1 mg/ml. Thereof 50 µl were first treated with 200µM IC **16** or DMSO for 30min then incubated with 10µM ABP **36** for 30 min. Labeling reaction was stopped by adding 4x sample buffer and proteins separated by SDS-PAGE. Gels were analyzed by using fluorescent scanner (TRIO+; GE Healthcare) at 546 nm excitation and 574 nm emission.

3.17 *In Vivo* labeling of GlpG.

GlpG expressing *E. coli* (BL21-Gold(DE3) pLysS strain; stratagene) cells were grown in 50 ml LB to OD = 0.6 and expression was induced by adding 1mM IPTG for 2h. Cells were harvested by centrifugation (45min, 4°C, 3000 x g), washed with 1 x PBS buffer, resuspended in 4ml 1 x PBS containing 1mM EDTA and incubated at 37°C for 30min. Thereof 100µl each were first treated with 200µM IC **16** or DMSO for 30min then incubated with 10µM ABP **36** for 30 min, subsequently lysed by adding SDS-sample buffer and proteins separated by SDS-PAGE. Gels were analyzed by using fluorescent scanner (TRIO+; GE Healthcare) at 546 nm excitation and 574 nm emission.

3.18 *In Vivo* inhibition of AarA in Providencia stuartii.

Providencia stuartii cells (strain DSM4539) were grown in LB media at 37°C to OD = 0.3, diluted (1:1) with LB containing 1mM EDTA and incubated for 30 min. Thereof 1ml each were transferred to 24well culture plates and incubated with 100µM AarA inhibitors or DMSO for 1h. Cells were then directly analyzed by microscopy (100x magnification).

3.19 ABP labeling of rhomboids in different detergents.

10 ng GlpG in a 50 mM HEPES buffer, pH 7.5 containing 5 x CMC of either OM, DM, DDM or TDM were incubated with ABP **36** at 37 °C in the dark while shaking, and samples taken after 10, 30 and 60 min. The labeling reaction was quenched by the

addition of 1 x sample buffer. The samples were separated by a 15 % Tris-Glycine SDS-polyacrylamide gel and visualized on a fluorescent scanner at 546 nm excitation and 574 nm emission. Fluorescent band intensity was determined densitometrically using ImageJ.

3.20 TAMRA-SE labeling rhomboid substrate peptide

Peptide (KKKMESTIATAAFGSPWQLIIIALLIIGTKKK; synthesized by PSL) was solved in reaction buffer (0,1 M sodium bicarbonate, pH 8.3) to 5 mg/ml. Amino-reactive Tetramethyl-rhodamine succinimidy ester (TAMRA-SE) was solved in anhydrous DMSO to 10 mg/ml. Solutions of peptide and dye were mixed in a 1:20 (v/v) ratio. Reaction was incubated for 1 h at room temperature with continuous stirring. Reaction was stopped by adding 1 equivalent (v/v) stop reagent (1.5 M hydroxylamine, pH 8.5,) followed by dialysis against rhomboid reaction buffer (50 mM HEPES, pH 7,4, 10 % glycerol, 0,05 % DDM) using a dialysis membrane with 1 kDa cutoff (SPECTRUMLABS).

3.21 Densitometry of fluorescent PAGE protein bands

Estimation of intensities of fluorescent bands was dano by using ImageJ software. Scanned gel pictures were analyzed in tiff formate.

3.22 Radiolabeled rhomboid substrate

The RNA template for *in vitro* translation of radiolabeled rhomboid substrate was prepared as described elsewhere (Stevenson et al, 2006). Briefly, TatA based substrate (TatA N-terminaly extended by four SG-repeats: (SG)4-TatA) was produced by *in vitro* transcription using wheat germ extract (Promega) according to manufacturer's instructions. 0.5 µl of radiolabeled substrate was incubated with purified rhomboid AarA und HEPES-buffer (10mM HEPES-NaOH pH 7.4, 10% glycerol, 0,05% DDM) or Ammonium acetate-buffer (50mM HEPES-NaOH pH 7.4, 10% glycerol, 300mM NaCl, 0,05% DDM) conditions for 30 min reaction was quenched by adding 4x sample buffer and separated on a 10 % Bis-Tris gel. Gels were fixed 2x 15 min in 0.25% glutaraldehyde in 0.4 M borate/phosphate buffer pH 6.2 (made by adding ~ 30 mL of 0.5 M Na_2HPO_4 in 1L 0.4 M boric acid) and

equilibrated 2x 15 min in 40% MeOH/10% AcOH. Gels were subsequently dried under vacuum at 80 °C for 40-50 min. Dried gels were exposed to phosphoimaging screen for 16 h and analyzed using Phospho imager (TRIO+; GE healthcare).

