Charles University in Prague

Faculty of Science

PhD study program: Biochemistry

Mgr. Jan Škerle

Substrátová specifita, mechanismus a regulace aktivity intramembránových proteas z rodiny rhomboidů

Substrate specificity, mechanism and activity regulation of the rhomboid family intramembrane proteases

Dissertation thesis

Supervisor: Kvido Stříšovský, Ph.D.

Prague, 2020

Prohlášení:

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

Praha, 27.03.2020

……………………….

Mgr. Jan Škerle

I hereby declare that Mgr. Jan Škerle contributed significantly to all scientific publications that are part of this dissertation. He performed experiments and took part in their planning; he interpreted results and participated in manuscript writing. His contributions to the individual publications are quantified below.

1. Substrate binding and specificity of rhomboid intramembrane protease revealed by substrate–peptide complex structures

Sebastian Zoll, Stancho Stanchev, Jakub Began, Jan Škerle, Martin Lepšík, Lucie Peclinovská, Pavel Majer, and Kvido Strisovsky (2014), The EMBO Journal. 33, 2408-2421. –15%

2. Sensitive versatile fluorogenic transmembrane peptide substrates for rhomboid intramembrane proteases

Anežka Tichá, Stancho Stanchev, Jan Škerle, Jakub Began, Marek Ingr, Kateřina Švehlová, Lucie Polovinkin, Martin Růžička, Lucie Bednárová, Romana Hadravová, Edita Poláchová, Petra Rampírová, Jana Březinová, Václav Kašička, Pavel Majer, and Kvido Strisovsky (2017), The Journal of Biological Chemistry. 292, 2703-2713. – 15%

3. General and modular strategy for designing potent, selective, and pharmacologically compliant inhibitors of rhomboid proteases

Anežka Tichá, Stancho Stanchev, Kutti R. Vinothkumar, David C. Mikles, Petr Pachl, Jakub Began, Jan Škerle, Kateřina Švehlová , Minh T.N. Nguyen, Steven H.L. Verhelst, Darren C. Johnson, Daniel A. Bachovchin, Martin Lepšík, Pavel Majer, and Kvido Strisovsky (2017), Cell Chemical Biology. 24, 1523–1536. – 15%

4. Membrane protein dimerization in cell-derived lipid membranes measured by FRET with MC simulations

Jan Škerle, Jana Humpolíčková, Nicholas Johnson, Petra Rampírová, Edita Poláchová, Monika Fliegl, Jan Dohnálek, Anna Suchánková, David Jakubec, and Kvido Strisovsky (2020), Biophysical Journal, accepted. – 50%

Prague, 27.03.2020

…………………………..

Kvido Stříšovský, Ph.D.

Table of contents

1   Abbreviations

ABI – abortive infection

AD – Alzheimer's disease

Aph1 – anterior pharynx defective 1

APP – amyloid-β (Aβ) precursor protein

AtRBL – *Arabidopsis thaliana* rhomboid-like proteases

BCAM – basal cell adhesion molecule

COP – coat protein complex

DCBLD2 – CUB (complement C1r/C1s, Uegf, Bmp1) and LCCL (Limulus clotting factor C,

Cochlin and Lgl1) domain-containing protein 2

DDR1 – discoidin domain-containing receptor 1

Der1 – degradation in the endoplasmic reticulum protein 1

EGF – epidermal growth factor

ER – endoplasmic reticulum

ERAD – ER-associated degradation

FCCS – fluorescence cross-correlation spectroscopy

FRET – Förster resonance energy transfer

GPMV – giant plasma membrane vesicle

IFAP – ichthyosis follicularis alopecia photophobia

IMPRs – intramembrane proteases

KIRREL1 – Kin of irregular chiasm-like protein 1

MBP – maltose binding protein

MC – Monte Carlo

PARL – presenilins-associated rhomboid-like protein

Pen2 – presenilin enhancer 2

PGAM5 – phosphoglycerate mutase 5

PINK1 – PTEN (phosphatase and tensin homolog) - induced kinase 1

PS1 – presenilin 1

PS2 – presenilin 2

Rce1 – Ras converting CAAX endopeptidase 1

RIP – regulated intramembrne proteolysis

S1P – site-1 protease

S2P – site-2 protease

SCAP – SREBP cleavage-activating protein

SPP – signal peptide peptidase

SPPLs – signal peptide peptidase like proteases

SREBP – sterol regulatory element-binding protein

Tat (pathway) – twin-arginine translocation pathway

TMH – transmembrane helix

2 Abstract

Intramembrane proteases from the rhomboid-like superfamily are enzymes widely distributed and conserved in all domains of life. They participate in many important processes such as membrane protein quality control or mitochondrial dynamics. Their activity is also linked with diseases like Parkinson's disease or cancer. This makes them potential therapeutic targets. In this work we tried to elucidate in more detail the mechanism of action of the main model intramembrane protease, GlpG from *E. coli*. We also focused on the mechanism of eukaryotic rhomboid RHBDL2, one of the four mammalian rhomboids, function of which is poorly understood. To acquire more detailed information about substrate-enzyme interaction, we synthesized a series of novel peptidyl-chloromethylketone inhibitors derived from natural rhomboid substrate TatA from *P. stuartii*. Crystal structure of the complex of GlpG with these inhibitors revealed four substrate binding subsites (S1 to S4) of the enzyme and explained its observed substrate specificity structurally. This study showed that substrate cleavage rate can be dramatically modified by changing the substrate sequence in positions P1 to P5. This helped us develop fluorogenic transmembrane peptide substrates for rhomboid proteases, which are usable in detergent and liposomes, and compatible with high-throughput screening. Using these substrates we showed that rhomboid proteases require almost the entire transmembrane domain of the substrate for efficient recognition and cleavage, and the enzyme probably interacts with the transmembrane domain of the substrate *via* a membrane-immersed exosite. Based on this knowledge we have designed novel and potent rhomboid inhibitors based on peptidyl-α-ketoamides. These compounds are active at nanomolar concentrations, and are selective for rhomboids. Crystal structures revealed that peptidyl-α-ketoamides bind the rhomboid covalently by mimicking the tetrahedral intermediate. Finally, by employing advanced fluorescence spectroscopy techniques (FRET and FCCS), we have investigated the behavior of the rhomboid protease RHBDL2 in a natural biomembrane. While it was previously thought that rhomboids are allosterically activated by dimerization, we found no evidence of RHBDL2 dimerization in natural membranes. Importantly, the approaches developed in this work are generally applicable to the assessment of dimerization of transmembrane proteins. In summary, the findings described in this thesis significantly contribute to the understanding of the mechanism of action of rhomboid proteases.

3 Abstrakt

Intramembránové proteasy z rodiny rhomboidů jsou v přírodě široce rozšířeny a vyskytují se ve všech doménách života. Podílejí se na mnoha důležitých procesech, jako například kontrola kvality membránových proteinů nebo mitochondriální dynamika. Jejich aktivita souvisí s nemocemi jako Parkinsonova nemoc nebo rakovina. Proto se rhomboidy jeví jako potenciální terapeutické cíle. V této práci jsme se snažili objasnit detailní mechanismus fungování modelové proteasy GlpG z *E. coli*. Zaměřili jsme se i na mechanismus rhomboidu RHBDL2, což je jeden ze čtyř eukaryotických rhomboidů, jejichž funkce není příliš prostudována. Pro pochopení vazby mezi substrátem a enzymem jsme připravili řadu nových peptidyl-chlormethylketonových inhibitorů odvozených od proteinu TatA, což je substrát rhomboidu v *P. stuartii*. Díky krystalové struktuře komplexů GlpG s těmito inhibitory jsme prozkoumali vazebná místa substrátu S1 až S4, což nám umožnilo objasnit strukturní podstatu substrátové specifity enzymu. Ukázali jsme, že rychlost štěpení substrátu může být významně ovlivněna modifikací sekvence substrátu na pozicích P1 až P5. Na základě těchto pozorování jsme vyvinuli fluorogenní transmembránový peptidový substrát rhomboidových proteas, který je použitelný jak v detergentu, tak v liposomech, a je vhodný pro testování s vysokou propustností. Pomocí těchto substrátů jsme dokázali, že rhomboidy pro efektivní zpracování substrátu vyžadují takřka kompletní transmembránovou část substrátu a že interakce mezi enzymem a substrátem nejspíš probíhá uvnitř membrány. Díky těmto znalostem se nám podařilo navrhnout silné inhibitory proteas z rodiny rhomboidů, jejichž základem jsou peptidyl-α-ketoamidy. Tyto inhibitory jsou aktivní v nanomolárních koncentracích a působí selektivně na proteasy z rodiny rhomboidů. Pomocí krystalových struktur jsme prokázali, že vazba peptidyl-α-ketoamidu na rhomboid je kovalentní, podobná tetraedrálnímu meziproduktu štěpení substrátu. Pomocí pokročilých metod fluorescenční spektroskopie (FRET a FCCS) jsme objasnili chování rhomboidové proteasy RHBDL2 v jejím přirozeném prostředí. Zatímco dosavadní výsledky naznačovaly možnost alosterické aktivace rhomboidů jejich dimerizací, my jsme nenašli žádné důkazy o dimerizaci RHBDL2 v bimembráně. Během této práce se nám podařilo vyvinout metodiku široce aplikovatelnou na studium dimerizace membránových proteinů. Všechny poznatky popsané v této práci významně přispívají k pochopení mechanismu fungování rhomboidových proteas.

4   Introduction

4.1   Proteases

Proteases have been studied by scientists for decades. They are involved in many complicated biological processes like apoptosis, wound healing, angiogenesis, cell migration and differentiation, tissue remodeling, neuronal outgrowth, hemostasis, morphogenesis and immunity [1], and their activity is also linked to many pathological conditions, neurodegenerative and cardiovascular diseases, cancer, arthritis and progeria [2]. One of the first discovered enzymes, the digestive protease pepsin (1836, Theodor Schwann), was also one of the first enzymes crystallized (1928, John H. Northrop). Moreover, the first X-ray diffraction pattern of a protein was acquired in 1934 using pepsin crystals [3]. Nowadays, many proteases are intensively studied as potential therapeutic targets [4] and biotechnological tools [5]. Based on the mechanism of their action, proteases are grouped into four main families: serine, cysteine, aspartyl and metalloproteases (Figure 1), with some specialized mechanistic variations, such as threonine peptidases (proteasomes). When cleaving a substrate, aspartyl proteases and metalloproteases use activated water molecule to attack the peptide bond, while cysteine and serine protease use water molecule to resolve a covalent intermediate formed between the enzyme and the substrate in the first step of catalysis.
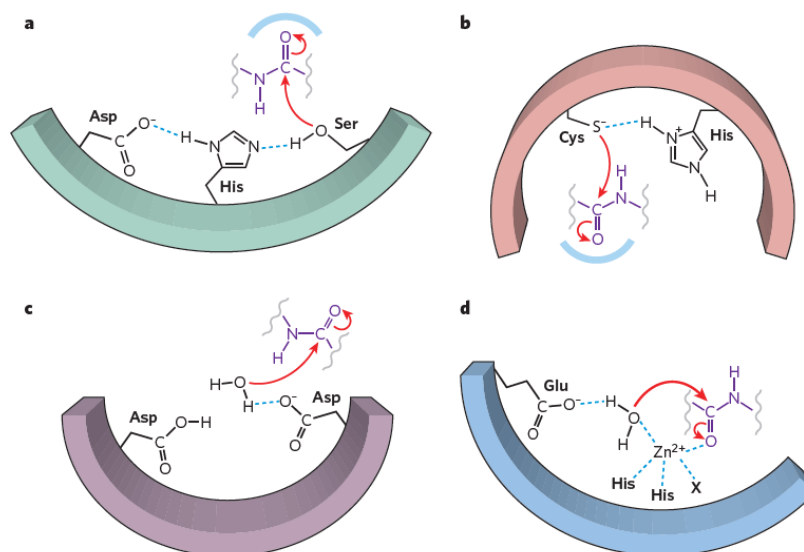
Figure 1: Schematic representation of the catalytic action of the four protease families. Serine proteases (a) have a catalytic triad (serine, histidine and aspartic acid). The hydroxyl group of serine acts as a nucleophile attacking the carbonyl of the peptide bond of the substrate, the nitrogen of histidine with a free electron pair can accept a proton from the hydroxyl group of serine, and the aspartic acid renders the nitrogen of histidine even more electronegative. A covalent intermediate between the enzyme and substrate occurs during the reaction. Cysteine proteases (b) also use nucleophile attack, in this case performed by the negatively charged sulfur of the catalytic cysteine, also including the enzyme – substrate covalent intermediate. Aspartyl proteases (c) have two aspartic acid residues coordinating a water molecule. In this case, the covalent intermediate does not occur and the proteolysis is performed in a single step. Metalloproteases (d) are the most variable protease family. In these enzymes, a coordinated metal, usually zinc, activates the catalytic water molecule. Adopted from [6].

4.2   Intramembrane proteolysis

Proteolysis within the membrane was first reported as an essential process in sterol homeostasis [7]. In the Golgi apparatus, the cytoplasmic domain of the mammalian sterol regulatory element-binding protein (SREBP) is cleaved off from the membrane domain by an intramembrane metalloprotease, later classified as a member of the site-2 protease (S2P) family, (reviewed in [8]). The liberated domain can then relocate to the nucleus and activate genes responsible for the synthesis of cholesterol and fatty acids [9]. This discovery established the principles of intramembrane proteolysis and since then, more intramembrane

proteases (IMPRs) have been identified and classified into three major groups based on their similarities: S2P zinc metalloproteases, rhomboid serine proteases, aspartyl IMPRs signal peptide peptidases (SPP) and presenilin. These major families are complemented by the recently discovered glutamate intramembrane proteases homologous to the Ras converting enzyme Rce1 [10]. In summary, intramembrane proteolysis (also called regulated intramembrane proteolysis, or RIP) is a process highly conserved from bacteria to mammals. RIP has an important role in many cellular processes like proliferation [11], differentiation [9], protein degradation [12], cell adhesion [13], lipid metabolism, transcriptional regulation and mitophagy [14].

4.2.1 Biological substrates and functions of IMPRs

More than 100 transmembrane protein substrates undergo RIP; these include growth factors and their receptors, cytokines and their receptors, cell adhesion proteins, viral proteins and signal peptides [15–17]. Among the most studied substrates of intramembrane proteases (IMPRs) are amyloid-β precursor [18], Notch [19,20], E-cadherin [13], tumor necrosis factor [21], interleukin-1 receptor type I and II [22–24], insulin-like growth factor 1 receptor [25], epidermal growth factor [26], receptor tyrosine kinase ErbB4 [27], p75 neurotrophin receptor [28–30], CD44 [31], triggering receptor expressed on myeloid cells-2 [32] and epithelial cell adhesion molecule [33]. Based on the wide diversity of substrates, RIP is proposed to be associated with many important physiological processes, such as haematopoiesis, embryonic development and regulation of nervous and immune systems, (reviewed in [34]).

IMPRs have very diverse sequences and even proteases from the same family and the same organism can have relatively low sequence similarity. Homologous IMPRs can also have diverse functions and different substrate(s) in different species, cells and organelles. On the other hand, it is known that some proteases can cleave synthetic model substrates or non-natural substrates from different organisms [35,36]. In principle, an IMPR substrate can be cleaved in three regions with respect to its membrane topology: close to one of the edges of the substrate TMH, which results in the release one of the products from the membrane, or

towards the substrate TMH center, which induces liberation of both cleavage products from the membrane. Each of these scenarios has different purpose and is related to the function of the substrate (see Figure 2) [37].
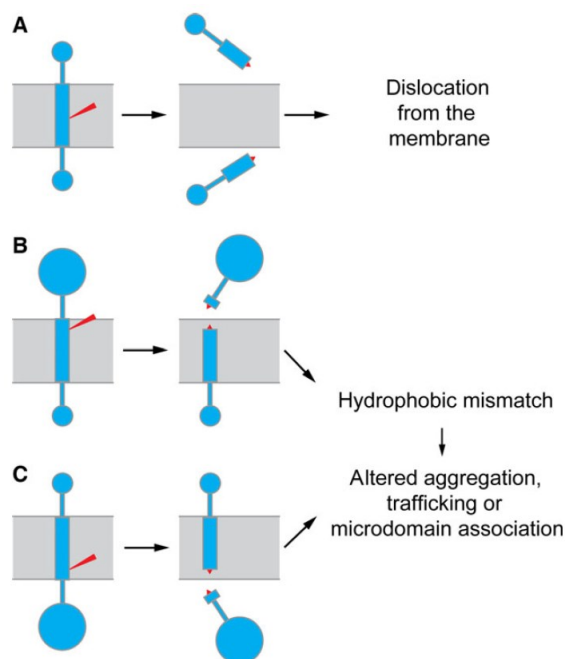
Figure 2: Three possible scenarios of transmembrane protein cleavage by an IMPR: (A) cleavage in the middle of the substrate's membrane region results in the dislocation of both cleavage product stubs from the hydrophobic environment of the lipid membrane; (B), (C) cleavage closer to the edge of the substrate's membrane region causes dislocation of only one of the products from the membrane. Adopted from [37].

Most IMPR substrates are transmembrane proteins with a single TMH, typically (but not always) containing a helix-destabilizing motif [9,38]. In some cases, insertion of a destabilizing amino acid residue such as Gly or Pro can even turn some non-substrates into substrates [39]. This indicates that IMPRs require substrates with helically unstable region(s), typically having one transmembrane segment, although IMPRs have been shown to cleave also polytopic membrane proteins [40,41]. Furthermore, some IMPRs do not require the destabilizing residues for substrate cleavage [42,43] and other IMPR substrates were shown to

have a tightly packed α-helix in the cleavage site [44,45]. All these facts together imply that RIP is a complex process and more research is required to understand it sufficiently well.

4.2.2 Site-2 proteases

Site-2 proteases (S2Ps) constitute a large family of intramembrane metalloproteases. They participate in the maturation of transcription factors, and usually require pre-cleavage of their substrate by another protease (also sometimes called S1P) [8,9,46]. As already mentioned, the first discovered IMPR was an S2P protease [7]. Its substrate, sterol regulatory element-binding protein (SREBP) – an intramembrane transcription factor, moves from the ER (in complex with SREBP cleavage-activating protein (SCAP)) to the Golgi apparatus via coat protein complex II (COP II) coated vesicles when the cell needs to synthetize more lipids. In the Golgi, two different proteases process this substrate in tandem. S1P cleaves SREBP in a luminal loop between its two TMHs; only the product of this cleavage can become a substrate of S2P [47] (see Figure 3). S2P homologs have also been identified in archaea [48] and prokaryotes [49], which shows that S2P proteases are evolutionarily very old. A mutation in human S2P is associated with the ichthyosis follicularis, alopecia and photophobia (IFAP) syndrome [50].
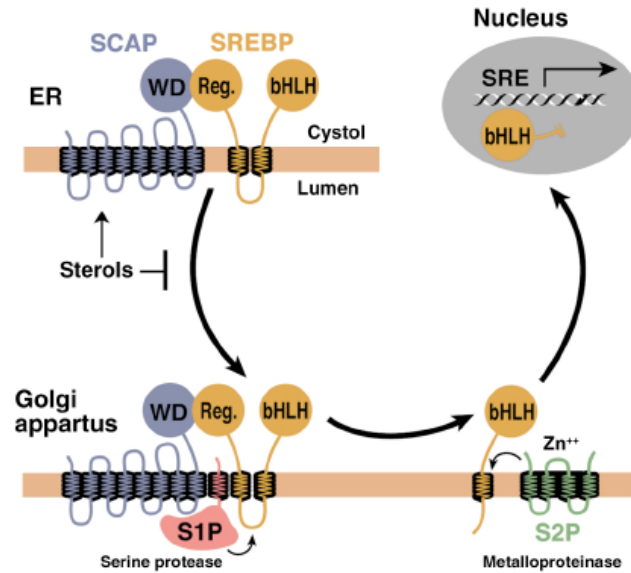
Figure 3: Schematic representation of SREBP cell fate. Adopted from [51].

As for other intramembrane proteases, it is difficult to fully understand the mechanism of action of S2Ps, because the substrate hydrolysis occurs in the membrane [48]. All the known S2P substrates are transcription factors with at least one type II TMH [52], which is the site of cleavage by S2P [50].

4.2.3 Aspartyl intramembrane proteases

Presenilins

Presenilins are key components of the γ-secretase complex whose activity is related to the pathogenesis of Alzheimer's disease (AD) [53–55], the most common neurodegenerative disease leading to dementia with no efficient treatment available so far. Gamma secretase is responsible for the cleavage of Amyloid-β (Aβ) precursor protein (APP) which contributes to the formation of toxic amyloid-β peptides which then aggregate into senile plaques [56]. There are two presenilin genes, PSEN1 and PSEN2, mutations of which are related to AD [57]. These genes encode polytopic membrane proteins presenilin 1 (PS1) and presenilin 2 (PS2),

respectively [58], which exhibit 67% sequence identity [59]. PS1 is a protein with 9 TMH, which contains two catalytic aspartate residues on TMH 6 and 7 (Figure 4:) and is localized mostly in ER and Golgi. Presenilins are synthesized in the form of inactive zymogens and have to be activated by endoproteolysis [60].
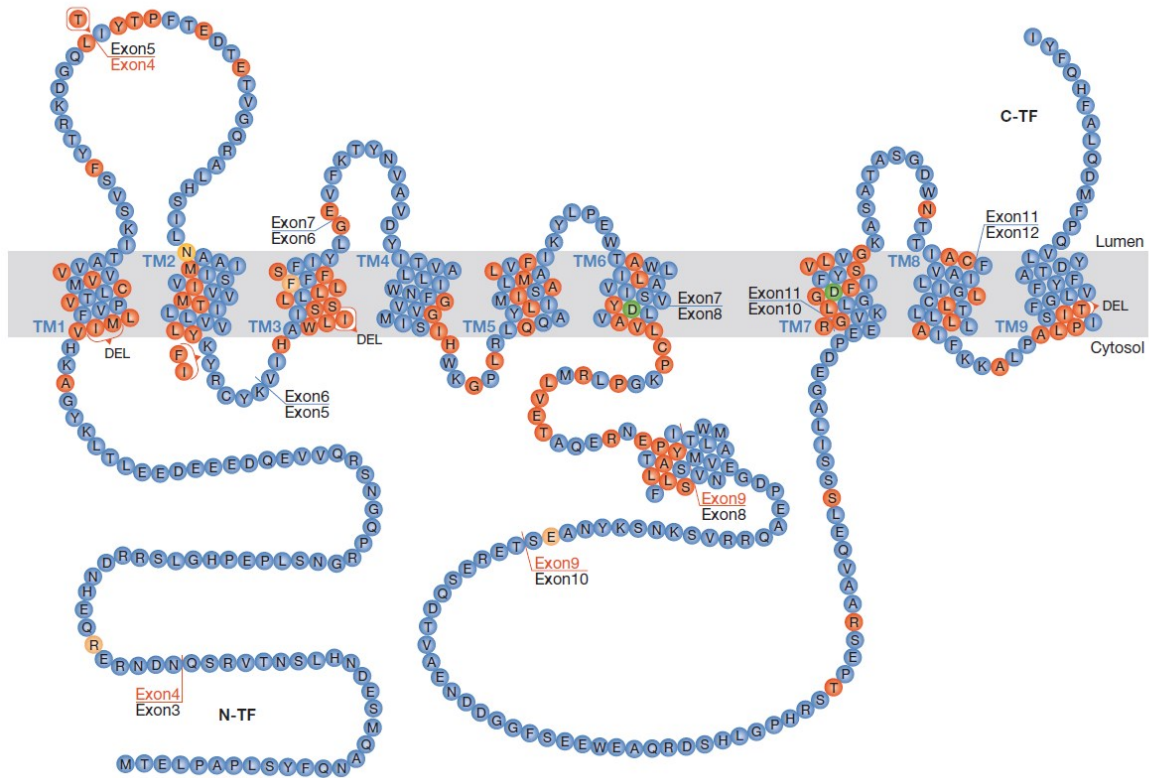
Figure 4: Schematic picture of PS1 protein topology with catalytic residues in green and mutations linked to AD in red. Adopted from [61].

Apart from presenilin, the γ-secretase complex contains three other components (Figure 5). Nicastrin helps with substrate recruitment [62], anterior pharynx defective 1 (Aph1) is an essential cofactor [63] and presenilin enhancer 2 (Pen2) subunit is important for proper complex maturation [64]. The stoichiometry of the subunits in the γ-secretase complex is 1:1:1:1 [65].
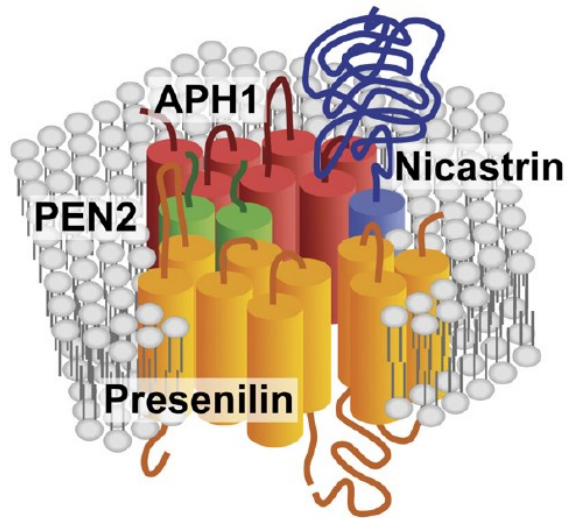
Figure 5: The γ-secretase complex with its subunits. Adopted from [55].

The γ-secretase complex has many substrates of various physiological functions, structure and localization. These are usually type I membrane proteins [66] responsible for signaling and regulation of cellular processes such as adhesion or migration. The complex also usually prefers membrane protein substrates after the shedding of ectodomains, rather than full-length proteins [67]. A total of 80 γ-secretase substrates have been identified so far, but the two most important and most studied substrates are the already mentioned APP and also Notch, a protein important for intercellular communication, gene regulation and cell differentiation [68].

Signal peptide peptidase like proteases (SPPLs)

Apart from eukaryotic organisms, presenilin homologs have been found also in organisms that do not possess other components of γ-secretase, including archaea [69], indicating that they represent members of a larger family of aspartyl IMPRs. In fact, other members of this family of IMPRs were discovered in 2002 using bioinformatics and

biochemical methods by three independent groups. They found that SPP and its homologs form a cluster of aspartyl IMPRs closely related to presenilins and that they share a conserved motif YD and LGLGD within their catalytic center (Figure 6). The family of SPP and SPP-like proteases (SPPLs) is conserved among eukaryotes, including protozoa, fungi, plants and animals [69–71]. Although these proteins share a similar architecture, they have evolved different physiological functions in different species [72]. SPPLs are localized in ER [73], Golgi [74] and lysosomes [75], but they can also occur on plasma membrane [76].
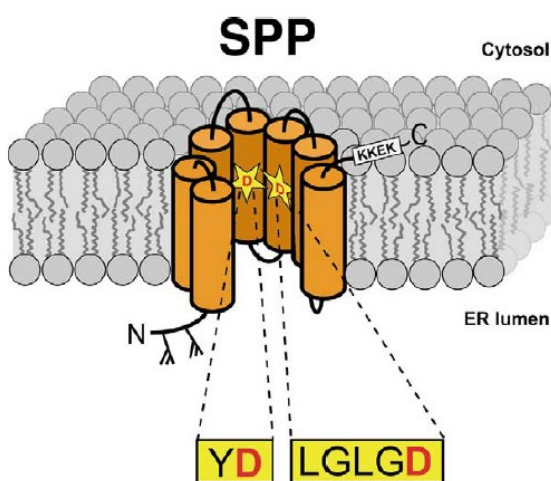
Figure 6: A model of SPP with the two catalytic aspartates in the YD and LGLGD motifs. Adopted from [77].

The name of SPPLs is derived from the ability of the founding member SPP to cleave signal peptides. However, SPPL substrates are much more diverse. For example, they play role in the immune system [78], participate in protein post-targeting [79], and their activity is important for maintaining membrane homeostasis [80]. Their substrates usually have to be pre-cleaved by other proteases (Figure 7) [72] and, despite their similarity to presenilins, they have an inverted topology and prefer substrates with the type II membrane orientation [81].
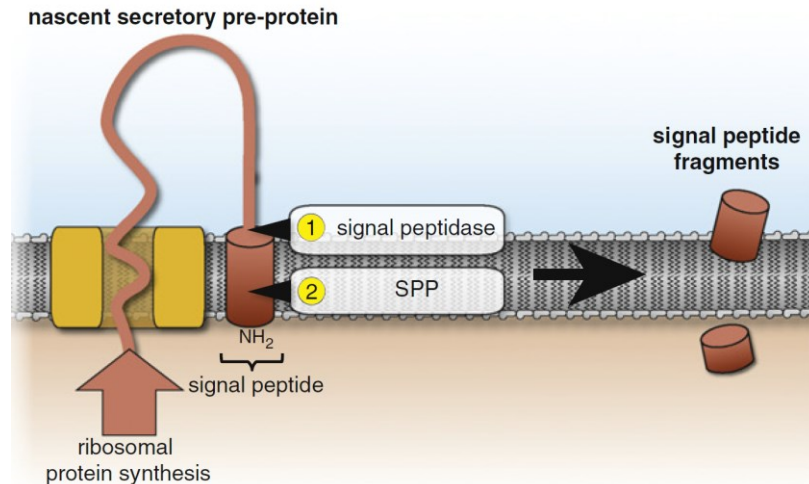
Figure 7: Schematic representation of SPP action and substrate processing. Adopted from [72].

4.2.4 Rce1

The class of glutamate IMPs is represented by a recently identified intramembrane protease Rce1 located in the ER. The overall structure of Rce1 with eight TMHs and also the catalytic site of this enzyme are distinct from the other IMPs, which makes Rce1 a founding member of a new IMP family [10]. Rce1 is also a member of the ABI (abortive infection) family of putative IMPRs. These proteins are involved in membrane protein anchoring in eukaryotes and their homologues in prokaryotes are probably involved in bacteriocin self-immunity [82]. It has been shown that Rce1 inactivation in mice leads to mislocalization of Ras proteins from the plasma membrane [83], development of lethal dilated cardiomyopathy [84], and it also interferes with the survival of photoreceptor cells in mice [85].

4.3 Rhomboids

The name of the rhomboid family of intramembrane proteases originates from the first discovered member – rhomboid-1 from *Drosophila melanogaster*. Mutations of the rhomboid-1 gene interfere with grow factor signaling [86] and result in a characteristically changed,

pointed head skeleton of the fly [87]. Since then rhomboids and their homologs have been identified in all domains of life and they are the most abundant and widespread family of IMPRs [88]. Interestingly, several psuedoproteases (i.e. proteolytically inactive proteins) show topological similarity to rhomboid proteases, such as iRhoms [89] and Derlins [90]. The usage of the term 'rhomboid-like' protein superfamily (Figure 8 and Figure 16) is then probably more accurate than using just the simple 'rhomboid' family [91].
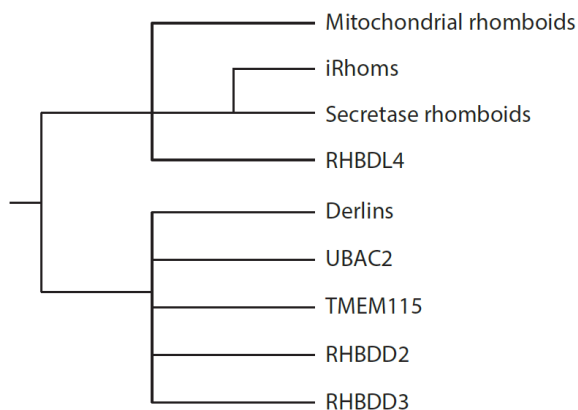
Figure 8: Classification of the rhomboid-like superfamily. The rhomboid-like cluster of proteins has been classified based on functional and sequence relationships [91], but this scheme does not represent evolutionary relations. Adopted from [91].

4.3.1 Rhomboid mechanism of action

It was initially very controversial whether proteolysis could be possible in a hydrophobic lipid environment. The research into rhomboid protease mechanism of action has revealed many details of how this may be possible. Rhomboid proteases have a catalytic dyad instead of the classical triad and the catalytic serine residue is hidden in a water-accessible pocket created by TMH's of the protease, which protects it from the lipid environment. This progress was possible mostly thanks to the reconstitution of rhomboid activity *in vitro* [90,92], and to the solving of high resolution X-ray crystallographic structures (Figure 9) of the model *E. coli* rhomboid protease GlpG (Figure 10) [93–95], which was the first structure of an intramembrane protease.
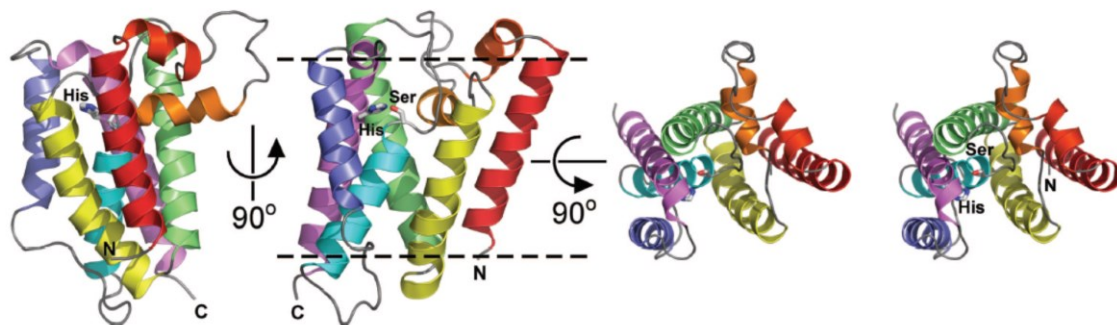
Figure 9: A ribbon diagram of the *E. coli* rhomboid protease GlpG crystal structure in three orientations. Adopted from [93].
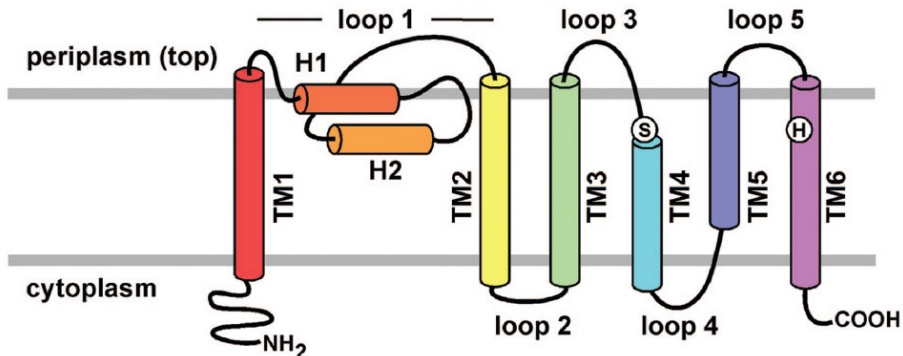
Figure 10: Schematic representation of a model rhomboid protease from *E. coli*: GlpG. The color code corresponds to Figure 9. Adopted from [93].

X-ray crystallography also suggested how the substrate may enter the active site of rhomboid. Crystallographic analysis of *E. coli* GlpG revealed alternative 'open' conformations of TMH5, which led to the proposal that a lateral 'substrate gate' is formed by TMH 2 and 5 of the protease [89]. This was confirmed by enzymatic [96] and biophysical [97] analyses. Similar structural disorder was detected also in the homologous *H. influenzae* GlpG [98]. The details of this mechanism are debated, such as whether lateral movement of TMH5 is really required for substrate cleavage by GlpG [99], or whether TMH2 and 5 just form an intramembrane binding site (an 'exosite') recognizing the TMH of the substrate. This substrate entry route is nevertheless strongly supported by the most recent observations [100].

Other efforts focused on the analysis of rhomboid protease substrates and their features important for the recognition by rhomboids. The influence of the biophysical properties of substrate TMH (Figure 11A) on rhomboid activity was studied in detail [101], and the results confirmed the importance of helix-destabilizing residues (see also section 4.2.1). Other research revealed that the cleavage site and rate are determined by a 'recognition motif' region between positions P4 and P2' (Figure 11B) [102]. A unifying picture emerges that rhomboid substrates are defined by two elements, a transmembrane region interacting with the rhomboid inside the membrane, and a recognition motif that interacts with the water-accessible active site of rhomboid. These two elements may require flexibility between them, which may be conferred by the helix-destabilizing residues. Because of the sequence diversity in the rhomboid-like superfamily, it is highly unlikely that a more exact common substrate-determining rule will be found. Rhomboid substrate specificity is probably driven by both mentioned elements with different contributions depending on the particular substrate [91]. According to some authors, rhomboid mechanism of action is also affected by their specific three-dimensional shape (Figure 12), which can bias surrounding lipids (and thus also protein diffusion) in an unusual way in comparison with a 'standard' membrane protein [103].
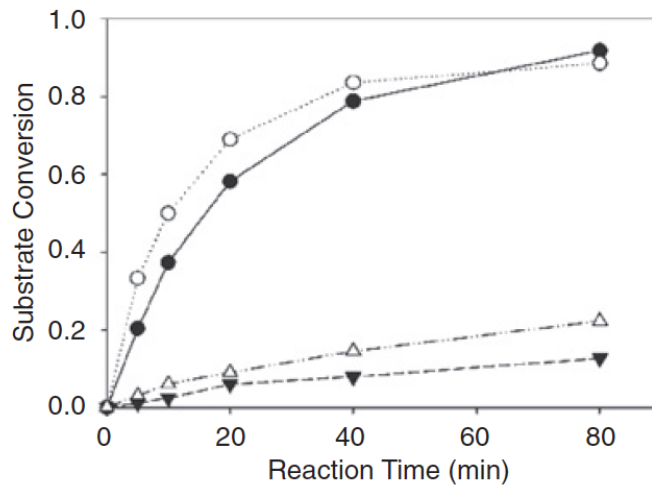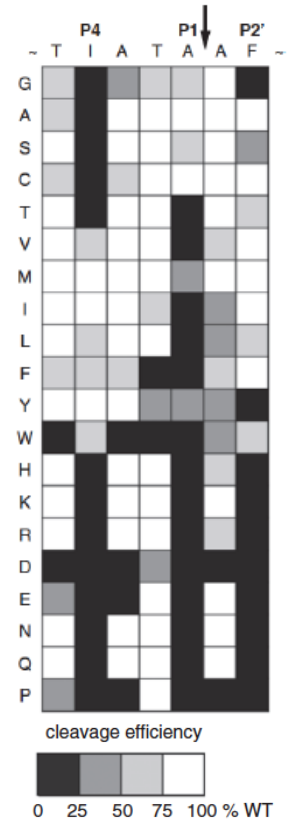
Figure 11: Schematic representation of rhomboid protease substrate specificity. (A) Sequence alignment of rhomboid substrate TatA TMD truncated variants and efficiency of their cleavage by bacterial rhomboid protease AarA. The enzyme requires a certain TMD length for efficient action. (B) Positional scanning mutagenesis of P5 to P2' sites of TatA, the natural substrate of bacterial rhomboid protease AarA. The mutations are ranked into four grades based on the severity of their effect on cleavage efficiency (depicted in shades of grey). Residues in positions P1, P4 and P2' in the substrate TatA had the biggest impact on cleavage efficiency. The cleavage site is marked by an arrow. Adopted from [102].
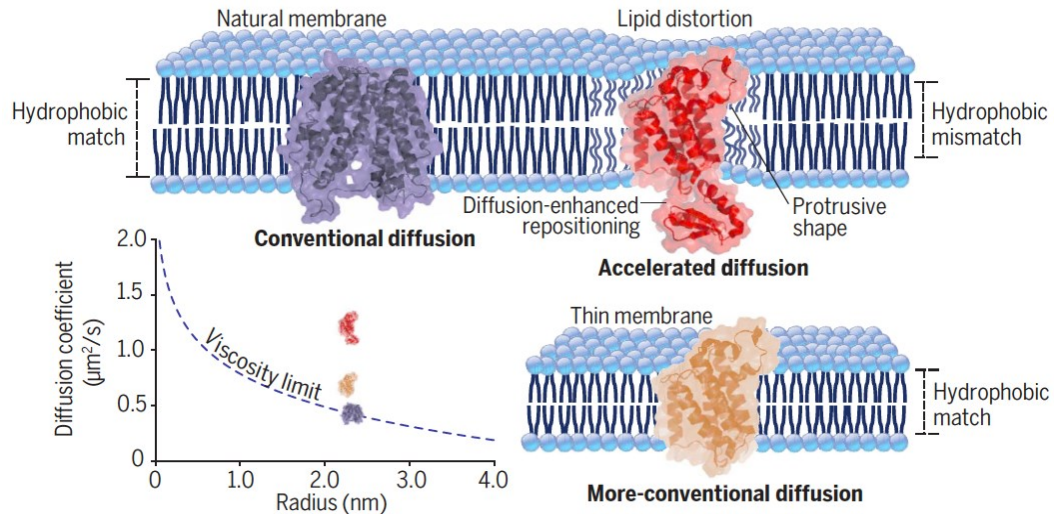
Figure 12: Lipid reorientation around the full-length rhomboid protease GlpG (red) induced by the rhomboid shape, and in the case of GlpG also affected by its N-terminal domain (orange: GlpG without the N-terminal domain). A standard membrane protein (purple) does not affect the arrangement of surrounding lipids. The diffusion coefficient of proteins is dependent on their shape (graph). Adopted from [103].

4.3.2 Biological role of selected rhomboid proteases in different species

Rhomboid proteases are present in all domains of life, and during evolution they presumably adopted widely different roles in different biological processes, from regulating protein secretion in bacteria to mitochondrial dynamics in eukaryotes [104–107]. Their activity has also been associated with pathological processes including cancer [108] and Parkinson's disease [109].

Despite the wide distribution of rhomboids in bacteria, with some bacteria containing more than one rhomboid, very little is known about their biological function in prokaryotic cells [110]. The best studied bacterial rhomboid from the functional point of view is the rhomboid AarA in *Providencia stuartii*. AarA is responsible for processing the TatA protein, thereby activating the twin-arginine translocation (TAT) pathway. TatA is activated by the removal of its N-terminal extension and this also mediates bacterial quorum sensing [111] (Figure 13).
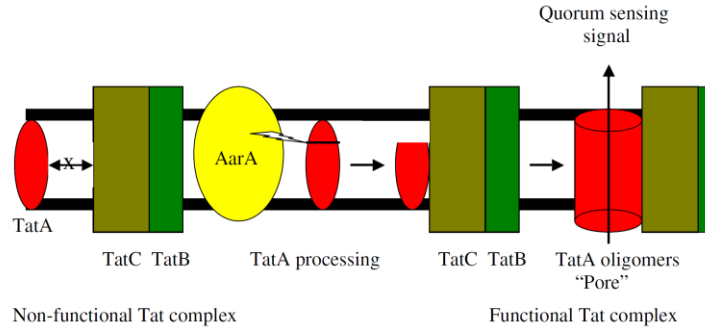
Figure 13: Model of TAT system activation in *P. stuartii* by the AarA rhomboid protease. The TatA protein can interact with the other subunits of the Tat system (TatB and TatC) only after its N-terminal extension is removed by the AarA rhomboid. Active Tat system is responsible for quorum sensing signal transmission by an unknown mechanism. Adopted from [110].

The *E. coli* rhomboid GlpG was identified in 2002 [35]. The gene coding for this enzyme is a member of the *glpEGR* operon, the function of which is associated with glycerol metabolism via *glpR* [112]. GlpG is the main model rhomboid for structural (Figure 10) and functional studies [97], but its natural substrates and its role in *E. coli* metabolism remain unknown [36,113,114]. GlpG can utilize a variety of model rhomboid substrates [35] ranging from bacterial to eukaryotic proteins, typically embedded in a hybrid scaffold composed of the substrate transmembrane region (e.g. of LacY, TatA, Gurken or Spitz) fused to a periplasmic maltose-binding protein domain and a cytoplasmic thioredoxin domain [102]. It was proposed that GlpG and two other bacterial rhomboid proteases are allosterically activated by dimerization, and allosteric regulation by dimerization has been suggested to potentially be a general property of rhomboid proteases [115]. However, this theory is in contrast with our own observation (see section 6.4), which is also supported by recently published independent research [116].

Plants contain a larger number of rhomboids than animals [117], but their biological roles are largely unknown. The exception are AtRBL8 and AtRBL9, which are probably responsible for flower development in *Arabidopsis thaliana* [118].

The roles of rhomboid proteases in eukaryotic parasites are understood relatively well. They participate in adhesion and invasion of apicomplexan parasites such as *Toxoplasma gondii* and *Plasmodium falciparum* [39], and for this reason they are investigated as potential therapeutic targets [119]. Their roles in other parasites such as *Cryptosporidium* are being studied [120].

Mitochondrial rhomboids constitute a special sort of eukaryotic rhomboids [121]. They are located in the inner mitochondrial membrane and participate in membrane remodeling in *Saccharomyces cerevisiae* [122] and membrane dynamics in *Drosophila melanogaster* [123]. Mouse mitochondrial rhomboid PARL and one of its substrates, PINK1 (Figure 14), are studied for their possible implication in Parkinson's disease [109]. Other PARL substrate, PGAM5, has been linked to apoptosis and sensing of mitochondrial damage [124]. PARL activity is influenced by mitochondrial membrane potential, possibly via translocation and topological effects on its substrates (Figure 14) [124].
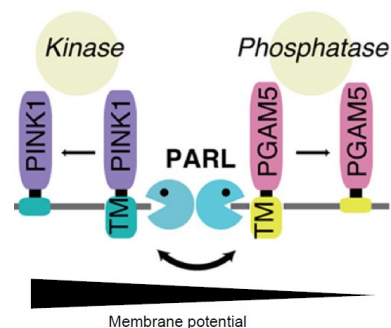
Figure 14: Proposed model of mitochondrial membrane dependent regulation of PARL-mediated cleavage of PINK1 and PGAM5. Adopted from [124].

Beside rhomboids in mitochondria, there are four other proteases of the rhomboid family in mammalian cells: RHBDL1 – 4 [125], all localized within the secretory pathway [40]. Their functions are poorly understood. RHBDL4 is localized to the endoplasmic reticulum (ER), the compartment where membrane and secretory proteins fold. RHBDL4 has only 6 TMHs, unlike the other RHBDLs, which have 7 TMHs (Figure 15 and Figure 16)

[125]. RHBDL4 is upregulated by ER stress, suggesting that it could act as a quality control protease [40]. Indeed, it has been shown to interact with p97/VCP, degrade some aberrant membrane proteins with unstable helices, and to participate in the process of ER-associated degradation (ERAD) [40]. A huge step forward was recently made in the identification of natural substrates of RHBDL2. Proteomics analysis showed that RHBDL2 cleaves several novel substrates including the interleukin-6 receptor, cell surface protease inhibitor Spint-1, the collagen receptor tyrosine kinase epithelial discoidin domain-containing receptor 1 (DDR1), N-cadherin, discoidin, CUB and LCCL domain-containing protein 2 (DCBLD2), Kin of IRRE-like protein 1 (KIRREL1), basal cell adhesion molecule (BCAM) and others [126]. The identified substrate repertoire and epithelial expression implicates RHBDL2 in epithelial homeostasis [126].
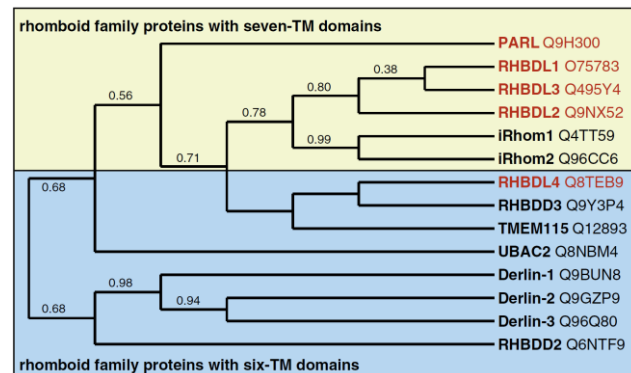
Figure 15: A phylogenetic tree of human rhomboid family proteins in the secretory pathway. Active rhomboid proteases and inactive pseudoproteases are highlighted in red and black, respectively. The comparison of similarity is based on the conserved regions (loop 1 and membrane domains 2, 4 and 6). Swissprot accession numbers are indicated. Adopted from [105].
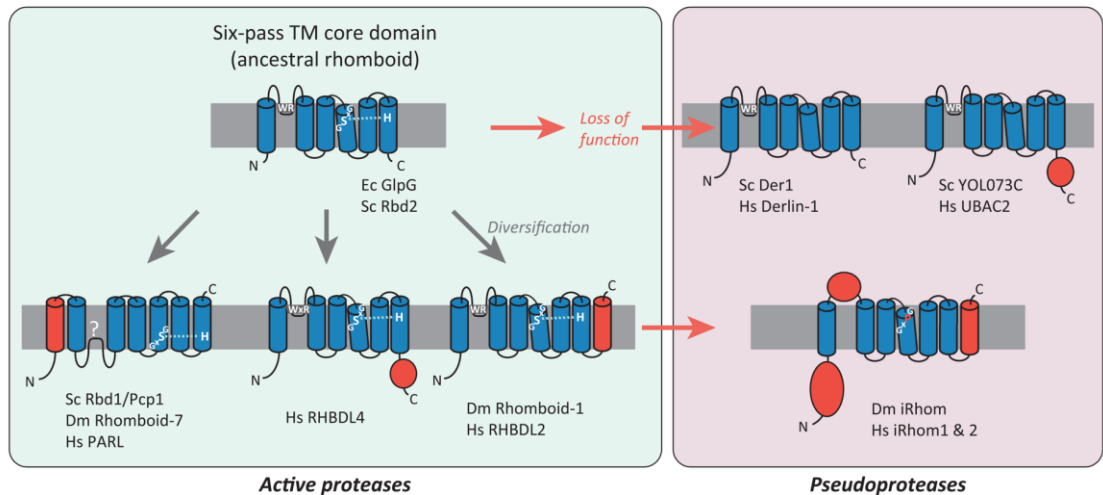
Figure 16: Schematic representation of the evolutionary relationship within the 'rhomboid-like superfamily'. *E. coli* (Ec) rhomboid GlpG is compared with members of the rhomboid-like superfamily from different species: *Homo sapiens* (Hs), *Drosophila melanogaster* (Dm) and *Saccharomyces cerevisiae* (Sc). Conserved six-pass TMH core is depicted in blue, with possible extra domain in red. The typical motif of the rhomboid active site (GxSG and H) and the L1 loop extension in the membrane (WR) are indicated and the GPxG motif typical for iRhoms is also shown. Presumable evolutionary relationships between members of the 'rhomboid-like superfamily' are represented by arrows. Adopted from [127].

The most important representatives of the proteolytically inactive rhomboid-like proteins are iRhoms and Derlins. iRhoms are phylogenetically closely related to active rhomboids (Figure 15). In comparison with active rhomboids, they have a slightly altered architecture with a large insertion into the L1 loop (Figure 16) [127] and they lack catalytic activity [89]. Their function is related to ER protein quality control [89] and ER to Golgi protein transport (Figure 17) [128,129]. It is however not clear what determines whether iRhom clients are exported or degraded [127]. Mutations in an iRhom coding gene have been shown to be linked with cancer [130].
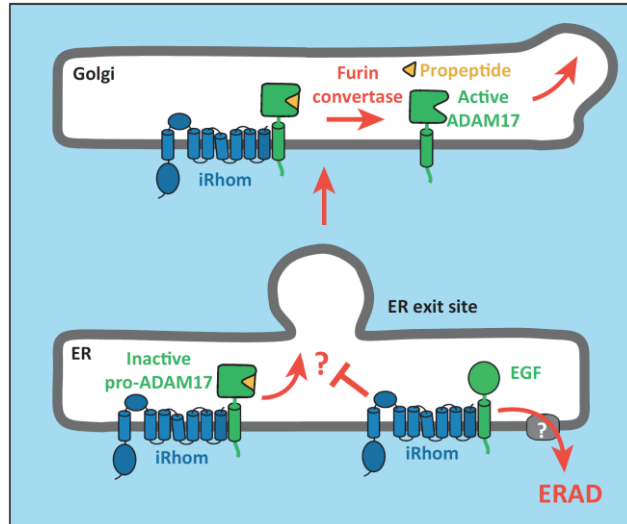
Figure 17: Schematic model of two functions of iRhoms in ER: ER to Golgi transport (top) and ERAD (bottom). Adopted from [127].

Derlins are in a more distant relationship with mammalian rhomboids than iRhoms (Figure 16) [131]. Their name originates from the first discovered derlin Der1 (degradation in the endoplasmic reticulum protein 1), which was discovered in *S. cerevisiae* [132]. This protein interacts with several ERAD-associated proteins [133], but the mechanistic details of its role in the ERAD process remain poorly understood. Its human ortholog, Derlin1, and its paralogs Derlin2 and 3 also play an important role in ERAD of many different proteins [134,135].

