Charles University in Prague

Faculty of Science

Study programme: Biochemistry

Study branch: Biochemistry

Analysis of substrate specificity and mechanism of GlpG, an intramembrane protease of the rhomboid family

Analysa substrátové specifity a mechanismu GlpG, intramembránové proteasy z rodiny rhomboidů

Bc. Lucie Peclinovská

Diploma Thesis

Supervisor: Ing. Kvido Stříšovský, PhD

Prague 2014

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze dne ……

……

Lucie Peclinovská

ACKNOWLEDGEMENTS

Not all those that I would like to thank speak English, so I will write my acknowledgements in Czech.

Můj dík patří především mému školiteli Kvidovi Stříšovskému za to, že ve mě vložil důvěru a nabídl práci na velice atraktivním projektu v moment, kdy jsem byla v nelehké situaci. Za jeho ochotu kdykoliv pomoci, vstřícný přístup a všechny rady, které mi pomohly nejen k vypracování této práce, ale zároveň mi rozšířily obecné vědecké povědomí.

Ráda bych poděkovala všem mým kolegům, jmenovitě pak: Jakubovi Beganovi, Janě Horákové, Nicholasu Johnsonovi, Stancho Stanchevovi, Janu Škerlemu, Kateřině Švehlové a Sebastianu Zollovi za vytváření přátelského kolektivu, kterého je radost být součástí a zároveň za veškerou jejich pomoc. Díky těmto lidem, společně s členy laboratoře docenta Konvalinky, se kterými sdílíme některé prostory a vybavení, se v práci vždy cítím 'jako doma'.

Děkuji spolupracujícím skupinám z Ústavu organické chemie a biochemie AV ČR, v.v.i., pod vedením Dr. Václava Kašičky, Dr. Pavla Majera a Doc. Josefa Cvačky za jejich pomoc při analýze vzorků, zejména pak Mgr. Růžičkovi, Ing. Součkovi, Ing. Blechové a Dr. Hubálkovi.

V neposlední řadě moc děkuji celé mé širší rodině za vytváření dokonalého zázemí během studia a práce na projektu. Moc si vážím Vaší podpory a je mi jasné, že bez Vás bych to jen těžko dokázala! No a samozřejmě děkuji těm, kdo jsou tu pro mě, když je potřebuji, motivují mě, jsou mými velkými vzory a nikdy mě nenechají zahálet - všem mým přátelům. A díky Cady!

Abstract

Membrane proteins of the rhomboid-family are evolutionarily widely conserved and include rhomboid intramembrane serine proteases and rhomboid-like proteins. The latter have lost their catalytic activity in evolution but retained the ability to bind transmembrane helices. Rhomboid-family proteins play important roles in intercellular signalling, membrane protein quality control and trafficking, mitochondrial dynamics, parasite invasion and wound healing. Their medical potential is steeply increasing, but in contrast to that, their mechanistic and structural understanding lags behind. Rhomboid protease GlpG from *E.coli* has become the main model rhomboid-family protein and the main model intramembrane protease - it was the first one whose X-ray structure was solved. GlpG cleaves single-pass transmembrane proteins in their transmembrane helix, but how substrates bind to GlpG and how is substrate specificity achieved is still poorly understood. This thesis investigates the importance of the transmembrane helix of the substrate in its recognition by GlpG using mainly enzyme kinetics and site-directed mutagenesis. We find that the transmembrane helix of the substrate contributes significantly to the binding affinity to the enzyme, hence to cleavage efficiency, but it also plays a role in cleavage site presentation to the active site of GlpG. Moreover, we identify four residues in transmembrane domains 2 and 5 of GlpG, whose mutations shift substrate specificity of GlpG, which means that they most likely interact with the topologically corresponding region of the substrate – its transmembrane helix. Taken together, our data support the model of the enzyme-substrate interaction where the initial contact between the two occurs at an intramembrane exosite of GlpG, which facilitates and is followed by the binding of the scissile-bond region of the substrate into rhomboid active site in a sequence-dependent manner. This acquired mechanistic knowledge allowed us to develop a fluorogenic transmembrane peptide substrate that is cleaved by several rhomboids, and which will find use for kinetics assays and high-throughput screening of rhomboid inhibitors.

Key-words: membrane protein, intramembrane protease, rhomboid, GlpG, substrate specificity

Abstrakt

Proteiny z rodiny rhomboidů jsou široce konservované v evoluci a obsahují dvě velké podskupiny, proteolyticky aktivní intramembránové proteasy serinového typu (rhomboidy) a rhomboidům podobné proteiny, které během evoluce ztratily proteolytickou aktivitu, ale zachovaly si schopnost vázat membránový helix jiných proteinů. Zástupci obou těchto skupin hrají významnou roli v celé řadě biologických procesů, jako například v buněčné signalizaci, v mechanismech podílejících se na kontrole kvality sbalení proteinů v endoplasmatickém retikulu, v mitochondriální dynamice a také v hojení ran. Ačkoliv medicinský potenciál proteinů z rodiny rhomobidů vzrůstá, pochopení strukturní podstaty jejich funkce pokulhává. Hlavním strukturním a mechanistickým modelem pro pro rodinu rhomboidů a vlastně pro intramembránové proteasy obecně se stal rhomboid GlpG z *E.coli*, který byl první intramembránovou proteasou jejíž rentgenová struktura byla vyřešena. Substráty GlpG jsou membránové proteiny s jedním transmembránovým helixem a ke štěpení dochází v jejich transmembránové části, ale mechanismus vazby substrátu na enzym ani podstata substrátové specifity nejsou plně pochopeny. Tato práce se zabývá charakterizací významu transmembránové domény substrátu pro jeho rozpoznání proteasou GlpG a využívá hlavně enzymové kinetiky a cílené mutageneze. Prokázali jsme, že transmembránový helix substrátu přispívá k vazebné afinitě a zároveň se podílí na správném prezntaci štěpného místa aktivnímu centru rhombodu. Identifikovali jsme čtyři aminokyseliny GlpG lokalizované na transmembránových helixech 2 a 5, jejichž mutace vedla ke změně specifity enzymu. Všechny mutované aminokyseliny GlpG se nachází uvnitř membrány, z čehož vyplývá, že k interakci enzymu se substrátem dochází také v oblasti transmembránového helixu substrátu. K prvnímu kontaktu mezi substrátem a enzymem tak dochází uvnitř membrány v místě jiném, než je aktivní místo enzymu ("exosite"). Tato interakce napomáhá nejen vazbě, ale zároveň se podílí na správné prezentaci štěpného motivu aktivnímu místu a spoluurčuje substrátovou specifitu enzymu. Na základě těchto výsledků byl vytvořen fluorogenní transmembránový peptidový substrát, který je štěpen několika různými rhomboidy a tím by měl usnadnit hledání a charakterizaci inhibitorů rhomboidů

Klíčová slova: membránový protein, intramembránová proteasa, rhomboid, GlpG, substrátová specifita

TABLE OF CONTENTS

1 ABBREVIATIONS

AD	Alzheimer's disease
ADAM	a disintegrin and metalloproteinase
Amp	ampicilin
DDM	n-Dodecyl-β-D-maltoside
DFP	diisopropyl fluorophosphate
DMSO	dimethylsulfoxide
EDTA	ethylendiamintetraacetic acid
EGF	epidermal growth factor
EGFR	epidermal growth factor receptor
ER	endoplasmic reticulum
FRET	fluorescence resonance energy transfer
HPLC	high-performance liquid chromatography
IMAC	immobilised metal ion affinity chromatography
IOCB	Instutute of Organic Chemistry and Biochemistry
IPTG	isopropyl-α-D-thiogalactopyranosid
IRHD	iRhom homology domain
MALDI	matrix-assisted laser desorption/ionization
MMP	matrix metalloproteinase
MS	mass spectrometry
MST	microscale thermophoresis
NiNTA	nickel-nitrilotriacetic acid
OPA	o-phthaldialdehyde
PBS	phosphate buffer saline

PMSF	phenylmethylsulphonyl fluorid
RFU	relative fluorescence units
S2P	site-2-protease
SDS	sodium dodecyl sulphate
SDS-PAG	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SPP	signal peptide peptidase
SREBP	sterol regulatory element-binding protein
TGF	transforming growth factor
TMH	transmembrane helix
TNF	tumor necrosis factor
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol

2 INTRODUCTION

This thesis investigates the mechanism of intramembrane proteolysis by the rhomboid protease GlpG. To place this topic into a wider context of biology, concise introduction will be given into membrane proteins and their importance in biology, into intramembrane proteases and their role in disease, and specialised structural and mechanistic aspects of their functions relevant for this thesis.

2.1 MEMBRANE PROTEINS

Every cell on Earth has a phospholipid membrane around it that separates its interior from the outside environment. Constant flux of matter, chemical energy and information across this membrane is indispensable for the maintenance of cellular homeostasis and adaptive processes essential for life. Besides cytoplasmic membrane, eukaryotes also contain an intricate system of internal membrane organelles, such as the endoplasmic reticulum (ER), Golgi apparatus, mitochondria or chloroplasts, which compartmentalise cellular metabolic processes. The specific functional character of membranes is mostly conferred by proteins associated with membranes. Their content varies - from ~25% in the myelin membrane (that serves primarily as an insulation), to up to ~75% in the inner mitochondrial membrane carrying respiratory chain (the powerhouse of the cell). Strikingly, about 25% of all proteins in a genome are stably associated with (or integrated in) phospholipid membranes, which illustrates their importance [1].

2.1.1 The defining biophysical characteristics

Physicochemical properties and the underlying molecular architecture of proteins reflect their surrounding biophysical environment. Soluble proteins reside in polar environment, which thermodynamically favours the exposure of their polar amino acids to the solvent and packing of the hydrophobic ones into the core of the folded protein globule. Their transmembrane counterparts face entirely different conditions. Biological membranes are phospholipid bilayers consisting of amphipathic phospholipid molecules. Phospholipids

are composed of hydrophobic fatty acid ´tails´ varying in the length and saturation, and the polar ´headgroups´. This amphiphilic character together with their geometrical parameters causes phospholipids spontaneously form bilayers in aqueous solvents, orientating their hydrophobic tails together and exposing their polar heads; this in effect generates a two-dimensional membrane separating two aqueous compartments. The length of an average phospholipid is around 3 nm and the average thickness of cytoplasmic membrane is around 6 nm, of which ca 3 nm is the thickness of the hydrophobic core and about 1.5 nm is the thickness of each of the two headgroup regions on the solvent-exposed faces of the bilayer [2] (Fig.1). Transmembrane proteins interact with the hydrophobic core of the membrane, which forces the respective segment of their polypeptide backbone into a secondary structure, in which the polar groups of the backbone are 'hidden' from the lipid phase by engaging in mutual hydrogen bonds, while hydrophobic side-chains are exposed to the hydrocarbon core of the bilayer (mostly α-helix, less frequently β-sheet). It is clear, therefore, that protein dynamics and protein-protein interactions are going to be fundamentally different between soluble and membrane proteins, and, for the latter, they will be strongly influenced by the lipid environment and by the phenomena at the lipid-water interface.

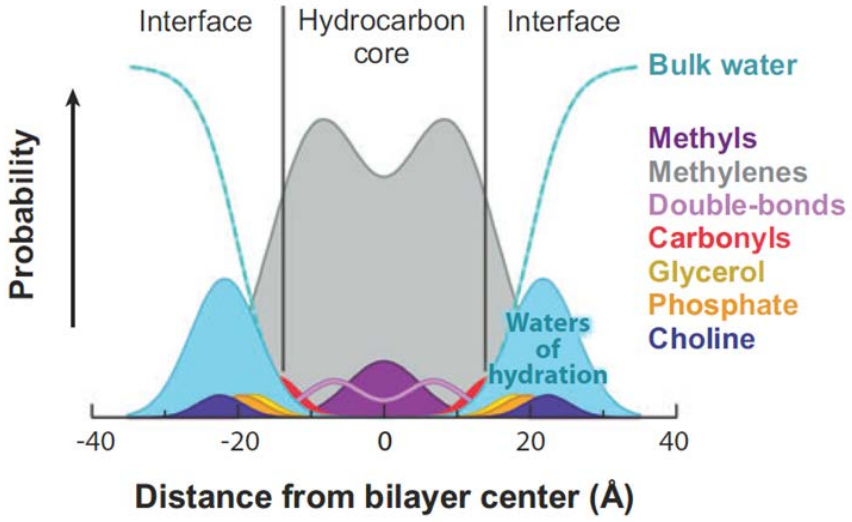


Figure 1: Schematic depiction of the polarity of a biological membrane and probability of occurrence of various chemical moieties along the bilayer. Area under the curve represents the number of constituent group per lipid. Adopted from [83].

Although the main lipid components of biological membranes are phospholipids, there are a myriad of other lipid species. The lipid composition varies according to the type of membrane but lipid distribution also strikingly differs between the two membrane leaflets. For example, phosphatidylserine and phosphatidylinositol occur mostly in the inner leaflet, while phosphatidylcholine, sphingomyelin and glycolipids are almost only in the outer leaflet of the membrane. This lipid asymmetry is generated during lipid biosynthesis and is maintained by dedicated enzymes – lipid flippases. Lipid asymmetry has an important role for the binding of various effector proteins of signalling pathways, such as protein kinase C that requires phosphatidylinositol for its activity, or proteins from phosphoinositide signalling. Free lateral movement of lipids and proteins is allowed by the fluid properties of biological membranes.

Membrane-associated proteins can be classified into peripheral membrane-proteins connected to the membrane by noncovalent interactions, partially anchored membrane proteins attached to the membrane via an amphipathic helix, glycosylphosphatidylinositol, fatty acid or prenyl chain, and integral membrane proteins also referred to as transmembrane proteins (Fig.2) that span the entire membrane via hydrophobic segments of their polypeptide chains (usually folded in α-helices, or less frequently β-sheets).As this study is focussed on intramembrane proteases and their membrane protein substrates, only in α-helical transmembrane proteins will be discussed further.

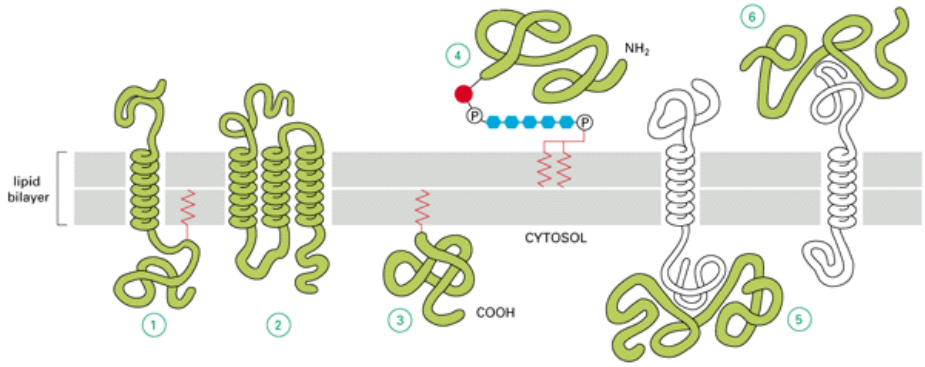


Figure 2: Membrane proteins are associated with lipid bilayer. Integral membrane proteins transverse the whole thickness of the membrane via their hydrophobic transmembrane helices, and are broadly divided into single-pass (1) or multi-pass (2) membrane proteins. Peripheral membrane proteins are bound to the membrane by a lipid anchor (prenyl group, fatty acid chain (3), or glycosylphosphatidylinositol (4)) or interact with the membrane or membrane proteins (5,6) noncovalently. Adopted from [1].

Transmembrane α-helices share a distinct distribution profile of amino acids along the membrane. Hydrophobic residues (Ala, Ile, Val and Leu) occur mostly in the middle of the membrane, aromatic residues (Tyr and Trp but not Phe) peak in the lipid-water interface regions and, not surprisingly, charged and polar residues are rare in transmembrane helices (Fig.3).

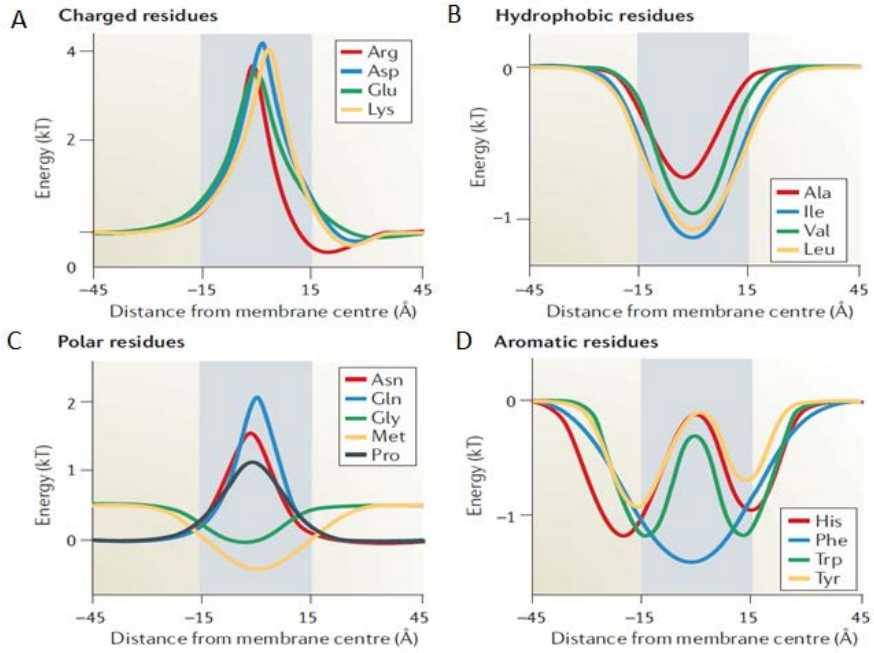


Figure 3: Lipid bilayer positioning preferences of amino acid within transmembrane helices of proteins. Graphs represent the localisations of divergent amino acids corresponding to their physiochemical properties (A-D). The central grey zone represents the hydrocarbon core of the membrane. Adopted from [84].

According to the "positive-inside" rule, positively charged residues (arginine and lysine) are found more frequently flanking the transmembrane helices at their cytoplasmic rather than non-cytoplasmic side [3][4]. This is partly due to the inherent bilayer asymmetry, where the negatively charged phosphatidylserine is the major lipid in the inner leaflet of the membrane. Whereas some lipids can ‘flip’ between the two leaflets of the bilayer, most

transmembrane proteins are stably integrated into the membrane in a specific orientation during their biogenesis.

The insertion of membrane proteins into the bilayer is mediated by a protein-conducting channel of the ER membrane (or plasma membrane in bacteria) called the translocon. While secreted proteins are fully translocated across the membrane, transmembrane helices of membrane proteins are instead shunted sideways into the lipid bilayer [5][6]. Their topology, that is, the number of transmembrane helices and their orientation in the membrane, is largely determined during their biogenesis and insertion into the membrane.Single-pass transmembrane proteins cross the membrane just once while multipass proteins are composed of several differently orientated transmembrane helices connected by loop regions that emerge from the membrane. Single-spanning membrane proteins are further classified according to their membrane orientation. Type I and type III membrane proteins translocate their N-terminus, while type II proteins translocate their C-terminus. As a result, the N-terminus of type I and III proteins is located in the periplasm or the ER lumen, while the N-terminus of type II proteins is cytoplasmic (Fig.4) [7]. Furthermore, type I membrane proteins also contain a cleavable, N-terminal signal sequence.

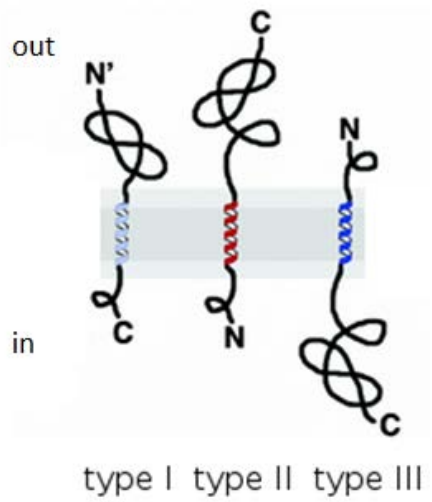


Figure 4: Three different types of membrane protein topologies. C-terminus of type I and type III membrane proteins is located in the cytoplasm, while their N-terminus faces the periplasm or ER lumen. Type I membrane proteins also have a cleavable N-terminal signal peptide. Type II proteins display opposite orientation. Adapted from [7].

2.1.2 Regulated proteolysis of membrane proteins at the cell surface - ectodomain shedding

Many signalling proteins are present at the cell surface as transmembrane precursors that require cleavage for activation or inactivation. Typical examples are growth factors (epidermal growth factor (EGF), transforming growth factor (TGFα)) or inflammatory cytokines (tumor necrosis factor (TNFα)). The process of cleavage of membrane protein domains and their liberation from the membrane is called ectodomain shedding [8]. It is mediated by two major protease groups, "a disintegrin and metalloproteinases" (ADAMs) and matrix metalloproteinases (MMPs). Ectodomain shedding is induced by many physiological and pharmacological stimuli such as cytokines [9] , growth factors [10], ceramide [11], cellular stress [12], calcium ionophores [13] but also by bacterial toxins [14], and has a downstream impact on other signalling pathways [15]. The sheddases are made in the form of zymogens that require regulated trafficking, propeptide cleavage and disulphide isomerisation for their activation [16]. They have been in the centre of pharmaceutical research for a number of years, but the regulation of their biogenesis and activation is not fully understood and new, unexpected regulatory principles have been discovered recently [17]. Metalloprotease sheddases were thought to be the major enzymes responsible for the cleavage of transmembrane protein precursors, until completely new and unexpected families of transmembrane proteases capable to cleave transmembrane protein substrates within the lipid bilayer were discovered.

2.2 INTRAMEMBRANE PROTEASES

Proteolysis is hydrolytic cleavage of peptide bonds, and it thus requires water molecules as reactants. It was therefore counter-intuitive and paradoxical that intramembrane proteases should catalyse this reaction inside the water-free interior of lipid membranes, and this concept was hard to accept. Nevertheless, several families of intramembrane proteases that do exactly this were discovered over the last 15 years. These unusual proteases use three different chemistries for catalysis. Site-2-protease (S2P) family are metalloenzymes, γ-secretase and signal peptide peptidases (SPPs) are aspartyl proteases, and rhomboids use a serine protease mechanism (Fig.5; page 16). Intramembrane proteases thus share similar

catalytic chemistry with, but are evolutionarily unrelated to their classical, soluble, globular counterparts.

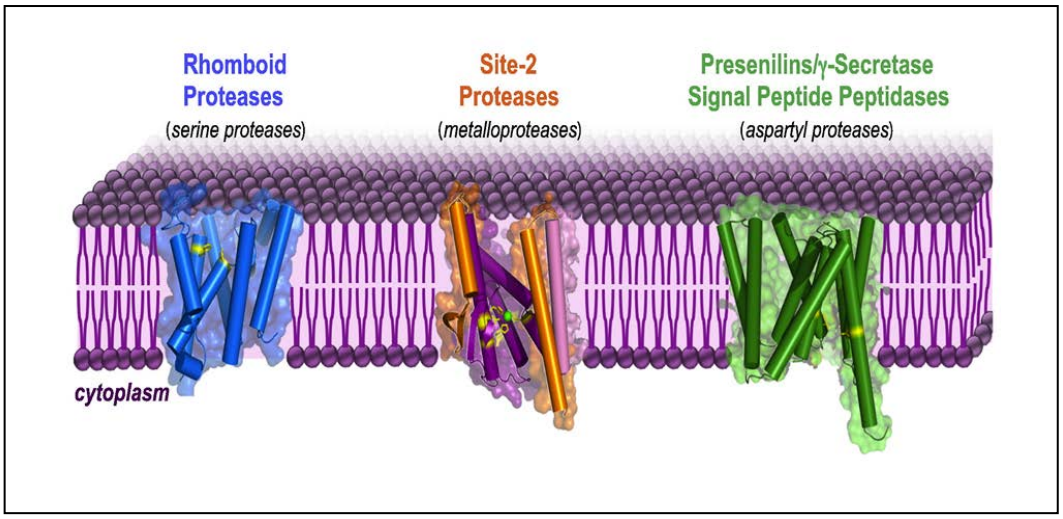


Figure 5: Crystal structures of prokaryotic homologs of the three catalytic types of intramembrane proteases are shown in different colours and shown embedded in a model membrane. *E.coli* rhomboid protease GlpG (PDB ID:2NRF), M.jannaschii S2P MJ0392 (PDB ID:4B4R) and M.marisnigri JR presenilin homologue (PDB ID: 4HYD). Adopted from [18].