3.23 Activity-based enzyme labeling in two steps using azide-alkyne cycloaddition

Enzymes were incubated with ABPs (from 100x DMSO stock solution) containing either alkyne or azide functional group for 30 min in HEPES-buffer conditions. Huisgen cycloaddition of fluorophore (TAMRA-alkyne or –azide respectively) was performed by adding click reagents TAMRA-N$_3$ 5mM in DMSO (=100x), TBTA 1.7mM in H$_2$O (=100x) which is a chelator for Cu$^+$ ions, CuSO$_4$ 100mM freshly prepared in H$_2$O (=100x) and reducing agent TCEP 100mM in H$_2$O (=100x). Click chemistry reaction was incubated for 1 h, quenched by adding 4x sample buffer and separated by SDS-PAGE. Gels were analyzed using a fluorescence scanner (TRIO+; GE healthcare) at 546 nm excitation and 574 nm emission.

4 Abbreviations

µM	micromolar
Å	Angstrom
ABP	activity-based probe
AMC	7-Amino-4-methylcoumarin
ACN	acetonitrile
CHCA	α-cyano-4-hydroxycinnamic acid
CMC	critical micellar concentration
C-terminus	Carboxy terminus
Da	Dalton
DCI	3,4-Dichloroisocoumarin
DDM	Dodecyl β-D-maltoside
DFP	diisopropyl fluorophosphonate
DHAP	2',4',-dihydroxyacetophenone
DM	Decyl β-D-maltoside
DMSO	Dimethyl sulfoxide
E. coli	Escherichia coli
EGF	epidermal growth factor
EGFR	epidermal growth factor receptor
ESI	electrospray ionization mass spectrometry
FA	formic acid
Form	formylated
FP	fluorophosphonate
FRET	Fluorescence resonance energy transfer
HABA	2-(4'-hydroxybenzeneazo)benzoic acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IC	isocoumarin

K	constant
kDa	kilo Dalton
K_M	Michaelis Menthen constant
L	liter
m/z	mass to charge
MALDI	matrix assisted lased disruption ionization
Mg	milligram
min	minute
ml	milliliter
mM	millimolar
MS	mass spectrometer
Mut	mutant
MW	molecular weight
Ng	nanogram
nM	nanomolar
N-terminus	amino terminus
OM	Octyl β-D-maltoside
P. stuartii	Providencia
PAGE	polyacrylamide gelelectrophoresis
PBS	phosphate buffered saline
PDF	peptide deformylase
RHBDL	rhomboid like
RNAse	Ribunuclease
S2P	site-2-protease
SA	sinapinic acid
SD	standard deviation
SDS	sodium dodecyl sulfate

SE	succinimidyl ester
SM	small molecule
T	time
TAMRA	Carboxytetramethylrhodamine
Tat	twin arginine transport
TDM	Tetradecyl β-D-maltoside
TFA	*Trifluoroacetic acid*
TMD	transmembrane domain
TOF	time of flight
v	velocity
v_{max}	maximal velocity
Wt	wild type

5 References

Adrain C, Freeman M (2012) New lives for old: evolution of pseudoenzyme function illustrated by iRhoms. *Nat Rev Mol Cell Biol* 13: 489-498

Adrain C, Strisovsky K, Zettl M, Hu L, Lemberg MK, Freeman M (2011) Mammalian EGF receptor activation by the rhomboid protease RHBDL2. *EMBO reports* 12: 421-427

Aebersold RH, Leavitt J, Saavedra RA, Hood LE, Kent SB (1987) Internal amino acid sequence analysis of proteins separated by one- or two-dimensional gel electrophoresis after in situ protease digestion on nitrocellulose. *Proceedings of the National Academy of Sciences of the United States of America* 84: 6970-6974

Angel TE, Aryal UK, Hengel SM, Baker ES, Kelly RT, Robinson EW, Smith RD (2012) Mass spectrometry-based proteomics: existing capabilities and future directions. *Chem Soc Rev* 41: 3912-3928

Baker RP, Urban S (2012) Architectural and thermodynamic principles underlying intramembrane protease function. *Nat Chem Biol* 8: 759-768

Baker RP, Young K, Feng L, Shi Y, Urban S (2007) Enzymatic analysis of a rhomboid intramembrane protease implicates transmembrane helix 5 as the lateral substrate gate. *Proceedings of the National Academy of Sciences of the United States of America* 104: 8257-8262

Banerjee S, Mazumdar S (2012) Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *Int J Anal Chem* 2012: 282574

Bannwarth L, Goldberg AB, Chen C, Turk BE (2012) Identification of exosite-targeting inhibitors of anthrax lethal factor by high-throughput screening. *Chem Biol* 19: 875-882

Bantscheff M, Lemeer S, Savitski MM, Kuster B (2012) Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal Bioanal Chem* 404: 939-965

Bayer ME, Leive L (1977) Effect of ethylenediaminetetraacetate upon the surface of Escherichia coli. *J Bacteriol* 130: 1364-1381

Ben-Shem A, Fass D, Bibi E (2007) Structural basis for intramembrane proteolysis by rhomboid serine proteases. *Proceedings of the National Academy of Sciences of the United States of America* 104: 462-466

Bogyo M, Baruch A, Jeffery DA, Greenbaum D, Borodovsky A, Ovaa H, Kessler B (2004) Applications for chemical probes of proteolytic activity. *Curr Protoc Protein Sci* Chapter 21: Unit 21 17

Bondar AN, del Val C, White SH (2009) Rhomboid protease dynamics and lipid interactions. *Structure* 17: 395-405

