Publication 1

Article

TRANSPARENT PROCESS OPEN ACCESS

THE EMBO JOURNAL

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.

Keywords intramembrane protease; rhomboid family; rhomboid protease; structure; substrate recognition

Subject Categories Membrane & Intracellular Transport; Post-translational Modifications, Proteolysis & Proteomics; Structural Biology

DOI 10.15252/embj.201489367 | Received 25 June 2014 | Revised 8 August 2014 | Accepted 19 August 2014 | Published online 12 September 2014

The EMBO Journal (2014) 33: 2408–2421

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-$COCH_2Cl$ (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

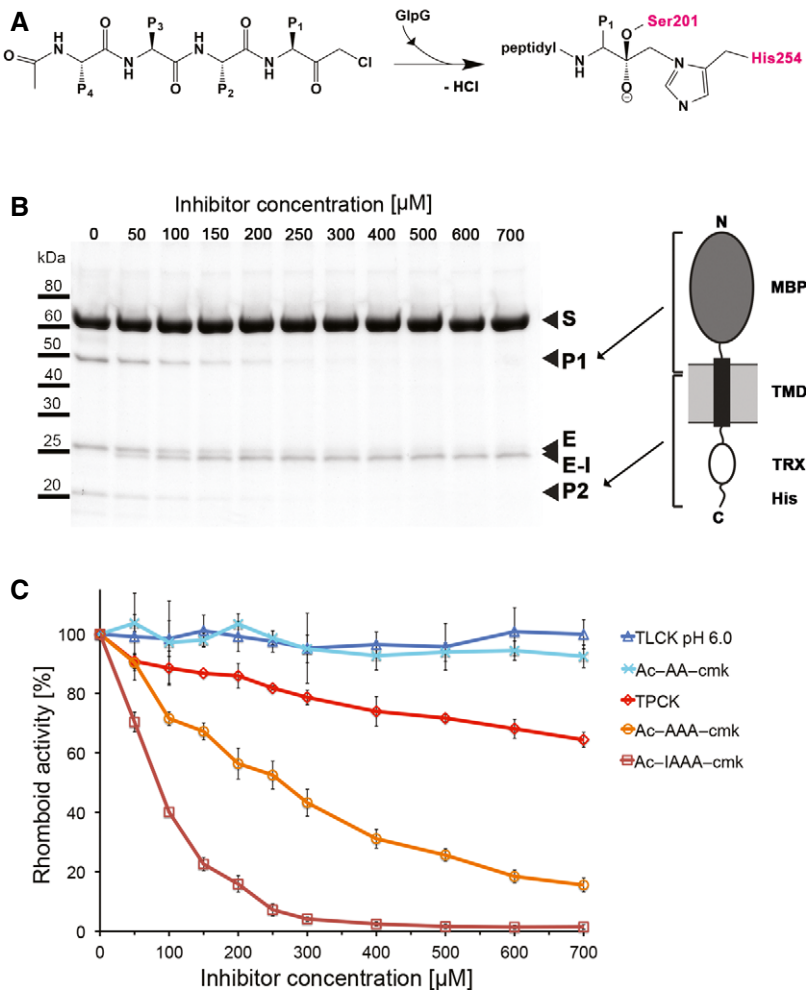


Figure 1. Tetrapeptidyl-chloromethylketones are mechanism-based inhibitors of rhomboid proteases.

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).

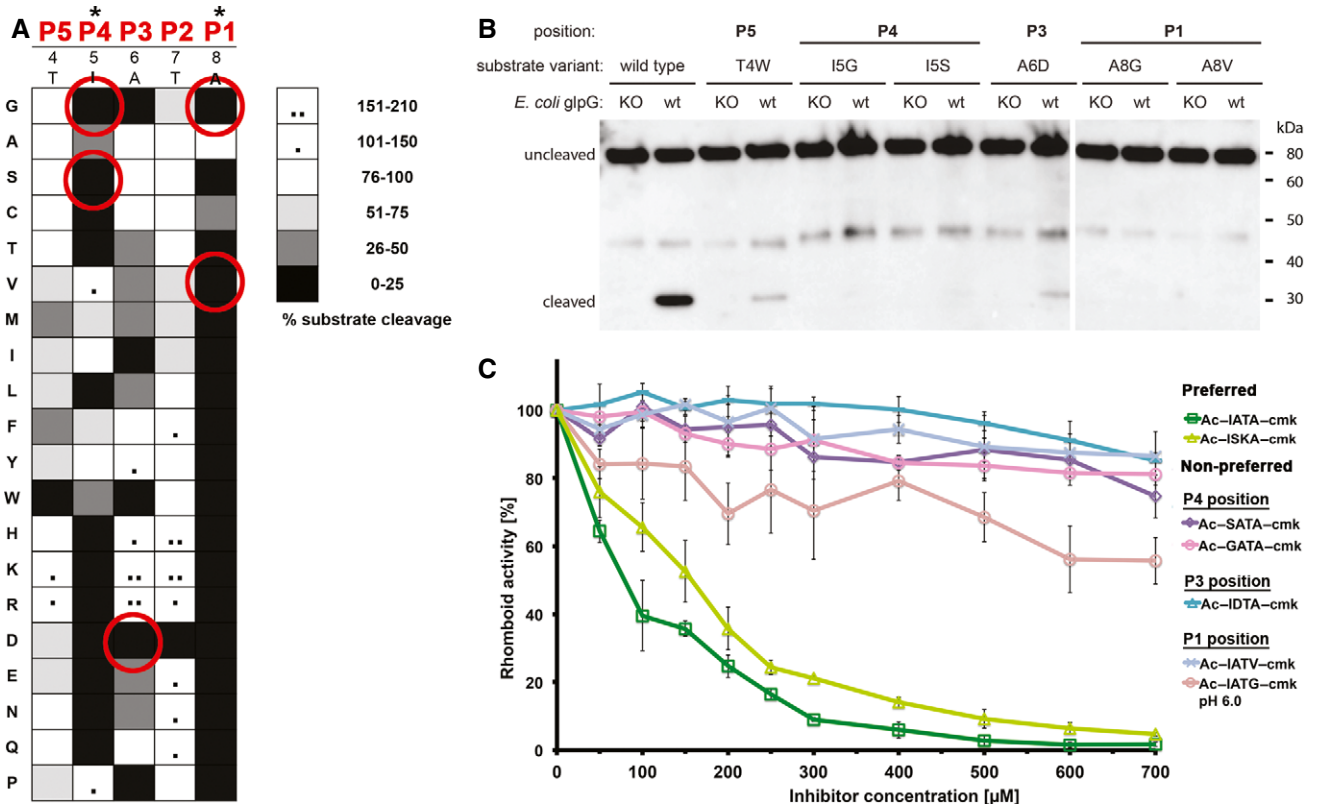


Figure 2. Tetrapeptidyl-chloromethylketones bind GlpG in a substrate-like manner.

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 showing 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

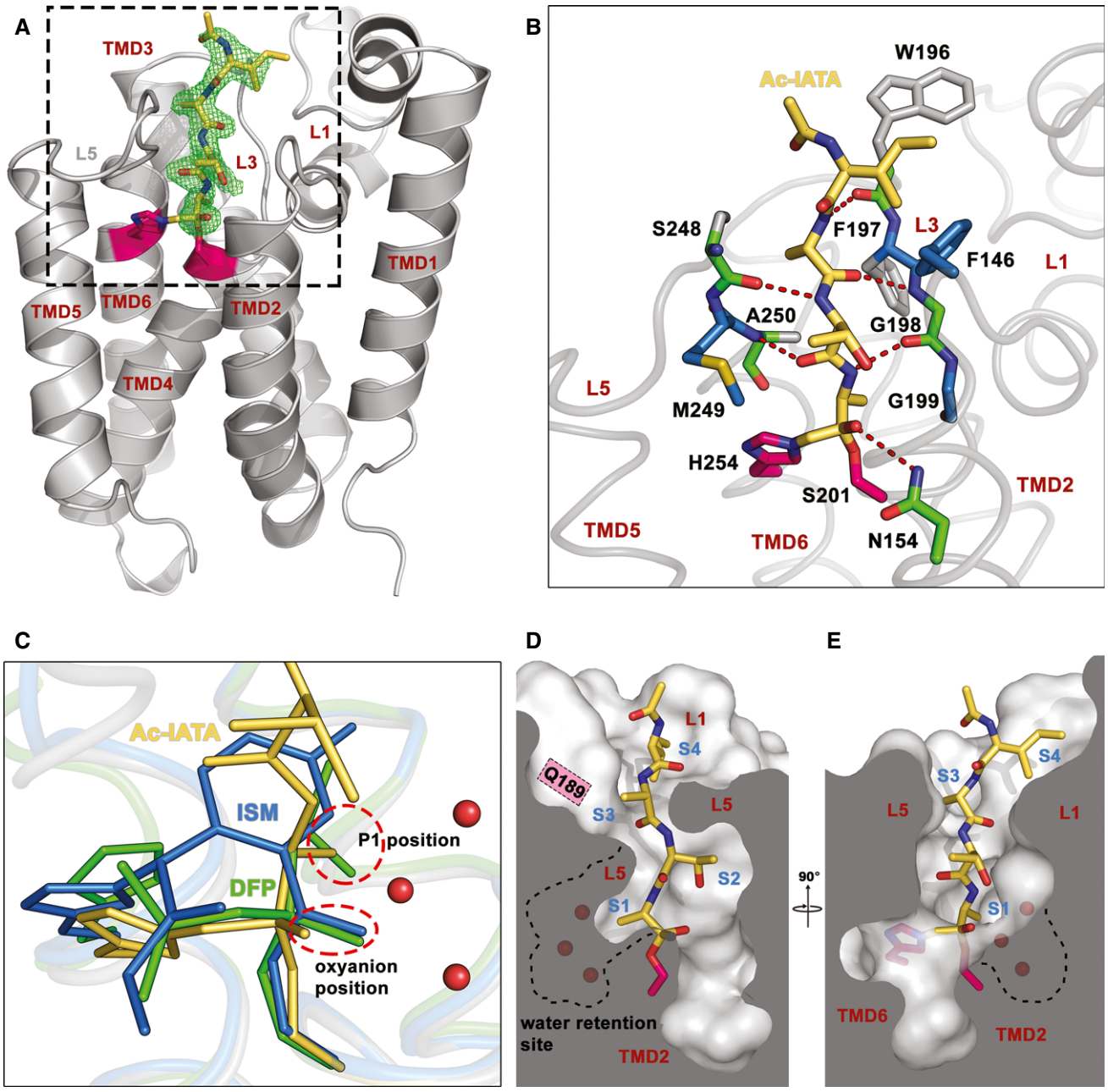


Figure 3. Crystal structure of GlpG complexed to Ac-IATA-cmk reveals the mode of substrate binding to GlpG.

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 the 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

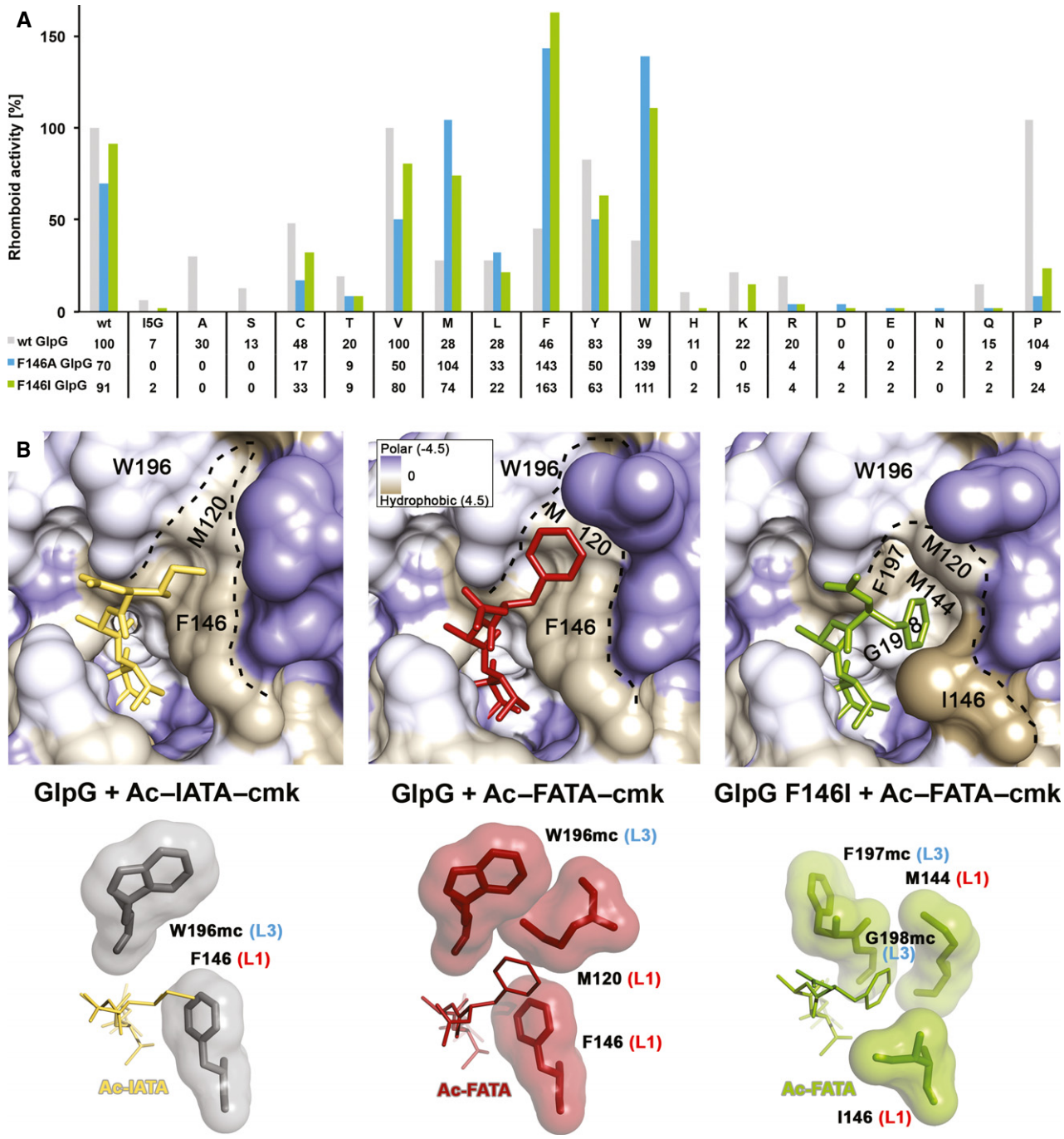


Figure 4. A patch of hydrophobic residues from the L1 loop forms the S4 subsite of GlpG.