Another paradoxical feature was the recognition and cleavage of rigid structure of transmembrane α-helices. Such secondary structures are generally refractile to proteolysis, and it was hypothesised that they need to be unfolded/destabilised during intramembrane proteolysis, but what determines substrates was long unclear. All initially identified substrates of intramembrane proteases were single-pass membrane proteins, but different intramembrane proteases were shown to have preferences for a particular substrate topology, which seemed correlated to the topology of their active site (Fig.6; page 17). Thus, S2P metalloproteases cleave type II single-pass membrane proteins, and most substrates of γ-secretase are type I transmembrane proteins. SPPs adopt opposite membrane topology compared to γ-secretases and they cleave substrates with opposite orientation, i.e. type II membrane proteins. Most rhomboid substrates are single-spanning helices of type I and III membrane proteins.

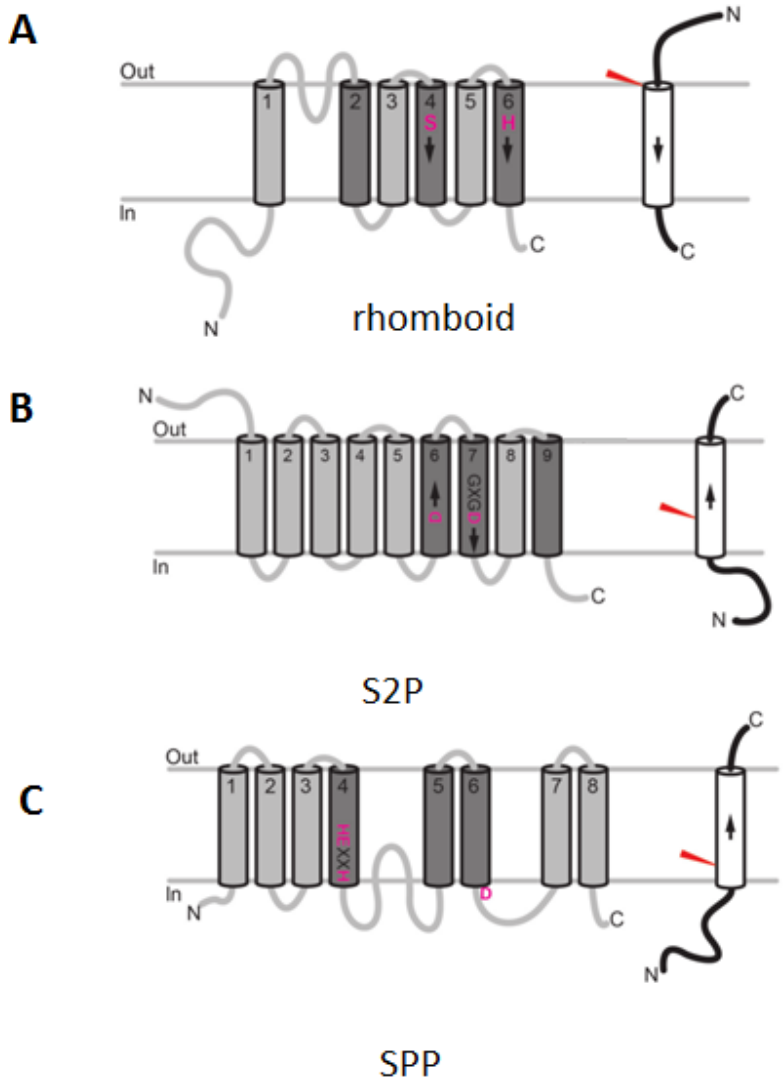


Figure 6: Catalytic types of intramembrane proteases. Schematic topological view of three catalytic types of intramembrane proteases; S2P metalloproteases (B), SPP aspartyl proteases(C), and rhomboids (A), serine proteases. Conserved regions are in dark grey, catalytic/active-site residues in magenta, and the cleavage of substrate is represented by red wedge. Adapted from [85].

2.2.1 Discovery and biological roles

Intramembrane proteases were discovered during studies of signalling, metabolism and human disease, and they have proven to be important biological regulators with a clear medical relevance. The first intramembrane protease was discovered during studies of the mechanism of sterol homeostasis. Sterol regulatory element-binding protein (SREBP) is a

transcription factor that regulates genes required for steroid lipid biosynthesis [19]. Mammalian SREBP is a transmembrane protein composed of transmembrane segment inside the membrane of endoplasmic reticulum (ER), an N-terminal transcription activator domain and a regulatory domain at the C-terminus [20]. It was found that when the level of cholesterol drops, two cleavages are required to activate SREBP [21]. First cleavage at "site-1" in its luminal loop region and then second cleavage at "site-2", which is located within its transmembrane domain. The site-2 cleavage was found to be catalysed by a novel ‘intramembrane’ metalloprotease termed site-2 protease [20]. Genomic studies revealed that S2Ps are found in all kingdoms of life, suggesting that they arose very early in evolution. Their functional range is also striking: all known S2P substrates are membrane-bound transcription factors regulating sterol biosynthesis, envelope stress responses in bacteria or ER stress response (unfolded protein response) [22]. Recently, mutations in the S2P gene were associated with human genetic diseases such as ichthyosis follicularis, alopecia and photophobia syndrome [23].

The aspartyl intramembrane protease γ-secretase was discovered independently during the study of Alzheimer’s disease (AD) as a gene whose mutations were associated with an early-onset familial AD (therefore the name presenilin for the catalytic subunit of the enzyme) [24], and during studies of developmental signalling via Notch receptors [25].The most striking roles of γ-secretase are the cleavage of the amyloid precursor protein functioning in Alzheimer’s disease and the participation on Notch signalling pathway which is critical for proper cell proliferation, positioning, differentiation and survival [25]. However, to date γ-secretase has more than 90 known substrates [26], and it appears to regulate also a variety of other cellular events such as cell migration, adhesion and cell fate determination. It participates also in the regulation of neurite outgrowth, axon guidance or formation and maintenance of synapses which are the most disrupted events during neurodegeneration [27].

Another highly conserved group of aspartyl intramembrane proteases are SPPs that are related to presenilin, the active subunit of the γ-secretase complex. The original SPP was isolated as a protein required for the degradation of signal peptides that had been cleaved-off by signal peptidase [28]. SPPs are absent from bacteria, while they are very common in higher organisms. Interestingly, the numbers of SPP paralogues differ among organisms leading to the hypothesis that they have evolved different functions. SPPs are implicated in immune surveillance by presentation of peptides via major histocompatibility

complex class I or by HLA-E molecules [29], they regulate B cell maturation [30], and promote cleavage and disposal of the cleaved-off signal peptides from the membrane [31].

Rhomboid discovery was rooted in developmental studies of *Drosophila melanogaster* (Fig.7). Genetic screens of embryonal patterning defects in the fruitfly identified a number of mutants with developmental abnormalities. Their subsequent analysis has led to a discovery of a number of developmental pathways (which won Christiane Nusslein-Volhardt and Eric Wieschaus a Nobel Prize) including the EGF receptor pathway. Several mutants clustered together phenotypically, including one with a rhomboid-shaped head skeleton.

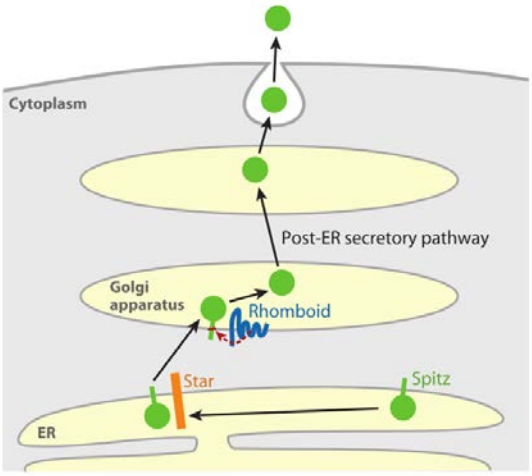


Figure 7: Biological function of *Drosophila* Rhomboid-1. Rhomboid-1 is localised in the Golgi apparatus, while its substrate Spitz, a transmembrane precursor of the EGF receptor ligand, is held in the endoplasmic reticulum (ER). The two do not meet until a third component, the trafficking chaperone Star, escorts Spit to the Golgi apparatus. Rhomboid-1 then cleaves Spitz in its transmembrane domain to release its signalling ectodomain so that it can activate the EGF receptor on neighbouring cells. Adopted from [86].

The *rhomboid* gene has ultimately been shown to be the main activator of the epidermal growth factor receptor (EGFR) pathway that acted in the signal sending cell as a catalytic entity promoting the release of EGF family growth factors that activate the EGF family growth factors from their transmembrane precursors, which can then activate the EGF receptor on neighbouring cells. The sequence of the *rhomboid* gene did not resemble any known genes, nor did it reveal any functional motifs except for the uninformative seven transmembrane domains. However, further research revealed that the EGFR ligands appeared to be cleaved in their transmembrane region, that Rhomboid was sensitive to some serine protease inhibitors, and Rhomboid activity depended on conserved serine,

histidine and asparagine located in transmembrane helices. These observations have collectively led to the proposal that rhomboids are novel intramembrane serine proteases [32][33].

2.2.2 Rhomboid-family proteins

Rhomboids are the largest group of intramembrane proteases and are highly conserved in all kingdoms of life. Rhomboid proteases use a variation of serine protease mechanism, having a catalytic dyad of serine and histidine, as opposed to the typical catalytic triad of serine, histidine and aspartate of classical, soluble serine proteases. Besides rhomboid proteases (further subdivided into secretases and PARLs [34]), rhomboid-like superfamily also comprises proteolytically inactive rhomboid-like proteins that are increasingly attracting interest (iRhoms and other inactive homologues [35], see later in this chapter) (Fig.8). The architectures of the rhomboid-like transmembrane domain are represented by the diversity of Rhomboid proteases. They share the transmembrane core domain made of six transmembrane helices (TMH) (6TMH, majority of bacterial rhomboids) but may possess one extra transmembrane helix either at the N-terminus (1+6TMH, PARL) or at the C-terminus (6+1TMH, AarA). In the next few paragraphs, a few examples of the functions of rhomboid-family proteins will illustrate their biological significance.

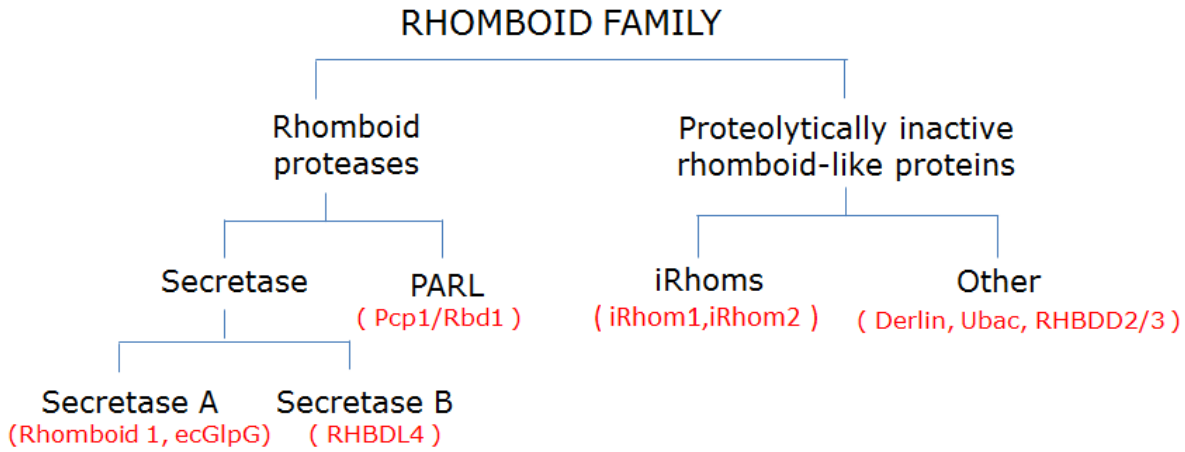


Figure 8: Rhomboid-family proteins. Rhomboid family comprises two big subgroups of proteins. The catalytically active rhomboid proteases are further subdivided based on their localisation and secondary structure. Rhomboid-like proteins have lost the protease activity but are still able to bind client proteins.

Concurrently with the work in *Drosophila*, genetic evidence revealed that a bacterial rhomboid protease is regulating intercellular signalling (also called quorum sensing [36]) in the Gram-negative opportunistic pathogen *Providencia stuartii*, which was quite startling as it suggested that rhomboids might be specific signalling enzymes. The molecular mechanism of AarA function in quorum sensing was not revealed until several years later. AarA cleaves its natural substrate, the TatA protein (a type III membrane protein), to activate a protein translocation machinery, the TAT system, which then produces an extracellular signal (Fig.9). Surprisingly, AarA and *Drosophila* Rhomboid-1 could complement each other functionally, which suggested that these two enzymes had overlapping substrate specificities [37][38].

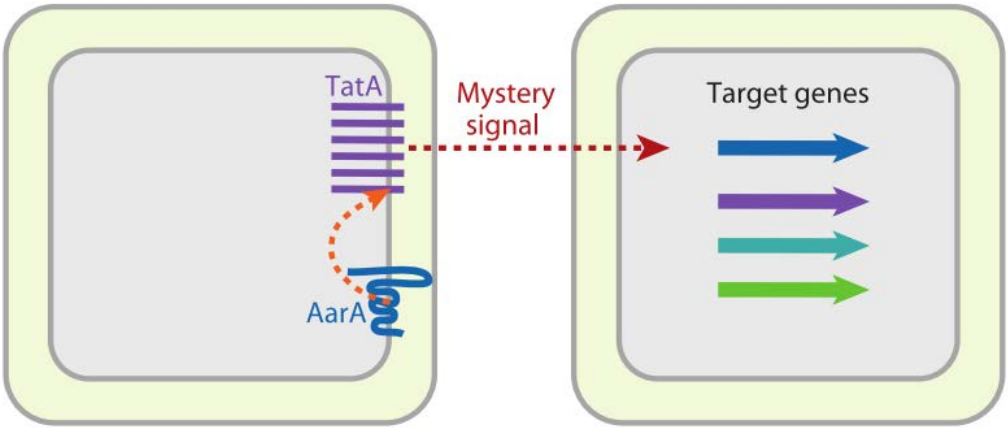


Figure 9: Biological function of the AarA rhomboid protease in *Providencia stuartii*. AraA constitutively cleaves its substrate TatA to activate a protein translocase that secretes a quorum sensing signal. The identity of the signal remains unknown. Adopted from [86].

The role of rhomboids in EGFR signalling is partially conserved also in *Caenorhabditis elegans* [39], but later research showed how very versatile their functions are. A mitochondrial rhomboid protease Rbd1 in yeast *Saccharomyces cerevisiae* regulates mitochondrial dynamics. The two physiological substrates of Rbd1, a dynamin-like GTPase Mgm1 and cytochrome c peroxidase Ccp1, are located in the inner mitochondrial membrane. The GTPase Mgm1 is involved in mitochondrial fusion (Fig.10; page 22); it has a transmembrane form and soluble form located in the intermembrane space, where the latter is generated by cleavage by Rbd1. The two forms must be in a roughly equimolar ratio to allow proper mitochondrial dynamics, and hence deletion or overexpression of Rbd1 both disrupt mitochondrial function [40].

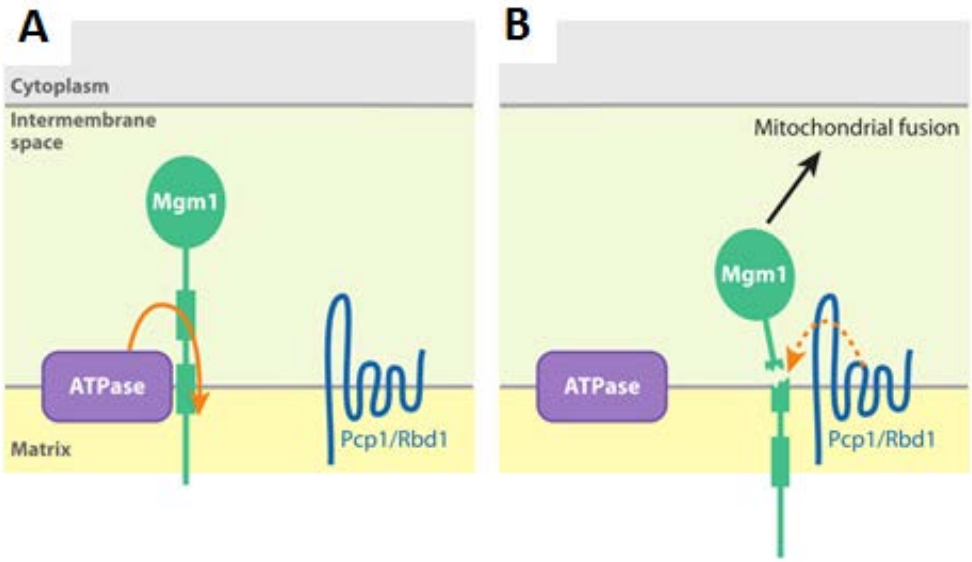


Figure 10: Role of rhomboid in the regulation of mitochondrial dynamics. Prior to the cleavage of Mgm1 by mitochondrial rhomboid protease Rbd1 (b) the ATPase mediated translocation of Mgm1 is required (a). This regulated cleavage influence mitochondrial dynamics and proper functioning of mitochondria. Adopted from [86].

The mouse homolog of the mitochondrial rhomboid called PARL has a slightly different function. Instead of affecting mitochondrial dynamics, mouse PARL acts as a suppressor of apoptosis. In response to cytokines, a Bcl-2 family protein Hax1 presents PARL with its substrate HtrA2. The PARL-cleaved soluble form of the HtrA2 protease is localised in the intermembrane space of mitochondria, preventing accumulation of pro-apoptotic protein Bax [41], which is important for the survival of lymphocytes and neurons. Mammalian genomes encode usually four bona fide rhomboid proteases that localise to the secretory pathway, RHBDLs 1-4. Of these, RHBDL2 is implicated in cell migration during wound healing [42], blood clotting [43] and EGFR signalling [44]. RHBDL4 has been suggested to be involved in apoptosis [45] and membrane protein quality control [46], while RHBDL1 and 3 have no known substrates, although they have all the sequence hallmarks of active rhomboid proteases.

Rhomboid proteases have also been intensely studied in eukaryotic apicomplexan parasites *Plasmodium* and *Toxoplasma* [47]. In this case, rhomboid-driven cleavage of adhesins, which are responsible for the first contact between the parasite cell surface and the host cell, is a crucial step in parasite invasion into the host cell (reviewed in [48]), which suggests that rhomboids could be medically relevant (Fig.11; page 23). Candidate

adhesin substrates have been identified to be cleaved efficiently by some parasite rhomboids. The diversity of rhomboids present in parasites suggests functions beyond invasion, for example *Toxoplasma gondii* rhomboid ROM1 appears to influence the intracellular growth of the parasite [49].

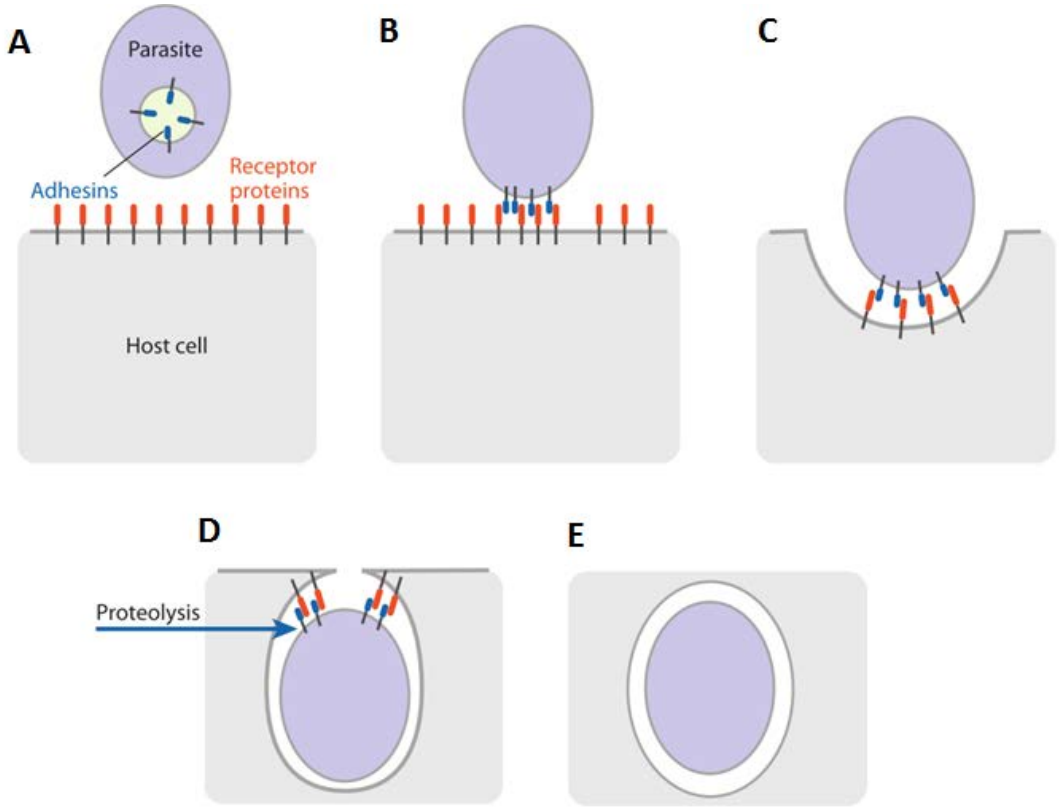


Figure 11: Host cell invasion by apicomplexan parasites. A) First, parasite's adhesins are delocalised on the cell surface to allow binding on host cell receptors (B). Parasite cells are then invaginated (C) and the whole process is terminated by the cleavage of adhesins (D). Adapted from [86].

In addition to the active rhomboid proteases, a number of homologs that apparently lost their catalytic residues – rhomboid pseudoproteases - are found in eukaryotic genomes (Fig.12; page 24). The iRhoms lack proteolytic activity because they lack of one or both active residues and have a conserved proline in a position just before the usual position of the active site serine, which potentially structurally disturbes the active site architecture [34]. Another characteristic feature of iRhoms is their extended N-terminal extramembrane domain and highly conserved loop connecting TMH1 and 2 known as iRhom homology domain (IRHD).

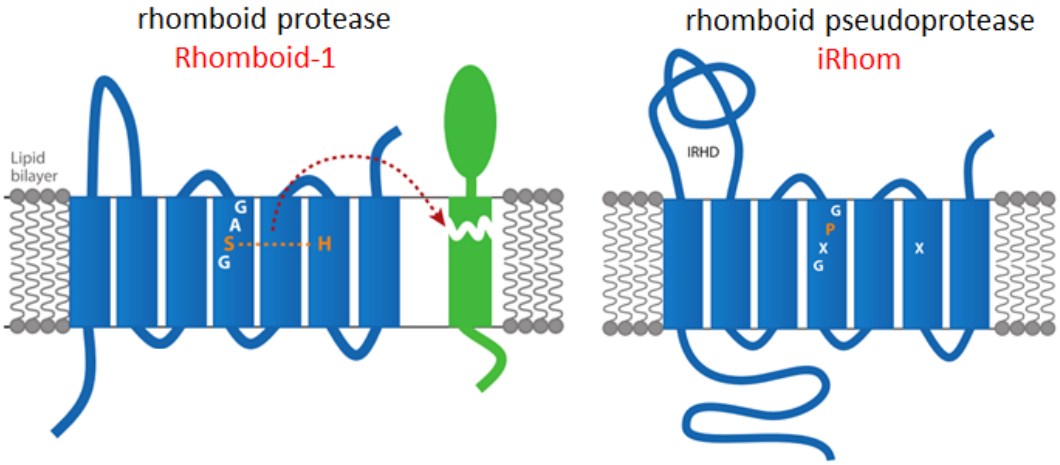


Figure 12: Topological similarities and differences between rhomboid proteases and the iRhoms. Compared to their proteolytically active rhomboid counterparts, the pseudoproteases iRhoms lack one or both catalytic aminoacids (represented in orange in picture of Rhomboid-1 and by X in picture of iRhom). Moreover, iRhoms have additional proline nearby that disrupts the active site architecture in rhomboids (orange P in iRhom figure) A highly conserved region called iRhom homology domain (IRHD), which is topologically equivalent to the L1 loop of rhomboids, is characteristic for iRhom group of rhomboid-like proteins. Adapted from [86].