Bornsen KO, Gass MA, Bruin GJ, von Adrichem JH, Biro MC, Kresbach GM, Ehrat M (1997) Influence of solvents and detergents on matrix-assisted laser desorption/ionization mass spectrometry measurements of proteins and oligonucleotides. *Rapid Commun Mass Spectrom* 11: 603-609

Brooks CL, Lazareno-Saez C, Lamoureux JS, Mak MW, Lemieux MJ (2012) Insights into substrate gating in H. influenzae rhomboid. *J Mol Biol* 407: 687-697

Brown MR, Richards RM (1965) Effect of ethylenediamine tetraacetate on the resistance of Pseudomonas aeruginosa to antibacterial agents. *Nature* 207: 1391-1393

Bucknall M, Fung KY, Duncan MW (2002) Practical quantitative biomedical applications of MALDI-TOF mass spectrometry. *J Am Soc Mass Spectrom* 13: 1015-1027

Cadene M, Chait BT (2000) A robust, detergent-friendly method for mass spectrometric analysis of integral membrane proteins. *Anal Chem* 72: 5655-5658

Cheng TL, Wu YT, Lin HY, Hsu FC, Liu SK, Chang BI, Chen WS, Lai CH, Shi GY, Wu HL (2011) Functions of rhomboid family protease RHBDL2 and thrombomodulin in wound healing. *J Invest Dermatol* 131: 2486-2494

Dealler SF, Hawkey PM, Millar MR (1988) Enzymatic degradation of urinary indoxyl sulfate by Providencia stuartii and Klebsiella pneumoniae causes the purple urine bag syndrome. *J Clin Microbiol* 26: 2152-2156

Deu E, Yang Z, Wang F, Klemba M, Bogyo M (2010) Use of activity-based probes to develop high throughput screening assays that can be performed in complex cell extracts. *PloS one* 5: e11985

Eichacker LA, Granvogl B, Mirus O, Muller BC, Miess C, Schleiff E (2004) Hiding behind hydrophobicity. Transmembrane segments in mass spectrometry. *The Journal of biological chemistry* 279: 50915-50922

Ejigiri I, Ragheb DR, Pino P, Coppi A, Bennett BL, Soldati-Favre D, Sinnis P (2012) Shedding of TRAP by a rhomboid protease from the malaria sporozoite surface is essential for gliding motility and sporozoite infectivity. *PLoS Pathog* 8: e1002725

Evans MJ, Cravatt BF (2006) Mechanism-based profiling of enzyme families. *Chemical reviews* 106: 3279-3301

Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246: 64-71

Florea BI, Verdoes M, Li N, van der Linden WA, Geurink PP, van den Elst H, Hofmann T, de Ru A, van Veelen PA, Tanaka K, Sasaki K, Murata S, den Dulk H, Brouwer J, Ossendorp FA, Kisselev AF, Overkleeft HS (2010) Activity-based profiling reveals reactivity of the murine thymoproteasome-specific subunit beta5t. *Chem Biol* 17: 795-801

Freeman M (1994) The spitz gene is required for photoreceptor determination in the Drosophila eye where it interacts with the EGF receptor. *Mech Dev* 48: 25-33

Freeman M (2008) Rhomboid proteases and their biological functions. *Annu Rev Genet* 42: 191-210

Fritsch MJ, Krehenbrink M, Tarry MJ, Berks BC, Palmer T (2012) Processing by rhomboid protease is required for Providencia stuartii TatA to interact with TatC and to form functional homo-oligomeric complexes. *Mol Microbiol* 84: 1108-1123

Greis KD (2007) Mass spectrometry for enzyme assays and inhibitor screening: an emerging application in pharmaceutical research. *Mass Spectrom Rev* 26: 324-339

Greis KD, Zhou S, Burt TM, Carr AN, Dolan E, Easwaran V, Evdokimov A, Kawamoto R, Roesgen J, Davis GF (2006) MALDI-TOF MS as a label-free approach to rapid inhibitor screening. *J Am Soc Mass Spectrom* 17: 815-822

Haedke U, Kuttler EV, Vosyka O, Yang Y, Verhelst SH (2012) Tuning probe selectivity for chemical proteomics applications. *Curr Opin Chem Biol*

Haque H, Russell AD (1974) Effect of chelating agents on the susceptibility of some strains of gram-negative bacteria to some antibacterial agents. *Antimicrob Agents Chemother* 6: 200-206

Harper JW, Hemmi K, Powers JC (1985) Reaction of serine proteases with substituted isocoumarins: discovery of 3,4-dichloroisocoumarin, a new general mechanism based serine protease inhibitor. *Biochemistry* 24: 1831-1841

Heal WP, Dang TH, Tate EW (2011) Activity-based probes: discovering new biology and new drug targets. *Chem Soc Rev* 40: 246-257

Henzel WJ, Billeci TM, Stults JT, Wong SC, Grimley C, Watanabe C (1993) Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proceedings of the National Academy of Sciences of the United States of America* 90: 5011-5015

Heynekamp JJ, Hunsaker LA, Vander Jagt TA, Royer RE, Deck LM, Vander Jagt DL (2008) Isocoumarin-based inhibitors of pancreatic cholesterol esterase. *Bioorg Med Chem* 16: 5285-5294