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.

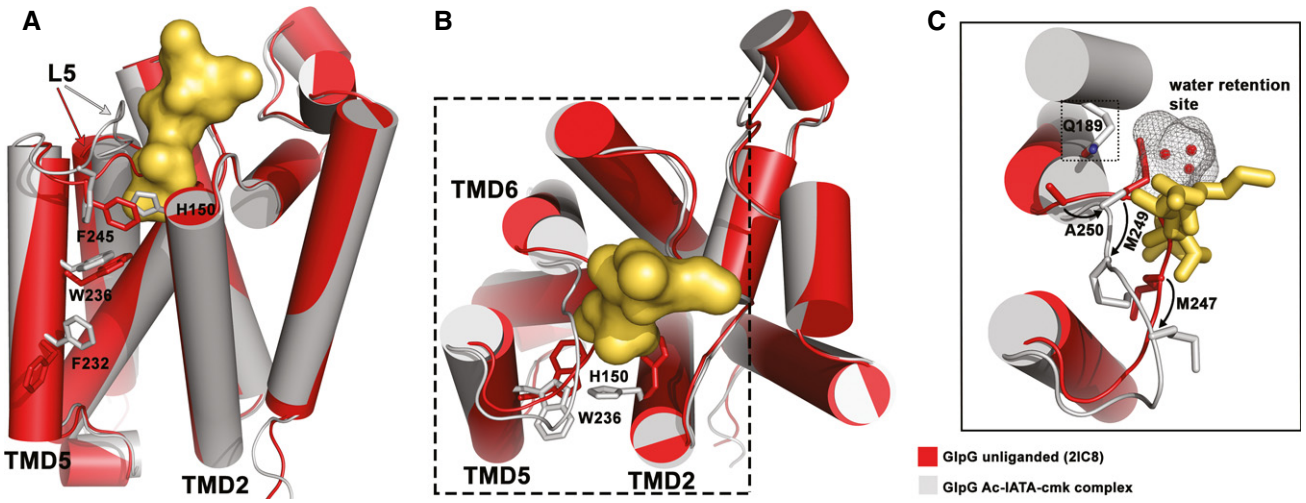


Figure 5. Binding of Ac-IATA-cmk to GlpG induces displacement of the L5 loop and side chain rotamer changes in TMD2 and 5.

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 diisopropylfuorophosphonate (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

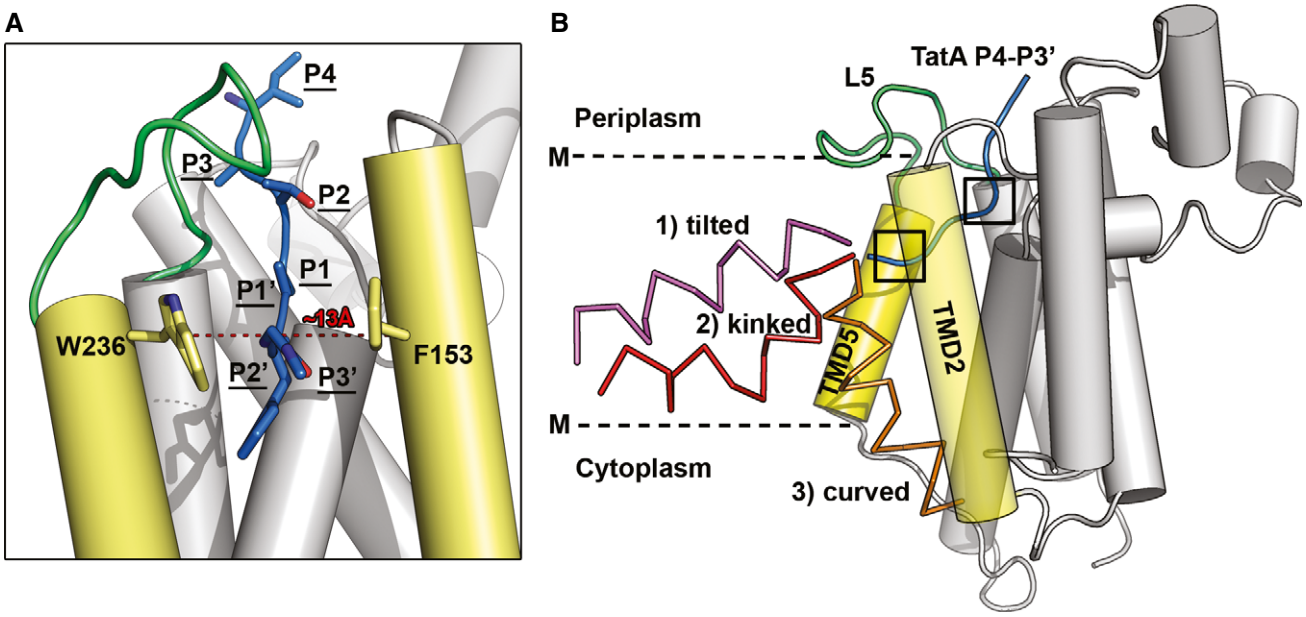


Figure 6. Molecular dynamics-based model of the Michaelis complex and possible interaction modes of substrate transmembrane domain.

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

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

Acknowledgements

We thank Jana Horáková and Martin Hubálek for mass spectrometry, Radko Souček for amino acid analysis and LC-MS, Zdeněk Voburka for N-terminal sequencing, Tobias Kloepper for help with rhomboid phylogeny and sequence alignment, Kateřina Švehlová for technical assistance, Vinothkumar Kutti Ragunath and Matthew Freeman for comments on the manuscript, and the beamline staff at the BESSY in Berlin and the ESRF in Grenoble for beamtime and support. Research in KS's lab is supported by Czech Science Foundation (project no. P305/11/1886), Ministry of Education, Youth and Sports of the Czech Republic (projects no. LK11206 and LO1302), EMBO Installation Grant (project no. 2329), Marie Curie Career Integration Grant (project no. 304154) and the National Subvention for Development of Research Organisations (RVO: 61388963) to the Institute of Organic Chemistry and Biochemistry (IOCB). SZ was supported by a post-doctoral fellowship from IOCB, JS was supported by a PhD grant project GA UK no. 232313 from Charles University in Prague, and ML by a Czech Science Foundation grant (project no. P208/12/G016).

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

Publication 2

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 292, NO. 7, pp. 2703–2713, February 17, 2017

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

Received for publication, October 14, 2016, and in revised form, January 5, 2017 Published, JBC Papers in Press, January 9, 2017, DOI 10.1074/jbc.M116.762849

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.

* This work was supported by EMBO Installation Grant 2329, Ministry of Education, Youth and Sports of the Czech Republic Projects LK11206 and LO1302, Marie Curie Career Integration Grant Project 304154 (to K. S.), and National Subvention for Development of Research Organizations RVO: 61388963 to the Institute of Organic Chemistry and Biochemistry. The authors declare that they have no conflicts of interest with the contents of this article.

[S] This article contains supplemental information.

[1] Both authors contributed equally to the results of this work.

[2] Supported by Ph.D. project number 232313 from the Grant Agency of Charles University (GA UK) in Prague.

[3] Supported by Ph.D. project number 170214 from the Grant Agency of Charles University (GA UK) in Prague.

[4] Supported by the Czech Science Foundation Grant P208-12-G016 (Center of Excellence).

[5] Present address: Institut de Biologie Structurale, 71 avenue des Martyrs, Grenoble, 38044, France.

[6] Supported by the Czech Science Foundation Grant 15-01948S.

[7] Recipient of the Purkyne Fellowship of the Academy of Sciences of the Czech Republic. To whom correspondence should be addressed: Institute of Organic Chemistry and Biochemistry, Flemingovo n. 2, Prague, 166 10, Czech Republic. Tel.: 420-220-183-468; E-mail: kvido.strisovsky@uochb.cas.cz.

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 readouts, 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-

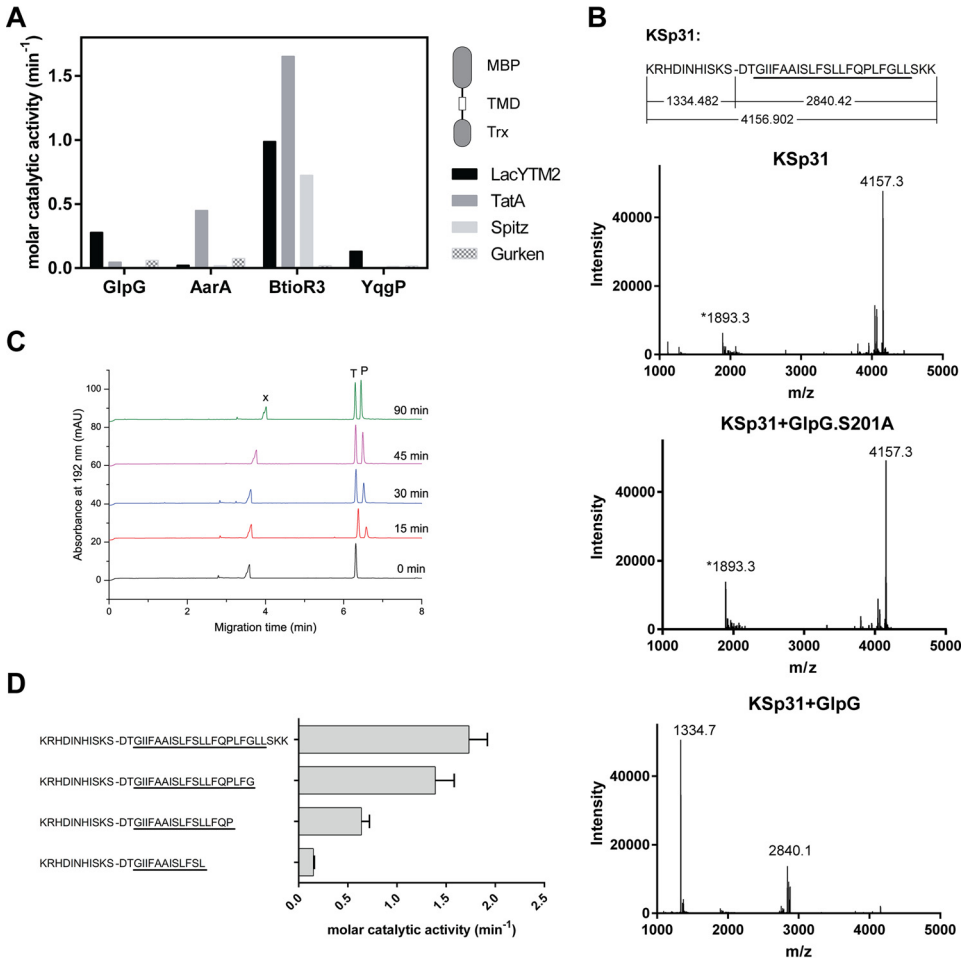


FIGURE 1. **Identification of a widely accepted transmembrane substrate for rhomboid proteases.** *A,* comparison of cleavage efficiency of model substrates LacYTM2, Gurken, TatA, and Spitz by bacterial rhomboid proteases GlpG (*E. coli*), AarA (*P. stuartii*), YqgP (*B. subtilis*), and BtioR3 (*B. thetaiotaomicron*) *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.