The iRhoms are highly conserved in metazoans, which implies important functions. Indeed the *Drosophila* and mouse iRhoms participate in regulating the EGFR pathway in interesting, but different ways. *Drosophila* iRhom downregulates the levels of EGF receptor ligands by shunting them into the ER associated degradation, while mouse iRhoms 1 and 2 are essential transport chaperones required for the maturation and activation of a membrane-tethered metalloprotease (ADAM17/TACE) whose role is to catalyse ectodomain shedding of many membrane proteins including some EGF receptor ligands and tumor necrosis factor TNFα.

Apart from iRhoms, there are also rhomboid-like pseudoproteases Derlins, TMEM151 and UBAC2. Their functions are less well understood, but it seems that they are all involved in membrane protein quality control in the ER [50][51][52]. Mechanistically, almost nothing is known about rhomboid pseudoproteases, but given their sequence and expectedly architectural/structural similarities to active rhomboids, the latter serve as useful initial structural and mechanistic models for the former.

The key steps to a better understanding of the biological functions of intramembrane protease families involve biochemical and structural analyses. These are very conveniently

conducted using model prokaryotic homologues of intramembrane proteases, because they are minimal architectural models and are usually easily available in recombinant form.

2.3 RHOMBOID PROTEASE GLPG FROM *E.COLI*, THE MAIN MODEL INTRAMEMBRANE PROTEASE

Rhomboid GlpG from *Escherichia coli* has come to the fore in intramembrane protease research as it was the first intramembrane protease whose structure has been solved. GlpG represents the simplest rhomboid architecture with six transmembrane helix core harbouring the catalytic unit. This core domain is expected to be similar in the rhomboid-like family, which makes GlpG a useful minimal model for the rhomboid-family proteins. In fact, GlpG has become the main model intramembrane protease and from a mechanistic perspective it is now the best understood one.

2.3.1 Structure of GlpG

Despite the accumulating biochemical, genetic and cell biological data on intramembrane proteases, the structural confirmation of their intramembrane remained unanswered until 2006 when a high-resolution (2.1 Å) crystallographic structure of the transmembrane core domain of *E. coli* rhomboid intramembrane protease GlpG was published [53], followed by the structure of *H. influenzae* GlpG [54]. The globular extramembrane domain at the N-terminus of GlpG was shown to be dispensable for activity [55] and so the structural understanding of the transmembrane core of GlpG is sufficient for understanding of its intramembrane cleavage ability.

The core domain of GlpG consists of six α-helices where the shorter TMH4 is in the centre of the bundle, surrounded by the other five TM helices (Fig.13; page 26). Importantly, one end of TMH4 is recessed about 10 Å below the membrane surface. This creates hydrophilic water accessible indentation at the bottom of which, at one end of TMH4, lies the catalytic serine 201. The active site is thus clearly buried below the level of the membrane, but water can apparently freely access it via the water-accessible indentation [56], which explains one mystery of intramembrane proteolysis. The second prominent structural characteristics of GlpG is the helical hairpin loop L1 connecting

TMH1 and 2 that extends sideways into the outer leaflet of the lipid bilayer. The L1 loop is conserved in rhomboids, it is very sensitive to mutations, and important for protease activity. However, its precise mechanistic role is unknown. According to a recent large-scale thermodynamics analysis, GlpG structure is stabilised mostly by van der Waals interactions and only by two peripheral hydrogen bonding clusters [57]. Glutamate 166, positioned near the cytosolic face of TMH2, tethers TMH1 and TMH3 by hydrogen bonds creating an apex at the bottom of the GlpG, and Asp268 stabilises GlpG by linking TMH3 and TMH4 to TMH6. Moreover, hydrogen bonds in L1 are also involved in GlpG stabilisation via histidines His141 and His145. Furthermore, packing interactions mediated by a GXXXG motif stabilise TMH4 and 6, and side-chain packing interactions between loop 3 and the TMH4 stabilise the transmembrane domain bundle. The finding that GlpG is stabilised mostly by a number of weak interactions has been interpreted as enabling the protease structure to be flexible and responsive to environment [57].

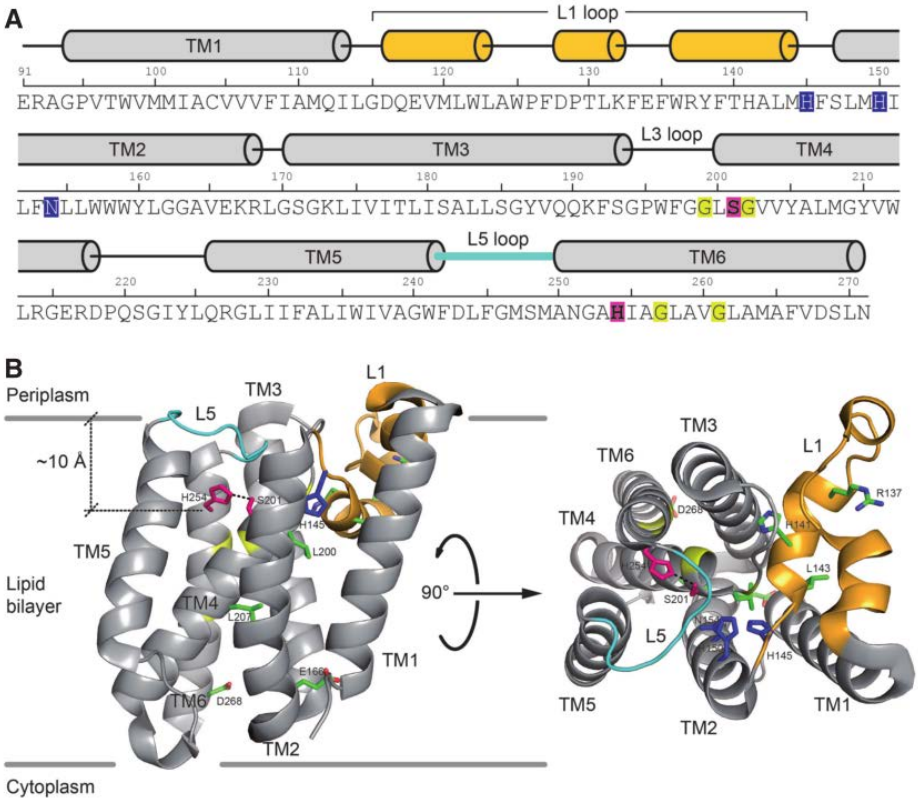


Figure 13: *E.coli* GlpG sequence and X-ray structure. A) Sequence of *E.coli* GlpG transmembrane helices are noticed by grey cylinders (TM1-TM6), conserved amino acids are represented by yellow and blue colour and active site amino acids are highlighted by pink. B) Lateral and top view of GlpG X-ray structure (PDB ID: 2XOV**)** is visualised using corresponding colours as in panel A. Active site (pink) is buried ~10 Å inside the lipid bilayer showing hydrogen bond (black dashed line) between S201 and H245. Amino acids important for thermodynamic stability of GlpG are represented by sticks in green. Adopted from [85].

The X-ray structures had revealed the conformation of the rhomboid protease, proven its intramembrane character, explained how water can access the site of proteolysis, but did not explain the mechanism how a substrate may interact with it and enter the active site. Since the catalytic serine is insulated from the lipid bilayer (where the transmembrane substrate resides) by the five transmembrane helices of GlpG, some conformational change is necessary in the protease to allow substrate access to the catalytic dyad.

2.3.2 Substrate access to the active site

According to two mechanistic principles, different models were proposed. Initially, based on the first structure of GlpG, it was hypothesised that substrate enters the active site within a large V-shaped gap between TMH1 and TMH3, gated by L1 loop. Opening of the L1 gate would however require a massive conformational change of the L1 loop, and possibly also of the entire active site [58] [59]. Considering the apparently high rigidity of L1 (low B-factors in X-ray structures), and the necessity of such a wide change in structure, this model was ultimately deemed unlikely.

The other two models were based on the observed conformational heterogeneity in the crystalline state that was observed in several structures determined with different detergents and crystallized in different space groups. Firstly, the loop 5 (connecting TMH5-TMH6), also called the ‘cap’, as it occludes the active side from the periplasmic side of membrane in the unliganded enzyme, was found disordered in several structures and in two conformations ("open" and "closed") in others [58]. Secondly, one structure observed similar heterogeneity in the TMH5. This gave rise to two models where substrate enters the active site via a portal between TMH5 and TMH2 uncapped by L5. According to these hypotheses the enzyme can adopt two conformational states - an open and a closed form. The open form might be generated by the movement of TMH5 away from TM2, causing a displacement of L5 (Fig.14; page 28). Or, alternatively, the opening of the enzyme might be masterminded by the lifting of L5 and removal of the bulky Phe245 from the active site entrance [59].

Although the exact details of the conformational changes accompanying substrate binding remain speculative, the implication of TMH2 and 5 in substrate entry was supported by further structural, enzymatic and biophysical studies. The region between TMH2 and 5 was shown to bind phospholipids and detergents, which predicts its exposure

to the hydrophobic membrane [60]. Furthermore, mutations in TMH2 and TMH5 can increase enzymatic activity up to 40-fold [61] while not impacting on the thermodynamic stability of GlpG [57]. However, there are two contradictory studies addressing the hypothesis of TMH5 working as a moveable ‘lateral gate’. One study suggested that restraining the mobility of the ‘gate’ by crosslinking TMH2 and TMH5 via disulphide formation between uniquely introduced cysteine’s impaired activity of GlpG , while a separate study from a different laboratory, using a different crosslinking method that was well controlled, did not show significant inhibition of substrate cleavage [62]. Despite the apparent controversy about the role of TMH5 as a moveable lateral gate, some role of TMH2 and 5 regions in substrate entry or binding seems to be consensual. Anyway, all proposed models remain hypotheses until the first structure of GlpG with a transmembrane substrate bound in its active site will be solved.

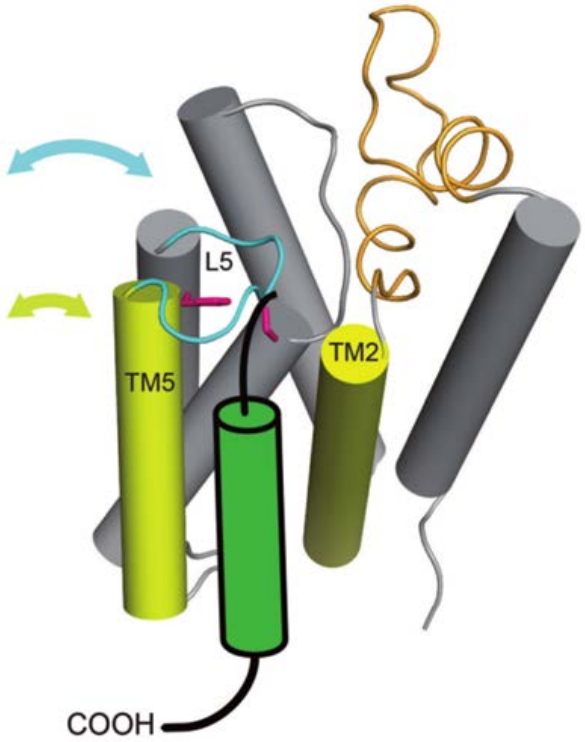


Figure 14: Potential models of substrate entrance into the active site. Loop 1 (L1, orange) is visualised as a former hypothesised substrate entrance site. Two recent models of potential conformational changes are indicated by flashes. First, displacement of loop 5 (L5, blue) and second, shift of transmembrane helix 5 (TMH5, yellow) sidewise from transmembrane helix 2 (TMH2, yellow). Latter is implied in opening of ‘lateral gate’ which might serve as a substrate entrance. Substrate is represented in green. Adopted from [85].

2.3.3 Catalytic mechanism

Rhomboids use a variation of serine protease mechanism possessing only a catalytic dyad of Ser201 and His154, whose mutations impair enzymatic activity. Although rhomboids are evolutionarily unrelated to the classical, soluble serine protease, the basic chemistry of proteolytic reaction will likely be similar, and we can use this conceptual framework for the model description of rhomboid cleavage mechanism (Fig.15). As with classical serine proteases, the reaction starts by the nucleophile attack of the carbonyl carbon of substrate by the hydroxyl of the catalytic serine. The resulting oxyanion is stabilised by the 'oxyanion hole' of the enzyme. Decomposition of this first intermediate yields the C-terminal cleavage product and a covalent intermediate, the acylenzyme. The catalytic histidine then activates a water molecule, which then attacks the acyl enzyme. The resulting second tetrahedral intermediate then collapses leading to the release of the N-terminal cleavage product and regenerated enzyme. Based on molecular dynamics simulations and consequent enzymatic experiments it was proposed that a site near the catalytic serine serves as a 'water retention site' that facilitates channelling of water molecules to the catalytic histidine in the second step of the catalytic cycle.

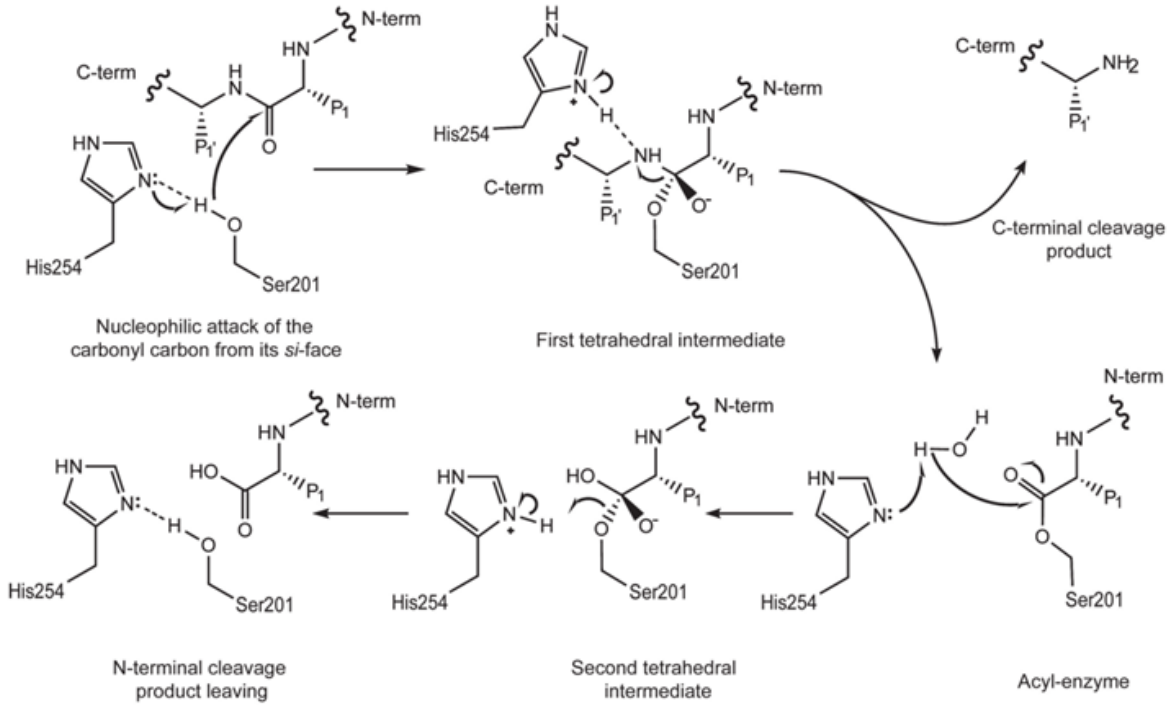


Figure 15: Rhomboid protease cleavage mechanism. Hydrogen bonds are shown as dotted lines, free electron pairs as double dots and bond rearrangements as curved arrows. Adopted from [85].

The precise structural changes in GlpG accompanying the catalytic cycle are unknown, but some initial clues have emerged from structural studies with small molecular inhibitors. The structure of GlpG with inhibitors diisopropyl fluorophosphate (DFP) [63] and isocoumarin [64] confirmed the formation of oxyanion hole made by Ser201 together with the side chains of His150 and Asn154 [65]. This is quite different from soluble serine proteases where the oxyanion hole is formed by main-chain amides. Another striking difference is the stereochemistry of nucleophilic attack of the prochiral carbonyl carbon of the scissile peptide bond [65]. Compared to the classical serine proteases, rhomboids are thought to attack the scissile bond from its *si*-face which leads the opposite configuration in the first intermediate. Lastly, movements of several amino acid side-chains have often been observed in inhibitor complex structures. The side chain of Tyr205 which was shown to interact with catalytic histidine, is rotated in the DFP or isocoumarin complexes, and rotation of Trp236 and Phe245 was also suggested important for the catalytic process [66].

2.3.4 Rhomboid substrate recognition

Two major questions arose with the discovery of intramembrane proteases. First, how the enzyme manages the retention of water in the hydrophobic surrounding and second, how transmembrane helical substrate accesses the active site and are cleaved. The first question was solved by X-ray crystallography and was described in the above sections, while the second one was addressed by multiple studies presented in this section.

Alpha helices are not easily accessible to the active sites of proteases, and it was thus hypothesised that destabilisation/unfolding of TMHs of substrates is required prior to proteolysis by intramembrane proteases. In line with this hypothesis, the very first analysis of rhomboid specificity showed that the replacement of first five TMH residues of a chimeric substrate by those of non-substrates abrogated substrate cleavage completely [47]. Conversely, introduction of destabilising residues (Gly, Thr, Ile) into non-substrates allowed cleavage by a panel of rhomboids (*Drosophila* Rhomboids1-3, YqgP of *B. subtillis*, *P. stuartii* AraA, and less efficiently by GlpG) [67] . However, these studies were using only cellular assays, making the analysis of cleavage sites and cleavage kinetics, which are crucial in investigation of enzyme mechanism and substrate specificity, difficult to interpret.

Reconstitution of rhomboid activity *in vitro* allowed more rigorous mechanistic studies. Akiyama et al. performed extensive domain-swapping experiment between a rhomboid substrate and non-substrate. The substrate was made of the artificial model substrate LacYTMH2 (the TMH2 of lactose permease), that had been shown to be cleaved by rhomboids previously [68], while the non-substrate was LacY TMH6 [69]. In this way they eventually focussed on the analysis of importance of TMH destabilising residues for substrate cleavage. It was found that TMH destabilising residues in substrate TMH improve cleavage by GlpG, but, surprisingly, even if they are very far from the actual cleavage site.

The discovery of the first natural substrate of a bacterial rhomboid protease enabled a comprehensive study of rhomboid substrate specificity [70]. Systematic comparative analysis using *in vitro* kinetics and *in vivo* steady state cleavage experiments revealed that rhomboids recognise two elements in substrates, the transmembrane domain and a short sequence motif that determines the position of the cleavage site. This recognition motif, identified in four known rhomboid substrates, consists of small residue in the P1 position and large hydrophobic residues in the P4 and P2' positions. Site-specificity of the AarA rhomboid was not affected even if moving this recognition motive up to 7 residues into the juxtamembrane region, suggesting that the position of cleavage does not depend on the position of destabilising residues. A similar motif was recognised by three different bacterial rhomboids (AarA, GlpG, YqgP) in four different substrates (TatA, Gurken, Spitz and LacYTMH2), which suggests that it could represent a relatively widespread property of rhomboid proteases. Strikingly, the TMH destabilising residues were indispensable when the cleavage occurred within or near the end of the TMH, but they were less important when cleavage site of TatA was moved out of the membrane into the juxtamembrane region by a linker.

The character of the recognition motif is relatively degenerate, and it is thus surprising that rhomboids nevertheless seem to be quite specific proteases. This implies that rhomboids may exert also significant recognition selectivity beyond the recognition motif, in the transmembrane region of the substrate. Besides the above described study [70], this idea is supported also by a recent study showing that RHBDL4 requires positively charged amino acids deep in substrate TMH [46]. A recent mechanistic model proposed, taht the first contact between the substrate and the enzyme occurs inside the membrane, distant from the cleavage site, at an intramembrane 'exosite' of rhomboid [70].

This model has been recently disputed [71] [72] , and it is currently unclear what proportion of substrate's TMH actually interacts with rhomboid, if rhomboids indeed have an intramembrane exosite and what is its character. An obvious hypothesis is that the exosite is formed by TMH2 and 5 protease since the mutations at these helices were reported to increase cleavage efficiency [66], mutations in TMH5 do not destabilise the protease thermodynamically [57], and the sequence of TMH5 is the most variable among GlpG homologues [34], possibly reflecting differences in substrate specificity (Fig 16).

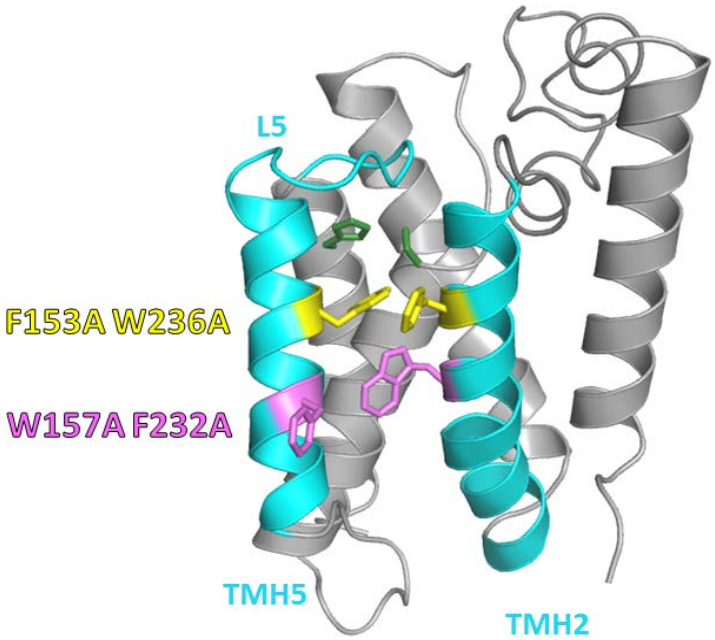


Figure 16: Hyperactive double mutations of GlpG. Two double mutants F153A/W236A (yellow) and W157A/F232A (pink) were designed to weaken the interaction between TMH5 and TMH2 which might led to the open conformation of the enzyme.TMH5 is displaced from TMH2 and thus also affects the position of loop 5 (L5) (all represented in blue). This conformational change exposes amino acids of the active site (green) to the substrate.

3 AIMS AND OBJECTIVES

This thesis focuses on addressing mechanistic hypothesis about substrate recognition by rhomboid protease GlpG using comparative enzyme kinetics and site-directed mutagenesis. Specifically I ask the following questions: What is the extent of substrate's transmembrane helix interaction with the enzyme? How important is the TMH interaction for substrate recognition? Where on the enzyme is the binding site for substrate TMH and which parts of substrate's TMH might be interacting with it? Answering these questions will further our understanding of intramembrane proteolysis by rhomboids and of membrane protein interactions in general, it will help our effort to characterise rhomboid-substrate interaction structurally, it will facilitate the design of rhomboid inhibitors and may improve our ability to predict rhomboid substrates, which would in turn facilitate the understanding of biological functions of rhomboids.

The main objectives of this thesis are:

- Expression and purification of four variants of GlpG: the wild-type enzyme, its inactive mutant S201A, and hyperactive double-mutants F153A/W236A and W157A/F232A.

- Expression and purification of four different substrate fusion proteins (TatA, Gurken, Spitz and LacYTMH2).

- Determination of the extent of interaction between substrate's TMH and GlpG by quantifying the effect of shortening the length substrate's TMH on cleavage kinetics and binding constant.

- Quantification of the activation effect of two reportedly hyperactive double mutants of GlpG F153A/W236A and W157A/F232A, on four different substrates that also differ in the sequence of their transmembrane helix.

- Design of a widely useable fluorogenic transmembrane peptide substrate for rhomboids.