Hill RB, Pellegrini L (2010) The PARL family of mitochondrial rhomboid proteases. *Seminars in cell & developmental biology* 21: 582-592

Hillenkamp F, Karas M, Beavis RC, Chait BT (1991) Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Anal Chem* 63: 1193A-1203A

Kateete DP, Katabazi FA, Okeng A, Okee M, Musinguzi C, Asiimwe BB, Kyobe S, Asiimwe J, Boom WH, Joloba ML (2012) Rhomboids of Mycobacteria: characterization using an aarA mutant of Providencia stuartii and gene deletion in Mycobacterium smegmatis. *PLoS One* 7: e45741

Kateete DP, Katabazi FA, Okeng A, Okee M, Musinguzi C, Asiimwe BB, Kyobe S, Asiimwe J, Boom WH, Joloba ML (2012) Rhomboids of Mycobacteria: characterization using an aarA mutant of Providencia stuartii and gene deletion in Mycobacterium smegmatis. *PloS one* 7: e45741

Kidd D, Liu Y, Cravatt BF (2001) Profiling serine hydrolase activities in complex proteomes. *Biochemistry* 40: 4005-4015

Klambt C (2002) EGF receptor signalling: roles of star and rhomboid revealed. *Curr Biol* 12: R21-23

Knuckley B, Jones JE, Bachovchin DA, Slack J, Causey CP, Brown SJ, Rosen H, Cravatt BF, Thompson PR (2010) A fluopol-ABPP HTS assay to identify PAD inhibitors. *Chem Commun (Camb)* 46: 7175-7177

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227: 680-685

Lazareno-Saez C, Arutyunova E, Coquelle N, Lemieux MJ (2013) Domain Swapping in the Cytoplasmic Domain of the Escherichia coli Rhomboid Protease. *J Mol Biol*

le Maire M, Champeil P, Moller JV (2000) Interaction of membrane proteins and lipids with solubilizing detergents. *Biochim Biophys Acta* 1508: 86-111

Lee ED, Mueck W, Henion J, Covey TR (1989) Real-time reaction monitoring by continuous-introduction ion-spray tandem mass spectrometry. *J. Am. Chem. Soc.* 111 (13), pp 4600–4604

Lee JR, Urban S, Garvey CF, Freeman M (2001) Regulated intracellular ligand transport and proteolysis control EGF signal activation in Drosophila. *Cell* 107: 161-171

Lei X, Li YM (2009) The processing of human rhomboid intramembrane serine protease RHBDL2 is required for its proteolytic activity. *J Mol Biol* 394: 815-825

Lemberg MK, Freeman M (2007) Functional and evolutionary implications of enhanced genomic analysis of rhomboid intramembrane proteases. *Genome Res* 17: 1634-1646

Lemberg MK, Menendez J, Misik A, Garcia M, Koth CM, Freeman M (2005) Mechanism of intramembrane proteolysis investigated with purified rhomboid proteases. *EMBO J* 24: 464-472

Letzel T (2008) Real-time mass spectrometry in enzymology. *Anal Bioanal Chem* 390: 257-261

Lichtenthaler SF, Haass C, Steiner H (2012) Regulated intramembrane proteolysis-- lessons from amyloid precursor protein processing. *J Neurochem* 117: 779-796

Lichtenthaler SF, Haass C, Steiner H (2011) Regulated intramembrane proteolysis-- lessons from amyloid precursor protein processing. *Journal of neurochemistry* 117: 779-796

Liesener A, Karst U (2005) Monitoring enzymatic conversions by mass spectrometry: a critical review. *Anal Bioanal Chem* 382: 1451-1464

Liesener A, Perchuc AM, Schoni R, Wilmer M, Karst U (2005) Screening for proteolytic activities in snake venom by means of a multiplexing electrospray ionization mass spectrometry assay scheme. *Rapid Commun Mass Spectrom* 19: 2923-2928

Lingwood D, Simons K (2010) Lipid rafts as a membrane-organizing principle. *Science* 327: 46-50

Lohi O, Urban S, Freeman M (2004) Diverse substrate recognition mechanisms for rhomboids; thrombomodulin is cleaved by Mammalian rhomboids. *Curr Biol* 14: 236-241

Lopez D, Kolter R (2010) Functional microdomains in bacterial membranes. *Genes Dev* 24: 1893-1902

Maegawa S, Koide K, Ito K, Akiyama Y (2007) The intramembrane active site of GlpG, an E. coli rhomboid protease, is accessible to water and hydrolyses an extramembrane peptide bond of substrates. *Mol Microbiol* 64: 435-447

Mayer U, Nusslein-Volhard C (1988) A group of genes required for pattern formation in the ventral ectoderm of the Drosophila embryo. *Genes Dev* 2: 1496-1511

Meissner C, Lorenz H, Weihofen A, Selkoe DJ, Lemberg MK (2011) The mitochondrial intramembrane protease PARL cleaves human Pink1 to regulate Pink1 trafficking. *Journal of neurochemistry* 117: 856-867