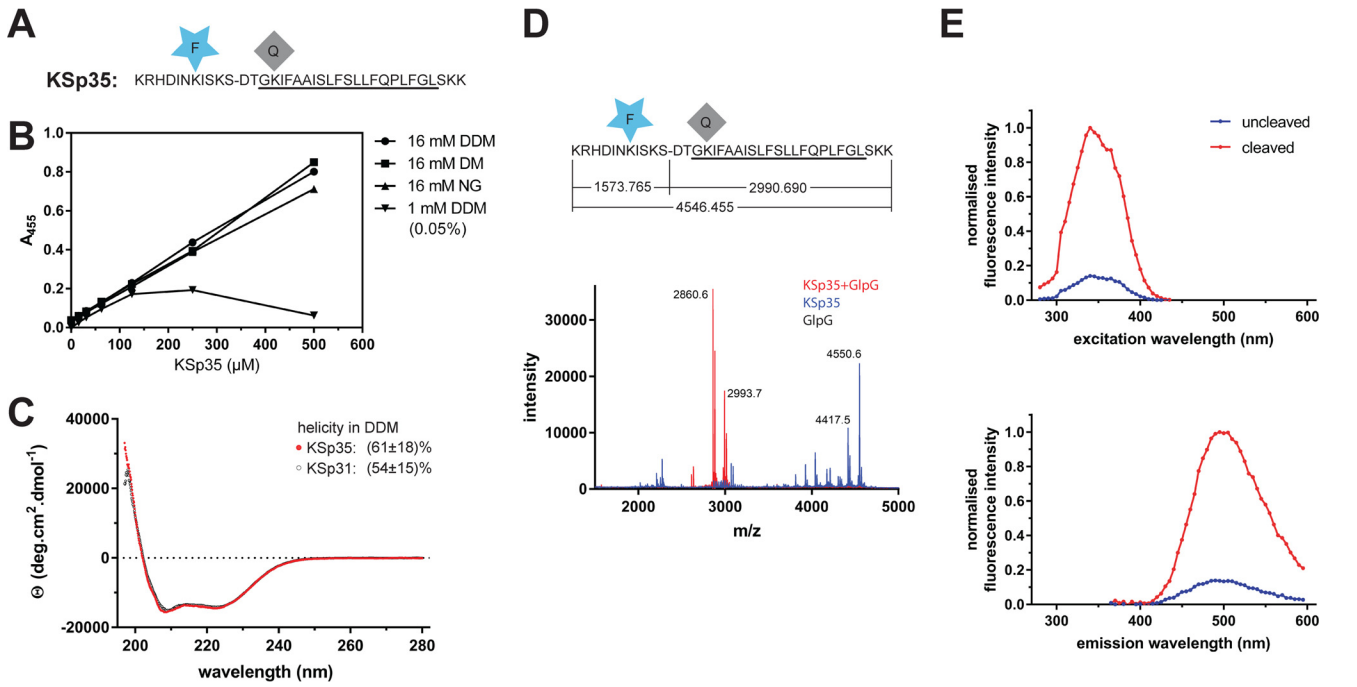


FIGURE 2. **Fluorogenic transmembrane peptide substrate based on LacYTM2.** *A,* fluorogenic variant of the LacYTM2 transmembrane helix-derived peptide (KSp31) with the P5 and P4′ positions replaced by Glu-EDANS and Lys-DABCYL, respectively, yielding fluorogenic substrate KSp35. *B,* solubility of KSp35 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

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. 3*A*). 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. 3*B*). 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. 3*C*), without an obvious impact on the secondary structure content of GlpG (Fig. 3*D*), 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 $n(M)$.

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

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 $n(E)$.

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

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 $[\text{DDM}]^{-1}$). 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. 3*C* ($\log a^n = n \times \log a$). The logarithmic plot (Fig. 3*C*, *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 $\sim[\text{DDM}]^{-1}$), 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 $\sim[\text{DDM}]^{-0.4}$). 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

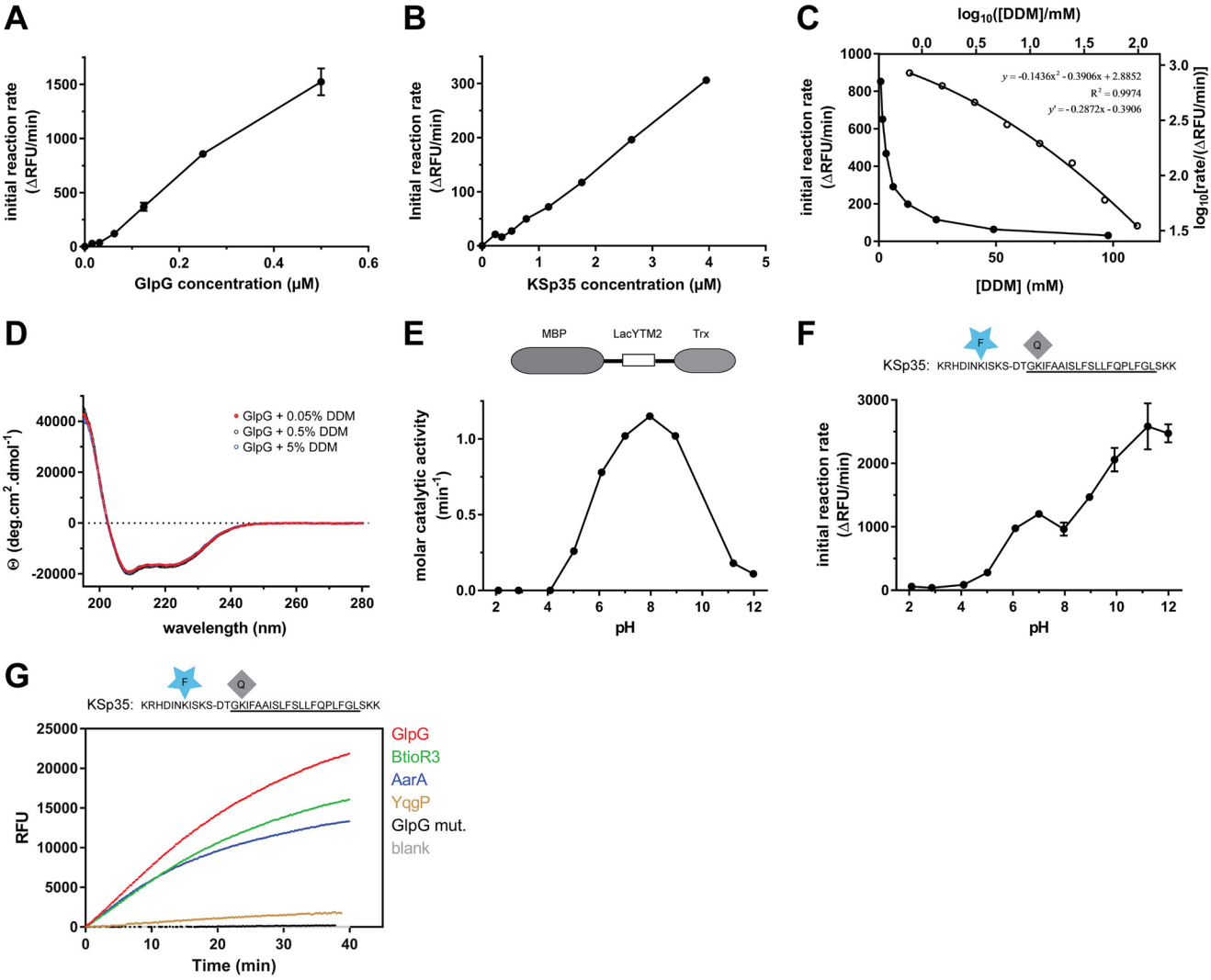


FIGURE 3. **Kinetic characterization of fluorogenic transmembrane peptide substrate KSp35 in the detergent micelle system.** *A*, dependence of the initial reaction rate on enzyme concentration. The fluorogenic substrate KSp35 (10 μM) was incubated with varying concentrations of GlpG in a reaction buffer 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 × 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_3PO_4, 40 mM CH_3COOH, and 40 mM H_3BO_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_3PO_4, 40 mM CH_3COOH, and 40 mM H_3BO_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).

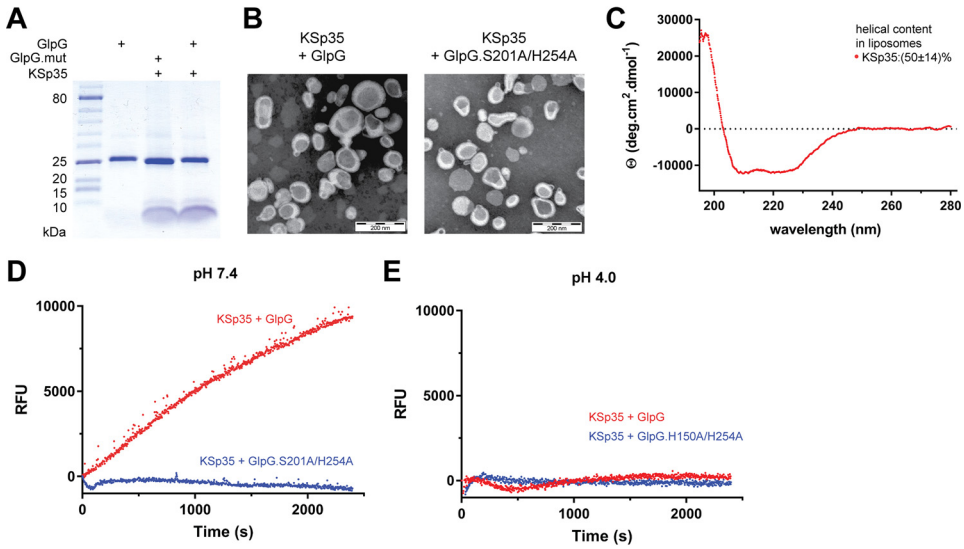


FIGURE 4. **The use of the transmembrane peptide substrate in liposomes.** *A*, KSp35 was reconstituted into liposomes (LUVs) formed from *E. coli* polar lipid extract in the presence of GlpG or its inactive mutant S201A at pH 4.0. The resulting large unilamellar vesicles were analyzed by SDS-PAGE. *B,* 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 substrates.

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*). Negative 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}

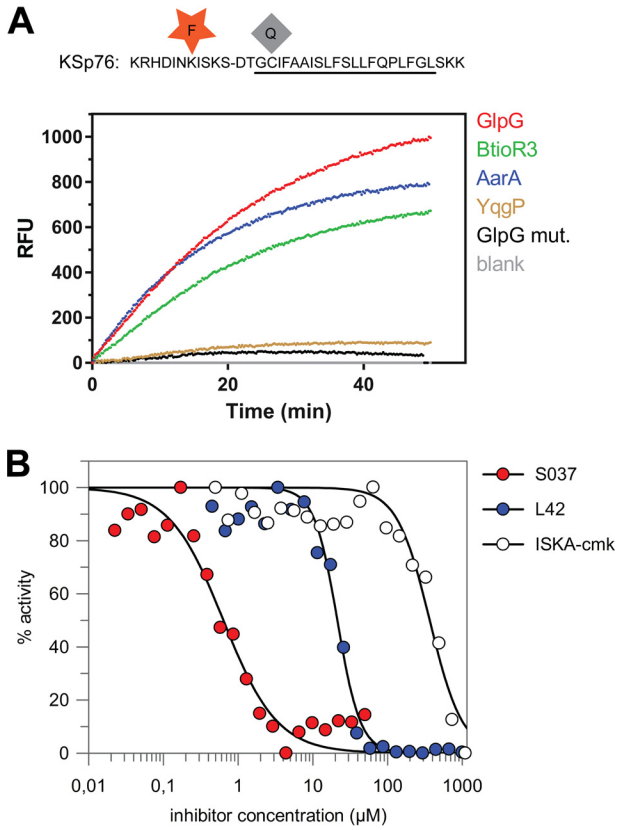


FIGURE 5. **Red-shifted variant of the LacYTM2-based fluorogenic substrate.** *A,* modification of Lys in the P5 position of KSp31 by the red-shifted TAMRA fluorophore and P4′ Cys by a dark quencher QXL610 yields highly fluorogenic substrate KSp76 that is efficiently cleaved by rhomboid proteases GlpG, AarA, YqgP, and BtioR3 at identical concentrations to those used in Fig. 3*G*. 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. 6*C*). 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 ± 38, 12.4 ± 1.6, and 0.64 ± 0.08 μM, respectively (Fig. 5*B*), 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 modifying 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. 6*A*). Combining all five mutations yielded a mutant substrate (RVRHA) that was cleaved 64-fold better than the wild type substrate (Fig. 6*A*), 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. 6*B*). 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. 6*B*).