4 MATERIALS AND METHODS

4.1 CHEMICALS, BUFFERS AND OTHER MATERIAL

2-Amino-2-hydroxymethyl-propane-1,3-diol (Tris)	Promega (USA)
2-mercaptoethanol	Sigma-Aldrich, USA
acetic acid	Penta, Czech Republic
acetonitrile	Penta, Czech Republic
acrylamide	USB, USA
ampicilin (Amp)	Sigma-Aldrich, USA
amylose resin	New England BioLabs, USA
boric acid	Penta, Czech Republic
bromphenol blue	Serva, Germany
E. coli C41(DE3)	John E. Walker, Cambridge, UK [73]
cobalt(II) chloride	Sigma-Aldrich, USA
Complete™, EDTA free	Roche, Switzerland
dimethylsulfoxide (DMSO)	Sigma-Aldrich, USA
D-Maltose	Sigma-Aldrich, USA
ethanol	Lach-Ner., Czech Republic
ethylendiamintetraacetic acid (EDTA)	Sigma-Aldrich, USA
glycerol	Penta, Czech Republic
glycine	USB, USA
hydrochloride acid	Penta, Czech Republic
imidazole	Sigma-Aldrich, USA
InstantBlue Coomassie™	Expedon, UK
isopropanol	Penta, Czech Republic

isopropyl-α-D-thiogalactopyranosid (IPTG)	Biosynth AGm, Switzerland
LB broth	Sigma-Aldrich, USA
methanol	Penta, Czech Republic
N-2-hydroxyethylpiperazine-N-2--ethane sulfonic acid (HEPES)	Sigma-Aldrich, USA
n-Dodecyl-β-D-maltoside (DDM)	Affymetrix, USA
nickel(II) sulphate NiSO4	Sigma-Aldrich, USA
o-phthaldialdehyde (OPA)	Sigma-Aldrich, USA
phenylmethylsulphonyl fluorid (PMSF)	Sigma-Aldrich, USA
potassium dihydrogen phosphate	Sigma-Aldrich, USA
potassium hydroxide	Sigma-Aldrich, USA
Protein Ladder 10-205 kDa	New England BioLabs, USA
SingleQuant Assay kit	Serva, Germany
sodium acetate	Penta, Czech Republic
sodium chloride	Penta, Czech Republic
sodium dodecyl sulphate (SDS)	Sigma-Aldrich, USA
sodium hydroxide	Penta, Czech Republic
sodium phosphate dibasic	Sigma-Aldrich, USA
tricine	Sigma-Aldrich, USA
Triton X 100	Sigma-Aldrich, USA
Trizma^{TM} (Tris base)	Sigma-Aldrich, USA

SDS-PAGE sample buffer (6x):
3.5 mM Tris; 30% (v/v) glycerol; 1% (w/v) SDS; 6% (v/v) μL 2-merkaptoethanol; 2 μM bromfenol blue; water up to 10 mL; pH 6.8

SDS-PAGE running buffer:
Tris-Glycine : 25 mM Tris, 250 mM glycin, 0.1% (w/v) SDS

Tris-Tricine: 10 mM Tris base, 10 mM Tricine, 0.1% (w/v) SDS

TALON™ W buffer:
20mM HEPES-NaOH pH 7.4 ; 10% (v/v) glycerol ; 300 mM NaCl; 25 mM imidazole ; 0.05% (w/v) DDM

TALON™ E buffer:
20 mM HEPES-NaOH pH 7.4 ; 10% (v/v) glycerol ; 300 mM NaCl; 75 mM imidazole ; 0.05% (w/v) DDM

Buffer A:
20 mM HEPES-NaOH pH 74 ; 10% (v/v) glycerol ; 100 mM NaCl

Isolation buffer B:
20 mM HEPES-NaOH pH 7.4 ; 10% (v/v) glycerol ; 300 mM NaCl; 10 mM imidazole

nickel-nitrilotriacetic acid (NiNTA) equilibration buffer:
20 mM HEPES-NaOH pH 7.4 ; 10% (v/v) glycerol; 300 mM NaCl; 10 mM imidazole ; 0.05% (w/v) DDM

NiNTA W buffer:
20 mM HEPES-NaOH pH 7.4 ; 10% (v/v) glycerol ; 300 mM NaCl; 35 mM imidazole ; 0.05% (w/v) DDM

NiNTA E buffer:
20 mM HEPES-NaOH pH 7.4 ; 10% (v/v) glycerol ; 300 mM NaCl; 100 mM imidazole ; 0.05% (w/v) DDM

Microscale thermophoresis (MST) buffer:
25 mM Tris; 150 mM NaCl; 0.05% (w/v) DDM

Phosphate buffer saline (PBS):
10 mM Na_2HPO_4; 1.8 mM KH_2PO_4; 137 mM NaCl; 2.7 mM KCl

All buffers were filtered using 0.22 μm membranes (Millipore).

LB medium: 20 g of LB Broth (Sigma-Aldrich) into 1 L of deionised water, autoclaved

4.2 INSTRUMENTS

ÄKTA Explorer FPLC	Amersham Pharmacia Biotech, Sweden
ÄKTA Prime	Amersham Pharmacia Biotech, Sweden
Allegra X-15R	Beckman Coulter, USA
Avanti J-30I	Beckman Coulter, USA
Beckman J2-MI,	Beckman Coulter, USA
Biofuge Pico	Hereaus Instruments, Germany
Ellectrophoreses cell	BIO-RAD, France
EmulsiFlex®-C3	Avestin, Canada
Eppendorf centrifuge 5424	Eppendorf, Germany
EPSON perfection scanner V37	EPSON, Japan
HiLoad™ 16/60 Superdex™ 200	GE Healthcare, UK
Innova 4300 rotary incubator	New Brunswick Scientific, USA
KS 260 basic shaker	IKA, Germany
Megafuge 2.0R	Hereaus Instruments, Germany
microbiological incubator	Memmert GmbH, Germany
MF Millipore membrane	Millipore, USA
MonolithNT.LabelFree	Nanotemper technologies, Germany
NanoDrop ND-1000 spectrometer	Thermo Scientific, USA
Ni-NTA HiTrap Superflow 1 mL	Quiagen, USA
Optima L-90K	Beckman Coulter, USA
pH meter Unicam 9450	Unicam, UK
pH meter Unicam 9450	Unicam, UK
Rolling mixer SRT6D	Stuart®, UK
Sephadex G25-M	GE Healthcare, UK
Soniprep 150 sonicator	MSE, USA

Sterivex^{TM} 0.22 μm Filter Unit	Millipore, USA
HisTALON^{TM} Superflow	Clontech Laboratories, USA
TECAN infiniteM1000	Tecan group, Switzerland
Techne termostat	Cambridge, UK
Thermomixer comfort	Eppendorf, Germany
UV-VIS Spectrophotometer	UNICAM UV500, Unicam UK
VIVAspin 50ml	Sartorius Stedim Biotech GmbH, Germany

4.3 METHODS

4.3.1 Protein expression

4.3.1.1 Transformation of *E.coli* expression host by recombinant plasmid DNA

Plasmid DNA (about 200 ng in 1 μL) was pipetted into 25 μL of calcium chloride competent cells (*E.coli* C41(DE3) for the expression of GlpG and *E.coli* MC4100, glpG::cat (Stříšovský, unpublished), for the expression of rhomboid substrates) and incubated on ice for 30 min. Cells were then heat-shocked at 42°C for 45 s in a water-bath and then incubated 2 min on ice. After this incubation, 250 μL of pre-warmed SOC medium (Invitrogen) was added to each tube and tubes were incubated in an Eppendorf Thermomixer for 1 h at 37°C and 450 rpm. Cells were then plated at different dilutions on agar plates containing the selection antibiotic ampicilin (Amp) (Sigma Aldrich) at 100 μg/mL. Agar plates were then incubated overnight in a microbiological incubator (Memmert GmbH) at 37°C to let the transformed, ampicilin-resistant cells, grow into well-separated colonies.

4.3.1.2 Recombinant protein expression in *E.coli*

Several representative colonies were inoculated into 100 mL of LB media with ampiciline (Sigma Aldrich) at 100 μg/mL the day after transformation and, this inoculum was grown

overnight in a thermostated orbital rotary shaker (Innova 4300, New Brunswick Scientific) at 37°C and 220 rpm. Next morning, the culture reached the density of OD_{600} = ~6 and the inoculum was diluted into eight 2 L Erlenmeyer flasks containing 1 L of autoclaved LB medium with ampicilin (100 μg/μl) each to a final OD_{600} of 0.05. These expression cultures were incubated under the same conditions as the overnight inoculum until OD_{600} reached 0.6. Cells were then induced by the addition of 0.5 mM IPTG (Biosynth AG) and incubated overnight at 20°C and 220 rpm. Samples for analysis (1 mL) were taken before and after induction, OD_{600} was measured, both samples were centrifuged to recover the cell pellet, and resuspended in PBS to the same final density corresponding to OD_{600} of 5. These samples were used for expression efficiency verification by SDS-PAGE (as described in 3.2.5.). To harvest the cells for the isolation of the expressed recombinant protein, the overnight cell cultures were centrifuged at 6000xg at 4°C for 15 min and processed further as described in the next section.

4.3.1.3 Isolation of membrane proteins

Cell pellet from 8 L of culture after expression (see 3.2.1.2.) was dissolved in 240 mL of Buffer A (see 3.1.1.1) with 1 mM EDTA (Sigma-Aldrich) and 1mM PMSF (Sigma-Aldrich). Cells were homogenised in a Dounce homogeniser and then disrupted by 3 passages through EmulsiFlex®-C3 (Avestin) operated at a homogenisation pressure of 1100 bar and at 4°C with additional cooling of the cell suspension on ice. The resulting cell lysate was centrifuged at 10000xg for 30 min at 4°C (Avanti J-30I, Beckman Coulter) and supernatant was ultracentrifuged at 100000xg for 2 h at 4°C (Optima L-90K, Beckman Coulter). The pellet from this ultracentrifugation step represented the enriched membrane fraction. It was then homognised by Dounce homogeniser in 10 mL of Isolation buffer B (for composition see 3.1.1.1) with a mix of protease inhibitors (Complete™, EDTA free 1 tbl/17 mL of buffer). For the purification of rhomboid on the Immobilised metal ion affinity chromatography co (IMAC) cobalt resin TALON™, membranes were resuspended in the Isolation buffer B without imidazole. Protein concentration in the suspension of membranes was measured by SingleQuant kit (Serva) and the membrane fraction was diluted to a final total protein concentration of 5 mg/mL by the same buffer with protease inhibitors. Membrane proteins were solubilised by adding an aliquot of a 20% (w/v) stock solution of n-dodecyl-β-D-maltopyranoside (DDM, Anatrace) detergent to obtain a final

concentration of 1.5% (w/v) DDM (which is ~67x above its critical micelle concentration). Membranes for solubilisation were incubated on a rotary shaker (Rolling mixer SRT6D, Stuart®) for 1 h at laboratory temperature, and then ultracentrifuged at 100000xg for 30 min at 4°C (Optima L-90K, Beckman Coulter) to remove unsolubilised proteins and precipitates and isolate the DDM-solubilised protein in the supernatant. Samples from all centrifugation steps were collected and analysed by SDS-PAGE to monitor isolation efficiency.

4.3.2 Affinity chromatography

4.3.2.1 Immobilised metal ion affinity chromatography on nickel column (NiNTA)

His-tagged substrates were purified on NiNTA column. All purification steps were done in the cold room (~4°C). NiNTA HiTrap Superflow 1 mL column (Qiagen) was equilibrated with 50 mL of buffer C at a flow rate of 1 mL/min using a peristaltic pump. All solubilised protein was loaded on the column at a flow rate of 0.5 mL/min and the effluent was then continuously pumped through the column at 1 mL/min overnight. Next day, column was first washed by 50 mL of buffer C (1 mL/min) and then by the same volume of NiNTA W buffer (1 mL/min). His-tagged protein was eluted in five steps by 3 mL of NiNTA E buffer. Purity of elution fractions was monitored by SDS-PAGE.

4.3.2.2 Amylose affinity chromatography

If the substrates were not pure enough after NiNTA purification, amylose resin purification was carried out. One milliliter of amylose resin was equilibrated by buffer A containing 0.05% (w/v) DDM at a flow rate of 1mL/min, and the protein was loaded in a closed loop continuously overnight using the same flow rate. The subsequent washing step comprised 5 mL of buffer A with 0.05% (w/v) DDM, and elution was done by three 2 mL of buffer A with 0.05% (w/v) DDM and 10 mM maltose. Samples were analysed by SDS-PAGE. Pure proteins were flash frozen in liquid nitrogen and stored at -20°C.

4.3.2.1 Immobilised metal ion affinity chromatography on cobalt column (TALON™)

Full-length GlpG is unstable in higher concentrations of imidazole, therefore TALON™ column chromatography was used because the elution steps could be done in lower imidazole concentrations compared to NiNTA chromatography. HisTALON™ Superflow 1 mL column (Clontech Laboratories) was washed by 50 mL of TALON™ wash buffer at 1 mL/min by peristaltic pump. The DDM-solubilised proteins were loaded on the column at 0.5 mL/min, and then 'circulation loading' was set up overnight, as described above. Washing steps were done next day. Column was washed by 50 mL of TALON™ W buffer at a flow rate of 1 mL/min and then by the same volume of buffer C (1 mL/min). Protein was eluted in five steps by 3 mL of TALON™ E buffer which contained 73 mM imidazole. Samples were analysed by SDS-PAGE and protein-containing fractions were buffer-exchanged into buffer A with 0.05% (w/v) DDM using size exclusion chromatography (desalting) column Sephadex G25-M (GE Healthcare).

4.3.3 Size exclusion chromatography

Gel-permeation chromatography was performed on ÄKTA Explorer. HiLoad™ 16/60 Superdex™ 200 (GE Healthcare) was equilibrated by filtered and degassed buffer A containing 0.05% (w/v) DDM. The protein to be purified was concentrated and centrifuged at 10000xg for 15 min at 4°C (Allegra X-15R, Beckman Coulter) prior to injection on the column. Proteins were eluted using the same buffer at a flow rate of 0.8 mL/min, at a maximal pressure of 0,5 MPa with spectrophotometric detection at 280 nm. Protein-containing fractions were collected and analysed by SDS-PAGE. Pure enzymes were flash frozen in liquid nitrogen and stored at -20°C.

4.3.4 Determination of protein concentration

Protein concentration was routinely (during isolation and concentration steps) determined by amido-black binding assay [74] in a kit format (SingleQuant, Serva), which is very tolerant to detergents and other substances, because the protein is first precipitated from solution, which removes most of the potentially interfering substances. However, since this

method, as all dye-binding assays, gives somewhat sequence-dependent values, to obtain accurate protein concentrations of purified proteins required for enzymological analyses we used quantitative amino acid analysis. This was carried out by Ing. Souček at the analytical service group of the Institute of Organic Chemistry and Biochemistry (IOCB).

4.3.5 Concentration of protein solutions

Proteins were concentrated by ultrafiltration using VIVAspin concentrators (Sartorius Stedim Biotech GmbH) with 30 kDa molecular weight cut off. Centrifugation was performed at 10000xg at 4°C (Allegra X-15R, Beckman Coulter) until the required volume was reached. Protein solutions before chromatographic steps were concentrated to maximum 5 mg/mL. Pure protein solutions were concentrated to 1 mg/mL as determined by SingleQuant protein concentration determination assay (Serva).

4.3.6 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins were analysed on commercial pre-cast gradient polyacrylamide gels – 4-20% Tris-Glycine or 10-20% Tris-Tricine (Bio-Rad). Electrophoresis cell (Bio-Rad) was set up and filled with SDS-PAGE running buffer (3.1.1.1) appropriate for the gel. Samples were mixed with 1X SDS-PAGE sample buffer, incubated at 65°C for 15min and loaded on the gel. Ten μL of samples was loaded on the 15-well comb gels and 45 μL was loaded on 10-well gels. Electrophoresis proceeded at a constant voltage of 200 V for 45 min for Tris-Glycine and at 100 V for 1.5 h for Tris-Tricine gels. Five μL of Protein Ladder marker 10-250 kDa (New England BioLabs) was loaded on each gel to estimate the molecular weight of the analysed proteins.

4.3.6.1 Protein staining in gels

Proteins in gels were visualized by colloidal Coomassie staining using the InstantBlue kit (Expedeon). Gels were incubated on an orbital shaker at laboratory temperature for 1 h then they were washed by distilled water and, as recommended by the manufacturer, heated in microwave oven for 1 min at maximal power to accelerate the staining

procedure. The background staining was then reduced by washing the stained gel in distilled water for 1 h on an orbital shaker at laboratory temperature.

4.3.6.1 Densitometry

Stained and washed gels were inserted into a plastic foil and scanned using a photographic transmission scanner (EPSON perfection V37). Densitometric analyses of protein bands in the images of scanned gels were conducted in ImageQuant 8.0 software (GE Healthcare).

4.3.7 Peptide handling

All peptides were synthesized by solid-phase chemistry in the medicinal chemistry service group of IOCB (headed by Dr. Pavel Majer), purified by reversed-phase high-performance liquid chromatography (HPLC) and analysed by mass-spectrometry. The pure, lyophilized peptides were stored in a dessicator at room temperature. To prepare stock solutions of the peptides, the lyophilized powders were dissolved in anhydrous dimethylsulphoxide (DMSO, Invitrogen) to 10 mM and stored at -20°C. Accurate concentration of these peptide solutions was determined by quantitative amino acid analysis carried out by the medicinal chemistry service group (Ing. Souček).

4.3.8 Peptide solubility determination

O-phthaldialdehyde (OPA) reacts with primary amines to form fluorescent product. Our peptides contain a number of lysines which can react with OPA and this derivatisation reagent can thus serve to determine the relative concentration of our pure peptides in solutions [75]. Serial two-fold dilutions of peptides KSp26-31 in 20 mM HEPES buffer with 0.05% (w/v) DDM and final 10% (v/v) DMSO were prepared so that peptide concentrations ranged from 250 μM down to 31.5 μM. Samples were incubated for two hours at 37°C and then centrifuged at 18407xg for 15 min at laboratory temperature in an Eppendorf centrifuge 5424 (Eppendorf). In the meantime, O-phthaldialdehyde (OPA) (Sigma-Aldrich) was dissolved in the mixture of 25 μL of acetonitrile and 25 μL 400 mM boric acid solution titrated to pH 9.7. After complete dissolution sample was diluted 100x by 400 mM boric acid solution titrated to pH 9.7 with potassium hydroxide containing

0.2% (v/v) 2-mercaptoethanol. Forty μL of supernatants after centrifugation were conjugated with 40 μL of OPA solution and fluorescence was measured directly in a plate reader (TECAN) at 340 nm excitation and 455 nm emission wavelength.

4.3.9 In vitro cleavages before SDS-PAGE analysis and matrix-assisted laser desorption/ionization (MALDI)

In this experiment, 100 μM peptides KSp26-31 were added to 50 mM HEPES buffer with 0.05 % (w/v) DDM and reaction was started by adding GlpG to 3.2 μM. Final concentration of DMSO was kept at 10% (v/v). WT GlpG and its inactive mutant S201A as a negative control were used. Samples were incubated 2 h at 37°C in the Thermomixer at 1400 rpm. Samples were then loaded on Tris-Tricine SDS-PAGE or analysed by MALDI in the mass spectrometry service group (Doc. Cvačka) at IOCB.

4.3.10 In vitro cleavages for capillary electrophoreses (CE)

In this experiment, 250 μM peptides were added into 20 mM HEPES with 0.05 % (w/v) DDM and DMSO was adjusted to be 10% (v/v) final concentration. Reaction was started by adding 2.6 μM WT GlpG. Samples were incubated at 37°C in the Thermomixer at 1400 rpm. Fractions of 20 μL were collected each 15 min for up to 2 hours and added to micro tubes with 5 μL of 50 mM HCl which stopped the rhomboid reaction. CE was performed by the service and research group of Dr.Kašička at IOCB.

4.3.11 Microscale Thermophoresis (MST)

Microscale thermophoresis is a method of measuring protein-ligand or protein-protein interactions in solution, and it is based on thermophoresis – the motion of molecules in a temperature gradient. Thermophoresis is characteristic for each studied system and depends on its several physical and chemical properties, such as hydration shell, charge, size and conformation. Interaction of protein with another molecule (protein/ligand) can influence these properties and change its migration in the temperature gradient. Microscale thermophoresis is performed in a capillary where a steep temperature gradient is created by

an infrared laser. Migration of molecules in the resulting thermal gradient is then detected by protein UV-fluorescence measurement [76]. Although microscale thermophoresis is a relatively new method, it has already been amply used with integral membrane proteins such as GPCRs. It seemed well-suited for K_D determination of our truncated peptide series, also because it allows measurement in solution in the presence of detergent micells and it can be a label-free method.

Solutions of 5 mM peptides in the MST buffer with 20 % (v/v) DMSO were prepared. Two-fold serial dilutions were done fifteen times yielding 16 serial samples of 1 mM to 30.52 μM peptide and final volume of 10 μL each. Then 170 μL of enzyme solution was prepared containing 3.4 μM GlpG.S201T in MST buffer with 20 % (v/v) DMSO. Ten μL of enzyme solution was added to each 10 μL of peptide solution. The final sample for MST measurements contained peptides at 500 μM to 15.26 μM, 20 % (v/v) DMSO and 1.7 μM GlpG mutant S201T. Samples were centrifuged at 10000xg for 5 min at laboratory temperature (Eppendorf centrifuge 5424, Eppendorf), supernatants were loaded into standard treated capillaries and measured on MonolithNT. LabelFree using 15 % LED power (allowing fluorescence measurement) and 80 % MST power (the power of laser creating temperature gradient).

4.3.12 Role of TMH in specificity – molar cleavage efficiency determination

Time dependent cleavage of four substrates varying in TMH sequence with four different enzymes (WT GlpG, inactive mutant S201A GlpG and two hyperactive mutants F153A/W236A and W157A/F232A) was followed. To allow quantitative comparisons we used the same concentration of substrates ~5 μM in all reactions. As cleavage efficiency of the analysed enzymes differed, they were used at different concentrations so that the conversion of substrate was lower than 30% at least for three time-points during the first 2 h. Pre-tests were done to find out optimal concentration of enzyme with which the conversion of substrate would be in estimated range in 2 hours. Finally the cleavage assays were done with concentrations of enzymes described in Table 1 (page 46).

Samples were analysed on SDS-PAGE gel, stained, scanned and evaluated densitometrically using ImageQuant 8.0 software as follows.

Table 1.Final concentration of enzymes used for time dependent cleavage assay: 5 μM substrates were mixed with different concentrations of enzymes visualised in this table and samples were taken each 15 min to map substrate conversion.

GlpG variants	substrates		
c (μM)	LacY	Gurken	TatA
WT GlpG	0,621	0,636	0,211
W157A/F232A	0,687	0,239	0,343
F153A/W236A	0,013	0,253	0,366

Data processing

Substrate conversion was calculated for each lane separately to avoid loading mistakes. We have first tested the linear range of Coomassie staining using a pure protein standard. Given the simple stoichiometry of the proteolytic reaction (Form. 1A), substrate conversion at a given time can be calculated from molar concentrations of one of the products and uncleaved substrate at the given time (Form. 1B). The intensity of Coomassie staining (obtained from densitometry) is proportional to the mass of protein in the band of SDS PAGE. These values can be converted to molar ratios, but since in our case the substrate and product 1 have similar molecular weights, this correction was omitted for simplicity, and densitometric values were directly used for the calculation of substrate conversion. The intensity of the uncleaved substrate band was divided by the sum of this value with the intensity of the band of the cleavage product 1 (Form.1B).

A

$$S \rightarrow P_1 + P_2$$

B

$$\alpha = \frac{[P_1]}{[S_0]} = \frac{[P_1]}{[P_1]+[S]} \approx \frac{I_{P1}}{I_{P1}+I_S}$$

Formula 1: Conversion of substrate. A) Simplified reaction scheme. B) SDS-PAGE gel was scanned and evaluated densitometrically. Conversion of substrate was calculated for each lane separately to avoid loading inaccuracies. S, substrate, P, product; $[S_0]$, substrate concentration at time zero, [S], substrate concentration at a given time, I_{P1}, intensity of product P1 band, I_S, intensity of substrate band.

By definition, when multiplied by 100, substrate conversion ranges between 0 and 100%. To ascertain kinetics condition approximating the initial rate conditions, only substrate conversions lower than 30% were used for further analysis. From substrate conversion in %, molar concentrations of the converted substrate were calculated based on the known concentration of substrate before cleavage we obtained from amino acid analysis. These molar concentrations of consumed substrate (equal to the concentration of formed product 1) were plotted against time to obtain a curve whose slope at time zero corresponded to the initial rate of the reaction. Although concentration of substrates were kept constant for all experiments, different concentration of enzymes had to be used because they differred in activity, while substrate-enzyme ratio was kept at 9:1 or higher. Initial reaction rates were divided by the concentration of the enzyme used (from amino acid analysis), which yielded molar catalytic activity in min^{-1} for all substrate-enzyme pairs that was used in all subsequent comparisons of the studied GlpG mutants.

4.3.13 Fluorescence measurement using plate reader

4.3.13.1 Solubility test and background determination

Solubility of KSp35 in the assay buffer was tested. Dilution series (100; 75; 50; 30; 10 and 1 μM) was prepared into 120 μL of 50 mM HEPES containing 0.05% (w/v) DDM. Final concentration of DMSO was 20% (v/v). Samples were incubated at 37 °C for 2h and then centrifuged at 18407xg at laboratory temperature (Eppendorf centrifuge 5424, Eppendorf). Supernatant was pipetted into transparent 96-well plates in duplicates (50-μL/well) and absorbance at 480 nm was measured on TECAN reader.