Michel V, Bakovic M (2007) Lipid rafts in health and disease. *Biol Cell* 99: 129-140

Miller MB, Bassler BL (2001) Quorum sensing in bacteria. *Annu Rev Microbiol* 55: 165-199

Moller JV, le Maire M (1993) Detergent binding as a measure of hydrophobic surface area of integral membrane proteins. *The Journal of biological chemistry* 268: 18659-18672

O'Donnell RA, Hackett F, Howell SA, Treeck M, Struck N, Krnajski Z, Withers-Martinez C, Gilberger TW, Blackman MJ (2006) Intramembrane proteolysis mediates shedding of a key adhesin during erythrocyte invasion by the malaria parasite. *J Cell Biol* 174: 1023-1033

Osenkowski P, Ye W, Wang R, Wolfe MS, Selkoe DJ (2008) Direct and potent regulation of gamma-secretase by its lipid microenvironment. *The Journal of biological chemistry* 283: 22529-22540

Pierrat OA, Strisovsky K, Christova Y, Large J, Ansell K, Bouloc N, Smiljanic E, Freeman M (2011) Monocyclic beta-lactams are selective, mechanism-based inhibitors of rhomboid intramembrane proteases. *ACS Chem Biol* 6: 325-335

Pils B, Schultz J (2004) Inactive enzyme-homologues find new function in regulatory processes. *J Mol Biol* 340: 399-404

Ploscher M, Granvogl B, Zoryan M, Reisinger V, Eichacker LA (2009) Mass spectrometric characterization of membrane integral low molecular weight proteins from photosystem II in barley etioplasts. *Proteomics* 9: 625-635

Powers JC, Asgian JL, Ekici OD, James KE (2002) Irreversible inhibitors of serine, cysteine, and threonine proteases. *Chemical reviews* 102: 4639-4750

Prox J, Rittger A, Saftig P (2012) Physiological functions of the amyloid precursor protein secretases ADAM10, BACE1, and presenilin. *Experimental brain research Experimentelle Hirnforschung Experimentation cerebrale* 217: 331-341

Puente XS, Sanchez LM, Overall CM, Lopez-Otin C (2003) Human and mouse proteases: a comparative genomic approach. *Nat Rev Genet* 4: 544-558

Rather PN, Orosz E (1994) Characterization of aarA, a pleiotrophic negative regulator of the 2'-N-acetyltransferase in Providencia stuartii. *J Bacteriol* 176: 5140-5144

Rather PN, Parojcic MM, Paradise MR (1997) An extracellular factor regulating expression of the chromosomal aminoglycoside 2'-N-acetyltransferase of Providencia stuartii. *Antimicrob Agents Chemother* 41: 1749-1754

Rawson RB, Zelenski NG, Nijhawan D, Ye J, Sakai J, Hasan MT, Chang TY, Brown MS, Goldstein JL (1997) Complementation cloning of S2P, a gene encoding a putative metalloprotease required for intramembrane cleavage of SREBPs. *Mol Cell* 1: 47-57

Ren G, Blum G, Verdoes M, Liu H, Syed S, Edgington LE, Gheysens O, Miao Z, Jiang H, Gambhir SS, Bogyo M, Cheng Z (2011) Non-invasive imaging of cysteine

cathepsin activity in solid tumors using a 64Cu-labeled activity-based probe. *PloS one* 6: e28029

Ruiz N, Falcone B, Kahne D, Silhavy TJ (2005) Chemical conditionality: a genetic strategy to probe organelle assembly. *Cell* 121: 307-317

Scheerle RK, Grassmann J, Letzel T (2012) Real-time ESI-MS of enzymatic conversion: impact of organic solvents and multiplexing. *Anal Sci* 28: 607-612

Schroder B, Saftig P (2010) Molecular insights into mechanisms of intramembrane proteolysis through signal peptide peptidase (SPP) *Biochem J* 427: e1-3

Seddon AM, Curnow P, Booth PJ (2004) Membrane proteins, lipids and detergents: not just a soap opera. *Biochim Biophys Acta* 1666: 105-117

Serim S, Haedke U, Verhelst SH (2012) Activity-based probes for the study of proteases: recent advances and developments. *ChemMedChem* 7: 1146-1159

Sherratt AR, Blais DR, Ghasriani H, Pezacki JP, Goto NK (2012) Activity-based protein profiling of the Escherichia coli GlpG rhomboid protein delineates the catalytic core. *Biochemistry* 51: 7794-7803

Sherratt AR, Braganza MV, Nguyen E, Ducat T, Goto NK (2009) Insights into the effect of detergents on the full-length rhomboid protease from Pseudomonas aeruginosa and its cytosolic domain. *Biochim Biophys Acta* 1788: 2444-2453

Shi G, Lee JR, Grimes DA, Racacho L, Ye D, Yang H, Ross OA, Farrer M, McQuibban GA, Bulman DE (2011) Functional alteration of PARL contributes to mitochondrial dysregulation in Parkinson's disease. *Hum Mol Genet* 20: 1966-1974

Siefert SA, Sarkar R (2012) Matrix metalloproteinases in vascular physiology and disease. *Vascular* 20: 210-216