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. 6*C*). The impact of the modifications of the P5 to P1 region on selectivity against other bacterial rhomboid proteases is particularly striking (Fig. 6*D*), 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. 6*D*) 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-

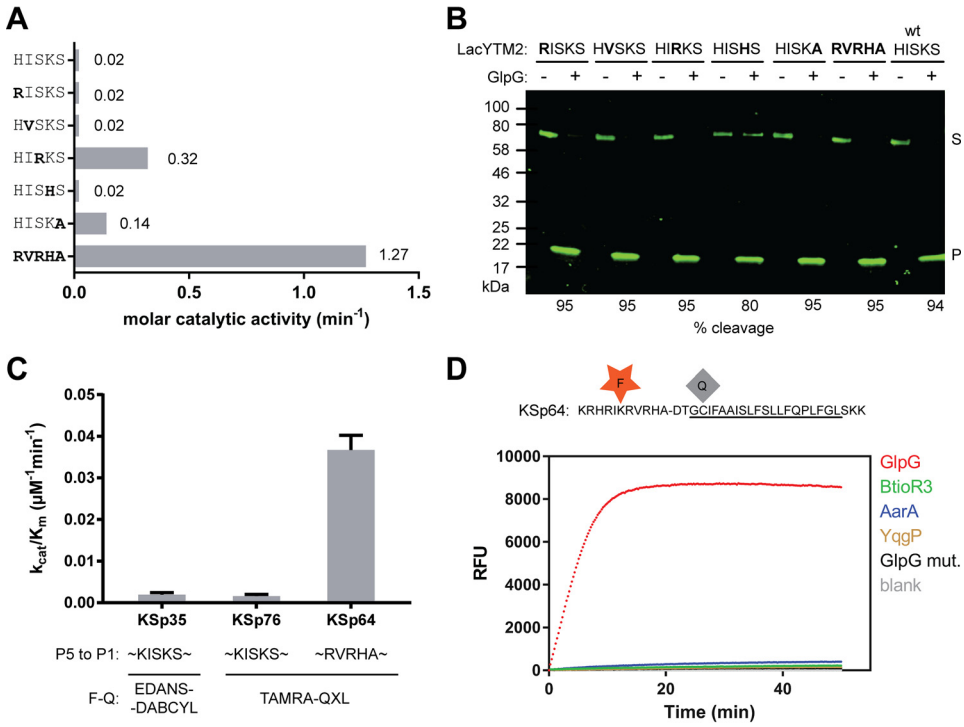


FIGURE 6. **The effect of non-prime side substitutions on the catalytic parameters and selectivity of rhomboid substrates.** *A,* preferred amino acids in the P5 to P1 positions of the LacYTM2 transmembrane substrate improve its cleavage by GlpG. The LacYTM2 embedded in the MBP-thioredoxin chimera (18) 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

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 (InstantBlue, 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-

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).

Author Contributions—K. S. conceived and coordinated the study, designed experiments, and wrote the paper with the input of A. T., M. I., S. S., J. B., J. S., M. R., and V. K. S. S. and P. M. designed and S. S. performed all chemical syntheses. M. R. and V. K. designed and performed all capillary eletrophoresis analyses. M. I. analyzed kinetics data, R. H. performed electron microscopy, and L. B. designed and performed all circular dichroism measurements. J. B. performed and evaluated experiments shown in Fig. 6, *A* and *B*. JŠ performed and evaluated experiments shown in Fig. 4, *A*, *D*, and *E*. K. Š designed, performed, and analyzed data shown in Fig. 1*A*. P. R. and E. P. contributed to experiments shown in Fig. 6, *A*, *B*, and *D*. L. P. established the fluorogenic assay and performed and evaluated experiments shown in Fig. 1D. J. Březinová contributed to all mass spectrometry experiments, and A. T. designed, performed, and evaluated all other kinetics and inhibition measurements that are the basis of this manuscript.

Acknowledgments—We thank Steven Verhelst (University of Leuven, Belgium) for his kind gift of isocoumarin S037, Matthew Freeman (Oxford University, United Kingdom) for his kind gift of the inhibitor L42, Zdeněk Voburka and Radko Souček for amino acid analyses, Mirka Blechová for peptide synthesis and purification, and Blanka Collis for critical reading of the manuscript.

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

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

Sensitive Versatile Fluorogenic Transmembrane Peptide Substrates for Rhomboid Intramembrane Proteases

Anezka Tichá, Stancho Stanchev, Jan Skerle, Jakub Began, Marek Ingr, Katerina Svehlová, Lucie Polovinkin, Martin Ruzicka, Lucie Bednárová, Romana Hadravová, Edita Poláchová, Petra Rampírová, Jana Brezinová, Václav Kasicka, Pavel Majer and Kvido Strisovsky

J. Biol. Chem. 2017, 292:2703-2713.
doi: 10.1074/jbc.M116.762849 originally published online January 9, 2017

SUPPLEMENTAL INFORMATION

Sensitive versatile fluorogenic transmembrane peptide substrates for rhomboid intramembrane proteases

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

[1]The Institute of Organic Chemistry and Biochemistry of the 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]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]Tomas Bata University in Zlín, Faculty of Technology, Department of Physics and Materials Engineering, nám. T.G. Masaryka 5555, 76001, Zlín, Czech Republic
[$]Present address: Institut de Biologie Structurale, 71 avenue des Martyrs, Grenoble, 38044, France
*corresponding author: kvido.strisovsky@uochb.cas.cz, Institute of Organic Chemistry and Biochemistry, Flemingovo n. 2, Prague, 166 10, Czech Republic, tel.: +420 220 183 468
[#]equal contribution

Running title: *Fluorogenic substrates for rhomboid proteases*

ABBREVIATIONS

ACN	Acetonitrile
AcOH	Acetic acid
DCM	Dichloromethane
DIC	Diisopropylcarbdodiimide
DIEA	Diisopropylethylamine
DMF	Dimethylformamide
DTT	Dithiothreitol
FA	Formic acid
HATU	2-(1H-(7-azabenzotriazol-1-yl))-1,1,3,3-tetramethyluronium hexafluorophosphate
HBTU	2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HOBt	1-hydroxybenzotriazol
iPrOH	Isopropylalcohol
TCEP	*Tris*(2-carboxyethyl)phosphine
TFA	Trifluoroacetic acid
TFE	Trifluoroethanol

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Chemical Synthesis

EDANS/Dabcyl labelled peptides

The major part of peptide synthesis was performed on Tenta Gel S Rinkamide resin (substitution 0.24 mmol/g) on PS3 peptide synthesizer (Protein technologies, USA), using Fmoc-chemistry protocol at a scale of 0.1 mmol. The coupling of Fmoc-Glu(EDANS)-OH and Fmoc-Lys(Dabcyl)-OH was performed manually at the same scale in fourfold excess, using HBTU (4 eqv) / HOBt (4 eqv) / DIEA (6 eqv) in 3-4 ml DMF. The peptide was cleaved from the resin by the mixture of trifluoroacetic acid / triisopropylsilane / water = 95:2.5:2.5 for 3 hours. The mixture was concentrated under nitrogen and the crude peptide was precipitated with cold ether. The precipitate was washed by ether, dried and purified by RP HPLC.

TAMRA/QXL labelled peptides

Synthesis of TAMRA

Trimellitic anhydride (**1**) (4 g, 21 mmol) and 3-dimethylaminophenol **(2)** (5.8 g, 42 mmol) were refluxed in toluene (220 mL) for 12 hours as described (1). Upon cooling to room temperature, the resulting precipitate was collected by filtration, and further purified by flash chromatography (silica gel, dichloromethane/methanol/acetic acid = 8:1.9:0.1) to afford pure mixture of 5(6)-carboxytetramethylrhodamine (TAMRA) as a dark purple solid.

Coupling of TAMRA to ε-amino group of the lysine at 7th position.
Fmoc-Lys(Mmt)-OH was coupled overnight manually to the rest of the peptide on the resin in 4-fold excess, using DIC (7 eqv) / HOBt (5 eqv) method in 3 mL DMF. The peptide-containing resin was washed with DMF (6 × 2 mL), isopropanol (6 × 2 mL), DMF (6 mL) and DCM (10 mL). Mmt as a very acid labile orthogonal protecting group was removed by treatment of the resin with a mixture of DCM / TFE / AcOH (7:2:1) in 4-5 cycles by 15 min each and washing with pure DCM between cycles until no red-orange methoxytrityl cations in the eluate were detected (2). The resin was washed with DCM (10 mL) and dried at vacuum. The ε-deprotected lysine was proved by the ninhydrin test. After soaking with DMF (3-4 mL) for 10 min, the resin was ready for coupling of TAMRA (4 eqv), using HATU (4 eqv) / HOBt (4 eqv) / DIEA (6 eqv) method in 3 – 4 mL DMF for 4-6 h. Then the resin was washed with DMF (6 × 2 mL), isopropanol (6 × 2 mL), DCM (10 mL) and was dried in the vacuum. Peptide synthesis was continued in the automated peptide synthesizer, and the peptides were cleaved off the resin by 5-6 mL of the TFA/TIS/ethanedithiol/thioanisole/water mixture (90:2:2.5:3:2.5) for 3.5 h. The peptides were processed and purified as described for the EDANS-Dabcyl labelled ones above.

Labelling with QXL610 vinylsulfone
The purified deprotected peptide was used for labelling of the cysteine at 15 position with a dark quencher QXL610 with vinylsulfone (VS) reactive group as described (3). The 37-mer TAMRA-peptide (13 mg, 2.9 μmol) was dissolved in 14 mL Tris-buffer (pH 7.8), containing 0.24 M TCEP (48 μL). The QXL610VS (3 mg, 5.9 μmol), dissolved in 360 μL DMSO was added and the mixture was stirred for 2 – 3 h at room temperature. When the starting peptide was depleted (80 – 90 % conversion as judged by HPLC analysis), the reaction was quenched by addition of DTT (22 mg, 0.14 mmol) to a final concentration 10 mM and stirred for 20 min. The solution was frozen and lyophilized and the double labelled TAMRA/QXL peptide was isolated by RP HPLC on a C_{18} Watrex Reprosil semipreprative column 250 × 10 (gradient 30 – 80 % B) at flow 3 mL/min.

HPLC analysis and purification
Chromatographic conditions: Module Jasco PU 1580 Series (Jasco, Japan) with a preparative C_{18} column (250 × 20 mm, 10 μm particle size , Watrex International, Inc, San Francisco, California, USA), using water + 0.1 % trifluoroacetic acid (A) / acetonitrile (B) as mobile phases, at 10 mL/min flow rate. Elution gradient: from 30 % B to 80 %B over 60 minutes at room temperature. The elution was monitored by absorption at 210 nm, using UV – VIS 1575 detector. The fractions corresponding to the major peak (desired peptide) were combined, frozen and lyophilized, giving the pure peptide. Analytical runs were done on the same module using Watrex C_{18}, 250 × 4.6 mm, 5 μm particle size column (Watrex International, Inc, San Francisco, California, USA) at 1 mL/min flow rate using various solvent gradients ranging from solvent A to solvent B as defined above over 30 minutes at room temperature.

LC-MS
Agilent Technologies Liquid Chromatograph coupled with TOF 6230 ESI-MS detector. Gradient: 2 %B to 100 %B over 10 min on a 1.7 μm particle size C_{18}, 100 × 2.1 mm RP HPLC column (Waters) at a flow rate of 0.3 mL/min, where A = water + 0.1%FA and B = ACN + 0.1%FA.

NMR spectroscopy
NMR spectra were acquired on a Bruker AV-400 MHz at room temperature.

SUPPLEMENTAL RESULTS

Compound characterisation data

5(6)-Carboxytetramethylrhodamine; 5(6)-TAMRA

Yield 1.3 g (14 %)

^{1}H NMR (401 MHz, DMSO-d6) δ (ppm) = 8.69 (s, 1H, aromatic), 8.40 – 8.28 (m, 1H, aromatic), 7.59 (d, J = 7.9 Hz, 1H, aromatic), 7.19 – 6.85 (m, 6H, aromatic), 3.24 (s, 12H, CH3).

ESI-MS: Monoisotopic mass 430.15; found $[M+H]^+$ 431.2, $[M+Na]^+$ 453.2.

KSp31 peptide

Sequence: KRHDINHISKSDTGIIFAAISLFSLLFQPLFGLLSKK

Purified by RP HPLC using gradient of 50 – 100 % B; (A = 1 % CH_3CN, 1 % iPrOH, 98 % water + 0.1 % TFA; B = 35 % CH_3CN, 60 % iPrOH, 5 % water + 0.1 % TFA). The product elutes at 27 min.

Yield: 20.3 mg

LC-MS: One peak at 9.58 min, m/z 4153.36 calculated for $C_{195}H_{313}N_{51}O_{49}$; found 4153.3: $[M+3H]^{3+}$1385.4, $[M+4H]^{4+}$ 1039.34, $[M+5H]^{5+}$ 831.66.

KSp35 peptide

Sequence: KRHDINE(Edans)ISKSDTGK(Dabcyl)IFAAISLFSLLFQPLFGLSKK

Yield: 10.4 mg

LC-MS: One peak at 9.39 min, m/z 4546.44 calculated for $C_{215}H_{328}N_{54}O_{53}S_1$; found 4546.45: $[M+4H]^{4+}$ 1137.62, $[M+5H]^{5+}$ 910.30, $[M+6H]^{6+}$ 758.75, $[M+7H]^{7+}$.

KSp64 peptide

Sequence: KRHRIK(Tamra)RVRHADTGC(QXL610)IFAAISLFSLLFQPLFGLSKK

Yield: 2.5 mg

LC-MS: One peak at 5.34 min, m/z calculated for $C_{223}H_{346}N_{64}O_{48}S$ plus QXL610 fragment (formula not released by the manufacturer) as 5204.6; found 5204.7: $[M+4H]^{4+}$ 1302.18, $[M+5H]^{5+}$ 1041.94, $[M+6H]^{6+}$ 868.45.

KSp76 peptide

Sequence: KRHDINK(Tamra)ISKSDTGC(QXL610)IFAAISLFSLLFQPLFGLSKK

Yield: 4.2 mg;

LC-MS: One peak at 9.45 min, m/z calculated for $C_{207}H_{321}N_{51}O_{52}S$ plus QXL610 fragment as 4916.8; found 4917.28: $[M+4H]^{4+}$ 1230.33, $[M+5H]^{5+}$ 984.47, $[M+6H]^{6+}$ 820.55.