5   Aims of the study

This work is a complex study of mechanism and specificity of intramembrane proteases from the rhomboid-like superfamily. It is focused on the main rhomboid model GlpG from the bacterium *E. coli* and on one of the four human secretase rhomboids, RHBDL2. GlpG was studied mainly *in vitro* to obtain detailed information about its substrate specificity, binding properties and kinetics parameters of the cleavage reaction, while RHBDL2 was studied *in vivo* with emphasis on its behavior in a natural biomembrane.

Specific research aims:

- Investigation of substrate binding, specificity and reaction mechanism of GlpG.

- Study of dimerization properties of RHBDL2 in natural lipid membrane.

6   Publications

1. Substrate binding and specificity of rhomboid intramembrane protease revealed by substrate–peptide complex structures

Sebastian Zoll, Stancho Stanchev, Jakub Began, Jan Škerle, Martin Lepšík, Lucie Peclinovská, Pavel Majer, and Kvido Strisovsky (2014), The EMBO Journal. 33, 2408-2421.

$IF_{2014} = 10.434$

2. Sensitive versatile fluorogenic transmembrane peptide substrates for rhomboid intramembrane proteases

Anežka Tichá, Stancho Stanchev, Jan Škerle, Jakub Began, Marek Ingr, Kateřina Švehlová, Lucie Polovinkin, Martin Růžička, Lucie Bednárová, Romana Hadravová, Edita Poláchová, Petra Rampírová, Jana Březinová, Václav Kašička, Pavel Majer, and Kvido Strisovsky (2017), The Journal of Biological Chemistry. 292, 2703-2713.

$IF_{2016} = 4.125$

3. General and modular strategy for designing potent, selective, and pharmacologically compliant inhibitors of rhomboid proteases

Anežka Tichá, Stancho Stanchev, Kutti R. Vinothkumar, David C. Mikles, Petr Pachl, Jakub Began, Jan Škerle, Kateřina Švehlová, Minh T.N. Nguyen, Steven H.L. Verhelst, Darren C. Johnson, Daniel A. Bachovchin, Martin Lepšík, Pavel Majer, and Kvido Strisovsky (2017), Cell Chemical Biology. 24, 1523–1536.

$IF_{2016} = 6.743$

4. Membrane protein dimerization in cell-derived lipid membranes measured by FRET with MC simulations

Jan Škerle, Jana Humpolíčková, Nicholas Johnson, Petra Rampírová, Edita Poláchová, Monika Fliegl, Jan Dohnálek, Anna Suchánková, David Jakubec, and Kvido Strisovsky (2020), Biophysical Journal, accepted

6.1 Substrate binding and specificity of rhomboid intramembrane protease revealed by substrate–peptide complex structures

Background

Rhomboid protease GlpG from the bacterium *E. coli* is intensively studied as the main model rhomboid for structural and mechanistic studies. The structure of this enzyme was well known, but our understanding of rhomboid mechanism was limited by the lack of information about the enzyme-substrate complex. To elucidate this, we determined X-ray structures of GlpG in co-crystals with peptidyl-chloromethylketone inhibitors derived from the natural rhomboid substrate TatA from *P. stuartii*.

Summary

Using biochemical analyses, we confirmed that binding of the peptidyl-chloromethylketone inhibitor to GlpG is similar to substrate binding. With the help of X-ray crystallography, we identified the S1 to S4 subsites of the protease. We found that S1 subsite co-creates a cavity with the previously proposed water retention site. Surprisingly, the L1 loop, which is a typical feature of rhomboids, helps create the S4 subsite. Its function may be similar also in other members of the rhomboid family. Finally, we created a molecular dynamics-based model of the Michaelis complex of GlpG with the substrate bound in the active site.

My contribution

I investigated GlpG specificity *in vitro* by conducting a complete positional scanning mutagenesis of the P5 to P1 region of the TatA substrate (Figure 18), analyzed the data and participated in writing of the relevant parts of the manuscript.
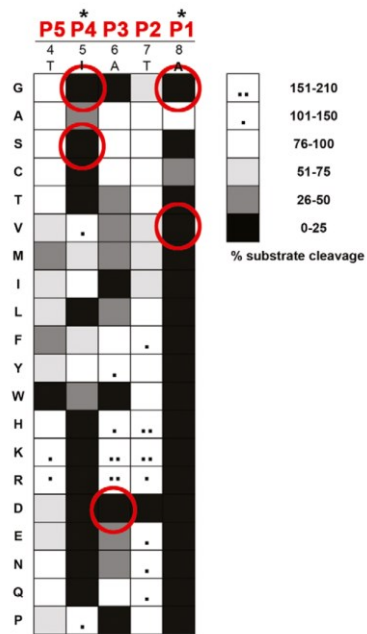
Figure 18: Specificity matrix of GlpG preferences on TatA variants *in vitro*.

Substrate binding and specificity of rhomboid intramembrane protease revealed by substrate–peptide complex structures. Sebastian Zoll, Stancho Stanchev, Jakub Began, Jan Škerle, Martin Lepšík, Lucie Peclinovská, Pavel Majer, and Kvido Strisovsky (2014), The EMBO Journal. 33, 2408-2421.

Substrate binding and specificity of rhomboid intramembrane protease revealed by substrate–peptide complex structures

Sebastian Zoll[1,†], Stancho Stanchev[1], Jakub Began[1,2], Jan Škerle[1,3], Martin Lepšík[1], Lucie Peclinovská[1,3], Pavel Majer[1] & Kvido Strisovsky[1,*]

Abstract

The mechanisms of intramembrane proteases are incompletely understood due to the lack of structural data on substrate complexes. To gain insight into substrate binding by rhomboid proteases, we have synthesised a series of novel peptidyl-chloromethylketone (CMK) inhibitors and analysed their interactions with *Escherichia coli* rhomboid GlpG enzymologically and structurally. We show that peptidyl-CMKs derived from the natural rhomboid substrate TatA from bacterium *Providencia stuartii* bind GlpG in a substrate-like manner, and their co-crystal structures with GlpG reveal the S1 to S4 subsites of the protease. The S1 subsite is prominent and merges into the 'water retention site', suggesting intimate interplay between substrate binding, specificity and catalysis. Unexpectedly, the S4 subsite is plastically formed by residues of the L1 loop, an important but hitherto enigmatic feature of the rhomboid fold. We propose that the homologous region of members of the wider rhomboid-like protein superfamily may have similar substrate or client-protein binding function. Finally, using molecular dynamics, we generate a model of the Michaelis complex of the substrate bound in the active site of GlpG.

Introduction

Cleavage of transmembrane domains (TMDs) by intramembrane proteases has emerged as an important and evolutionarily widespread signalling and quality control mechanism with medical significance (Brown *et al*, 2000; Lemberg, 2011), but a full understanding of the biological roles and design of pharmacological interventions against intramembrane proteases requires a greater knowledge of their mechanism and structure. Intramembrane proteases are very different from the classical water soluble proteases, since they evolved independently and operate in a distinct biophysical environment—at the interface of lipid membrane and aqueous solvent (Strisovsky, 2013). Although the crystal structures of prokaryotic homologues of all four catalytic types of intramembrane proteases have been solved (Wang *et al*, 2006; Feng *et al*, 2007; Li *et al*, 2013; Manolaridis *et al*, 2013), mechanistic understanding is limited by the lack of structures of enzyme–substrate complexes.

Rhomboids are serine proteases—probably the best characterised intramembrane proteases as regards structure and mechanism. Rhomboid proteases are widely conserved and regulate many biological processes including intercellular signalling, mitochondrial dynamics, invasion of eukaryotic parasites and membrane protein quality control (Lemberg, 2013). In addition, the recently discovered rhomboid-like proteins that share a similar scaffold, but are devoid of enzymatic activity, have emerged as important regulators of membrane protein quality control (Greenblatt *et al*, 2011; Zettl *et al*, 2011) and trafficking (Adrain *et al*, 2012). Non-catalytic rhomboid-like proteins regulate growth factor signalling (Zettl *et al*, 2011), inflammatory signalling via tumour necrosis factor in macrophages (Adrain *et al*, 2012) and NK-cell signalling (Liu *et al*, 2013), which illustrates their wide medical importance. In contrast to the advances in the biology of the non-protease rhomboid-family proteins, their mechanistic understanding lags behind. The only current source of structural information about rhomboid-family proteins are the bacterial rhomboid proteases.

The structures of bacterial rhomboid proteases published over the last 8 years have provided the first glimpses into the molecular architecture of an intramembrane protease. However, the mechanism of action and the structural basis of substrate specificity of rhomboids remain unresolved, largely due to the absence of structural analyses of rhomboid–substrate complexes. The recently

---

1 Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Prague, Czech Republic
2 Department of Microbiology, Faculty of Science, Charles University, Prague, Czech Republic
3 Department of Biochemistry, Faculty of Science, Charles University, Prague, Czech Republic
  *Corresponding author. Tel: +420 220 183 468; E-mail: kvido.strisovsky@uochb.cas.cz
  †Present address: Department of Biochemistry, University of Oxford, Oxford, UK

published structures of GlpG bound to various small, mechanism-based inhibitors (Vinothkumar *et al*, 2010, 2013; Xue & Ha, 2012; Vosyka *et al*, 2013) have served as models for speculations on substrate binding, but their utility in this respect is limited since the inhibitors are relatively small and structurally very different from peptide or protein substrates.

Here, we report crystal structures of a rhomboid intramembrane protease in complex with substrate-derived peptides, providing the first direct structural view of rhomboid specificity and catalytic mechanism. We show that tetrapeptidyl-chloromethylketone inhibitors bind the *Escherichia coli* rhomboid protease GlpG in a way that mimics the substrate, which allows us to map the specificity determining pockets of GlpG with confidence. Unexpectedly, the S4 subsite (which binds to the P4 residue of the substrate) is formed by the residues from the L1 loop, a conspicuous but enigmatic structural feature of rhomboid proteases (Wang *et al*, 2007; Bondar *et al*, 2009; Baker & Urban, 2012). Using site-directed mutagenesis, quantitative enzymatic assays and structural analyses, we demonstrate the plasticity of the S4 subsite. Furthermore, our work has implications for the recently discovered proteolytically inactive members of the rhomboid-like family (such as iRhoms or Derlins). It suggests that their domains topologically corresponding to the L1 loop of rhomboids may have client-binding roles. Finally, using molecular modelling and dynamics, we generate an extended model of our complex structure comprising the P4 to P3′ fragment of a bound substrate, allowing us to speculate about the mode of interaction of substrate's transmembrane domain with rhomboid.

Results

The inhibitory properties of peptidyl-chloromethylketones

One of the problems complicating structural analyses of rhomboid–substrate complexes is the relatively low affinity of rhomboids for their substrates (Dickey *et al*, 2013). To overcome this hurdle and gain insight into rhomboid substrate binding, we developed mechanism-based irreversible inhibitors modified with a peptide derived from a natural rhomboid substrate. The currently used rhomboid inhibitors, isocoumarins, phosphonofluoridates and monocyclic β-lactams (Vinothkumar *et al*, 2010, 2013; Pierrat *et al*, 2011; Xue & Ha, 2012; Xue *et al*, 2012), were unsuitable as warheads because the stereochemical similarity of peptidyl conjugates of isocoumarins and β-lactams to the acyl enzyme intermediate would be limited, and phosphonofluoridates have proven difficult to synthesise in the desired sequence diversity. We therefore turned our attention to peptidyl-chloromethylketones (CMKs) (Fig 1A), whose complexes with serine proteases resemble the tetrahedral transition state intermediate (Mac Sweeney *et al*, 2000; Malthouse, 2007) and which are readily synthesisable. The commercially available CMKs TLCK (N-α-tosyl-L-lysine chloromethylketone) and TPCK (N-α-tosyl-L-phenylalanine chloromethylketone) had shown only weak inhibition of YqgP and *Drosophila* rhomboid 1 (Urban *et al*, 2001; Urban & Wolfe, 2005), but we reasoned that this could have been due to their unsuitable P1 residues (Lys or Phe), since P1 residues with large side chains are not tolerated in substrates by several rhomboids including GlpG (Strisovsky *et al*, 2009; Vinothkumar *et al*, 2010).

We have first examined the inhibitory properties of tetrapeptidyl-CMK Ac-IleAlaAlaAla-COCH$_2$Cl (abbreviated as Ac-IAAA-cmk henceforth) based on the well-characterised bacterial rhomboid substrate TatA (Stevenson *et al*, 2007; Strisovsky *et al*, 2009). Like all other peptidyl-CMKs used in this study, this compound was stable in aqueous solution for more than 4 h (Supplementary Fig S1) and was soluble in rhomboid assay buffer up to 1 mM concentration (data not shown), allowing robust inhibition measurements. The compound Ac-IAAA-cmk inhibited GlpG in a concentration- and time-dependent manner (Fig 1B and Supplementary Fig S2A), and mass-spectrometric analysis indicated that it formed a stoichiometric (1:1) complex with the enzyme, which was dependent on the catalytic residues Ser201 and His254 (Supplementary Fig S2B). Upon reaction of Ac-IAAA-cmk with wild-type (wt) GlpG, but not with its S201A and H254A mutants, a faster migrating species on SDS-PAGE arose (Fig 1B and Supplementary Fig S2B). A similar effect has been observed recently upon disulphide cross-linking of TMDs 2 and 5 in GlpG (Xue & Ha, 2013), which suggested that Ac-IAAA-cmk may be cross-linking two TMDs of GlpG. The mass shift of GlpG in the presence of Ac-IAAA-cmk was consistent with the formation of the inhibitor–enzyme complex and elimination of a leaving group of approximately 36 Da (consistent with the molecular weight of HCl). This behaviour was analogous to how CMKs react with classical serine proteases, and we concluded that Ac-IAAA-cmk acted as a mechanism-based inhibitor of GlpG, forming a covalent adduct with the catalytic dyad residues, thus cross-linking TMDs 4 and 6. Furthermore, N-terminal truncation analysis of Ac-IAAA-cmk revealed that the inhibitory potency markedly decreased with progressive truncation of peptidyl chain of the inhibitor (Fig 1C).

Tetrapeptidyl-chloromethylketone inhibitors bind GlpG in a substrate-like manner

To assess whether our peptidyl-CMKs bound to rhomboid in a manner similar to the parent substrate, we analysed the sensitivity of the substrate and inhibitors to identical amino acid changes. We first investigated the subsite preferences of GlpG in the context of the TatA substrate *in vitro* by conducting a complete positional scanning mutagenesis of its P5 to P1 region. The P1 position was the most restrictive one, where GlpG strongly preferred small amino acids with non-branched side chain, such as Ala or Cys (Fig 2A and Supplementary Fig S3); the second most restrictive position was P4 with preference for hydrophobic residues. Positions P5, P3 and P2 were much less restrictive, with P2 accepting almost any amino acid with little impact on cleavage efficiency. Interestingly, aspartate inhibited cleavage profoundly anywhere between P1 to P4 positions, and glycine was not tolerated well at P1, P3 and P4 positions. To verify these results in biological membranes, we introduced some of the strongest inhibitory mutations in the context of full-length TatA into a chimeric substrate construct based on fusions with maltose-binding protein and thioredoxin (Strisovsky *et al*, 2009) and tested the cleavability of the mutants by endogenous GlpG *in vivo*. Consistently, mutations in the P4 position (I5S or I5G), the P3 position (A6D) and the P1 position (A8G or A8V) led to a dramatic decrease in substrate cleavage to nearly undetectable levels, as documented by Western blotting (Fig 2B), confirming our *in vitro* inhibition data.

Having defined the positional sequence preferences of GlpG in a substrate, we determined whether the peptidyl-CMK inhibitors
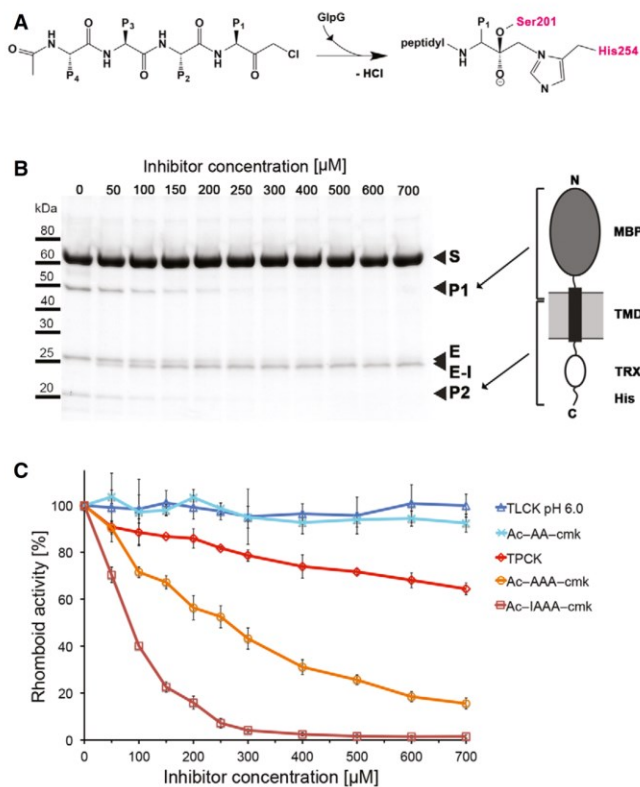
Figure 1.   Tetrapeptidyl-chloromethylketones are mechanism-based inhibitors of rhomboid proteases.

A   Scheme of a tetrapeptidyl-chloromethylketone and mechanism of its reaction with the catalytic dyad of GlpG. In the final adduct, the inhibitor has lost chlorine and is covalently bound to serine 201 and histidine 254.

B   SDS gel showing the inhibition of GlpG by increasing concentrations of Ac-IAAA-cmk. The identity of cleavage products P1 and P2 is illustrated by the schematic drawing of chimeric TatA that was used as substrate (right panel). Reaction of the inhibitor with catalytic residues links TMDs 4 and 6 of GlpG, resulting in a faster migrating band in SDS-PAGE (left panel). MBP, maltose-binding protein; TRX, thioredoxin; S, substrate; P1, product 1; P2, product 2, E, enzyme; E-I, enzyme inhibitor complex.

C   The inhibition properties of chloromethylketones depend on the length and sequence of their peptidyl chain. All compounds were pre-incubated with GlpG for 180 min and reacted with the TatA substrate for 30 min as described in Materials and Methods. The assays were performed in triplicate, and data points show average ± standard deviation.

showed the same specificity, implying a similar binding mechanism. We focussed on the amino acid changes in positions P4, P3 and P1 of TatA that strongly impaired substrate cleavage by GlpG both *in vitro* (Fig 2A) and *in vivo* (Fig 2B): I5S, I5G, A6D, A8V and A8G. These amino acid changes were introduced into the TatA-derived parent compound Ac-IATA-cmk, and inhibitory properties of the resulting compounds were compared at a range of concentrations and fixed pre-incubation time. While all the amino acid changes that impaired cleavage of mutant TatA substrates (I5S, I5G, A6D, A8V and A8G) also profoundly worsened the inhibitory properties of the variant peptidyl-CMKs, those amino acid changes that did not negatively affect cleavage of mutant substrate (T7A and A6S/T7K) had no impact on the inhibitory properties of the respective CMK derivatives (Figs 1C and 2C, and Supplementary Fig S4). This demonstrates

that TatA-derived peptidyl-CMKs bind GlpG in a substrate-like manner and can hence be used as substrate mimetics in crystallographic experiments.

The GlpG:Ac-IATA-cmk complex structure reveals substrate interactions in the active site

The experiments described above provided us with validated tools for structural characterisation of rhomboid–substrate interaction. We co-crystallised Ac-IATA-cmk with the transmembrane core of the wild-type GlpG rhomboid protease and solved the complex structure at 2.1 Å resolution (data collection and refinement statistics in Supplementary Table S1). The electron density for the whole inhibitor was clearly defined and allowed unambiguous model building (Fig 3A).

Figure 2. Tetrapeptidyl-chloromethylketones bind GlpG in a substrate-like manner.

A Specificity matrix of GlpG preferences on TatA variants *in vitro*. Preferences for TatA positions P5–P1 (residues 4–8) are displayed in shades of grey. Substrates show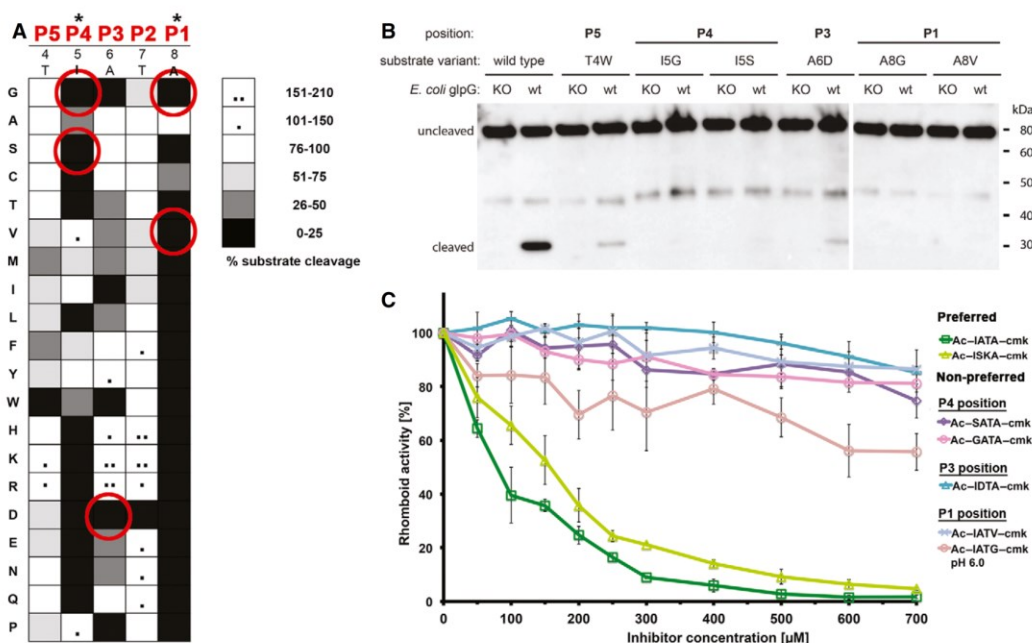ing increased cleavage are additionally marked with black dots. GlpG preferences are most stringent for positions P1 and P4 of TatA. The assays have been done in duplicates and representative source data are shown in Supplementary Fig S3.

B *In vivo* cleavage efficiency of TatA variants with mutations not tolerated *in vitro*. Consistent with the *in vitro* assay, substrates with mutations T4W (P5 position), I5G or I5S (P4 position), A6D (P3 position) and A8G or A8V (P1 position) are refractory to cleavage or show severely inhibited cleavage by GlpG in biological membranes.

C Correlation of effects of amino acid changes in inhibitors with corresponding mutations in substrates. Amino acids that are not tolerated in TatA by GlpG *in vitro* and *in vivo* cause a loss of inhibitory property in the respective inhibitors Ac-GATA-cmk, Ac-SATA-cmk, Ac-IATG-cmk, Ac-IATV-cmk and Ac-IDTA-cmk. The parent compound Ac-IATA-cmk, having the same P1–P4 sequence as wild-type TatA, or its variant Ac-ISKA-cmk harbouring mutations innocuous in the substrate, inhibit GlpG efficiently. The assays have been done in independent triplicates and plotted as average ± standard deviation. Representative source data are shown in Supplementary Fig S4.

The inhibitor is anchored in the active site by two covalent bonds to the catalytic dyad residues S201 and H254, confirming that the CMK warhead reacts as expected. The peptidyl part of the inhibitor fills the active site lying wedged between loops 5 (L5) and 3 (L3), forming a parallel β-sheet with the latter (Fig 3B). The carbonyl oxygen of the CMK warhead forms a weak hydrogen bond to the side chain amido group of N154, but not to the main chain amides of S201 or L200, unlike previously observed in isocoumarin (ISM) and diisopropylfluorophosphonate (DFP) inhibitor complexes (Vinothkumar *et al*, 2010; Xue & Ha, 2012). This minor difference could be a consequence of the covalent binding of the CMK to both the catalytic serine and histidine, which might slightly distort the carbonyl oxygen from the position it would adopt in the natural (singly bonded) tetrahedral intermediate (Mac Sweeney *et al*, 2000). Nevertheless, the position of the P1 carbonyl oxygen is similar to the position of the ISM benzoyl carbonyl (Vinothkumar *et al*, 2010) and DFP phosphonyl oxygens (Xue & Ha, 2012) (Fig 3C), suggesting that the double binding of the CMK warhead to the catalytic dyad is unlikely to affect the conformation of the tetrapeptide ligand in the active site significantly.

The peptide ligand is further stabilised in the active site by hydrogen bonds of its backbone with the backbone carbonyl and amido groups of residues S248/A250 of the L5 loop, and residues G198/W196 of the L3 loop (Fig 3B). Side chain and main chain atoms in each position of the ligand are also engaged in van der Waals interactions with residues of the L3 loop (P1 → G199, P3 → F197), the L5 loop (P2 → M249) and the L1 loop. The terminal P4 isoleucine of the ligand has the right orientation and distance to be considered to interact with the aromatic ring of F146 of the L1 loop via a CH–π interaction (Fig 3B), a weak hydrogen bond with a dominant dispersive character (Brandl *et al*, 2001; Plevin *et al*, 2010). These numerous interactions run along the entire length of the peptide, and, although relatively weak individually, they collectively contribute to the productive positioning of the peptide in the active site in a significant way. This may explain why N-terminal truncations of Ac-IAAA-cmk led to a dramatic progressive decrease in inhibitory potency (Fig 1C).

Since we observed weak sequence preferences also at the P5 position of the substrate (Fig 2A), we solved the GlpG complex

38
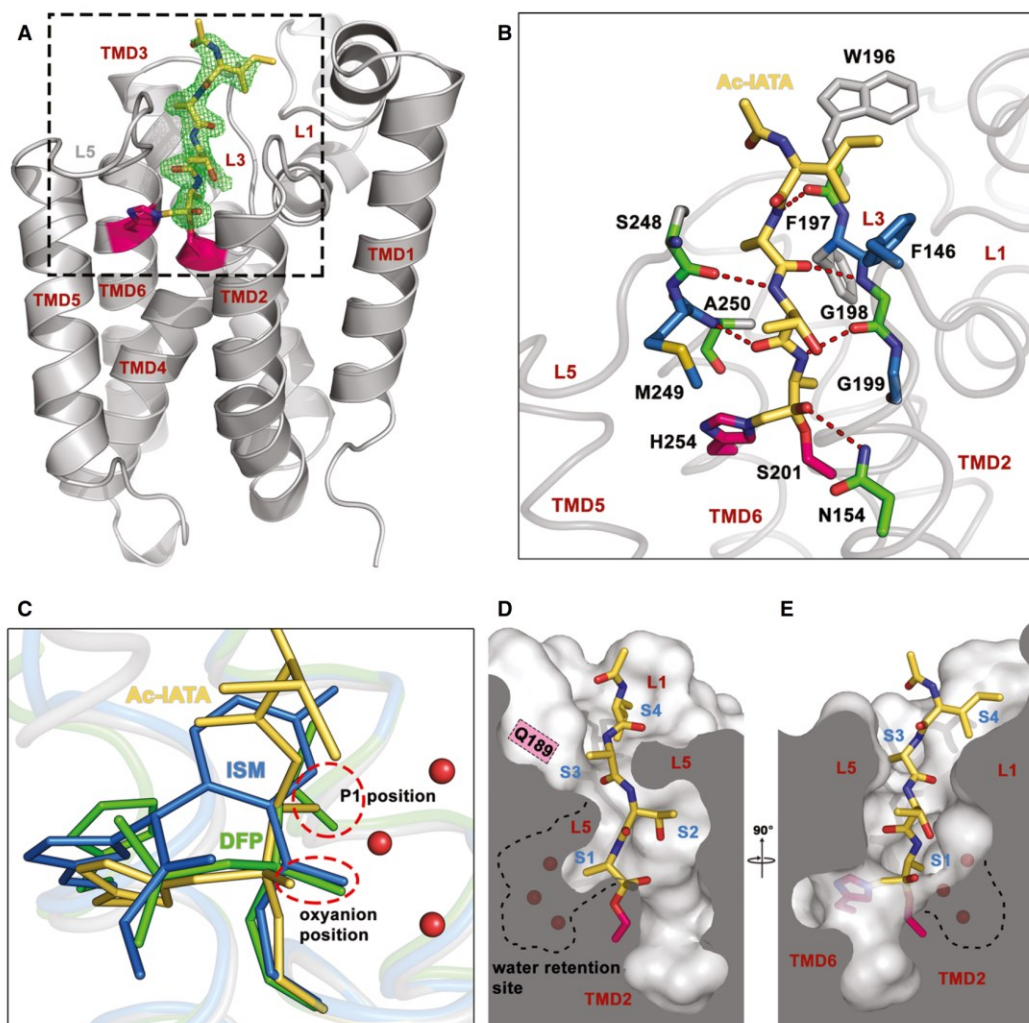
Figure 3. Crystal structure of GlpG complexed to Ac-IATA-cmk reveals the mode of substrate binding to GlpG.

A   Overall structure of the GlpG:Ac-IATA-cmk complex. Ac-IATA-cmk (yellow) is covalently bound to the catalytic histidine and serine (pink). An F_o–F_c simulated annealing omit map, calculated at 2.1 Å and contoured at 3 σ, is shown 2 Å around Ac-IATA-cmk. Crystallographic statistics is shown in Supplementary Table S1.

B   Interactions between Ac-IATA-cmk and GlpG. GlpG residues forming hydrogen bonds with Ac-IATA-cmk are coloured green. These residues are additionally engaged in van der Waals (vdW) contacts; residues making vdW contacts only are depicted in blue. All interactions were calculated using the program Ligplot.

C   Superposition of the GlpG:Ac-IATA-cmk complex with the isocoumarin (ISM) and diisopropylfluorophosphonate (DFP) complexes (PDB-IDs 2XOV and 4H1D, respectively). The side chain of the P1 alanine superimposes well with one of the DFP isopropyl groups and points into the S1 subsite. The ISM ring points away from the P1 alanine, but still into the cavity. The oxyanion position is occupied by the superimposing DFP phosphonyl and the ISM benzoyl carbonyl oxygens. The P1 alanine carbonyl oxygen of Ac-IATA-cmk points slightly away from them and forms a hydrogen bond with N154 (see panel B).

D, E   Ac-IATA-cmk bound in the active site of GlpG. Shown are views from two different angles. Positions of the S1–S4 subsites are indicated. A dashed line contours the water-filled S1 subsite. Water molecules are depicted as red spheres. ISM, 7-amino-4-chloro-3-methoxy-isocoumarin; DFP, diisopropyl fluorophosphonate.

with the pentapeptide Ac-TIATA-cmk to get insight into their structural basis. However, no additional electron density for the P5 threonine could be observed in this structure, and the overall orientation of the P1–P4 residues was the same as in Ac-IATA-cmk complex (Supplementary Fig S5). These findings indicate that substrate residues beyond P4 are unlikely to interact with GlpG significantly and are completely solvent-exposed. This is consistent with the observation that only hydrophobic amino acids

are not tolerated well in the P5 position of the substrate (Fig 2A).

Substrate-binding subsites in GlpG

The structure of Ac-IATA-cmk complex with GlpG reveals substrate interactions in the active site of a rhomboid protease, allowing us to correlate them to the observed amino acid preferences in the TatA substrate from which Ac-IATA-cmk is derived (Fig 2A). GlpG shows a strict requirement for a small P1 residue, strongly preferring alanine and less well accepting cysteine and serine (Fig 2A). The side chain of the P1 alanine in Ac-IATA-cmk is bound into a well-formed S1 subsite, corresponding to the one proposed earlier (Vinothkumar *et al*, 2010) (Fig 3C). The S1 subsite is the proximal part of a deeper cavity, whose distal part has a strongly hydrophilic character with negative surface electrostatic potential (Supplementary Fig S6) and contains three conspicuous conserved water molecules present in all structures of GlpG from different crystallisation conditions and space groups (Wang *et al*, 2006; Ben-Shem *et al*, 2007; Vinothkumar, 2011). It was recently proposed that this region constitutes a 'water retention site' in GlpG that facilitates channelling of water molecules from the aqueous environment into the body of the hydrophobic protease to confer catalytic efficiency (Zhou *et al*, 2012; Fig 3D and E). The mechanistic implications of its proximity to the S1 subsite will be discussed later.

In contrast to the P1 position, P2 and P3 positions in TatA are relatively insensitive to residue changes (Fig 2A). Consistent with this, both S2 and S3 subsites are large and open enough to accommodate residues of any size. While the S2 subsite is half-open to the periplasm, S3 subsite resembles a mere notch in the rim of the active site of GlpG, through which the side chain of the P3 alanine of Ac-IATA-cmk points towards Q189 (Fig 3D and E). The P4 isoleucine of the bound peptide interacts with the aromatic ring of F146, possibly via a CH–π bond. This interaction defines the S4 subsite as a recessed area on the periplasmic face of GlpG, the borders and bottom of which are delineated mainly by residues of the L1 loop with some contribution from the side chain of W196 in the L3 loop. This patch is unusual because it is fully solvent-exposed, yet strongly hydrophobic in nature (Fig 4), which suggests functional importance. Indeed, the character of the S4 subsite provides a structural explanation of the preference for large and hydrophobic residues and the intolerance for polar residues in the P4 position of TatA (Fig 2A).

The S4 subsite is plastically formed by residues of the L1 loop

As P4 residue crucially contributes to substrate recognition by several rhomboids (Strisovsky *et al*, 2009), strongly influencing mainly the k_{cat} of the reaction (Dickey *et al*, 2013), we examined the functional and structural properties of S4 subsite in greater detail. The mutation of F146 to alanine was reported to inactivate GlpG without substantially affecting its thermodynamic stability (Baker & Urban, 2012), which was previously difficult to explain. Since F146 interacts with the P4 residue side chain of the substrate, we hypothesised that mutations in F146 could actually affect P4 specificity of GlpG. To test this hypothesis, we engineered complementary enzyme and substrate mutants by introducing hydrophobic residues of different side chain volumes to position 146 of GlpG

(F146A and F146I) and by testing their activity against all 20 possible mutations in the P4 position of TatA substrate. Indeed, the F146A mutant was not inactive as previously reported (Baker & Urban, 2012), but it rather showed a shift in specificity for the P4 residue. TatA variants with smaller residues in P4 position (e.g. A, C, V) were cleaved less efficiently by both the F146A and F146I mutants than by wt GlpG, while TatA variants with larger hydrophobic side chains in P4 position (such as M, F, W) were cleaved significantly better by F146A and F146I mutants than by wt GlpG (Fig 4A and Supplementary Fig S7).

To understand the properties of S4 subsite structurally, we determined the structures of wt GlpG and its F146I mutant complexed to Ac-FATA-cmk (2.9 and 2.55 Å resolution, respectively, Supplementary Table S1) and compared the ligand-binding mode to the parent structure of GlpG and Ac-IATA-cmk complex. Interestingly, the P4 residue of the ligand binds GlpG in a slightly different way in the three complexes (Fig 4B), illustrating the plasticity of S4 subsite. In wt GlpG, the isoleucine of Ac-IATA-cmk interacts with the main chain atoms of W196 of the L3 loop and the side chain of F146 (Fig 4B), while the ring of the P4 phenylalanine of Ac-FATA-cmk is accommodated additionally by the side chain of M120 contributing to the hydrophobic patch that constitutes the S4 subsite (Fig 4B). In the F146I mutant of GlpG, the P4 phenylalanine points down into a well-formed, hydrophobic pocket and engages in contacts with the main chain atoms of F197 and G198 of L3 loop and the side chains of I146 and M144 of L1 loop (Fig 4B). Our structural analyses therefore reveal a function for the L1 loop in rhomboid specificity determination: the S4 subsite is plastically formed by the side chains of three L1 loop residues, aided by the main chain atoms of L3 loop. This finding is consistent with the observations that mutations at the L1-L3 loop interface often lead to a significant decrease in GlpG activity (Baker & Urban, 2012).

Structural changes upon inhibitor binding—implications for rhomboid mechanism

The previously published inhibitor-bound complex structures of GlpG (Vinothkumar *et al*, 2010, 2013; Xue & Ha, 2012; Xue *et al*, 2012) were useful first approximations for uncovering the structural changes involved in GlpG catalysis, but the small size and chemical dissimilarity of the inhibitors to a polypeptide limited their use as models for substrate binding. The present structures of GlpG with substrate-derived peptides resemble the tetrahedral intermediate and the acylenzyme, thus allowing us to characterise more accurately structural changes during catalysis.

Alignment of the unliganded and Ac-IATA-cmk complex structures of GlpG (Fig 5A and B) reveals that only minor TMD movements occur in the complex. TMD6 is slightly turned inwards in the ligand-bound state, but this may be the consequence of the double binding of the CMK warhead to both H254 and S201 (Mac Sweeney *et al*, 2000). The lateral movement of TMD5, thought to be required for substrate access (Baker *et al*, 2007), is negligibly small in the Ac-IATA-cmk complex structure. However, since our ligands include neither the TMD of the substrate nor the prime-side residues, which would probably co-localise with the top of TMD5 in the enzyme–substrate complex, we cannot exclude the possibility of larger TMD5 movements in other phases of the catalytic cycle of rhomboid. The most dramatic secondary structure changes involve

	wt	I5G	A	S	C	T	V	M	L	F	Y	W	H	K	R	D	E	N	Q	P
■ wt GlpG	100	7	30	13	48	20	100	28	28	46	83	39	11	22	20	0	0	0	15	104
■ F146A GlpG	70	0	0	0	17	9	50	104	33	143	50	139	0	0	4	4	2	2	2	9
■ F146I GlpG	91	2	0	0	33	9	80	74	22	163	63	111	2	15	4	2	2	0	2	24

GlpG + Ac–IATA–cmk   GlpG + Ac–FATA–cmk   GlpG F146I + Ac–FATA–cmk
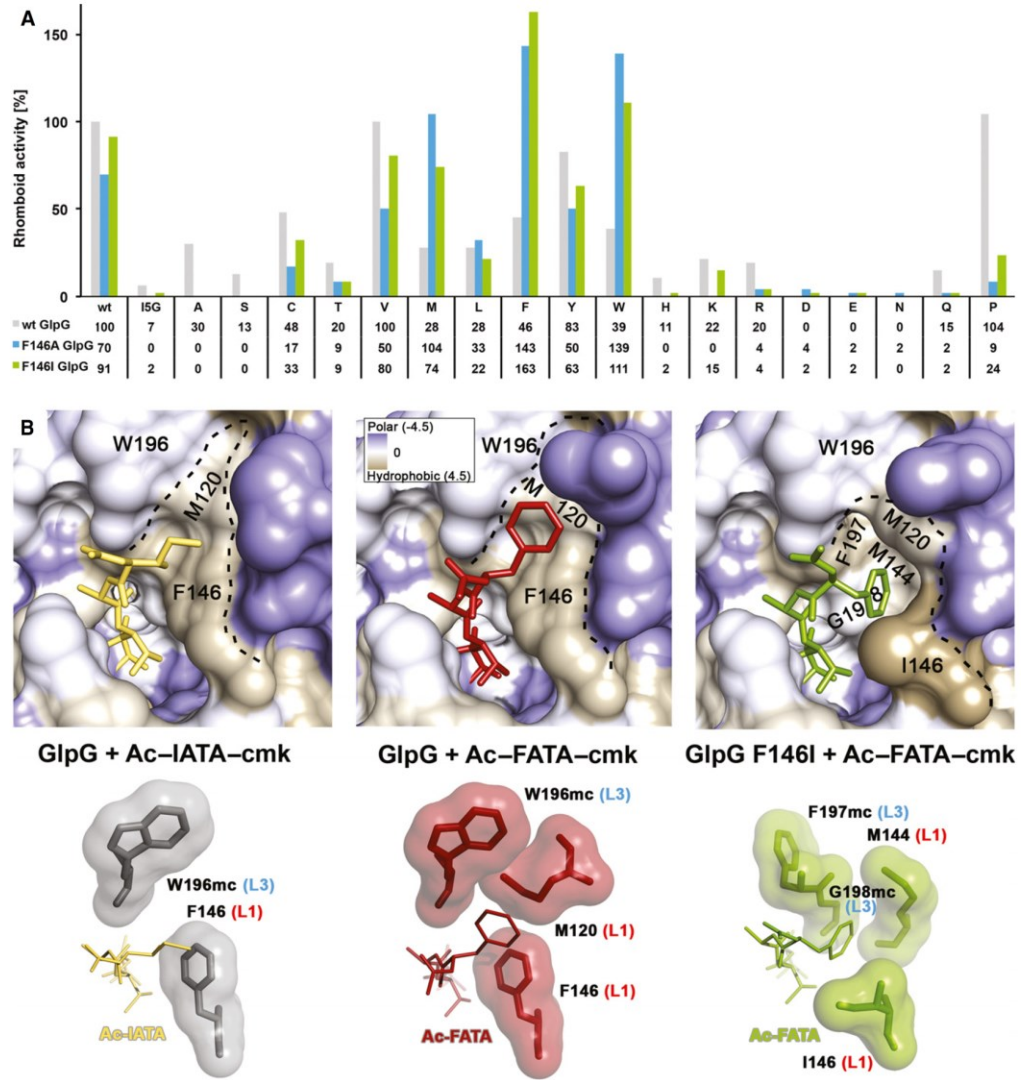
Figure 4. A patch of hydrophobic residues from the L1 loop forms the S4 subsite of GlpG.

A   Compensatory effect of mutations in the S4 subsite of GlpG and the P4 position of TatA. Mutation of the S4 subsite residue F146 to the smaller hydrophobic residues alanine or isoleucine is nearly or completely inactivating only for substrate variants with small- to medium-sized side chains (A, S, C, T, V, I) in the P4 position. This attenuating effect of the GlpG S4 subsite mutants can be compensated by a mutation of the TatA P4 residue to a residue with a bigger side chain. Hence, while wild-type (F146) GlpG cleaves the TatA substrate with a large P4 residue (such as I5W) very poorly, the activity can be fully recovered by replacing the bulky phenylalanine 146 in the S4 subsite of the enzyme by a small hydrophobic side chain (such as F146A). The assays have been conducted three times independently, and representative data are shown (source data in Supplementary Fig S7).

B   The S4 subsite in the complex structures GlpG:Ac-IATA-cmk, GlpG:Ac-FATA-cmk and GlpG_F146I:Ac-FATA-cmk. Crystallographic statistics are shown in Supplementary Table S1. Upper panels: close-up view of the S4 subsite with the surface of GlpG coloured according to the Kyte-Doolittle hydrophobicity scale (Kyte & Doolittle, 1982). The S4 subsite is a surface-exposed hydrophobic patch formed by residues from the L1 loop. Lower panels: residues making vdW interactions with the bound tetrapeptide are shown as contact surfaces. For residues W196, F197 and G198 of the L3 loop, only main chain atoms (mc) are engaged in interactions. Residues F146, M120, M144 and I146 of the L1 loop make vdW contacts with their side chains. The isoleucine in the P4 position of Ac-IATA forms a CH-π interaction with the aromatic side chain of F146.

Figure 5. Binding of Ac-IATA-cmk to GlpG induces displacement of the L5 loop and side chain rotamer changes in TMD2 and 5.

A   Structural alignment of unliganded GlpG and the GlpG:Ac-IATA-cmk complex, side view. Unliganded GlpG (PDB-ID 2IC8) is coloured in red, the complex in grey/yellow. Ac-IATA-cmk is represented as contact surface. Residues with different side chain rotamers are shown for TMD2, TMD5 and the L5 loop.
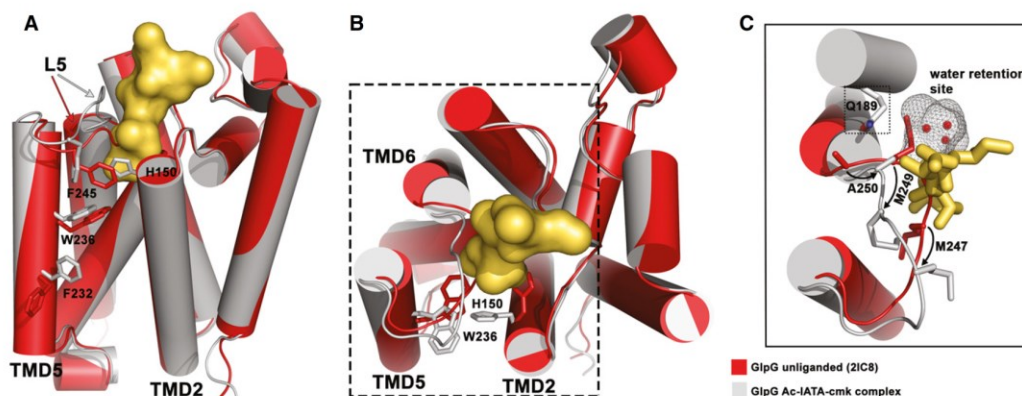B   Top view. F232 and F245 are omitted for clarity.
C   Displacement of the L5 loop causes a positional shift for M247, M249 and A250. Both methionines form the S2 subsite in the complex. A250 adopts the position in the complex that was occupied by M249 in the apoenzyme. The presence of the shorter side chain of alanine between Q189 and the 'water retention site' might play a role in facilitating the relay of water molecules for the reaction (see Discussion).

the L5 loop: it caps the active site in the apoenzyme while swinging upwards and shifting laterally upon binding of Ac-IATA-cmk (Fig 5A and B).

In addition to secondary structure changes, we detect several pronounced rotamer changes in residues of TMD2, TMD5 and L5 loop, which may indicate the importance of these residues for the catalytic mechanism. The movement of the L5 loop inflicts a positional change on the side chains of M247 and M249 (Fig 5C), having profound impact on S1 and S2 subsite formation and potentially also on catalysis (see Discussion). Upon binding of Ac-IATA-cmk, M249 shifts and becomes engaged in van der Waals interactions with the methyl group of threonine in the P2 position of the substrate, while the original position of M249 in the unliganded enzyme is adopted by A250 in the complex structure (Fig 5C). Methionine 247 fills the centre of the active site in the apoenzyme, while in the complex structure, it moves to the entrance of the active site, where it confines the S2 subsite together with H150. In the apoenzyme, the side chain of H150 fills the space that corresponds to the S2 cavity, swinging far out from this position upon binding of Ac-IATA-cmk. If H150 stayed in its original position, it would sterically clash with the side chain of the P2 threonine (Fig 5A and B), suggesting that the role of H150 in catalysis may be more dynamic than previously thought.

Several other conspicuous rotamer changes occur in the Ac-IATA-cmk complex. The L5 residue F245 obstructs the entrance to the active site at the level of the catalytic dyad residues in the apoenzyme, while in the complex structure, it has rotated to the side (Fig 5A). Given the position of F245 and the fact that F245A mutation results in a modest enhancement of proteolytic activity (Baker & Urban, 2012), it is suggestive that rotation of F245 may be required for substrate entry into the active site. The indole ring of W236 of TMD5 has rotated 180° in the complex when compared to the apoenzyme, thus allowing the formation of an internal cavity

thought to represent the S2′ subsite (Vinothkumar *et al*, 2010, 2013) (Fig 5A and B). It is noteworthy that this cavity forms even in the absence of prime-side residues in our complex or in complexes with small molecular inhibitors, isocoumarins and β-lactams (Vinothkumar *et al*, 2010, 2013; Xue & Ha, 2012). Finally, residue F232 of TMD5 is also found in a different conformation in the complex structure than in the apoenzyme, closing the gap to TMD2 residue W157 (Fig 5A). Since the F232A mutation has been shown to result in increased enzymatic activity (Baker & Urban, 2012), it is possible that F232 directly or indirectly participates in substrate binding.

Molecular dynamics reveals active site interactions of the substrate in the Michaelis complex

Besides revealing the substrate-binding subsites on GlpG, crystal structures of the peptidyl-CMK complexes enabled us to investigate rhomboid mechanism in closer detail. We used the complex structures, molecular modelling and molecular dynamics (MD) to create a model of the Michaelis complex of rhomboid protease and the substrate spanning the P4 to P3′ subsites. The model was validated by monitoring (i) the root-mean-square deviation (RMSD) of protein and substrate backbone (Supplementary Fig S8A) and (ii) hydrogen bonds (H-bonds) at the non-prime side of the substrate during the MD run. Throughout MD simulations, H-bonds between the L3/L5 loop and the substrate backbone, as present in the crystal structure (Fig 3B), were retained (Supplementary Fig S8B). Furthermore, we observed (i) the formation of H-bonds between the catalytic dyad residues, (ii) the scissile bond carbonyl carbon and the S201 side chain oxygen coming into close spatial proximity compatible with nucleophilic attack, and (iii) formation of H-bonds between the P1 carbonyl oxygen and residues thought to form the oxyanion hole (Supplementary Fig S8B). The interactions (iii) involved mainly the H-bonds by the N154 side chain nitrogen and by the S201 main

chain amide. The former H-bond was stable, while the latter one was transient, and the previously observed H-bond to L200 main chain amide (Vinothkumar *et al*, 2010) could not be detected. During MD simulations, H150 transiently flipped back into the position it adopts in the unliganded enzyme (data not shown), suggesting that H150 (and maybe also L200) may hydrogen-bond to the negatively charged oxyanion that forms in the tetrahedral intermediate (but is absent from the Michaelis complex). Overall, the carbonyl oxygen of the P1 residue adopts a similar orientation in our MD simulations as found in the complex structure with diisopropylfluorophosphonate (DFP), deemed to mimic the tetrahedral intermediate (Xue & Ha, 2012) (Supplementary Fig S8C). This finding makes us confident that our MD model of the Michaelis complex (Fig 6A) is realistic, allowing us to examine the interactions of the prime-side residues with GlpG and estimate the likely exit position of the unwound C-terminus of the substrate from the body of GlpG.

The MD model of the Michaelis complex reveals the likely interactions of the P2′ residue, which is important for substrate recognition by *P. stuartii* AarA and *E. coli* GlpG rhomboids (Strisovsky *et al*, 2009; Dickey *et al*, 2013). The major ensemble (92%) of conformations of the P2′ phenyl of TatA (Supplementary Fig S8D and E) snugly fits into the previously proposed S2′ subsite (Vinothkumar *et al*, 2010, 2013). The 'back wall' of the subsite is formed by residues of TMD4 deeply buried within the core of the enzyme (Supplementary Fig S8E). The bulk of this interaction interface is provided by Y205, assisted by V204, M208 and A233, all of which make van der Waals contacts to the P2′ residue of the substrate. Phenylalanine 245, located at the tip of L5 loop, constitutes the

'roof' above the S1′ and S2′ subsites, making van der Waals contacts with the P1′ and P2′ residues (Supplementary Fig S8E). Amino acids F153 and W157 of TMD2 and W236 of TMD5 form the outer rim of the active site cavity that opens to the lipid bilayer, making van der Waals contacts to the P2′ residue as well as to the glycine in P3′ position (Supplementary Fig S8E). This arrangement suggests that F153, W157 and W236 could directly interact with the substrate as opposed to having just an indirect 'gating' role in limiting the mobility of TMD5, as proposed earlier (Baker *et al*, 2007).

Our data indicate that the full extent of the enzyme–substrate interactions in the active site of GlpG comprises a stretch of seven consecutive residues of the substrate in an extended conformation, from the P4 to P3′ position (I5 to G11 in TatA) (Fig 6A). The P3′ glycine marks the end of the unwound part of the TatA substrate, suggesting that its transmembrane helical part begins just after the helix-destabilising proline in P4′. The P3′ glycine exits the active site of GlpG within or just above the plane of the Cα atoms of residues W236 and F153. It was recently reported that intramolecular disulphide cross-linking of a W236C/F153C mutant of GlpG via 1,2-ethanediyl bismethanethiosulfonate (M2M) does not impair enzyme activity (Xue & Ha, 2013), suggesting that substrate accesses the active site above these residues (above the M2M cross-link). That report is compatible with our MD simulations, since the Cα–Cα distance between W236 and F153 is 12.5 ± 0.6 Å, which matches the calculated distance of 13 Å between the Cα atoms of the M2M-cross-linked cysteine pair mutant, calculated from the respective MD model (Fig 6A).

In conclusion, our crystallographic, biochemical and molecular dynamics data reveal for the first time substrate interactions in the
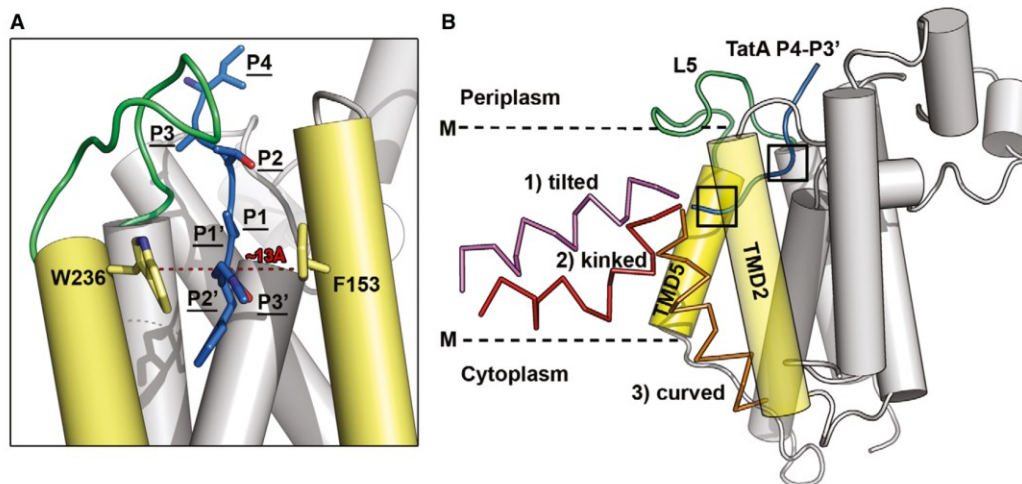
Figure 6. Molecular dynamics-based model of the Michaelis complex and possible interaction modes of substrate transmembrane domain.