Speers AE, Cravatt BF (2004) Profiling enzyme activities in vivo using click chemistry methods. *Chem Biol* 11: 535-546

Srinivasan P, Coppens I, Jacobs-Lorena M (2009) Distinct roles of Plasmodium rhomboid 1 in parasite development and malaria pathogenesis. *PLoS Pathog* 5: e1000262

Steinkamp T, Liesener A, Karst U (2004) Reaction monitoring of enzyme-catalyzed ester cleavage by time-resolved fluorescence and electrospray mass spectrometry: method development and comparison. *Anal Bioanal Chem* 378: 1124-1128
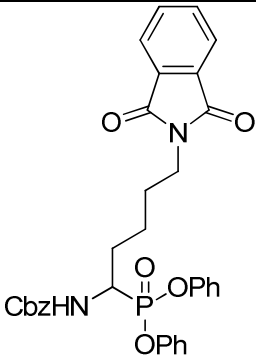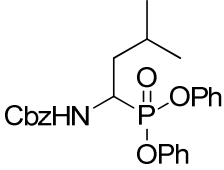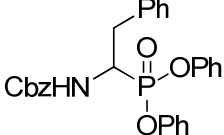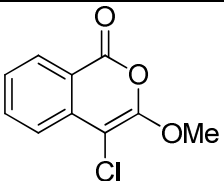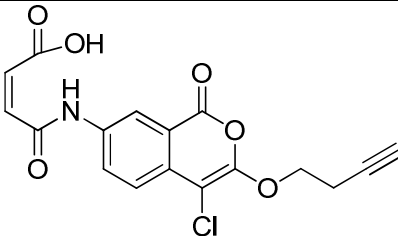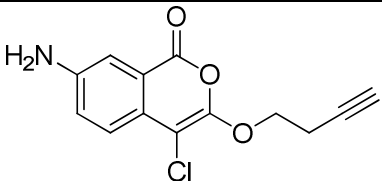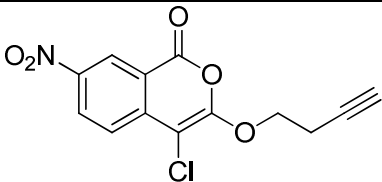
Stevenson LG, Strisovsky K, Clemmer KM, Bhatt S, Freeman M, Rather PN (2007) Rhomboid protease AarA mediates quorum-sensing in Providencia stuartii by activating TatA of the twin-arginine translocase. *Proceedings of the National Academy of Sciences of the United States of America* 104: 1003-1008
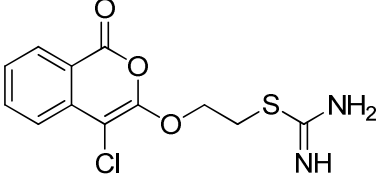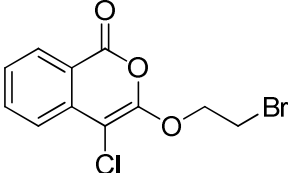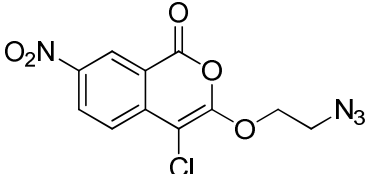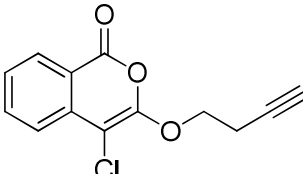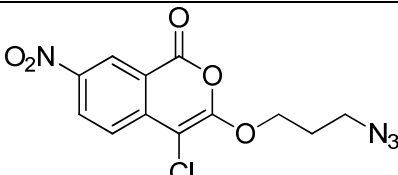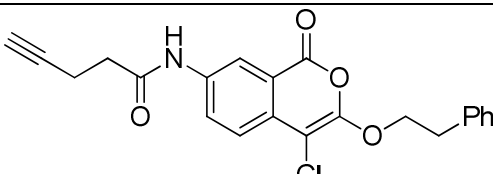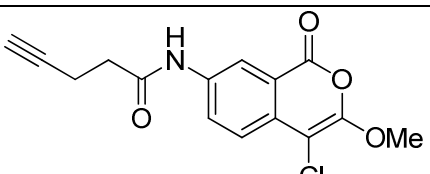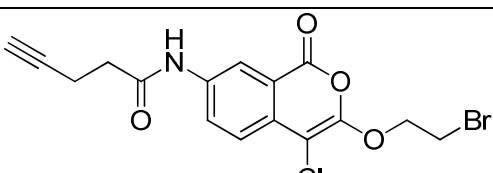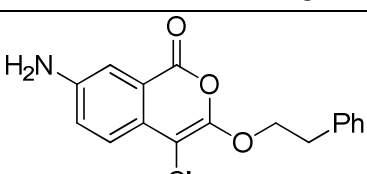
Strisovsky K, Sharpe HJ, Freeman M (2009) Sequence-specific intramembrane proteolysis: identification of a recognition motif in rhomboid substrates. *Mol Cell* 36: 1048-1059