4.3.13.2 Cleavage of fluorogenic substrates by panel of different rhomboids

Purified rhomboid proteases (WT GlpG, AraA, Bacteroides rhomboids 1, 2 and inactive GlpG S201A as a negative control) were diluted in 20 mM HEPES buffer with 0.05% (w/v) DDM in Eppendorf tubes to a final concentration of 0.8 μM rhomboid and final volume of 117 μL, and 9 μL of DMSO was added to this solution. Samples were pre-incubated in a Thermomixer at 37°C for 15 min. In the meantime, the 96-well plate and TECAN fluorescence reader were equilibrated to 37°C. Then the reaction solution was

pipetted into 96-well plates (48.75 μL/well) in duplicates. Addition of 1.25 μL of 1 mM fluorogenic substrate KSp35 in DMSO into each well resulted in 10% (v/v) final concentration of DMSO and 25 μM substrate. Samples were measured directly on TECAN fluorescence reader using excitation at 335 nm and emission wavelength of 493 nm.

5 RESULTS

5.1 PROTEIN EXPRESSION AND PURIFICATION

The recombinant rhomboid protease variants WT GlpG, its inactive mutant S201A and two double mutants, F153A/W236A and W157A/F232A, were each expressed in 8 L of *E.coli* expression cells C41(DE3) [77]. Cells were lysed and membrane proteins were solubilised from the isolated membrane fraction by 1.5% (w/v) DDM, and GlpG variants were then purified on the cobalt IMAC TALON^{TM} column. This matrix was chosen because it exhibits a lower level of unspecific binding than NiNTA and requires lower concentrations of imidazole for elution (because it has lower affinity to His-tag than Ni-NTA). This is important, because full-length GlpG is unstable and tends to precipitate in higher concentrations of imidazole needed for the elution from the NiNTA column. Nevertheless, to minimize the contact of purified protein with imidazole, buffer was exchanged directly after the cobalt IMAC chromatography. Purified enzymes were then concentrated and further purified via size exclusion chromatography on Superdex 200.

For substrates TatA, LacY, Gurken and Spitz, *glpG* knock-out *E.coli* had to be used for expression to prevent their possible cleavage by the endogenous GlpG. The rest of the solubilisation protocol was identical to the one used for GlpG (see 4.2.1.1-3). Fractions from the isolation steps were analysed by SDS-PAGE using Tris-Glycine gels (Fig.17 A; page 50). Substrates were purified by two sequential affinity chromatography steps. As all the substrates were C-terminally His-tagged, NiNTA was used as a first purification step (Fig.17 B; page 50). If the eluted protein was not pure enough for our purposes, it was purified to homogeneity on amylose resin (New England Biolabs) via the maltose binding protein tag located at the N-terminus of the substrates [70].

After the isolation and purification steps, the pure enzymes and substrates were obtained in purity (Fig.18; page 50) and yields summarised in the table below (Tab.2; page 51)

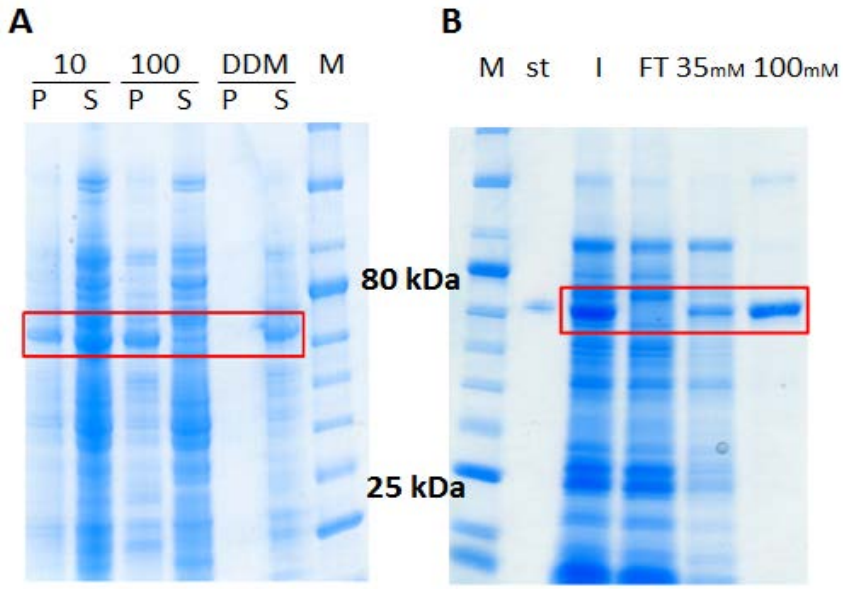


Figure 17: Isolation and purification of the recombinant chimeric substrate TatA: A) After cell disruption, low speed centrifugation (10000xg) was done to exclude undisrupted cells or aggregates (P10) from the cell lysate (S10). Membrane fraction (P100) was obtained by ultracentrifugation (100000xg) of the cell lysate and after the solubilisation of the membrane pellet in 1.25% (w/v) DDM, second ultracentrifugation was performed to obtain solubilised protein (S_{DDM}). **B)** SDS-PAGE (Tris-Glycine gel) of the purification of TatA on NiNTA. Solubilised protein (S_{DDM}) was purified on NiNTA. All of the His-tagged substrate protein from the input (I) to the column was bound to the resin, and virtually no substrate was lost in the flow through (FT) fraction. Unspecifically bound proteins were washed out by buffer containing 35 mM imidazole (35 mM). Protein elution was done by the loading buffer with 100 mM imidazole. The purified His-tagged substrate protein is represented in the red box. The upper band in the same lane is an SDS-stable dimer as characterized by mass spectrometry (data not

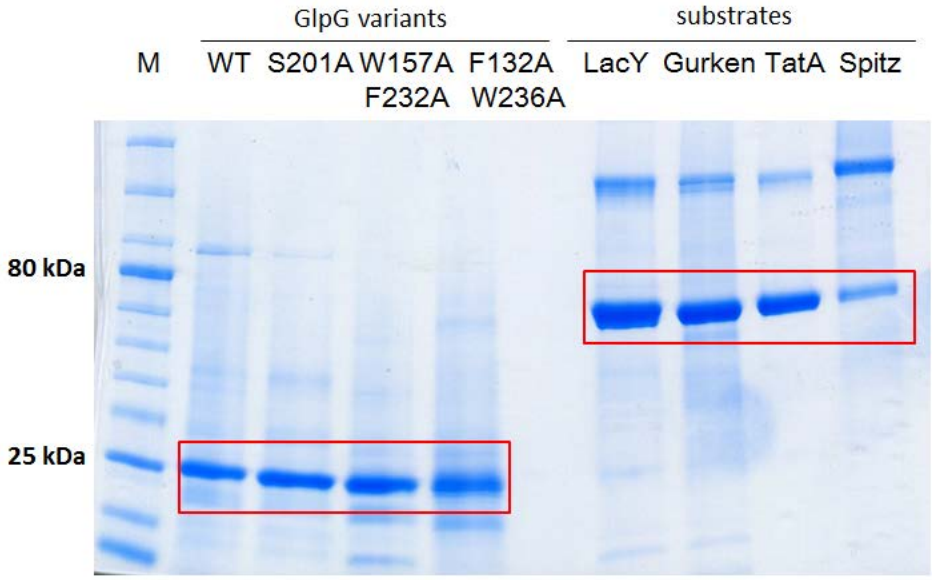


Figure 18: Isolated and purified GlpG variants and substrates. ~5 µg of purified proteins were loaded on Tris-Glycine gradient (10-20%) gel and SDS-PAGE was performed. Upper band present in all substrate lanes is an SDS-stable dimer.

Table 2: Yields of proteins after expression in 8 L of *E.coli* culture and purification on TALON™ and SEC for enzymes and NiNTA and Amylose resin for substrates.

pure protein	concentration (mg/mL)	volume (mL)	yield (mg)
TatA	1	8	8
LacY	1	4,5	4,5
Gurken	1,06	7	7,42
Spitz	1	3,25	3,25
WT GlpG	1	2,25	2,25
GlpG S201A	1	1	1
GlpG F153A /W236A	0,97	0,8	0,78
GlpG W157A/F232A	0,98	0,5	0,49

5.2 THE CONTRIBUTION OF SUBSTRATE'S TRANSMEMBRANE HELIX TO THE EFFICIENCY AND PRECISION OF RHOMBOID CATALYSIS

To address if and how important is the substrate TMH for recognition by rhomboid, we first decided to investigate the effect of C-terminally truncating the TMH of peptide substrates on the kinetics of their cleavage by GlpG. We designed synthetic peptides based on the sequence of artificial substrate LacYTMH2, which we knew was very well cleaved by GlpG (and several other rhomboids). The peptide with full length TMH contained LacYTMH2 sequence from P11 to P'26 (peptide pKS31) (Tab.3; page 52). Variants of this parent substrate that were C-terminally truncated by 3 or 4 amino acids (corresponding to about 1 turn of a transmembrane helix) were commissioned with the IOCB peptide synthesis laboratory. The expected N-terminal cleavage product peptide (pKS34) was synthesised as a positive control for analytical purposes.

Table 3: **Synthetic peptide substrates based on LacYTMH2.** Peptides KSp30-26 derived from the KSp31 with full length TMH were made shorter from their C-termini. TMH amino acids are underlined and the cleavage site is represented by dash. KSp34 is the expected N-terminal cleavage product.

peptide	sequence
KSp34	KRHDINHISKS (N-terminal cleavage product)
KSp31	KRHDINHISKS-DTGIIFAAISLFSLLFQPLFGLLSKK (full length TMH)
KSp30	KRHDINHISKS-DTGIIFAAISLFSLLFQPLFG
KSp29	KRHDINHISKS-DTGIIFAAISLFSLLFQP
KSp28	KRHDINHISKS-DTGIIFAAISLFSL
KSp27	KRHDINHISKS-DTGIIFAAISL
KSp26	KRHDINHISKS-DTGIIFAA

5.2.1 Peptide solubility determination

Any TMH containing peptides will be considerably hydrophobic, so prior to their kinetic characterisation we first needed to ascertain their solubility in our activity assay buffer. We diluted each peptide from its 10 mM stock solution in anhydrous DMSO into the assay buffer at a range of concentration between 0 and 250 µM so that the total DMSO concentration remained constant at 10% (v/v), incubated the solutions at room temperature for 1 h and centrifuged away any insoluble aggregates in a table-top microcentrifuge. We then detected the peptides remaining in the supernatant using derivatisation of their free amino groups by o-phthaldialdehyde (OPA), which becomes fluorescent upon reaction with primary amines [78]. Thus, after conjugation with our peptides that contain a number of primary amino groups at several lysine residues, this reagent allows the determination of relative peptide concentration. The fluorescence of the OPA derivatives of our post-centrifugation supernatants should be linear with the nominal concentration of the peptide if the given peptide is soluble over the examined concentration range.

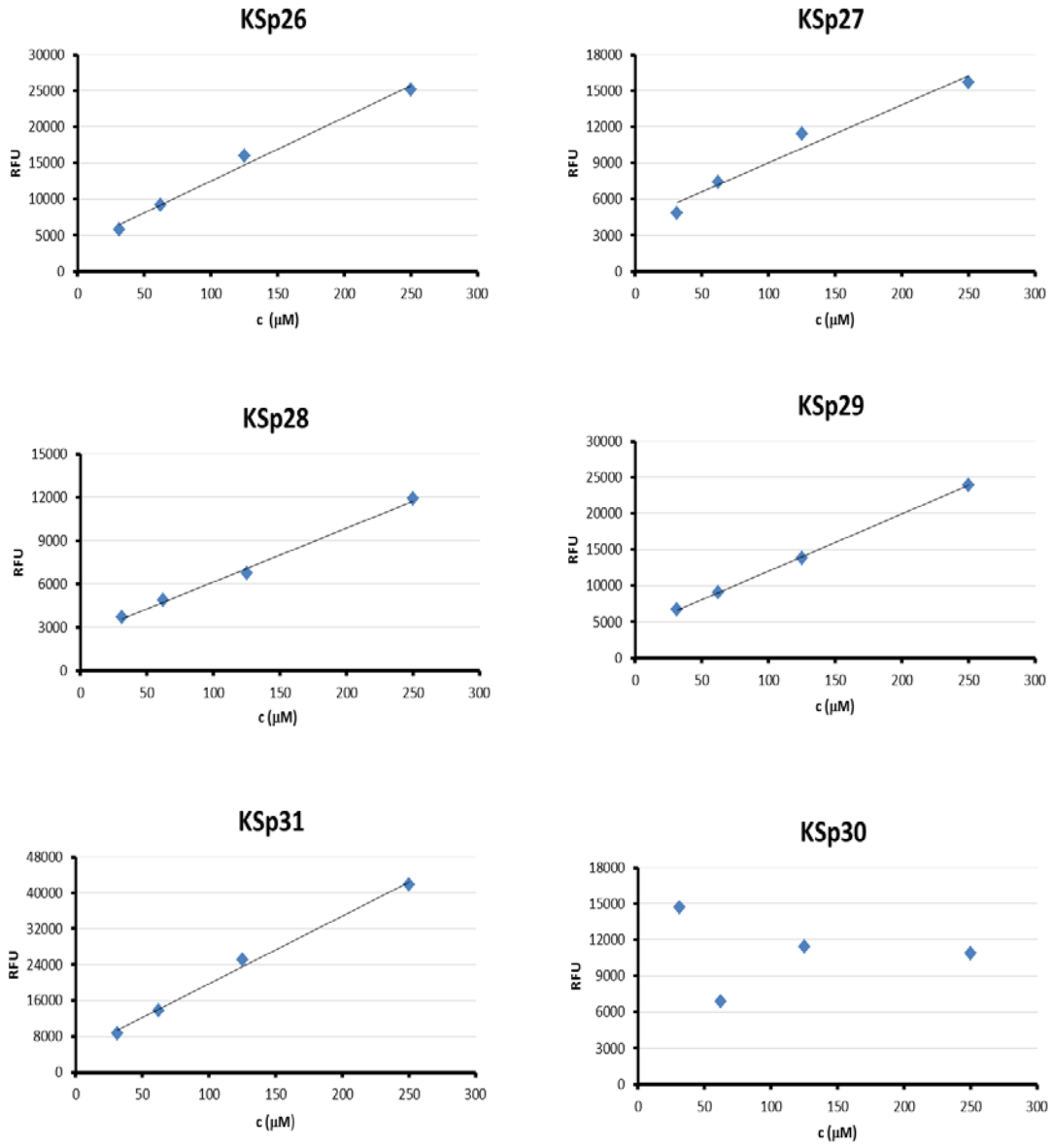


Figure 19: Solubility of KSp26-31 in the activity assay buffer containing 0.05% (w/v) DDM, representative graphs. Samples were prepared and incubated 2 hours at 37°C. After that all samples were centrifuged at 18407xg for 15 min at laboratory temperature. Supernatant was pipetted into 96-well plate and conjugated with o-phthaldialdehyde. Fluorescence at 340nm excitation and 455nm emission wave length was measured.

The OPA assay performed with our series of peptides has shown a linear tendency for all peptides except for KSp30, meaning that peptides KSp26-29 and KSp31 were soluble in the range of 31.25 µM to 250 µM (Fig.19). Peptide KSp30 was excluded from further analysis because of its limited solubility.

5.2.2 In vitro cleavage efficiency of substrate peptides with progressively truncated transmembrane domain

5.2.2.1 Qualitative characterisation by mass spectrometry

To verify the site of cleavage of the series of the C-terminally truncated LacYTMH2 derived peptides, we exposed them to wild type GlpG and to its inactive mutant S201A and analysed the reaction mixtures by MALDI mass spectrometry (MS). The expected cleavage product KSp34 (based on experiments with the full-length protein substrate [70]) was used as a standard. Briefly, samples with 25 μM peptides KSp26-31 and 0.8 μM GlpG in the assay buffer 20 mM HEPES pH 8 were incubated at 37°C for 2 h. Inactive GlpG mutant S201A at 0.8 μM was used as a negative control. A preliminary analysis of the reaction mixtures by Tris-Tricine SDS PAGE, which should be able to resolve peptides down to ~1 kDa size [79], already suggested that peptide KSp31 was cleaved (Fig.20).

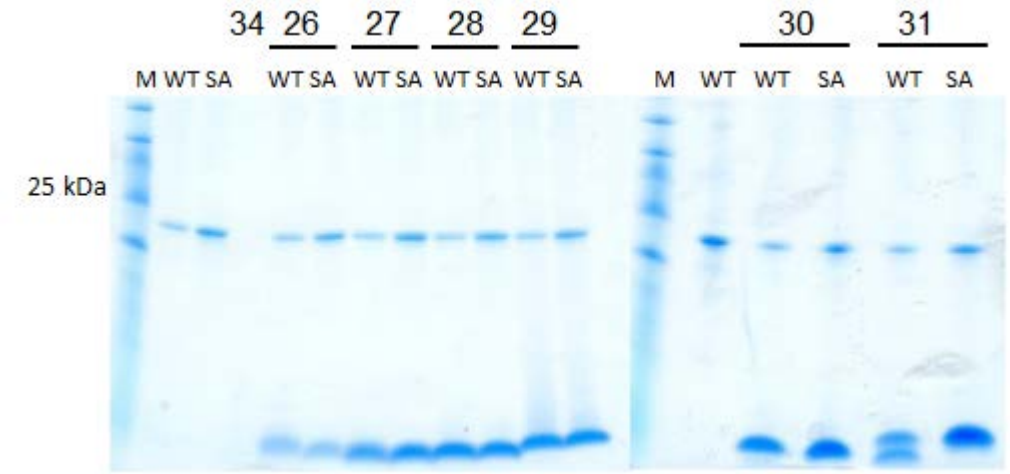


Figure 20: Reactions of peptides KSp26-31 with WT GlpG and its inactive mutant S201A. After 2 h incubation of peptides with the enzymes, reactions were subjected to 10%-20% Tris-Tricine SDS-PAGE analysis. Cleavage product was clearly observed for the longest peptide KSp31.

Cleavage reactions of the whole peptide series were characterised by MALDI MS in three independent replicates, giving essentially identical results. We did observe the formation of the cleavage product corresponding to KSp34 (Tab.4; page 55) after reaction with WT

GlpG but not with its S201A mutant, but only for peptides KSp27-31 (Fig.21; page 56). Since mass spectrometry is not a quantitative method, to compare the cleavage kinetics of these peptide series, we first attempted to separate the cleavage products from substrates by reversed-phase HPLC using a C_{18} column. However, we persistently observed the occurrence of serious artifacts such as peak broadening or ghost peaks (data not shown) that precluded identification and quantitation. These artifacts were probably due to the strongly hydrophobic character of the peptides and the presence of relatively high concentration of the detergent (DDM), and we abandoned this method. Since our preliminary data showed that electrophoretic separation of the cleavage products may be feasible, we resorted to the use of analytical capillary electrophoresis with spectrophotometric deection

Table 4. Sequence and relative molecular mass of KSp31 and its cleavage products.

peptide	sequence	Mr
KSp31	KRHDINHISKSDTGIIFAAISLFSLLFQPLFGLLSKK	~4156
N-terminal cleavage product	KRHDINHISKS	~1334
C-terminal clevage product	DTGIIFAAISLFSLLFQPLFGLLSKK	~2822

Colour of peptides corresponds to the colour of the arrow for each peptide characterised by MALDI analyses (Fig.21; page 56).

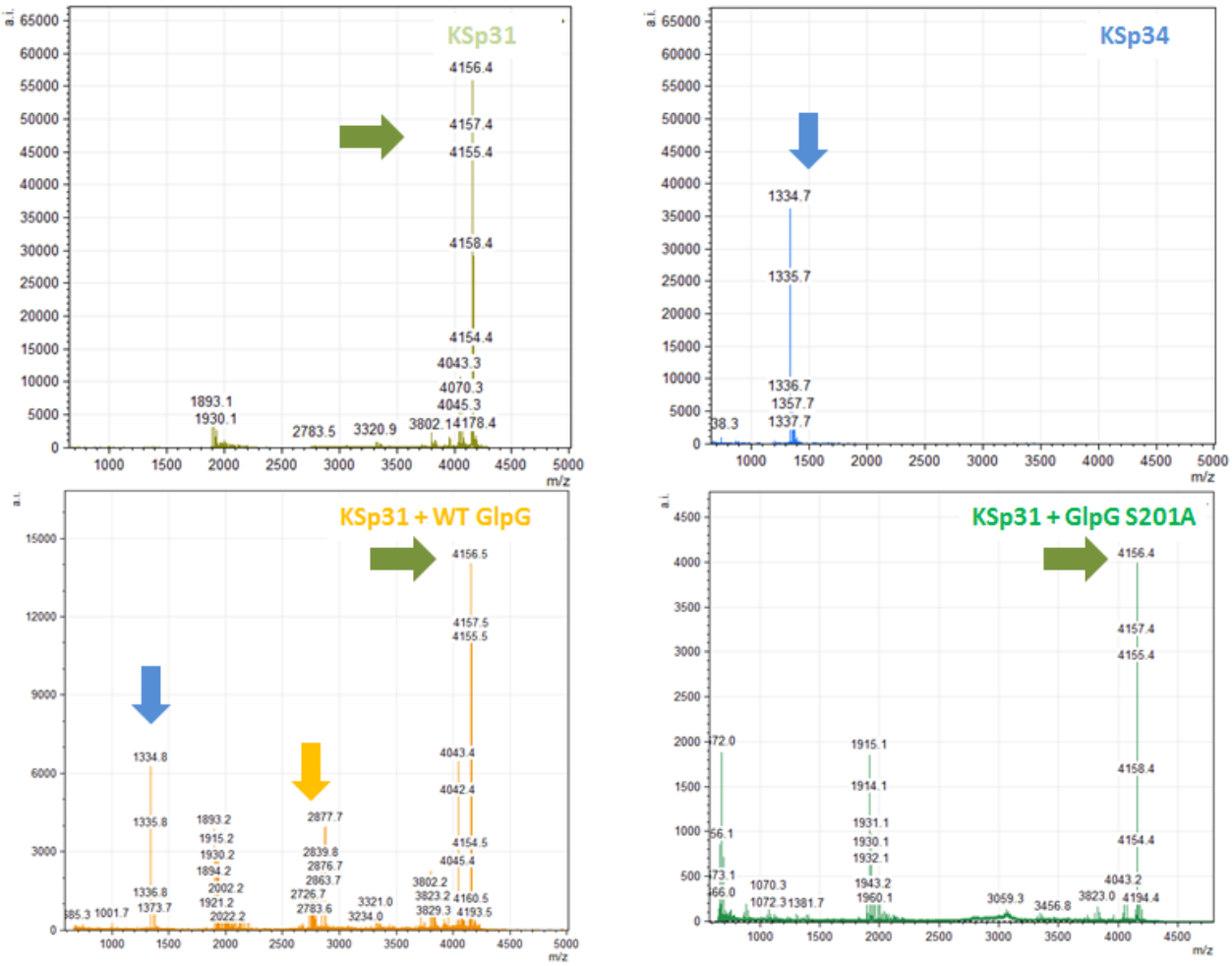


Figure 21: Mass spectrometric (MALDI) analysis of KSp31 cleavage by GlpG. The light green spectrum shows a window of the mass spectrum of pure KSp31 and the blue one the spectrum of pure KSp34, the expected cleavage product of KSp31. After incubation of KSp31 with WT GlpG 2 hrs at 37°C reactions were analysed by mass spectrometry. A fragment of an identical mass to KSp34 (blue flash) as well as C-terminal cleavage product (yellow flash) were formed in the presence of GlpG but not with its S201A inactive mutant (dark green).

5.2.2.1 Quantitative comparison of cleavage rates by capillary electrophoresis

We failed to obtain separation of the hydrophobic, TMH containing cleavage products of our peptide series by capillary electrophoresis (CE), but we could accurately determine the concentration of the relatively hydrophilic N-terminal cleavage product corresponding to KSp34, which we had shown was produced from peptides KSp27-31. To correct for injection inaccuracies, tyramine was used as an inner standard that was added to each sample prior to its analysis by CE. The example electrophoretogram in Figure 22A (page 57) demonstrates the separation of the inner standard tyramine and KSp34. The

calibration curve on pure KSp34 was linear up to 160 µM KSp34, and the lower detection limit was about 20 µM (Fig.22 A, B).