A   The molecular dynamics model for the active site bound unwound part of TatA comprising positions P4-P3′. The P3′ residue exits the active site of GlpG between W236 (TMD5) and F153 (TMD2).

B   A cartoon model of the full transmembrane TatA substrate interacting with GlpG. The substrate continues by the N-terminus of its helical transmembrane domain from the point of where its P3′ residue 'exits' the active site of GlpG in the Michaelis complex model. This arrangement suggests three principally different orientations of TatA TMD that are shown in the illustrative cartoon in different colours. Detailed views of the boxed areas around TatA P4-P3′ segments are shown in Supplementary Fig S8D and E.

active site of an intramembrane protease, explain the observed substrate specificity of rhomboid proteases structurally and reveal a role in substrate binding for the hitherto enigmatic conserved element of the rhomboid fold—the L1 loop. Besides providing new insights into intramembrane protease mechanism, our work raises testable mechanistic hypotheses that, if confirmed, could facilitate development of selective rhomboid inhibitors.

Discussion

Understanding of the mechanism and specificity of intramembrane proteases would be significantly advanced by high-resolution structural characterisation of substrate binding, but it has long been an unattained goal. Rhomboids, the most structurally characterised intramembrane proteases, have so far been co-crystallised only with small molecular mechanism-based inhibitors (Vinothkumar *et al*, 2010, 2013; Xue & Ha, 2012; Xue *et al*, 2012; Vosyka *et al*, 2013) useful for only indirect inferences about mechanism and specificity (Vinothkumar *et al*, 2010, 2013). We have developed a new series of peptidic chloromethylketone inhibitors based on a natural bacterial rhomboid substrate sequence (*Providencia stuartii* TatA) (Stevenson *et al*, 2007) and solved X-ray structures of their complexes with GlpG, thus providing the first structural insight into substrate binding to rhomboids. We reveal the subsites for the P1–P4 residues that had been demonstrated to be crucial for substrate recognition and efficient catalysis (Strisovsky *et al*, 2009; Dickey *et al*, 2013). Furthermore, we show that the S4 subsite is formed by residues of the highly conserved but previously enigmatic L1 loop, leading us to propose that the domain topologically equivalent to the L1 loop may have evolved for client-protein recruitment in rhomboid-like pseudoproteases.

Rhomboid substrate binding—the unwound, the destabilising and the helical

The peptidyl-CMKs used in this study exhibit identical specificity requirements to natural substrates, validating their ability to provide mechanistic insight. We can now use our data in combination with previous structural and biochemical work to propose a plausible working model of the enzyme–substrate complex. Our work shows that the non-helical P4 to P3′ segment of the substrate is in contact with the active site cleft of GlpG. The importance of the P4, P1 and P2′ positions in the substrate (Strisovsky *et al*, 2009) was recently confirmed by showing that they determine the k_{cat} of rhomboid cleavage (Dickey *et al*, 2013). These residues have only a negligible impact on K_M (Dickey *et al*, 2013), suggesting that they do not make a major contribution to the overall binding energy between a full transmembrane substrate and the enzyme. This in turn implies that the overall interaction area of rhomboid–substrate complex is significantly larger than the segment containing the P4 to P2′ residues, and the majority of overall binding energy of the substrate is probably contributed by the part of its TMD directly contacting the enzyme. The mode of binding of substrate TMD is unknown, but our structures and MD models provide a solid framework to reflect on it.

To propose a structure-based conceptual model of a full transmembrane substrate complex with GlpG, we took advantage of the recent solution NMR structure of *E. coli* TatA (Rodriguez *et al*, 2013). A homology model of *P. stuartii* TatA that we generated shows that the region spanning residues P13 (P4′ position) to F27 is α-helical and about 22 Å long. The estimated hydrophobic thickness of GlpG molecule from the point of exit of the P3′ residue to the cytoplasmic boundary of the membrane is about 13 Å (Fig 6B), and manual docking of *P. stuartii* TatA TMD region P13 (P4′ position) to F27 into a representative structure of the Michaelis complex model suggests that the TatA TMD would 'stick out' of the membrane.

Such hydrophobic mismatch would be energetically unfavourable, and different ways of alleviating it can be envisaged, for example (i) tilting of substrate TMD in the membrane or (ii) minimising the solvent-exposed hydrophobic surface area of substrate TMD by its interaction with GlpG. In the first scenario (i), a tilted but straight TMD of the substrate (Fig 6B) would have virtually no interaction interface with the transmembrane region of GlpG (unless GlpG is also tilted in the membrane accordingly) and might therefore be less likely. However, a tilted orientation with a kinked α-helix would still allow some interaction with the transmembrane region of GlpG, making it perhaps more likely (Fig 6B). In the second scenario (ii), a slight 'inward' curving of the substrate transmembrane helix that would allow its alignment and interaction with TMD2 of GlpG (which is also slightly bent) might provide a larger interaction interface and shield much of the 'mismatched' TMD from the solvent (Fig 6B). Indeed, such a mechanism has been described in cases where positive mismatch is bigger than 4 Å (Lewis & Engelman, 1983). Interestingly, introducing transmembrane helix-destabilising residues at several positions along the TMD of an artificial rhomboid substrate increases its cleavage efficiency by GlpG (Akiyama & Maegawa, 2007; Moin & Urban, 2012), but this effect has been difficult to explain (Ha, 2009). Now, our conceptual models of the complex where substrate TMD is kinked or bent (Fig 6B) would both be consistent with and explain these observations.

Structural changes in rhomboid accompanying substrate binding

Crystal structures of model intramembrane proteases suggest that substrate access to their catalytic residues may be conformationally regulated (Strisovsky, 2013). Based on the alternative conformation of one molecule in the asymmetric unit of a crystal structure of GlpG (Wu *et al*, 2006), substrate access to rhomboid protease had been suggested to be governed by a 'gating' mechanism. In analogy to the translocon (Van den Berg *et al*, 2004), this mechanism should involve a large dislocation of TMD5 to make the core of the enzyme accessible laterally from the lipid bilayer (Wu *et al*, 2006; Baker *et al*, 2007). Mutations in residue pairs W236A/F153A and F232A/W157A, designed to weaken the contacts between TMD2 and 5, increased enzymatic activity, supposedly by opening the TMD5 gate (Baker *et al*, 2007), which was further supported by enzymatic and thermodynamic studies (Baker & Urban, 2012; Moin & Urban, 2012). In contrast, other authors showed that preventing large lateral movement of TMD5 by chemically cross-linking TMDs 2 ad 5 in a W236C/F153C mutant does not abrogate the activity of GlpG. This suggests that a 'gating' movement of TMD5 may not actually be required for substrate binding, and it leaves the mechanism of substrate access to rhomboid controversial.

Our structures of the peptidyl-CMK complexes show that the L5 loop has to be displaced significantly to allow binding of substrate

to the active site, but we do not observe any significant movement of the adjoining TMD5. Since our peptide ligands comprise only the non-prime-side residues and capture the reaction at the stage of the tetrahedral or acylenzyme intermediate, we explored rhomboid–substrate interactions at the prime side and possible involvement of TMD5 by molecular modelling and dynamics. The results show that a large lateral movement of TMD5 is not required for the formation of the acylenzyme nor the Michaelis complex with the P4 to P3′ segment of the substrate. Our data are thus compatible with the published cross-linking data suggesting that major movements of TMD5 are not required for substrate access (Xue & Ha, 2013). We cannot formally exclude the possibility of a large TMD5 movement in the earlier phases of a transmembrane substrate binding. However, the positions of residues W236 and F153, which we observe in the Michaelis complex model (Fig 6A and Supplementary Fig S8E), suggest that they may directly interact with the substrate, rather than just acting as 'openers' of the TMD5 gate. These results collectively imply that the lateral gate opening analogy with the translocon (Wu *et al*, 2006; Baker *et al*, 2007) may not be entirely correct and that substrate access mechanism to rhomboid merits further investigation.

Several other conspicuous movements of side chains accompany ligand binding, among which H150 is worth highlighting. Histidine 150 flips out completely from its position in the unliganded enzyme to make space for the P2 residue of the ligand, which can be almost any amino acid type (Fig 2A). In this conformation, however, the side chain of H150 cannot make a hydrogen bond to the carbonyl oxygen of the substrate. This dislocation of H150 could well be partly due to the chloromethylketone warhead binding to the catalytic dyad and slightly distorting the carbonyl oxygen (Fig 3C; Mac Sweeney *et al*, 2000). Indeed, our MD simulations of the Michaelis complex suggest that the side chain of H150 can occasionally flip to its original position (M. Lepšík, S. Zoll, K. Strisovsky, unpublished observations), although this may be less likely in substrates with larger P2 residues. Interestingly, the side chain of H150 occupies a similar position in the crystal structure of GlpG complex with 2-phenylethyl 2-(4-azanyl-2-methanoyl-phenyl) ethanoate (Vosyka *et al*, 2013) as it does in our Ac-IATA-cmk complex, but it is covalently bound to the inhibitor. In summary, these observations collectively indicate that the role of H150 in catalysis may be more dynamic than previously thought and may extend beyond oxyanion hole formation.

Water access to the catalytic site—a key open question

To better understand intramembrane proteolysis, one of the key aspects to consider is the mechanism of water supply to the catalytic site immersed in the hydrophobic environment of the lipid bilayer. It was recently proposed, based on molecular dynamics and mutagenesis data, that GlpG employs a specific mechanism to channel water molecules from bulk solution to an internal 'water retention site' near the catalytic dyad (Zhou *et al*, 2012). Our structural data are consistent with this concept and offer a plausible mechanistic interpretation based on several observations. First, the 'water retention site' forms a continuous cavity with the S1 subsite of GlpG. Although the whole cavity is quite large, only alanine and to lower extent also cysteine or serine are accepted in the P1 position of the substrate. One explanation could be that the strongly negative electrostatic

potential of this cavity (Supplementary Fig S6) disfavours binding of negatively charged residues and residues with longer aliphatic side chains than that of alanine. Polar natural amino acids other than serine are likely to be either too large to be accommodated (K, R, H) or might engage in hydrogen bonds to the water molecules inside the retention site, thus perturbing the described dynamic hydrogen bonding network (Zhou *et al*, 2012). Such interference could result in (i) structural destabilisation of the enzyme–substrate complex or (ii) impaired catalysis as water molecules may not effectively access the catalytic site to be used in the deacylation step. The latter mechanism is experimentally testable, since one would predict that a substrate with a P1 residue of a suitable character larger than an alanine could be trapped at the acyl-enzyme stage, bound to the catalytic serine. However, given the structural restrains of the cavity and the structural properties of genetically encoded amino acids, testing this hypothesis might require the use of unnatural amino acids. Our structural analyses also rationalise why glycine is poorly tolerated in the P1 position of a substrate and the corresponding peptidyl-CMK. The poor tolerance cannot be due to steric hindrance because glycine has no side chain, but it can be caused by a higher degree of rotational freedom endowed by glycine, which could prevent optimal alignment of the ligand's polypeptide chain for hydrogen bonding to the L3 loop backbone in a parallel β-strand and productive exposure of the scissile bond to the catalytic residues.

A second observation relates to glutamine 189 that had been proposed to channel water molecules to the water retention site (besides S185, H141 and S181). The side chain of the P3 residue of the substrate/inhibitor points directly at Q189 (Fig 3D). We can thus speculate that substitution of the P3 alanine in Ac-IATA-cmk by a residue that can either sterically interfere with Q189 (e.g. W in Fig 2A) or form direct or water-mediated hydrogen bonds with Q189 (e.g. D, E, N in Fig 2A) could result in a loss of proteolytic activity due to the interference with water channelling into the retention site. Third, residue M249 from the L5 loop protrudes right in between Q189 and water molecules in the water retention site, again potentially interfering with water channelling to the water retention site. Upon ligand binding, the L5 loop is displaced, and the position of M249 side chain is adopted by the side chain of A250, which may 'unblock' the pathway from Q189 to the water retention site (Fig 5C). Although necessarily speculative, the mechanism of water access control supported by the above observations deserves further investigation, also because if proven correct it could represent a unique rhomboid-specific mechanism exploitable in the design of selective rhomboid inhibitors.

L1 loop—a prominent feature of the rhomboid fold—binds substrate

We find that the S4 subsite of GlpG is, unexpectedly, formed by a patch of hydrophobic but solvent-exposed residues from the L1 loop. This interaction surface is plastic, and substitution of the P4 residue requires adjustment of residues in the S4 subsite to maximise the number of van der Waals contacts and preserve catalytic efficiency (Fig 4). Notably, the only other structurally characterised rhomboid, GlpG from *Haemophillus influenzae* (Lemieux *et al*, 2007), contains a similar solvent-exposed hydrophobic patch formed mainly by L61 (EcGlpG equivalent F146), V59 (EcGlpG eq. M144) and M35 (EcGlpG eq. M120) (Supplementary Fig S9A), allowing for substrate

45

interactions comparable to the ones observed in the S4 subsite of EcGlpG (Supplementary Fig S9B). In fact, most GlpG homologues harbour hydrophobic residues at the positions corresponding to F146, M144 and M120 of EcGlpG (Supplementary Fig S9C), suggesting that this specificity feature is more widely conserved.

Given how large and diverse the rhomboid protease family is [less than 15% of sequence identity in the conserved region (Koonin *et al*, 2003)], it is expected that substrate specificity and S4 subsite preferences may differ among phylogenetic clusters of rhomboids. Nevertheless, some key features of rhomboid architecture are likely to be used for a similar purpose even in distant homologues. It has been recently suggested that rhomboids are dimeric (Sampathkumar *et al*, 2012), and that natural substrates induce dimer-dependent allosteric activation of the enzyme (Arutyunova *et al*, 2014). The molecular details of the dimerisation interface and the basis for the allosteric regulation are unknown (Strisovsky & Freeman, 2014), but it is attractive to speculate that either of them may involve the L1 loop. Notably, this region of rhomboid architecture, topologically corresponding to the L1 loop, is present in Derlins, and has expanded in size and been conserved in iRhoms (called iRhom homology domain) (Lemberg & Freeman, 2007). Taking the implications of our work evolutionarily further, we speculate that the L1 loop region may have evolved for the interaction with client proteins also in iRhoms and other proteins of the rhomboid-like superfamily (Freeman, 2014).

Materials and Methods

Chemical synthesis

Peptidyl-chloromethylketone inhibitors were prepared by coupling of the protected N-α-acetyl-peptide fragment and the corresponding chloromethylketone derived from the C-terminal (P1) amino acid synthesised analogously with previously described methods (Thomson & Denniss, 1973; Owen & Voorheis, 1976; Jahreis *et al*, 1984; Hauske *et al*, 2009). Acidolabile *tert*-butyl type groups were used for protection of side chain functionalities. The resulting peptidyl-chlomethylketones were then deprotected by trifluoroacetic acid and purified by reversed-phase HPLC. Identity of all compounds was confirmed by mass spectrometry on Waters Micromass ZQ ESCi multimode ionisation mass-spectrometer, using ESI-ionisation method (ESI-MS) and NMR (Bruker AV-400 MHz, data collected at room temperature). Stability of the compounds in aqueous buffers was analysed by reversed-phase HPLC with UV and ESI-MS detection (Supplementary Fig S1), and their solubility was checked using Millipore low-binding hydrophilic centrifugal filters and HPLC with UV detection. Full experimental details on chemical synthesis and analytical characterisation of all synthesised compounds are included in Supplementary Information.

Protein expression and purification

Recombinant GlpG core domain for crystallography was expressed, solubilised in n-decyl-β-D-maltoside (DM, Anatrace) and purified essentially as described (Wang *et al*, 2006; Vinothkumar *et al*, 2010) with minor modifications detailed in the Supplementary Information. For purification of full-length GlpG used in inhibition

assays, n-dodecyl-β-D-maltoside (DDM, Anatrace) was used instead of DM. Imidazole from the Ni-NTA elution buffer was removed by dialysis into the rhomboid reaction buffer (50 mM Tris (pH 7.4), 100 mM NaCl, 25 mM EDTA, 10% (v/v) glycerol and 0.05% (w/v) DDM). Purification of GlpG mutants (S201A, H254A, F146I and F146A) was performed in the same way. The recombinant chimeric substrate based on TatA TMD was expressed in *glpG* knock-out *E. coli* and purified by Ni-NTA and amylose affinity chromatography as described (Strisovsky *et al*, 2009).

Rhomboid activity assays

To analyse sequence preferences of GlpG, the panel of *P. stuartii* TatA mutants in positions 4–8 (Strisovsky *et al*, 2009) was PCR-amplified and *in vitro*-transcribed and translated in the presence of radioactive [^{35}S]-L-Met as described (Strisovsky *et al*, 2009) with minor modifications detailed in the Supplementary Information. All mutant TatA variants were used at equimolar concentrations as judged by autoradiography. The substrates were exposed to purified recombinant full-length GlpG (20 ng/μl) in 16-μl reactions in a buffer containing 50 mM HEPES pH 7.4, 0.5 M NaCl, 10% (v/v) glycerol, 5 mM EDTA and 0.05% (w/v) DDM. After 40 min incubation at 37°C, the reactions were stopped by transfer on ice and addition of SDS-PAGE sample buffer. Reaction products were separated on 12% BisTris-MES SDS-PAGE (NuPAGE, Invitrogen), and substrate conversion was analysed by radiography and densitometry as described (Strisovsky *et al*, 2009) using ImageQuant 8.0 software (GE Healthcare).

For evaluating GlpG activity *in vivo*, recombinant chimeric MBP-TatAtmd-Trx substrates (Strisovsky *et al*, 2009) were expressed in the wild-type *E. coli* MC4100 encoding endogenous GlpG and in its *glpG::tet* mutant derivative at 37°C under conditions specified in the Supplementary Information, and 3 h after induction, substrate cleavage was analysed by Western blotting.

Inhibition assays

For inhibition assays, the purified MBP-TatAtmd-Trx fusion protein encompassing amino acids 1–50 of *P. stuartii* TatA (Strisovsky *et al*, 2009) was used as substrate. Purified full-length GlpG (5.4 μM) was preincubated with peptidyl-chloromethylketone inhibitors at different concentrations (50–700 μM) for 3 h at 37°C in reaction buffer containing 50 mM Tris (pH 7.4), 100 mM NaCl, 25 mM EDTA, 10 % (v/v) glycerol and 0.05 % (v/v) DDM. The cleavage reaction was started by adding substrate in fivefold molar excess over the enzyme, and let proceed for 30 min at 37°C, after which it was stopped by the addition of SDS-PAGE sample buffer and transfer on ice. Reaction products were resolved by 4–20% Tris-Glycine SDS-PAGE (Bio-Rad) and Coomassie stained (Instant Blue, Expedeon, UK). Substrate conversion was quantified densitometrically from the scanned stained gels using the ImageQuant 8.0 software (GE Healthcare).

Crystallisation and structure solution

For co-crystallisation, N-terminally truncated GlpG core domain was complexed with chloromethylketone inhibitors overnight. Excess inhibitor was then removed using desalting columns packed with

Sephadex G-25 (PD-10, GE Healthcare), and the completion of complex formation was confirmed by MALDI-MS. The complex was concentrated to 6 mg/ml, mixed with crystallisation buffer in a 1:1 ratio and crystallised by the sitting drop method at 20°C. Crystal diffraction was measured at 100 K using synchrotron radiation at BESSY (Berlin, Germany) and ESRF (Grenoble, France), and structures were solved using molecular replacement. For detailed crystallisation, freezing and measurement conditions and for details on structure solution and refinement, see Supplementary Information. Figures were generated with PyMol (Schrodinger, 2012).

Methods for plasmids and mutagenesis and modelling of the Michaelis complex are fully described in Supplementary Information.

Accession codes

The coordinates of the X-ray structures presented in this paper have been deposited with the Protein Data Bank under identifiers 4QO2, 4QO0 and 4QNZ.

Supplementary information for this article is available online: http://emboj.embopress.org

Author contributions

SZ designed, conducted and evaluated all X-ray crystallographic experiments and inhibition assays and co-wrote the paper. SS and PM designed and carried out all chemical syntheses. JS and JB characterised the specificity of GlpG and its mutants, LP contributed key reagents, and ML performed all molecular dynamics simulations. KS conceived and led the project, designed experiments and evaluated the data, and KS and SZ wrote the manuscript with input from all co-authors.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Adrain C, Zettl M, Christova Y, Taylor N, Freeman M (2012) Tumor necrosis factor signaling requires iRhom2 to promote trafficking and activation of TACE. *Science* 335: 225–228

Akiyama Y, Maegawa S (2007) Sequence features of substrates required for cleavage by GlpG, an *Escherichia coli* rhomboid protease. *Mol Microbiol* 64: 1028–1037

Arutyunova E, Panwar P, Skiba PM, Gale N, Mak MW, Lemieux MJ (2014) Allosteric regulation of rhomboid intramembrane proteolysis. *EMBO J* 33: 1869–1881

Baker RP, Young K, Feng L, Shi Y, Urban S (2007) Enzymatic analysis of a rhomboid intramembrane protease implicates transmembrane helix 5 as the lateral substrate gate. *Proc Natl Acad Sci USA* 104: 8257–8262

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

Ben-Shem A, Fass D, Bibi E (2007) Structural basis for intramembrane proteolysis by rhomboid serine proteases. *Proc Natl Acad Sci USA* 104: 462–466

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

Brandl M, Weiss MS, Jabs A, Suhnel J, Hilgenfeld R (2001) C-H..pi-interactions in proteins. *J Mol Biol* 307: 357–377

Brown MS, Ye J, Rawson RB, Goldstein JL (2000) Regulated intramembrane proteolysis: a control mechanism conserved from bacteria to humans. *Cell* 100: 391–398

Dickey SW, Baker RP, Cho S, Urban S (2013) Proteolysis inside the membrane is a rate-governed reaction not driven by substrate affinity. *Cell* 155: 1270–1281

Feng L, Yan H, Wu Z, Yan N, Wang Z, Jeffrey PD, Shi Y (2007) Structure of a Site-2 Protease Family Intramembrane Metalloprotease. *Science* 318: 1608–1612

Freeman M (2014) The rhomboid-like superfamily: molecular mechanisms and biological roles. *Annu Rev Cell Dev Biol* doi: 10.1146/annurev-cellbio-100913-012944

Greenblatt EJ, Olzmann JA, Kopito RR (2011) Derlin-1 is a rhomboid pseudoprotease required for the dislocation of mutant alpha-1 antitrypsin from the endoplasmic reticulum. *Nat Struct Mol Biol* 18: 1147–1152

Ha Y (2009) Structure and mechanism of intramembrane protease. *Semin Cell Dev Biol* 20: 240–250

Hauske P, Meltzer M, Ottmann C, Krojer T, Clausen T, Ehrmann M, Kaiser M (2009) Selectivity profiling of DegP substrates and inhibitors. *Bioorg Med Chem* 17: 2920–2924

Jahreis G, Smalla K, Fittkau S (1984) Thermitase – eine thermostabile Serinprotease. II. Synthese von substratanalogen Peptidchlormethylketonen als potentielle irreversible Enzyminhibitoren. *J Prakt Chem* 326: 41–47

Koonin EV, Makarova KS, Rogozin IB, Davidovic L, Letellier MC, Pellegrini L (2003) The rhomboids: a nearly ubiquitous family of intramembrane serine proteases that probably evolved by multiple ancient horizontal gene transfers. *Genome Biol* 4: R19

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157: 105–132

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

Lemberg MK (2011) Intramembrane proteolysis in regulated protein trafficking. *Traffic* 12: 1109–1118

Lemberg MK (2013) Sampling the membrane: function of rhomboid-family proteins. *Trends Cell Biol* 23: 210–217

Lemieux MJ, Fischer SJ, Cherney MM, Bateman KS, James MN (2007) The crystal structure of the rhomboid peptidase from Haemophilus influenzae provides insight into intramembrane proteolysis. *Proc Natl Acad Sci USA* 104: 750–754

Lewis BA, Engelman DM (1983) Bacteriorhodopsin remains dispersed in fluid phospholipid bilayers over a wide range of bilayer thicknesses. *J Mol Biol* 166: 203–210

Li X, Dang S, Yan C, Gong X, Wang J, Shi Y (2013) Structure of a presenilin family intramembrane aspartate protease. *Nature* 493: 56–61

Liu J, Liu S, Xia M, Xu S, Wang C, Bao Y, Jiang M, Wu Y, Xu T, Cao X (2013) Rhomboid domain-containing protein 3 is a negative regulator of TLR3-triggered natural killer cell activation. *Proc Natl Acad Sci USA* 110: 7814–7819

Mac Sweeney A, Birrane G, Walsh MA, O'Connell T, Malthouse JP, Higgins TM (2000) Crystal structure of delta-chymotrypsin bound to a peptidyl chloromethyl ketone inhibitor. *Acta Crystallogr D Biol Crystallogr* 56: 280–286

Malthouse JP (2007) 13C- and 1H-NMR studies of oxyanion and tetrahedral intermediate stabilization by the serine proteinases: optimizing inhibitor warhead specificity and potency by studying the inhibition of the serine proteinases by peptide-derived chloromethane and glyoxal inhibitors. *Biochem Soc Trans* 35: 566–570

Manolaridis I, Kulkarni K, Dodd RB, Ogasawara S, Zhang Z, Bineva G, O'Reilly N, Hanrahan SJ, Thompson AJ, Cronin N, Iwata S, Barford D (2013) Mechanism of farnesylated CAAX protein processing by the intramembrane protease Rce1. *Nature* 504: 301–305

Moin SM, Urban S (2012) Membrane immersion allows rhomboid proteases to achieve specificity by reading transmembrane segment dynamics. *ELife* 1: e00173.

Owen MJ, Voorheis HP (1976) Active-site-directed inhibition of the plasma-membrane carrier transporting short-chain, neutral amino acids into trypanosoma brucei. *Eur J Biochem* 62: 619–624

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

Plevin MJ, Bryce DL, Boisbouvier J (2010) Direct detection of CH/pi interactions in proteins. *Nat Chem* 2: 466–471

Rodriguez F, Rouse SL, Tait CE, Harmer J, De Riso A, Timmel CR, Sansom MS, Berks BC, Schnell JR (2013) Structural model for the protein-translocating element of the twin-arginine transport system. *Proc Natl Acad Sci USA* 110: E1092–E1101

Sampathkumar P, Mak MW, Fischer-Witholt SJ, Guigard E, Kay CM, Lemieux MJ (2012) Oligomeric state study of prokaryotic rhomboid proteases. *Biochim Biophys Acta* 1818: 3090–3097

Schrodinger LLC. (2012) The PyMOL Molecular Graphics System, Version 1.5.0.4.

Stevenson LG, Strisovsky K, Clemmer KM, Bhatt S, Freeman M, Rather PN (2007) Rhomboid protease AarA mediates quorum-sensing in Providencia stuartii by activating TatA of the twin-arginine translocase. *Proc Natl Acad Sci USA* 104: 1003–1008

Strisovsky K, Sharpe HJ, Freeman M (2009) Sequence-specific intramembrane proteolysis: identification of a recognition motif in rhomboid substrates. *Mol Cell* 36: 1048–1059

Strisovsky K (2013) Structural and mechanistic principles of intramembrane proteolysis–lessons from rhomboids. *FEBS J* 280: 1579–1603

Strisovsky K, Freeman M (2014) Sharpening rhomboid specificity by dimerisation and allostery. *EMBO J* 33: 1847–1848

Thomson A, Denniss IS (1973) The reaction of active-site inhibitors with elastase using a new assay substrate. *Eur J Biochem* 38: 1–5

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

Urban S, Wolfe MS (2005) Reconstitution of intramembrane proteolysis in vitro reveals that pure rhomboid is sufficient for catalysis and specificity. *Proc Natl Acad Sci USA* 102: 1883–1888

Van den Berg B, Clemons WM Jr, Collinson I, Modis Y, Hartmann E, Harrison SC, Rapoport TA (2004) X-ray structure of a protein-conducting channel. *Nature* 427: 36–44

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

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

Vinothkumar KR, Pierrat OA, Large JM, Freeman M (2013) Structure of rhomboid protease in complex with beta-lactam inhibitors defines the S2′ cavity. *Structure* 21: 1051–1058

Vosyka O, Vinothkumar KR, Wolf EV, Brouwer AJ, Liskamp RM, Verhelst SH (2013) Activity-based probes for rhomboid proteases discovered in a mass spectrometry-based assay. *Proc Natl Acad Sci USA* 110: 2472–2477

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

Wang Y, Maegawa S, Akiyama Y, Ha Y (2007) The role of L1 loop in the mechanism of rhomboid intramembrane protease GlpG. *J Mol Biol* 374: 1104–1113

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

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

Xue Y, Ha Y (2012) Catalytic mechanism of rhomboid protease GlpG probed by 3,4-dichloroisocoumarin and diisopropyl fluorophosphonate. *J Biol Chem* 287: 3099–3107

Xue Y, Ha Y (2013) Large lateral movement of transmembrane helix S5 is not required for substrate access to the active site of rhomboid intramembrane protease. *J Biol Chem* 288: 16645–16654

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

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

48

6.2 Sensitive versatile fluorogenic transmembrane peptide substrates for rhomboid intramembrane proteases

Background

Rhomboid-like superfamily is widespread and conserved in all domains of life. Its members participate in important biological processes like EGF receptor signaling, membrane protein quality control and mitochondrial dynamics. This makes them potential drug targets, but their potent and selective inhibitors were still missing. Unfortunately, current methods of studying rhomboid activity *in vitro* were incompatible with high-throughput screening and detailed kinetic analysis. To fill this gap, we designed a robust fluorogenic transmembrane peptide substrate platform for continuous activity assays.

Summary

We used the second transmembrane helix of the polytopic membrane protein LacY (LacYTM2) as a basis for the development of an internally quenched transmembrane fluorogenic substrate with positions P5 and P4' replaced by Glu-EDANS and Lys-DABCYL. The LacYTM2 substrate was cleaved most efficiently by four diverse rhomboid proteases from all the tested model substrates. The fluorogenic transmembrane substrates derived from LacYTM2 can be used both in detergent micelles and in liposomes (Figure 19), without losing their helical structure. The use of the EDANS-DABCYL fluorophore-quencher pair and the red-shifted fluorophores such as carboxytetramethylrhodamine (TAMRA) makes these substrates applicable in high-throughput screening. Importantly, the cleavage efficiency and selectivity of this substrate can be dramatically increased by mutations of its residues at the P5 to P1 positions, enabling design of tailor-made substrate variants for individual rhomboids.
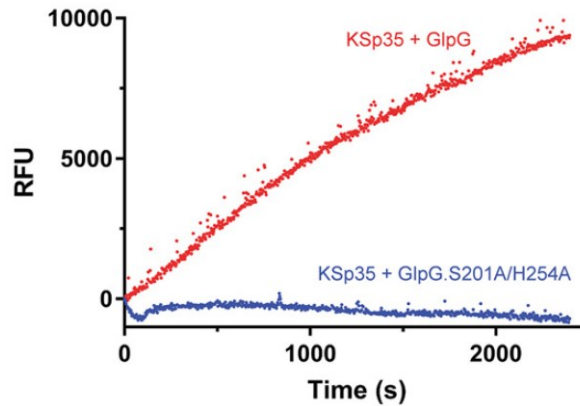
Figure 19: Activity of GlpG in liposomes detected by the KSp35 fluorogenic substrate.

My contribution

I investigated the behavior and applicability of fluorogenic substrates in liposomes and I also cloned and purified some proteins used for the identification of the best starting substrate scaffold. I analyzed the results and participated in writing of relevant parts of the manuscript.

Sensitive versatile fluorogenic transmembrane peptide substrates for rhomboid intramembrane proteases. Anežka Tichá, Stancho Stanchev, Jan Škerle, Jakub Begar, Marek Ingr, Kateřina Švehlová, Lucie Polovinkin, Martin Růžička, Lucie Bednárová, Romana Hadravová, Edita Poláchová, Petra Rampírová, Jana Březinová, Václav Kašička, Pavel Majer, and Kvido Strisovsky (2017), The Journal of Biological Chemistry. 292, 2703-2713

Sensitive Versatile Fluorogenic Transmembrane Peptide Substrates for Rhomboid Intramembrane Proteases*[S]

Anežka Tichá[‡§1], Stancho Stanchev[‡1], Jan Škerle[‡¶2], Jakub Began[‡‖3], Marek Ingr[¶**4], Kateřina Švehlová[‡], Lucie Polovinkin[‡¶5], Martin Růžička[‡¶6], Lucie Bednárová[‡], Romana Hadravová[‡], Edita Poláchová[‡], Petra Rampírová[‡], Jana Březinová[‡], Václav Kašička[‡6], Pavel Majer[‡], and Kvido Strisovsky[‡7]

From the ‡Institute of Organic Chemistry and Biochemistry of the Czech Academy of Science, Flemingovo n. 2, Prague 166 10, the ¶Department of Biochemistry, Faculty of Science, Charles University, Hlavova 2030/8, Prague 128 43, the ‖Department of Genetics and Microbiology, Faculty of Science, Charles University, Viničná 5, Prague 128 44, the §First Faculty of Medicine, Charles University, Kateřinská 32, Prague 121 08, and the **Department of Physics and Materials Engineering, Tomas Bata University in Zlín, Faculty of Technology, nám. T.G. Masaryka 5555, 76001, Zlín, Czech Republic

Edited by George N. DeMartino

Rhomboid proteases are increasingly being explored as potential drug targets, but their potent and specific inhibitors are not available, and strategies for inhibitor development are hampered by the lack of widely usable and easily modifiable *in vitro* activity assays. Here we address this bottleneck and report on the development of new fluorogenic transmembrane peptide substrates, which are cleaved by several unrelated rhomboid proteases, can be used both in detergent micelles and in liposomes, and contain red-shifted fluorophores that are suitable for high-throughput screening of compound libraries. We show that nearly the entire transmembrane domain of the substrate is important for efficient cleavage, implying that it extensively interacts with the enzyme. Importantly, we demonstrate that in the detergent micelle system, commonly used for the enzymatic analyses of intramembrane proteolysis, the cleavage rate strongly depends on detergent concentration, because the reaction proceeds only in the micelles. Furthermore, we show that the catalytic efficiency and selectivity toward a rhomboid substrate can be dramatically improved by targeted modification of the sequence of its P5 to P1 region. The fluorogenic substrates that we describe and their sequence variants should find wide use in the detection of activity and development of inhibitors of rhomboid proteases.

Rhomboid intramembrane proteases are evolutionarily widespread and regulate important biological processes including growth factor secretion (1, 2), mitochondrial dynamics (3), invasion of the malaria parasite (4), and membrane protein quality control (5). Rhomboid proteases are increasingly being explored as potential drug targets (6–9), but their selective and potent inhibitors are lacking (reviewed in Ref. 10). Rhomboid inhibitor discovery and development are complicated by the lack of widely usable and easily modifiable *in vitro* activity assays.

Rhomboid activity assays have traditionally relied on recombinant transmembrane protein substrates and gel-based read-outs, but such assays are unsuitable for high-throughput screening. A fluorogenic substrate for the *Providencia stuartii* rhomboid protease AarA lacking most of the transmembrane domain of the parent substrate Gurken is cleaved very poorly by other rhomboids including the main model rhomboid protease GlpG of *Escherichia coli* (11). Other published variants of fluorogenic substrates can be used only in liposomes (12) or involve large fluorescent protein moieties making them dependent on expression in a biological system and photochemically less variable (13), which may be important for high-throughput screening of compound libraries where bright red-shifted fluorophores are preferred (14). Moreover, each of the described rhomboid substrates has been used only with one or two related rhomboid proteases, and a strategy to design widely usable or specific substrates has been lacking. Other types of activity assays employing MALDI mass spectrometry (15) and fluorescence polarization (16) have been reported, but MALDI is a low-throughput method that requires sophisticated instrumentation, and fluorescence polarization assays are based on competition of small molecular activity probes with inhibitors and are prone to detergent artifacts (16), making both of these methods unfit for routine kinetics measurements or high-throughput screening.

In view of these limitations, we have sought to develop a robust fluorogenic transmembrane peptide substrate platform for continuous activity assays that would capture all the native enzyme-substrate interactions, be applicable to both the detergent micelle system and liposomes, and would be easily adapt-
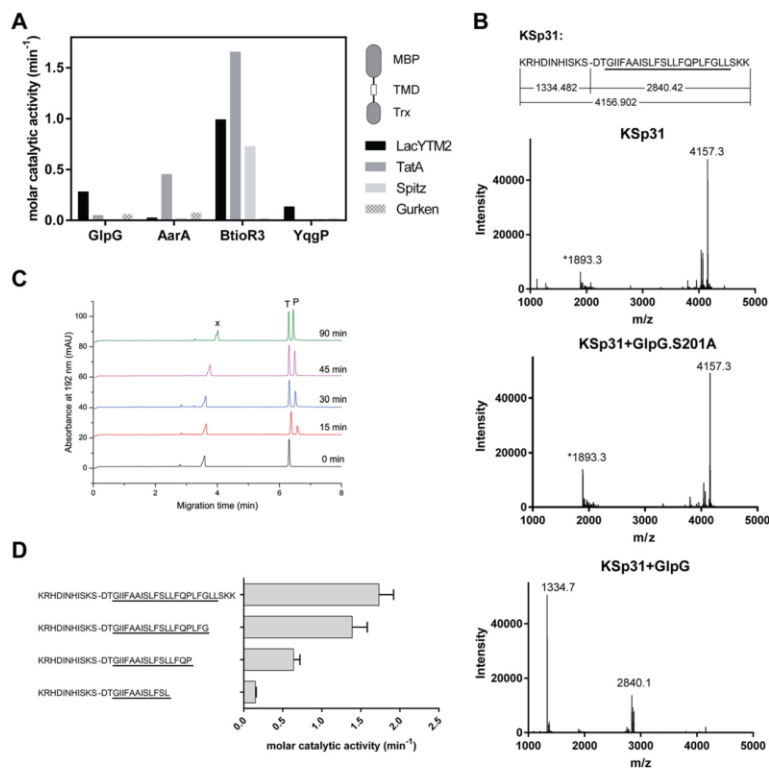
51

FIGURE 1. **Identification of a widely accepted transmembrane substrate for rhomboid proteases.** *A,* comparison of cleavage efficiency of model substrates LacYTM2, Gurken, TatA, and Spitz by bacterial rhomboid proteases GlpG (*E. coli*), AarA (*P. stuartii*), YqgP (*B. subtilis*), and BtioR3 (*B. thetaiotaomicron*) *in vitro.* Equal concentrations of purified recombinant substrates were exposed to purified recombinant rhomboid proteases. Cleavage products were separated by SDS-PAGE, stained, and quantified densitometrically to determine initial reaction rates, which were converted to molar catalytic activities to allow comparisons. Displayed values are representative of two independent experiments. *B,* cleavage of synthetic LacYTM2 transmembrane peptide KSp31 by GlpG. Purified synthetic peptide KSp31 was incubated with purified recombinant GlpG or its inactive mutant S201T in the presence of 0.05% (w/v) DDM, and the reaction mixtures were analyzed by MALDI mass spectrometry. The theoretical molecular masses of the expected cleavage products at the native cleavage site are denoted *below* the peptide sequence, and unambiguously match those experimentally determined and displayed in the mass spectra. The *star*-marked peak with molecular mass of 1893.3 is an unidentified minor contaminant in the preparation of KSp31. *C,* monitoring of cleavage of peptide substrate KSp31 by rhomboid protease GlpG using CE. The N-terminal cleavage product (P) of KSp31 was separated by free-flow CE in the background electrolyte composed of 100 mM H_3PO_4 and 69 mM Tris, pH 2.5, in bare fused silica capillary at separation voltage +25 kV. Samples for CE were prepared by mixing 20 μl of reaction mixture at selected reaction times (0–90 min) with 2 μl of 2.2 mM tyramine (*T*) as an internal standard. Samples were injected into the capillary by 20 mbar pressure for 10 s. Quantitative analysis was based on the ratio of corrected (migration time normalized) peak areas of peptides of interest and the internal standard. Analyses were performed in triplicate. *P,* cleaved N-terminal peptide; *X,* system peak; *D,* the importance of the transmembrane domain of the substrate for its recognition and cleavage by rhomboid. A series of synthetic peptides covering LacYTM2 with progressive truncations of its transmembrane domain from the C terminus was exposed to GlpG and initial rates of cleavage were quantified by capillary electrophoresis as denoted in *panel C.*

able to diverse rhomboid proteases. Because solid phase synthesis of transmembrane peptides and their purification are non-trivial, and their solution behavior often unpredictable, we place emphasis on choosing a robust system and characterizing it thoroughly, and present a generalizable framework for rhomboid substrate design.

Results and Discussion

LacYTM2 Is a Widely Accepted Rhomboid Substrate—To identify a substrate widely accepted by diverse rhomboid proteases, we have measured the efficiency of cleavage of four common model rhomboid substrate transmembrane domains (*P. stuartii* TatA, *Drosophila melanogaster* Gurken and Spitz, and *E. coli* LacYTM2) embedded in a chimeric construct by

four unrelated rhomboid proteases (*E. coli* GlpG, *Bacillus subtilis* YqgP, *P. stuartii* AarA, and *Bacteroides thetaiotaomicron* rhomboid 3 (BtioR3)) (Fig. 1*A*). Comparison of the efficiencies of cleavage (molar catalytic activities) revealed that the substrate containing the second transmembrane (TM)[8] helix of *E. coli* LacY protein (LacYTM2) (17) was the most "promiscuous" substrate.

Although it is well accepted that the region around the scissile bond, mainly P4 to P2', is key for the turnover efficiency of

---

[8] The abbreviations used are: TM, transmembrane; DDM, *n*-dodecyl-β-D-maltopyranoside; DM, *n*-decyl-β-D-maltopyranoside; CMC, critical micellar concentration; LUV, large unilamellar vesicles; MBP, maltose-binding protein; TAMRA, tetramethylrhodamine; CE, capillary electrophoresis; BGE, background electrolyte.
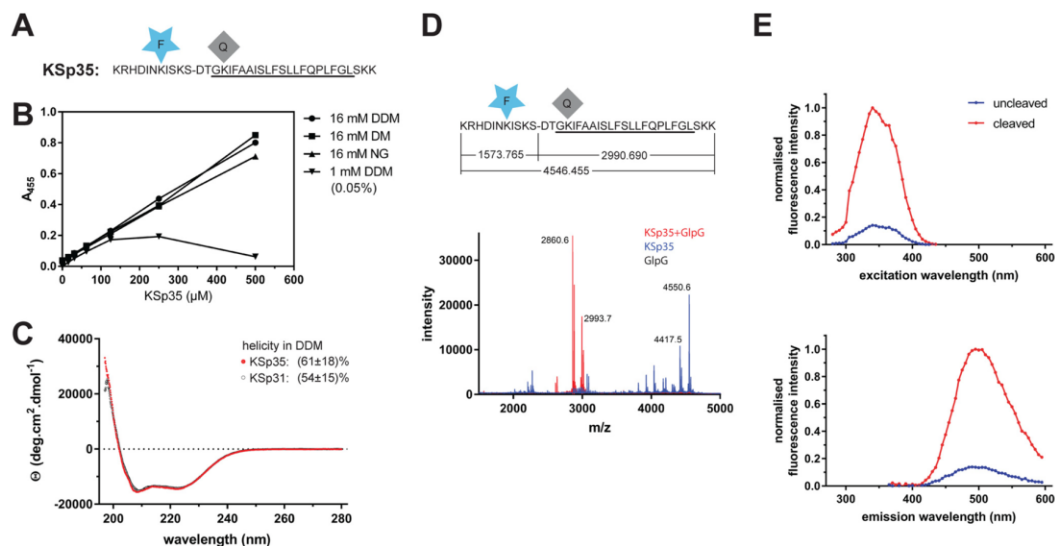
52

FIGURE 2. **Fluorogenic transmembrane peptide substrate based on LacYTM2.** *A,* fluorogenic variant of the LacYTM2 transmembrane helix-derived peptide (KSp31) with the P5 and P4' positions replaced by Glu-EDANS and Lys-DABCYL, respectively, yielding fluorogenic substrate KSp35. *B,* solubility of KSp35 in 16 mM detergents DDM, DM, and nonyl glucoside (*NG*) and at 1 mM DDM. Note that the concentration of DDM micelles is about 100 μM at 16 mM DDM and about 10 μM at 1 mM DDM. The peptide was dissolved to the indicated concentration by dilution from a 10 mM stock solution in DMSO, and after a 2-h incubation at 37 °C the solution was centrifuged at 21,130 × *g* for 20 min. The absorbance of the supernatant at 455 nm indicated the concentration of the chromophore in solution. *C,* circular dichroism spectra of LacYTM2-derived transmembrane peptide KSp31 and its fluorogenic variant KSp35 in detergent micelles. Peptides were reconstituted into 0.5% (w/v) DDM to 135 μM (KSp31) and 82 μM (KSp35) concentrations. The spectra show similarly significant helical content for both peptides. *D,* identification of the cleavage site in KSp35 by GlpG. Purified 95 μM KSp35 was incubated with 26 μM GlpG for 20 h and analyzed by MALDI. The *red peak* of the mass of 2993.7 corresponds well to the expected size of the C-terminal cleavage product of 2990.690. The second peak lower by 130 Da is visible in both the *blue* and *red traces* is probably a deletion product of chemical synthesis lacking a C-terminal lysine. This variant has proven difficult to purify away, but it is cleaved by GlpG and probably does not influence the kinetics properties of the substrate significantly (see Fig. 1*D*). *E,* excitation and emission spectra of KSp35 and their change upon cleavage by rhomboid GlpG measured in detergent micelles. The spectra of 10 μM KSp35 substrate in reaction buffer (20 mM HEPES, pH 7.4, 150 mM NaCl, 0.05% (w/v) DDM, 10% (v/v) DMSO) were measured at 37 °C. Excitation wavelengths ranged from 235 to 435 nm with a 10-nm increment and the emission was measured at 493 nm. The emission wavelengths ranged from 365 to 595 nm with a 10-nm increment and excitation at 335 nm.

rhomboid substrates (12, 18), the role of the TM domain of the substrate for recognition and catalysis by rhomboid is less well understood. We have thus next evaluated the importance of the transmembrane region of LacYTM2 for the recognition by *E. coli* GlpG, the main model rhomboid protease, by synthesizing a peptide covering the whole transmembrane region and adjacent juxtamembrane segments of LacYTM2, and a series of its C terminally truncated variants. The full-length LacYTM2 transmembrane peptide KSp31 was cleaved by GlpG efficiently and highly specifically at the expected Ser-Asp cleavage site (Fig. 1*B*). The kinetics of cleavage were monitored by capillary electrophoresis (Fig. 1*C*). The cleavage rate decreased significantly upon truncating the TM helix of LacYTM2 peptide by more than 5 amino acids from the C terminus (Fig. 1*D*), suggesting that most of the TM domain of the substrate is important for the interaction with and recognition by rhomboid. Thus, to develop a widely accepted fluorogenic substrate that would faithfully mimic all the relevant enzyme-substrate interactions including the intramembrane ones, we have used the full-length LacYTM2 transmembrane domain peptide KSp31 as a starting point.

Fluorogenic Transmembrane Peptide Substrate Based on LacYTM2, Basic Properties—To generate a fluorogenic variant of the LacYTM2 peptide, we have replaced the P5 and P4' positions in KSp31 by Glu-EDANS and Lys-DABCYL to yield KSp35 (Fig. 2*A*). Previously published mutagenic analyses show that these positions are not critical for recognition by rhomboid (18, 19), and they are sufficiently close for Förster resonance energy transfer (FRET) to occur. The KSp35 peptide was soluble up to 500 μM (Fig. 2*B*) in frequently used detergents at 16 mM decyl maltoside (DM), nonyl glucoside (NG), and dodecyl maltoside (DDM). At a total DDM concentration of 16 mM (0.82%(w/v)), the concentration of micelles is about 110 μM, suggesting a partitioning ratio of more than 1 molecule of the substrate per micelle. When DDM was kept at only 1 mM (0.05% (w/v)) total concentration, which yields about 6–10 μM micelles, the solubility of KSp35 became limited to about 100 μM (Fig. 2*B*), indicating that the upper limit of the partitioning ratio is about 10–20 molecules of KSp35 per DDM micelle. The solubility of KSp35 in the absence of detergent was negligible (not shown). Circular dichroism of KSp35 in 0.5% (w/v) DDM (Fig. 2*C*) showed a significant content of α-helical structure (61 ± 18%), which is consistent with the transmembrane character of the peptide and comparable with the helical content of the parent peptide KSp31 (54 ± 15%). Cleavage of KSp35 by GlpG occurred at the expected cleavage site (Fig. 2*D*), and was accompanied by an increase in fluorescence at 495 nm (Fig. 2*E*), demonstrating that FRET between the donor and acceptor is occurring in the uncleaved peptide. Collectively, the above results show that KSp35 is a realistic model reflecting all the

53

important interactions between a rhomboid protease and its transmembrane substrate.

Kinetic Characterization of the LacYTM2-based Substrate KSp35 in Detergent Micelle System—In the detergent-solubilized state, most commonly used to study the biochemistry of intramembrane proteolysis, the reaction catalyzed by rhomboid protease occurs in detergent micelles due to the hydrophobicity of both enzyme and substrate. The system is thus microheterogeneous, the effective concentrations of the reactants depend on the volume of the micellar milieu and on the partitioning of reaction components between free solution and the micelles. To characterize the kinetic behavior of the new fluorogenic transmembrane substrates in light of these features of the micellar system, steady-state kinetics was measured with 10 μM substrate, 0.4 μM enzyme, and 0.05% (w/v) DDM, always keeping the concentrations of two components constant and varying the third one around the stated values. At 0.05% (w/v) DDM, the concentration of detergent monomers is 980 μM and micelle concentration about 6–10 μM, calculated assuming critical micellar concentration (CMC) of 0.17 mM (20) and aggregation number between 78 and 149 (20). The molar ratio of enzyme:substrate:micelles is thus 4:100:60–100. In these conditions, assuming that all the reaction partners are evenly distributed among micelles, the average number of substrate molecules per micelle is about 1.5, and only up to 4% of micelles carry an enzyme molecule (micelles containing more than one enzyme molecule are strongly improbable).

The cleavage reactions were started by either mixing two preheated solutions containing substrate or enzyme preincubated with detergent, or adding the DMSO-dissolved substrate into the rest of the preheated reaction mixture. In either case, progress curves are linear from the beginning, which implies that the redistribution of the adsorbed molecules among the micelles is significantly faster than substrate cleavage itself. In accordance with this, the reaction rate is proportional to enzyme concentration within the 0–0.6 μM range (Fig. 3A). Within this concentration range, few enzyme molecules are randomly distributed among many more micelles, providing in principle equal conditions for each enzyme molecule. A similar principle can also explain the observation that the dependence of the reaction rate on substrate concentration is linear in the 0–4 μM range (Fig. 3B). At the upper limit of 4 μM substrate, all micelles can be populated by one (or less likely more) substrate molecule, the linear dependence, furthermore, suggests that this substrate concentration is still below the apparent Michaelis constant of this process.