SUPPLEMENTAL REFERENCES

1. Yeo, D. S. Y., Srinivasan, R., Uttamchandani, M., Chen, G. Y. J., Zhu, Q., and Yao, S. Q. (2003) Cell-permeable small molecule probes for site-specific labeling of proteins. *Chem. Commun.*, 2870-2871
2. Matysiak, S., Boldicke, T., Tegge, W., and Frank, R. (1998) Evaluation of monomethoxytrityl and dimethoxytrityl as orthogonal amino protecting groups in fmoc solid phase peptide synthesis. *Tetrahedron Lett.* **39**, 1733-1734
3. Bayburt, T. H., Vishnivetskiy, S. A., McLean, M. A., Morizumi, T., Huang, C. C., Tesmer, J. J. G., Ernst, O. P., Sligar, S. G., and Gurevich, V. V. (2011) Monomeric Rhodopsin Is Sufficient for Normal Rhodopsin Kinase (GRK1) Phosphorylation and Arrestin-1 Binding. *J. Biol. Chem.* **286**, 1420-1428

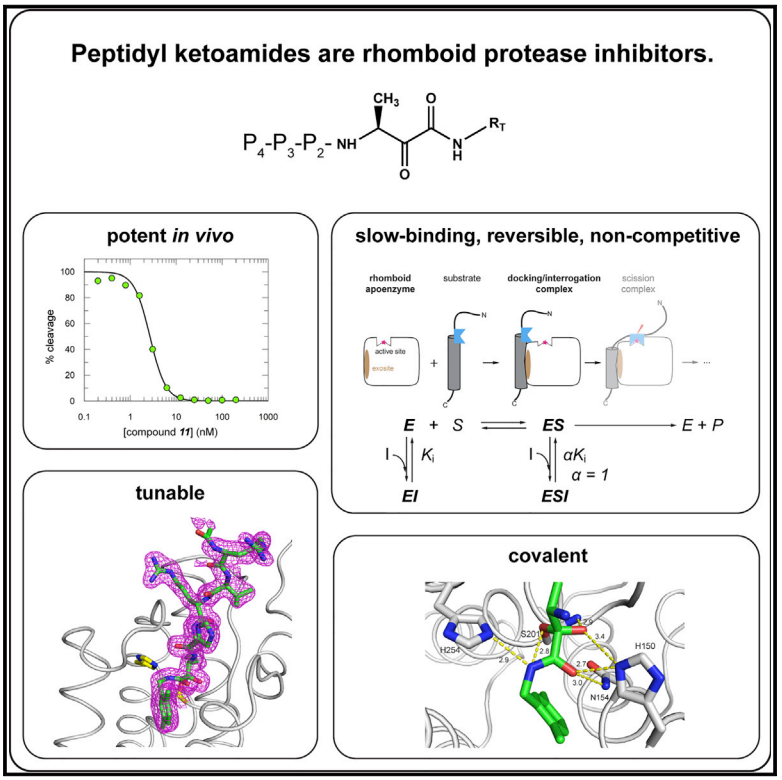
Publication 3

Article

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

Tichá et al., 2017, Cell Chemical Biology *24*, 1523–1536
December 21, 2017
https://doi.org/10.1016/j.chembiol.2017.09.007

Cell Chemical Biology
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 dynamics, 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}] ~5–10 μM) (Pierrat et al., 2011), and β-lactones are not very potent (apparent IC_{50} of ~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

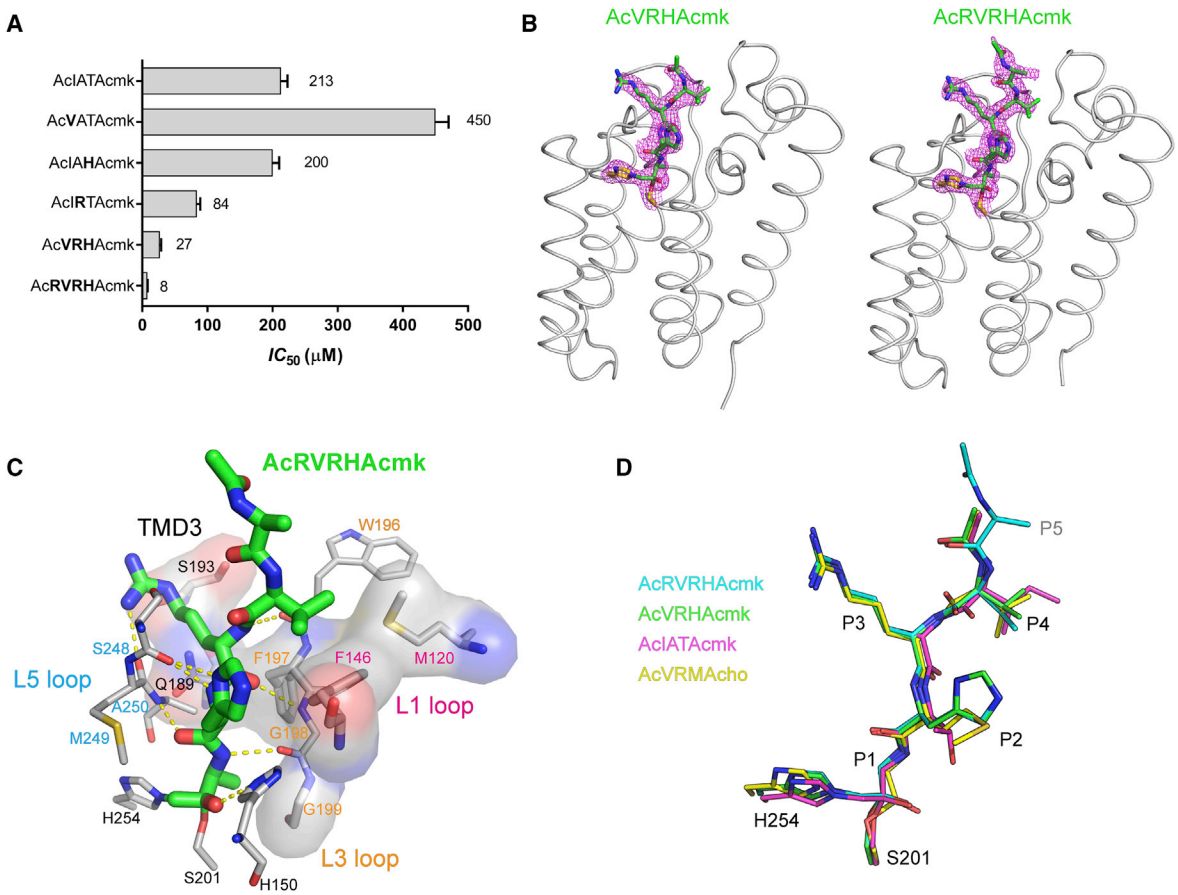


Figure 1. The Potency of Substrate-Derived Inhibitors Can be Improved by Modifying the Amino acid Sequence of the P5 to P1 Region

(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

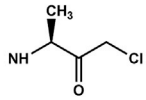
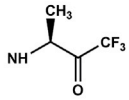
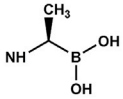
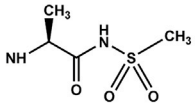
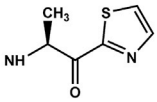
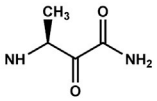
warhead	Ala-X	IC_{50}
chloromethylketone		8 μM
trifluoromethylketone		> 1 mM
boronate		8 μM
acylsulfonamide		> 1 mM
thiazolylketone		> 1 mM
α-ketoamide		203 μM

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, 2013a; 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 stable 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 ($\alpha = 1$) (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.

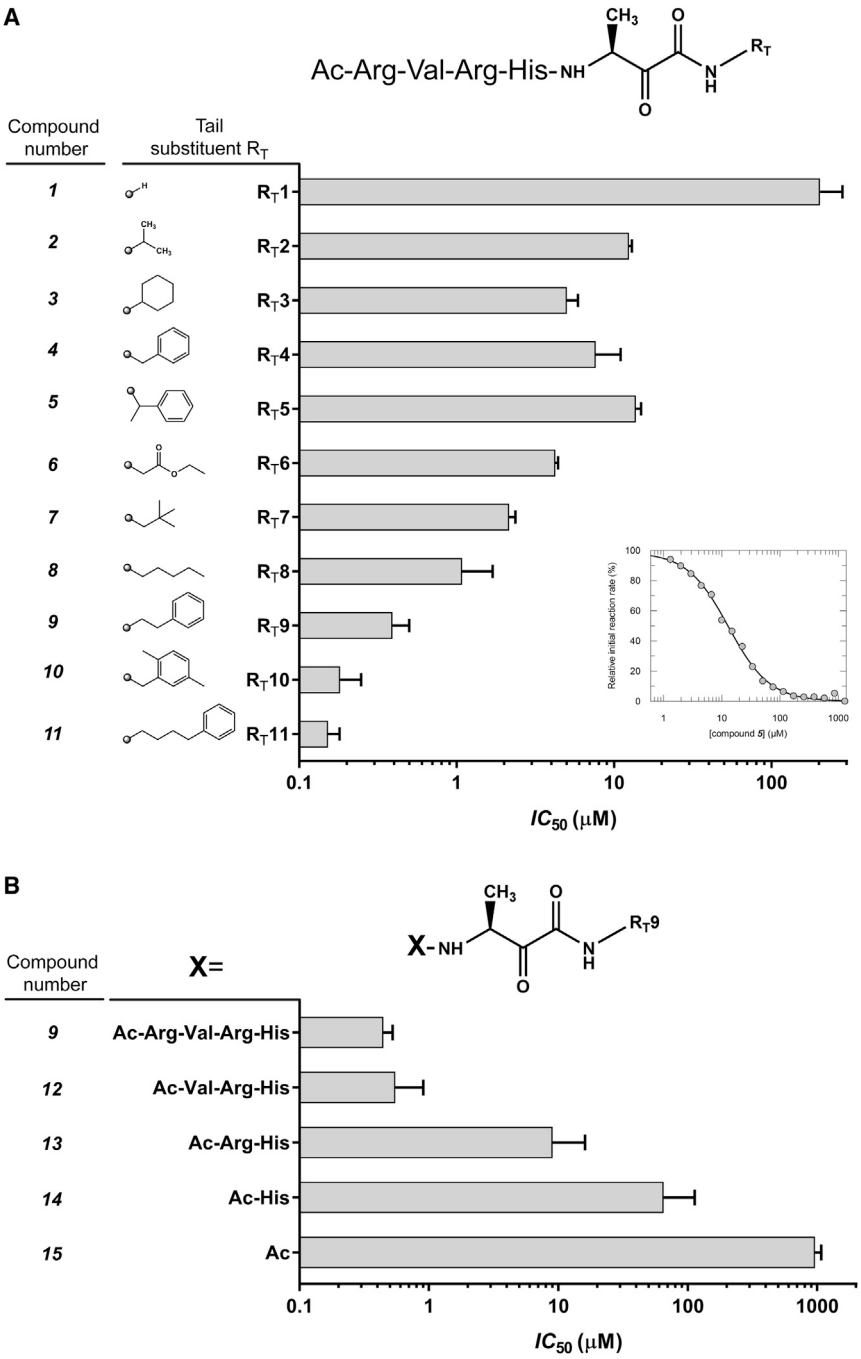


Figure 3. Modification of the Prime-Side Substituent at the Amide Group of Peptidyl Ketoamides Enhances Their Potency by Orders of Magnitude

(A) A screen of the effect of the tail substituent RT on the inhibitory properties of ketoamide inhibitors of GlpG based on the parent compound Ac-RVRHA-$CONH_2$. The apparent IC_{50} values of all compounds were measured in 0.05% (w/v) DDM and 10 μM KSp35 (Ticha et al., 2017) with 1 hr preincubation. The IC_{50} values of the most effective compounds **9**, **10**, and **11** are three orders of magnitude lower than that of the parent compound **1**. The reported values are best-fit means with SD representative of 2–3 measurements. The inset shows a typical inhibition curve.

(B) The significance of the peptidyl part in compound **9**. The peptidyl part of **9** was progressively truncated from the N terminus, and the apparent IC_{50} values of all compounds were measured in 0.05% (w/v) DDM and 10 μM KSp35 (Ticha et al., 2017) with 1 hr preincubation. The reported values are best-fit means with SD representative of 2–3 measurements.