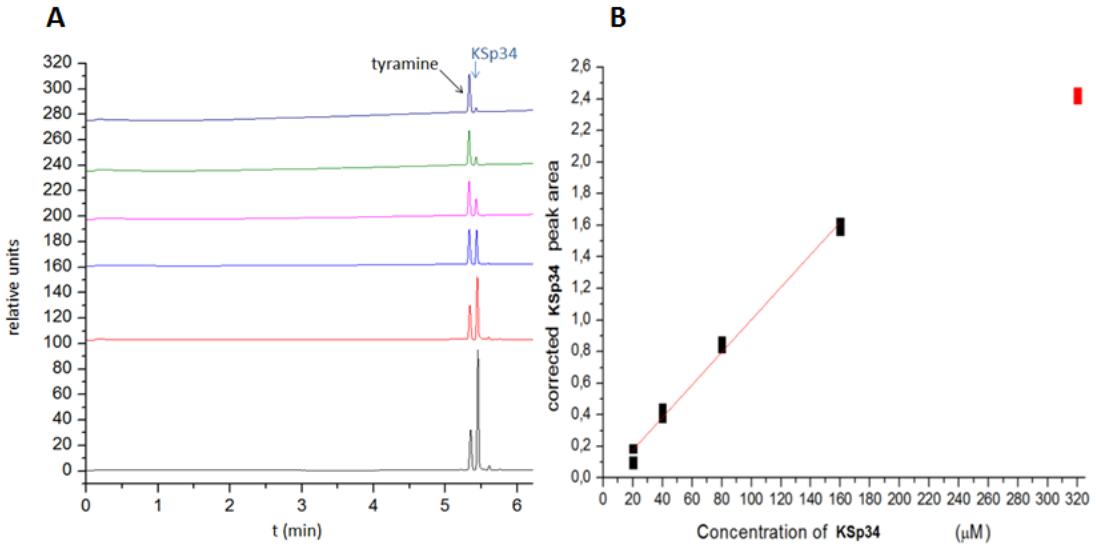


Figure 22: Calibration curve of peptide KSp34 (N-terminal cleavage product) determined by capillary electrophoresis. Two-fold serial dilutions of KSp34 from 320 µM down to 10 µM were analysed by CE. Absorbance was measured at 192 nm. Peak of KSp34 was integrated and normalized to the inner standard tyramine. A) example electrophoretogram; B) the calibration curve

GlpG activity has been shown to be pH-sensitive, becoming negligible at and below pH 4 [72]. We tested the inactivation efficiency of pH 4 by our newly developed fluorogenic substrate KSp35 (which will be described later in section 5.4) - it occurred practically instantly after the addition of HCl (details in Methods and data not shown). We exploited this property of GlpG to stop the cleavage reactions in time-course experiments after the indicated time by adding 5 µL of 50 mM HCl into 20 µL of the sample, which instantly adjusted pH to 4.

To compare the cleavage efficiency of the peptide series KSp27-31, cleavage reactions were prepared in the assay buffer containing 250 µM peptides, incubated at 37°C and 20 µL aliquots were withdrawn after 15 min intervals and quenched by HCl as described above. Samples were then processed by the CE service group to determine the concentration of the N-terminal cleavage product corresponding to KSp34 that was formed by GlpG. An illustrative overlay of electrophoretogram for the time-course of KSp31 cleavage is shown below (note that since the curves are plotted stacked upon each other in

one graph, the y axis absorbance units are relative) (Fig.23). Each electrophoretogram contains the peak of the inner standard and the peak representing the N-terminal cleavage product, where the latter increases with progressing reaction time, as expected.

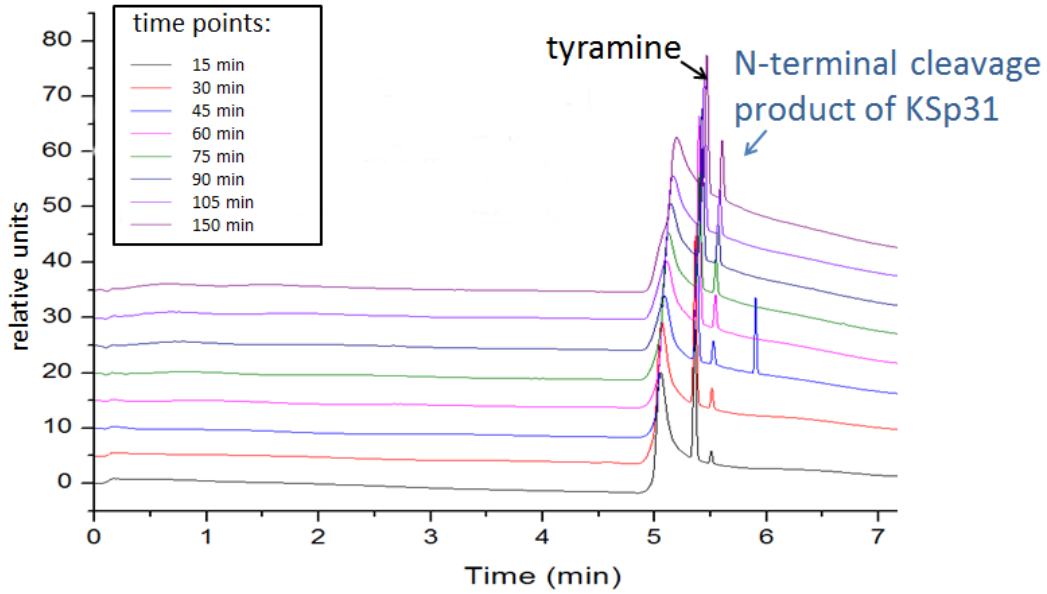


Figure 23: Electrophoretogram from capillary electrophoresis of the time-course of KSp31 cleavage by WT GlpG. All absorbance (at 220 nm) traces were plotted to scale in the same graph to visualise the increasing production of the N-terminal cleavage product. Samples were taken each 15 min and the curves are colour-coded. The lowest curve represents the reaction sample after 15 min and each upper curve represents 15 min step

The integration of the peak area of the N-terminal product and its comparison to the calibration curve gave us the concentrations of the cleavage product produced by GlpG after indicated reaction times. Peptide KSp27 was cleaved so poorly that at the sensitivity of our CE assay it was not possible to quantify it reproducibly. All reactions were prepared once, but they were measured in three independent runs with the coefficient of variation lower than 4% for all analysed samples.

The resulting concentration of the N-terminal cleavage product of all peptides where it was measurable (KSp28, 29 and 31) was plotted against time (Fig.24; page 59).

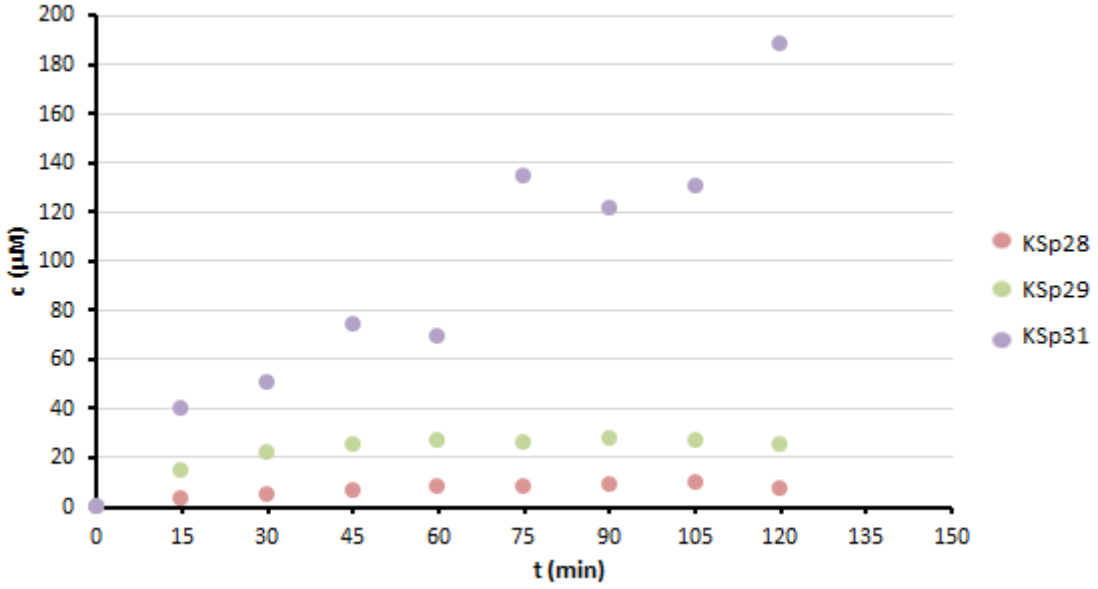


Figure 24: Capillary electrophoresis analysis of cleavage of KSp28, 29 and 31 by WT GlpG. To compare the time dependent cleavage of peptides KSp28, 29 and 31 by WT GlpG, samples were taken each 15 min. Concentration of the formed N-terminal cleavage product was determined by capillary electrophoresis and plotted against time for all peptides. The efficiency of cleavage was found to be proportional to the length of TMH of the substrate. Samples were measured in three independent runs and the coefficient of variation was lower than 4% for all analysed samples.

It is evident that peptide KSp31 (having the full-length transmembrane helix) was cleaved with much higher efficiency compared to its TMH-truncated variants KSp29 and 28. In other words, we observed an inverse correlation between the length of the TMH of the substrate peptide and the rate of its cleavage by GlpG. Even though the KSp34-like cleavage product of KSp27 was detected by MALDI MS, it was formed at levels too low for reliable quantification by CE. Surprisingly, also a different peak appeared on CE for KSp27 and KSp26 (Fig.25; page 60), suggesting that these peptides were cleaved at a different cleavage site than KSp28-31. We have not managed to identify the cleavage product unambiguously yet, but MALDI MS analyses suggested it could be peptide of ~1100kDa. This surprising finding suggests that the transmembrane region of the substrate co-determines the cleavage site specificity of rhomboid.

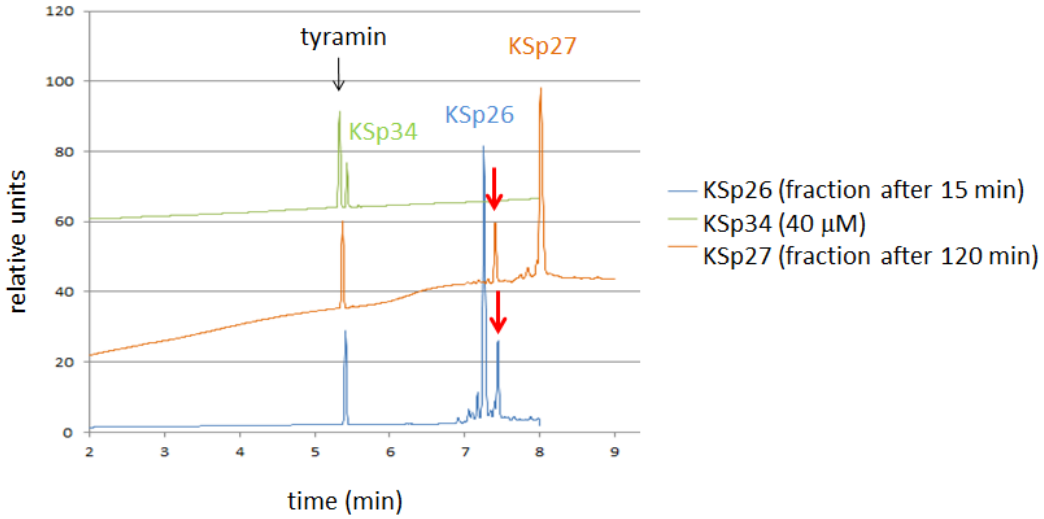


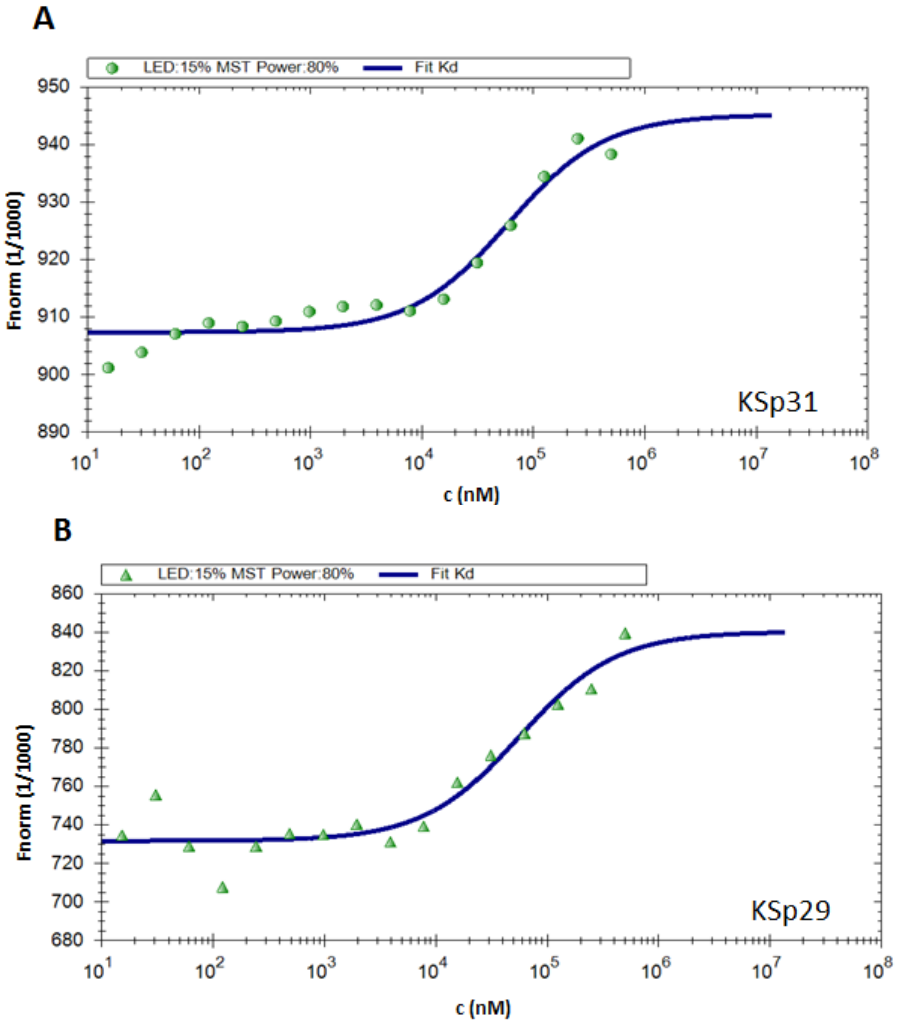
Figure 25: Cleavage of KSp26 and 27 by WT GlpG revealed an uncharacterised cleavage product. Electrophoretograms of KSp26 cleavage reaction after 15 min (blue), KSp27 cleavage reaction after 2 hours (orange) and pure 40 µM N-terminal cleavage product (green) were plotted into same graph (Y axis represents relative absorbance units). Electrophoretograms of KSp26 and 27 do not contain a peak for the N-terminal cleavage product but rather another, uncharacterised, peak has appears (highlighted by a red arrow). Uncleaved peptides KSp26 and KSp27 are represented by dominant peaks in each curve. Two different time-points were chosen to demonstrate different efficiency with which both peptides are converted into the cleavage products. Approximately the same conversion of KSp26 was reached eight times faster compared to KSp27.

5.2.3 Determination of binding constants by microscale thermophoresis (MST)

Given the technical complexity of measuring the kinetics of cleavage of transmembrane peptides, we were not able to determine the apparent Michaelis-Menten parameters of the truncated peptide series KSp27-31 by which we could better understand the role of the TMH of the substrate. Since we conducted our rate comparison at 250 µM substrate peptides, and the apparent K_M for TatA substrate by GlpG is about 100 µM [72], we estimate that we were above the K_M for KSp31, the full-length LacYTMH2 derived substrate. To find out if the strong defect in cleavage efficiency upon shortening of the TMH of the substrate may be at least partly caused by a lower substrate binding constant, we set out to measure the K_D of these peptides towards an inactive mutant of GlpG

(S201T) that was shown to have exactly the same 3D structure and thermodynamic stability as WT GlpG [80][57].

Of the methods that are in principle available to measure protein-ligand interactions, we chose a new method that has recently become available at the IOCB, microscale thermophoresis (MST), because it is a label-free method that can detect ligand binding events at low concentrations in solution by measuring the intrinsic fluorescence of tryptophan residues, and can be used also for membrane proteins in detergent micelles. MST is sensitive to changes in the hydration shell and molecular size of the analyte in a complex manner, so some interaction produce a strong response in thermophoretic behaviours, while others less so. For a brief description of the physical principle behind MST refer to the Methods section 4.3.11. To measure the apparent dissociation constant of our peptide series, 1 mM peptide solutions were prepared in the measurement buffer and diluted 1:1 in serial dilutions sixteen times so that the final concentration of peptides was from 500 μM to 15.26 μM and the final concentration of enzyme was kept constant at 1.7 μM. Representative MST data for KSp26-31 (without KSp30) and GlpG S201T are presented below (Fig.26 A, B; C-E page 62).



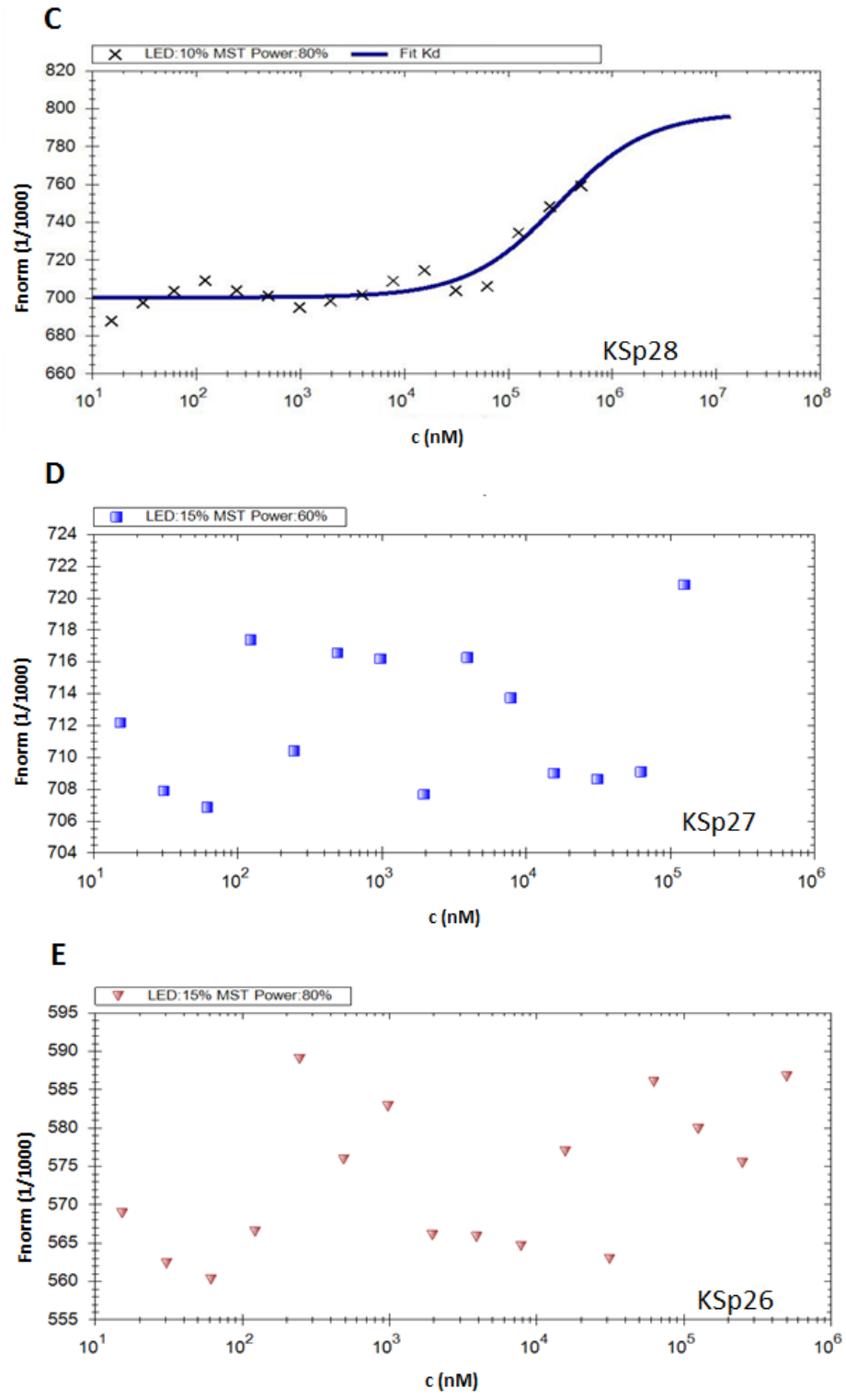


Figure 26: Representative graphs of MST measurement of K_D for peptides KSp26, 27, 28, 29 and 31 with GlpG S201T. A) A representative graph of KD measurement of KSp31. K_D for KSp31-GlpG S201T was determined by three independent measurements to be ~40 µM. B) For peptide KSp29 the increase in K_D was observed to be around 10 µM but it was shown to be 6 times higher for the KSp28 (C). No trend was observed for KSp26 and 27(D, E). Nevertheless our measurement was limited by the solubility of peptides so we can better compere the dissociation constants than determine precise values.

We have detected a reproducible MST response for peptide KSp31 from three independent measurements, yielding a K_D value of 40 μM. Given the solubility limit of KSp31 (and also of the other peptides of the series), we were not able to determine the exact value of K_D because we did not obtain enough datapoints in the upper plateau phase of the binding curve. Nevertheless, the measurements were reproducible (done at least in duplicates) and we did observe a clear trend of decreasing affinity with decreasing TMH length: the K_D value determined from the MST data for KSp28 was about six times higher than for KSp31. Thus, we found that the affinity of the substrate to GlpG depends on the length of its transmembrane helix, which at least partly explains our CE data. Interestingly, we were not able to observe any MST response to peptides KSp26 and 27, although we had observed their relatively efficient cleavage by CE. Although counter-intuitive, it could be explained by the possibility that as these peptides may be cleaved at a different site from KSp28-31, the interaction of KSp26 and 27 with GlpG may elicit a dramatically less robust MST response, than the 'correct' interaction of KSp31

5.2.4 Substrate binding at TMH2 and 5 in GlpG: lateral gate opening or intramembrane exosite?

Having found that the TMH of the substrate is responsible for the affinity to GlpG, it implied that it interacted with a specific site on GlpG. Current consensus in the field is that substrates enter GlpG between its TMH2 and TMH5. Mutations at the interface of TMH2 and TMH5, especially two double mutants, F153A (TMH5) and W236A (TMH2), and the second one W157A (TMH5) and F232A (TMH2), were reported to increase GlpG activity dramatically [81]. These amino acids mediate interactions between TMH2 and TMH5 and the effect of these double mutants was interpreted as that the weakening of TMH2 and TMH5 interactions results in the 'lateral opening' or 'lateral gating' of the enzyme to facilitate substrate access, in analogy with the function of the Sec translocase. However, other studies have shown that mobility of TMH5 is not necessary for activity of GlpG [62]. We therefore hypothesised that the residues F153, W236, W157 and F232 could also mediate the interaction with the TMH of the substrate. Their mutation to alanine could thus result in a change of substrate specificity, which could be an alternative explanation of the observed 'activation effect' of the mentioned double mutants. The prediction of this model

would be that the ‘activation effect’ of the double mutants would show a different magnitude for different substrates (that differ in the sequence of their TMH).

To test this hypothesis, we compared the initial reaction rates of cleavage of the four main model rhomboid substrates TatA, LacYTM2, Gurken and Spitz by the WT GlpG and its two double mutants at the interface of TMH2 and TMH5, F153A/W236A (also termed ‘upper mutant’ henceforth), and W157A/F232A (also termed ‘lower mutant’ henceforth). Concentrations of substrates were kept constant at ~5 μM, while the concentration of enzymes had varied depending on the cleavage efficiency of each substrate. Samples were taken after each 10 min and the reaction was stopped by the SDS-PAGE loading buffer. After the gels were stained by Coomassie blue, washed and scanned (Fig.27), the intensity of substrate bands was densitometrically quantified using the ImageQuant software (GE Healthcare). This experiment required a lot of pre-testing and only the data corresponding to the optimal cleavage conditions were evaluated. An example of the evaluation and data processing, the gel of WT GlpG with TatA substrate is shown.