An important phenomenon is observed when the dependence of the initial rate on detergent concentration is measured. At concentrations above the CMC, the reaction rate rapidly decreases as DDM concentration grows (Fig. 3C), without an obvious impact on the secondary structure content of GlpG (Fig. 3D), suggesting that the effect is caused primarily by the increase in the volume of the micellar phase and consequent decrease of the effective concentrations of both substrate and enzyme. Indeed, mathematical consideration suggests that when substrate and enzyme concentrations are significantly lower than the concentration of micelles (*i.e.* at high DDM concentrations), the probability of location of a substrate molecule

on the same micelle as the enzyme molecule is inversely proportional to the concentration of DDM. Under these conditions, the fraction of substrate-occupied micelles, f_{SM}, is equal to the ratio of the numbers of substrate molecules, $n(S)$, and micelles $n(M)$.

$$f_{SM} = n(S)/n(M) \qquad \text{(Eq. 1)}$$

The mean number of micelles occupied by both the enzyme and substrate molecules, $n(ESM)$, is then given by this fraction multiplied by the number of enzyme molecules $n(E)$.

$$n(ESM) = f_{SM} \times n(E) = n(S) \times n(E)/n(M) \qquad \text{(Eq. 2)}$$

Hence, when the DDM concentration is increased at constant $n(S)$ and $n(E)$, then $n(ESM)$ reflecting the reaction rate decreases in accord with the growing value of $n(M)$. This causes the proportional decrease of the reaction rate (in other words, the reaction rate is proportional to $[DDM]^{-1}$). To inspect whether this model is correct, one can conveniently determine the power of the measured rate dependence on DDM concentration by taking a logarithm of the data from Fig. 3C (log $a^n = n \times \log a$). The logarithmic plot (Fig. 3C, *open circles, right* and *upper axes*) can be satisfyingly ($R^2 = 0.9974$) fitted by a second-order polynomial, yielding equation: $y = -0.1436x^2 - 0.3906x + 2.8852$, whose derivative $y' = -0.2872x - 0.3906$ indicates the power of DDM concentration on which the reaction rate depends. This analysis shows that for high DDM concentrations the derivative indeed tends to -1 (for $x = 2$, $y' = -0.965$; thus rate $\sim [DDM]^{-1}$), which is in accordance with the above assumption, whereas for the lower end of DDM concentrations the absolute value of the power decreases (for $x = 0$, $y' = -0.3906$; thus rate $\sim [DDM]^{-0.4}$). This is consistent with a model that upon decreasing the detergent concentration (while still being above the CMC), the density of the adsorbed molecules in the micellar phase increases, whereas total concentration of micelles decreases, which leads to less frequent collisions between them and thus less effective redistribution of the adsorbed molecules among the micelles. Possibly, the redistribution efficiency might also be insufficient because of the higher reaction rate caused by the higher reactant concentrations.

Although the reaction kinetics of intramembrane proteases in liposomes has been described in terms of interfacial kinetics (12, 21), that is, expressing the kinetic constants in relationship to the volume or molar fraction of the lipidic phase, (22, 23), the kinetic effects related to the reaction occurring in detergent micelles have surprisingly not yet been considered in enzyme kinetics studies on rhomboid proteases (12, 13) nor other intramembrane proteases, yet they are evidently important for the interpretation of kinetics measurements. Our data show that for reliable and meaningful measurement of apparent Michaelis-Menten kinetics parameters, the micelle concentration must not be limiting the solubility of the substrate, and the detergent concentration must be kept constant. The latter point also means that having a stock solution of the substrate dissolved in detergent (at a higher concentration than intended in the reaction mixture, which frequently can occur during purification and concentration) may lead to underestimation of reaction rates at high substrate concentrations due to a possibly
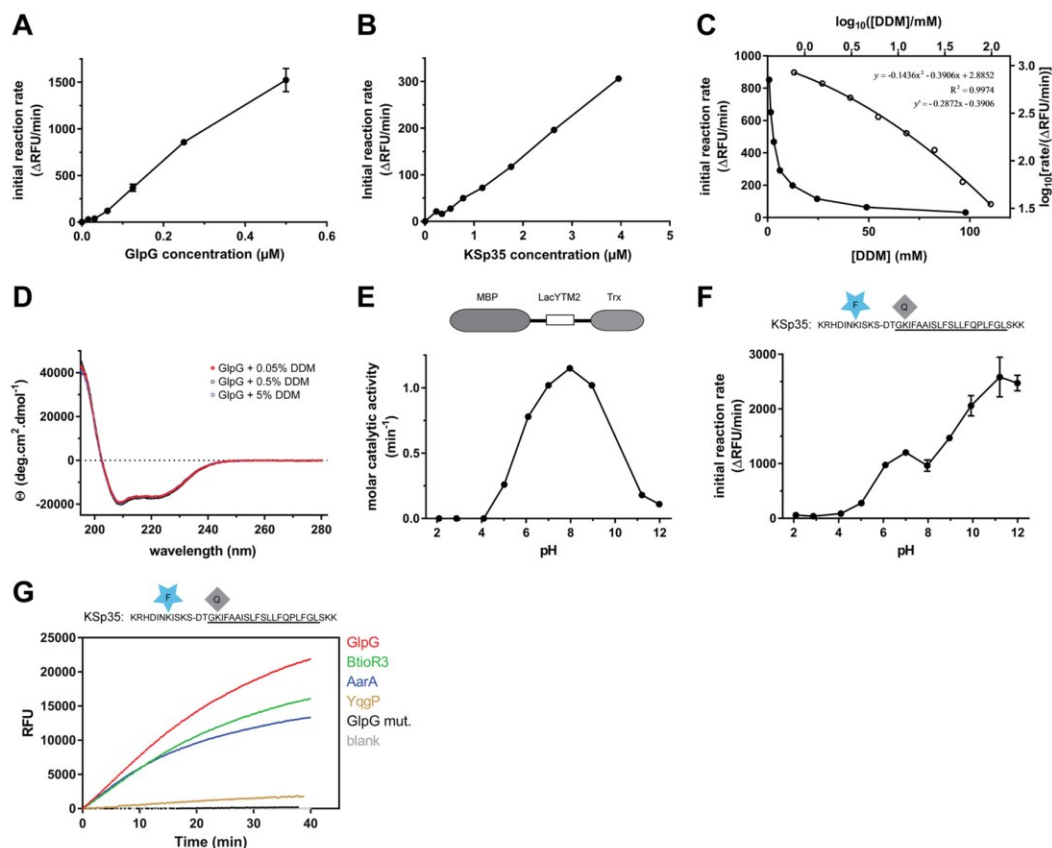
54

FIGURE 3. Kinetic characterization of fluorogenic transmembrane peptide substrate KSp35 in the detergent micelle system. *A*, dependence of the initial reaction rate on enzyme concentration. The fluorogenic substrate KSp35 (10 μM) was incubated with varying concentrations of GlpG in a reaction buffer composed of 20 mM HEPES, pH 7.4, 150 mM NaCl, 0.05% (w/v) DDM, and 10% (v/v) DMSO, and initial reaction rates were measured by following fluorescence at 493 nm. The displayed values are means from duplicate measurements with 2 $\times$ S.D. *B*, dependence of the initial reaction rate on substrate concentration. The rhomboid protease GlpG (0.4 μM) was incubated with varying concentrations of the fluorogenic substrate KSp35 in a reaction buffer composed of 20 mM HEPES, pH 7.4, 150 mM NaCl, 0.05% (w/v) DDM, 10% (v/v) DMSO, and the initial reaction rates were measured by following fluorescence at 493 nm. Representative values from one of three independent experiments are shown. *C*, dependence of the initial reaction rate on detergent concentration (*solid circles, left* and *lower axes*). The fluorogenic substrate KSp35 (10 μM) was incubated with 0.4 μM GlpG at varying concentrations of DDM in a reaction buffer composed of 20 mM HEPES, pH 7.4, 150 mM NaCl, 10% (v/v) DMSO, and initial reaction rates were measured by following fluorescence at 493 nm. Representative values from one of three independent experiments are shown. The *open circles* (*right* and *upper axes*) represent the same plot at the logarithmic scale. When this plot is fitted by second-order polynomial, the equation $y = -0.1436x^2 - 0.3906x + 2.8852$ is obtained, the derivative of which, $y' = -0.2872x - 0.3906$, is equal to the power of DDM concentration with which the reaction rate decreases. For high DDM concentrations the derivative tends to -1 (for $x = 2$, $y' = -0.965$), whereas for lower DDM concentrations the absolute value of the power decreases (for $x = 0$, $y' = -0.3906$). *D*, overall secondary structure of GlpG is not affected by high concentrations of DDM. CD spectra of GlpG at 0.05, 0.5, and 5% (w/v) (98 mM) DDM were recorded and show no variation in the secondary structure content of GlpG depending on DDM concentration. *E*, the pH dependence of GlpG activity on the LacYTM2-derived chimeric substrate MBP-LacYTM2-Trx. The substrate (2 μM) was incubated with 0.1 μM GlpG in a broad pH range buffer (38) composed of 40 mM H$_3$PO$_4$, 40 mM CH$_3$COOH, and 40 mM H$_3$BO$_3$ adjusted to pH values between 2 and 12, and initial reaction rates were measured by SDS-PAGE and densitometry as described under "Experimental Procedures." *F*, the pH dependence of cleavage of the fluorogenic LacYTM2-derived substrate KSp35 by GlpG. The substrate (10 μM) was incubated with 0.4 μM GlpG in a broad pH range buffer (38) composed of 40 mM H$_3$PO$_4$, 40 mM CH$_3$COOH, and 40 mM H$_3$BO$_3$ adjusted to pH values between 2 and 12, and initial reaction rates were measured by recording fluorescence at 493 nm. *G*, selectivity of the fluorogenic substrate KSp35 for diverse bacterial rhomboid proteases. The purified recombinant rhomboid proteases GlpG, AarA, YqgP (all at 0.4 μM), and BtioR3 (at 0.04 μM) were incubated with 10 μM KSp35 in a reaction buffer composed of 20 mM HEPES, pH 7.4, 150 mM NaCl, 0.05% (w/v) DDM, and 10% (v/v) DMSO, and progress curves were measured by recording the increase in fluorescence at 493 nm.

significant increase of detergent concentration in the final reaction mixture, as shown in Fig. 3*C*. This could result in pseudo-Michaelis kinetics and yield falsely low K_m values. Practical implications are that 1) exact detergent concentrations must be known in any kinetics measurements, and 2) it is advantageous to have the substrate stock solution dissolved in a detergent-free medium or at a detergent concentration lower or equal to

that used in the final assay buffer. The transmembrane substrates presented in this article, generated by chemical synthesis, are in principle avoiding this problem, because their stock solutions are detergent-free dissolved in anhydrous dimethyl sulfoxide. Alternatively, they can be reconstituted into a detergent of choice via disaggregation in hexafluoroisopropanol, as described by Deber *et al.* (24).
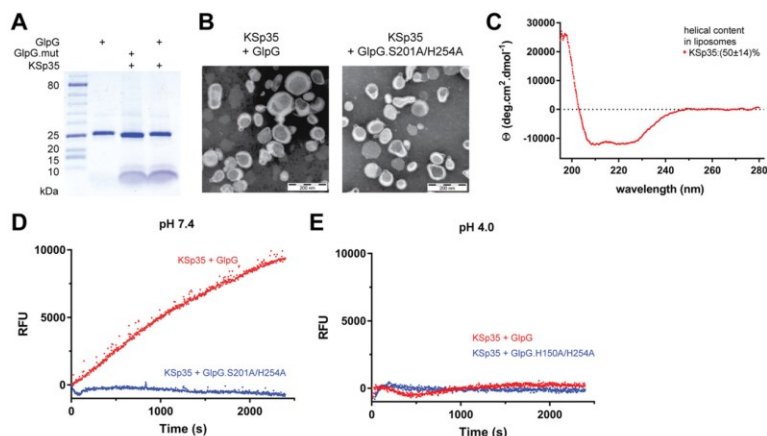
55

FIGURE 4. **The use of the transmembrane peptide substrate in liposomes.** *A,* KSp35 was reconstituted into liposomes (LUVs) formed from *E. coli* polar lipid extract in the presence of GlpG or its inactive mutant S201A at pH 4.0. The resulting large unilamellar vesicles were analyzed by SDS-PAGE. *B,* the shape, lamellarity, and approximate size distribution of the KSp35+GlpG containing proteoliposomes formed at pH 4.0 were characterized by transmission electron microscopy. *C,* the integration of KSp35 into liposomes and its secondary structure content were analyzed by electronic CD. The substrate KSp35 (3 μM) was reconstituted with 2 mg/ml of *E. coli* polar lipid extract yielding an approximate peptide:lipid weight ratio of 1:500. *D,* activity of GlpG in liposomes detected by the KSp35 fluorogenic substrate. The substrate was co-reconstituted with wild type GlpG or its S201A/H254A mutant in a 30:1 molar ratio into LUVs made of *E. coli* polar lipid extract at pH 4.0, proteoliposomes were collected by ultracentrifugation and resuspended in 10 mM HEPES, 150 mM NaCl, pH 7.4, to start the cleavage reaction, which was then followed by measuring fluorescence at 493 nm. *E,* wild type GlpG or its H150A/H254A mutant were co-reconstituted with the substrate KSp35 in a 30:1 molar ratio into LUVs made of *E. coli* polar lipid extract at pH 4.0, proteoliposomes were collected by ultracentrifugation, resuspended in 50 mM sodium acetate, 150 mM NaCl, pH 4.0, and fluorescence was followed at 493 nm.

The pH dependence of cleavage rate of the unmodified LacYTM2 transmembrane segment in the context of an MBP-thioredoxin fusion protein shows a relatively broad maximum around pH 9, with substantial activity of GlpG between pH 6 and 11 and negligible activity below pH 4 and at pH 12 (Fig. 3*E*), which is largely in agreement with previous studies (12, 13). The dependence of the cleavage rate of KSp35 on pH also shows that GlpG is completely inactive at pH values below and up to 4, but the initial reaction rate of KSp35 cleavage then appears to grow up to pH 12 (Fig. 3*F*). This effect cannot be ascribed to the pH-dependent change of EDANS fluorescence (data not shown), and could possibly be due to effects of pH on the conformational dynamics of KSp35. However, this is not a concern because in most cases measurements are performed at a physiologicaly relevant pH near neutral. The apparent catalytic efficiency k_{cat}/K_m of GlpG against KSp35 measured at pH 7.4 and 0.05%(w/v) DDM is $(2.0 \pm 0.5) \times 10^{-3}$ min^{-1} μM^{-1}, which is comparable with the values reported for the TatA substrate by Dickey *et al.* (12) and Arutyunova *et al.* (13) obtained in similar conditions. Importantly, the LacYTM2-derived fluorogenic peptide substrate KSp35 is cleaved efficiently by unrelated recombinantly purified bacterial rhomboids GlpG, AarA, and BtioR3, and modestly by YqgP (Fig. 3*G*), which demonstrates its wide usability, surpassing any other currently available rhomboid substrate.

Use of the Transmembrane Peptide Substrate in Liposomes—Because the natural environment of rhomboid proteases is the lipid membrane, we next tested whether the fluorogenic peptide substrate KSp35 can also be used in liposomes. We co-reconstituted KSp35 with GlpG or its inactive mutant S201A/H254A at pH 4 into large unilamellar vesicles (LUVs) formed from *E. coli* polar lipid extract, and confirmed the composition of the resulting proteoliposomes by SDS-PAGE (Fig. 4*A*). Neg-ative stain transmission electron microscopy showed that both empty LUVs and proteoliposomes containing KSp35 in the presence or absence of GlpG or its inactive mutant S201A/H254A had similar morphology and size distribution both at pH 7 and 4 (Fig. 4*B*). The CD spectrum of LUV-reconstituted KSp35 showed helicity of 50 ± 14% (Fig. 4*C*), which is consistent with its transmembrane helix prediction. GlpG is inactive at pH 4 (Fig. 3, *E* and *F*), and, consistently, fluorescence of proteoliposomes containing KSp35 and GlpG at pH 4 was at a constant background level (Fig. 4*E*). Upon neutralization to pH 7.4, time-dependent increase of fluorescence at 495 nm was observed in the presence of wild type GlpG but not in the presence of its active-site mutant S201A/H254A (Fig. 4*D*). These results collectively demonstrate that the LacYTM2-based fluorogenic transmembrane substrate KSp35 is widely usable both in detergent micelles or liposomes and with diverse rhomboid proteases.

A Red-shifted Variant of the Fluorogenic Transmembrane Substrate for Rhomboids—Large compound libraries for high-throughput screening can often contain compounds that absorb in the UV region (14), and fluorogenic substrates operating at red-shifted wavelengths are less affected by such compound interference. Because EDANS is excited in the UV region, and is thus prone to interference in library screening, we have modified the LacYTM2 peptide backbone by instead attaching the red-shifted TAMRA fluorophore to a Lys introduced into the P5 position and a compatible dark quencher QXL610 to a Cys introduced into the P4' position (Fig. 5*A*) to yield KSp76. This red-shifted fluorogenic substrate is cleaved by several bacterial rhomboid proteases with efficiencies similar to its UV variant KSp35. The apparent catalytic efficiency k_{cat}/K_m of GlpG cleaving KSp76 is $(1.6 \pm 0.5) \times 10^{-3}$ min^{-1}
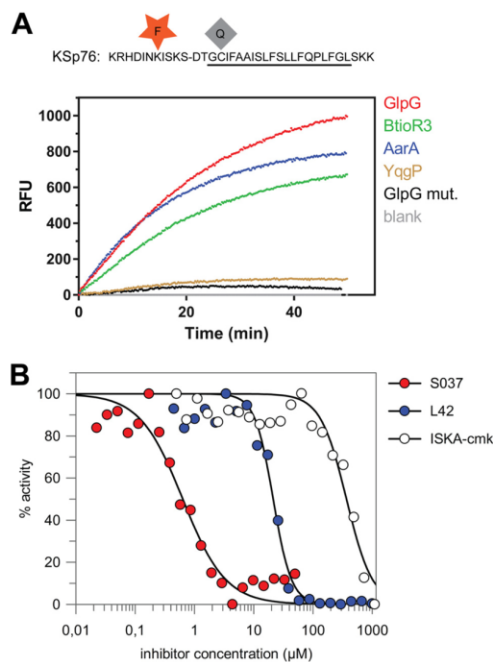
56

FIGURE 5. **Red-shifted variant of the LacYTM2-based fluorogenic substrate.** *A,* modification of Lys in the P5 position of KSp31 by the red-shifted TAMRA fluorophore and P4′ Cys by a dark quencher QXL610 yields highly fluorogenic substrate KSp76 that is efficiently cleaved by rhomboid proteases GlpG, AarA, YqgP, and BtioR3 at identical concentrations to those used in Fig. 3G. Excitation wavelength was 553 nm, and emission was followed at 583 nm. *B,* the red-shifted fluorogenic substrate KSp76 allows measurement of inhibition by compounds that absorb in the UV region, such as isocoumarin, and is thus suitable for high-throughput screening. The dose-response curves of the chloromethylketone ISKAcmk, β-lactam L42, and isocoumarin S037 were measured after a 60-min preincubation of enzyme with inhibitor. The curves were fitted in GraFit 7 to yield apparent IC$_{50}$ values.

μM^{-1}, which is similar to the EDANS variant KSp35 ((2.0 $\pm$ 0.5) $\times$ 10^{-3} min^{-1} μM^{-1}) under identical reaction conditions within experimental error (Fig. 6C). The utility of this red-shifted variant of the LacYTM2 substrate is demonstrated by measuring the inhibition curves of chloromethylketone ISKA-cmk (19), β-lactam L42 (11), and isocoumarin S037 (25, 26). Using a 60-min enzyme + inhibitor preincubation time, the measurements yielded apparent IC$_{50}$ values of 370 $\pm$ 38, 12.4 $\pm$ 1.6, and 0.64 $\pm$ 0.08 μM, respectively (Fig. 5B), which are largely in agreement with published values measured in other assay systems and otherwise comparable conditions (11, 15, 19).

Efficiency and Selectivity of the Substrates Can Be Tuned by Varying Their Non-prime Side Amino Acid Sequence—One of the problems with current rhomboid protease assays is that there has been little rationale about how to modify the substrates to improve their kinetic properties and adapt them for different rhomboid proteases. Recent enzymatic analyses (12, 18) have shown that the region between the P4 and P2′ residues determines the k_{cat} of the cleavage reaction, suggesting that selective substrates for rhomboids could be designed by modi-

fying the P4 to P2′ region appropriately. A recent mutagenic study of the TatA substrate and structural analysis of a derived rhomboid·substrate-peptide complex revealed amino acids at the P5 to P1 positions of TatA that are preferred by GlpG (19). We tested the impact of these substitutions in the context of the LacYTM2 substrate.

Although single mutations of the P5 amino acid to the preferred Arg, P4 amino acid to Val, and P2 amino acid to His did not improve the cleavage of the purified recombinant MBP-LacYTM2-Trx substrate *in vitro*, mutation of the P1 amino acid to Ala improved the cleavage of mutant 7-fold, and mutation of the P3 residue to Arg improved the cleavage of mutant 16-fold (Fig. 6A). Combining all five mutations yielded a mutant substrate (RVRHA) that was cleaved 64-fold better than the wild type substrate (Fig. 6A), which shows that the effects of the preferred substitutions are additive. When analyzed for cleavage *in vivo*, it turns out that already the wild type MBP-LacYTM2-Trx substrate is such a good substrate of GlpG that it is turned over from 94% (Fig. 6B). The effects of the preferred P5 to P1 mutations thus cannot be assessed in this context as they all exhibit similarly high steady-state turnover (Fig. 6B).

To test this effect in our fluorogenic substrates, we have modified the TAMRA-based LacYTM2-derived fluorogenic substrate by changing the P5 to P1 segment from HISKS to RVRHA to yield KSp64, and compared the kinetic properties of both substrates. The analysis revealed that catalytic efficiency k_{cat}/K_m of GlpG cleaving KSp64 is (3.7 $\pm$ 0.4) $\times$ 10^{-2} min^{-1} μM^{-1}, which is 23-fold higher than that of the original red-shifted LacYTM2 substrate KSp76 ((1.6 $\pm$ 0.5) $\times$ 10^{-3} min^{-1} μM^{-1}) (Fig. 6C). The impact of the modifications of the P5 to P1 region on selectivity against other bacterial rhomboid proteases is particularly striking (Fig. 6D), with the initial reaction rate of KSp64 cleavage by GlpG being about 50-fold higher than that of AarA (measured from data displayed in Fig. 6D) and even higher for the other tested rhomboid proteases, revealing a straightforward strategy for designing selective rhomboid substrates.

In summary, we report novel sensitive versatile fluorogenic transmembrane peptide substrates for rhomboid intramembrane proteases that are usable both in detergent micelles and liposomes, are cleaved by diverse rhomboid proteases, and contain a red-shifted fluorophore suitable for high-throughput screening assays. Furthermore, we provide a strategy how to adapt these substrates to individual rhomboid proteases by modifying their P5 to P1 residues, and we demonstrate that controlling the detergent concentration is important for obtaining accurate kinetic data. We expect that the substrates we describe and sequence variants thereof will enable facile detection of activity and development of inhibitors of rhomboid proteases.

Experimental Procedures

General Biochemicals—Lipids were from Avanti Polar Lipids, detergents from Anatrace, buffers and other biochemicals were from Sigma or other suppliers as specified below.

DNA Constructs and Cloning—The expression constructs for rhomboid proteases GlpG, YqgP, and AarA and chimeric MBP-TMD-Trx substrate constructs where TMD = LacYTM2, Gur-
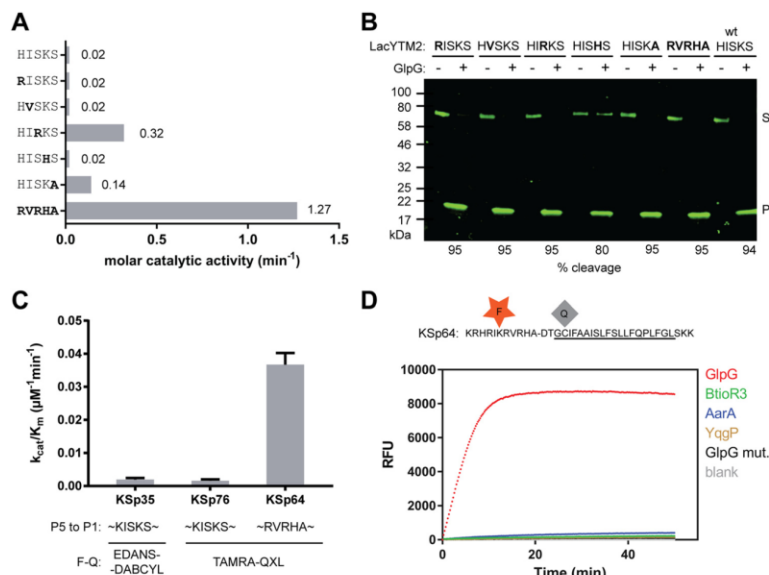
57

FIGURE 6. **The effect of non-prime side substitutions on the catalytic parameters and selectivity of rhomboid substrates.** *A,* preferred amino acids in the P5 to P1 positions of the LacYTM2 transmembrane substrate improve its cleavage by GlpG. The LacYTM2 embedded in the MBP-thioredoxin chimera (18) was point-mutated in the P5 to P1 positions according to the sequence preferences of *E. coli* GlpG (19). The recombinant substrates were expressed in *E. coli* ΔglpG, purified, and molar catalytic activity of GlpG in cleaving each of the substrates was determined using gel-based assay (see "Experimental Procedures" for details). The concentration of substrate was always 1.47 μM, concentration of DDM was 0.5%(w/v), the concentration of GlpG was 0.8 μM for wild type substrate (HISKS). and for the RISKS, HVSKS, and HISHS mutants the concentration was 0.08 μM for the HISKA mutant and 0.016 μM for the HIRKS and RVRHA variants (to ensure reliable measurement of the initial reaction rate). Representative values from one of three independent experiments are shown. *B,* the effects of the preferred amino acids in the P5 to P1 region of LacYTM2 on the steady-state level of cleavage by GlpG in biological membranes *in vivo.* Plasmids encoding individual mutant versions of the chimeric mutant LacYTM2 substrates described above were transformed into *E. coli* MC4100 expressing endogenous GlpG, and 2 h after induction of expression of the substrates, the cell lysates were analyzed by immunoblotting using antibody against His tag, located at the C terminus of the constructs. Detection by near-infrared laser scanning, exhibiting linearity over 6 orders of magnitude, enabled reliable quantitation. Integration of product and substrate band intensities yielded steady-state substrate conversion values that are listed *below* the image. A representative experiment is displayed. *C,* apparent kinetic parameters of fluorogenic rhomboid substrates derived from LacYTM2. Initial reaction rates at very low substrate concentrations were used to calculate catalytic efficiency values (k_{cat}/K_m) of substrates KSp35, KSp64, and Ksp76 cleaved by GlpG at 0.5% (w/v) DDM. The reaction buffer was 20 mM HEPES, pH 7.4, 150 mM NaCl, 10% (v/v) DMSO, enzyme concentration was 0.4 μM, and substrate concentration ranged from 0.5 to 20 μM. Note that a mere optimization of the P5 to P1 region of the substrate increases the catalytic efficiency (k_{cat}/K_m) of its cleavage by GlpG by 23-fold. *D,* influence of the optimization of the P5 to P1 region on the selectivity of a transmembrane substrate for rhomboids. KSp76 underwent cleavage by rhomboid proteases GlpG, AarA, YqgP, and BtioR3 at the same concentrations as described in the legends to Figs. 3*G* and 5*A*. Note that optimization of the P5 to P1 region of the substrate increases the selectivity for GlpG dramatically.

ken, TatA, or Spitz as described previously (18). The expression construct for rhomboid protease BtioR3 was generated by PCR amplification of the entire ORF encoding the Q8A3X2 (Uniprot ID) protein from *B. thetaiotaomicron* genomic DNA (purchased from ATCC), and its cloning as a C terminally His-tagged construct into pET25b+M as described previously (27). Mutations of the TatA and LacYTM2 recognition motif in the MBP-TMD-Trx construct were generated by overlap assembly PCR (28) and isothermal assembly (29). All constructs were verified by DNA sequencing.

Chemical Synthesis—All reagents were acquired from commercial sources and used without purification. Protected amino acids and amino acid derivatives were purchased from Iris Biotech (Marktredwitz, Germany). Trimellitic anhydride and 3-dimethylaminophenol were from Sigma, QXL610 vinylsulfone was from AnaSpec (Fremont, CA), and *N*-(9-fluorenyl) methoxycarbonyl (Fmoc)-Glu(EDANS)-OH from Merck KGaA (Darmstadt, Germany). The detailed synthetic procedures, analytical methods, and compound characterization data are included in the supporting information.

Protein Expression and Purification—Bacterial rhomboid proteases AarA, GlpG, BtioR3, and YqgP and the active site mutant GlpG.S201A were overexpressed in *E. coli* C41(DE3) (30) as full-length, C terminally His-tagged proteins from a modified pET25b+ vector (27). The cultures were grown at 37 °C in LB medium to A_{600} of 0.4 and induced by 1 mM isopropyl 1-thio-β-D-galactopyranoside. The expression was continued overnight at 20 °C. Cells were harvested, resuspended in buffer A (25 mM HEPES, pH 7.4, 100 mM NaCl, 10%(v/v) glycerol, 1 mM PMSF), and lysed by 2 to 3 passes through Avestin EmulsiFlex-C3. Cell debris was removed by a low-speed centrifugation. Cellular membranes were isolated by a 2-h centrifugation at 100,000 × *g* and were solubilized in 1.5%(w/v) DDM (solubilization grade, Anatrace) in Buffer B (25 mM HEPES, pH 7.4, 300 mM NaCl, 10%(v/v) glycerol, 10 mM imidazole, EDTA-free Complete Protease Inhibitor mixture (Roche Applied Science)) at room temperature for 1 h. Solubilized proteins were isolated by centrifugation at 100,000 × *g* for 30 min and loaded onto nickel-nitrilotriacetic acid HiTrap IMAC HP 1-ml columns (GE Healthcare). Nonspecifically bound proteins were

58

washed off with Buffer C (25 mM HEPES, pH 7.4, 300 mM NaCl, 10% (v/v) glycerol, 0.05% (w/v) DDM) containing 10, 50, and 125 mM imidazole. The protein was eluted with Buffer C containing 250 to 500 mM imidazole. The peak fractions were buffer exchanged into 25 mM HEPES, pH 7.4, 150 mM NaCl, 10% (v/v) glycerol, and 0.05% (w/v) DDM on a HiPrep 26/10 desalting column (GE Healthcare). If needed, proteins were concentrated using Vivaspin ultrafiltration spin cells with 30-kDa MWCO. Protein concentration was determined from absorbance at 280 nm, and the final concentration of DDM was determined as described (31).

Capillary Electrophoresis (CE)—Analyses of standard peptides and enzymatically cleaved peptide substrates were performed on an Agilent CE 7100 instrument (Agilent, Waldbronn, Germany) equipped with photodiode array UV-visible detector operating in the 190–600 nm range. Electropherograms were acquired at 192, 205, and 214 nm and absorbance data at 192 nm were selected for quantitative evaluation due to the highest signal to noise ratio. CE analyses were carried out in a bare fused silica capillary with polyimide outer coating (internal diameter 50 μm, outer diameter 375 μm, effective length to the detector 40 cm, total length 48.5 cm, supplied by Polymicro Technologies, Phoenix, AZ). Peptides were analyzed as cations in acidic background electrolyte (BGE) composed of 100 mM H_3PO_4, 69 mM Tris, pH 2.5. For highly hydrophobic peptides, this BGE was modified by the addition of 0.05% (w/v) DDM. The temperature of the air-cooled capillary was set to 20 °C and the sample carousel was kept at the same temperature using a circulating water bath. Prior to each CE run, the capillary was successively washed with 100 mM sodium dodecyl sulfate, ethanol, 1 M NaOH, water, 1 M HCl, and the BGE, to remove any possible carryover of hydrophobic peptides and detergents from the previous run. All washes were done at 8 bar pressure for 30 s. Peptide standards used for identification of cleavage products were solubilized in DMSO at 1 mM concentration and mixed with 50 mM HEPES buffer containing 0.05% (w/v) DDM, resulting in 50 μM peptide concentration.

The enzymatic cleavage reactions were carried out in 20 mM HEPES, pH 7.4, with 0.05%(w/v) DDM and 10% (v/v) DMSO, with 250 μM peptide substrate and 2.6 μM full-length GlpG at 37 °C. To measure the initial reaction rates, fractions were collected every 15 min for up to 2 h and the reaction was terminated by the addition of 10 mM HCl. Samples for CE were prepared by mixing 20 μl of peptide solutions with 2 μl of 2.2 mM tyramine (internal standard for quantitative analysis). Sample solutions were injected into the capillary by 20 mbar pressure for 10 s. Separations were performed at +25 kV (anode at the capillary injection end). The electrode vessels were replenished with fresh BGE after each run. All analyses were performed in triplicate. Quantitative analysis was based on the ratio of corrected (migration time normalized) peak areas of peptides of interest and the internal standard (tyramine) (32).

Mass Spectrometry—The analysis of enzymatic cleavage products of transmembrane peptides was carried out using MALDI-TOF mass spectrometry on an UltrafleXtreme™ MALDI-TOF/TOF mass spectrometer (Bruker Daltonics, Germany) with α-cyano-4-hydroxycinnamic acid matrix using a thin-layer method (33). For routine quality control during peptide synthesis, mass spectra were acquired on a Waters Micromass ZQ ESCi multimode ionization mass spectrometer, and LTQ Orbitrap XL (Thermo Fisher Scientific) for HR-MS experiments, in both cases using ESI(+) ionization.

Gel-based Assay for Rhomboid Activity—For gel-based assays used in Fig. 1, the purified recombinant full-length maltose-binding protein thioredoxin fusion proteins harboring the transmembrane domains of TatA, LacYTM2, Gurken, and Spitz (18) were used as substrates. The reaction was carried out in 50 mM Tris, pH 7.4, 100 mM NaCl, 10% (v/v) glycerol, 0.05% (w/v) DDM, and 5 μM substrate. Enzyme concentrations varied to ensure adequate conditions for measurement of initial reaction rates for each enzyme-substrate combination. Time courses were measured by withdrawing 10-μl aliquots from the reaction mixture after 10, 20, 30, 40, 50, 60, and 120 min from the start of the reaction, and stopping the reaction by the addition of SDS-PAGE sample buffer. The reaction mixtures were analyzed by SDS-PAGE, Coomassie staining (Instant-Blue, Expedeon, UK), and densitometry as described (19), and initial reaction rates were converted to molar catalytic activities defined as the number of substrate molecules converted by a molecule of the enzyme per unit of time (consistent with the definition by IUPAC (34, 35)). Variations in conditions used for measurements in Fig. 6 are denoted in the figure legend.

The *in vivo* assay of rhomboid activity was carried out essentially as described (19). Cleavage products were detected by SDS-PAGE and Western blotting using primary anti-penta-His mouse monoclonal antibody (Thermo) and IRDye 800CW goat anti-mouse fluorescent secondary antibody (LiCor). Densitometry was done in ImageStudio software (LiCor) and substrate conversion (α) was calculated from band intensities as $\alpha_\tau = [P]/[S] + [P]$, where $[P]$ and $[S]$ are product and substrate concentrations at time τ, which are proportional to the fluorescence intensity of the product and substrate bands at time τ, because the monoclonal antibody binds to the substrate or product in a constant molar ratio irrespective of their molecular weights.

Fluorescence Assay for Rhomboid Activity—The fluorescence assay of rhomboid activity was performed at 37 °C in 96-well black HTS plates (Greiner Bio-One). The reaction conditions were typically as follows: 20 mM HEPES, pH 7.4, 150 mM NaCl, 0.05% (w/v) DDM, 12% (v/v) DMSO, and 10 μM fluorogenic peptide substrate in a final volume of 50 μl, unless noted otherwise. Concentrations of stock solutions of peptide substrates and inhibitors (if applicable) were determined by quantitative amino acid analysis. Fluorescence was read continuously in a plate reader (Tecan Infinite M1000). Excitation and emission wavelengths were 335 and 493 nm, respectively, for the EDANS-DABCYL substrate, and 553 and 583 nm for the TAMRA-QXL610 substrates. Data were evaluated in i-Control (Tecan), Excel (Microsoft), GraphPad Prism 7 (GraphPad Software, Inc.), and GraFit 7 (Erithacus Software, Ltd.) software.

Inhibition Assays—The inhibition assay was carried out in 20 mM HEPES, pH 7.4, 150 mM NaCl, 12% (v/v) DMSO, 0.05% (w/v) DDM at 37 °C in 96-well black HTS plates (Greiner Bio-one). Purified recombinant full-length GlpG (0.4 μM) was pre-

59

incubated with each inhibitor at different concentrations for 1 h at 37 °C. The cleavage reaction was started by adding 10 μM KSp76 and fluorescence was read continuously to measure initial reaction rates as described above.

Reconstitution into Liposomes—E. coli polar lipids (20 mg), with optionally 0.1 mg of Lissamine Rhodamine B-labeled phosphatidylethanolamine (16:0) (Avanti Polar Lipids) added for visibility, were dried in a glass test tube by manual rotation under a nitrogen stream. Residual traces of solvent were removed by overnight incubation in a vacuum chamber (Binder). The resulting lipid film was hydrated in 5 ml of 50 mM acetate, 150 mM NaCl, pH 4.0, by 2 min vortexing followed by a 1-h incubation in a horizontal shaker at 200 rpm and 37 °C, and 3 cycles of freezing in liquid nitrogen and thawing in a 37 °C water bath. The lipid suspension was then extruded through a 200-nm pore membrane by 19 strokes in an Avanti Mini Extruder (Avanti Polar Lipids).

For reconstitution of proteins and peptides into liposomes, these unilamellar LUVs were solubilized in DM to a final ratio of 1.5:1 detergent:lipid, and incubated for 1 h at room temperature under gentle rotation. This mixture was diluted to a final lipid concentration of 2 mg/ml in 50 mM acetate, 150 mM NaCl, pH 4.0, and protein (GlpG or its inactive mutant) dissolved in detergent was added to a final concentration of 8 μg/ml; alternatively, the stock solution of substrate peptide KSp35 in DMSO was diluted to 10 μM. The resulting mixture was incubated at room temperature for 1 h under gentle mixing by inversion. Detergent was removed by overnight dialysis against 500-fold excess of 50 mM acetate, 150 mM NaCl, pH 4, followed by 5 h dialysis against 500-fold excess the same buffer, using 10-kDa MWCO dialysis membranes, which allowed reconstitution of proteoliposomes. These were extruded through 200-nm pore filters 9 times to ensure reproducible size distribution and lamellarity. These final proteoliposomes were harvested by ultracentrifugation (250,000 $\times$ g for 1 h at 4 °C), and resuspended in 10 mM HEPES, pH 7.4, 150 mM NaCl to a concentration of about 33 mg/ml of lipids. The morphology and size distribution of proteoliposomes was analyzed by electron microscopy.

Transmission Electron Microscopy—Liposome samples were negatively stained with 2% phosphotungstic acid on carbon-coated electron microscopy grids and analyzed with a JEOL JEM-1011 device at 80 kV beam acceleration voltage.

CD Spectroscopy—Protein and peptide samples were dissolved in 50 mM phosphate buffer at the indicated concentrations and in the presence of detergent as indicated, or reconstituted in LUVs made of *E. coli* polar lipids and extruded by 100-nm filters to minimize light scattering. Electronic circular dichroism spectra were collected by a Jasco 815 spectrometer (Tokyo, Japan) in the spectral 195–280 nm range using a cylindrical 0.02-cm quartz cell with 0.1-nm step resolution, 5 nm/min scanning speed, 16 s response time, and 1 nm spectral band. After baseline correction, the spectra were expressed as molar ellipticity per residue θ (deg cm^2 dmol^{-1}). Numerical analysis of the secondary structure and secondary structure assignment were performed using a CDPro software package and CONTIN program (36, 37).

References

1. Urban, S., Lee, J. R., and Freeman, M. (2002) A family of rhomboid intramembrane proteases activates all *Drosophila* membrane-tethered EGF ligands. *EMBO J.* **21,** 4277–4286
2. Lee, J. R., Urban, S., Garvey, C. F., and Freeman, M. (2001) Regulated intracellular ligand transport and proteolysis control EGF signal activation in *Drosophila. Cell* **107,** 161–171
3. McQuibban, G. A., Saurya, S., and Freeman, M. (2003) Mitochondrial membrane remodelling regulated by a conserved rhomboid protease. *Nature* **423,** 537–541
4. O'Donnell, R. A., Hackett, F., Howell, S. A., Treeck, M., Struck, N., Krnajski, Z., Withers-Martinez, C., Gilberger, T. W., and Blackman, M. J. (2006) Intramembrane proteolysis mediates shedding of a key adhesin during erythrocyte invasion by the malaria parasite. *J. Cell Biol.* **174,** 1023–1033
5. Fleig, L., Bergbold, N., Sahasrabudhe, P., Geiger, B., Kaltak, L., and Lemberg, M. K. (2012) Ubiquitin-dependent intramembrane rhomboid protease promotes ERAD of membrane proteins. *Mol. Cell* **47,** 558–569
6. Riestra, A. M., Gandhi, S., Sweredoski, M. J., Moradian, A., Hess, S., Urban, S., and Johnson, P. J. (2015) A *Trichomonas vaginalis* rhomboid protease and its substrate modulate parasite attachment and cytolysis of host cells. *PLoS Pathog.* **11,** e1005294
7. Etheridge, S. L., Brooke, M. A., Kelsell, D. P., and Blaydon, D. C. (2013) Rhomboid proteins: a role in keratinocyte proliferation and cancer. *Cell Tissue Res.* **351,** 301–307
8. Chan, E. Y., and McQuibban, G. A. (2013) The mitochondrial rhomboid protease: its rise from obscurity to the pinnacle of disease-relevant genes. *Biochim. Biophys. Acta* **1828,** 2916–2925
9. Song, W., Liu, W., Zhao, H., Li, S., Guan, X., Ying, J., Zhang, Y., Miao, F., Zhang, M., Ren, X., Li, X., Wu, F., Zhao, Y., Tian, Y., Wu, W., *et al.* (2015) Rhomboid domain containing 1 promotes colorectal cancer growth through activation of the EGFR signalling pathway. *Nat. Commun.* **6,** 8022
10. Strisovsky, K. (2016) Why cells need intramembrane proteases: a mechanistic perspective. *FEBS J.* **283,** 1837–1845
11. Pierrat, O. A., Strisovsky, K., Christova, Y., Large, J., Ansell, K., Bouloc, N., Smiljanic, E., and Freeman, M. (2011) Monocyclic β-lactams are selective, mechanism-based inhibitors of rhomboid intramembrane proteases. *ACS Chem. Biol.* **6,** 325–335
12. Dickey, S. W., Baker, R. P., Cho, S., and Urban, S. (2013) Proteolysis inside the membrane is a rate-governed reaction not driven by substrate affinity. *Cell* **155,** 1270–1281

ASBMB

60

13. Arutyunova, E., Panwar, P., Skiba, P. M., Gale, N., Mak, M. W., and Lemieux, M. J. (2014) Allosteric regulation of rhomboid intramembrane proteolysis. *EMBO J.* **33,** 1869–1881

14. Simeonov, A., Jadhav, A., Thomas, C. J., Wang, Y., Huang, R., Southall, N. T., Shinn, P., Smith, J., Austin, C. P., Auld, D. S., and Inglese, J. (2008) Fluorescence spectroscopic profiling of compound libraries. *J. Med. Chem.* **51,** 2363–2371

15. Vosyka, O., Vinothkumar, K. R., Wolf, E. V., Brouwer, A. J., Liskamp, R. M., and Verhelst, S. H. (2013) Activity-based probes for rhomboid proteases discovered in a mass spectrometry-based assay. *Proc. Natl. Acad. Sci. U.S.A.* **110,** 2472–2477

16. Wolf, E. V., Zeißler, A., Vosyka, O., Zeiler, E., Sieber, S., and Verhelst, S. H. (2013) A new class of rhomboid protease inhibitors discovered by activity-based fluorescence polarization. *PLoS ONE* **8,** e72307

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

18. Strisovsky, K., Sharpe, H. J., and Freeman, M. (2009) Sequence-specific intramembrane proteolysis: identification of a recognition motif in rhomboid substrates. *Mol. Cell* **36,** 1048–1059

19. Zoll, S., Stanchev, S., Began, J., Skerle, J., Lepšik, M., Peclinovská, L., Majer, P., and Strisovsky, K. (2014) Substrate binding and specificity of rhomboid intramembrane protease revealed by substrate-peptide complex structures. *EMBO J.* **33,** 2408–2421

20. VanAken, T., Foxall-VanAken, S., Castleman, S., and Ferguson-Miller, S. (1986) Alkyl glycoside detergents: synthesis and applications to the study of membrane proteins. *Methods Enzymol.* **125,** 27–35

21. Kamp, F., Winkler, E., Trambauer, J., Ebke, A., Fluhrer, R., and Steiner, H. (2015) Intramembrane proteolysis of β-amyloid precursor protein by γ-secretase is an unusually slow process. *Biophys. J.* **108,** 1229–1237

22. Scheel, G., Acevedo, E., Conzelmann, E., Nehrkorn, H., and Sandhoff, K. (1982) Model for the interaction of membrane-bound substrates and enzymes: hydrolysis of ganglioside GD1a by sialidase of neuronal membranes isolated from calf brain. *Eur. J. Biochem.* **127,** 245–253

23. Parry, G., Palmer, D. N., and Williams, D. J. (1976) Ligand partitioning into membranes: its significance in determining K_m and K_s values for cytochrome P-450 and other membrane bound receptors and enzymes. *FEBS Lett.* **67,** 123–129

24. Rath, A., and Deber, C. M. (2013) Design of transmembrane peptides: coping with sticky situations. *Methods Mol. Biol.* **1063,** 197–210

25. Wolf, E. V., Zeissler, A., and Verhelst, S. H. (2015) Inhibitor fingerprinting of rhomboid proteases by activity-based protein profiling reveals inhibitor selectivity and rhomboid autoprocessing. *ACS Chem. Biol.* **10,** 2325–2333

26. Haedke, U., Küttler, E. V., Vosyka, O., Yang, Y., and Verhelst, S. H. (2013) Tuning probe selectivity for chemical proteomics applications. *Curr. Opin. Chem. Biol.* **17,** 102–109

27. Lemberg, M. K., Menendez, J., Misik, A., Garcia, M., Koth, C. M., and Freeman, M. (2005) Mechanism of intramembrane proteolysis investigated with purified rhomboid proteases. *EMBO J.* **24,** 464–472

28. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77,** 51–59

29. Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison C. A., 3rd, and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–345

30. Miroux, B., and Walker, J. E. (1996) 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,** 289–298

31. Urbani, A., and Warne, T. (2005) A colorimetric determination for glycosidic and bile salt-based detergents: applications in membrane protein research. *Anal. Biochem.* **336,** 117–124

32. Solínová, V., Kasicka, V., Koval, D., Barth, T., Ciencialová, A., and Záková, L. (2004) Analysis of synthetic derivatives of peptide hormones by capillary zone electrophoresis and micellar electrokinetic chromatography with ultraviolet-absorption and laser-induced fluorescence detection. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **808,** 75–82

33. Fenyo, D., Wang, Q., DeGrasse, J. A., Padovan, J. C., Cadene, M., and Chait, B. T. (2007) MALDI sample preparation: the ultra thin layer method. *J. Vis. Exp.* **192,** 10.3791/192

34. Nomenclature Committee of the International Union of Biochemistry (1979) Units of enzyme-activity, recommendations 1978. *Eur. J. Biochem.* **97,** 319–320

35. Nomenclature Committee of the International Union of Biochemistry (1983) Symbolism and terminology in enzyme-kinetics, recommendations 1981. *Biochem. J.* **213,** 561–571

36. Sreerama, N., and Woody, R. W. (2000) Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* **287,** 252–260

37. Provencher, S. W., and Glöckner, J. (1981) Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* **20,** 33–37

38. Mongay, C., and Cerda, V. (1974) Britton-Robinson buffer of known ionic-strength. *Anal. Chim.* **64,** 409–412

61

6.3 General and modular strategy for designing potent, selective, and pharmacologically compliant inhibitors of rhomboid proteases

Background

Function of rhomboid proteases is associated with diseases like Parkinson's disease, malaria and cancer. Their usage as therapeutic targets is limited by the lack of suitable inhibitors that could be used in biological studies and as templates for further drug development. To fill this gap in search for efficient rhomboid inhibitors, we tested inhibition potency of oligopeptides equipped with electrophilic warheads.

Summary

We discovered that peptidyl-α-ketoamides substituted at the ketoamide nitrogen by hydrophobic groups are potent rhomboid inhibitors active in the nanomolar range. These inhibitors bind the enzyme covalently, in a substrate-like manner, and the inhibition is selective and reversible. We also showed that excellent properties of these inhibitors, which work well also *in vivo* (Figure 20), can be tailored for individual rhomboid enzymes by an optimization of the peptide sequence and by choosing a suitable ketoamide substituent. Thus, we developed a platform for the design of specific and potent rhomboid inhibitors.
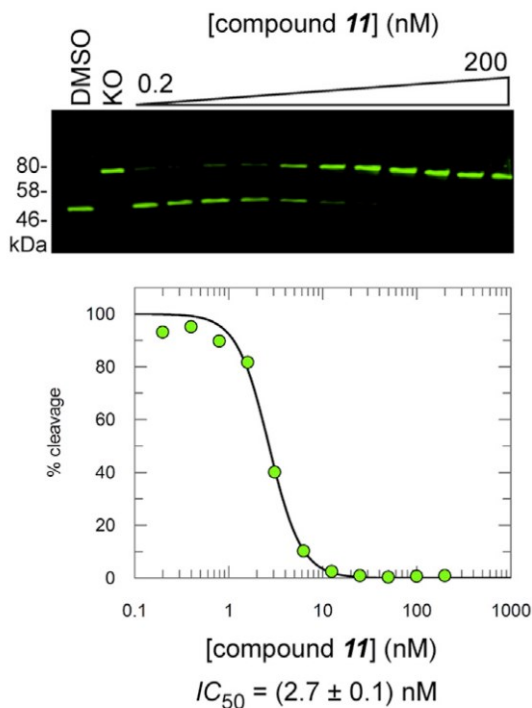
Figure 20: Inhibition of endogenous GlpG by compound No *11* in the membranes of live *E. coli*.

My contribution

I participated in the development of an *in vivo* assay in permeabilized bacterial cells and participated in the characterization of some of the compounds both *in vitro* and *in vivo*.

General and modular strategy for designing potent, selective, and pharmacologically compliant inhibitors of rhomboid proteases. Anežka Tichá, Stancho Stanchev, Kutti R. Vinothkumar, David C. Mikles, Petr Pachl, Jakub Began, Jan Škerle, Kateřina Švehlová, Minh T.N. Nguyen, Steven H.L. Verhelst, Darren C. Johnson, Daniel A. Bachovchin, Martin Lepšík, Pavel Majer, and Kvido Strisovsky (2017), Cell Chemical Biology. 24, 1523–1536

Cell Chemical Biology

General and Modular Strategy for Designing Potent, Selective, and Pharmacologically Compliant Inhibitors of Rhomboid Proteases

Graphical Abstract
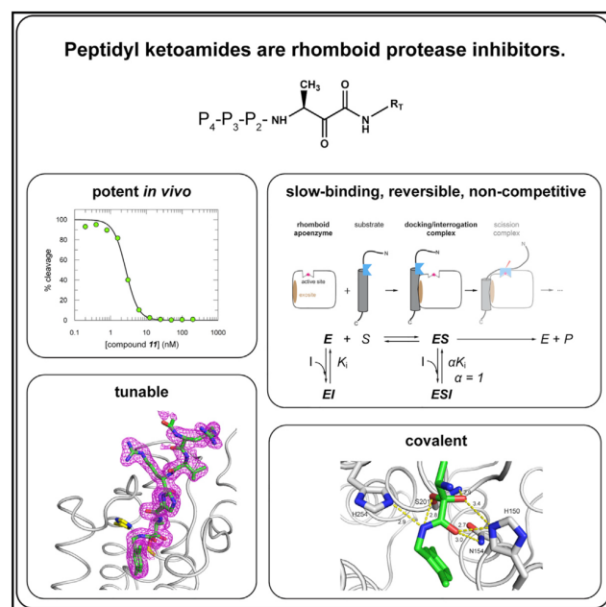
Authors

Anežka Tichá, Stancho Stanchev, Kutti R. Vinothkumar, ..., Martin Lepšík, Pavel Majer, Kvido Strisovsky

Correspondence

kvido.strisovsky@uochb.cas.cz

In Brief

Ticha et al. discover rhomboid protease inhibitors that are unprecedentedly potent and selective. They are based on a pharmacologically compliant chemotype and can thus serve as hitherto unavailable specific tools for cell biology or can yield lead compounds targeting rhomboids in medically relevant contexts such as malaria or Parkinson's disease.