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-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

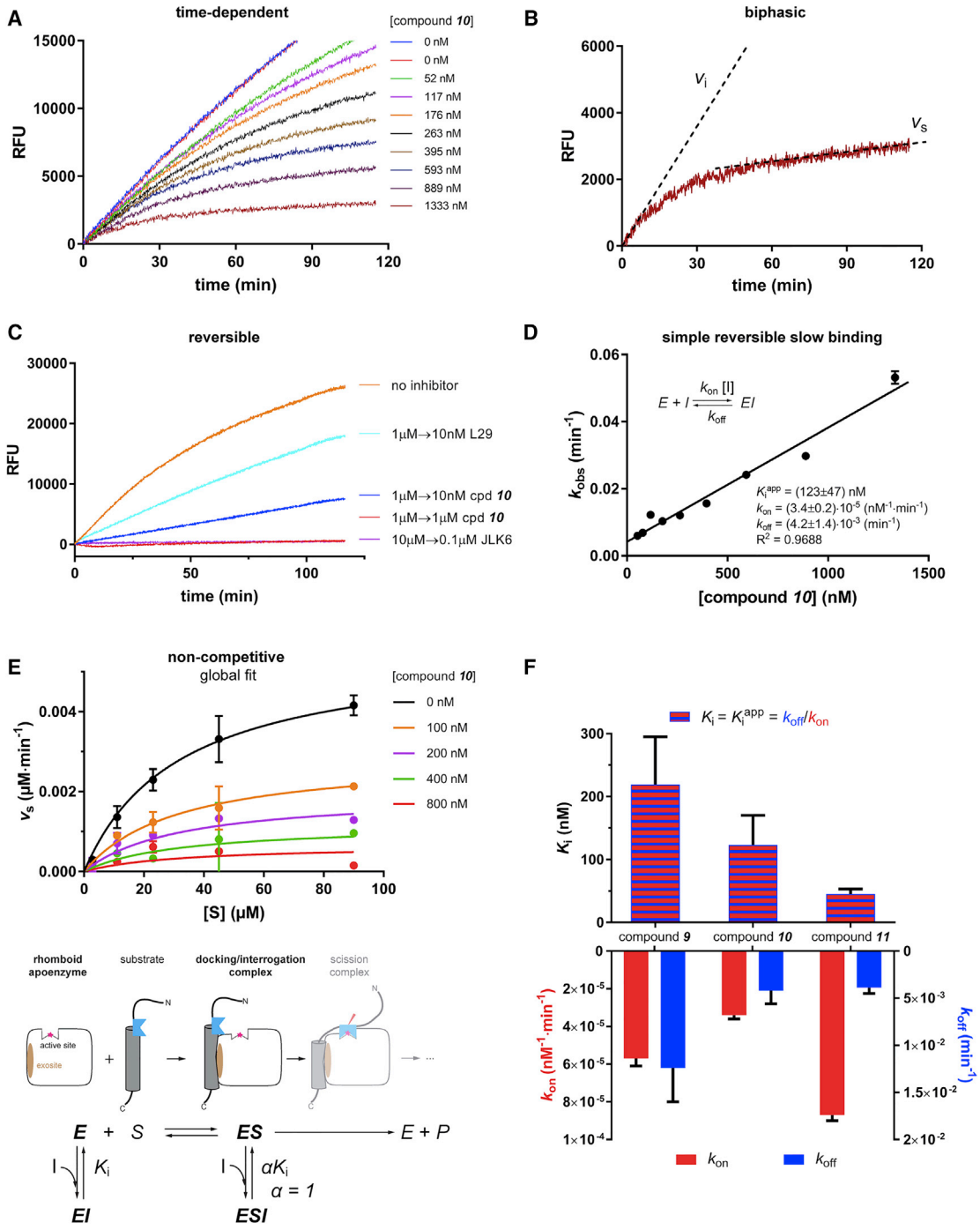


Figure 4. Mechanism of Inhibition of Rhomboid Protease GlpG by Peptidyl Ketoamides Analyzed Using Fluorogenic Transmembrane Peptide Substrates

(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

(legend continued on next page)

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

Compound	GlpG				YqgP
	K_i (nM)	k_{on} (10^{-6} nM^{-1} $\cdot min^{-1}$)	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 proteases 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.

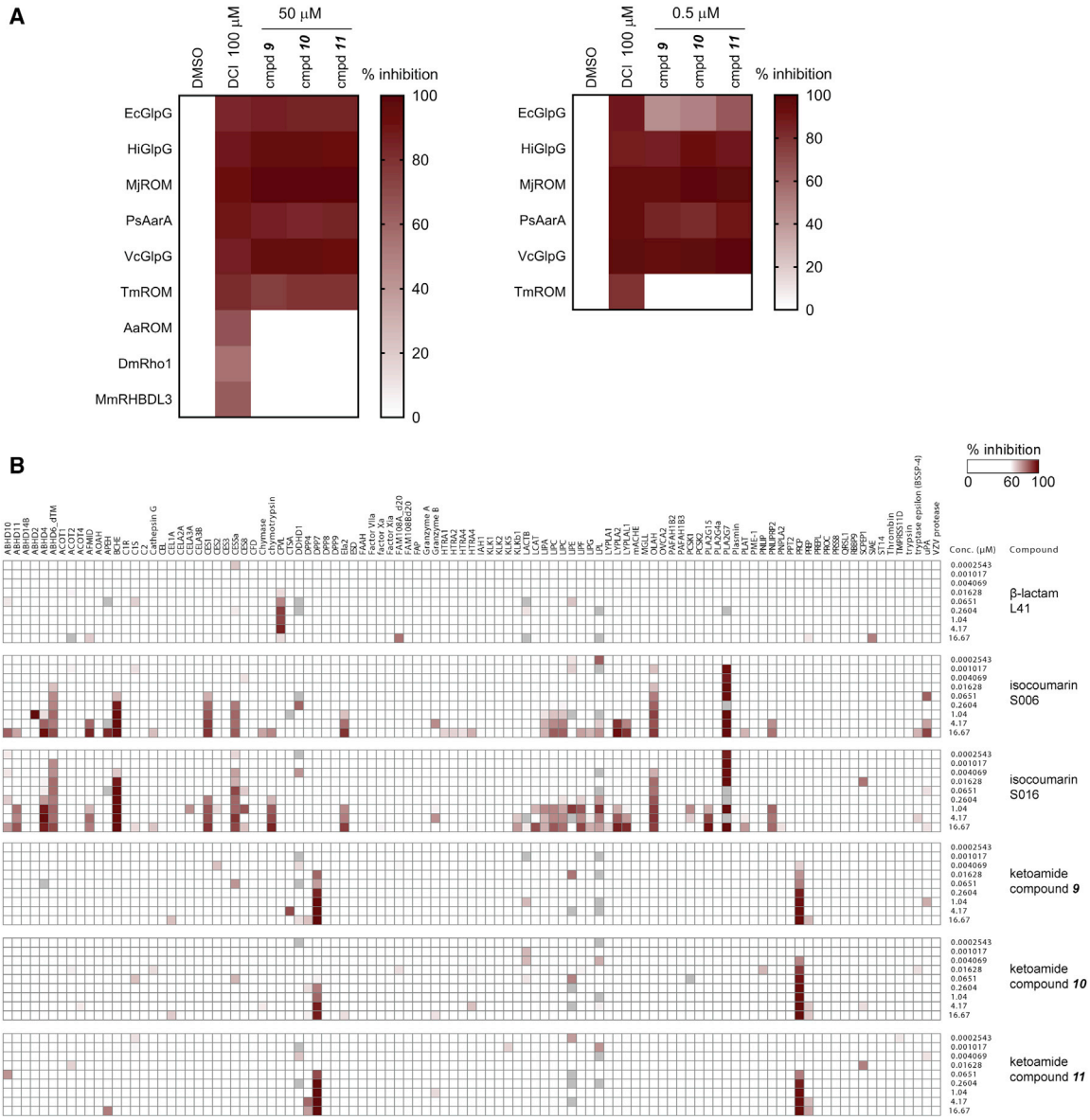


Figure 5. Selectivity of Peptidyl Ketoamides

(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,

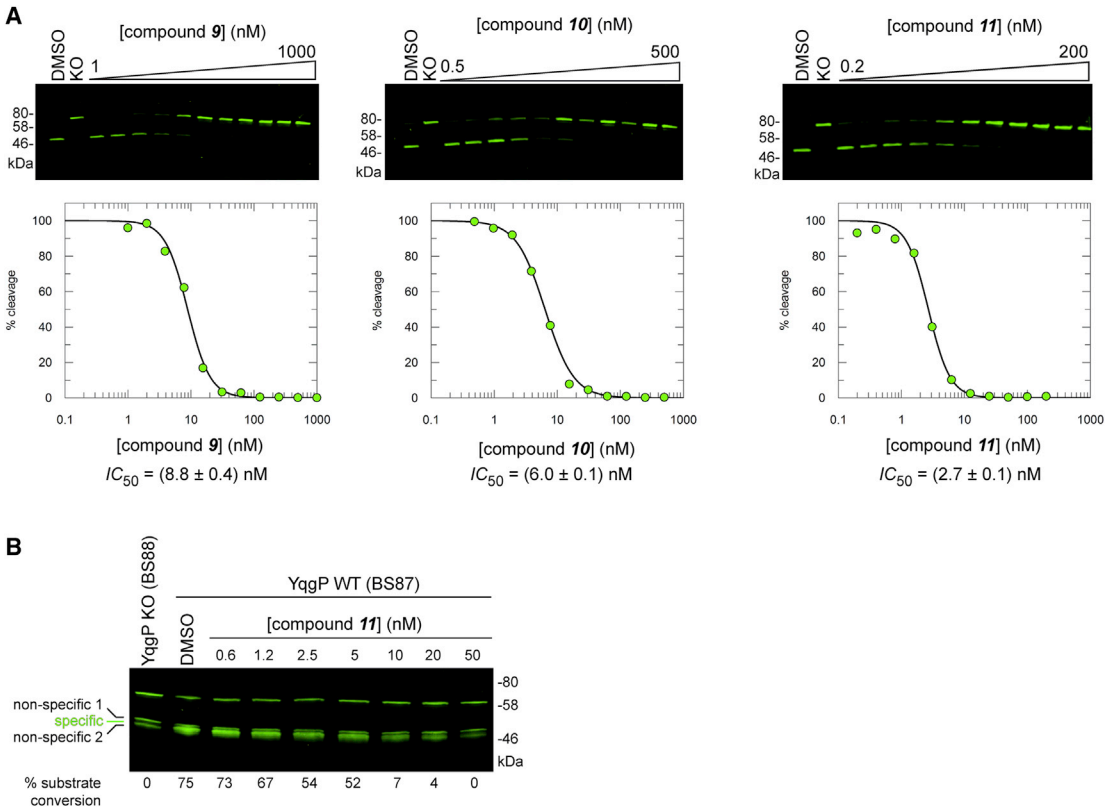


Figure 6. Peptidyl Ketoamides Potently Inhibit Rhomboid Activity in the Membranes of Living Cells

(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)

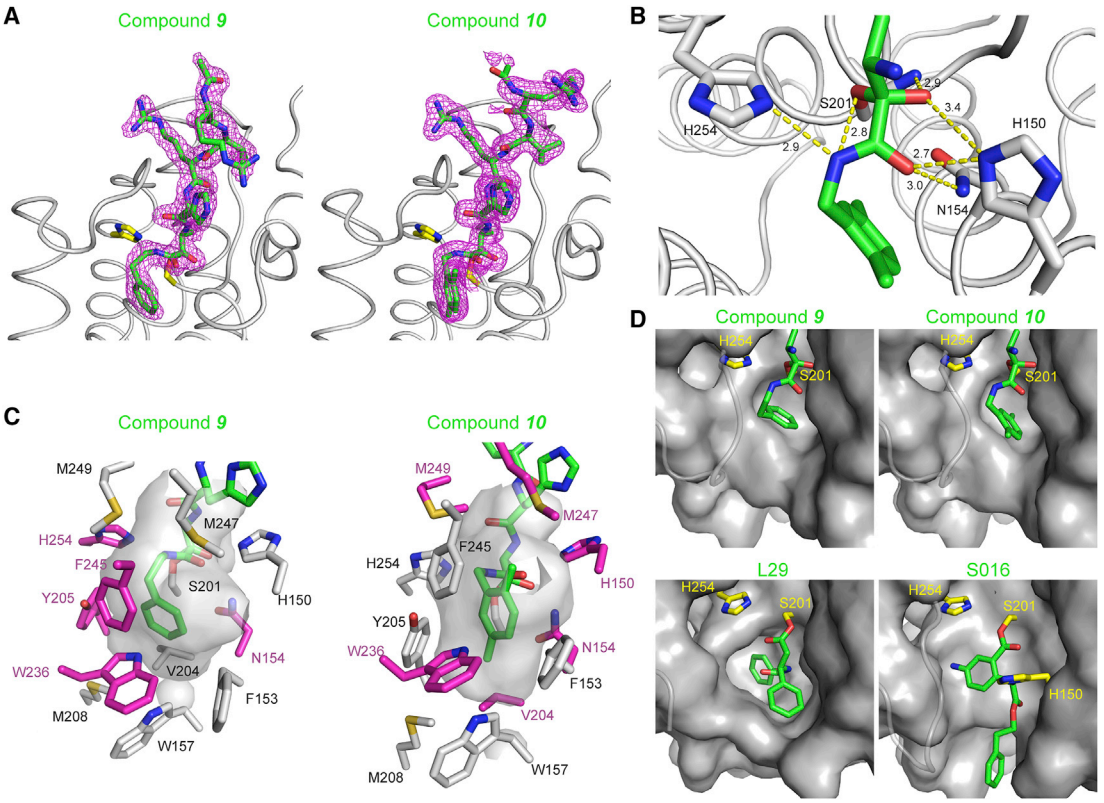


Figure 7. N-Substituted Peptidyl Ketoamides Bind GlpG in a Substrate-like Manner and Occupy the S4 to S2′ Subsites of the Rhomboid Active Site

(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 non-prime 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

SUPPLEMENTAL INFORMATION

Supplemental Information includes one table and supplemental text and can be found with this article online at https://doi.org/10.1016/j.chembiol.2017.09.007.