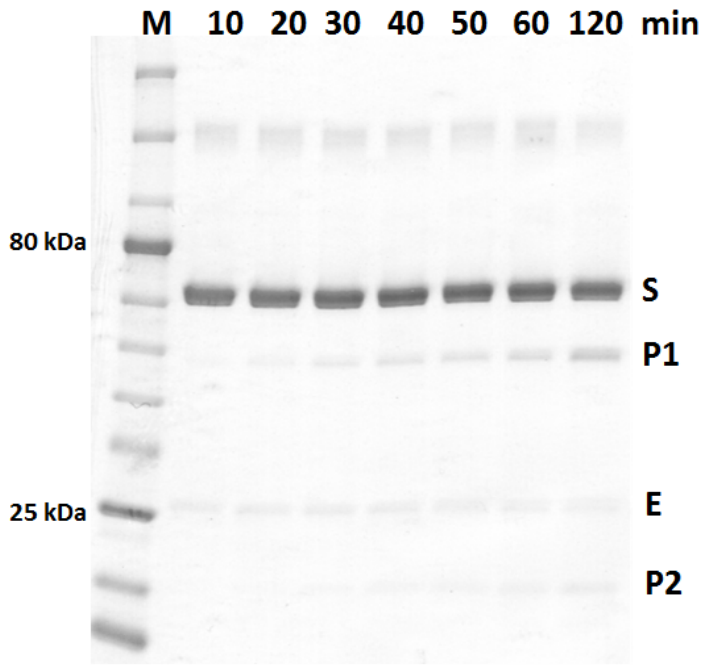


Figure 27: SDS-PAGE on Tris-Glycine gel of time dependent cleavage of TatA substrate by GlpG WT. Time dependent cleavage assay was done. Samples were collected each 10 min and reaction was stopped by SDS-PAGE sample buffer. Substrate (S) conversion into N-terminal cleavage product (P1) and C-terminal cleavage product (P2) was observed. There are also bands corresponding to the enzyme (E) and upper band is an SDS-stable substrate dimer.

The linearity of Coomassie staining was verified by densitometry of a dilution series of a standard pure TatA substrate and the staining was linear up to 3 µg of protein per band. To eliminate the possible loading inaccuracies between gel lanes, substrate conversion was calculated for each lane separately as $\alpha = I_{P1}/(I_{P1}+I_S)$ (see data processing section 4.3.12). Substrate conversion was recalculated into micro-molar values by its multiplication by the exact concentration of substrate in the reaction calculated from on the concentration of stock solution of substrate as determined by amino quantitative amino acid analysis. These values were plotted against time (Fig.28), and the slope of these curves at time zero represents the initial reaction rate of proteolysis.

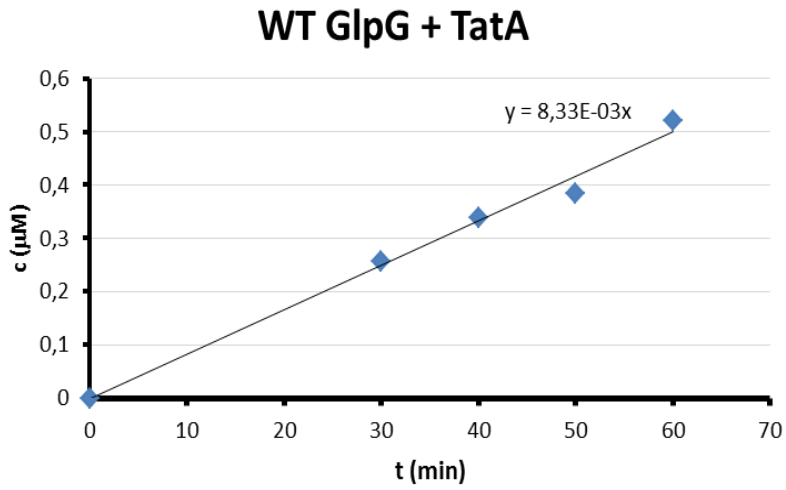


Figure 28: Determination of initial rate of reaction for pair WT GlpG and substrate TatA. Molar conversion of substrate was plotted against time and initial rate of reaction was determined as the slope of linear regression through the data points.

The substrate was in molar excess over the enzyme by 9:1 or higher to approximate steady-state conditions. Only the data from the first ~30% conversion of the substrate were used for the determination of the initial reaction rate by linear regression. The initial reaction rate was divided by the concentration of the enzyme (calculated from amino acid analysis) yielding molar catalytic activity. The whole process was repeated for all substrate-enzyme pairs. Since the activity of the mutants varied on different substrates, enzyme concentration was set after pretesting differently for some enzyme-substrate pairs. The cleavage efficiency of Spitz substrate was too low to be evaluated densitometrically after SDS-PAGE. As we needed to maintain a molar excess of substrate over the enzyme (at least 9:1 in our setup), it was not possible to improve cleavage efficiency of Spitz by

increasing enzyme concentration. The calculation of molar catalytic activity for each substrate-enzyme pair allowed us to compare their activities. To facilitate the interpretation of the data, the molar catalytic activities of double mutants were divided by that of wild type enzyme and the relative activities of these mutants on three different substrates are shown in Figure 29.

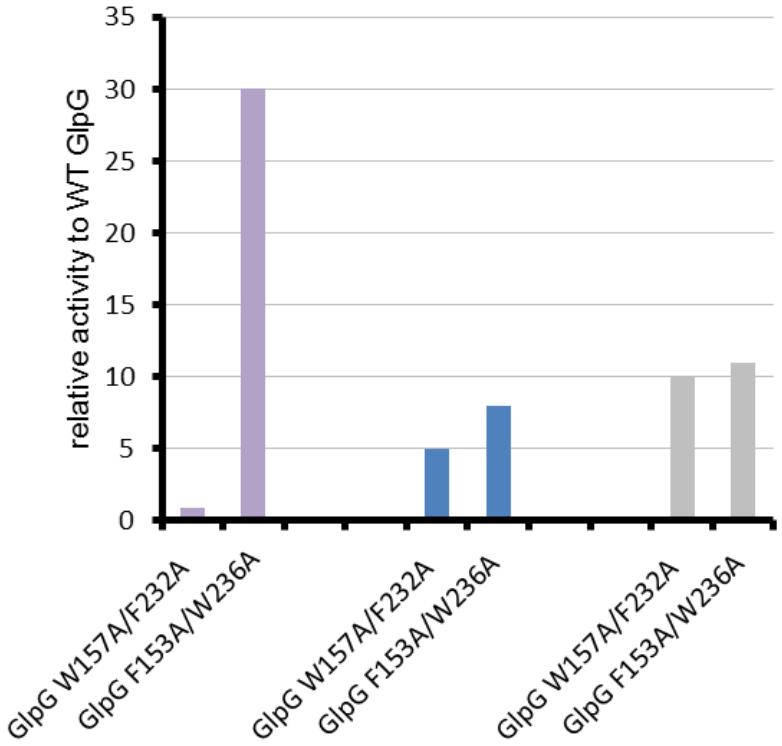


Figure 29: Activation effect of W157A/F232A and F153A/W236A mutants of GlpG on three different substrates. Activation effect of double mutants is substrate-specific. Compared to the WT GlpG, both mutants display better cleavage efficiency for all substrates except for the pair GlpG W157A/F232A - LacY.

Except for the lower mutant of GlpG (W157A/F232A), which cleaves the LacY substrate with the same efficiency as the WT GlpG, the 'activation effect' was observed for both mutants on all substrates. The highest activation effect was observed for the pair of the upper mutant (F153A/W236A) with LacY substrate where molar catalytic activity has increased 30-fold compared to WT GlpG. Importantly, the magnitude of the activation effect differs between the three substrates, and while the upper and lower mutants have the same effect on TatA and Gurken, they have a markedly different effect on LacYTM2. This suggests that our hypothesis may be correct: their mutations of the residues at the interface

between TMH2 and 5 result in a change of GlpG substrate specificity implying that these residues may directly interact with substrate TMH. To better visualise the change of substrate specificity of the mutants, the data can be replotted with respect to the enzyme used (Fig30).

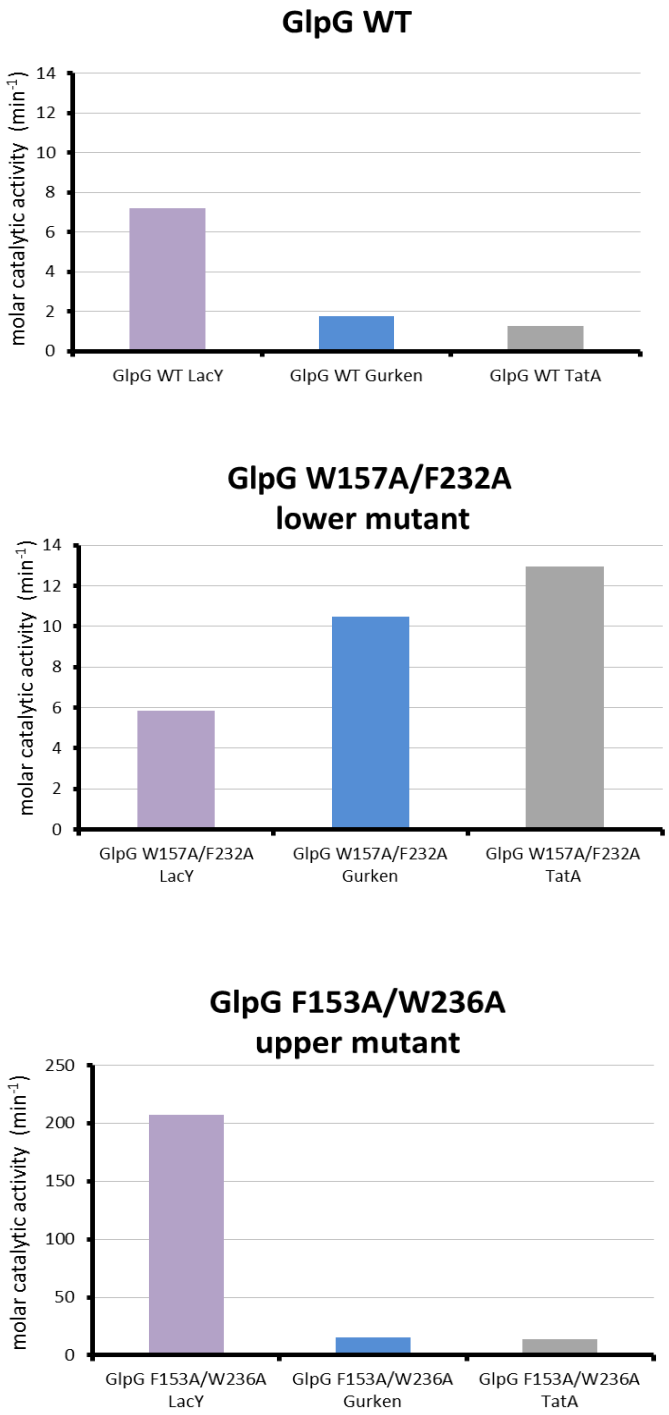


Figure 30: Comparison of specificity of different GlpG variants. Specificity of WT GlpG was determined. Mutant F153A/W2236A has had the same specificity as WT GlpG even thought the activation effect of cleavage has varied. Surprisingly, mutant W157A/F232A displays significantly different specificity from WT GlpG.

LacY was shown to be the best substrate of WT GlpG followed by Gurken and TatA. The profiles in each graph show that the upper mutant (F153A/W236A) has similar substrate preferences as WT GlpG, while the lower mutant (W157A/F232A) displays markedly different, nearly opposite substrate preferences. Although the densitometric evaluation was not possible for the Spitz substrate due to very low cleavage efficiency, its visual analysis indicated that the cleavage efficiency by F153A/W236A would be much higher than by WT GlpG while W157A/F232A would cleave Spitz with about the same efficiency as would WT GlpG (data not shown), which would result in a similar profile to LacYTM2 substrate.

5.3 DEVELOPMENT OF FLUOROGENIC TRANSMEMBRANE SUBSTRATE FOR RHOMBOIDS

Surprisingly, no widely useable, reliable, fluorogenic or chromogenic transmembrane peptide substrate that would be useable for enzyme kinetics and high-throughput screening of inhibitors has been reported for rhomboids to date. A short fluorogenic peptide substrate has been published for AarA [82], but this TatA derived peptide, which lacks most of the transmembrane helix of the parent TatA protein, is not cleaved by GlpG and only poorly so by other bacterial rhomboids (data not shown). GlpG is currently the only rhomboid protease amenable to crystallographic analysis, and inhibitor development would be thus conveniently done on GlpG, but the absence of a quantitative, continuous, high-throughput activity assay is limiting.

Based on the cleavage of the four model protein substrates described above and the cleavage of the LacYTM2-derived peptide series (KSp26-31) described earlier (section 5.2.2), LacYTM2-derived peptide KSp31 was chosen as the best TMH peptide template for the development of a more generally usable fluorogenic substrate. As a pilot test, KSp31 was modified by introducing a commercially available fluorescence resonance energy transfer (FRET) pair EDANS and DABCYL as the fluorophore and quencher, into the P5 and P4´ positions, via amino acid derivatives, respectively, directly during solid state peptide synthesis (done by the IOCB peptide synthesis service) (Tab.5; page 69). These positions were chosen because our specificity analysis of GlpG (unpublished data from the Strisovsky laboratory) shows that they are not very sensitive to substitutions.

Table 5: Sequence of fluorogenic substrate KSp35. Two mutations of KSp31 peptide were done (represented in bold) to allow the attachment of FRET pair. EDANS fluorophore was bound to P5 position (grey E) and DABCYL quencher was conjugated at P4' position (yellow K). The cleavage site is represented by dash.

peptide	sequence
KSp31	KRHDIN**H**ISKS-DTG**I**IFAAISLFSLLFQPLFGLSK
KSp35	KRHDIN**E**ISKS-DTG**K**IFAAISLFSLLFQPLFGLSK | | EDANS DABCYL

First, the solubility of KSp35 in the assay buffer (20 mM HEPES buffer containing 0.05% (w/v) DDM and 10% (v/v) DMSO) was tested by making serial dilutions of KSp35 to 100 μM; 75 μM; 50 μM; 30 μM; 10 μM and 1 μM. After incubation at 37°C for 2 h, tubes were centrifuged at 14000 rpm for 15 min and absorbance of the supernatant was measured at 480 nm. Linear relationship between absorbance and nominal concentration in the supernatant indicated that KSp35 was soluble in the whole concentration range tested (Fig.31).

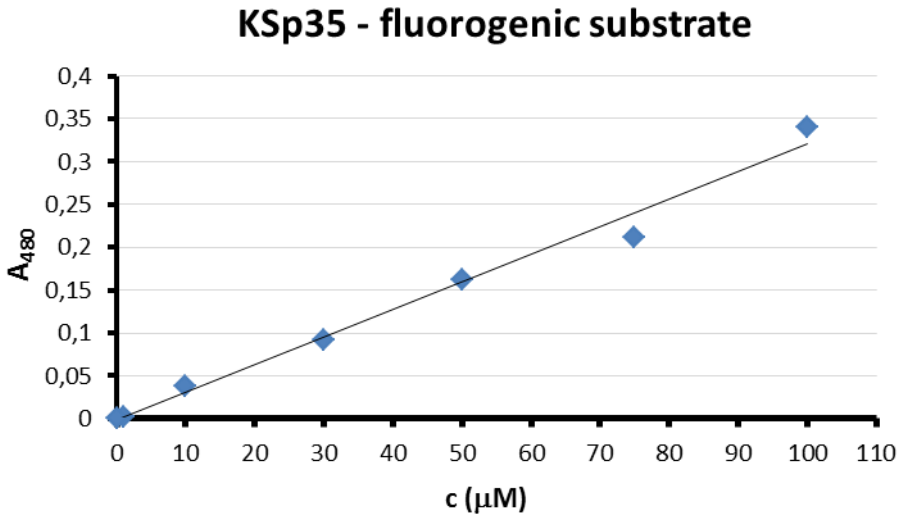


Figure 31: Solubility of fluorogenic substrate. Linear tendency of absorbance measured at 480 nm confirmed solubility of fluorogenic peptide in the range of concentration form 1-100 μM **(**100 μM; 75 μM; 50 μM; 30 μM; 10 μM and 1 μM).

Second, background fluorescence of the substrate was measured at the excitation maximum of EDANS corresponding to 335 nm for same concentrations 100 μM; 75 μM; 50 μM; 30 μM; 10 μM and 1 μM (Fig.32). Fluorescence increases linearly with the concentration up to 30 μM. This concentration was determined as the maximal concentration to be used in all cleavage assays performed with this fluorogenic substrate.

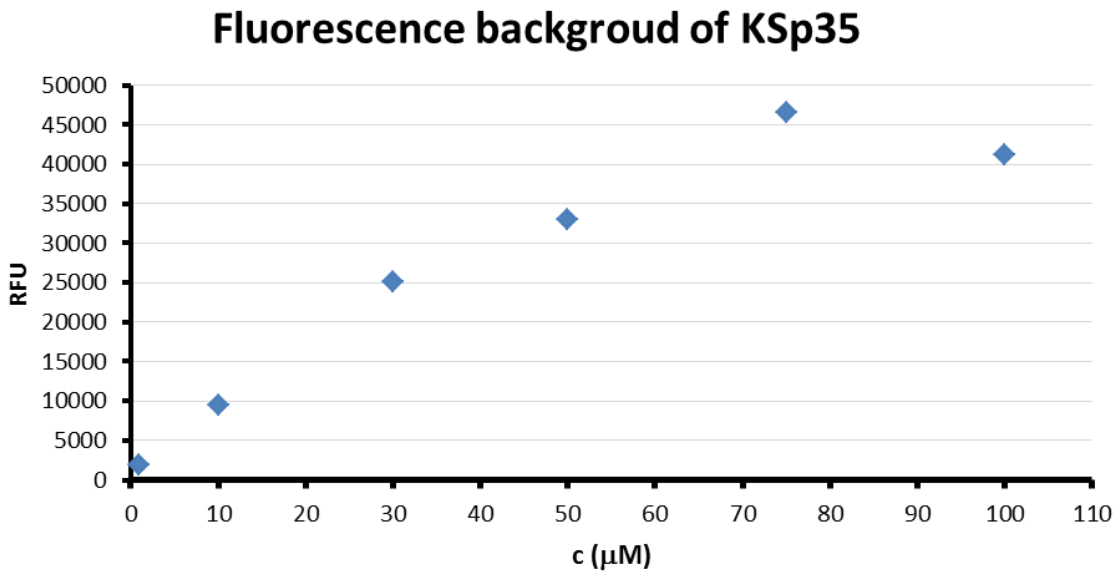


Figure 32: Fluorescent background of KSp35. Fluorescence at excitation wave length 335 nm and emission wave length 493nm was measured for concentrations 100 μM; 75 μM; 50 μM; 30 μM; 10 μM and 1 μM of KSp35. Linear range was observed up to the concentration 30 μM peptide.

Third, KSp35 was tested for cleavage with a panel of bacterial rhomboids. Twenty five μM of substrate was exposed to 0.8 μM rhomboids AarA (*P.stuartii*), GlpG (*E.coli*) and two rhomboids from *Bacteroides thetaiotaomicron*. Inactive mutant S201A of GlpG was used as a negative control (Fig.33; page 71). Substrate was shown to be cleaved by all used active rhomboids, albeit at different efficiencies, which suggests that it will be a widely useable rhomboid substrate.

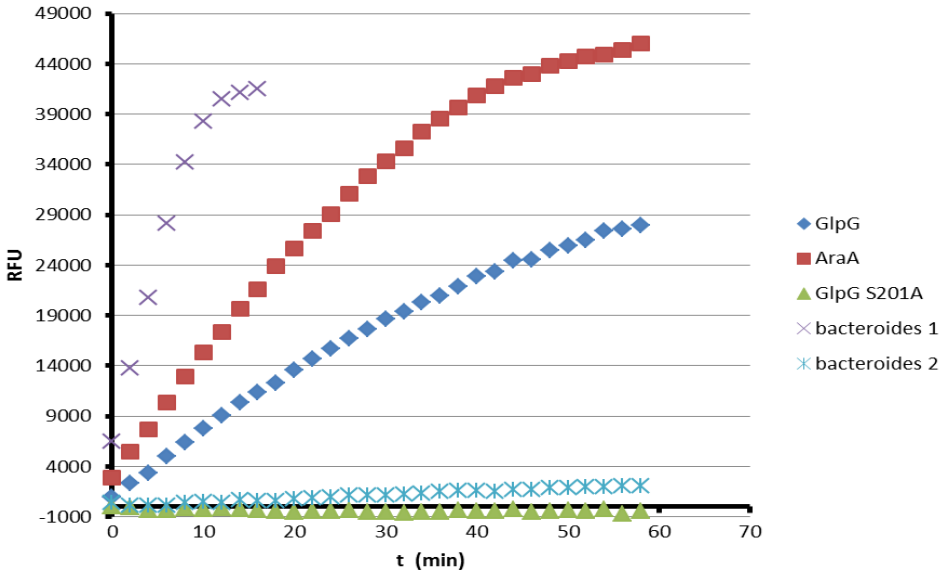


Figure 33: Cleavage of fluorogenic peptide KSp35 by different rhomboids. KSp35 was shown to be cleaved by a panel of different rhomboids. Inactive GlpG S201A was used as a negative control. Fluorescence measurement was performed using excitation wave length 335 nm and emission wave length 493nm

6 DISCUSSION

The *E.coli* rhomboid protease GlpG is at present the main structural and mechanistic model for the whole rhomboid family of proteins, and analysis of substrate recognition and cleavage mechanism by GlpG is thus crucial for a better understanding of rhomboid proteases and rhomboid-like proteins and their interaction with client proteins. An X-ray structure of GlpG with the substrate bound would elucidate rhomboid-substrate interactions but as it has not been solved so far, all models of this interaction are based on enzymological studies and on the structures of unliganded GlpG or GlpG complexes with generic serine protease small-molecular inhibitors [71][82]. The extent and nature of rhomboid-substrate interactions and the basis of rhomboid substrate specificity are still a matter of contention. This thesis addresses both of these disputed issues using kinetics measurements and site-directed mutagenesis.

From the structure of GlpG it is obvious that the active site is positioned inside the membrane, encircled by transmembrane helices from all sides and protected by the loop 5 from the upper side (Fig.13; page26). A conformational change seems to be required so that the substrate can access the active site. Since the substrates are transmembrane, single TMH proteins, a substrate will approach the enzyme laterally from the lipid membrane, and one would expect that the first contact between enzyme and substrate would occur inside the lipid bilayer. One study demonstrated that three bacterial rhomboids require two elements in their substrates for recognition, the transmembrane helix and a short sequence motif (recognition motif) that determines the site of cleavage [70]. In contrast, other studies show that there is no sequence recognition, and the main determinant of rhomboid substrates is helical instability in their TMH [81] [69]. Other major study has shown that catalytic efficiency of rhomboids is mainly due to the differences in k_{cat}, which is determined by the recognition motif of the substrate [70] mentioned above, and that the K_M of rhomboid proteases is relatively high, around 100 μM [72], concluding that the interaction between the enzyme and substrate occurring at the TMH level is nearly negligible [72]. As a result of these findings, one extreme model in the field suggests that the TMH of the substrate does not contribute to rhomboid specificity at all and it does not even bind the protease [62].

We therefore sought to clarify the importance of the transmembrane helix of substrate for its recognition by rhomboid, and determine the extent of this interaction. We

designed a series of C-terminally truncated peptides derived from a model substrate LacYTM2 varying in the TMH length, and measured their cleavage efficiency by wild type GlpG. Our results showed that the cleavage efficiency of the substrate increased proportionally with the length of its TMH, and that the full-length-TMH peptide KSp31 was cleaved with much higher efficiency than the rest of the series. From this result it seems that full length TMH is markedly preferred by GlpG. Moreover, substrate peptide variants lacking 16 (KSp26) and 13 (KSp27) amino acids from the C-terminus of the TMH produced different cleavage products than the rest of the series. In the case of KSp26, only the incorrect product was formed (as observed by MALDI and CE), but for KSp27, the correct cleavage product was formed in very small amounts besides the incorrect cleavage product formed from KSp26. Thus, the TMH of the substrate influences not only the efficiency of cleavage, but also the selection of cleavage site, probably by directing the proper positioning of substrate's recognition motif into the active site of the enzyme.