Highlights

- N-substituted peptidyl α-ketoamides are nanomolar inhibitors of rhomboid proteases

- Peptidyl ketoamides inhibit rhomboids covalently, reversibly, and non-competitively

- The peptide and ketoamide substituent independently modulate potency and selectivity

- Peptidyl ketoamides are selective for rhomboids, sparing most human serine proteases

CrossMark

CellPress

Article

General and Modular Strategy for Designing Potent, Selective, and Pharmacologically Compliant Inhibitors of Rhomboid Proteases

Anežka Tichá,[1,4,9] Stancho Stanchev,[1,9] Kutti R. Vinothkumar,[5] David C. Mikles,[1] Petr Pachl,[1] Jakub Began,[1,3] Jan Škerle,[1,2] Kateřina Švehlová,[1] Minh T.N. Nguyen,[6] Steven H.L. Verhelst,[6,7] Darren C. Johnson,[8] Daniel A. Bachovchin,[8] Martin Lepšík,[1] Pavel Majer,[1] and Kvido Strisovsky[1,10,*]

[1]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo n. 2, Prague 166 10, Czech Republic
[2]Department of Biochemistry, Faculty of Science, Charles University, Hlavova 2030/8, Prague 128 43, Czech Republic
[3]Department of Genetics and Microbiology, Faculty of Science, Charles University, Viničná 5, Prague 128 44, Czech Republic
[4]First Faculty of Medicine, Charles University, Kateřinská 32, Prague 121 08, Czech Republic
[5]Medical Research Council Laboratory of Molecular Biology, Cambridge Biomedical Campus, Francis Crick Avenue, Cambridge CB2 0QH, UK
[6]Leibniz Institute for Analytical Sciences ISAS, Otto-Hahn-Strasse 6b, 44227 Dortmund, Germany
[7]KU Leuven - University of Leuven, Herestraat 49, Box 802, 3000 Leuven, Belgium
[8]Chemical Biology Program, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, Box 428, New York, NY 10065, USA
[9]These authors contributed equally
[10]Lead Contact
*Correspondence: kvido.strisovsky@uochb.cas.cz
https://doi.org/10.1016/j.chembiol.2017.09.007

SUMMARY

Rhomboid-family intramembrane proteases regulate important biological processes and have been associated with malaria, cancer, and Parkinson's disease. However, due to the lack of potent, selective, and pharmacologically compliant inhibitors, the wide therapeutic potential of rhomboids is currently untapped. Here, we bridge this gap by discovering that peptidyl α-ketoamides substituted at the ketoamide nitrogen by hydrophobic groups are potent rhomboid inhibitors active in the nanomolar range, surpassing the currently used rhomboid inhibitors by up to three orders of magnitude. Such peptidyl ketoamides show selectivity for rhomboids, leaving most human serine hydrolases unaffected. Crystal structures show that these compounds bind the active site of rhomboid covalently and in a substrate-like manner, and kinetic analysis reveals their reversible, slow-binding, non-competitive mechanism. Since ketoamides are clinically used pharmacophores, our findings uncover a straightforward modular way for the design of specific inhibitors of rhomboid proteases, which can be widely applicable in cell biology and drug discovery.

INTRODUCTION

Rhomboid intramembrane proteases are evolutionarily conserved proteins with numerous important biological functions, including growth factor secretion, regulation of mitochondrial dy-

namics, and membrane protein quality control (Fleig et al., 2012). As such, they are being increasingly explored as potential drug targets, for example, for malaria (Baker et al., 2006; Lin et al., 2013; O'Donnell et al., 2006), cancer (Song et al., 2015), Parkinson's disease (Meissner et al., 2015), and diabetes (reviewed in Chan and McQuibban, 2013). These efforts are, however, hindered by the lack of selective and potent rhomboid inhibitors that could be used for cell biological studies, validation of therapeutic potential of rhomboids, and as templates for drug development (Strisovsky, 2016a). As explained elsewhere in more detail (Strisovsky, 2016a), the currently used inhibitors of rhomboid proteases suffer from drawbacks, making them unsuitable for these purposes. Isocoumarins are highly reactive and lack selectivity (Harper et al., 1985; Powers et al., 2002; Powers et al., 1989), β-lactams have limited potency *in vivo* (half maximal inhibitory concentration [IC$_{50}$] $\sim$5–10 μM) (Pierrat et al., 2011), and β-lactones are not very potent (apparent IC$_{50}$ of $\sim$40 μM) (Wolf et al., 2013). Furthermore, no rational strategy for modulation of their selectivity exists for any of these inhibitor classes. Here, we address both of these bottlenecks.

The principles of the mechanism and specificity of a protease determine to a large extent the strategies for inhibitor development (Drag and Salvesen, 2010). Rhomboids are serine proteases with a Ser-His catalytic dyad (Wang et al., 2006), and they recognize their transmembrane substrates in a two-tier process. It is assumed that first a portion of the transmembrane domain of the substrate docks into an intramembrane interaction site of rhomboid within the plane of the lipid bilayer, upon which a linear segment of the substrate (possibly generated by local unfolding of the top of the substrate's transmembrane helix) interacts with the water-exposed active site (reviewed in Strisovsky, 2016a; Strisovsky, 2013). This "recognition motif" encompasses the P4 to P2' (Schechter and Berger, 1967) residues of the substrate (Strisovsky et al., 2009), it largely determines the k_{cat} of the

reaction (Dickey et al., 2013), and thus modulates selectivity toward a given rhomboid protease (Ticha et al., 2017). Recent reports have shown that peptidyl chloromethylketones (Zoll et al., 2014) and peptidyl aldehydes (Cho et al., 2016) are weakly inhibiting rhomboid proteases at medium to high micromolar concentrations, but they lack selectivity and their potency is insufficient for use as research tools.

Inspired by the current knowledge of the rhomboid protease mechanism (reviewed in Strisovsky, 2016a), we set out to explore the chemical space of oligopeptides equipped with electrophilic warheads in search of new rhomboid inhibitors of greater potency. Our systematic analysis resulted in the discovery of a modular scaffold based on peptidyl-ketoamide substituted with hydrophobic groups that represents a novel class of potent and selective rhomboid inhibitors. The *in vivo* activity of these compounds is in the low nanomolar range, which is up to three orders of magnitude more potent than any other currently known rhomboid inhibitors. Furthermore, we gained insight into the mode of binding of peptidyl ketoamides by solving their co-crystal structures with rhomboid protease, and we present strategies to modify their selectivity and potency on a systematic basis. We expect this compound class to find a widespread use in cell biology in rhomboid protease related contexts and to provide templates for the development of drugs targeting rhomboid proteases.

RESULTS

The Potency of Substrate-Derived Peptidyl Chloromethylketone Inhibitors Can be Markedly Enhanced by Optimizing the Amino Acid Sequence of the P5 to P1 Region

Rhomboid proteases exhibit discernible sequence preferences in the P5 to P2′ region of their substrates (Strisovsky et al., 2009; Zoll et al., 2014). To gain insight into these preferences and their possible interactions, we have generated tetra- and pentapeptidyl chloromethylketones (CMK or cmk henceforth) harboring amino acids preferred in positions P5 to P1 by the *Escherichia coli* rhomboid GlpG (GlpG henceforth), using the sequence background of the *Providencia* TatA (Stevenson et al., 2007), represented by the parent compound Ac-IATA-cmk. We measured the inhibitory properties of this series of compounds using a newly developed *in vitro* assay employing a fluorogenic transmembrane peptide substrate (Ticha et al., 2017) that faithfully represents a native rhomboid substrate. The effects of the mutations were additive, and the inhibitor containing the most favored amino acids in positions P5 to P1 (Ac-RVRHA-cmk) is approximately 26-fold more potent than the parent compound Ac-IATA-cmk (Figure 1A).

To provide mechanistic explanation for the observed increase in inhibitory potency, we determined the structures of GlpG in complex with Ac-RVRHA-cmk and Ac-VRHA-cmk (Figure 1B). The side chain of Arg in the P5 position of Ac-RVRHA-cmk could not be modeled due to poor electron density, and the two structures are otherwise virtually identical; superposition of all corresponding Cα atoms yields a root-mean-square deviation of 0.19 Å per atom (using the SSM method as implemented in CCP4MG v2.10.4; Krissinel and Henrick, 2004; Mitchell et al., 1990). Both inhibitors interact with the L3 and L5 loops via

main-chain hydrogen bonds, and via hydrogen bonds involving the side chains of the strongly preferred Arg and His in the P3 and P2 position, respectively (Figure 1C). These interactions are not observed in the structure of the parent compound Ac-IATA-cmk (Zoll et al., 2014), suggesting that they contribute to the higher potency of Ac-RVRHA-cmk and Ac-VRHA-cmk over Ac-IATA-cmk. The interactions of the residues in the P4, P3, and P2 positions with the enzyme are structurally independent, explaining why the effects of substitutions in these positions are additive (Figure 1A). The overall binding mode of both compounds into the rhomboid active site is similar to the binding mode of peptide aldehyde Ac-VRMA-cho (Cho et al., 2016) (Figure 1D). Collectively, these data show that rhomboid subsite preferences are additive in the context of an active site targeted inhibitor and that sequence optimization in this region can significantly increase the inhibitory potency of the compounds.

A Screen of Covalent Reversible Warheads for Inhibition of Rhomboid

Since the sequence-optimized chloromethylketones are poor inhibitors with low micromolar IC_{50}, we searched for alternative, more suitable electrophilic warheads that might improve the inhibitory potency. Furthermore, we reasoned that extending the inhibitor to the prime side of the active site might offer additional binding energy. We therefore synthesized a series of compounds based on the Ac-RVRHA sequence equipped with a selection of electrophilic, reversibly binding warheads commonly used for serine proteases in pharmacological settings (reviewed in Hedstrom, 2002; Walker and Lynas, 2001), including trifluoromethylketones, boronates, acylsulfonamides, thiazolylketones, and ketoamides (Figure 2), the last three of which can be extended into the prime side. We measured the apparent IC_{50} values of these compounds, and while trifluoromethylketones, acylsulfonamides, and thiazolylketones showed none or very weak inhibition in the millimolar range, the apparent IC_{50} of the boronate was 8 μM and of the ketoamide 203 μM under identical reaction conditions (Figure 2). Although the peptidyl boronate was the best of the series, it was still a relatively weak inhibitor comparable with the parent chloromethylketone, and it was not clear how to further improve its potency. The ketoamide was about 25-fold less potent, but since it could be extended to the prime side by a modification at the ketoamide nitrogen (Chatterjee et al., 1999; Liu et al., 2004), we next focused our attention on this class of compounds.

Extensions at the Prime Side of Peptidyl Ketoamides Greatly Enhance Their Inhibitory Potency

We hypothesized that extending the peptidyl ketoamides to the prime side of the active site might increase their potency, since the P2′ residue (hydrophobic in case of GlpG) was shown to be important for substrate recognition by rhomboids (Dickey et al., 2013; Strisovsky et al., 2009), and interactions of the substrate transmembrane domain beyond P2′ potentiate substrate cleavage in a detergent micelle assay (Ticha et al., 2017). We synthesized a series of peptidyl ketoamides based on the Ac-RVRHA sequence, bearing a mostly hydrophobic "tail" of increasing size at the ketoamide nitrogen (Figure 3A) that could reach far into the prime side of the rhomboid active site. The tail substituent indeed had a dramatic effect on the potency of
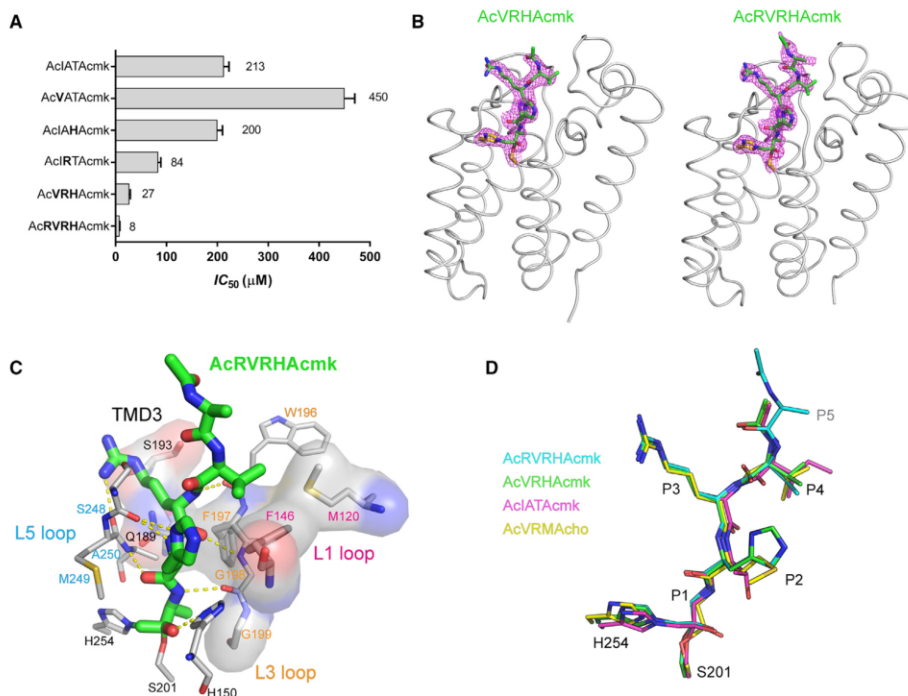
Figure 1. The Potency of Substrate-Derived Inhibitors Can be Improved by Modifying the Amino acid Sequence of the P5 to P1 Region

(A) The parent inhibitor Ac-IATA-cmk was modified by introducing strongly preferred amino acids (Zoll et al., 2014) into the P4, P3, P2, and P5 positions to yield the listed compounds. Their apparent IC_{50} values were measured with 1 hr preincubation using 10 μM fluorogenic substrate KSp35 and 0.05% (w/v) DDM. The reported values are best-fit means with SD representative of 2–3 measurements.

(B) The sequence-optimized peptidyl chloromethylketones were soaked into the native crystals of GlpG and structures of the complexes were solved by X-ray diffraction (for statistics, see Table S1). In the displayed structures, the catalytic dyad is shown as yellow sticks and the inhibitors are shown as green sticks surrounded by the $2mF_o - DF_c$ electron density map contoured at 1σ and shown 1.6 Å around the stick model. Note that in the Ac-RVRHA-cmk structure (right), the side chain of the Arg residue in the P5 position of the inhibitor has not been modeled due to poor or missing electron density peaks.

(C) Interactions of RVRHA-cmk with GlpG were analyzed by Ligplot+ (Laskowski and Swindells, 2011). Ligands are shown as thick sticks with carbons in green, proteins as thin sticks with carbons in gray, hydrogen bonds as yellow dashed lines, and amino acids involved in van der Waals contacts are highlighted as transparent surfaces. The inhibitor forms covalent bonds with S201 and H254 via the chloromethylketone warhead, and it hydrogen bonds with the backbone of residues 196–198 from the L3 loop and residues 248–250 from the L5 loop. van der Waals contacts with the inhibitor are formed by the residues from the L3 loop of GlpG, by S193 and Q189 from TMD3, and by F146 and M120 from the L1 loop that pack against the Val side chain of the P4 position of the inhibitor, as observed previously (Zoll et al., 2014).

(D) The conformations of peptide inhibitors bound to the active site of GlpG were compared by performing structural alignment of the complexes of VRHAcmk (PDB: 5MT7), RVRHAcmk (PDB: 5MT8), IATAcmk (PDB: 4QO2), and VRMAcho (PDB: 5F5B) in PyMOL (Schrodinger, 2012). Note that the structure of RVRHAcmk suggests where the P5 amino acid points, but the density for this side chain is not visible beyond its β-carbon. The backbone of the ligands in all complexes has virtually identical conformation with the exception of the distortion of the oxyanion by the chloromethylketone, and the biggest differences are found in the conformation of the P2 side chain, which is not surprising, because almost any side chain can be accommodated in this position (Zoll et al., 2014).

the inhibitors *in vitro* (Figure 3A). The most effective compound of the series, bearing a 4-phenyl-butyl tail (compound **11**), already displayed about 1,000-fold lower IC_{50} than the parent compound **1**. The IC_{50} of **11** reaches half of the enzyme concentration used in the assay, suggesting that **11** is a potent inhibitor of GlpG.

Next, we examined the relative importance of the peptidyl part for the inhibitory potency. We generated a series of progressively N-terminally truncated variants of **9** and measured their inhibitory potency against GlpG (Figure 3B). Removing the P5 Arg from **9** to yield **12** had virtually no effect on IC_{50} (0.44 versus 0.55 μM),

while removing the P5 and P4 residues in **13** led to a ~20-fold decrease in potency in comparison with the parent compound **9** (IC_{50} changes from 0.44 to 9 μM). Removing three residues (from P5 to P3) in **14** led to a dramatic ~150-fold loss of potency, yielding a weak inhibitor with about 65 μM IC_{50}, and the absence of the P5 to P2 residues in **15** resulted in a total ~2,250-fold reduction in potency compared with **9** and IC_{50} higher than 1 mM. This experiment demonstrates that the non-prime (P4 to P1) and prime sides of the inhibitor contribute to its potency almost equally. The P5 residue can be omitted with only a
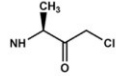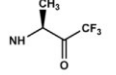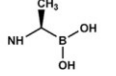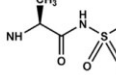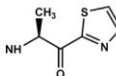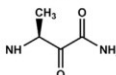
Ac-Arg-Val-Arg-His-Ala-X

warhead	Ala-X	IC_{50}
chloromethylketone		8 μM
trifluoromethylketone		> 1 mM
boronate		8 μM
acylsulfonamide		> 1 mM
thiazolylketone		> 1 mM
α-ketoamide		203 μM

Figure 2. A Screen of Electrophilic Warheads for the Inhibition of Rhomboid Proteases

The optimized parent sequence Ac-RVRHA was linked to electrophilic warheads commonly used for targeting serine proteases (reviewed in Hedstrom, 2002; Walker and Lynas, 2001). The apparent IC_{50} values of the compounds were measured in 0.05% DDM using 10 μM substrate KSp35 (Ticha et al., 2017) with 1 hr preincubation. Given are the mean values of 2–3 measurements.

marginal effect on inhibitory potency, which can be probably compensated by a suitable prime side tail substituent.

Ketoamides are known to be covalent reversible inhibitors of soluble serine proteases with a classical catalytic triad (Liu et al., 2004). Since rhomboids are unusual serine proteases using only a Ser-His dyad for catalysis (Wang et al., 2006), we investigated the mechanism of rhomboid inhibition by these compounds more closely. Progress curves measured at varying inhibitor concentrations (Figure 4A) had biphasic character; especially at the highest inhibitor concentrations tested, the reaction rate decreased over approximately the first hour and became more or less constant over the next hour (Figures 4A and 4B). This indicates that inhibition was time dependent, which is typical for slow-binding inhibitors (Copeland, 2013b). In addition, upon rapid dilution of inhibitor-saturated enzyme to a subinhibitory concentration, the reaction rate was partially recovered (Figure 4C), together indicating that peptidyl ketoamides exhibit slow-binding reversible behavior (Copeland, 2013; Singh et al., 2011).

The slow-binding reversible inhibition mechanism can be formally divided into two steps. First, an initial encounter complex (EI) forms, and then a slow step leads to the much more sta-

ble EI* complex (E + I ↔ EI ↔ EI*), usually involving a significant conformational change of the enzyme (Copeland, 2013a). To analyze the contribution of each of these two steps to the mechanism of inhibition of rhomboids by peptidyl ketoamides, we investigated the concentration and time dependence of inhibition kinetics by **10**. The "bending" of biphasic progress curves (Figures 4A and 4B) reflects the rate of "onset of inhibition" described by the rate constant k_{obs}, which can be obtained from progress curve data using non-linear fitting to Equation 1:

$$[P] = v_s t + \frac{(v_i - v_s)}{k_{obs}}[1 - \exp(-k_{obs}t)], \qquad \text{(Equation 1)}$$

where $[P]$ is the concentration of the reaction products, v_i is the initial reaction rate in the first phase of the biphasic progress curve, and v_s is the steady-state reaction rate (Figure 4B). Analysis of progress curves from Figure 4A showed that v_i was independent of inhibitor concentration, and the plot of k_{obs} against inhibitor concentration fitted well to a linear dependence (Figure 4D). Both phenomena are typical for simple (single-step) slow-binding inhibition (E + I ↔ EI) (Figure 4D); in other words, peptidyl ketoamides behave as "regular" reversible inhibitors but with very low rate constants for association and dissociation (Copeland, 2013a; Morrison, 1982), leading to the slow-binding kinetics. The application of this model yields the apparent inhibitory constant K_i^{app} (i.e., not taking into account the inhibition modality and the influence of the substrate) for **10** of (123 ± 47) nM (Figure 4D).

The true inhibitory constant K_i, which is an important, substrate-independent property of an inhibitor, can be calculated from the apparent inhibitory constant K_i^{app}, depending on the inhibitory modality and kinetic parameters of the substrate used. Global non-linear regression fitting of Michaelis curves measured in the presence of increasing concentrations of **10** (plotting v_s against $[S]$) shows that the experimental data are best described by a non-competitive inhibition model (Figure 4E). This inhibition mode means that the inhibitor can bind both to the free enzyme and to the enzyme-substrate complex; in this case specifically, the affinities of the inhibitor to both forms of the enzyme are equal (α = 1) (Copeland, 2013a). Although non-competitive modality is non-typical for slow-binding inhibitors, it is conceivable why it is plausible in the case of peptidyl ketoamides and rhomboids. Several studies have suggested that substrate recognition by rhomboid proteases proceeds in two steps, via a docking/interrogation complex, where only a part of substrate's transmembrane domain interacts with rhomboid, followed by the interaction of the recognition motif with the active site forming the scission-competent complex (Cho et al., 2016; Strisovsky, 2016a, 2016b; Strisovsky et al., 2009) (Figure 4E). Since the active site is unoccupied in the docking complex, binding of an active site-directed inhibitor is possible (Figure 4E), resulting in non-competitive behavior. Under this mechanism of inhibition, the true K_i is identical to K_i^{app} (Copeland, 2013a; Purich, 2010). Similar progress curve analyses of **9** and **11** yield their k_{on} and k_{off} rate constants, their K_i^{app} ($K_i^{app} = k_{off}/k_{on}$), and the true K_i values of (219 ± 76) nM and (45 ± 8) nM, respectively (Figure 4F and Table 1). In summary, this kinetic analysis shows that the peptidyl ketoamides described here are high-affinity inhibitors of rhomboid proteases unprecedented in the literature.

A

Ac-Arg-Val-Arg-His-NH—[CH₃][O][RT][O]

Compound number	Tail substituent R_T	
1		R_T1-
2		R_T2-
3		R_T3-
4		R_T4-
5		R_T5-
6		R_T6-
7		R_T7-
8		R_T8-
9		R_T9-
10		R_T10-
11		R_T11-

IC_{50} (μM)

B

X-NH—[CH₃][O][R_T9][O]

Compound number	X=
9	Ac-Arg-Val-Arg-His-
12	Ac-Val-Arg-His-
13	Ac-Arg-His-
14	Ac-His-
15	Ac-

IC_{50} (μM)

based probe (ABP) competition assays that enable a more general and substrate-independent measurement of inhibitory potency, because they rely solely on the competition between a fluorescently labeled activity-based probe and the tested inhibitor (Nguyen et al., 2015; Serim et al., 2012). The assays we employed used fluorophosphonate ABPs that target the catalytic serine of a wide-range of serine hydrolases (Bachovchin et al., 2014), including rhomboids (Xue et al., 2012), and are thus very practical general detection reagents even for serine hydrolases for which sensitive substrates might not be available.

First, we tested **9**, **10**, and **11** against a panel of bacterial and eukaryotic rhomboid proteases (Wolf et al., 2015), and found that all three compounds potently competed with ABP labeling of rhomboids from bacterium *Providencia stuartii* (AarA), archaebacterium *Methanocaldococcus jannaschii* (MjROM), and three closely related rhomboids from bacteria *E. coli* (EcGlpG), *Haemophilus influenzae* (HiGlpG), and *Vibrio cholerae* (VcROM). Compounds **9**, **10**, and **11** outcompeted the ABP even at a concentration of 500 nM, suggesting that they were potent inhibitors of these rhomboid proteases. In contrast, none of these compounds were able to compete with the ABP labeling of rhomboid protease from the bacterium *Aquifex aeolicus* (AaROM), rhomboids from *Drosophila* (DmRho1) and mouse (MmRHBDL3), and they only partially inhibited labeling of rhomboid protease from bacterium *Thermotoga maritima* (TmROM) at 50 μM (Figure 5A). These data demonstrate that already these first-generation

Selectivity of Peptidyl Ketoamides

Any enzyme inhibitors to be used as specific tools for cell biology or as starting points for drug development must show sufficient level of selectivity toward their intended target. This is particularly important for compounds that react with the catalytic nucleophile common to many serine hydrolases. Only limited tests of selectivity have been conducted for the currently used rhomboid inhibitors isocoumarins, β-lactams and β-lactones, at best interrogating them against trypsin or chymotrypsin (Pierrat et al., 2011; Vosyka et al., 2013). To map the selectivity of peptidyl ketoamides more objectively and widely, we employed activity-
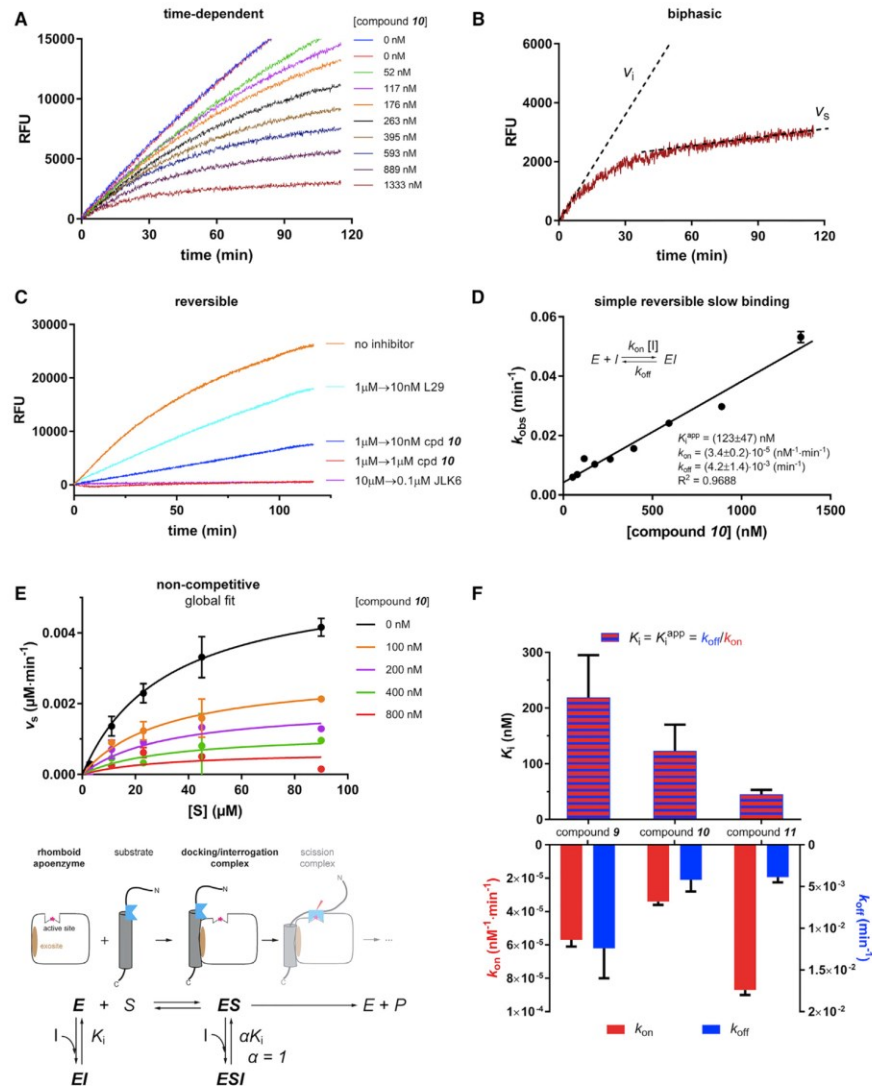
Figure 4. Mechanism of Inhibition of Rhomboid Protease GlpG by Peptidyl Ketoamides Analyzed Using Fluorogenic Transmembrane Peptide Substrates

(A) Progress curves in the presence of increasing concentrations of compound **10** show biphasic character, which is typical for slow-binding inhibitors (Copeland, 2013a; Morrison, 1982). GlpG (0.5 nM) was incubated with 25 μM substrate KSp93 in the presence of 0.05% (w/v) DDM and 0–1,333 nM **10**. Fluorescence at 493 nm was followed to monitor substrate cleavage.

(B) Biphasic progress curves characterized by an initial reaction rate (v_i) and steady state reaction rate (v_s). The progress curve at 1,333 nM compound **10** from the experiment in (A) is shown in detail, and both reaction rates obtained from non-linear regression into Equation (1) are shown as dotted lines.

(C) Reversibility of inhibition by ketoamides was assessed by the rapid dilution method (Harper et al., 1985; Harper and Powers, 1985). Compound **10** (1 μM) was pre-incubated with 0.4 μM GlpG, 0.05% (w/v) DDM at 37°C for 1 hr, leading to complete inhibition. This solution was then rapidly diluted 100-fold either into the reaction buffer containing 10 μM substrate KSp64 (Ticha et al., 2017) (yielding final 10 nM inhibitor) or into the reaction buffer with 10 μM substrate KSp64 and 1,000 nM **10**. For comparison, we used β-lactam L29 (Pierrat et al., 2011) at 1 μM and isocoumarin JLK6 (Vinothkumar et al., 2010) at 10 μM as known reversible

Table 1. Summary of the Inhibition Properties of Compounds 9–11

Compound	GlpG					YqgP
	K_i (nM)	k_{on} (10^{-6} nM^{-1}·min^{-1})	k_{off} (10^{-3} min^{-1})		IC$_{50}$ *In Vivo* (nM)	IC$_{50}$ *In Vivo* (nM)
9	220 ± 80	5.7 ± 0.4	12.0 ± 0.4	8.8 ± 0.4		ND
10	120 ± 50	3.4 ± 0.2	4.2 ± 1.4	6.0 ± 0.1		ND
11	45 ± 8	8.7 ± 0.3	3.9 ± 0.6	2.7 ± 0.1		~5–10

Values for GlpG are reported as means ± SD.

peptidyl ketoamides can discriminate between diverse rhomboid proteases.

We next examined peptidyl ketoamides for their possible off-target effects on other serine proteases. To get a representative picture of the selectivity of peptidyl ketoamides, we employed a recently developed EnPlex technology, which allows multiplex analysis of ABP competition with about 100 human serine hydrolases, mostly proteases (Bachovchin et al., 2014). Profiling of 9, 10, and 11 showed that in the concentration range where they inhibit rhomboid proteases, they fail to inhibit most of the tested human serine hydrolases with the exception of prolylcarboxypeptidase (PRCP) and the sequence related dipeptidyl-peptidase 7 (DPP7) (Figure 5B). To put this into the context of the current generation of rhomboid inhibitors, isocoumarins S006 and S016 (Vosyka et al., 2013) hit about a dozen serine hydrolases in the same concentration range. The β-lactam L41 (Pierrat et al., 2011) inhibited appreciably only one enzyme (predicted serine carboxypeptidase CPVL), but it is much less potent on rhomboids than 9, 10, and 11, and it does not inhibit GlpG completely *in vivo* (Pierrat et al., 2011). The selectivity profile of ketoamide inhibitors of rhomboids is similar to the profile of clinically used ketoamide inhibitors of the hepatitis C protease (Bachovchin et al., 2014), indicating that the rhomboid-targeting N-modified peptidyl ketoamides are sufficiently selective with minimal risk of cross-reactivity against other serine proteases.

Peptidyl Ketoamides Potently Inhibit Rhomboids in Living Cells

Having established the mechanism of rhomboid inhibition by peptidyl ketoamides in detergent micelles, and having shown that 9, 10, and 11 are able to inhibit potently rhomboid prote-

ases from several Gram-negative bacteria (Figure 5A), we next tested whether the inhibitors will be able to target rhomboid proteases embedded in their native lipid bilayer in live cells. First, we expressed the model substrate derived from LacYTM2 in *E. coli* expressing endogenous levels of GlpG, incubated the bacterial cultures in the presence of increasing concentrations of 9, 10, and 11, and detected the steady-state levels of substrate processing by quantitative near-infrared western blotting (Figure 6A). The calculated substrate conversion values relative to the uninhibited reaction were plotted against the inhibitor concentration yielding the *in vivo* IC$_{50}$ values. Strikingly, the most effective compound 11 had an *in vivo* IC$_{50}$ value of 2.7 nM, which is three orders of magnitude lower than any other currently known rhomboid inhibitors (Cho et al., 2016; Pierrat et al., 2011).

We then extended the range of organisms to *Bacillus subtilis*, a representative of Gram-positive bacteria, which have a thick cell wall and include major pathogens such as *Staphylococcus*, *Listeria*, *Streptococcus*, and others. Since the endogenous substrate of the *B. subtilis* rhomboid protease YqgP is unknown, and no robust and rescuable phenotypes have been reported for YqgP, we focused on inhibition of cleavage of a model substrate. Of the common model rhomboid substrates, YqgP cleaves LacYTM2 reasonably well (Ticha et al., 2017). We have thus expressed MBP-LacYTM2-Trx (Strisovsky et al., 2009) from the ectopic *xkdE* locus (Gerwig et al., 2014) in the wild-type *B. subtilis* 168 (BS87) and its *yqgP* deletion mutant (BS88) on an otherwise rhomboid-free background. Although the substrate was to some extent truncated by unknown processes in the *ΔyqgP* strain, a specific, closely co-migrating rhomboid-generated N-terminal cleavage product (Figure 6B) was produced in the YqgP wild-type strain BS87 but not in the *ΔyqgP*

and irreversible inhibitors of rhomboid proteases, respectively. Activity recovery was followed by measuring fluorescence over the course of 120 min with excitation at 553 nm and emission at 583 nm.

(D) Progress curves of KSp93 cleavage at increasing concentrations of 10 measured under (A) were analyzed by non-linear regression as described for slow-binding inhibition (Copeland, 2013a; Morrison, 1982) using GraphPad Prism version 7.02 for Windows (GraphPad Software, La Jolla, California, USA) to yield the rate constant for the onset of inhibition, k_{obs}. The linear character of the dependence of k_{obs} on inhibitor concentration is typical for a simple slow-binding mechanism (inset), and its linear regression allows determination of the underlying apparent inhibitory constant K_i^{app} and its constituent rate constants k_{on} and k_{off} (inset). The k_{obs} values are reported as best-fit mean ± SD.

(E) The influence of inhibitor concentration on the apparent K_M and k_{cat} suggests the mode of inhibition by compound 10. Michaelis curves at the indicated inhibitor concentrations were measured by plotting v_S (measured as in Figure 4B) against substrate concentration using 1 nM GlpG, 0.15% (w/v) DDM and highly sensitive substrate KSp96. The data were globally fitted to the models of competitive, non-competitive (figure top), uncompetitive, and mixed inhibition as implemented in GraphPad Prism 7.02, and their statistical analysis yielded the non-competitive mechanism (figure bottom) as the best fit. The middle of the figure shows a schematic mapping of this mechanism onto the consensual model of substrate recognition by rhomboid proteases. The data points in the Michaelis plots (figure top) represent means ± SD of duplicate measurements.

(F) Summary of inhibition kinetics parameters of compounds 9, 10, and 11. The apparent inhibitory constants K_i^{app} (blue-red striped columns) and the constituent rate constants k_{on} (red columns) and k_{off} (blue columns) were determined from progress curve analysis as shown in (A and C) (note that $K_i^{app} = k_{off}/k_{on}$). For non-competitive inhibitors, the true inhibitory constant K_i equals K_i^{app}. Note that 11 is a highly potent inhibitor with K_i of (45 ± 8) nM. Graphs show best-fit means with SDs.
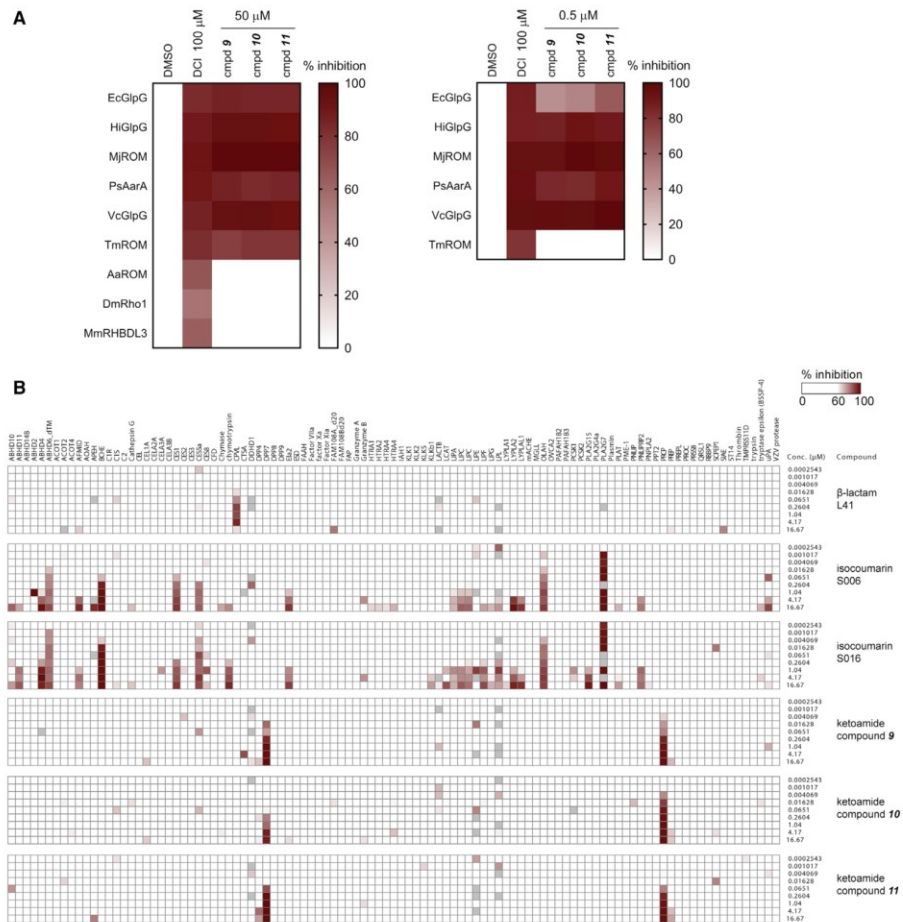
Figure 5. Selectivity of Peptidyl Ketoamides

(A) Selectivity of compounds **9**, **10**, and **11** for nine rhomboid proteases was profiled using activity-based probe competition assay at 50 μM and 0.5 μM concentration. The upper limit of enzyme concentration was 0.4 μM.

(B) Selectivity of compounds **9**, **10**, and **11** against human serine hydrolases was analyzed using EnPlex as described (Bachovchin et al., 2014).

strain BS88 (Figure 6B). In the absence of any inhibitors, MBP-LacYTM2-Trx was cleaved to about 75% conversion by the endogenous YqgP, and addition of **11** into the growth media completely inhibited substrate cleavage at 50 nM (Figure 6B), indicating that the compound can penetrate the Gram-positive cell wall easily. Moreover, since compound **11** also inhibits several homologs of GlpG (Figure 5A), it is safe to assume that YqgP orthologs in other *Bacilli*, *Lactobacilli*, *Staphylococci*, and *Listeria* might be equally susceptible to inhibition by the described inhibitors, and compound **11** and its analogs can be directly used for chemical proteomics and cell biological studies of rhomboid proteases in Gram-positive bacteria.

N-Modified Peptidyl Ketoamides Bind the Rhomboid Active Site in a Substrate-like Manner Occupying the S4 to S2′ Subsites

To understand why peptidyl ketoamides are such efficient rhomboid inhibitors and to establish the basis for structure-guided design of their improved variants, we determined the co-crystal structures of GlpG with **9** and **10** (Figure 7A). The complexes were formed by soaking the inhibitors into apoenzyme crystals, and the structures were solved using diffraction data to 2.16 and 1.78 Å resolution, respectively, allowing detailed comparison of their binding modes. In both cases, the pentapeptide RVRHA binds the active site cavity as an extended β strand,
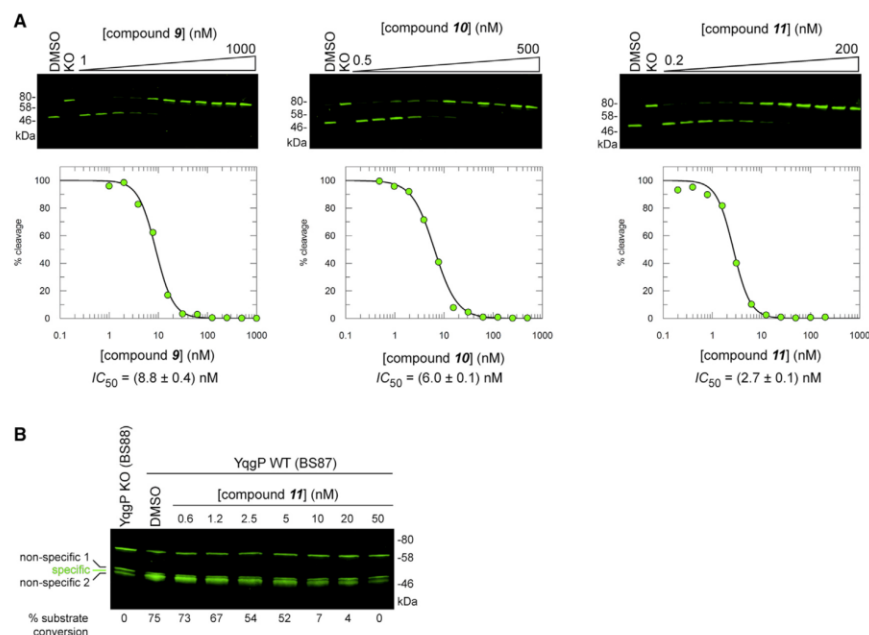
Figure 6. Peptidyl Ketoamides Potently Inhibit Rhomboid Activity in the Membranes of Living Cells

(A) Inhibition of endogenous GlpG by compounds **9**, **10**, and **11** in the membranes of live *E. coli*. The substrate MBP-FLAG-LacYTM2-Trx (Strisovsky et al., 2009) was expressed in wild-type *E. coli* NR698 with genetically permeabilized outer membrane (Ruiz et al., 2005) in the presence of increasing concentrations of inhibitors as described in STAR Methods. Substrate cleavage was measured in cell lysates by immunoblotting for FLAG and quantified using near-infrared fluorescence. The reported *in vivo* IC_{50} values are best-fit means with SD representative of 2–3 measurements. DMSO, dimethylsulfoxide vehicle control; KO, *E. coli glpG::tet*.

(B) Inhibition of endogenous YqgP by compound **11** in the membranes of live *B. subtilis*. The substrate AmyE$_{SP}$-MBP-FLAG-LacYTM2-Trx-HA was expressed in *Bacillus subtilis* 168 (*ydcA::neo, xdkE::AmyE$_{SP}$-MBP-LacYTM2-Trx(erm, lin)*) (BS87) in the presence of increasing concentrations of inhibitors as described in STAR Methods. Substrate cleavage was detected in cell lysates by immunoblotting for FLAG and detection by near-infrared fluorescence. Unspecific cleavage of the substrate was corrected for by substracting the intensity of the unspecific bands formed in the YqgP knockout control cells (BS88) from the product band and the closely co-migrating unspecific bands observed in the YqgP positive cells (BS87). This treatment was necessary because the specific cleavage product could not be resolved sufficiently well from the non-specific bands to be integrated separately. DMSO, dimethylsulfoxide vehicle control; YqgP KO, *Bacillus subtilis* 168 (*ydcA::neo, yqgP::tet, xdkE::AmyE$_{SP}$-MBP-LacYTM2-Trx(erm, lin)*) (BS88).

virtually identically to the binding mode of Ac-RVRHA-cmk (Figure 1C). We do observe electron density for the side chain of arginine in the P5 position in both structures, but its conformation differs between **9** and **10** (Figure 7A), and it is influenced by crystal contacts with the same residue from a neighboring molecule in the crystal (data not shown).

In both inhibitors, the ketoamide warhead is covalently bonded via its proximal carbon to the side-chain oxygen of the catalytic S201, and it engages in a network of six hydrogen bonds in the active site (Figure 7B). The oxyanion formed by the proximal carbonyl oxygen accepts hydrogen bonds from His150 and the main-chain amide nitrogen of the catalytic serine, and the distal ketoamide carbonyl oxygen accepts hydrogen bonds from both H150 and N154, thus amply saturating the hydrogen-bonding groups engaged in the stabilization of the oxyanion (Cho et al., 2016). Furthermore, the ketoamide nitrogen donates a hydrogen bond to H254 and to the S201 side-chain oxygen covalently bound to the warhead. The resulting network

of six hydrogen bonds (Figure 7B) probably helps position the ketoamide warhead in the proximity of the hydroxyl of the catalytic S201 to enhance its chemical reactivity in a conformation-dependent manner.

The tail substituents of **9** and **10** (R_T9 and R_T10) interact with the prime side of GlpG, buried in a cavity delimited by the side chains of amino acids F245, M247, and M249 from the L5 loop, W236 from TMD5, F153 and W157 from TMD2, and residues V204, M208, Y205, H254, H150, and N154 (Figure 7C). The different sets of residues making van der Waals contacts with each ketoamide tail are shown in magenta. The NH group of the side chain of W236 seems to form a weak H-π bond with the phenyl ring of the tail of both compounds, and F245 engages in $\pi \ldots \pi$ stacking against the dimethylbenzyl in R_T10 (Figure 7C). Structural alignment of both ketoamide complexes to the complex of the β-lactam L29 (Vinothkumar et al., 2013) and isocoumarin S016 (Vosyka et al., 2013) (Figure 7D) shows that the tails of **9** and **10** bind in a similar area (the S2′ subsite)
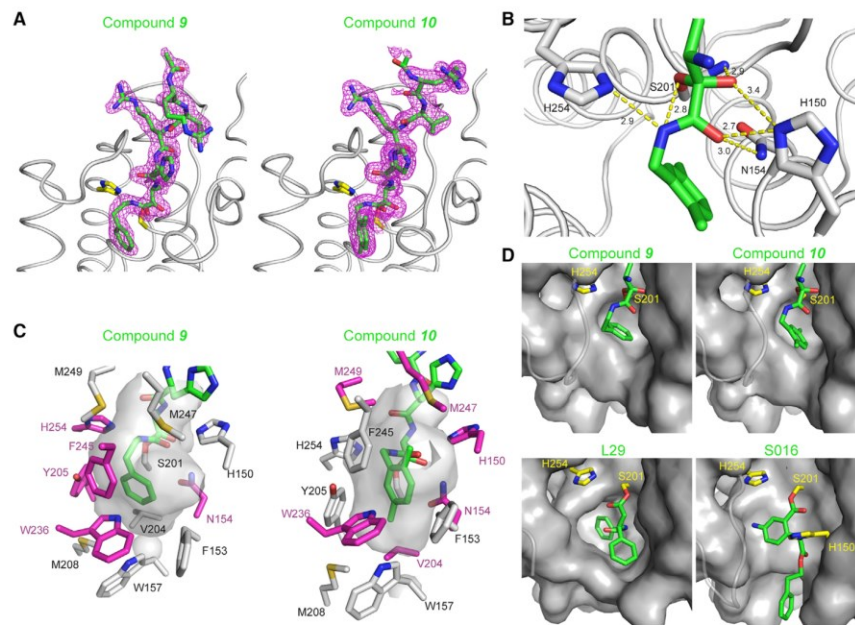
Figure 7. N-Substituted Peptidyl Ketoamides Bind GlpG in a Substrate-like Manner and Occupy the S4 to S2′ Subsites of the Rhomboid Active Site

(A) Electron density map and ligand stick model of **9** and **10** in complex with GlpG. Compounds **9** and **10** were soaked into the native crystals of GlpG, and the structures of the complexes were solved by X-ray diffraction (for statistics, see Table S1). The catalytic dyad is shown as yellow sticks and the inhibitors as green sticks surrounded by the $2mF_o - DF_c$ electron density map contoured at 1σ and shown 1.6 Å around the inhibitor model. In the complex of **9**, the electron density for the Arg residue in the P5 position was weaker, and the side chain has been modeled in a different conformation than in the complex of **10**, which was solved to a higher resolution and where the side chain of the Arg in the P5 position is defined clearly.

(B) Hydrogen bond engagement by the warhead of compound **10** in the active site of GlpG was analyzed using the HBplus program (McDonald and Thornton, 1994) implemented in Ligplot+ (Laskowski and Swindells, 2011) with default criteria (donor … acceptor [D … A] distance cutoff of 3.9 Å; donor … acceptor-acceptor antecedent [D … A-AA] angle of 90°).

(C) Interaction pattern of inhibitor tails in the S2′ site of GlpG. The cavity surrounding the tails of **9** and **10** is shown as an inverse surface, and the side chains lining the cavity are shown as sticks. The residues engaged in van der Waals interactions (identified by Ligplot+) with the tails of the inhibitors are shown in magenta.

(D) Comparison of binding modes of the S2′ binding moieties in compounds **9**, **10**, L29 (Pierrat et al., 2011), and S016 (Vosyka et al., 2013) in the respective complex structures PDB: 5MT6, 5MTF, 3ZMI, and 3ZEB. Protein is showed as a gray surface, catalytic dyad carbons in yellow, and ligand carbons in green. The L5 loop residues 245–250 are shown as semitransparent loops for clarity. All structures are oriented in the same way.

as the significantly larger groups of inhibitors L29 and S016. This alignment shows that the prime side of the GlpG active site is rather malleable, and larger or branched tails could be accommodated at the amide nitrogen of peptidyl ketoamides. This is likely to provide additional selectivity or binding energy and delineates one possible direction of further development of ketoamides as rhomboid inhibitors. The results presented here open the door to systematic development of rhomboid protease inhibitors in medically relevant contexts such as malaria (Baker et al., 2006; O'Donnell et al., 2006), Parkinson's disease (Chu, 2010; Meissner et al., 2015), and cancer (Song et al., 2015).

DISCUSSION

Here, we discover that peptidyl ketoamides bearing a substantial hydrocarbon modification at the ketoamide nitrogen are efficient

inhibitors of rhomboid intramembrane proteases, superior to any known rhomboid inhibitors in selectivity and by up to three orders of magnitude in potency. We also show that both of these properties are tunable by optimization of the peptide sequence and the character of the ketoamide "tail" substituent, defining a platform for the development of specific and potent rhomboid inhibitors. Since ketoamides are clinically used pharmacophores (Njoroge et al., 2008), our discovery of this pharmacologically compliant chemotype for rhomboid proteases enables the design of rhomboid inhibitors for cell biological and pharmacological use.

Structural analysis of peptidyl ketoamides complexed to GlpG reveals that they bind in a substrate-like manner, occupying the P4 to P2′ subsites (Figure 7A). The presence of residues in the P5 and P6 positions has been reported to improve the inhibition potency of peptidyl aldehydes significantly, but these residues

could not be observed in any co-crystal structures (Cho et al., 2016). We do observe weak electron density for the side chain of Arg in the P5 position, but its conformation in the final crystallographic models of the complexes of **9** and **10** differs, indicating some degree of flexibility, and it is probably influenced by crystal contacts. In addition, the P5 residue does not contribute significantly to the inhibition potency of **9** (Figure 3B) and is thus dispensable.

The binding mode of peptidyl ketoamides suggests that they can access the rhomboid active site from bulk solvent, and probably do not need prior partitioning into the membrane. They are covalent (Figure 7) and reversible (Figure 4C), and their kinetics of binding to rhomboid is adequately described by a one-step slow-binding mechanism (Figure 4D). Their inhibition modality is non-competitive (Copeland, 2013a) (Figure 4E), implying that they can bind to the free enzyme as well as the docking/interrogation complex during rhomboid catalysis (Strisovsky, 2016a). This is consistent with the proposed mechanism of inhibition of rhomboid protease GlpG by peptidyl aldehydes (Cho et al., 2016).

For the development of peptidyl ketoamides as rhomboid inhibitors, subsite preferences of the given rhomboid protease must be mapped efficiently. This could be achieved using classical positional scanning peptide libraries starting from a known substrate sequence. Given that the effects of the amino acids in the P5–P1 positions are additive (Figure 1A), the optimal substrate could be formed by combining the single subsite preferences identified in the positional scan. An alternative method for mapping subsite preferences at both the prime and nonprime sides could be multiplex substrate profiling using designed peptide libraries and mass spectrometry (O'Donoghue et al., 2012), although its application to rhomboids has not been tested yet.

The second module determining the potency and selectivity is the tail substituent at the ketoamide nitrogen. Here, the effects of flexibility versus rigidity, branching, and polarity of the substituents need to be investigated to explore the available chemical and conformational space. A more speculative direction of further improvement of the inhibitors may involve cyclization via the tail substituent and the P2 residue, which seems sterically possible and unobstructive in the enzyme-inhibitor complex (Figure 7A). Such cyclization could improve the potency of the inhibitor by conformationally restricting it near the bound conformation.

Finally, peptidyl ketoamides have been used clinically to treat hepatitis C infection (boceprevir, telaprevir) (Njoroge et al., 2008), suggesting that both the intracellular availability and metabolic stability of rhomboid-targeting peptidyl ketoamides can most likely be modified for compliance with pharmacological needs. The potential of rhomboid inhibitors in pharmacologically relevant settings has yet to be proven, but it currently seems that inhibitors of *Plasmodium* rhomboids might be therapeutic for malaria (Baker et al., 2006; Lin et al., 2013), inhibitors of the human mitochondrial rhomboid protease PARL might stimulate mitophagy (Meissner et al., 2015) and thus be disease-modifying in the context of Parkinson's disease (Chan and McQuibban, 2013), and inhibitors of human RHBDL4 could be targeting EGF receptor signaling by transforming growth factor α in colorectal cancer (Song et al., 2015). Specific rhomboid protease inhibitors such

as those that we describe here will serve as key tools for the validation and exploitation of these and other upcoming therapeutic opportunities involving rhomboid proteases.