Sturtevant MA, Roark M, Bier E (1993) The Drosophila rhomboid gene mediates the localized formation of wing veins and interacts genetically with components of the EGF-R signaling pathway. *Genes Dev* 7: 961-973

Tang J, Hernandez G, LeMaster DM (2004) Increased peptide deformylase activity for N-formylmethionine processing of proteins overexpressed in Escherichia coli: application to homogeneous rubredoxin production. *Protein Expr Purif* 36: 100-105

Timmer JC, Zhu W, Pop C, Regan T, Snipas SJ, Eroshkin AM, Riedl SJ, Salvesen GS (2009) Structural and kinetic determinants of protease substrates. *Nat Struct Mol Biol* 16: 1101-1108

Todd AE, Orengo CA, Thornton JM (2002) Sequence and structural differences between enzyme and nonenzyme homologs. *Structure* 10: 1435-1451

Urban S (2010) Taking the plunge: integrating structural, enzymatic and computational insights into a unified model for membrane-immersed rhomboid proteolysis. *Biochem J* 425: 501-512

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

Urban S, Wolfe MS (2005) Reconstitution of intramembrane proteolysis in vitro reveals that pure rhomboid is sufficient for catalysis and specificity. *Proceedings of the National Academy of Sciences of the United States of America* 102: 1883-1888

Vaara M (1992) Agents that increase the permeability of the outer membrane. *Microbiol Rev* 56: 395-411

Vinothkumar KR (2011) Structure of rhomboid protease in a lipid environment. *J Mol Biol* 407: 232-247

Vinothkumar KR, Strisovsky K, Andreeva A, Christova Y, Verhelst S, Freeman M (2010) The structural basis for catalysis and substrate specificity of a rhomboid protease. *EMBO J* 29: 3797-3809

Vosyka O, Vinothkumar KR, Wolf EV, Brouwer AJ, Liskamp RM, Verhelst SH (2013) Activity-based probes for rhomboid proteases discovered in a mass spectrometry-based assay. *Proceedings of the National Academy of Sciences of the United States of America*

Wang Y, Ha Y (2007) Open-cap conformation of intramembrane protease GlpG. *Proceedings of the National Academy of Sciences of the United States of America* 104: 2098-2102
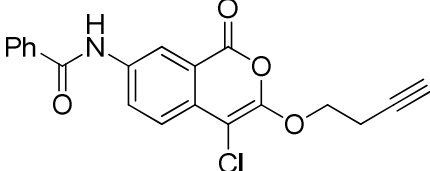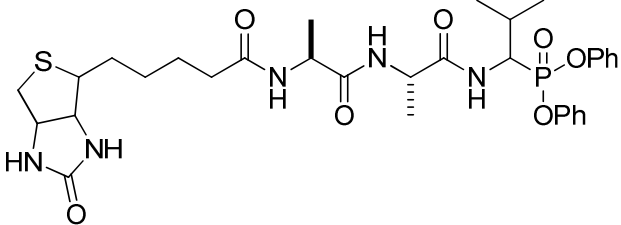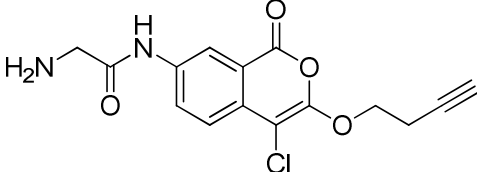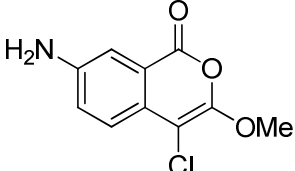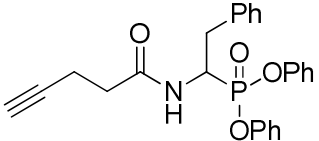
Wang Y, Maegawa S, Akiyama Y, Ha Y (2007) The role of L1 loop in the mechanism of rhomboid intramembrane protease GlpG. *J Mol Biol* 374: 1104-1113

Wang Y, Zhang Y, Ha Y (2006) Crystal structure of a rhomboid family intramembrane protease. *Nature* 444: 179-180

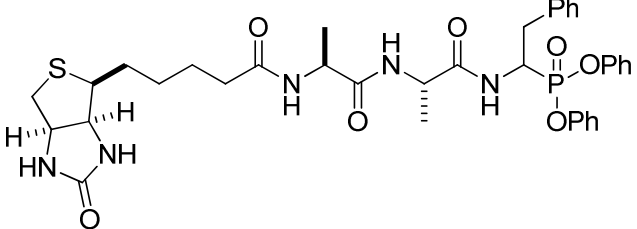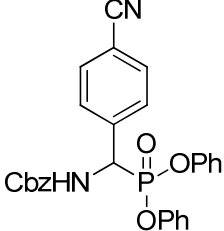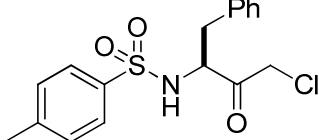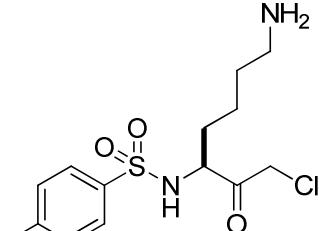Wolfe MS (2009) gamma-Secretase in biology and medicine. *Semin Cell Dev Biol* 20: 219-224