To identify whether the substrate's TMH influenced the efficiency of its binding to rhomboid, we investigated the K_D of the C-terminally TMH-truncated peptides that could be correctly cleaved by GlpG, using microscale thermophoresis (MST). A low K_D value means a higher stability of the enzyme-substrate complex; the binding constant of the longest peptide KSp31 was determined to be around ~40 μM and the approximate K_D of the shortest cleaved peptide was about six-fold higher. We observed a clear negative correlation between the binding data and the resulting (approximate but reproducible) K_D and the length of the TMH of the peptides, although we were not able to measure the exact K_D of the shorter peptides because of their weak binding and limited solubility, which made it impossible to measure whole sigmoidal binding curve. Surprisingly, the peptides whose cleavage site was different from the rest of the series (KSp26-27), did not elicit any binding response in MST, although both were cleaved efficiently by GlpG. This may be due to the fact that they are most likely cleaved at a site different from the other peptides of the series, and this different interaction may not elicit a robust MST response. Thus, the data from microscale thermophoresis and CE showed the importance of the TMH of the substrate for the binding constant to the enzyme, but also for presenting the substrate's recognition motif region to the enzyme active site.

Based on the above results we decided to develop fluorogenic substrate using the very well and accurately cleaved KSp31. This peptide was conjugated with EDANS-DABCYL FRET pair by the IOCB medicinal chemistry group. Although the

peptide was soluble up to 100 μM, the useful range of concentrations was limited to 30 μM due to the autofluorescent background. The substrate (KSp35) was shown to be widely useable, as it was cleaved by four significantly different bacterial rhomboids. This substrate is currently used for routine IC_{50} measurement during inhibitor development in the laboratory of the thesis advisor.

Once the importance of substrate's TMH for the recognition by rhomboid was confirmed, we asked how the interactions with substrate's TMH influence rhomboid specificity. There are two major mechanistic models proposing different conformational changes required for substrate binding. Both of them are based on X-ray structures. One model suggests the existence of two conformational stages of the enzyme - "open" and "closed" [65][66]. Two double mutants were designed to induce open conformation of the enzyme. Both double mutants had mutations at the interface between TMH2 and 5, which is the hypothesised place of substrate entry into the active site. Both double mutants exhibited increased enzymatic activity compared to the wild type enzyme (for more details see 2.3.4). This 'activation effect' had been attributed only to the conformational change (to the opening of enzyme) [65][66]. A recently published study has however shown that a mutant of GlpG in which TMH 2 and 5 are demonstrably cross-linked (and which thus cannot adopt the 'open' conformation), can cleave the Gurken substrate with the same efficiency as the wild type enzyme [62] . Our working hypothesis therefore was that the enhanced activity of the 'activation mutants' is conferred not by the opening of enzyme, but by a change of substrate specificity.

In fact, initial reaction rate measurements in the present work have shown that the activation effect of both double mutants differs depending on the used substrate. The 'upper' mutant F153A/W236A has shown a 30-fold better cleavage of the LacYTM2 substrate, 14-fold higher cleavage of Gurken, and 11-fold better cleavage of TatA compared to the wild type enzyme, which means it has a similar substrate specificity as the wild type GlpG. In contrast, the 'lower' mutant W157A/F232A has shown no change in the cleavage efficiency of LacYTM2, 5-fold better cleavage of Gurken and 10-fold increased activity on TatA substrate compared to the wild type enzyme, which represents a markedly different substrate specificity to that of wild-type GlpG (and of the upper mutant). Overall, our analysis shows that the interaction of rhomboid with the transmembrane helix of the substrate contributes significantly to the affinity to GlpG, to the cleavage-site selection and to the overall substrate specificity of GlpG. These

observations are consistent with the hypothesis that an intramembrane exosite of rhomboid, which binds the TMD of the substrate and that is physically separated from the active site [70], is one of the two key elements determining its substrate specificity.

To dissect these issues in more detail, we would suggest to perform an experiment where the TMH of a good substrate could be gradually replaced by the sequence of a poor substrate TMH. Another possibility is to slide a four amino acid sequence window (representing ~one turn of an α-helix) of a poor substrate along the TMH sequence of a good substrate using site-directed mutagenesis and to observe the effect on the cleavage rate and affinity, or, ideally, on K_M, k_{cat} and K_D. Ultimately a structural approach using a combination of X-ray crystallography and NMR, would be addressing all these questions objectively and at high spatial and temporal resolution, and they are technically challenging and, are well beyond the scope of the present work.

7 CONCLUSIONS

1) Eight transmembrane proteins (four different substrates and four variants of the GlpG rhomboid protease), were overexpressed in *E.coli*. Substrates were purified to homogeneity via their affinity tags by NiNTA and amylose resin affinity chromatography and rhomboid enzymes were purified by TALON^{TM} and size exclusion chromatography.

2) Transmembrane helix (TMH) of the substrate plays a role in cleavage site recognition. Only those substrate-derived peptides with 13 and more TMH amino acids were cleaved at the same position as the full-length parent substrate LacYTMH2.

3) The cleavage rate of substrates depends proportionally on their TMH length

4) Substrate TMH increases the affinity of the substrate to the rhomboid enzyme. Substrate peptides exhibiting cleavage at the correct cleavage site display a decrease in their K_D (i.e. increase in affinity) with the increasing length of their TMH.

5) Residues W236, F153, F232 and W157 in GlpG may interact directly with substrate TMH and substrate's TMH contributes to substrate specificity of rhomboid via this interaction. In other words, the mentioned residues could be forming the intramembrane exosite for substrate TMH, which contributes to rhomboid specificity.

6) Based on the results summarised in conclusions 2)-4) we developed soluble and sensitive fluorogenic peptide substrate useable for several bacterial rhomboids and suitable for high throughput screening for rhomboid inhibitors.

8 REFERENCES

[1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, in book Molecular Biology of the Cell 5th edition, 5th editon, Garland Science, (1994), 629-42

[2] J.L. Popot, D.M. Engelman, Helical membrane protein folding, stability, and evolution, Annu. Rev. Biochem., 69 (2000) 881–922.

[3] G. Heijne, The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology, EMBO J., 5 (1986) 3021–7.

[4] J. Nilsson, B. Persson, G. von Heijne, Comparative analysis of amino acid distributions in integral membrane proteins from 107 genomes, Proteins, 60 (2005) 606–16.

[5] N.N. Alder, A.E. Johnson, Cotranslational membrane protein biogenesis at the endoplasmic reticulum, J. Biol. Chem., 279 (2004) 22787–90.

[6] T.A. Rapoport, V. Goder, S.U. Heinrich, K.E.S. Matlack, Membrane-protein integration and the role of the translocation channel, Trends Cell Biol., 14 (2004) 568–75.

[7] M. Higy, T. Junne, M. Spiess, Topogenesis of membrane proteins at the endoplasmic reticulum, Biochemistry, 43 (2004) 12716–22.

[8] J. Arribas, A. Borroto, Protein ectodomain shedding, Chem. Rev., 102 (2002) 4627–38.

[9] J.C. Becker, R. Dummer, A.A. Hartmann, G. Burg, R.E. Schmidt, Shedding of ICAM-1 from human melanoma cell lines induced by IFN-gamma and tumor necrosis factor-alpha Functional consequences on cell-mediated cytotoxicity, J. Immunol., 147 (1991) 4398–401.

[10] S. V Subramanian, M.L. Fitzgerald, M. Bernfield, Regulated shedding of syndecan-1 and -4 ectodomains by thrombin and growth factor receptor activation, J. Biol. Chem., 272 (1997) 14713–20.

[11] V. Matthews, B. Schuster, S. Schütze, I. Bussmeyer, A. Ludwig, C. Hundhausen, T. Sadowski, P. Saftig, D. Hartmann, K.-J. Kallen, S. Rose-John, Cellular cholesterol depletion triggers shedding of the human interleukin-6 receptor by ADAM10 and ADAM17 (TACE), J. Biol. Chem., 278 (2003) 38829–39.

[12] M.L. Fitzgerald, Z. Wang, P.W. Park, G. Murphy, M. Bernfield, Shedding of syndecan-1 and -4 ectodomains is regulated by multiple signaling pathways and mediated by a TIMP-3-sensitive metalloproteinase, J. Cell Biol., 148 (2000) 811–24.

[13] S.M. Dethlefsen, G. Raab, M.A. Moses, R.M. Adam, M. Klagsbrun, M.R. Freeman, Extracellular calcium influx stimulates metalloproteinase cleavage and secretion of heparin-binding EGF-like growth factor independently of protein kinase C, J. Cell. Biochem., 69 (1998) 143–53.

[14] Y. Chen, A. Hayashida, A.E. Bennett, S.K. Hollingshead, P.W. Park, Streptococcus pneumoniae sheds syndecan-1 ectodomains through ZmpC, a metalloproteinase virulence factor, J. Biol. Chem., 282 (2007) 159–67.

[15] K. Hayashida, A.H. Bartlett, Y. Chen, P.W. Park, Molecular and cellular mechanisms of ectodomain shedding, Anat. Rec. (Hoboken)., 293 (2010) 925–37.

[16] H. Nagase, R. Visse, G. Murphy, Structure and function of matrix metalloproteinases and TIMPs, Cardiovasc. Res., 69 (2006) 562–73.

[17] C. Adrain, M. Zettl, Y. Christova, N. Taylor, M. Freeman, Tumor necrosis factor signaling requires iRhom2 to promote trafficking and activation of TACE, Science, 335 (2012) 225–8.

[18] S. Urban, Mechanisms and cellular functions of intramembrane proteases, Biochim. Biophys. Acta, 1828 (2013) 2797–800.

[19] M.S. Brown, J. Ye, R.B. Rawson, J.L. Goldstein, Regulated intramembrane proteolysis: a control mechanism conserved from bacteria to humans, Cell, 100 (2000) 391–8.

[20] R.B. Rawson, N.G. Zelenski, D. Nijhawan, J. Ye, J. Sakai, M.T. Hasan, T.Y. Chang, M.S. Brown, J.L. Goldstein, Complementation cloning of S2P, a gene encoding a putative metalloprotease required for intramembrane cleavage of SREBPs, Mol. Cell, 1 (1997) 47–57.

[21] X. Wang, R. Sato, M.S. Brown, X. Hua, J.L. Goldstein, SREBP-1, a membrane-bound transcription factor released by sterol-regulated proteolysis, Cell, 77 (1994) 53–62.

[22] T. Vaccari, H. Lu, R. Kanwar, M.E. Fortini, D. Bilder, Endosomal entry regulates Notch receptor activation in Drosophila melanogaster, J. Cell Biol., 180 (2008) 755–62.

[23] O.M. Andersen, J. Reiche, V. Schmidt, M. Gotthardt, R. Spoelgen, J. Behlke, C.A.F. von Arnim, T. Breiderhoff, P. Jansen, X. Wu, K.R. Bales, R. Cappai, C.L. Masters, J. Gliemann, E.J. Mufson, B.T. Hyman, S.M. Paul, A. Nykjaer, T.E. Willnow, Neuronal sorting protein-related receptor sorLA/LR11 regulates processing of the amyloid precursor protein, Proc. Natl. Acad. Sci. U. S. A., 102 (2005) 13461–6.

[24] B. De Strooper, P. Saftig, K. Craessaerts, H. Vanderstichele, G. Guhde, W. Annaert, K. Von Figura, F. Van Leuven, Deficiency of presenilin-1 inhibits the normal cleavage of amyloid precursor protein, Nature, 391 (1998) 387–90.

[25] R. Kopan, M.X.G. Ilagan, The canonical Notch signaling pathway: unfolding the activation mechanism, Cell, 137 (2009) 216–33.

[26] A. Haapasalo, D.M. Kovacs, The many substrates of presenilin/γ-secretase, J. Alzheimers. Dis., 25 (2011) 3–28.

[27] H.W. Querfurth, F.M. LaFerla, Alzheimer's disease, N. Engl. J. Med., 362 (2010) 329–44.

[28] A. Weihofen, K. Binns, M.K. Lemberg, K. Ashman, B. Martoglio, Identification of signal peptide peptidase, a presenilin-type aspartic protease, Science, 296 (2002) 2215–8.

[29] F. El Hage, V. Stroobant, I. Vergnon, J.-F. Baurain, H. Echchakir, V. Lazar, S. Chouaib, P.G. Coulie, F. Mami-Chouaib, Preprocalcitonin signal peptide generates a cytotoxic T lymphocyte-defined tumor epitope processed by a proteasome-independent pathway, Proc. Natl. Acad. Sci. U. S. A., 105 (2008) 10119–24.

[30] D.R. Beisner, P. Langerak, A.E. Parker, C. Dahlberg, F.J. Otero, S.E. Sutton, L. Poirot, W. Barnes, M.A. Young, S. Niessen, T. Wiltshire, U. Bodendorf, B. Martoglio, B. Cravatt, M.P. Cooke, The intramembrane protease Sppl2a is required for B cell and DC development and survival via cleavage of the invariant chain, J. Exp. Med., 210 (2013) 23–30.

[31] M.K. Lemberg, B. Martoglio, Requirements for signal peptide peptidase-catalyzed intramembrane proteolysis, Mol. Cell, 10 (2002) 735–44.

[32] S. Urban, J.R. Lee, M. Freeman, Drosophila rhomboid-1 defines a family of putative intramembrane serine proteases, Cell, 107 (2001) 173–82.

[33] J.D. Wasserman, S. Urban, M. Freeman, A family of rhomboid-like genes: Drosophila rhomboid-1 and roughoid/rhomboid-3 cooperate to activate EGF receptor signaling, Genes Dev., 14 (2000) 1651–63.

[34] M.K. Lemberg, M. Freeman, Functional and evolutionary implications of enhanced genomic analysis of rhomboid intramembrane proteases, Genome Res., 17 (2007) 1634–46.

[35] Y. Christova, C. Adrain, P. Bambrough, A. Ibrahim, M. Freeman, Mammalian iRhoms have distinct physiological functions including an essential role in TACE regulation, EMBO Rep., 14 (2013) 884–90.

[36] L.G. Stevenson, K. Strisovsky, K.M. Clemmer, S. Bhatt, M. Freeman, P.N. Rather, Rhomboid protease AarA mediates quorum-sensing in Providencia stuartii by activating TatA of the twin-arginine translocase, Proc. Natl. Acad. Sci. U. S. A., 104 (2007) 1003–8.

[37] M. Gallio, G. Sturgill, P. Rather, P. Kylsten, A conserved mechanism for extracellular signaling in eukaryotes and prokaryotes, Proc. Natl. Acad. Sci. U. S. A., 99 (2002) 12208–13.

[38] S. Urban, D. Schlieper, M. Freeman, Conservation of intramembrane proteolytic activity and substrate specificity in prokaryotic and eukaryotic rhomboids, Curr. Biol., 12 (2002) 1507–12.

[39] A. Dutt, S. Canevascini, E. Froehli-Hoier, A. Hajnal, EGF signal propagation during C elegans vulval development mediated by ROM-1 rhomboid, PLoS Biol., 2 (2004) e334.

[40] M. Herlan, F. Vogel, C. Bornhovd, W. Neupert, A.S. Reichert, Processing of Mgm1 by the rhomboid-type protease Pcp1 is required for maintenance of mitochondrial morphology and of mitochondrial DNA, J. Biol. Chem., 278 (2003) 27781–8.

[41] J.-R. Chao, E. Parganas, K. Boyd, C.Y. Hong, J.T. Opferman, J.N. Ihle, Hax1-mediated processing of HtrA2 by Parl allows survival of lymphocytes and neurons, Nature, 452 (2008) 98–102.

[42] J.C. Pascall, K.D. Brown, Intramembrane cleavage of ephrinB3 by the human rhomboid family protease, RHBDL2, Biochem. Biophys. Res. Commun., 317 (2004) 244–52.

[43] O. Lohi, S. Urban, M. Freeman, Diverse substrate recognition mechanisms for rhomboids; thrombomodulin is cleaved by Mammalian rhomboids, Curr. Biol., 14 (2004) 236–41.

[44] C. Adrain, K. Strisovsky, M. Zettl, L. Hu, M.K. Lemberg, M. Freeman, Mammalian EGF receptor activation by the rhomboid protease RHBDL2, EMBO Rep., 12 (2011) 421–7.

[45] Y. Wang, W. Song, S. Li, X. Guan, S. Miao, S. Zong, S.S. Koide, L. Wang, GC-1 mRHBDD1 knockdown spermatogonia cells lose their spermatogenic capacity in mouse seminiferous tubules, BMC Cell Biol., 10 (2009) 25.

[46] L. Fleig, N. Bergbold, P. Sahasrabudhe, B. Geiger, L. Kaltak, M.K. Lemberg, Ubiquitin-dependent intramembrane rhomboid protease promotes ERAD of membrane proteins, Mol. Cell, 47 (2012) 558–69.

[47] S. Urban, M. Freeman, Substrate specificity of rhomboid intramembrane proteases is governed by helix-breaking residues in the substrate transmembrane domain, Mol. Cell, 11 (2003) 1425–34.

[48] L.D. Sibley, Intracellular parasite invasion strategies, Science, 304 (2004) 248–53.

[49] F. Brossier, G.L. Starnes, W.L. Beatty, L.D. Sibley, Microneme rhomboid protease TgROM1 is required for efficient intracellular growth of Toxoplasma gondii, Eukaryot. Cell, 7 (2008) 664–74.

[50] B.N. Lilley, H.L. Ploegh, A membrane protein required for dislocation of misfolded proteins from the ER, Nature, 429 (2004) 834–40.

[51] E.J. Greenblatt, J.A. Olzmann, R.R. Kopito, Derlin-1 is a rhomboid pseudoprotease required for the dislocation of mutant α-1 antitrypsin from the endoplasmic reticulum, Nat. Struct. Mol. Biol., 18 (2011) 1147–52.

[52] M.H. Smith, H.L. Ploegh, J.S. Weissman, Road to ruin: targeting proteins for degradation in the endoplasmic reticulum, Science, 334 (2011) 1086–90.

[53] V.K. Lazarov, P.C. Fraering, W. Ye, M.S. Wolfe, D.J. Selkoe, H. Li, Electron microscopic structure of purified, active gamma-secretase reveals an aqueous intramembrane chamber and two pores, Proc. Natl. Acad. Sci. U. S. A., 103 (2006) 6889–94.

[54] A. Ben-Shem, D. Fass, E. Bibi, Structural basis for intramembrane proteolysis by rhomboid serine proteases, Proc. Natl. Acad. Sci. U. S. A., 104 (2007) 462–6.

[55] C. Lazareno-Saez, E. Arutyunova, N. Coquelle, M.J. Lemieux, Domain Swapping in the Cytoplasmic Domain of the Escherichia coli Rhomboid Protease, J. Mol. Biol., 425 (2013) 1127–42.

[56] S. Maegawa, K. Koide, K. Ito, Y. Akiyama, 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 (2007) 435–47.

[57] R.P. Baker, S. Urban, Architectural and thermodynamic principles underlying intramembrane protease function, Nat. Chem. Biol., 8 (2012) 759–68.

[58] M.J. Lemieux, S.J. Fischer, M.M. Cherney, K.S. Bateman, M.N.G. James, The crystal structure of the rhomboid peptidase from Haemophilus influenzae provides insight into intramembrane proteolysis, Proc. Natl. Acad. Sci. U. S. A., 104 (2007) 750–4.

[59] Y. Wang, Y. Ha, Open-cap conformation of intramembrane protease GlpG, Proc. Natl. Acad. Sci. U. S. A., 104 (2007) 2098–102.

[60] R.L. Lieberman, M.S. Wolfe, Membrane-embedded protease poses for photoshoot, Proc. Natl. Acad. Sci. U. S. A., 104 (2007) 401–2.

[61] S. Urban, R.P. Baker, In vivo analysis reveals substrate-gating mutants of a rhomboid intramembrane protease display increased activity in living cells, Biol. Chem., 389 (2008) 1107–15.

[62] Y. Xue, Y. Ha, Large lateral movement of transmembrane helix S5 is not required for substrate access to the active site of rhomboid intramembrane protease, J. Biol. Chem., (2013) 0–20.

[63] Y. Xue, Y. Ha, Catalytic mechanism of rhomboid protease GlpG probed by 3,4-dichloroisocoumarin and diisopropyl fluorophosphonate, J. Biol. Chem., 287 (2012) 3099–107.

[64] K.R. Vinothkumar, K. Strisovsky, A. Andreeva, Y. Christova, S. Verhelst, M. Freeman, The structural basis for catalysis and substrate specificity of a rhomboid protease, EMBO J., 29 (2010) 3797–809.

[65] Z. Wu, N. Yan, L. Feng, A. Oberstein, H. Yan, R.P. Baker, L. Gu, P.D. Jeffrey, S. Urban, Y. Shi, Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry, Nat. Struct. Mol. Biol., 13 (2006) 1084–91.

[66] R.P. Baker, K. Young, L. Feng, Y. Shi, S. Urban, Enzymatic analysis of a rhomboid intramembrane protease implicates transmembrane helix 5 as the lateral substrate gate, Proc. Natl. Acad. Sci. U. S. A., 104 (2007) 8257–62.

[67] S. Urban, D. Schlieper, M. Freeman, Conservation of intramembrane proteolytic activity and substrate specificity in prokaryotic and eukaryotic rhomboids, Curr. Biol., 12 (2002) 1507–12.

[68] S. Maegawa, K. Ito, Y. Akiyama, Proteolytic action of GlpG, a rhomboid protease in the Escherichia coli cytoplasmic membrane, Biochemistry, 44 (2005) 13543–52.

[69] Y. Akiyama, S. Maegawa, Sequence features of substrates required for cleavage by GlpG, an Escherichia coli rhomboid protease, Mol. Microbiol., 64 (2007) 1028–37.

[70] K. Strisovsky, H.J. Sharpe, M. Freeman, Sequence-specific intramembrane proteolysis: identification of a recognition motif in rhomboid substrates, Mol. Cell, 36 (2009) 1048–59.

[71] Y. Xue, S. Chowdhury, X. Liu, Y. Akiyama, J. Ellman, Y. Ha, Conformational change in rhomboid protease GlpG induced by inhibitor binding to its S' subsites, Biochemistry, 51 (2012) 3723–31.

[72] S.W. Dickey, R.P. Baker, S. Cho, S. Urban, Proteolysis inside the membrane is a rate-governed reaction not driven by substrate affinity, Cell, 155 (2013) 1270–81.

[73] B. Miroux, J.E. Walker, Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels, J. Mol. Biol., 260 (1996) 289–98.

[74] W. Schaffner, C. Weissmann, A rapid, sensitive, and specific method for the determination of protein in dilute solution, Anal. Biochem., 56 (1973) 502–14.

[75] J.R. Benson, P.E. Hare, O-phthalaldehyde: fluorogenic detection of primary amines in the picomole range Comparison with fluorescamine and ninhydrin, Proc. Natl. Acad. Sci. U. S. A., 72 (1975) 619–22.

[76] S.A.I. Seidel, P.M. Dijkman, W.A. Lea, G. van den Bogaart, M. Jerabek-Willemsen, A. Lazic, J.S. Joseph, P. Srinivasan, P. Baaske, A. Simeonov, I. Katritch, F.A. Melo, J.E. Ladbury, G. Schreiber, A. Watts, D. Braun, S. Duhr, Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions, Methods, 59 (2013) 301–15.

[77] B. Miroux, J.E. Walker, Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels, J. Mol. Biol., 260 (1996) 289–98.

[78] T.M. Joys, H. Kim, o-Phthalaldehyde and the fluorogenic detection of peptides, Anal. Biochem., 94 (1979) 371–7.

[79] H. Schägger, G. von Jagow, Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa, Anal. Biochem., 166 (1987) 368–79.

[80] K.R. Vinothkumar, Structure of rhomboid protease in a lipid environment, J. Mol. Biol., 407 (2011) 232–47.

[81] S.M. Moin, S. Urban, Membrane immersion allows rhomboid proteases to achieve specificity by reading transmembrane segment dynamics, Elife, 1 (2012) e00173.

[82] O. a Pierrat, K. Strisovsky, Y. Christova, J. Large, K. Ansell, N. Bouloc, E. Smiljanic, M. Freeman, Monocyclic β-lactams are selective, mechanism-based inhibitors of rhomboid intramembrane proteases, ACS Chem. Biol., 6 (2011) 325–35.

[83] S.H. White, G. von Heijne, How translocons select transmembrane helices, Annu. Rev. Biophys., 37 (2008) 23–42.

[84] G. von Heijne, Membrane-protein topology, Nat. Rev. Mol. Cell Biol., 7 (2006) 909–18.

[85] K. Strisovsky, Structural and mechanistic principles of intramembrane proteolysis--lessons from rhomboids, FEBS J., 280 (2013) 1579–603.

[86] M. Freeman, Rhomboid proteases and their biological functions, Annu. Rev. Genet., 42 (2008) 191–210.

Svoluji k zapůjčení této práce pro studijní účely a prosím, aby byla řádně vedena evidence zapůjčovatelů.

Jméno a příjmení S adresou	Číslo OP	Datum vypůjčení	Poznámka