SIGNIFICANCE

Intramembrane proteases of the rhomboid family are widely conserved and have been implicated in malaria, colon cancer, and Parkinson's disease. They represent potentially attractive drug targets, but until now, no specific, potent, and pharmacologically compatible inhibitors have been available. Here, we discover that peptidyl ketoamides are the first such potent and specific inhibitors of rhomboid proteases, and we delineate a general modular way for their design against diverse rhomboid enzymes. This discovery can have a broad impact on the cell biology of rhomboid proteases and on drug discovery targeting this family of enzymes in the context of infectious diseases, cancer, and neurodegeneration.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENTS AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Constructs and Cloning
  - Protein Expression and Purification
  - Chemical Synthesis
  - Protein Crystallography
  - Rhomboid Activity and Inhibition Assays
  - Inhibitor Selectivity Profiling
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPORTING CITATIONS

The following references appear in the Supplemental Information: Bastiaans et al., 1997; Cao et al., 2010; Coste et al., 1994; D'Andrea and Scola, 2008; Dondoni and Perrone, 1993; Semple et al., 2000; Souček and Urban, 1995; Tulla-Puche et al., 2008; Venkatraman et al., 2006; Yin et al., 2007.

REFERENCES

Ahlrichs, R., Bär, M., Häser, M., Horn, H., and Kölmel, C. (1989). Electronic structure calculations on workstation computers: the program system turbomole. Chem. Phys. Lett. 162, 165–169.

Bachovchin, D.A., Koblan, L.W., Wu, W., Liu, Y., Li, Y., Zhao, P., Woznica, I., Shu, Y., Lai, J.H., Poplawski, S.E., et al. (2014). A high-throughput, multiplexed assay for superfamily-wide profiling of enzyme activity. Nat. Chem. Biol. 10, 656–663.

Baker, R.P., Wijetilaka, R., and Urban, S. (2006). Two Plasmodium rhomboid proteases preferentially cleave different adhesins implicated in all invasive stages of malaria. PLoS Pathog. 2, e113.

Bastiaans, H.M.M., vanderBaan, J.L., and Ottenheijm, H.C.J. (1997). Flexible and convergent total synthesis of cyclotheonamide B. J. Org. Chem. 62, 3880–3889.

Cao, H., Liu, H., and Domling, A. (2010). Efficient multicomponent reaction synthesis of the schistosomiasis drug praziquantel. Chemistry 16, 12296–12298.

Chan, E.Y., and McQuibban, G.A. (2013). The mitochondrial rhomboid protease: its rise from obscurity to the pinnacle of disease-relevant genes. Biochim. Biophys. Acta 1828, 2916–2925.

Chatterjee, S., Dunn, D., Tao, M., Wells, G., Gu, Z.Q., Bihovsky, R., Ator, M.A., Siman, R., and Mallamo, J.P. (1999). P2-achiral, P'-extended alpha-ketoamide inhibitors of calpain I. Bioorg. Med. Chem. Lett. 9, 2371–2374.

Cho, S., Dickey, S.W., and Urban, S. (2016). Crystal structures and inhibition kinetics reveal a two-stage catalytic mechanism with drug design implications for rhomboid proteolysis. Mol. Cell 61, 329–340.

Chu, C.T. (2010). A pivotal role for PINK1 and autophagy in mitochondrial quality control: implications for Parkinson disease. Hum. Mol. Genet. 19, R28–R37.

Copeland, R.A. (2013a). Reversible modes of inhibitor interactions with enzymes. In Evaluation of Enzyme Inhibitors in Drug Discovery (John Wiley), pp. 57–121.

Copeland, R.A. (2013b). Slow binding inhibitors. In Evaluation of Enzyme Inhibitors in Drug Discovery (John Wiley), pp. 203–244.

Coste, J., Frerot, E., and Jouin, P. (1994). Coupling N-methylated amino-acids using pybrop and pyclop halogenophosphonium salts - mechanism and fields of application. J. Org. Chem. 59, 2437–2446.

D'Andrea, S., and Scola, P.M. (2008). Inhibitors of hepatitis C virus. Bristol-Myers Squibb Company. US patent US2008107623 (A1), filed October 25, 2007, and published May 8, 2008.

Dickey, S.W., Baker, R.P., Cho, S., and Urban, S. (2013). Proteolysis inside the membrane is a rate-governed reaction not driven by substrate affinity. Cell 155, 1270–1281.

Dondoni, A., and Perrone, D. (1993). 2-Thiazolyl alpha-amino ketones - a new class of reactive intermediates for the stereocontrolled synthesis of unusual amino-acids. Synthesis (Stuttg), 1162–1176.

Drag, M., and Salvesen, G.S. (2010). Emerging principles in protease-based drug discovery. Nat. Rev. Drug Discov. 9, 690–701.

Eggert, U.S., Ruiz, N., Falcone, B.V., Branstrom, A.A., Goldman, R.C., Silhavy, T.J., and Kahne, D. (2001). Genetic basis for activity differences between vancomycin and glycolipid derivatives of vancomycin. Science 294, 361–364.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. Sect. D Biol. Crystallogr. 60, 2126–2132.

Evans, P.R. (2011). An introduction to data reduction: space-group determination, scaling and intensity statistics. Acta Crystallogr. Sect. D Biol. Crystallogr. 67, 282–292.

Fanfrlik, J., Holub, J., Ruzickova, Z., Rezac, J., Lane, P.D., Wann, D.A., Hnyk, D., Ruzicka, A., and Hobza, P. (2016). Competition between halogen, hydrogen and dihydrogen bonding in brominated carboranes. ChemPhysChem 17, 3373–3376.

Fleig, L., Bergbold, N., Sahasrabudhe, P., Geiger, B., Kaltak, L., and Lemberg, M.K. (2012). Ubiquitin-dependent intramembrane rhomboid protease promotes ERAD of membrane proteins. Mol. Cell 47, 558–569.

Gerwig, J., Kiley, T.B., Gunka, K., Stanley-Wall, N., and Stulke, J. (2014). The protein tyrosine kinases EpsB and PtkA differentially affect biofilm formation in Bacillus subtilis. Microbiology 160, 682–691.

Gibson, D.G. (2011). Enzymatic assembly of overlapping DNA fragments. Methods Enzymol. 498, 349–361.

Grimme, S. (2006). Semiempirical GGA-type density functional constructed with a long-range dispersion correction. J. Comput. Chem. 27, 1787–1799.

Harper, J.W., and Powers, J.C. (1985). Reaction of serine proteases with substituted 3-alkoxy-4-chloroisocoumarins and 3-alkoxy-7-amino-4-chloroisocoumarins: new reactive mechanism-based inhibitors. Biochemistry 24, 7200–7213.

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

Hedstrom, L. (2002). Serine protease mechanism and specificity. Chem. Rev. 102, 4501–4524.

Jensen, F. (2006). Introduction to Computational Chemistry (Wiley).

Kabsch, W. (2010). Xds. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 125–132.

Klamt, A., and Schüürmann, G. (1993). COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. J. Chem. Soc. Perkin Trans. 2, 799–805.

Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr. Sect. D Biol. Crystallogr. 60, 2256–2268.

Laskowski, R.A., and Swindells, M.B. (2011). LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. J. Chem. Inf. Model 51, 2778–2786.

Lebedev, A.A., Young, P., Isupov, M.N., Moroz, O.V., Vagin, A.A., and Murshudov, G.N. (2012). JLigand: a graphical tool for the CCP4 template-restraint library. Acta Crystallogr. Sect. D Biol. Crystallogr. 68, 431–440.

76

Lee, C., Kang, H.J., Hjelm, A., Qureshi, A.A., Nji, E., Choudhury, H., Beis, K., de Gier, J.W., and Drew, D. (2014). MemStar: a one-shot *Escherichia coli*-based approach for high-level bacterial membrane protein production. FEBS Lett. *588*, 3761–3769.

Lemberg, M.K., Menendez, J., Misik, A., Garcia, M., Koth, C.M., and Freeman, M. (2005). Mechanism of intramembrane proteolysis investigated with purified rhomboid proteases. EMBO J. *24*, 464–472.

Lin, J.W., Meireles, P., Prudencio, M., Engelmann, S., Annoura, T., Sajid, M., Chevalley-Maurel, S., Ramesar, J., Nahar, C., Avramut, C.M., et al. (2013). Loss-of-function analyses defines vital and redundant functions of the *Plasmodium* rhomboid protease family. Mol. Microbiol. *88*, 318–338.

Liu, Y., Stoll, V.S., Richardson, P.L., Saldivar, A., Klaus, J.L., Molla, A., Kohlbrenner, W., and Kati, W.M. (2004). Hepatitis C NS3 protease inhibition by peptidyl-alpha-ketoamide inhibitors: kinetic mechanism and structure. Arch. Biochem. Biophys. *421*, 207–216.

McCoy, A.J. (2007). Solving structures of protein complexes by molecular replacement with Phaser. Acta Crystallogr. Sect. D Biol. Crystallogr. *63*, 32–41.

McDonald, I.K., and Thornton, J.M. (1994). Satisfying hydrogen bonding potential in proteins. J. Mol. Biol. *238*, 777–793.

Meissner, C., Lorenz, H., Hehn, B., and Lemberg, M.K. (2015). Intramembrane protease PARL defines a negative regulator of PINK1- and PARK2/Parkin-dependent mitophagy. Autophagy *11*, 1484–1498.

Miroux, B., and Walker, J.E. (1996). 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*, 289–298.

Mitchell, E.M., Artymiuk, P.J., Rice, D.W., and Willett, P. (1990). Use of techniques derived from graph-theory to compare secondary structure motifs in proteins. J. Mol. Biol. *212*, 151–166.

Morrison, J.F. (1982). The slow-binding and slow, tight-binding inhibition of enzyme-catalysed reactions. Trends Biochem. Sci. *7*, 102–105.

Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. Sect. D Biol. Crystallogr. *67*, 355–367.

Nguyen, M.T., Van Kersavond, T., and Verhelst, S.H. (2015). Chemical tools for the study of intramembrane proteases. ACS Chem. Biol. *10*, 2423–2434.

Njoroge, F.G., Chen, K.X., Shih, N.Y., and Piwinski, J.J. (2008). Challenges in modern drug discovery: a case study of boceprevir, an HCV protease inhibitor for the treatment of hepatitis C virus infection. Acc. Chem. Res. *41*, 50–59.

O'Donnell, R.A., Hackett, F., Howell, S.A., Treeck, M., Struck, N., Krnajski, Z., Withers-Martinez, C., Gilberger, T.W., and Blackman, M.J. (2006). Intramembrane proteolysis mediates shedding of a key adhesin during erythrocyte invasion by the malaria parasite. J. Cell Biol. *174*, 1023–1033.

O'Donoghue, A.J., Eroy-Reveles, A.A., Knudsen, G.M., Ingram, J., Zhou, M., Statnekov, J.B., Greninger, A.L., Hostetter, D.R., Qu, G., Maltby, D.A., et al. (2012). Global identification of peptidase specificity by multiplex substrate profiling. Nat. Methods *9*, 1095–1100.

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

Powers, J.C., Kam, C.M., Narasimhan, L., Oleksyszyn, J., Hernandez, M.A., and Ueda, T. (1989). Mechanism-based isocoumarin inhibitors for serine proteases: use of active site structure and substrate specificity in inhibitor design. J. Cell. Biochem. *39*, 33–46.

Powers, J.C., Asgian, J.L., Ekici, O.D., and James, K.E. (2002). Irreversible inhibitors of serine, cysteine, and threonine proteases. Chem. Rev. *102*, 4639–4750.

Purich, D.L. (2010). Kinetic behavior of enzyme inhibitors. In Enzyme Kinetics: Catalysis & Control (Elsevier), pp. 485–574.

Ruiz, N., Falcone, B., Kahne, D., and Silhavy, T.J. (2005). Chemical conditionality. Cell *121*, 307–317.

Schechter, I., and Berger, A. (1967). On the size of the active site in proteases. I. Papain. Biochem. Biophys. Res. Commun. *27*, 157–162.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675.

Schrodinger. (2012). The PyMOL Molecular Graphics System, Version 1.5.0.4 (Schrödinger LLC).

Semple, J.E., Owens, T.D., Nguyen, K., and Levy, O.E. (2000). New synthetic technology for efficient construction of alpha-hydroxy-beta-amino amides via the Passerini reaction. Org. Lett. *2*, 2769–2772.

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

Singh, J., Petter, R.C., Baillie, T.A., and Whitty, A. (2011). The resurgence of covalent drugs. Nat. Rev. Drug Discov. *10*, 307–317.

Song, W., Liu, W., Zhao, H., Li, S., Guan, X., Ying, J., Zhang, Y., Miao, F., Zhang, M., Ren, X., et al. (2015). Rhomboid domain containing 1 promotes colorectal cancer growth through activation of the EGFR signalling pathway. Nat. Commun. *6*, 8022.

Souček, M., and Urban, J. (1995). An efficient method for preparation of optically active N-protected α-amino aldehydes from N-protected α-amino alcohols. Collect. Czechoslov. Chem. Commun. *60*, 693–696.

Stevenson, L.G., Strisovsky, K., Clemmer, K.M., Bhatt, S., Freeman, M., and Rather, P.N. (2007). Rhomboid protease AarA mediates quorum-sensing in *Providencia stuartii* by activating TatA of the twin-arginine translocase. Proc. Natl. Acad. Sci. USA *104*, 1003–1008.

Strisovsky, K. (2013). Structural and mechanistic principles of intramembrane proteolysis - lessons from rhomboids. FEBS J. *280*, 1579–1603.

Strisovsky, K. (2016a). Rhomboid protease inhibitors: emerging tools and future therapeutics. Semin. Cell Dev. Biol. *60*, 52–62.

Strisovsky, K. (2016b). Why cells need intramembrane proteases - a mechanistic perspective. FEBS J. *283*, 1837–1845.

Strisovsky, K., Sharpe, H.J., and Freeman, M. (2009). Sequence-specific intramembrane proteolysis: identification of a recognition motif in rhomboid substrates. Mol. Cell *36*, 1048–1059.

Ticha, A., Stanchev, S., Skerle, J., Began, J., Ingr, M., Svehlova, K., Polovinkin, L., Ruzicka, M., Bednarova, L., Hadravova, R., et al. (2017). Sensitive versatile fluorogenic transmembrane peptide substrates for rhomboid intramembrane proteases. J. Biol. Chem. *292*, 2703–2713.

Tulla-Puche, J., Torres, A., Calvo, P., Royo, M., and Albericio, F. (2008). N,N,N',N'-Tetramethylchloroformamidinium hexafluorophosphate (TCFH), a powerful coupling reagent for bioconjugation. Bioconj. Chem. *19*, 1968–1971.

Venkatraman, S., Bogen, S.L., Arasappan, A., Bennett, F., Chen, K., Jao, E., Liu, Y.T., Lovey, R., Hendrata, S., Huang, Y., et al. (2006). Discovery of (1R,5S)-N-[3-amino-1-(cyclobutylmethyl)-2,3-dioxopropyl]- 3-[2(S)-[[[(1,1-dimethylethyl)amino]carbonyl]amino]-3,3-dimethyl-1-oxobutyl]- 6,6-dimethyl-3-azabicyclo[3.1.0]hexan-2(S)-carboxamide (SCH 503034), a selective, potent, orally bioavailable hepatitis C virus NS3 protease inhibitor: a potential therapeutic agent for the treatment of hepatitis C infection. J. Med. Chem. *49*, 6074–6086.

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

Vinothkumar, K.R., Pierrat, O.A., Large, J.M., and Freeman, M. (2013). Structure of rhomboid protease in complex with beta-lactam inhibitors defines the S2' cavity. Structure *21*, 1051–1058.

Vosyka, O., Vinothkumar, K.R., Wolf, E.V., Brouwer, A.J., Liskamp, R.M., and Verhelst, S.H. (2013). Activity-based probes for rhomboid proteases discovered in a mass spectrometry-based assay. Proc. Natl. Acad. Sci. USA *110*, 2472–2477.

Walker, B., and Lynas, J.F. (2001). Strategies for the inhibition of serine proteases. Cell. Mol. Life Sci. *58*, 596–624.

Wang, Y., and Ha, Y. (2007). Open-cap conformation of intramembrane protease GlpG. Proc. Natl. Acad. Sci. USA *104*, 2098–2102.

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

Wolf, E.V., Zeissler, A., Vosyka, O., Zeiler, E., Sieber, S., and Verhelst, S.H. (2013). A new class of rhomboid protease inhibitors discovered by activity-based fluorescence polarization. PLoS One *8*, e72307.

Wolf, E.V., Zeissler, A., and Verhelst, S.H. (2015). Inhibitor fingerprinting of rhomboid proteases by activity-based protein profiling reveals inhibitor selectivity and rhomboid autoprocessing. ACS Chem. Biol. *10*, 2325–2333.

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

Yin, J., Gallis, C.E., and Chisholm, J.D. (2007). Tandem oxidation/halogenation of aryl allylic alcohols under Moffatt-Swern conditions. J. Org. Chem. *72*, 7054–7057.

Zoll, S., Stanchev, S., Began, J., Skerle, J., Lepsik, M., Peclinovska, L., Majer, P., and Strisovsky, K. (2014). Substrate binding and specificity of rhomboid intramembrane protease revealed by substrate-peptide complex structures. EMBO J. *33*, 2408–2421.

78

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit anti-DYKDDDDK	Cell Signaling Technology	Cat#2368
Monoclonal ANTI-FLAG® M2 antibody produced in mouse	Sigma	Cat#F1804
Donkey anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, DyLight 800	Invitrogen	Cat#SA5-10044
Bacterial and Virus Strains		
E. coli NR698	Laboratory of Tom Silhavy (Princeton)	(Ruiz et al., 2005)
E. coli NR698$\bar{A}$glpG::tet	Laboratory of Matthew Freeman (Oxford)	(Pierrat et al., 2011)
Bacillus subtilis 168	Bacillus Genetic Stock Center	
Bacillus subtilis 168 *ydcA::neo*	This work	BS2
Bacillus subtilis 168 *ydcA::neo, yqgP::tet*	This work	BS4
Bacillus subtilis 168 *ydcA::neo, xdkE::MBP-LacYTM2-Trx(erm, lin)*	This work	BS87
Bacillus subtilis 168 *ydcA::neo, yqgP::tet, xdkE::MBP-LacYTM2-Trx(erm, lin)*	This work	BS88
E. coli C41(DE3)	Lucigen	Cat#60452-1
Chemicals, Peptides, and Recombinant Proteins		
h-KRHDIN(E-edans)ISKSDTG(K-dabcyl) IFAAISLFSLLFQPLFGLSKK-nh$_2$	(Ticha et al., 2017)	KSp35
h-KRHRI(E-edans)RVRHADTG(K-dabcyl) IFAAISLFSLLFQPLFGLSKKR-nh$_2$	This work	KSp93
h-KRHRI(K-tamra)RVRHADTG(C-qxl610) IFAAISLFSLLFQPLFGLSKKR-nh$_2$	(Ticha et al., 2017)	KSp64
h-KRHRINRVR(E-edans)ADTG(K-dabcyl) IFAAISLFSLLFQPLFGLSKKR-nh$_2$	This work	KSp96
TAMRA-XP	Thermo Fisher Scientific	Cat#88318
Revert Total Protein Stain Kit	LI-COR, Inc.	Cat#926-11010
3,4-dichloroisocoumarin (DCI)	Sigma	D7910
AcRVRHAcmk	This paper	
AcVRHAcmk	This paper	
AcIATAcmk	Zoll et al., 2014	
AcVATAcmk	This paper	
AcIAHAcmk	This paper	
AcIRTAcmk	This paper	
AcRVRHA-trifluoromethylketone	This paper	
AcRVRHA-boronate	This paper	
AcRVRHA-acylsulfonamide	This paper	
AcRVRHA-thiazolylketone	This paper	
AcRVRHA-CONH$_2$	This paper	
Ac-RVRHA-CONH-isopropyl	This paper	
Ac-RVRHA-CONH-cyclohexyl	This paper	
Ac-RVRHA-CONH-benzyl	This paper	
Ac-RVRHA-CONH-methylbenzyl	This paper	
Ac-RVRHA-CONH-ethyloxyacetyl	This paper	

(Continued on next page)

79

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Ac-RVRHA-CONH-neopentyl	This paper	
Ac-RVRHA-CONH-pentyl	This paper	
Ac-RVRHA-CONH-phenylethyl	This paper	
Ac-RVRHA-CONH-(2,4-dimethyl)benzyl	This paper	
Ac-RVRHA-CONH-phenylbutyl	This paper	
Ac-RHA-CONH-phenylethyl	This paper	
Ac-RHA-CONH-phenylethyl	This paper	
Ac-A-CONH-phenylethyl	This paper	
Deposited Data		
Crystal structure of GlpG bound to AcRVRHAcmk	This paper	PDB: 5MT8
Crystal structure of GlpG bound to AcVRHAcmk	This paper	PDB: 5MT7
Crystal structure of GlpG bound to Ac-RVRHA-CONH-phenylethyl	This paper	PDB: 5MT6
Crystal structure of GlpG bound to Ac-RVRHA-CONH-(2,4-dimethyl)benzyl	This paper	PDB: 5MTF
Recombinant DNA		
pMALp2E_MBP-LacYTM2-Trx	(Zoll et al., 2014)	pKS506
pD881-SR	DNA2.0 Inc.	
pD881-SR_MBP-LacYTM2-Trx	This work	pPR61
pET25b+M_GlpG	(Lemberg et al., 2005)	
pGP886_pxyl-AmyE$_{SP}$-MBP-FLAG-LacYTM2-Trx-HA	This work	pPR200
Software and Algorithms		
XDS	(Kabsch, 2010)	http://xds.mpimf-heidelberg.mpg.de/html_doc/downloading.html
AIMLESS	(Evans, 2011)	http://www.ccp4.ac.uk/html/aimless.html
COOT 0.8.6	(Emsley and Cowtan, 2004)	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
Jligand	(Lebedev et al., 2012)	http://www.ysbl.york.ac.uk/mxstat/JLigand/
Refmac_5.8.0158	(Murshudov et al., 2011)	
Ligplot+	(Laskowski and Swindells, 2011)	http://www.ebi.ac.uk/thornton-srv/software/LigPlus/download.html
Phaser	(McCoy, 2007)	http://www.ccp4.ac.uk/html/phaser.html
Pymol 1.8.4.0	Schrodinger, LLC	https://www.pymol.org/
Turbomole 7	(Ahlrichs et al., 1989)	http://www.turbomole.com/
ImageJ 1.49	(Schneider et al., 2012)	https://imagej.net/Welcome
Image Studio Lite Version 5.2	LI-COR, Inc.	https://www.licor.com/bio/products/software/image_studio_lite/download.html
GraphPad Prism 7.02	GraphPad Software, Inc.	https://www.graphpad.com/scientific-software/prism/
GraFit 7	Erithacus Software Ltd.	http://www.erithacus.com/grafit/

CONTACT FOR REAGENTS AND RESOURCE SHARING

Further information and requests for reagents may be directed to, and will be fulfilled by the corresponding author Kvido Strisovsky (kvido.strisovsky@uochb.cas.cz).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Escherichia coli K12 strain NR698 (Ruiz et al., 2005), which has the MC4100 background with the *imp4213* allele carried from BE100 (Eggert et al., 2001) (an in-frame deletion of amino acids 330-352 of LptD) is a gift of Dr. Tom Silhavy (Princeton University). A GlpG-free variant was created by deleting *glpG* using a tetracyclin marker (Pierrat et al., 2011).

To generate a rhomboid activity free *Bacillus subtilis*, the *ydcA::neo* mutant (BS2, this work) of the wild type B. *subtilis* 168 strain (Bacillus Genetic Stock Center, USA) was modified by deleting the entire the *yqgP* gene and replacing it with a tetracyclin resistance gene using homologous recombination, yielding strain BS4 (*ydcA::neo, yqgP::tet*). Both modifications were verified by genomic PCR of the disrupted locus and Sanger sequencing of the amplified region.

METHOD DETAILS

Constructs and Cloning

To generate a model rhomboid substrate for *in vivo* activity assays in E. *coli*, the MBP-LacYTM2-Trx-Stag-Histag construct was PCR-amplified from pKS506 (Ticha et al., 2017) and cloned into the SapI linearized plasmid pD881-SR (DNA2.0 Inc., Newark, USA) using isothermal assembly (Gibson, 2011), yielding construct pPR61. For expression in B. *subtilis*, the substrate was modified by replacing the MBP signal peptide by the signal peptide from B. *subtilis* AmyE, and the AmyE$_{sp}$-MBP$_{mat}$-FLAG-LacYTM2-Trx-HA construct was cloned into the XbaI, SalI digested plasmid pGP886 (Gerwig et al., 2014) (gift of Dr. Libor Krasny, Prague, CR) using isothermal assembly (Gibson, 2011) to yield construct pPR200. This construct was linearized by ScaI and integrated into the *xkdE* locus of BS2 and BS4 using erythromycin-lincomycin selection yielding strains BS87 (BS2 *xkdE::Pxyl-LacYTM2(erm)*) and BS88 (BS4 *xkdE::Pxyl-LacYTM2(erm)*).

Protein Expression and Purification

The E. *coli* GlpG for crystallisation was expressed in E. *coli* C41(DE3) (Miroux and Walker, 1996) in PASM 5052 medium as described (Lee et al., 2014). Membrane isolation, purification by metal affinity chromatography, cleavage by chymotrypsin to produce GlpG transmembrane core domain and gel filtration chromatography were carried out as described previously (Vinothkumar et al., 2010; Zoll et al., 2014). The E. *coli* GlpG for inhibition studies was expressed in E. *coli* C41(DE3) (Miroux and Walker, 1996) in LB medium, and solubilised and purified in 0.05% (w/v) DDM as described (Ticha et al., 2017). Other rhomboid proteases were expressed and purified as reported previously (Wolf et al., 2015).

Chemical Synthesis

All reagents were acquired from commercial sources and used without purification. Protected amino acids and amino acid derivatives were purchased from Iris Biotech (Marktredwitz, Germany), Sigma-Aldrich (St. Louis, MO, U.S.A), Thermo Fischer Scientific (Waltham, Massachusets, U.S.A) and Fluorochem (Hadfield, Derbyshire, UK). Further details on chemical syntheses as well as compound characterisation data by mass spectrometry and NMR are available as Supplemental Information (Methods S1).

Protein Crystallography

Crystals of truncated wild type GlpG apoenzyme were obtained by mixing a solution of 2 - 3 M ammonium chloride or sodium chloride, 0.1 M Bis-Tris, pH 7.0 with protein (4-6 mg/mL) at ratio of 1:1 in hanging drops at 22°C (Vinothkumar et al., 2010; Wang and Ha, 2007). Inhibitors were diluted from 10 mM stock solutions in anhydrous DMSO into buffer resembling the mother liquor to yield final 1 mM inhibitor and 10% DMSO just before soaking. For the chloromethylketone inhibitors, the crystals were incubated with inhibitors at 0.3-0.5 mM concentrations for 24 h. The ketoamide inhibitors were incubated at final concentrations of 0.3-0.5 mM for 30-120 min. All crystals were cryo-protected by adding 25% (v/v) glycerol to the mother liquor and flash frozen in liquid nitrogen.

Data sets of the CMKs and **9** were collected at the I02 beam line at the Diamond Light Source (Harwell) and the data set of **10** was collected at BESSY (Berlin, Germany). Diffraction data were indexed, integrated and scaled with XDS (Kabsch, 2010) and AIMLESS (Evans, 2011). For the structures with inhibitor bound, the coordinates of GlpG (PDB 2XOV) with residues 245-249 (of Loop 5) omitted were used as an input model for Phaser (McCoy, 2007). Restrained refinement was carried out with Refmac (Murshudov et al., 2011) followed by manual model building in COOT (Emsley and Cowtan, 2004). In the final step, TLS was used using the enzyme and the inhibitor peptide as one group (Murshudov et al., 2011). The model, library and link files of the inhibitors were generated with Jligand (Lebedev et al., 2012). In the structures of Ac-(R)VRHA-cmk, H150 was modelled to hydrogen bond to the chloromethylketone oxygen. An additional density was observed close to M149 and H150 raising the possibility that the H150 residue could be also in an alternative conformation, but modeling the alternative conformation or both conformations of H150 was not conclusive in explaining the density. Other similar datasets of CMKs obtained by soaking show that this density might perhaps represent a bound ion, but due to ambiguity we have left the density unmodelled.

In order to find the best possible fit of the molecules of **9** and **10** to the experimental electron densities, quantum mechanical calculations were performed. The model systems comprised the whole inhibitors in their tetrahedral intermediate form with methoxy group representing the S201 side-chain. These models were made in several variants: i) *cis/trans* isomers of the ketoamide proximal/distal carbonyls, ii) *cis/trans* isomers of the distal carbonyl/NH, and iii) different rotameric forms of the His side chain in the P2 position of the inhibitors. All these variants were optimized in Turbomole ver. 7 program (Ahlrichs et al., 1989) using DFT-D3 method (Grimme, 2006) at B-LYP/DZVP level (Fanfrlik et al., 2016; Jensen, 2006) and COSMO implicit solvent model (Klamt and Schüürmann, 1993). Their intrinsic stabilities were assessed by comparing the final energies, and the conformer with the lowest energy was built into the electron density and chosen as a model for crystallographic refinement.

Noncovalent interactions between the ligands and protein were detected using Ligplot+ (Laskowski and Swindells, 2011) and hydrogen bonds were defined by canonical geometrical criteria (Laskowski and Swindells, 2011; McDonald and Thornton, 1994).

Structural alignments and all structure figures were made with Pymol (Schrodinger, 2012). The coordinates of the structures presented in this manuscript have been deposited in the PDB under the following IDs: 5MT7 (Ac-VRHA-cmk), 5MT8 (Ac-RVRHA-cmk), 5MT6 (compound **9**) and 5MTF (compound **10**). Data collection and refinement statistics are listed in Table S1.

Rhomboid Activity and Inhibition Assays

The activity of GlpG *in vitro* was determined as reported (Ticha et al., 2017). Concentrations of stock solutions of peptide substrates and inhibitors were determined by quantitative amino acid analysis. The IC_{50} and reversibility measurements were performed in 20 mM HEPES, pH 7.4, 150 mM NaCl, 0.05%(w/v) DDM, 12%(v/v) DMSO, and other kinetic measurements in 50 mM potassium phosphate, pH 7.4, 150 mM NaCl, 0.05%(w/v) DDM, 10%(v/v) DMSO, 0.05%(w/v) PEG8000, and 20%(v/v) glycerol unless noted otherwise. The reaction mixture typically consisted of 10 μM fluorogenic peptide substrate and the measurements were performed without enzyme-inhibitor pre-incubation unless noted otherwise. Note that the fluorogenic substrates used in Figure 4 had nearly identical amino acid sequences but for the point of attachment of the fluorophore or the identity of the fluorophore and quencher (see Key Resources Table).

For measuring the inhibition of GlpG *in vivo*, the *E. coli* strain NR698 with genetically permeabilised outer membrane (Ruiz et al., 2005) and its *glpG* knock-out derivative KS69 (*glpG::tet*) were used as described (Pierrat et al., 2011) with the following modifications. The chimeric substrate encoding LacY transmembrane domain 2 inserted between maltose binding protein and thioredoxin (Strisovsky et al., 2009) was expressed under control of rhamnose promoter (construct pPR61). To evaluate the *in vivo* inhibition by ketoamides, the NR698 cells were inoculated to the density of $OD_{600} = 0.05$ and grown to $OD_{600} = 0.6$ at 37°C. The cells were then incubated with increasing concentrations of inhibitor for 15 min at room temperature, and expression of the chimeric substrate was induced by adding 1 mM L-rhamnose. Cells were grown for further 4 h at 25°C, after which steady-state level of substrate cleavage was evaluated by western blotting with near-infrared fluorescence detection as described (Ticha et al., 2017).

For measuring the inhibition of YqgP *in vivo*, the *B. subtilis* strains BS87 and its *yqgP* knock-out derivative BS88, generated in this work (see Constructs and Cloning section), were used as follows. The chimeric LacYTM2 substrate AmyE$_{sp}$-MBP-FLAG-LacYTM2-Trx-HA (this work) was expressed under control of xylose promoter from the *xdkE* genomic locus. Fresh LB medium, supplemented with appropriate antibiotic, was inoculated with a few colonies of the *B. subtilis* strain grown overnight on selective LB agar plate and pre-culture was grown for 2 h at 37°C to $OD_{600} = 1$. Pre-culture was then diluted with fresh LB medium to the density of $OD_{600} = 0.05$. At this point, the expression of LacYTM2 was induced by adding 1% (w/v) D-(+)-xylose (Sigma), rhomboid inhibitors were added at a range of concentrations, and the cultures were further incubated for 2.5 h at 37°C (reaching $OD_{600} \sim 1$). Steady-state conversion of the substrate was evaluated by western blotting with near-infrared fluorescence detection as described (Ticha et al., 2017), substracting the intensity of non-specific bands, closely co-migrating with the specific rhomboid-formed N-terminal cleavage product of the substrate.

Inhibitor Selectivity Profiling

For inhibitor selectivity profiling against rhomboid proteases (Wolf et al., 2015), 400 ng of a purified protein preparation of *E. coli* GlpG was diluted in 30 μL of reaction buffer (20 mM HEPES, pH 7.4, with 0.05% (w/v) DDM). For other rhomboids, amounts were taken that gave similar labeling intensity during profiling. Rhomboids were incubated for 30 min at room temperature with the indicated concentration of compound, 100 μM DCI as positive control, or an equal volume of DMSO as negative control. Next, TAMRA-FP serine hydrolase probe (Thermo Fisher #88318) was added to a final concentration of 1 ìM and incubated for 2 h at 37°C in the dark. The reaction was stopped by addition of 4× Laemmli buffer and the reaction mixture was resolved on 15% SDS-PAGE. Gels were scanned on a Typhoon Trio+ and analyzed using ImageJ. The intensity of each rhomboid protease band calculated by ImageJ was normalized against its corresponding DMSO-treated counterpart (100% activity) to indicate the residual activity left after inhibition. The remaining activity was used to calculate the percentage of inhibition depicted in the heatmap. Selectivity profiles against human serine hydrolases were determined by EnPlex as described previously (Bachovchin et al., 2014).

QUANTIFICATION AND STATISTICAL ANALYSIS

Enzyme kinetics and inhibition data were analysed in GraphPad Prism v7.02 using in-built algorithms. Means and standard deviations have been derived from the best fit of the data, or based on three independent measurements, as specified, unless noted otherwise. Quantitative western blots were evaluated using near infrared detection with the IRDye 800CW secondary antibody on a LiCor Odyssey CLx infrared scanner with normalisation to total protein using the Revert total protein stain (LiCor).

DATA AND SOFTWARE AVAILABILITY

All crystallographic coordinates of the protein structures presented in this manuscript have been deposited in and will be freely available from the Protein Data Bank (www.rcsb.org) under the following identifiers: 5MT7, 5MT8, 5MT6 and 5MTF.

6.4 Membrane protein dimerization in cell-derived lipid membranes measured by FRET with MC simulations

Background

Limited information is available on the behaviour of proteins from rhomboid-like superfamily in native biomembranes. It has been published that rhomboid proteases dimerize in detergent micelles, which results in their allosteric activation [115], but information about their aggregation state in biomembranes has been absent. By combining several biophysical methods employing fluorescence reporters attached to proteins, we focused on both strong and weak interactions and we also excluded the potential effect of varying lipid composition in different cell compartments.

Summary

We used Foester resonance energy transfer (FRET) and fluorescence correlation spectroscopy (FCCS) combined with Monte Carlo (MC) simulation to address dimerization of human rhomboid protease RHBDL2 in its native biomembrane. Using this non-invasive approach, we found no evidence of rhomboid dimerization in membrane. By adapting methods commonly used for soluble proteins to 2D membrane environment, we developed a novel approach for membrane protein dimerization studies. To study the oligomeric state of RHBDL2 during its maturation in ER, our approach was also complemented by colocalization analysis of fluorescently labeled proteins (Figure 21).
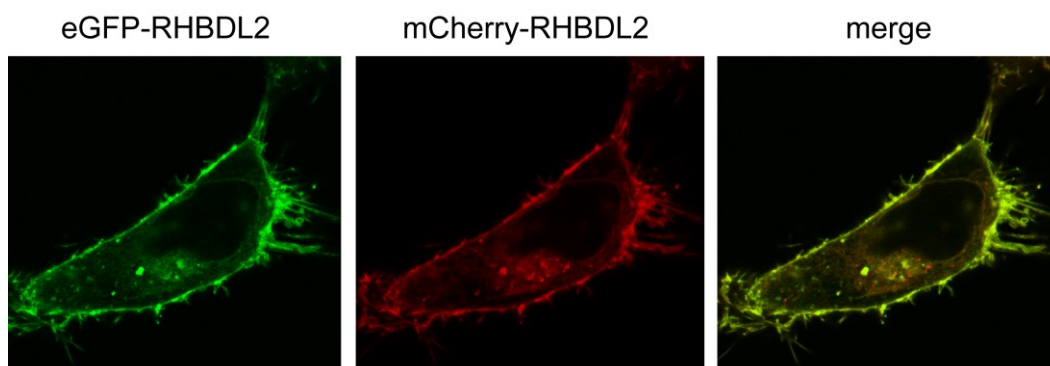
Figure 21: Localization of human rhomboid RHBDL2 in fusion with fluorescent protein on the plasma membrane.

My contribution

I participated in planning and design of some of the experiments. I also designed and cloned most of the used constructs and performed a majority of the biological experiments. I participated in data analysis and writing of the manuscript.

Membrane protein dimerization in cell-derived lipid membranes measured by FRET with MC simulations. Jan Škerle, Jana Humpolíčková, Nicholas Johnson, Petra Rampírová, Edita Poláchová, Monika Fliegl, Jan Dohnálek, Anna Suchánková, David Jakubec, and Kvido Strisovsky (2020), Biophysical Journal, accepted

Title

Membrane protein dimerization in cell-derived lipid membranes measured by FRET with MC simulations

Authors

J. Škerle[$,1,2], J. Humpolíčková[$,1,*], N. Johnson[1], P. Rampírová[1], E. Poláchová[1,3], M. Fliegl[1], J. Dohnálek[1,4], A. Suchánková[1], D. Jakubec[1], and K. Strisovsky[1,*]

[1]Institute of Organic Chemistry and Biochemistry, Czech Academy of Science, Flemingovo n. 2, Prague, 166 10, Czech Republic

[2]Department of Biochemistry, Faculty of Science, Charles University, Hlavova 2030/8, Prague, 128 43, Czech Republic

[3]First Faculty of Medicine, Charles University, Kateřinská 32, Prague, 121 08, Czech Republic

[4]University of Chemistry and Technology Prague, Technická 5, Prague, 166 28, Czech Republic

[$]Equal contribution

*Authors for correspondence

Running Title

Rhomboid protease is monomeric in lipids

Keywords

fluorescence, single-molecule, energy transfer, dimerization, membrane protein, lipid membrane

Abstract

Many membrane proteins are thought to function as dimers or higher oligomers, but measuring membrane protein oligomerization in lipid membranes is particularly challenging. Förster resonance energy transfer (FRET) and fluorescence cross-correlation spectroscopy (FCCS) are non-invasive, optical methods of choice that have been applied to the analysis of dimerization of single-spanning membrane proteins. However, the effects inherent to such two-dimensional systems, such as excluded volume of polytopic transmembrane proteins, proximity FRET, and rotational diffusion of fluorophore dipoles, complicate interpretation of FRET data and have not been typically accounted for. Here, using FRET and FCCS we introduce a method to measure surface protein density and to estimate the apparent Förster radius, and we use Monte Carlo (MC) simulations of the FRET data to account for the proximity FRET effect occurring in confined 2D environments. We then use FRET to analyze the dimerization of human rhomboid protease RHBDL2 in giant plasma membrane vesicles (GPMVs). We find no evidence for stable oligomers of RHBDL2 in GPMVs of human cells even at concentrations that highly exceed endogenous expression levels. This indicates that the rhomboid transmembrane core is intrinsically monomeric. Our findings will find use in the application of FRET and FCS for the analysis of oligomerization of transmembrane proteins in cell derived lipid membranes.

Statement of Significance

Membrane proteins often function as dimers or higher oligomers. Measuring membrane protein oligomerization in native lipid membranes is thus very important, but it is also particularly challenging. Here we address key problems and develop a method to measure surface protein density and to estimate the apparent Förster radius, and we use Monte Carlo simulations of the FRET data to account for the proximity FRET effect occurring in confined 2D environments. Using this method we find that rhomboid protease RHBDL2 is monomeric in lipid membranes. Our findings will find use in the application of FRET and FCS for the analysis of oligomerization of transmembrane proteins in native cellular lipid membranes.

Introduction

Membrane proteins frequently form functionally important homo or hetero-oligomeric complexes, and their surrounding milieu, i.e. lipid membrane, can shape their properties and interactions profoundly (1). It is thus important to study membrane protein interactions directly in their native membranes in a minimally invasive way. Optical, fluorescence-based methods appear ideal for this purpose. Fluorescence resonance energy transfer (FRET) is a powerful technique for measuring distances at the nanometer scale and thus it can potentially address protein oligomerization. FRET reports on the vicinity of protein that carries an energy donor in the presence of an acceptor, typically provided by fluorescent protein reporters, fused genetically to the protein of interest. In the three dimensional space (in solution), FRET response is observed almost exclusively when the proteins of interest are indeed physically interacting in oligomers. However, in a two dimensional confinement of membranes, protein density can be elevated to such a level that even if the two proteins do not physically interact in a complex, energy can be transferred to multiple acceptor molecules that are in close proximity of the donor (2). Studying protein oligomers within cellular lipid membranes by FRET is therefore ideal for situations of low protein density, i.e. the interaction has to be rather strong. Measuring dissociation constants of weaker interactions requires higher protein surface density, which results in increased FRET 'background' arising merely from the proximity of non-interacting fluorescence acceptors (3-6). We aimed at analyzing and compensating these and other limitations to allow usage of fluorescence techniques for the study of oligomerization of polytopic membrane proteins spanning a wide range of dissociation constants in live cell derived lipid membranes.

Already under mild overexpression conditions, common in cell-biological experiments, the 'proximity-induced FRET' effect mentioned above may become significant, and it has to be properly taken into account to avoid spurious results. This problem can be ameliorated by the use of FRET controls, i.e. validated known interacting and non-interacting protein pairs. In addition to that, protein density level in the membrane has to be evaluated and only samples with similar surface densities of donor and acceptor labelled proteins should be mutually compared. In most membrane structures of living cells, direct protein surface density determination is not feasible, since the membrane area from which the fluorescence signal is collected cannot be measured. Therefore, fluorescence intensity of the donor and acceptor is used instead for relative comparison (7). Even such a comparison, however, requires careful interpretation because the level of FRET is influenced not only by

oligomerization and density-induced proximity, but also possibly by the length of the linker between the protein of interest and the fluorescent probe. In addition, it is affected by the excluded area of studied proteins (8). Importantly, global protein surface density at the membrane can be determined in spherical giant plasma membrane-derived vesicles (GPMVs) formed from cells expressing proteins of interest (9-12). Finally, recent work analyzing the FRET efficiency signatures of non-interacting membrane proteins both experimentally and by Monte Carlo (MC) simulations indicates how the overall FRET efficiency should be corrected for the 'proximity-induced FRET' effect mentioned above (2).

We show that if FRET is combined with quantification of lateral concentrations of the proteins and combined with MC simulations, the true interaction can be distinguished from the proximity effect, and quantification of oligomerization directly in the membrane becomes possible. Understanding the interaction quantitatively helps evaluate its importance, i.e. the value of dissociation constants refers to the protein concentration at which the oligomerization occurs. Knowing this, we can decide whether the interaction of endogenous proteins is their intrinsic property, or whether other physical forces may need to be involved to explain oligomerization, if it occurs in cells, such as other proteins, cytoskeleton or lipid arrangement.

Here we focus on intramembrane proteases of the rhomboid family, proteins formed by six or seven α-helical transmembrane segments (reviewed in 13). These enzymes occur widely across evolution and regulate EGF receptor signaling, mitochondrial dynamics, mitophagy, apoptosis, or invasion of apicomplexan parasites, and are relevant for a growing number of diseases including malaria, cancer and Parkinson's disease (reviewed in 14). Based largely on *in vitro* experiments in detergent-solubilized state of three bacterial rhomboid proteases, these enzymes have been proposed to form stable dimers and thereby become allosterically activated (15). If indeed rhomboid protease activity was regulated by dimerization, this could be significant for the interpretation of biological mechanisms involving rhomboids. Thus, to clarify this issue, here we zoom in on the well-characterized human rhomboid protease RHBDL2 (16-21), which is localized to the plasma membrane and consists of seven transmembrane helices, and analyze its oligomeric status in cell-derived lipid membranes.

Strong interactions are detectable by single molecule approaches such as fluorescence cross-correlation spectroscopy (FCCS) or photobleaching, but to observe weaker interactions

in the native membranes, higher protein densities are required. To investigate the oligomeric status of RHBDL2 as a model transmembrane protein, we use human cells expressing fluorescently labelled RHBDL2 to create GPMVs, which have simple geometry and show homogeneous distribution of the observed molecules. In addition to FCCS, we then employ FRET in individual vesicles in combination with Monte Carlo (MC) simulations to investigate the oligomeric state of RHBDL2. In doing so we build on the MC-FRET approach introduced by Johannsson *et al* (22, 23) and later applied to lipid clustering by others (24, 25). Within this context we develop and implement several novel considerations that are essential when addressing large membrane inclusions (such as RHBDL2) linked to fluorescent proteins. We focus on the direct determination of protein surface density in the 2D environment of GPMVs, we address the issue of not being in the isotropic dynamic limit in these conditions, and we show that the commonly used Förster radius (R_0) cannot be satisfactorily used in FRET quantification. Finally, we use MC simulations to model the FRET measurements in GPMVs, thus accounting for the lack of *à priori* knowledge of the dissociation constant and the geometry of the putative dimer. Comparing the simulation read-out with the measured data then allows drawing conclusions on the dimerization of the membrane protein.

In summary, we address several key limitations of using FRET for oligomerization studies of polytopic membrane proteins in live cell derived membranes and demonstrate their use by showing the lack of dimerization of rhomboid protease RHBDL2, which we confirm independently also by cell biological 'relocalization' experiments. Our results can find use in the application of FRET and FCCS for the analysis of oligomerization of transmembrane proteins in lipid membranes.

Materials and Methods

Materials

Oligonucleotides were from Sigma-Aldrich (USA), and restriction endonucleases and other enzymes for DNA cloning were from New England Biolabs (USA). All other chemicals were from Sigma Aldrich unless stated otherwise.

Cloning and constructs

The fluorescence reporters eGFP and mCherry were cloned as fusions to the N-terminus of human RHBDL2 (NCBI Reference Sequence NP_060291.2) in the pEGFP vector (18). For relocalization experiments, DNA sequence encoding the peptide KDEL preceded by a $(GS)_3$ linker was cloned at the 3' end of the RHBDL2 gene in the eGFP/mCherry-RHBDL2 constructs. The constructs encoding fluorescently tagged human GCPII were generated by fusing eGFP or mCherry and a 17 amino acid linker to the N-terminus of human GCPII (NCBI Reference Sequence NP_004467.1) harbouring mutations L4A/L5A (26, 27) that impair endocytosis. The control constructs for RHBDL2 encoding its N-terminal cytosolic domain (R2Ncyto) of 72 amino acids fused to eGFP or mCherry and equipped with a His-tag (eGFP-R2Ncyto-His$_6$ and mCherry-R2Ncyto-His$_6$) were cloned into the pET25b+ vector for bacterial expression. The constructs encoding His$_6$-eGFP or His$_6$-mCherry were kind gifts of Evzen Boura and had been generated by cloning the eGFP or mCherry encoding fragments into the pHis2 vector downstream of a His$_6$ purification tag and TEV cleavage site (28), yielding a spacer of 18 amino acids between the His tag and the fluorescent protein. All constructs were verified by Sanger sequencing.

Protein expression and purification

The His-tagged fusions eGFP-R2Ncyto-His$_6$, mCherry-R2Ncyto-His$_6$, eGFP-GCPIINcyto-His$_6$, and mCherry-GCPIINcyto-His$_6$ were expressed from T7 driven vectors in *E.coli* BL21(DE3) at 20 °C, inducing by 0.5 mM IPTG for 12 hrs. Cells were broken by 3 passages through Emulsiflex C3 (Avestin, Canada) in the presence of 1 mM serine protease inhibitor phenylmethylsulfonylfluoride, and the insoluble fraction was removed by centrifugation at $15000 \times g$ for 30 min at 4 °C. The His-tagged proteins were purified from the supernatant by metal-chelate affinity chromatography using NiNTA agarose (Qiagen), and eluted into 20 mM HEPES pH 8.0, 300 mM NaCl, and 10% (w/v) glycerol using 250 mM imidazole, which was immediately removed by desalting into PBS pH 7.4 using a PD-10 desalting column (GE Healthcare). Protein concentration was determined from absorbance at 280 nm, and purified proteins were flash-frozen into liquid nitrogen and stored at −80 °C. The identity of purified proteins was further validated by quantitative amino acid analysis, mass spectrometry and SDS-PAGE with fluorescence scanning.

Cell culture

Page **7** of **35**

91

HeLa cells (ATCC) were cultured in DMEM medium (Invitrogen) supplemented with 10% foetal calf serum (Thermo) at 37 °C and 10% CO_2. For transfection, 2×10^5 cells were seeded per well of a 4-chamber dish (Cellvis, cat. no. D35C4-20-1.5-N) and transfected by FuGene6 (Promega). The eGFP-RHBDL2 and mCherry-RHBDL2 mutant fusion construct plasmids were transfected at 50:50 ng, and 50:500 ng DNA per well for donor:acceptor ratios of approximately 1:1 and 1:10, respectively, for FLIM-FRET experiments, and 50:50 ng (donor:acceptor ratio of approximately 1:1) for FCCS experiments. The eGFP-GCPII and mCherry-GCPII mutant fusion construct plasmids were transfected at 250:250 ng, 250:750 ng and 75:750 ng DNA per well for donor:acceptor ratios of approximately 1:1, 1:3 and 1:10, respectively, for FLIM-FRET experiments, and 250:250 ng (donor:acceptor ratio of approximately 1:1) for FCCS experiments. The transfection mixtures were complemented with empty vector pcDNA3.1 to 1 μg of total DNA mass. Natural heterogeneity in expression levels within the cell population was exploited to choose cells with a range of fluorescence intensities of expressed reporters. For the relocalization experiments, 50 ng of FP-RHBDL2 vector with corresponding 50 ng FP-RHBDL2-KDEL fusion construct plasmids were transfected.

Spontaneously immortalized human keratinocytes HaCaT (29) (item no. 300493, Cell Lines Service, Germany) were cultured in DMEM medium (Invitrogen). To deplete endogenous RHBDL2, HaCaT cells were transduced with a recombinant lentivirus expressing the shRNA #01 targeting RHBDL2 (17) as described (16) and selected for puromycin resistance.

Immunoblotting

The integrity of the expressed fusion proteins and the expression levels of RHBDL2 were examined by immunoblotting. Cell lysates were separated by 4-20% gradient Tris-Glycine SDS PAGE (BioRad), electroblotted onto a PVDF membrane (Immobilon-FL, Millipore), optionally stained by the Revert protein stain (Li-Cor) and scanned using infrared fluorescence scanner Odyssey CLx (Li-Cor), destained, blocked in the Casein Blocker solution (Thermo) for 1 hr at 25°C, and exposed to primary antibodies over night at 4°C. Primary antibodies were from rabbit (with the exception of αGCPII-02 (30) which was from mouse), and were used at the following concentrations: α-RHBDL2 (Proteintech, cat. no. 12467-1-AP) at 1:500, α-eGFP (Cell Signaling Technology, cat. no. 2956) at 1:1000, α-RFP

(Thermo Fischer Scientific, cat. no. R10367) at 1:3000, αGCPII-02 at 1:5000 and secondary antibody Donkey α-Rabbit IgG (H+L) or Donkey α-Mouse IgG (H+L) cross-adsorbed and conjugated to DyLight 800 (Thermo Fischer Scientific, cat. no. SA5-10044) at 1:10000. Secondary antibody fluorescence was visualized using near-infrared fluorescence scanner Odyssey CLx (Li-Cor), and optionally quantified using Image Studio™ Lite (Li-Cor).

GPMV preparation

GPMVs were prepared as described elsewhere ([31]). Briefly, HeLa cells transiently expressing eGFP and mCherry fusions to human RHBDL2 were washed by the GPMV buffer (10 mM HEPES, 150 mM NaCl, 2 mM $CaCl_2$, pH 7.4). Then 2 mM solution of N-ethylmaleimide in the GPMV buffer was added and the cells were incubated for minimum 1 hour at 37 °C. Once GPMVs were formed, they were immediately used for microscopy experiments *in situ*. GPMVs containing GCPII constructs were produced using the chloride salt method as described ([10]).