AUTHOR CONTRIBUTIONS

K.S. conceived and coordinated the study, designed and evaluated experiments, and wrote the paper with the input of the co-authors. P.M. and S.S. designed and S.S. performed all chemical syntheses, and A.T. designed, performed, and evaluated all kinetics and inhibition measurements with the help of K.Š. and J.Š. in the initial stages. K.R.V., D.C.M., and P.P. performed all crystallographic experiments and evaluated the data with the input of M.L. M.T.N.N. and S.H.L.V. performed selectivity profiling against rhomboid proteases, J.B. performed all experiments on *B. subtilis*, and D.C.J. and D.A.B. performed and evaluated the EnPlex experiments.

ACKNOWLEDGMENTS

We thank Mirka Blechová and Lenka Monincová for peptide synthesis and HPLC/MS analyses, Radko Souček for amino acid analysis, Jiří Brynda for help with crystallographic data analysis, Pavel Srb and Marek Ingr for help and advice on data fitting and kinetics, Petra Rampírová for DNA cloning

and laboratory assistance, Libor Krásný and Tom Silhavy for reagents and advice, and Colin Adrain and Cyril Bařinka for critical reading of the manuscript. K.S. was a recipient of a Purkyne Fellowship of the Academy of Sciences of the Czech Republic and acknowledges support also from EMBO (installation grant no. 2329), Ministry of Education, Youth and Sports of the Czech Republic (project nos. LK11206 and LO1302), Marie Curie Career Integration Grant (project no. 304154), Gilead Sciences & IOCB Research Centre, and the National Subvention for Development of Research Organisations (RVO: 61388963) to the Institute of Organic Chemistry and Biochemistry. P.P. acknowledges support from the Ministry of Education of the Czech Republic (program "NPU I"), project no. LO1304, and J.S. and J.B. acknowledge support from the Grant Agency of Charles University (GA UK) in Prague (PhD grant project nos. 232313 and 170214). K.R.V. was supported by an MRC grant (MC_U105184322) as part of R. Henderson's group. D.A.B. was supported by Josie Robertson Foundation and the MSKCC (core grant P30 CA008748), and D.C.J. by the NIH (T32 GM115327-Tan). M.L. was supported by the Czech Science Foundation (grant number P208/12/G016). M.T.N.N. and S.H.L.V. are supported by the Deutsche Forschungsgemeinschaft and the Ministerium für Innovation, Wissenschaft und Forschung des Landes Nordrhein-Westfalen.

Received: April 13, 2017
Revised: August 19, 2017
Accepted: September 18, 2017
Published: October 26, 2017

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.

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.

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 NR698Ä*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*)

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.

Publication 4

Title page

Rhomboid intramembrane protease YqgP licenses bacterial membrane protein quality control as adaptor of FtsH AAA protease.

[1, 2, #]Jakub Began, [3, #]Baptiste Cordier, [1,4]Jana Březinová, [3]Jordan Delisle, [1]Rozálie Hexnerová, [1]Pavel Srb, [1]Petra Rampírová, [1]Milan Kožíšek, [5]Mathieu Baudet, [5]Yohann Couté, [3]Anne Galinier, [1,6]Václav Veverka, [3,7]Thierry Doan* and [1]Kvido Strisovsky*

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

[2]Department of Genetics and Microbiology, Faculty of Science, Charles University, Viničná 5, Prague, 128 44, Czech Republic

[3]Laboratoire de Chimie Bactérienne (LCB), Institut de Microbiologie de la Méditerranée (IMM), Aix Marseille Univ, CNRS, UMR 7283, 31 Chemin Joseph Aiguier, 13402, Marseille Cedex 20, France

[4]Department of Biochemistry, Faculty of Science, Charles University, Albertov 6, Prague, 128 43, Czech Republic

[5]Univ. Grenoble Alpes, CEA, Inserm, IRIG-BGE, 38000 Grenoble, France

[6]Department of Cell Biology, Faculty of Science, Charles University, Viničná 7, Prague, 128 00, Czech Republic

[7]Laboratoire d'Ingénierie des Systèmes Macromoléculaires Chimie Bactérienne (LISM), Institut de Microbiologie de la Méditerranée (IMM), Aix Marseille Univ, CNRS, UMR 7255, 31 Chemin Joseph Aiguier, 13402, Marseille Cedex 20, France

[#]equal contribution

*Authors for correspondence (email: kvido.strisovsky@uochb.cas.cz, tdoan@imm.cnrs.fr)

Keywords

Intramembrane protease, rhomboid, membrane transporter, proteostasis, ER-associated degradation

Running title

Rhomboid controls MgtE proteostasis.

Abstract

Magnesium homeostasis is essential for life and depends on magnesium transporters, whose activity and ion selectivity need to be tightly controlled. Rhomboid intramembrane proteases pervade the prokaryotic kingdom, but their functions are largely elusive. Using proteomics, we find that *Bacillus subtilis* rhomboid protease YqgP interacts with the membrane-bound ATP-dependent processive metalloprotease FtsH and cleaves MgtE, the major high-affinity magnesium transporter in *B. subtilis*. MgtE cleavage by YqgP is potentiated in conditions of low magnesium and high manganese or zinc, thereby protecting *B. subtilis* from Mn^{2+}/Zn^{2+} toxicity. The N-terminal cytosolic domain of YqgP binds Mn^{2+} and Zn^{2+} ions and facilitates MgtE cleavage. Independently of its intrinsic protease activity, YqgP acts as a substrate adaptor for FtsH, a function that is necessary for degradation of MgtE. YqgP thus unites protease and pseudoprotease function, hinting at the evolutionary origin of rhomboid pseudoproteases such as Derlins that are intimately involved in eukaryotic ER-associated degradation (ERAD). Conceptually, the YqgP-FtsH system we describe here is analogous to a primordial form of 'ERAD' in bacteria and exemplifies an ancestral function of rhomboid-superfamily proteins.

Introduction

Rhomboids are intramembrane serine proteases widespread across the tree of life. In eukaryotes, rhomboid proteases regulate signalling via the epidermal growth factor receptor, mitochondrial quality control or invasion of malaria parasites into the host cells (reviewed in Urban, 2016). Rhomboid proteases also pervade the prokaryotic kingdom (Kinch & Grishin, 2013, Koonin, Makarova et al., 2003), suggesting sufficient biological significance for evolutionary conservation and expansion. However, their functions in bacteria are much less well understood than in eukaryotes. Notably, in eukaryotes, the rhomboid superfamily includes a number of pseudoproteases of important biological roles, such as iRhoms and Derlins, but the mechanistic and evolutionary aspects of rhomboid pseudoprotease functions have not been clarified (Ticha, Collis et al., 2018).

In the Gram-negative eubacterium *Providencia stuartii*, the rhomboid protease AarA is required to process a pro-form of TatA, a component of the Twin-Arginine Translocase secretion apparatus (Stevenson, Strisovsky et al., 2007), but this function does not seem to be widely conserved because most bacteria encode a mature form of TatA. In the archaebacterium *Haloferax volcanii*, a rhomboid protease is involved in protein glycosylation of the S-layer (Costa, Cerletti et al., 2018, Parente, Casabuono et al., 2014), but the molecular mechanism has not been uncovered yet. Similarly, the main model rhomboid protease GlpG of *Escherichia coli*, widely distributed in Gram-negative bacteria, is extremely well characterised structurally and mechanistically (Baker & Urban, 2012, Strisovsky, Sharpe et al., 2009, Zoll, Stanchev et al., 2014), yet its biological function is unknown. A recent report proposed a role of GlpG in the survival of a pathogenic strain of *E. coli* in the mouse gut (Russell, Richards et al., 2017), but the mechanism has not been demonstrated and it is not clear how specific the phenotype is, because rescue experiment has not been reported. Current knowledge of functions of rhomboid proteases in Gram-positives is similarly sketchy. An isolated report indicated that YqgP may play a role

in cell division and may be required for glucose export (hence the therein coined alternative name GluP) (Mesak, Mesak et al., 2004), but rescue experiment was not reported and the involved substrates were not investigated, leaving the molecular mechanisms of these functions unknown.

Therefore, to shed light on one of the significantly populated subclasses of bacterial rhomboid proteases in a model Gram-positive organism, we focused on *Bacillus subtilis* YqgP and used quantitative proteomics to identify its substrates and interactors. We find that YqgP cleaves the high-affinity magnesium transporter MgtE and that YqgP interacts with the membrane-anchored metalloprotease FtsH. At low extracellular concentration of magnesium cations and high concentration of manganese or zinc cations, cleavage of MgtE by YqgP is potentiated, and the globular N-terminal cytosolic domain of YqgP represents the manganese/zinc sensing unit. The second molecular role of YqgP is presenting MgtE or its cleavage products as substrates to FtsH, for which the proteolytic activity of YqgP is dispensable but its unoccupied active site is essential. YqgP thus fulfils both protease and pseudoprotease functions in tandem with FtsH, representing an ancestral proteolytic platform dedicated to the regulated degradation of polytopic membrane proteins, functionally equivalent to regulatory ERAD in eukaryotes. Our results shed light on the evolution of membrane proteostasis control in response to environmental stimuli, and on the arising of pseudoproteases, which are surprisingly common in the rhomboid superfamily (Adrain & Freeman, 2012, Freeman, 2014).

Methods

Materials and chemicals

All common chemicals were from Sigma Aldrich or VWR unless otherwise indicated. Oligonucleotides were from Sigma Aldrich and Eurogentec; restriction and DNA modification enzymes were from New England Biolabs.

Plasmids and DNA cloning

All DNA constructs were created by PCR and isothermal assembly (Gibson, 2011), and verified by Sanger sequencing. All constructs used in this work are summarised in Table EV1.

Bacterial strains, media and growth conditions

All *B. subtilis* strains used in this study were derived from the *168 trpC2+* strain (Nicolas, Mader et al., 2012), and construction details of the strains are listed in Table EV2. For routine work, *B. subtilis* strains were streaked onto solid LB agar plate (supplemented with appropriate antibiotic) from -80°C stock solution and grown overnight at 37°C. Subsequently, single colony was inoculated into the fresh LB medium and cultivated until the pre-culture reached OD_{600} around 1.0. Culture was then diluted in the fresh growth medium (LB or M9) to starting OD_{600} of 0.01 to 0.025 and cultivated until the exponential

phase was reached at OD_{600} ranging from 0.6 to 1.5. Depending on the experimental setup, protein expression from an ectopic site was induced either by adding 1%(w/v) D-xylose (for constructs under the control of xylose promoter, P_{xyl}) or IPTG ($P_{hyperspank}$ promoter) at the start of cultivation as indicated, at 37°C, until OD_{600} reached 1, unless stated otherwise. For regular growth, both *B. subtilis* and *E. coli* cells were cultivated in LB medium (Merck) at 37°C. For *in vivo* activity assays, *B. subtilis* cells were cultivated in modified M9 minimal medium (low- or high- Mg^{2+} content) composed of M9 salts solution (6.8 g/l $Na_2HPO_4.2H_2O$, 3 g/l KH_2PO_4, pH 7.4, 0.5 g/l NaCl, 1 g/l NH4Cl), 0.5 % (w/v) D-glucose, 10 µM (low Mg^{2+}) or 1 mM $MgSO_4$ (high Mg^{2+}), 1 µM $MnCl_2$, 100 µM $CaCl_2$, 10 µM $FeSO_4$ and 18 amino acids including all genetically encoded ones except lysine and tyrosine (Harwood & Cutting, 1990). All minimal media in which toxicity assays were performed were prepared in water of ultrapure quality, processed through Milli-Q® Reference Water Purification System (Merck) and having resistivity of at least 18.2 MΩ.cm (at 25°C) and TOC values below 5 ppb. For translational shut-off chase experiments, *B. subtilis* cultures were grown to mid-exponential phase, in LB medium at 30°C. At this point, expression of *yqgP* from the $P_{hyperspank}$ promoter was induced by adding 0.1 mM IPTG for 30 min. At a given time-point, proteosynthesis was blocked by the addition of tetracycline to a final concentration of 20 µg/ml. Cation toxicity assays were performed by adding concentrations of $MnCl_2$, $ZnCl_2$, $CoSO_4$ or $NiCl_2$ into M9 minimal medium at various growth time points, as specified in the result section. *Bacilli* cultures for SILAC-based proteomic experiments were cultured in modified M9 minimal medium (see above) supplemented with either naturally occurring "light" ($^{12}C_6{}^{14}N_2$) or "heavy" ($^{13}C_6{}^{15}N_2$) isotopic variants of L-lysine (Silantes GmbH, Germany).

Proteins uniformly labelled by ^{15}N or $^{15}N^{13}C$ for NMR studies were overexpressed in *E. coli* BL21(DE3) (YqgP NTD) or *E. coli* Lemo21(DE3) (YqgP CTD), grown in M9 minimal medium containing partial M9 salts solution (6.8 g/l $Na_2HPO_4.2H_2O$, 3 g/l KH_2PO_4, pH 7.4, 0.5 g/l NaCl), 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.5% (w/v) D-glucose (U-$^{13}C_6$, 99%, Cambridge Isotope Laboratories), and 1 g/l NH_4Cl (^{15}N, 99%, Cambridge Isotope Laboratories).