Wu Z, Yan N, Feng L, Oberstein A, Yan H, Baker RP, Gu L, Jeffrey PD, Urban S, Shi Y (2006) Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry. *Nat Struct Mol Biol* 13: 1084-1091

Xue Y, Chowdhury S, Liu X, Akiyama Y, Ellman J, Ha Y (2012) Conformational change in rhomboid protease GlpG induced by inhibitor binding to its S' subsites. *Biochemistry* 51: 3723-3731

Xue Y, Ha Y (2012) Catalytic mechanism of rhomboid protease GlpG probed by 3,4-dichloroisocoumarin and diisopropyl fluorophosphonate. *The Journal of biological chemistry* 287: 3099-3107

Ye J, Rawson RB, Komuro R, Chen X, Dave UP, Prywes R, Brown MS, Goldstein JL (2000) ER stress induces cleavage of membrane-bound ATF6 by the same proteases that process SREBPs. *Mol Cell* 6: 1355-1364

Zettl M, Adrain C, Strisovsky K, Lastun V, Freeman M (2011) Rhomboid family pseudoproteases use the ER quality control machinery to regulate intercellular signaling. *Cell* 145: 79-91

Zhou Y, Moin SM, Urban S, Zhang Y (2012) An internal water-retention site in the rhomboid intramembrane protease GlpG ensures catalytic efficiency. *Structure* 20: 1255-1263

6 Supplementary

Table 2: Structures of compounds used in this study.

Cmp	Structure
1	(phthalimide-terminated alkyl chain; CbzHN–CH–P(=O)(OPh)OPh)
2	(isobutyl; CbzHN–CH–P(=O)(OPh)OPh)
3	(Ph–CH₂; CbzHN–CH–P(=O)(OPh)OPh)
4	(isochromen-1-one; 3-OMe, 4-Cl)
5	(HOOC–CH=CH–C(=O)–NH–isochromenone; 4-Cl, 3-O-but-3-ynyloxy)
6	(H₂N–isochromenone; 4-Cl, 3-O-but-3-ynyloxy)
7	(O₂N–isochromenone; 4-Cl, 3-O-but-3-ynyloxy)

8	isochromen-1-one structure: 4-chloro-1-oxo-1H-isochromen-3-yl O-CH₂CH₂-S-C(=NH)NH₂
9	4-Cl, 1-oxo-isochromen-3-yl-O-CH₂CH₂-Br
10	7-O₂N, 4-Cl, 1-oxo-isochromen-3-yl-O-CH₂CH₂-N₃
11	4-Cl, 1-oxo-isochromen-3-yl-O-CH₂-C≡CH (but-3-ynyl)
12	7-O₂N, 4-Cl, 1-oxo-isochromen-3-yl-O-CH₂CH₂CH₂-N₃
13	HC≡C-CH₂CH₂-C(=O)-NH-[7]-4-Cl-1-oxo-isochromen-3-yl-O-CH₂CH₂-Ph
14	HC≡C-CH₂CH₂-C(=O)-NH-[7]-4-Cl-1-oxo-isochromen-3-yl-OMe
15	HC≡C-CH₂CH₂-C(=O)-NH-[7]-4-Cl-1-oxo-isochromen-3-yl-O-CH₂CH₂-Br
16	7-H₂N, 4-Cl, 1-oxo-isochromen-3-yl-O-CH₂CH₂-Ph

17	
18	
19	
20 (JLK-6)	
21	
22	
23	
24 (TPCK)	
25 (TLCK)	

26	CbzHN~S(=O)(=O)F (CbzHN–CH₂CH₂–SO₂F)
27	CbzHN–CH(CH₃)–CH₂–SO₂F
28	CbzHN–CH(CH(CH₃)₂)–CH₂–SO₂F
29	CbzHN–CH(CH₂CH(CH₃)₂)–CH₂–SO₂F
30	CbzHN–CH(CH₂Ph)–CH₂–SO₂F, Ph
31	BocHN–CH(CH₂Ph)–C(=O)–NH–CH₂CH₂–SO₂F, Ph
32	CbzHN–CH(CH₂Ph)–SO₂–NH–CH₂CH₂–SO₂F, Ph
33	BocHN–CH(CH₂Ph)–C(=O)–NH–CH(CH₂CH(CH₃)₂)–CH₂–SO₂F, Ph
34	BocHN–CH(CH₂CH(CH₃)₂)–C(=O)–NH–CH(CH₂CH(CH₃)₂)–CH₂–SO₂F
35	N₃–CH(CH₂Ph)–C(=O)–NH–CH(CH₂Ph)–SO₂–NH–CH₂CH₂–SO₂F, Ph, Ph
36	H_2N-substituted 4-chloro-1-oxo-1H-isochromen-3-yl–O–CH₂CH₂–(1,2,3-triazol-4-yl), N-linked propyl–NH–C(=O)–aryl rhodamine (with $-O_2C$, dimethylamino, N^+–methyl groups)

<u>Acknowledgement</u>