GUV preparation

GUVs were prepared by electroformation ([32]). Chloroform-lipid mixture was prepared so that total lipid concentration was 5 μg/μL, containing 75 mol % of POPC, 20 mol % cholesterol and 5 mol % DGS-NTA(Ni) (all lipids purchased from Avanti Polar Lipids, Alabaster, AL). Ten μL of the mixture was spread on two ITO coated glass electrodes each. The electrodes were dried under vacuum overnight and then assembled in parallel into a homemade Teflon holder containing 5 mL of 600 mM sucrose solution. For the electroformation, 10 Hz, harmonically oscillating voltage of 1 V peak-value was applied to the electrodes for 1 hour in an incubator set to 60 °C. For the imaging, BSA-coated 4-chamber glass bottom dish (In Vitro Scientific) were used, and 100 μL of GUVs were mixed with 100 μL iso-osmotic buffer (25 mM Tris pH 8, 10 mM $MgCl_2$, 20 mM imidazole, 261.5 mM NaCl, 2 mM β-mercaptoethanol) containing the His-tagged fluorescent proteins.

For the experiments that were carried out to validate the FCCS based lateral concentration measurement, we prepared GUVs consisting of POPC and ATTO488- and ATTO647N-DOPE (ATTO-TEC GmbH, Siegen, Germany). The amount of the acceptor

labelled lipid was varied so that it was sufficient for the single molecule measurement as well as for the FRET, i.e. from 0.01 to 10 %. The amount of donor was 0.01 %.

Microscopy

All the microscopy images were acquired on the LSM 780 confocal microscope (Zeiss, Jena, Germany) using 40×/1.2 water objective. For the FLIM-FRET and FCCS measurements the external tau-SPAD detectors equipped with time-correlated single photon counting (TCSPC) electronics (Picoquant, Berlin, Germany). For FCCS experiments, eGFP and ATTO488 was excited with the 490 nm line of the Intune laser (Zeiss) pulsing at 40 MHz repetition frequency and mCherry and ATTO647N was excited continuously at 561 and 633 nm, respectively. The excitation light was focused on the apical membrane of giant liposomes. The precise positioning was checked by maximizing fluorescence intensity and the apparent molecular brightness ([33]). Fluorescence intensity, collected by the same objective lens, was re-focused on the pinhole (1 airy unit) and the re-collimated light behind the pinhole was split on the external tau-SPADs in front of which emission band pass filters 520/45, 600/52, and 679/41 for eGFP and ATTO488, mCherry and ATTO647N signal, respectively, were placed. The intensity of the excitation light at the back aperture of the objective was 2, 6, and 0.5 µW for 490 nm, 561 nm, and 633 nm laser line, respectively. The collected data were correlated by home-written script in Matlab (Mathworks, Natick, MA) according to a described algorithm ([34], [35]).

To avoid detector crosstalk, i.e. the bleed through from the GFP channel to the mCherry channel, the red channel fluorescence signal was split according to its TCSPC pattern (exponential for the signal generated by the pulsed Intune laser and flat for the 561 or 633 nm continuous wave laser) into two contributions and only the signal assigned to the flat TCSPC profile was correlated (Fig. S1). The data were processed as described ([35], [36]). The FLIM-FRET data were acquired in the equatorial plane of giant liposomes. The equatorial plane was chosen as it is well defined, easy to find and the measurement is not affected by small drifts of focus that are possible during FRET measurements. The data were accumulated during fast repetitive frame scans. Frame-sequentially, the acceptor signal was collected. The mean fluorescence signal of the donor and acceptor was used to obtain information on their concentrations by extrapolation of the calibration measurement as described in the Results section. The excitation intensity of the 561 and 633 nm laser was

decreased to 0.5 and 0.05 µW, respectively, in order to minimize photobleaching of the acceptor.

Monte Carlo (MC) simulations

For each FLIM measurement carried out on a single GPMV/GUV, an MC simulation was executed. The simulation algorithm is illustrated step by step in Fig. S2. The input parameters for the simulation were the surface concentrations of donor- and acceptor-labelled proteins that were obtained directly during FLIM acquisition with the help of the calibration measurement (see further). The simulations were run for the following variable parameters: i) Förster radius (R_{0app}) or ii) dissociation constant, K_D, and excluded radius, L_0, when oligomerization was addressed.

MC simulations were performed on the grid of 10000×10000 pixels. Size of the pixel was set to be $0.01 \times R_0$. First, the total amount of proteins in the simulated field was calculated. Based on the assumed K_D, overall numbers of monomers and dimers were calculated. In our planar system, the dissociation constant K_D is defined as follows:

$$K_D = \frac{c_M^2}{c_D},$$

(1)

where c_D and c_M are surface concentrations of dimer and monomer, respectively.

All the monomers and one of the partners from each dimer were placed to the field so that they did not overlap (each occupies a circular area of a diameter of an excluded distance), i.e. a random position was generated and if it was occupied, it was rejected and another one was generated. Subsequently, the partners in the dimers were located: a random, already localized molecule was selected and a new one was placed directly next to it (so that the distance of their centers equals the excluded distance) at a random angle. Once the molecule was located, a random number was generated that distinguishes whether the molecule is a donor or an acceptor based on the experimentally known ratio between those. The system uses periodic boundary.

Once the molecules were localized in the field, a random donor was selected and overall FRET rate Ω_i was determined according to the formula:

$$\Omega_i = \sum_j \left(R_0/R_{ij}\right)^6 \tau_D^{-1} + 0.5 * C_2 (R_0/R_C)^4 \tau_D^{-1},$$

(2)

where R_{ij} is the distance between the selected donor and all acceptors closer than the cut-off distance R_C ($10 \times R_0$), τ_D is the lifetime of the donor in the absence of acceptor, and C_2 is the number of the acceptors inside the circle of the Förster radius (22, 23, 37).

For the FRET efficiency calculations, 100 configurations of donors and acceptors within the simulated field are generated. For each configuration, 10 random donors are selected and their FRET rates generated according to Eq. **2**. For the overall FRET efficiency it holds: $\eta = 1 - \sum_i (1 + \Omega_i)^{-1}$), where the sum goes over all selected donors in all configurations.

For the fluorescence decay calculations, a random time Δt_i at which the energy transfer occurs was generated: $\Delta t_i = \frac{\ln\alpha}{\Omega_i}$, where α is a random number from the interval of 0 to 1. For each situation, i.e. each GPMV/GUV, 100 configurations of donors and acceptors were generated and at each configuration, 100 random excitations were performed. First, the probability that the donor does not transfer its energy to an acceptor is calculated by cumulative histogramming of Δt_i. Second, this probability is multiplied by the acceptor free donor decay (fit of the eGFP decay), and third, the resulting function is convolved with the instrumental response function.

The simulated efficiencies as well as the fluorescence decays for the given acceptor/donor surface densities were compared with the measured data by calculating the mean square displacement (MSD).

Experimental values of FRET efficiencies were obtained from the fits of experimental decays from the amplitude weighted average of the decay times:

$$\eta = 1 - \frac{\sum_i A_{DAi} \tau_{DAi}}{\sum_i A_{Di} \tau_{Di}},$$

(3)

where A_i and τ_i stay for amplitudes characteristic times of bi-exponential fits of the experimental decays of donor in the presence and absence of acceptors.

FRET based determination of surface acceptor concentration

According to Bauman and Fayer (37), the donor fluorescence decay $F(t)$ consists of two contributions:

$$F(t) = G^s(t) \sum_i \alpha_i \exp(-t/\tau_i),$$
(4)

i) the decay of the donor in the absence of acceptors characterized by the amplitudes α_i and the corresponding decay times τ_i, and ii) the probability $G^s(t)$, that a donor does not transfer its energy to an acceptor from the moment of having been excited until time t. $G^s(t)$ is a product of an intra and inter membrane leaflet energy transfer:

$$\ln G^s_{intra}(t) = -C_2 \Gamma(2/3)(t/\tau)^{1/3}, \qquad\qquad \text{and}$$
(5)

$$\ln G^s_{inter}(\mu) = -C_2/3 \, (d/R_0)^2 (2\mu/3)^{1/3} \int_0^{2/3\mu} (1 - e^{-s}) s^{-4/3} ds.$$
(6)

Here, C_2 is the number of acceptors within in a single leaflet in the circular area of R_0 in radius, d is the membrane thickness, Γ is the gamma function, $\mu = 3t(R_0/d)^6/2\tau$, $s = 2\mu\cos^6\theta_r/3$, where θ_r is the angle between the bilayer normal and a vector that connects positions of donor and acceptor dipoles.

We used the model to determine C_2 for a 2D membrane system, where ATTO488 and ATTO647N-DOPE served as donor and acceptor, respectively. First, we have determined amplitudes α_i and the decay times τ_i of the donor in the absence of acceptors by 2-exponential fitting of the donor fluorescence decay. Second, we have fitted the donor decay in the presence of acceptors with Eq. **4**. The fitting involved convolution with the experimental instrument response function and was implemented in Matlab.

Results

Investigating dimerization of human rhomboid protease RHBDL2 by fluorescence cross-correlation spectroscopy

The joint motion of proteins of interest provides the most convincing evidence of their physical interaction, and single-molecule techniques based on tracking individual fluorescently labeled proteins thus represent straightforward tools to address protein-protein oligomerization. Since single-molecule based approaches require protein concentrations low enough to distinguish individual molecules, they are applicable only to strongly interacting species (i.e. those that form significant fraction of oligomers at low total concentrations).

We have first employed fluorescence cross-correlation spectroscopy (FCCS) (38) to investigate the dimerization status of human RHBDL2 in live cell derived membranes of giant plasma membrane derived vesicles (GPMVs) (31). The FCCS technique analyzes fluorescence intensity fluctuations arising from individual molecular transits through the diffraction limited laser focus, and the temporal cross-correlation between the signal of fluorescence reporter labeled membrane proteins suggests their joint motion. We have performed the experiment in the apical membrane of GPMVs (Fig. 1J) derived from HeLa cells co-expressing eGFP-RHBDL2 and mCherry-RHBDL2. Both fusion proteins localized largely to the plasma membrane, were mostly full-length (Fig. 1A, D, G), and exhibited normal protease activity of RHBDL2 against a well-established model substrate (Fig. S3). Cross-correlation functions (Fig. 1M) did not suggest any interaction of RHBDL2 monomers at this concentration range. The same experiment was performed with murine GCPII (Fig. 1B, E, H, K, N), which is known to form dimers and localize to the plasma membrane (39, 40). As expected, the co-expressed eGFP and mCherry fusions of GCPII exhibited a positive cross-correlation amplitude (Fig. 1N), indicative of a dimer. In contrast, eGFP and mCherry fused to His-tagged N-terminal cytosolic domain of RHBDL2 (R2Ncyto) attached to the membrane via Ni-NTA lipid containing GUVs (Fig. 1C, F, I, L) did not show any cross-correlation (Fig. 1O), serving as negative, monomeric controls. More details on fits and the analysis of correlation amplitudes are given in SI (Fig. S4, Table S1). Taken together, these results indicated that if RHBDL2 dimers exist at all, their interaction is relatively weak, and their analysis thus requires methods that can access higher concentration ranges. The use of FCCS has been shown to be subjected solely to brightness of molecules and can principally address also higher concentrations provided that detectors can handle high fluorescence countrates and that effects of laser fluctuations can be corrected (41). This

approach represents a great potential in addressing protein-protein interactions, but cannot be achieved with most of the common microscope setups at the moment, including ours. Our detectors can work linearly to about 0.2 MHz, which means that we can observe maximum about 100 fluorophores (with brightness of 2-3 kHz per molecule, as is common for eGFP or mCherry) in focus (which corresponds to about 0.4×10^{-3} fluorophores per nm^2), and corrections of fluctuations of laser power, and fluctuation of 2D system, such as thermal motion of the membrane, have not been addressed.

Investigating weakly dimerizing polytopic membrane proteins by FRET

From a biophysical point of view, quantifying weaker protein-protein interactions by fluorescence techniques requires working at concentrations that are close to the dissociation constant, which may exceed maximum levels used in single-molecule experiments (typically fewer than about 0.4×10^{-3} fluorophores per nm^2). Therefore, spectroscopic techniques probing the vicinity of fluorescent probes are required, such as FRET. Working at high protein densities, with membrane-embedded entities of non-negligible excluded volume that are potentially forming dimers of unknown spatial orientation however brings specific problems, such as proximity-induced FRET, and the need to know the surface density of donors and acceptors at the site of measurement. To accommodate all the specifics of dealing with weakly interacting polytopic transmembrane proteins fused to fluorescent protein reporters, we discuss several underlying physical and methodological considerations below.

Determination of protein surface densities in the membrane of giant liposomes

Quantitative determination of dissociation constants requires simulations of FRET efficiencies by Monte Carlo approach (24), and for this, knowledge of surface densities of donors and acceptors is essential. Due to the large differences in geometry of 3D and 2D systems, calibration of fluorescence intensity versus known protein concentration feasible in 3D is not transferrable to a 2D system. Therefore, we employed a single molecule technique of fluorescence correlation spectroscopy (FCS) that enables counting fluorescent molecules within a diffraction-limited spot. The full length in half maximum (FWHM) of the spot light profile is given as FWHM $= \frac{0.61 * \lambda}{N_A}$, where λ is the excitation wavelength and N_A stands for the numerical aperture of the objective lens. We used FCS to measure fluctuations of

fluorescence intensity in the upper membrane of a liposome (its apical surface) (Fig. 2A) (42), and analysis of the intensity trace yielded the absolute number of fluorophores within this focal spot. Since the FRET measurement, for which we need to know surface densities of donors and acceptors, is made in the equatorial plane of a spherical liposome (GPMV) (Fig. 2A) at a lower excitation power, the relationship between the mean equatorial intensity at the excitation power used for FRET and the apical surface density needs to be established. FRET protein densities are much higher than the single molecule experiment requires, therefore the dependence between the equatorial signal and the apical protein density is linearly extrapolated from the FCS data (Fig. 2B).

In order to validate our approach, we decided to compare the results obtained by the FCS/extrapolation approach with a FRET based approach. In a planar system of donors and acceptors which are in the dynamic isotropic regime, such as ATTO488 and ATTO647N headgroup labeled lipids, the fluorescence decay of donors in the "sea" of acceptors has been theoretically described by Bauman and Fayer (37). For details see Methods. This analysis (Fig. S5) shows a reasonably good agreement between the two methods of acceptor surface density determination, indicating that our approach is a valid way of determination of protein surface densities in giant liposomes.

Determination of the apparent Förster radius

Förster radius represents the main characteristics of a FRET pair. It equals the distance where FRET efficiency drops to 50%, and it thus refers to the distances that can be addressed by the pair. Förster radius (R_0) can be determined as follows:

$$R_0 = \sqrt[6]{\frac{9\ln10}{128\,\pi^5 N_A}\frac{\kappa^2 Q_D}{n^4}J},$$

(7)

where N_A stays for the Avogadro's number, n is the refractive index, Q_D is the quantum yield of the donor in the absence of acceptor, J is the spectral overlap integral and κ^2 is the orientation factor kappa squared, which equals 2/3 if rotational depolarization upon excitation occurs much faster than transfer of energy and all orientations of the fluorophores are equally distributed (i.e. dynamic isotropic limit). Fluorescent proteins attached to membranes via fusions to integral membrane proteins are thought to rotate slowly compared to the donor fluorescence lifetime, and thus the dynamic limit may no longer hold for them. In addition,

membranes are naturally anisotropic, which prevents the fluorophores from sampling the entire rotational space, and together makes correct mathematical treatment of such a system exceedingly complicated. In these cases the distribution of acceptor transition dipole orientations becomes distance dependent as a direct consequence of the motion restrictions and system anisotropy. Therefore the appropriate physical description of the system would require knowledge of not only a mean value of the orientation factor (if we had a chance to access it), but of its entire donor/acceptor distance distribution. In single molecule studies the orientation factors were determined for some single donor-acceptor pairs (43, 44). In our laterally crowded system, however, a donor can transfer its energy to several acceptors with different probabilities. Experimentally, in such laterally crowded systems, an individual energy transfer cannot be separated from transfers to other neighboring acceptors, and this has to be taken into account.

These considerations imply that Förster radius obtained for isotropic dynamic limit cannot be automatically used for quantitative description of membrane attached fluorescent proteins. Moreover, even the simple equations that are commonly used when characterizing FRET are no more valid due to the need for quantum yield averaging. Despite that the established formalism for FRET efficiency η can be kept:

$$\eta = 1 - Q_{\mathrm{DA}}/Q_{\mathrm{D}}$$

(8)

where Q_{DA} and Q_{D} are quantum yields of donor in the presence and absence of the acceptor, respectively. For their ratio, it holds:

$$Q_{\mathrm{DA}}/Q_{\mathrm{D}} = 1/M \lim_{M \to \infty} \Sigma_M \left[1 + \Sigma_{i=1}^{N}\left(R_{0\mathrm{app}}/R_i\right)^6\right]^{-1},$$

(9)

where summing over i holds for all acceptors in the vicinity of a selected donor and summing over M stays for all donors in the system in all possible configurations. We introduce the apparent Förster radius, $R_{0\mathrm{app}}$, which formally behaves as a common Förster radius, but is not only specific for the given donor/acceptor pair, but for the entire situation, i.e. angular distribution of all transition dipole moments at given protein densities. It has to be emphasized that $R_{0\mathrm{app}}$ is an empirical constant that is principally influenced by the linker length between the transmembrane protein and the fluorescent protein, as well as to the given donor/acceptor surface density or heterogeneities in lateral protein organization.

In order to characterize protein-protein interaction by means of dissociation constant, R_{0app} has to be determined in a reference system that experimentally best resembles the system of our interest (which are GPMVs with transmembrane proteins fused to fluorescent proteins). In our approach, we decided for the reference system that consists of the N-terminal cytosolic extramembrane part of the studied transmembrane protein RHBDL2 (R2Ncyto), fused to the fluorescent protein and attached to the membrane of giant unilamellar vesicles (GUVs) via His-tag and a NTA-lipid anchor. Geometrically, both the reference system and the studied system represent planes with green and red fluorescing protein barrels separated by the cytosolic domain from the plane, i.e. the rotational motion of the barrels and their mutual geometry would be comparable. Therefore, we assume that the donor-acceptor distance-dependent κ^2 factor distribution would be similar to that in GPMVs containing polytopic membrane proteins fused to fluorescence proteins. The use of this approach would be restricted only to the systems where this assumption holds, i.e. those with similar rotational dynamics and geometry. It has to be pointed out that in any FRET experiments using established interacting or non-interacting membrane proteins as positive or negative controls, this assumption is tacitly expected to hold, because otherwise the differences in FRET efficiencies could be caused by a different κ^2 distribution rather than by interaction between the proteins of interest.

In order to estimate the R_{0app} for two fluorescent proteins attached to the membrane, we made use of the discussed proximity FRET effect. Knowing that the distribution of His$_6$-R2Ncyto-eGFP and His$_6$-R2Ncyto-mCherry attached to the membrane of giant unilamellar vesicles (GUVs) spiked with DGS-NTA(Ni) is homogeneous and that no significant interaction between the two proteins occurs, the combination of FRET with MC simulations allows estimation of the apparent Förster radius of this fluorophore pair. At given surface densities of donors and acceptors, and with knowledge of the excluded areas of the protein barrels, the apparent Förster radius is the only variable parameter required for the determination of FRET efficiency. We have thus prepared GUVs containing DGS-NTA(Ni) and measured FRET efficiency for various amounts of His-tagged fluorescent proteins added to the GUVs. For every GUV, the mean equatorial intensity was evaluated, the surface protein density of donors and acceptors was calculated from the extrapolation of the calibration measurement described above, and FRET efficiency based on fluorescence lifetime imaging (FLIM) was determined (Fig. 3A).

The surface densities of donors and acceptors obtained for each GUV were taken as input parameters for MC simulations that were employed to calculate a theoretical level of FRET efficiency for the given situation including the proximity FRET phenomenon. Varying the value of the apparent Förster radius (R_{0app}) and assuming that the donor (eGFP) and acceptor (mCherry) cannot come closer to each other than 30 Å ([4]), we have obtained the best agreement between our experimental data and the MC simulation for R_{0app} ranging between 50 and 58 Å ([Fig. 3]A, inset). Also when fitting our experimental data with the numerical model established by Snyder and Freire ([8]), we obtained the value of R_{0app} of 54 Å (for fixed value of the excluded distance of 30 Å). Both for the MC simulations and for the Snyder-Freire model, we used the closest donor acceptor distance of 30 Å. It is of note that we do not precisely know the distance, but based on the protein geometry it has to fall within the range of 25 to 35 Å. The numerical model gives corresponding values of R_{0app} 52 and 56 Å, which means that the error in the closest protein-protein distance estimate is much smaller than the range of R_{0app} obtained from our simulation.

Both the values we obtained from the FRET efficiency simulations as well as those we obtained from the numerical model are very close to the value of R_0 for the eGFP/mCherry pair at dynamic isotropic limit, 52 Å. In addition, we have performed the reference experiment also with a shorter linker of 18 amino acids instead of the whole 72 amino acid R2Ncyto and obtained very similar values of FRET efficiency suggesting that the effect of the linker length is negligible (see [Fig. S6]). Collectively this suggests that despite our concerns the FRET behavior of the fluorescence proteins barrels resembles the dynamic isotropic regime in our system.

Our MC simulations were not only adopted for the FRET efficiency calculations, but also for modeling of fluorescence decays. The advantage of the fluorescence lifetime imaging combined with the decay simulations over the efficiencies is that the whole decay bears significantly more complex information on the energy transfer. Therefore, even from a single GUV measurement, the information on Förster radius or interaction parameters (see later) can be drawn. To prove this, we have chosen a few single-GUV measurements and analyzed fluorescence decay data gathered during imaging of the GUVs in their equatorial plane. Comparing the measured decay with the simulated data revealed the best agreement for R_{0app} of 58 Å (see [Fig. 3]B), which is remarkably close to the values resulting from the analysis of FRET efficiencies ([Fig. 3]A).

Dimerization of RHBDL2 and GCPII

To address the thermodynamic propensity of the human rhomboid intramembrane protease RHBDL2 to form dimers in the natural lipid environment and at high protein densities, we have measured FRET in giant plasma membrane-derived vesicles (GPMVs) prepared from HeLa cells co-expressing eGFP and mCherry fusions to RHBDL2, using glutamyl carboxypeptidase II (GCPII, also known as NAALADase) as a dimeric positive control (39, 40) (Fig. 1). The use of the spherical GPMVs was crucial for quantification of protein surface densities in the area from which the signal was collected. In living cells, membrane proteins are synthetized at the endoplasmic reticulum (ER) membranes and are then trafficked to the plasma membrane. Since plasma membrane surface is not simply planar, but it is complicated by ruffles and numerous filopodia-like protrusions (Fig. 1G, H), plenty of signal intensity heterogeneities are visible when focusing on the plasma membrane adhering to the glass (Fig. 1G, H). This could be also due to numerous ER-plasma membrane contact sites. As a consequence, overall area of the plasma membrane cannot be easily measured in live cells. Thus, GPMVs with low fluorescence signal were used for the calibration of surface protein density by FCS as described above, and protein density in the apical surface of the highly fluorescent GPMVs used for FRET measurements was calculated from the mean equatorial fluorescence. These surface protein densities were used as input parameters for MC simulations, where the value of K_D and the closest donor-acceptor distance (excluded radius) were varied, and the results were compared to the experimental values of FRET efficiency (Fig. 4).

The results for RHBDL2 show that independently of the excluded radius of the protein (L_0 in Fig. 4B), experimental data match the simulations only at relatively high K_D values (low $-\log(K_D \times nm^2)$) of about 0.16 nm^{-2} (corresponding to about 4.5 mol % assuming lipid head group size is 60 $Å^2$). This is one or two orders of magnitude higher than the highest acceptor concentration achieved in our experiment, and a few more orders of magnitude higher than the concentration of the endogenous RHBDL2 in HeLa cells (Fig. S7). It is thus highly unlikely that the rhomboid protease scaffold of RHBDL2 be intrinsically dimeric in lipid membranes at physiological concentrations. For the positive control, GCPII, the simulation predicts apparent K_D value of $1.5 \times 10^{-5} - 1.5 \times 10^{-4}$ nm^{-2} (corresponding to 0.00045 – 0.0045 mol %, respectively, provided the lipid head group size is 60 $Å^2$), which is 3-4

orders of magnitude lower than that of RHBDL2, and in agreement with the known dimeric character of this protein ([39](), [40]()).

For both RHBDL2 and GCPII, we have also employed the analysis of fluorescence donor decays. We have selected several GPMVs with higher acceptor density to assure for high enough level of FRET. We have simulated the decay for the selected GPMV characterized by the surface concentrations of donors and acceptors. Similar to efficiency simulations, we have executed the simulations for increasing K_Ds and radii of excluded volumes (L_0). Eventually, the simulated data were compared with the experimental decay curve, which shows that the K_Ds as well as the excluded radii obtained from the decay analysis are in a good agreement with those obtained from the analysis of efficiencies ([Fig. 5]()).

It is worth noticing that the levels of FRET observed for RHBDL2 are comparable to those obtained with R2Ncyto, the negative control, for similar levels of acceptor density (the donor density was small in both situations, see [Table S2]()). This means that the observed FRET efficiency can be attributed entirely to the proximity phenomena in case of RHBDL2 and R2Ncyto. In contrast, in the case of GCPII, large FRET efficiencies (around 30 %) were observed already at protein densities almost an order of magnitude lower than the highest densities used with RHBDL2 and R2Ncyto. At the concentrations used for GCPII, the level of proximity FRET for a GCPII monomer would not exceed 5 %, for dimer it would be even lower ([2]()). Most of the FRET efficiency in GCPII experiments thus arises from the strong interaction between GCPII monomers.

Our analysis indicates that RHBDL2 is not dimeric to any significant degree in cell membrane derived vesicles at expression levels far exceeding the endogenous ones (17 to 56-fold higher, [Fig. S7]()). We however could not exclude that in other intracellular compartments such as the ER or Golgi, which have a different lipid composition and hydrophobic thickness, RHBDL2 may dimerize. Since interactions within the intracellular compartments of the secretory pathway are not easily directly addressable by FRET or FCCS, we resorted to a cell biological approach exploiting the cellular mechanism of retrieving proteins back to the ER from early Golgi via the KDEL tag ([45]()) and membrane-bound KDEL receptor ([46](), [47]()). In fact, KDEL tagging was shown to exert a strong dominant negative effect on the secretion of dimeric proteins such as TGFβ ([48]()), documenting feasibility of this approach. When eGFP-RHBDL2 and mCherry-RHBDL2 are co-expressed in HeLa cells, both constructs show

predominantly plasma membrane localization, notably labelling filamentous extrusions of the cell surface (Fig. 6A, Fig. S8A). When both of these constructs are equipped with a C-terminal, luminal KDEL tag, both show predominantly an ER localization with a complete overlap and complete loss of filopodia-like labelling (Fig. 6B, Fig. S8B). When, however, mCherry-RHBDL2-KDEL and eGFP-RHBDL2 are co-expressed, only mCherry-RHBDL2-KDEL relocalizes to the ER, while the localization of eGFP-RHBDL2 is barely affected and fluorescence of the two reporters overlaps only minimally (Fig. 6C, Fig. S8C). The same is true when eGFP-RHBDL2 is KDEL-tagged and mCherry-RHBDL2 is not (data not shown). These independent qualitative data strongly indicate the absence of stable dimers of RHBDL2 that could traffic together. In other words, RHBDL2 appears monomeric in all major membrane compartments of the secretory pathway where it normally resides.

Discussion

For addressing weak protein oligomerization in cellular membranes, FRET combined with FLIM represents one of few applicable optical non-invasive techniques. It has been shown that the expression levels of fluorescent protein acceptor fusion has to be kept under control as it may contribute to the overall FRET efficiency by the proximity effect (2). Having full control of lateral protein density does not seem to be possible in living cells, but it is possible in GPMVs derived from the cells of interest using an external calibration (4). To separate the impact of the proximity effect from the impact caused by protein oligomerization on the overall FRET efficiency we have utilized MC simulations.

A large fluorescent protein attached to a membrane-residing protein of interest, however, cannot freely rotate in all directions, and it rotates at the time-scales much slower that the donor fluorescence lifetime. Such fluorescent fusion membrane proteins hence do not fulfil the requirements for the dynamic isotropic limit that predicts the value of the orientation factor κ^2 to be 2/3. Knowledge of κ^2 and consequently of Förster radius is however essential for meaningful quantification of FRET. In situations that are neither isotropic nor dynamic, estimation of κ^2 is very difficult and usually requires understanding the dynamics of the involved proteins. Moreover, to calculate FRET efficiency under the static or intermediate conditions, the overall dependence of the κ^2 factor on the donor-acceptor distance has to be known (49). Here, we propose to use an apparent value of R_0 (R_{0app}) that would allow to compare systems resembling each other well as regards rotational dynamics and geometry of

the fluorophores. In this sense, we consider our system consisting of transmembrane proteins fused to fluorescence proteins comparable to the eGFP/mCherry donor/acceptor pair attached by a linker and His-tag to the membrane of GUVs. Our approach then simulates FRET efficiencies at experimentally obtained protein densities and compares those with the measured FRET efficiency values. The only optimized parameter in the simulations is the apparent Förster radius. Our results clearly indicate that the apparent Förster radius for eGFP and mCherry in our system falls into the range of $50 - 58$ Å, which is in a very good agreement with the value reported in literature for the dynamic isotropic limit, i.e. 52 Å ([50]). This result however does not prove that the pair is in the dynamic isotropic regime. Whether it is close to it or whether it is just a coincidence cannot be distinguished from our data. Our results however imply that a pair of similar barrel-like fluorescent proteins attached to a membrane plane via a transmembrane protein may be treated as if they were in dynamic isotropic regime as their geometrical situation is similar to our system. Being aware that the rotational rigidity caused by the close proximity of the fluorescent protein barrel and the lipid membrane can be significantly decreased by the presence of a flexible linker, we tested two different linker lengths: 18 and 72 amino acids. As we did not observe any difference between those two, we can conclude that already 18 residues probably allow maximal mobility of the fluorescence protein barrel.

The quantitative determination of the dissociation constant by the MC simulation can be generally applied also for any higher-order protein-protein interaction. It is, however, limited by the *a priori* knowledge of the stoichiometry of the protein complexes. When it is known, it can be straightforwardly implemented into the MC simulation. When stoichiometry is unknown, single molecule or other biochemical approaches can be employed to acquire this information.

Apart from FRET, we have also applied FCCS to address protein dimerization. In our hands the two techniques are more or less complementary. While FCCS is used for lower surface concentrations of proteins (typically up to approximately 0.4×10^{-3} protein fluorophores per nm^2), FRET, even though it can be applied generally in all concentration ranges, is beneficial especially in ranges were FCCS cannot be easily used in practice (typically above approximately 0.4×10^{-3} protein fluorophores per nm^2). Although the concentration range where FCCS can be applied can generally be much broader, theoretically unlimited ([41]), its limits are in practice set by the technical features of the used microscope, specifically, by the dynamic range of the detectors or by the ability to correct for the laser

fluctuations. In addition, the membrane thermally fluctuates, which would probably also add a noise component to the correlation curve. This may be negligible at low protein densities as its amplitude is very small, but it could become significant at high protein concentrations. The actual donor concentrations in experiments depicted in Fig. 4 are reported in Table S2.

In order to validate our method, we have applied it to the dimeric membrane protein GCPII. The data indicated that GCPII was dimeric in GPMVs, and the values of the dissociation constant of GCPII were 3-4 orders of magnitude lower than those of the rhomboid intramembrane protease RHBDL2, which was previously proposed to be dimeric (15). Interestingly, recent single-molecule photobleaching analysis of several rhomboid proteases and pseudoprotease by Kreuzberger and Urban (51) also found no evidence for dimerization in live cells. This study was however limited to low concentrations of the species compatible with single-molecule studies. Our study additionally provides data from the high concentration range of rhomboid, and is testing the possibility of weak interactions between rhomboid monomers. Since our study also finds no evidence for rhomboid dimerization, these two studies together strongly indicate that the rhomboid domain is intrinsically monomeric in membranes. This is particularly relevant for the related rhomboid pseudoproteases iRhoms (52) and Derlins (53) that are involved in inflammatory signaling (54) and ER associated degradation (55), respectively. Derlins were proposed to be dimeric based on detergent solubilisation and pull-down experiments (56), but it is highly likely that Derlins themselves do not dimerize in membranes. Indeed, dimerization of the rhomboid-family proteins could occur if there was an additional force that drives the partners to one another, such as dimerization of their extramembrane domains (57), interaction via a third partner, segregation to some specific lipid pools, or binding to juxtamembrane structures.

Conclusions

Here we adapt the usage of FRET for the analysis of dimerization of polytopic transmembrane proteins by taking into account proximity induced FRET, by careful estimation of the apparent Förster radius that can principally deviate from the value commonly used in isotropic conditions, and by employing Monte Carlo simulation to interpret FRET results. In addition, we have developed and validated a method for lateral concentration determination that is crucial for having distribution of donors and acceptors under the control. Using these methods, our biophysical and cell-biological experiments do

not provide any evidence for the dimerization of human rhomboid protease RHBDL2 in lipid membranes of live cells, which suggests that the transmembrane core of rhomboid protease is intrinsically monomeric, unlike proposed previously ([15]).

Author Contributions

KS, JH, and JŠ designed research, JŠ, JH, PR, NJ, DJ, AS, EP, MF and JD performed research, JH, JŠ, DJ, NJ and KS analyzed data, and JH and KS wrote the manuscript with input from co-authors.

Acknowledgements

References

1. Gupta, K., J. A. C. Donlan, J. T. S. Hopper, P. Uzdavinys, M. Landreh, W. B. Struwe, D. Drew, A. J. Baldwin, P. J. Stansfeld, and C. V. Robinson. 2017. The role of interfacial lipids in stabilizing membrane protein oligomers. Nature 541(7637):421-424.

2. King, C., S. Sarabipour, P. Byrne, D. J. Leahy, and K. Hristova. 2014. The FRET signatures of noninteracting proteins in membranes: simulations and experiments. Biophys. J. 106(6):1309-1317.

3. King, C., V. Raicu, and K. Hristova. 2017. Understanding the FRET Signatures of Interacting Membrane Proteins. J. Biol. Chem. 292(13):5291-5310.

4. Chen, L., L. Novicky, M. Merzlyakov, T. Hristov, and K. Hristova. 2010. Measuring the energetics of membrane protein dimerization in mammalian membranes. J. Am. Chem. Soc. 132(10):3628-3635.

5. Meyer, B. H., J. M. Segura, K. L. Martinez, R. Hovius, N. George, K. Johnsson, and H. Vogel. 2006. FRET imaging reveals that functional neurokinin-1 receptors are monomeric and reside in membrane microdomains of live cells. Proc Natl Acad Sci U S A 103(7):2138-2143.

6. Chakraborty, H., and A. Chattopadhyay. 2015. Excitements and challenges in GPCR oligomerization: molecular insight from FRET. ACS Chem Neurosci 6(1):199-206.

7. Strachotova, D., A. Holoubek, H. Kucerova, A. Benda, J. Humpolickova, L. Vachova, and Z. Palkova. 2012. Ato protein interactions in yeast plasma membrane revealed by fluorescence lifetime imaging (FLIM). Biochim. Biophys. Acta 1818(9):2126-2134.

8. Snyder, B., and E. Freire. 1982. Fluorescence energy transfer in two dimensions. A numeric solution for random and nonrandom distributions. Biophys. J. 40(2):137-148.

9. Sarabipour, S., N. Del Piccolo, and K. Hristova. 2015. Characterization of membrane protein interactions in plasma membrane derived vesicles with quantitative imaging Forster resonance energy transfer. Acc. Chem. Res. 48(8):2262-2269.

10. Del Piccolo, N., J. Placone, L. He, S. C. Agudelo, and K. Hristova. 2012. Production of plasma membrane vesicles with chloride salts and their utility as a cell membrane mimetic for biophysical characterization of membrane protein interactions. Anal. Chem. 84(20):8650-8655.

11. Cohen, S., H. Ushiro, C. Stoscheck, and M. Chinkers. 1982. A native 170,000 epidermal growth factor receptor-kinase complex from shed plasma membrane vesicles. J. Biol. Chem. 257(3):1523-1531.

12. Scott, R. E., R. G. Perkins, M. A. Zschunke, B. J. Hoerl, and P. B. Maercklein. 1979. Plasma-Membrane Vesiculation in 3t3-Cells and Sv3t3-Cells .1. Morphological and Biochemical Characterization. J. Cell Sci. 35(Feb):229-243.

13. Strisovsky, K. 2013. Structural and mechanistic principles of intramembrane proteolysis--lessons from rhomboids. FEBS J. 280(7):1579-1603.

14. Dusterhoft, S., U. Kunzel, and M. Freeman. 2017. Rhomboid proteases in human disease: Mechanisms and future prospects. Biochim Biophys Acta Mol Cell Res 1864(11 Pt B):2200-2209.

15. Arutyunova, E., P. Panwar, P. M. Skiba, N. Gale, M. W. Mak, and M. J. Lemieux. 2014. Allosteric regulation of rhomboid intramembrane proteolysis. EMBO J. 33(17):1869-1881.

16. Johnson, N., J. Brezinova, E. Stephens, E. Burbridge, M. Freeman, C. Adrain, and K. Strisovsky. 2017. Quantitative proteomics screen identifies a substrate repertoire of rhomboid protease RHBDL2 in human cells and implicates it in epithelial homeostasis. Sci. Rep. 7(1):7283.

17. Adrain, C., K. Strisovsky, M. Zettl, L. Hu, M. K. Lemberg, and M. Freeman. 2011. Mammalian EGF receptor activation by the rhomboid protease RHBDL2. EMBO Rep. 12(5):421-427.

18. Lohi, O., S. Urban, and M. Freeman. 2004. Diverse substrate recognition mechanisms for rhomboids; thrombomodulin is cleaved by Mammalian rhomboids. Curr. Biol. 14(3):236-241.

19. Noy, P. J., R. K. Swain, K. Khan, P. Lodhia, and R. Bicknell. 2016. Sprouting angiogenesis is regulated by shedding of the C-type lectin family 14, member A (CLEC14A) ectodomain, catalyzed by rhomboid-like 2 protein (RHBDL2). FASEB J. 30(6):2311-2323.

20. Cheng, T. L., Y. T. Wu, H. Y. Lin, F. C. Hsu, S. K. Liu, B. I. Chang, W. S. Chen, C. H. Lai, G. Y. Shi, and H. L. Wu. 2011. Functions of rhomboid family protease RHBDL2 and thrombomodulin in wound healing. J. Invest. Dermatol. 131(12):2486-2494.

21. Pascall, J. C., and K. D. Brown. 2004. Intramembrane cleavage of ephrinB3 by the human rhomboid family protease, RHBDL2. Biochem. Biophys. Res. Commun. 317(1):244-252.

22. Johansson, L. B. A., S. Engstrom, and M. Lindberg. 1992. Electronic-Energy Transfer in Anisotropic Systems .3. Monte-Carlo Simulations of Energy Migration in Membranes. Journal of Chemical Physics 96(5):3844-3856.

23. Engstrom, S., M. Lindberg, and L. B. A. Johansson. 1988. Monte-Carlo Simulations of Electronic-Energy Transfer in 3-Dimensional Systems - a Comparison with Analytical Theories. Journal of Chemical Physics 89(1):204-213.

24. Stefl, M., R. Sachl, J. Humpolickova, M. Cebecauer, R. Machan, M. Kolarova, L. B. Johansson, and M. Hof. 2012. Dynamics and size of cross-linking-induced lipid nanodomains in model membranes. Biophys. J. 102(9):2104-2113.

25. Sachl, R., M. Amaro, G. Aydogan, A. Koukalova, Mikhalyov, II, I. A. Boldyrev, J. Humpolickova, and M. Hof. 2015. On multivalent receptor activity of GM1 in cholesterol containing membranes. Biochim. Biophys. Acta 1853(4):850-857.

26. Mlcochova, P., C. Barinka, J. Tykvart, P. Sacha, and J. Konvalinka. 2009. Prostate-specific membrane antigen and its truncated form PSM'. Prostate 69(5):471-479.

27. Rajasekaran, S. A., G. Anilkumar, E. Oshima, J. U. Bowie, H. Liu, W. Heston, N. H. Bander, and A. K. Rajasekaran. 2003. A novel cytoplasmic tail MXXXL motif mediates the internalization of prostate-specific membrane antigen. Mol. Biol. Cell 14(12):4835-4845.

28. Dubankova, A., J. Humpolickova, M. Klima, and E. Boura. 2017. Negative charge and membrane-tethered viral 3B cooperate to recruit viral RNA dependent RNA polymerase 3D (pol). Sci. Rep. 7(1):17309.

29. Boukamp, P., R. T. Petrussevska, D. Breitkreutz, J. Hornung, A. Markham, and N. E. Fusenig. 1988. Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line. J. Cell Biol. 106(3):761-771.

30. Barinka, C., P. Mlcochova, P. Sacha, I. Hilgert, P. Majer, B. S. Slusher, V. Horejsi, and J. Konvalinka. 2004. Amino acids at the N- and C-termini of human glutamate carboxypeptidase II are required for enzymatic activity and proper folding. Eur. J. Biochem. 271(13):2782-2790.

31. Sezgin, E., H. J. Kaiser, T. Baumgart, P. Schwille, K. Simons, and I. Levental. 2012. Elucidating membrane structure and protein behavior using giant plasma membrane vesicles. Nat. Protoc. 7(6):1042-1051.

32. Humpolickova, J., I. Mejdrova, M. Matousova, R. Nencka, and E. Boura. 2017. Fluorescent Inhibitors as Tools To Characterize Enzymes: Case Study of the Lipid Kinase Phosphatidylinositol 4-Kinase IIIbeta (PI4KB). J. Med. Chem. 60(1):119-127.

33. Humpolickova, J., E. Gielen, A. Benda, V. Fagulova, J. Vercammen, M. Vandeven, M. Hof, M. Ameloot, and Y. Engelborghs. 2006. Probing diffusion laws within cellular membranes by Z-scan fluorescence correlation spectroscopy. Biophys. J. 91(3):L23-25.

34. Wahl, M., I. Gregor, M. Patting, and J. Enderlein. 2003. Fast calculation of fluorescence correlation data with asynchronous time-correlated single-photon counting. Opt. Express 11(26):3583-3591.

35. Gregor, I., and J. Enderlein. 2007. Time-resolved methods in biophysics. 3. Fluorescence lifetime correlation spectroscopy. Photochem Photobiol Sci 6(1):13-18.

36. Kapusta, P., R. Machan, A. Benda, and M. Hof. 2012. Fluorescence Lifetime Correlation Spectroscopy (FLCS): concepts, applications and outlook. Int. J. Mol. Sci. 13(10):12890-12910.

37. Baumann, J., and M. D. Fayer. 1986. Excitation Transfer in Disordered Two-Dimensional and Anisotropic 3-Dimensional Systems - Effects of Spatial Geometry on Time-Resolved Observables. Journal of Chemical Physics 85(7):4087-4107.

38. Schwille, P., F. J. Meyer-Almes, and R. Rigler. 1997. Dual-color fluorescence cross-correlation spectroscopy for multicomponent diffusional analysis in solution. Biophys. J. 72(4):1878-1886.

39. Ptacek, J., J. Nedvedova, M. Navratil, B. Havlinova, J. Konvalinka, and C. Barinka. 2018. The calcium-binding site of human glutamate carboxypeptidase II is critical for dimerization, thermal stability, and enzymatic activity. Protein Sci. 27(9):1575-1584.

40. Schulke, N., O. A. Varlamova, G. P. Donovan, D. Ma, J. P. Gardner, D. M. Morrissey, R. R. Arrigale, C. Zhan, A. J. Chodera, K. G. Surowitz, P. J. Maddon, W. D. Heston, and W. C. Olson. 2003. The homodimer of prostate-specific membrane antigen is a functional target for cancer therapy. Proc Natl Acad Sci U S A 100(22):12590-12595.

41. Laurence, T. A., S. Ly, F. Bourguet, N. O. Fischer, and M. A. Coleman. 2014. Fluorescence correlation spectroscopy at micromolar concentrations without optical nanoconfinement. J Phys Chem B 118(32):9662-9667.

42. Magde, D., W. W. Webb, and E. Elson. 1972. Thermodynamic Fluctuations in a Reacting System - Measurement by Fluorescence Correlation Spectroscopy. Phys. Rev. Lett. 29(11):705-&.

43. Wozniak, A. K., G. F. Schroder, H. Grubmuller, C. A. Seidel, and F. Oesterhelt. 2008. Single-molecule FRET measures bends and kinks in DNA. Proc Natl Acad Sci U S A 105(47):18337-18342.

44. Diez, M., B. Zimmermann, M. Borsch, M. Konig, E. Schweinberger, S. Steigmiller, R. Reuter, S. Felekyan, V. Kudryavtsev, C. A. Seidel, and P. Graber. 2004. Proton-powered subunit rotation in single membrane-bound F0F1-ATP synthase. Nat. Struct. Mol. Biol. 11(2):135-141.

45. Munro, S., and H. R. Pelham. 1987. A C-terminal signal prevents secretion of luminal ER proteins. Cell 48(5):899-907.

46. Lewis, M. J., D. J. Sweet, and H. R. Pelham. 1990. The ERD2 gene determines the specificity of the luminal ER protein retention system. Cell 61(7):1359-1363.

47. Lewis, M. J., and H. R. Pelham. 1990. A human homologue of the yeast HDEL receptor. Nature 348(6297):162-163.

48. Matsukawa, S., Y. Moriyama, T. Hayata, H. Sasaki, Y. Ito, M. Asashima, and H. Kuroda. 2012. KDEL tagging: a method for generating dominant-negative inhibitors of the secretion of TGF-beta superfamily proteins. Int. J. Dev. Biol. 56(5):351-356.
49. Dale, R. E., J. Eisinger, and W. E. Blumberg. 1979. Orientational Freedom of Molecular Probes - Orientation Factor in Intra-Molecular Energy-Transfer. Biophysical Journal 26(2):161-193.
50. Akrap, N., T. Seidel, and B. G. Barisas. 2010. Forster distances for fluorescence resonant energy transfer between mCherry and other visible fluorescent proteins. Anal. Biochem. 402(1):105-106.
51. Kreutzberger, A. J. B., and S. Urban. 2018. Single-Molecule Analyses Reveal Rhomboid Proteins Are Strict and Functional Monomers in the Membrane. Biophys. J. 115(9):1755-1761.
52. Zettl, M., C. Adrain, K. Strisovsky, V. Lastun, and M. Freeman. 2011. Rhomboid family pseudoproteases use the ER quality control machinery to regulate intercellular signaling. Cell 145(1):79-91.
53. Greenblatt, E. J., J. A. Olzmann, and R. R. Kopito. 2011. Derlin-1 is a rhomboid pseudoprotease required for the dislocation of mutant alpha-1 antitrypsin from the endoplasmic reticulum. Nat. Struct. Mol. Biol. 18(10):1147-1152.
54. Adrain, C., M. Zettl, Y. Christova, N. Taylor, and M. Freeman. 2012. Tumor necrosis factor signaling requires iRhom2 to promote trafficking and activation of TACE. Science 335(6065):225-228.
55. Neal, S., P. A. Jaeger, S. H. Duttke, C. Benner, K. G. C, T. Ideker, and R. Y. Hampton. 2018. The Dfm1 Derlin Is Required for ERAD Retrotranslocation of Integral Membrane Proteins. Mol. Cell 69(2):306-320 e304.
56. Goder, V., P. Carvalho, and T. A. Rapoport. 2008. The ER-associated degradation component Der1p and its homolog Dfm1p are contained in complexes with distinct cofactors of the ATPase Cdc48p. FEBS Lett. 582(11):1575-1580.
57. Lazareno-Saez, C., E. Arutyunova, N. Coquelle, and M. J. Lemieux. 2013. Domain swapping in the cytoplasmic domain of the Escherichia coli rhomboid protease. J. Mol. Biol. 425(7):1127-1142.

Figure Legends

Fig. 1: The experimental system used and fluorescence cross-correlation analysis of dimerization of RHBDL2 and GCPII.

(A-C) Construct schemes for eGFP-RHBDL2 and mCherry-RHBDL2 **(A)**, eGFP-GCPII and mCherry-GCPII **(B)**, and eGFP-R2Ncyto-His$_6$ and mCherry-R2Ncyto-His$_6$ **(C)**. **(D, E)** Western blots showing expression and integrity of eGFP-RHBDL2 and mCherry-RHBDL2 **(D)**, and eGFP-GCPII and mCherry-GCPII **(E)**. **(F)** Coommassie-stained SDS PAGE showing integrity of recombinant eGFP-R2Ncyto-His$_6$ and mCherry-R2Ncyto-His$_6$. **(G, H)** Live-cell GFP fluorescence showing subcellular localization of eGFP-RHBDL2 **(G)** and eGFP-GCPII **(H)** expressed in HeLa cells. Scale bars mark 10 μm. **(I)** Scheme of GUVs spiked with DGS-NTA(Ni). **(J-L)** Fluorescence images illustrating GPMVs containing eGFP-RHBDL2 and mCherry-RHBDL2 **(J)** (scale bar 2 μm), eGFP-GCPII and mCherry-GCPII **(K)** (scale bar 2 μm), and NTA-decorated GUVs containing surface-bound eGFP-R2Ncyto-His$_6$ or mCherry-R2Ncyto-His$_6$ **(L)** (scale bar 5 μm). **(M-O)**, Auto- and cross-correlation functions of eGFP-RHBDL2 (green) and mCherry-RHBDL2 (red) in GPMVs **(M)**, eGFP-GCPII (green) and mCherry-GCPII (red) in GPMVs **(N)**, and R2Ncyto-His$_6$ (green) and mCherry-R2Ncyto-His$_6$ (red) bound to DGS-NTA(Ni)-spiked GUVs **(O)**. Blue lines denote the cross-correlation functions. Representative measurements are displayed. A number of GPMVs/GUVs were measured for each construct pair, with detailed statistics of the FCCS experiment available in Fig. S4.

Fig. 2: Determination of protein surface densities in the membrane of giant liposomes.

(**A**) Scheme of the calibration experiment: FCCS determining the protein concentration was acquired in the apical membrane of the GUVs – the magenta laser light profile; the FRET experiment requiring the equatorial intensity of the fluorescent proteins was carried out in the middle section of the GUVs – orange ellipse. (**B**) Linear dependence of the fluorescent protein surface concentration on the mean equatorial intensity; green and red lines denote His$_6$-eGFP and for His$_6$-mCherry, respectively. The dotted lines show the extrapolation to the concentrations for which FCCS cannot be carried out.

Fig. 3: Determination of the R_{0app} for two fluorescent proteins attached to the membrane.

(**A**) FRET efficiency as a function of acceptor concentration: black squares denote the measured data; circles denote the simulated data with increasing apparent Förster radius. The best agreement between the measured and the simulated data was found for the apparent Förster radius is 50-58 Å. The black line is a fit of experimental data by a numerical model introduced by Snyder and Freire giving the value of R_{0app} = 54 Å. (**B**) MC simulation of the fluorescence decay (grey line) obtained for a selected GPMV; colored lines are obtained from the simulation with increasing apparent Förster radius. The best agreement between the data and the simulation was obtained for R_{0app} = 58 Å. The insets show the dependence of MSD (mean square deviation), calculated from the measured data and each of the simulated dependences, on the apparent Förster radius. Table S2 lists all concentrations of donor and acceptors.

Fig. 4: Analysis of dimerization of RHBDL2 and GCPII in GPMVs derived from HeLa cells by FRET measurements and MC simulations.

(**A**) Comparison of acceptor concentration dependent FRET efficiencies obtained experimentally (black squares) and from MC simulations (circles). The MC simulations were carried out for RHBDL2 and GCPII, at increasing K_D values, and for excluded radii ranging from 30 to 60 Å and from 20 to 70 Å for RHBDL2 and GCPII, respectively. Donor concentrations are not depicted for the sake of simplicity, but for every analyzed GPMV, they were used as input parameters for the MC simulations. (**B**) Heat-maps of $-\log(MSD)$ visualizing the agreement between the measured and the simulated data for various values of K_D and the excluded radii. Table S2 lists all concentrations of donor and acceptors. $pK_D = -\log(K_D)$; MSD – mean square deviation;

Fig. 5: Analysis of dimerization of RHBDL2 and GCPII in GPMVs derived from HeLa cells by FLIM-FRET measurements and MC simulations.
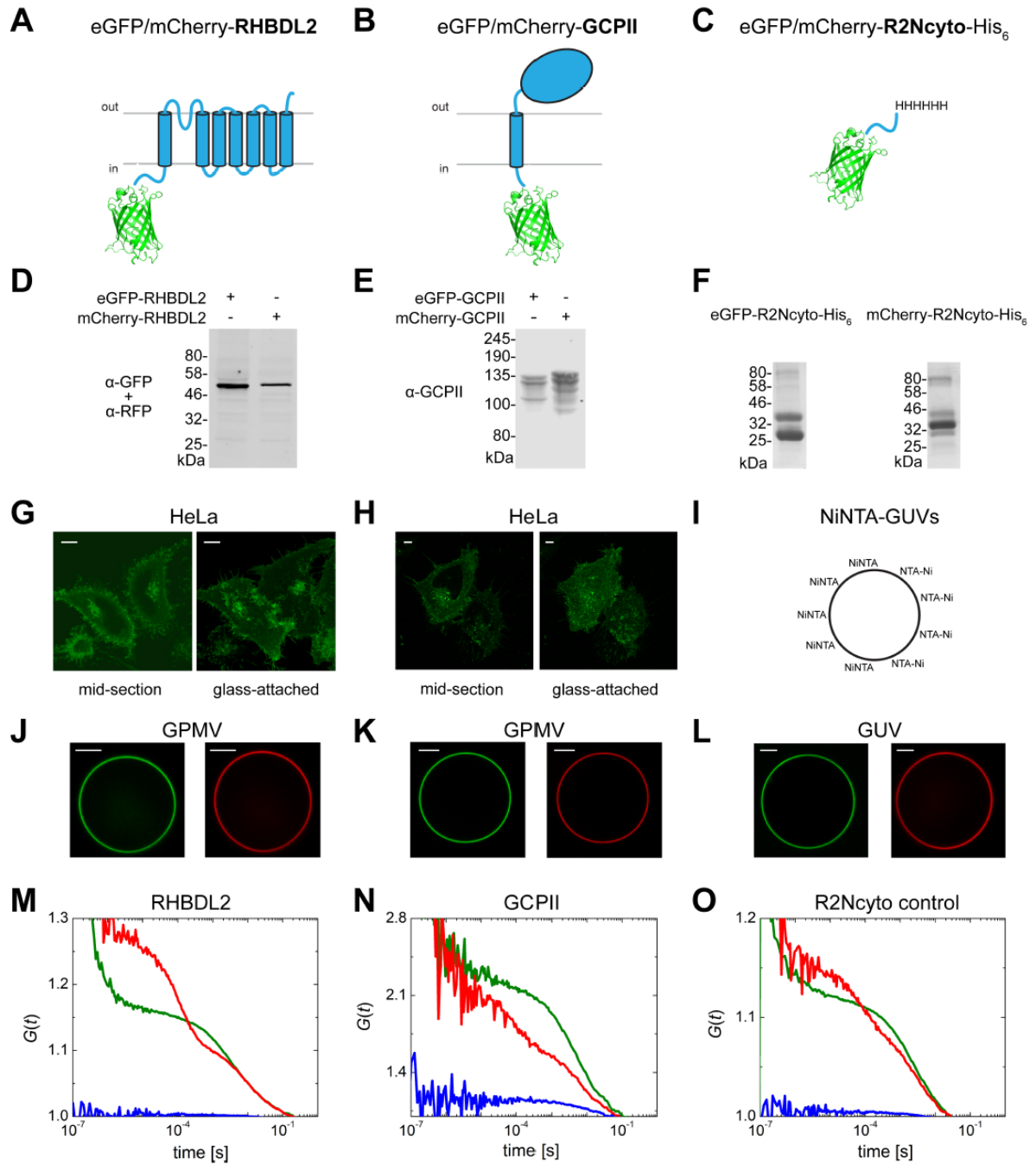
(**A**) Comparison donor fluorescence decay (grey line) obtained at given lateral concentration of donors and acceptors with decay curves obtained from MC simulations (colored lines). The MC simulations were carried out for RHBDL2 and GCPII, at increasing K_D and L_0 values. (**B**) Heat-maps of $-\log$ (MSD) visualizing the agreement between the measured and the simulated data for various values of K_D and the excluded radii L_0. $pK_D = -\log(K_D)$; MSD – mean square deviation;

Fig. 6: Relocalization analysis of RHBDL2 in live cells

Fluorescent constructs of human RHBDL2 fused to either eGFP or mCherry with or without the ER-retaining KDEL signal fused to the very C-terminus of each protein were co-expressed in HeLa cells, and live cell fluorescence was recorded 20-24 hrs after transfection. **A**, eGFP-RHBDL2 co-expressed with mCherry-RHBDL2; **B**, eGFP-RHBDL2-KDEL co-expressed with mCherry-RHBDL2-KDEL; **C**, eGFP-RHBDL2 co-expressed with mCherry-RHBDL2-KDEL. Note that while both fusions show strong plasma membrane localization including filopodia (A), KDEL tagging effectively relocalizes both fusions to the ER (B), while KDEL tagging of only one of the fusion proteins does not relocalize the other co-expressed one (C), meaning that the two fusion proteins do not stably interact with one another within the cell. Scale bars are 5 μm. Images of more cells from panels A – C are shown in Fig. S8.

A eGFP/mCherry-**RHBDL2**

out

in

B eGFP/mCherry-**GCPII**

out

in

C eGFP/mCherry-**R2Ncyto**-His$_6$

HHHHHH

D

eGFP-RHBDL2   +   -
mCherry-RHBDL2 -   +

α-GFP
+
α-RFP

80-
58-
46-
32-
25-
kDa

E

eGFP-GCPII   +   -
mCherry-GCPII -   +

α-GCPII

245-
190-
135-
100-
80-
kDa

F

eGFP-R2Ncyto-His$_6$ mCherry-R2Ncyto-His$_6$

80-
58-
46-
32-
25-
kDa

80-
58-
46-
32-
25-
kDa

G HeLa

mid-section   glass-attached

H HeLa

mid-section   glass-attached

I NiNTA-GUVs

NiNTA   NTA-Ni
NiNTA   NTA-Ni
NiNTA   NTA-Ni
NiNTA   NTA-Ni
NiNTA   NTA-Ni

J GPMV

K GPMV

L GUV

M RHBDL2

$G(t)$

1.3
1.2
1.1
1.0

10^{-7} 10^{-4} 10^{-1}
time [s]

N GCPII

$G(t)$

2.8
2.1
1.4

10^{-7} 10^{-4} 10^{-1}
time [s]

O R2Ncyto control

$G(t)$

1.2
1.1
1.0

10^{-7} 10^{-4} 10^{-1}
time [s]

A

B

1000*protein concentration [nm^{-2}]

mean equatorial intensity [a.u.]

A

RHBDL2

GCPII

B

123

A eGFP-RHBDL2 mCherry-RHBDL2 merge

B eGFP-RHBDL2-KDEL mCherry-RHBDL2-KDEL merge

C eGFP-RHBDL2 mCherry-RHBDL2-KDEL merge

7 Discussion

7.1 Substrate binding, specificity and reaction mechanism of rhomboid protease GlpG

The phenomenon of intramembrane proteolysis was first described about twenty years ago [7]. Since then, we have learn that this mechani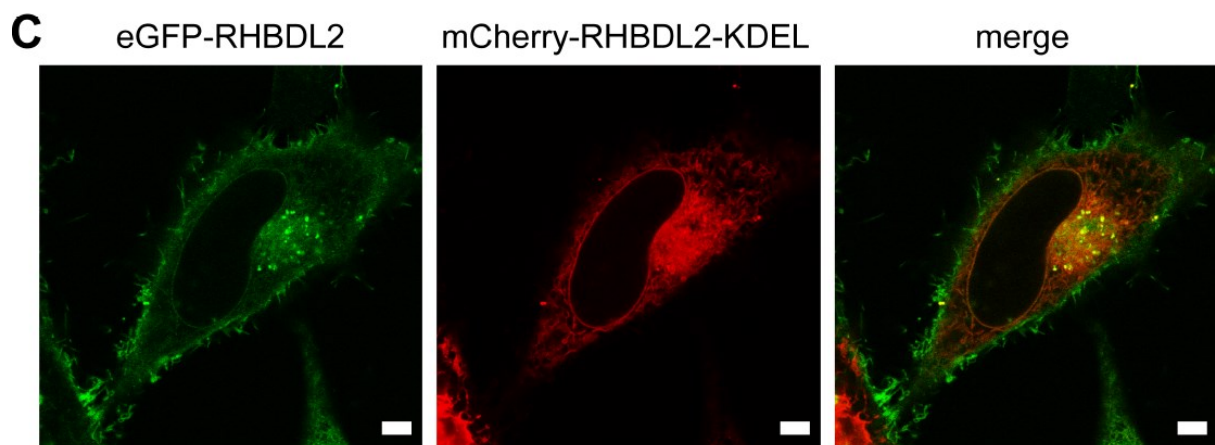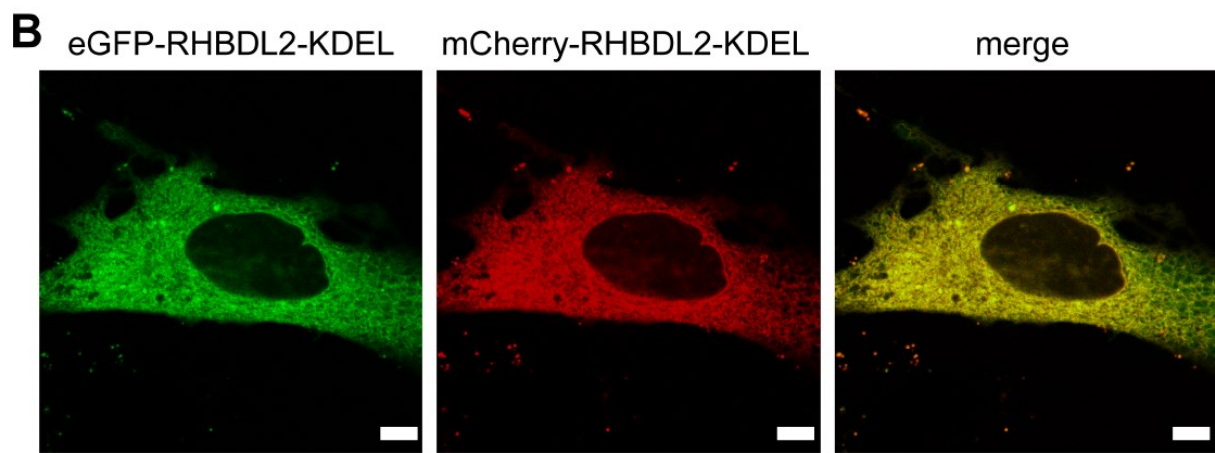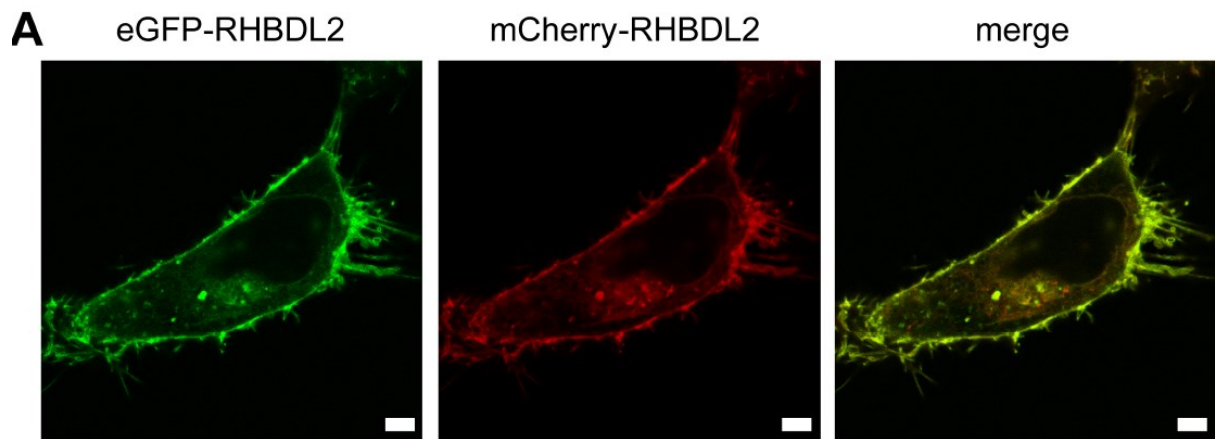sm is linked with many important cell biological processes, from protein quality control and degradation [40] to cell proliferation [11] and differentiation [9], and also with many diseases like Parkinson's disease [109] and cancer [108]. To develop effective drugs targeting intramembrane proteolysis we have to understand it mechanistically, and acquiring detailed information about substrate-enzyme interaction is key. To this end we have studied the main model intramembrane protease – *E. coli* GlpG from the rhomboid-like superfamily.

With the help of X-ray crystallography, we have solved high-resolution structures of GlpG in complexes with peptidyl-chloromethylketone inhibitors. We have identified the substrate residues at P1 – P4 subsites which are crucial for substrate recognition and catalysis [102]. We have also demonstrated that the S4 subsite is formed by the L1 loop, which is highly conserved among rhomboids and function of which was unknown. We speculate that its role in the enzyme-substrate recognition and cleavage reaction is also conserved in other members of the rhomboid superfamily. The combination of our structural data and biochemical analyses with previously published work [136,137] enabled us to outline a model of enzyme-substrate complex, which showed an interaction between the P4 to P3' segments of the substrate and the GlpG active site. Substrate residues at positions P4, P1 and P2' strongly influence the k_{cat} of the cleavage reaction, whereas they play minimal role in the substrate binding event, because the K_M remains almost unaffected [136]. Taken together, the interaction interface between the enzyme and substrate is much larger than just the segment comprising residues at positions P4 to P2' and this region of the substrate is responsible only for a small part of the total binding energy of the substrate. We significantly contributed to the clarification of the mechanism of substrate binding by rhomboid proteases. The full details of

substrate recruitment by the enzyme, such as the structure of the transmembrane helix of the substrate bound to rhomboid, however, still remain to be elucidated.

To build a structural model of GlpG with the full transmembrane domain of the substrate, we used an NMR structure of *E. coli* TatA [137]. Our model of the homologous *P. stuartii* TatA showed that its TMH is about 22 Å long, flanked by residues P13 and F27. In contrast, the thickness of the GlpG membrane portion is about 13 Å. Manual docking of this substrate TMH to the GlpG enzyme showed that a part of the TatA substrate TMD would stick out of the membrane (Figure 22). This situation is very unlikely due to the energy demand of such hydrophobic mismatch. Most probably, the substrate would try to avoid this by one of the two possible strategies. In the first case, the substrate TMD could be tilted or kinked to be completely hidden in the membrane; in this case the interaction of the transmembrane portion of the substrate with the enzyme would be minimal. In the second scenario, the substrate TMD could be curved and interact with GlpG to avoid interaction with the hydrophilic environment outside the membrane (Figure 22). The interface between GlpG and its substrate would thus be much larger in the second scenario, which is supported by our data and which would thus be more likely. In addition, similar behavior has been described for a region in a different protein a with a hydrophobic mismatch larger than 4 Å [138].
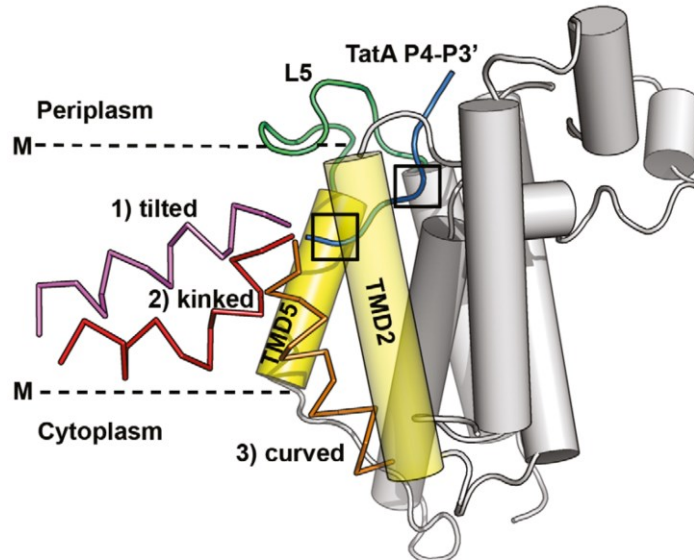
Figure 22: Structural model of GlpG protease with the TatA substrate TMH. Three different possibilities of substrate TMH orientation within a membrane are distinguished by colors. Adopted from [139].

We propose that the interaction interface between the substrate and the enzyme is large, by showing the importance of the transmembrane region of the substrate for its recognition and catalysis using a series of C-terminally truncated transmembrane peptide substrates. These substrates are derivatives of LacYTM2, which is readily cleaved by diverse rhomboid proteases. These peptides were cleaved by the enzyme only in their full transmembrane versions, but their derivatives truncated by more than 5 residues from the transmembrane C-terminus were poorly utilized [140]. We tested whether LacYTM2 could also be used as a fluorogenic substrates when residues at the P5 and P4' positions were modified by the EDANS/DABCYL FRET pair. These positions are not critical for rhomboid cleavage [102,139] and GlpG action was not affected by their mutations. The 7 amino acid residues distance between the two fluorophores is short enough for FRET to occur in the uncleaved substrate (EDANS is quenched by DABCYL) and cleavage is thus accompanied with an increase in fluorescence of EDANS. Such substrate can also be used in liposomes (Figure 19), where it is cleaved similarly as in detergent micelles. We also verified with CD spectroscopy that the helical structure of the TMH is unaffected. In addition, we experimentally confirmed that changing the sequence of the substrate according to GlpG

specificity [139] results in a more efficient GlpG-specific cleavage. By mutating all five residues at positions P5 to P1 of the fluorogenic LacYTM2 derivate, we reached 23-fold more efficient cleavage as compared to the wild type. We have also shown that by using this modification approach substrates can be easily adapted for efficient cleavage by different rhomboid proteases.

Furthermore, we used the information about GlpG substrate specificity to generate a set of novel peptidyl-α-ketoamide inhibitors. We identified ketoamide as a promising electrophilic warhead, which exceeds other electrophilic reversibly binding warheads commonly used for serine proteases that we tested (trifluoromethylketones, boronates, acylsulfonamides, thiazolylketones) [141,142]. We experimentally confirmed that a modification of the peptidyl-α-ketoamides at their prime side of the active site increases their ability to inhibit rhomboids, presumably because the hydrophobic (in case of GlpG) P2' residue of the substrate is important for rhomboid substrate recognition [102,136]. We have confirmed the selectivity of the peptidyl-α-ketoamides by a high-throughput assay for testing inhibitor potency and specificity [143], and a kinetic and structural analysis of our compounds revealed that they inhibit the rhomboid in a substrate-like manner and that the inhibition is covalent and reversible. They inhibited *E. coli* GlpG *in vivo* with up to 2 nM IC50. The inhibitors are modular because the peptidyl part of the inhibitor and the ketoamide warhead modification can be tailored separately, yielding selective inhibitors. In summary, we generated a new class of potent and selective rhomboid inhibitors that are superior to all other currently used rhomboid inhibitors and have a straightforward application in biology and drug discovery.

7.2 Dimerization properties of membrane proteases from rhomboid family

Dimerization of rhomboid proteases was reported as a mechanism regulating their activity *in vitro* [115]. However, this report was based only on experiments in a solubilized state in detergent micelles and the *in vivo* situation was not studied. The importance of this topic is confirmed by the fact that the same question was also addressed by other researchers.

Their observation is consistent with our results that the rhomboid behaves as a monomeric entity in natural lipid membrane [116]. By an integration of biophysical fluorescent techniques (FRET, FCCS), we focused on registration of strong and also weak interactions between two proteins. Our constructs of fluorescently labeled human rhomboid RHBDL2 were designed with respect to their usage on membrane surface (Figure 21). We paid special attention to the linker between the protein of interest and the fluorescent marker, as the distance between the fluorophores is one of FRET intensity determinants. We also needed to avoid misinterpretation and false positive results of FRET experiments caused by the 'proximity effect' [144] of overexpressed protein in high density. In the first place, we had the protein expression and localization under strict control and then we also used the Monte Carlo simulations to confirm the relevance of our observations. During its maturation, a membrane protein is exposed to varying lipid environment in different cellular organelles [145] and this can of course also affect the protein oligomeric state [146]. Taking this scenario into consideration, we also focused on the dimeric state of RHBDL2 in endoplasmic reticulum. Using this approach, we examined the oligomeric state of human rhomboid RHBDL2 under about 20-fold overexpression above physiological concentration, and we found no evidence of its dimerization in natural biomembranes after or during its maturation. This suggests, contrary to the current belief based on *in vitro* experiments, that dimerization may not be a universal regulatory mechanism in rhomboid proteases.

8   Conclusions

In this work we focused on the mechanism, specificity and potential drug targeting of intramembrane proteases from the rhomboid family using biochemical and biophysical approaches. By X-ray crystallography combined with enzymatic assay and *in silico* modeling, we focused on the interaction between *E. coli* rhomboid protease GlpG and its substrate. We also used the non-invasive spectroscopic techniques FRET/FCCS in combination with MC simulations to investigate the oligomerization of human rhomboid RHBDL2 in native cell-derived biomembranes.

We elucidated the specificity of GlpG by positional scanning mutagenesis of rhomboid substrate TatA. Our observations confirmed the importance of substrate residues at positions P4 and P1. We found that some substrate mutations inhibit cleavage, but GlpG was also surprisingly more efficient in cleaving some mutant substrates. From our structural data on GlpG in complex with TatA-based peptidyl-chloromethylketone inhibitors, we identified the S1 to S4 subsites of the protease and elucidated the role of the conserved L1 loop, which co-forms the S4 subsite. Furthermore, we used molecular dynamics to build a model of GlpG with the TatA substrate bound in its active site.

We also focused on rhomboid proteases as potential therapeutic targets for inhibitor development. Detailed information about GlpG interaction with substrate helped us design a variety of fluorogenic substrates based on the sequence of the promiscuous rhomboid substrate TatATM2. By employing functional rhomboid cleavage assay, we confirmed the applicability of our fluorogenic substrates in the environment of detergent micelles and also in liposomes. Our substrates can be also easily adapted for different rhomboid proteases by mutating residues in the peptidyl part of the substrate at the P1 to P5 positions. By applying principles similar to those for the fluorogenic substrates, we further developed peptidyl-α-ketoamide inhibitors of rhomboids. These inhibitors bind the enzyme in a substrate-like manner, covalently and reversibly. The inhibition is selective and the inhibitors work in nanomolar range. Their advantage is that they can be easily adapted to different rhomboid

proteases by inhibitor warhead modification and also by alteration of amino acid residues in the peptidyl part.

We analyzed *in vivo* the behavior of human rhomboid RHBLD2. By spectroscopic techniques FCCS and FRET in combination with *in vivo* imagining of rhomboid localization and MC simulation, we investigated the tendency of rhomboid to dimerize. Although rhomboid dimerization and related activation was already described in detergent micelles *in vitro*, we found no evidence for occurrence of this phenomenon *in vivo*. The methodological approach we have used to study RHBDL2 is generally applicable to any eukaryotic membrane protein.

9 References

1. López-Otín C, Bond JS (2008) Proteases: multifunctional enzymes in life and disease. *J Biol Chem* **283**: 30433–30437.

2. Quesada V, Ordóñez GR, Sánchez LM, Puente XS, López-Otín C (2009) The Degradome database: mammalian proteases and diseases of proteolysis. *Nucleic Acids Res* **37**: D239-43.

3. Dodson G, Chothia C (1984) Protein crystallography: Fifty years of pepsin crystals. *Nature* **309**: 309–309.

4. Southan C (2001) A genomic perspective on human proteases as drug targets. *Drug Discov Today* **6**: 681–688.

5. Tavano OL (2013) Protein hydrolysis using proteases: An important tool for food biotechnology. *J Mol Catal B Enzym* **90**: 1–11.

6. Erez E, Fass D, Bibi E (2009) How intramembrane proteases bury hydrolytic reactions in the membrane. *Nature* **459**: 371–378.

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

8. Wolfe MS, Kopan R (2004) Intramembrane proteolysis: Theme and variations. *Science (80-)* **305**: 1119–1123.

9. Brown MS, Ye J, Rawson RB, Goldstein JL (2000) Regulated intramembrane proteolysis: a control mechanism conserved from bacteria to humans. *Cell* **100**: 391–398.

10. Manolaridis I, Kulkarni K, Dodd RB, Ogasawara S, Zhang Z, Bineva G, O'Reilly N, Hanrahan SJ, Thompson AJ, Cronin N, et al. (2013) Mechanism of farnesylated CAAX protein processing by the intramembrane protease Rce1. *Nature*.

11. Yoshida R, Nagata M, Nakayama H, Niimori-Kita K, Hassan W, Tanaka T, Shinohara M, Ito T (2013) The pathological significance of Notch1 in oral squamous cell carcinoma. *Lab Investig* **93**: 1068–1081.

12. Avci D, Lemberg MK (2015) Clipping or Extracting: Two Ways to Membrane Protein Degradation. *Trends Cell Biol* **25**: 611–622.

13. Marambaud P, Shioi J, Serban G, Georgakopoulos A, Sarner S, Nagy V, Baki L, Wen P, Efthimiopoulos S, Shao Z, et al. (2002) A presenilin-1/γ-secretase cleavage releases the E-cadherin intracellular domain and regulates disassembly of adherens junctions. *EMBO J* **21**: 1948–1956.

14. Shanbhag R, Shi G, Rujiviphat J, McQuibban GA (2012) The emerging role of proteolysis in mitochondrial quality control and the etiology of Parkinson's disease. *Parkinsons Dis* **2012**:.

15. Jurisch-Yaksi N, Sannerud R, Annaert W (2013) A fast growing spectrum of biological functions of γ-secretase in development and disease. *Biochim Biophys Acta - Biomembr* **1828**: 2815–2827.

16. McCarthy J V., Twomey C, Wujek P (2009) Presenilin-dependent regulated intramembrane proteolysis and γ-secretase activity. *Cell Mol Life Sci* **66**: 1534–1555.

17. P. Hurst T, Coleman-Vaughan C, Patwal I, V. McCarthy J (2016) Regulated intramembrane proteolysis, innate immunity and therapeutic targets in Alzheimer's disease. *AIMS Mol Sci* **3**: 138–157.

18. De Strooper B, Saftig P, Craessaerts K, Vanderstichele H, Guhde G, Annaert W, Von Figura K, Van Leuven F (1998) Deficiency of presenilin-1 inhibits the normal cleavage of amyloid precursor protein. *Nature* **391**: 387–390.

19. De Strooper B, Annaert W, Cupers P, Saftig P, Craessaerts K, Mumm JS, Schroeter EH, Schrijvers V, Wolfe MS, Ray WJ, et al. (1999) A presenilin-1-dependet g-secretase-like protease mediates release of Notch intracelluar domain. *Nature* **398**: 518–522.

20. Guruharsha KG, Kankel MW, Artavanis-Tsakonas S (2012) The Notch signalling system: recent insights into the complexity of a conserved pathway. *Nat Rev Genet* **13**: 654.

21. Fluhrer R, Grammer G, Israel L, Condron MM, Haffner C, Friedmann E, Böhland C, Imhof A, Martoglio B, Teplow DB, et al. (2006) A γ-secretase-like intramembrane cleavage of TNFα by the GxGD aspartyl protease SPPL2b. *Nat Cell Biol* **8**: 894–896.

22. Elzinga BM, Twomey C, Powell JC, Harte F, McCarthy J V. (2009) Interleukin-1 receptor type 1 is a substrate for γ-secretase- dependent regulated intramembrane proteolysis. *J Biol Chem* **284**: 1394–1409.

23. Kuhn PH, Marjaux E, Imhof A, De Strooper B, Haass C, Lichtenthaler SF (2007) Regulated intramembrane proteolysis of the interleukin-1 receptor II by α-, β-, and γ-secretase. *J Biol Chem* **282**: 11982–11995.

24. Twomey C, Qian S, McCarthy J V. (2009) TRAF6 promotes ubiquitination and regulated intramembrane proteolysis of IL-1R1. *Biochem Biophys Res Commun* **381**: 418–423.

25. McElroy B, Powell JC, McCarthy J V. (2007) The insulin-like growth factor 1 (IGF-1) receptor is a substrate for γ-secretase-mediated intramembrane proteolysis. *Biochem Biophys Res Commun* **358**: 1136–1141.

26. Sturtevant MA, Roark M, Bier E (1993) The Drosophila rhomboid gene mediates the

localized formation of wing veins and interacts genetically with components of the EGF-R signaling pathway. *Genes Dev* **7**: 961–973.

27. Ni C-Y (2001) gamma -Secretase Cleavage and Nuclear Localization of ErbB-4 Receptor Tyrosine Kinase. *Science (80-)* **294**: 2179–2181.

28. Powell JC, Twomey C, Jain R, McCarthy J V. (2009) Association between Presenilin-1 and TRAF6 modulates regulated intramembrane proteolysis of the p75NTRneurotrophin receptor. *J Neurochem* **108**: 216–230.

29. Berghoff J, Jaisimha AV, Duggan S, MacSharry J, McCarthy J V. (2015) Gamma-secretase-independent role for cadherin-11 in neurotrophin receptor p75 (p75NTR) mediated glioblastoma cell migration. *Mol Cell Neurosci* **69**: 41–53.

30. Zampieri N, Xu CF, Neubert TA, Chao M V. (2005) Cleavage of p75 neurotrophin receptor by α-secretase and γ-secretase requires specific receptor domains. *J Biol Chem* **280**: 14563–14571.

31. Lammich S, Okochi M, Takeda M, Kaether C, Capell A, Zimmer AK, Edbauer D, Walter J, Steiner H, Haass C (2002) Presenilin-dependent intramembrane proteolysis of CD44 leads to the liberation of its intracellular domain and the secretion of an Aβ-like peptide. *J Biol Chem* **277**: 44754–44759.

32. Wunderlich P, Glebov K, Kemmerling N, Tien NT, Neumann H, Walter J (2013) Sequential proteolytic processing of the triggering receptor expressed on myeloid cells-2 (TREM2) protein by ectodomain shedding and γ-secretase- dependent intramembranous cleavage. *J Biol Chem* **288**: 33027–33036.

33. Maetzel D, Denzel S, Mack B, Canis M, Went P, Benk M, Kieu C, Papior P, Baeuerle PA, Munz M, et al. (2009) Nuclear signalling by tumour-associated antigen EpCAM. *Nat Cell Biol* **11**: 162–171.

34. McCarthy AJ, Coleman-Vaughan C, McCarthy J V. (2017) Regulated intramembrane proteolysis: emergent role in cell signalling pathways. *Biochem Soc Trans* **45**: 1185–1202.

35. Urban S, Schlieper D, Freeman M (2002) Conservation of intramembrane proteolytic activity and substrate specificity in prokaryotic and eukaryotic rhomboids. *Curr Biol* **12**: 1507–1512.

36. Akiyama Y, Maegawa S (2007) Sequence features of substrates required for cleavage by GlpG, an Escherichia coli rhomboid protease. *Mol Microbiol* **64**: 1028–1037.

37. Strisovsky K (2016) Why cells need intramembrane proteases - a mechanistic perspective. *FEBS J* **283**: 1837–1845.

38. Lemberg MK, Martoglio B (2002) Requirements for signal peptide peptidase-catalyzed intramembrane proteolysis. *Mol Cell* **10**: 735–744.

39. Urban S, Freeman M (2003) Substrate specificity of rhomboid intramembrane proteases is governed by helix-breaking residues in the substrate transmembrane domain. *Mol Cell* **11**: 1425–1434.

40. Fleig L, Bergbold N, Sahasrabudhe P, Geiger B, Kaltak L, Lemberg MK (2012) Ubiquitin-Dependent Intramembrane Rhomboid Protease Promotes ERAD of Membrane Proteins. *Mol Cell* **47**: 558–569.

41. Began J, Cordier B, Březinová J, Delisle J, Hexnerová R, Srb P, Rampírová P, Kožíšek M, Baudet M, Couté Y, et al. (2020) Rhomboid intramembrane protease YqgP licenses bacterial membrane protein quality control as adaptor of FtsH <scp>AAA</scp> protease. *EMBO J.*

42. Wolfe MS (2009) Intramembrane Proteolysis. *Chem Rev* **109**: 1599–1612.

43. Kopan R, Ilagan MXG (2004) Opinion: γ-Secretase: proteasome of the membrane? *Nat Rev Mol Cell Biol* **5**: 499–504.

44. Barrett PJ, Song Y, Van Horn WD, Hustedt EJ, Schafer JM, Hadziselimovic A, Beel AJ, Sanders CR (2012) The Amyloid Precursor Protein Has a Flexible Transmembrane Domain and Binds Cholesterol. *Science (80-)* **336**: 1168–1171.

45. Linser R, Salvi N, Briones R, Rovó P, de Groot BL, Wagner G (2015) The membrane anchor of the transcriptional activator SREBP is characterized by intrinsic conformational flexibility. *Proc Natl Acad Sci* **112**: 12390–12395.

46. Urban S, Freeman M (2002) Intramembrane proteolysis controls diverse signalling pathways throughout evolution. *Curr Opin Genet Dev* **12**: 512–518.

47. Espenshade PJ, Hughes AL (2007) Regulation of Sterol Synthesis in Eukaryotes. *Annu Rev Genet* **41**: 401–427.

48. Feng L, Yan H, Wu Z, Yan N, Wang Z, Jeffrey PD, Shi Y (2007) Structure of a site-2 protease family intramembrane metalloprotease. *Science* **318**: 1608–1612.

49. Zhou R, Cusumano C, Sui D, Garavito RM, Kroos L (2009) Intramembrane proteolytic cleavage of a membrane-tethered transcription factor by a metalloprotease depends on ATP. *Proc Natl Acad Sci U S A* **106**: 16174–16179.

50. Rawson RB (2013) The site-2 protease. *Biochim Biophys Acta - Biomembr* **1828**: 2801–2807.

51. Horton JD, Goldstein JL, Brown MS (2002) SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. *J Clin Invest* **109**: 1125–1131.

52. Ott CM, Lingappa VR (2002) Integral membrane protein biosynthesis: why topology is hard to predict. *J Cell Sci* **115**:.

53. Beel AJ, Sanders CR (2008) Substrate specificity of γ-secretase and other intramembrane proteases. *Cell Mol Life Sci* **65**: 1311–1334.

54. Lleó A (2008) Activity of gamma-secretase on substrates other than APP. *Curr Top Med Chem* **8**: 9–16.

55. Parks AL, Curtis D (2007) Presenilin diversifies its portfolio. *Trends Genet* **23**: 140–150.

56. Li N, Liu K, Qiu Y, Ren Z, Dai R, Deng Y, Qing H (2016) Effect of Presenilin Mutations on APP Cleavage; Insights into the Pathogenesis of FAD. *Front Aging Neurosci* **8**: 51.

57. St George-Hyslop PH (2000) Molecular genetics of Alzheimer's disease. *Biol Psychiatry* **47**: 183–199.

58. Clark RF, Hutton M, Fuldner M, Froelich S, Karran E, Talbot C, Crook R, Lendon C, Prihar G, He C, et al. (1995) The structure of the presenilin 1 (S182) gene and identification of six novel mutations in early onset AD families. *Nat Genet* **11**: 219–222.

59. Levy-Lahad E, Wasco W, Poorkaj P, Romano DM, Oshima J, Pettingell WH, Yu CE, Jondro PD, Schmidt SD, Wang K, et al. (1995) Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science (80-)* **269**: 973–977.

60. Thinakaran G, Borchelt DR, Lee MK, Slunt HH, Spitzer L, Kim G, Ratovitsky T, Davenport F, Nordstedt C, Seeger M, et al. (1996) Endoproteolysis of presenilin 1 and accumulation of processed derivatives in vivo. *Neuron* **17**: 181–190.

61. Hardy J (2007) Putting presenilins centre stage. Introduction to the Talking Point on the role of presenilin mutations in Alzheimer disease. *EMBO Rep* **8**: 134–135.

62. Chávez-Gutiérrez L, De Strooper B (2016) Probing γ-secretase-substrate interactions at the single amino acid residue level. *EMBO J* **35**: 1597–1599.

63. De Strooper B (2003) Aph-1, Pen-2, and Nicastrin with Presenilin generate an active gamma-Secretase complex. *Neuron* **38**: 9–12.

64. Holmes O, Paturi S, Selkoe DJ, Wolfe MS (2014) Pen-2 Is Essential for γ-Secretase Complex Stability and Trafficking but Partially Dispensable for Endoproteolysis. *Biochemistry* **53**: 4393–4406.

65. Kimberly WT, LaVoie MJ, Ostaszewski BL, Ye W, Wolfe MS, Selkoe DJ (2003) -Secretase is a membrane protein complex comprised of presenilin, nicastrin, aph-1, and pen-2. *Proc Natl Acad Sci* **100**: 6382–6387.

66. Struhl G, Adachi A (2000) Requirements for Presenilin-Dependent Cleavage of Notch and Other Transmembrane Proteins. *Mol Cell* **6**: 625–636.

67. Haapasalo A, Kovacs DM (2011) The many substrates of presenilin/γ-secretase. *J Alzheimers Dis* **25**: 3–28.

68. Mumm JS, Kopan R (2000) Notch Signaling: From the Outside In. *Dev Biol* **228**: 151–165.

69. Ponting CP, Hutton M, Nyborg A, Baker M, Jansen K, Golde TE (2002) Identification of a novel family of presenilin homologues. *Hum Mol Genet* **11**: 1037–1044.

70. Weihofen A, Binns K, Lemberg MK, Ashman K, Martoglio B (2002) Identification of Signal Peptide Peptidase, a Presenilin-Type Aspartic Protease. *Science (80-)* **296**: 2215–2218.

71. Grigorenko AP, Moliaka YK, Korovaitseva GI, Rogaev EI (2002) Novel class of polytopic proteins with domains associated with putative protease activity. *Biochemistry (Mosc)* **67**: 826–835.

72. Voss M, Schröder B, Fluhrer R (2013) Mechanism, specificity, and physiology of signal peptide peptidase (SPP) and SPP-like proteases. *Biochim Biophys Acta - Biomembr* **1828**: 2828–2839.

73. Urny J, Hermans-Borgmeyer I, Gercken G, Chica Schaller H (2003) Expression of the presenilin-like signal peptide peptidase (SPP) in mouse adult brain and during development. *Gene Expr Patterns* **3**: 685–691.

74. Martin L, Fluhrer R, Reiss K, Kremmer E, Saftig P, Haass C (2008) Regulated Intramembrane Proteolysis of Bri2 (Itm2b) by ADAM10 and SPPL2a/SPPL2b. *J Biol Chem* **283**: 1644–1652.

75. Behnke J, Schneppenheim J, Koch-Nolte F, Haag F, Saftig P, Schröder B (2011) Signal-peptide-peptidase-like 2a (SPPL2a) is targeted to lysosomes/late endosomes by a tyrosine motif in its C-terminal tail. *FEBS Lett* **585**: 2951–2957.

76. Urny J, Hermans-Borgmeyer I, Schaller HC (2006) Cell-surface expression of a new splice variant of the mouse signal peptide peptidase. *Biochim Biophys Acta - Gene Struct Expr* **1759**: 159–165.

77. Lemberg MK, Martoglio B (2004) On the mechanism of SPP-catalysed intramembrane proteolysis; conformational control of peptide bond hydrolysis in the plane of the membrane. *FEBS Lett* **564**: 213–218.

78. El Hage F, Stroobant V, Vergnon I, Baurain J-F, Echchakir H, Lazar V, Chouaib S, Coulie PG, Mami-Chouaib F (2008) 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**: 10119–10124.

79. Kapp K, Schrempf S, Lemberg MK, Dobberstein B (2013) Post-Targeting Functions of Signal Peptides.

80. Wu C-M, Chang MD-T (2004) Signal peptide of eosinophil cationic protein is toxic to cells lacking signal peptide peptidase. *Biochem Biophys Res Commun* **322**: 585–592.

81. Friedmann E, Lemberg MK, Weihofen A, Dev KK, Dengler U, Rovelli G, Martoglio B (2004) Consensus Analysis of Signal Peptide Peptidase and Homologous Human Aspartic Proteases Reveals Opposite Topology of Catalytic Domains Compared with Presenilins. *J Biol Chem* **279**: 50790–50798.

82. Kjos M, Snipen L, Salehian Z, Nes IF, Diep DB (2010) The abi proteins and their involvement in bacteriocin self-immunity. *J Bacteriol* **192**: 2068–2076.

83. Bergo MO, Ambroziak P, Gregory C, George A, Otto JC, Kim E, Nagase H, Casey PJ, Balmain A, Young SG (2002) Absence of the CAAX endoprotease Rce1: effects on cell growth and transformation. *Mol Cell Biol* **22**: 171–181.

84. Bergo MO, Lieu HD, Gavino BJ, Ambroziak P, Otto JC, Casey PJ, Walker QM, Young SG (2004) On the physiological importance of endoproteolysis of CAAX proteins: heart-specific RCE1 knockout mice develop a lethal cardiomyopathy. *J Biol Chem* **279**: 4729–4736.

85. Christiansen JR, Kolandaivelu S, Bergo MO, Ramamurthy V (2011) RAS-converting enzyme 1-mediated endoproteolysis is required for trafficking of rod phosphodiesterase 6 to photoreceptor outer segments. *Proc Natl Acad Sci* **108**: 8862–8866.

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

87. Jürgens G, Wieschaus E, Nüsslein-Volhard C, Kluding H (1984) Mutations affecting the pattern of the larval cuticle inDrosophila melanogaster. *Wilhelm Roux's Arch Dev Biol* **193**: 283–295.

88. Koonin E V, Makarova KS, Rogozin IB, Davidovic L, Letellier M-C, Pellegrini L (2003) The rhomboids: a nearly ubiquitous family of intramembrane serine proteases that probably evolved by multiple ancient horizontal gene transfers. *Genome Biol* **4**: R19.

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

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

91. Freeman M (2014) The Rhomboid-Like Superfamily: Molecular Mechanisms and Biological Roles. *Annu Rev Cell Dev Biol* **30**: 235–254.

92. Urban S, Wolfe MS (2005) Reconstitution of intramembrane proteolysis in vitro reveals that pure rhomboid is sufficient for catalysis and specificity. *Proc Natl Acad Sci U S A*

102: 1883–1888.

93.   Ben-Shem A, Fass D, Bibi E (2007) Structural basis for intramembrane proteolysis by rhomboid serine proteases. *Proc Natl Acad Sci U S A* **104**: 462–466.

94.   Lemieux MJ, Fischer SJ, Cherney MM, Bateman KS, James MNG (2007) The crystal structure of the rhomboid peptidase from *Haemophilus influenzae* provides insight into intramembrane proteolysis. *Proc Natl Acad Sci* **104**: 750–754.

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

96.   Baker RP, Young K, Feng L, Shi Y, Urban S (2007) Enzymatic analysis of a rhomboid intramembrane protease implicates transmembrane helix 5 as the lateral substrate gate Rosanna. *Proc Natl Acad Sci* **104**: 8257–8262.

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

98.   Brooks CL, Lazareno-Saez C, Lamoureux JS, Mak MW, Lemieux MJ (2011) Insights into Substrate Gating in H. influenzae Rhomboid. *J Mol Biol* **407**: 687–697.

99.   Xue Y, Ha Y (2013) Large lateral movement of transmembrane helix S5 is not required for substrate access to the active site of rhomboid intramembrane protease. *J Biol Chem* **288**: 16645–16654.

100.   Cho S, Baker RP, Ji M, Urban S (2019) Ten catalytic snapshots of rhomboid intramembrane proteolysis from gate opening to peptide release. *Nat Struct Mol Biol* **26**: 910–918.

101.   Moin SM, Urban S (2012) Membrane immersion allows rhomboid proteases to achieve specificity by reading transmembrane segment dynamics. *Elife* **1**: e00173.

102.   Strisovsky K, Sharpe HJ, Freeman M (2009) Sequence-specific intramembrane proteolysis: identification of a recognition motif in rhomboid substrates. *Mol Cell* **36**: 1048–1059.

103.   Kreutzberger AJB, Ji M, Aaron J, Mihaljević L, Urban S (2019) Rhomboid distorts lipids to break the viscosity-imposed speed limit of membrane diffusion. *Science (80-)*.

104.   Freeman M (2008) Rhomboid Proteases and their Biological Functions. *Annu Rev Genet* **42**: 191–210.

105.   Bergbold N, Lemberg MK (2013) Emerging role of rhomboid family proteins in mammalian biology and disease. *Biochim Biophys Acta - Biomembr* **1828**: 2840–2848.

106.   Hill RB, Pellegrini L (2010) The PARL family of mitochondrial rhomboid proteases. *Semin Cell Dev Biol* **21**: 582–592.

107. Dowse TJ, Soldati D (2005) Rhomboid-like proteins in Apicomplexa: phylogeny and nomenclature. *Trends Parasitol* **21**: 254–258.

108. Etheridge SL, Brooke MA, Kelsell DP, Blaydon DC (2013) Rhomboid proteins: A role in keratinocyte proliferation and cancer. *Cell Tissue Res* **351**: 301–307.

109. Chan EYL, McQuibban GA (2013) The mitochondrial rhomboid protease: Its rise from obscurity to the pinnacle of disease-relevant genes. *Biochim Biophys Acta - Biomembr* **1828**: 2916–2925.

110. Rather P (2013) Role of rhomboid proteases in bacteria. *Biochim Biophys Acta - Biomembr* **1828**: 2849–2854.

111. Stevenson LG, Strisovsky K, Clemmer KM, Bhatt S, Freeman M, Rather PN (2007) Rhomboid protease AarA mediates quorum-sensing in Providencia stuartii by activating TatA of the twin-arginine translocase. *Proc Natl Acad Sci U S A* **104**: 1003–1008.

112. Yang B, Larson TJ (1998) Multiple promoters are responsible for transcription of the glpEGR operon of Escherichia coli K-12. *Biochim Biophys Acta* **1396**: 114–126.

113. Maegawa S, Ito K, Akiyama Y (2005) Proteolytic Action of GlpG, a Rhomboid Protease in the *Escherichia coli* Cytoplasmic Membrane [†]. *Biochemistry* **44**: 13543–13552.

114. Clemmer KM, Sturgill GM, Veenstra A, Rather PN (2006) Functional characterization of Escherichia coli GlpG and additional rhomboid proteins using an aarA mutant of Providencia stuartii. *J Bacteriol* **188**: 3415–3419.

115. Arutyunova E, Panwar P, Skiba PM, Gale N, Mak MW, Lemieux MJ (2014) Allosteric regulation of rhomboid intramembrane proteolysis. 1–13.

116. Kreutzberger AJB, Urban S (2018) Single-Molecule Analyses Reveal Rhomboid Proteins Are Strict and Functional Monomers in the Membrane. *Biophys J* **115**: 1755–1761.

117. Knopf RR, Adam Z (2012) Rhomboid proteases in plants - still in square one? *Physiol Plant* **145**: 41–51.

118. Knopf RR, Feder A, Mayer K, Lin A, Rozenberg M, Schaller A, Adam Z (2012) Rhomboid proteins in the chloroplast envelope affect the level of allene oxide synthase in *Arabidopsis thaliana*. *Plant J* **72**: 559–571.

119. Düsterhöft S, Künzel U, Freeman M (2017) Rhomboid proteases in human disease: Mechanisms and future prospects. *Biochim Biophys Acta - Mol Cell Res* **1864**: 2200–2209.

120. M. Santos J, Graindorge A, Soldati-Favre D (2012) New insights into parasite rhomboid proteases. *Mol Biochem Parasitol* **182**: 27–36.

121. Esser K, Tursun B, Ingenhoven M, Michaelis G, Pratje E (2002) A Novel Two-step Mechanism for Removal of a Mitochondrial Signal Sequence Involves the mAAA Complex and the Putative Rhomboid Protease Pcp1. *J Mol Biol* **323**: 835–843.

122. McQuibban GA, Saurya S, Freeman M (2003) Mitochondrial membrane remodelling regulated by a conserved rhomboid protease. *Nature* **423**: 537–541.

123. McQuibban GA, Lee JR, Zheng L, Juusola M, Freeman M (2006) Normal mitochondrial dynamics requires rhomboid-7 and affects Drosophila lifespan and neuronal function. *Curr Biol* **16**: 982–989.

124. Sekine S, Kanamaru Y, Koike M, Nishihara A, Okada M, Kinoshita H, Kamiyama M, Maruyama J, Uchiyama Y, Ishihara N, et al. (2012) Rhomboid protease PARL mediates the mitochondrial membrane potential loss-induced cleavage of PGAM5. *J Biol Chem* **287**: 34635–34645.

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

126. Johnson N, Březinová J, Stephens E, Burbridge E, Freeman M, Adrain C, Strisovsky K (2017) Quantitative proteomics screen identifies a substrate repertoire of rhomboid protease RHBDL2 in human cells and implicates it in epithelial homeostasis. *Sci Rep* **7**: 1–13.

127. Lemberg MK (2013) Sampling the membrane: function of rhomboid-family proteins. *Trends Cell Biol* **23**: 210–217.

128. Adrain C, Zettl M, Christova Y, Taylor N, Freeman M (2012) Tumor necrosis factor signaling requires iRhom2 to promote trafficking and activation of TACE. *Science* **335**: 225–228.

129. McIlwain DR, Lang PA, Maretzky T, Hamada K, Ohishi K, Maney SK, Berger T, Murthy A, Duncan G, Xu HC, et al. (2012) iRhom2 regulation of TACE controls TNF-mediated protection against Listeria and responses to LPS. *Science* **335**: 229–232.

130. Blaydon DC, Etheridge SL, Risk JM, Hennies H-C, Gay LJ, Carroll R, Plagnol V, McRonald FE, Stevens HP, Spurr NK, et al. (2012) RHBDF2 Mutations Are Associated with Tylosis, a Familial Esophageal Cancer Syndrome. *Am J Hum Genet* **90**: 340–346.

131. Greenblatt EJ, Olzmann JA, Kopito RR (2011) Derlin-1 is a rhomboid pseudoprotease required for the dislocation of mutant α-1 antitrypsin from the endoplasmic reticulum. *Nat Struct Mol Biol* **18**: 1147–1152.

132. Knop M, Finger A, Braun T, Hellmuth K, Wolf DH (1996) Der1, a novel protein specifically required for endoplasmic reticulum degradation in yeast. *EMBO J* **15**: 753–763.

133. Carvalho P, Goder V, Rapoport TA (2006) Distinct ubiquitin-ligase complexes define

convergent pathways for the degradation of ER proteins. *Cell* **126**: 361–373.

134. Christianson JC, Olzmann JA, Shaler TA, Sowa ME, Bennett EJ, Richter CM, Tyler RE, Greenblatt EJ, Harper JW, Kopito RR (2011) Defining human ERAD networks through an integrative mapping strategy. *Nat Cell Biol* **14**: 93–105.

135. Oda Y, Okada T, Yoshida H, Kaufman RJ, Nagata K, Mori K (2006) Derlin-2 and Derlin-3 are regulated by the mammalian unfolded protein response and are required for ER-associated degradation. *J Cell Biol* **172**: 383–393.

136. Dickey SW, Baker RP, Cho S, Urban S (2013) Proteolysis inside the Membrane Is a Rate-Governed Reaction Not Driven by Substrate Affinity. *Cell* **155**: 1270–1281.

137. Rodriguez F, Rouse SL, Tait CE, Harmer J, De Riso A, Timmel CR, Sansom MSP, Berks BC, Schnell JR (2013) Structural model for the protein-translocating element of the twin-arginine transport system. *Proc Natl Acad Sci* **110**: E1092–E1101.

138. Lewis BA, Engelman DM (1983) Bacteriorhodopsin remains dispersed in fluid phospholipid bilayers over a wide range of bilayer thicknesses. *J Mol Biol* **166**: 203–210.

139. Zoll S, Stanchev S, Began J, Skerle J, Lepšík M, Peclinovská L, Majer P, Strisovsky K (2014) Substrate binding and specificity of rhomboid intramembrane protease revealed by substrate-peptide complex structures. *EMBO J* 1–14.

140. Tichá A, Stanchev S, Škerle J, Began J, Ingr M, Švehlová K, Polovinkin L, Růžička M, Bednárová L, Hadravová R, et al. (2017) Sensitive versatile fluorogenic transmembrane peptide substrates for rhomboid intramembrane proteases. *J Biol Chem* **292**: 2703–2713.

141. Hedstrom L (2002) Serine protease mechanism and specificity. *Chem Rev* **102**: 4501–4523.

142. Walker B, Lynas JF (2001) Strategies for the inhibition of serine proteases. *Cell Mol Life Sci* **58**: 596–624.

143. Bachovchin DA, Koblan LW, Wu W, Liu Y, Li Y, Zhao P, Woznica I, Shu Y, Lai JH, Poplawski SE, et al. (2014) A high-throughput, multiplexed assay for superfamily-wide profiling of enzyme activity. *Nat Chem Biol* **10**: 656–663.

144. King C, Sarabipour S, Byrne P, Leahy DJ, Hristova K (2014) The FRET signatures of noninteracting proteins in membranes: simulations and experiments. *Biophys J* **106**: 1309–1317.

145. Voelker DR (1991) Organelle biogenesis and intracellular lipid transport in eukaryotes. *Microbiol Rev* **55**: 543–560.

146. Goddard AD, Dijkman PM, Adamson RJ, Watts A (2013) Lipid-Dependent GPCR Dimerization. In, *Methods in Cell Biology* pp 341–357. Academic Press Inc.