Protein expression and purification

Proteins for structural and calorimetric studies were prepared by overexpression in *E. coli* BL21Star™(DE3) (Invitrogen) and Lemo21(DE3) (Wagner, Klepsch et al., 2008) (New England Biolabs) strains grown in LB medium. This included the C-terminally 6×Histidine-tagged TEV site-containing N-terminal extramembrane domain of YqgP [6×His-TEVsite-YqgP(1-177); NTD], and the N-terminally 6×Histidine-tagged TEV-site containing C-terminal extramembrane domain of YqgP [YqgP(384-507)-TEVsite-6×His; CTD] expressed from the pET25b or pHIS-2 vector, respectively. The N-terminal cytosolic extramembrane domain of MgtE [MgtE(2-275)] for antibody production was overexpressed in *E. coli* Lemo21(DE3) as a GST-6×Histidine-TEVsite-MgtE(2-275) construct from the pGEX6P1 vector. His_6-YqgP(1-170) and His_6-YqgP(385-507) for antibody production were expressed in *E. coli* BL21(DE3)pLysS from pRSET-A vector and purified using metal-chelate affinity chromatography.

Typically, protein production was induced with 0.5 mM IPTG at 25°C with aeration and gentle shaking overnight. Bacterial cultures were harvested by centrifugation at 5000×g for 10 min at 4°C, and

cells were resuspended in isolation buffer composed of 20 mM HEPES, pH 7.4, 300 mM NaCl, 10%(v/v) glycerol, 10 mM imidazole, 1 mM PMSF and 1× cOmplete™ EDTA-free Protease Inhibitor Cocktail (Roche). Cells were lysed by three passages through the high-pressure homogenizer Emulsiflex®-C3 (Avestin, Inc.), and lysates were cleared from the cell debris by centrifugation (15000×g, 30 min, 4°C). Supernatants were loaded on Ni-NTA Agarose (Qiagen) equilibrated in isolation buffer and His-tagged protein was eluted specifically with increasing concentration of imidazole in elution buffer (20 mM HEPES, pH 7.4, 300 mM NaCl, 10%(v/v) glycerol). The GST-tagged MgtE(2-275) domain was additionally purified using Glutathione Sepharose® 4 Fast Flow (Merck) and specifically eluted by 10 mM reduced glutathione into 20 mM HEPES, pH 7.4, 300 mM NaCl, 10%(v/v) glycerol. The 6×Histidine tags from YqgP domains and GST-6×Histidine tags from MgtE domain were removed by incubation with 6×Histidine tagged TEV protease at protease-to-substrate ratio of ~1:200 (w/w), in 5 mM Tris-HCl, pH 8, 1 mM EDTA, 1 mM β-mercaptoethanol for 16 h at 25°C. The cleaved protein was then buffer-exchanged into isolation buffer (this paragraph) using PD-10 desalting column (GE Healthcare Life Sciences), and loaded onto Ni-NTA agarose. The flow-through fraction containing tag-free YqgP or MgtE domains were collected, concentrated by ultrafiltration (Vivaspin, Sartorius) and used for further applications.

In vitro translation

Reference fragments of *B. subtilis* MgtE encompassing the first 300, 315, 330, 340, 355 and 370 residues of MgtE were generated by PCR amplification of the respective coding regions from *B. subtilis* 168 genomic DNA, and *in vitro* transcribed and translated as described previously (Lemberg & Martoglio, 2003, Strisovsky et al., 2009). Briefly, PCR was performed using set of *mgtE*-specific primers containing SP6 RNA polymerase promoter and ribosome binding site in the forward primer, and stop codon in the reverse primer. Messenger RNAs were purified by LiCl precipitation and translated using Wheat Germ Extract (Promega). The resulting crude translation mixtures containing the MgtE reference fragments were separated using SDS-PAGE, and visualised by polyclonal anti MgtE(2-275) antibody and immunoblotting.

Production of primary antibodies

Purified His_6-YqgP(1-170), His_6-YqgP(385-507) and MgtE(2-275) were used to raise rabbit polyclonal antibodies (Agro-Bio, France). Forty two days after immunization, the sera were collected and purified by affinity chromatography on an Affi-Gel 10 (Bio-Rad) column containing covalently immobilized antigen as described before (Campo & Rudner, 2006). Briefly, approximately 2 mg of purified antigen was dialyzed into coupling buffer (20 mM HEPES pH 8, 200 mM NaCl, 10% (v/v) glycerol) and coupled to 1 ml Affigel-10 resin (BioRad) as described by the manufacturer. After coupling, the antigen resin was washed with 100 mM glycine pH 2.5 and neutralized with 1×PBS to remove all uncoupled protein. Twenty ml of antisera were batch-adsorbed to the antigen-resin overnight at 4°C. The resin was loaded into a 5 ml column and washed with 20 ml 1×PBS, 50 ml 1×PBS 500 mM NaCl and 10 ml 0.2×PBS. The

affinity-purified antibodies were then eluted with 100 mM glycine pH 2.5. The peak fractions were pooled and dialyzed with three changes into 1×PBS with 50% (v/v) glycerol and stored at -20°C.

Immunoblotting

B. subtilis cultures in exponential phase (OD_{600} of ~1) were harvested by centrifugation (5000 ×g, 10 min, 25°C). Cell pellet was resuspended in lysis buffer containing 20 mM Tris-HCl, pH 7.5, 10 mM EDTA, 1 mM PMSF, and 1x cOmplete™ EDTA-free Protease Inhibitor Cocktail (Roche), 1 mg/ml lysozyme, 2500 U Pierce Universal Nuclease (cat no. 88700, ThermoFisher) and incubated at 37°C, for 10 min. Reducing SDS-PAGE loading buffer was added and the mixture was heated to 60°C for 10 min with vigorous shaking. Samples were separated on Tris-glycine SDS-PAGE and electroblotted onto a nitrocellulose membrane (Serva). Equivalent total protein amounts were loaded as estimated by the OD_{600} values, and after electrotransfer, total protein in each lane was quantified using the Revert total protein stain (LI-COR®). The following antibodies were used for immunodetection: anti-Thioredoxin antibody produced in rabbit (Merck), polyclonal anti-MgtE (this study), polyclonal anti-$YqgP_{NTD}$ (this study), and polyclonal anti-$YqgP_{CTD}$ (this study) antibodies all raised in rabbit (Agro-Bio, France), and rabbit polyclonal anti-FtsH (gift from Prof. Thomas Wiegert, Zittau, Germany). Donkey anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody DyLight800 conjugate (Invitrogen) was used for visualisation by near infrared laser scanning on the Odyssey® CLx Imaging System (LI-COR®) according to manufacturer's instructions. The resulting fluorescence intensity data for individual proteins were optionally used for quantification, as specified in each case. For occasional chemiluminiscent detection (as noted in figure legends), goat anti-rabbit IgG conjugated to horseradish peroxidase (HRP) (Dako) was used as a secondary antibody, and HRP activity was detected by the Western Lightning Pro kit (PerkinElmer).

Rhomboid activity assay in vivo

B. subtilis cells overexpressing particular YqgP variant under the control of P_{xyl} promoter were cultivated as described above. The amounts of the uncleaved and cleaved forms of MgtE or model substrate TatA were measured by quantitative fluorescence immunoblotting as described in the previous section. Since the integrated intensity of fluorescence signal for a given band is proportional to the molarity of the antigen, substrate conversion at steady state was quantified as the ratio of fluorescence intensities of bands corresponding to the cleavage product (P) divided by the sum of signals for the full-length substrate (S) and the cleavage product (P), i.e. signal(P)/[signal(S) +signal(P)]. Relative 'specific activities' at steady state *in vivo* were then calculated by dividing substrate conversion by the signal from the expressed rhomboid enzyme (YqgP variant) corrected for the total protein signal in that gel lane.

SILAC-based quantitative proteomics and data evaluation

For SILAC-based quantitative proteomics, the BS50 and BS51 (Table EV2) strains of *B. subtilis* auxotrophic for lysine were cultivated in the medium described in section *Bacterial strains, media and growth conditions*. Equal amounts of heavy and light cell cultures (based on optical density at 600 nm, OD_{600}) were mixed at harvest and centrifuged. Cell pellets were then resuspended in lysis buffer containing 20 mM HEPES, pH 7.4, 100 mM NaCl, 10% (v/v) glycerol, 1 mM EDTA, 50 µg/ml lysozyme and 0.4x MS-SAFE Protease and Phosphatase Inhibitor cocktail (Merck) and incubated at 37°C, for 15 min and on ice for 15 min, afterwards. Cells were lysed by sonication using Sonopuls HD2200 device (Bandelin) set to 4×30 s pulse cycle and 30% power. Cell debris were removed by centrifugation (15000×g, 30 min, 4°C). Crude membranes were isolated by ultracentrifugation (100 000×g, 1.5h, 4°C) and washed sequentially with 0.1 M Na_2CO_3 and 1 M NaCl.

B. subtilis transmembrane protein-enriched fractions were separated on 4-20% gradient Tris-glycine SDS-PAGE system (Bio-Rad). Two SILAC experiments with swapped labelling were conducted. Experiment 1 [marked “heavy-to-light” or (H/L)]: heavy BS50 (YqgP), light BS51 (YqgP.S288A); and Experiment 2 [marked “light-to-heavy” or (L/H)]: light BS50, heavy BS51]. Both experiments were resolved in separate gel lanes, which were subsequently sliced into 5 fractions (A-E) each (Fig. 1), and individually digested using a standard in-gel digestion protocol. Briefly, every gel slice was first destained with 25 mM ammonium bicarbonate in 50% (v/v) acetonitrile (ACN, Merck), dehydrated and shrunk using acetonitrile (ACN) in a reducing environment of 5 mM 1,4-dithiothreitol (DTT, Merck) and incubated for 30 min at 65 °C. Fractions were then alkylated using 12.5 mM iodoacetamide (Merck), shrunk and digested using 0.1 µg trypsin at 37 °C overnight. Tryptic peptides were extracted with 60% (v/v) ACN in 0.5% (v/v) trifluoroacetic acid (TFA, Merck), dried and reconstituted in 20 µl of 2% (v/v) ACN with 0.1% (v/v) formic acid (FA, Merck) and analysed via LC-MS/MS.

The LC-MS/MS analyses were performed using UltiMate 3000 RSLCnano system (Dionex) coupled to a TripleTOF 5600 mass spectrometer with a NanoSpray III source (AB Sciex). After injection, the peptides were trapped and desalted in 5% (v/v) ACN/ 0.1 % (v/v) FA at a flow rate of 5 µL/min on an Acclaim® PepMap100 column (5 µm, 2 cm × 100 µm ID, Thermo Scientific) for 5 minutes. The separation of peptides was performed on an Acclaim® PepMap100 analytical column (3 µm, 25 cm × 75 µm ID, Thermo Scientific) using a gradient from 5% (v/v) to 18% (v/v) ACN, over 12 min, with a subsequent rise to 95 % (v/v) ACN/ 0.1% (v/v) FA, over 30 min. TOF MS scans were recorded from 350 to 1250 m/z and up to 18 candidate ions per cycle were subjected to fragmentation. Dynamic exclusion was set to 10 s after one occurrence. In MS/MS mode, the fragmentation spectra were acquired within the mass range of 100 – 1600 m/z.

The quantitative mass spectrometric data files were processed and analysed using MaxQuant (v1.5.2.8) (Cox & Mann, 2008). The search was performed using a Uniprot/Swissprot *B. subtilis* database (downloaded 17/05/15) with common contaminants included. Enzyme specificity was set to trypsin, with methionine oxidation as a variable modification. Cysteine carbamidomethylation was considered as a fixed modification. The heavy SILAC label was set to K8, the minimal peptide length to 6, and 2 missed cleavages were allowed. Proteins were considered as identified if they had at least one unique peptide, and quantified if they had at least one quantifiable SILAC pair. Transmembrane topology

predictions were obtained using Phobius (Käll, Krogh et al., 2004) and additional sequence visualisations were obtained using QARIP software (Ivankov, Bogatyreva et al., 2013) (Fig. 1).

Quantitative proteomics for interactor identification

All procedures were performed essentially as described (Doan, Morlot et al., 2009). For crude membrane preparation, cells were grown in LB medium, at 37°C. At OD_{600} of 0.6, 50 mL of culture were harvested for membrane preparation. Membrane proteins were solubilized by the addition of the detergent NP-40 to a final concentration of 1% (v/v). The soluble fraction was mixed with 25 µl anti-GFP antibody resin (Chromotek, Germany) and rotated for 3h at 4°C. The resin was pelleted at 3000×g and the supernatant was removed. After washes, immunoprecipitated proteins were eluted by the addition of 85.5 µl of SDS PAGE sample buffer and heated for 15 minutes at 50°C. The eluted material (the IP) was transferred to a fresh tube and 2-mercaptoethanol was added to a final concentration of 10% (v/v).

Proteins extracted from the co-immunoprecipitation eluates were resolved using NuPAGE 4-12% system (ThermoFisher Scientific), stained with Coomassie blue (R250, Bio-Rad) and in-gel digested using modified trypsin (Promega, sequencing grade) as described, previously (Casabona, Vandenbrouck et al., 2013). Of note, samples from replicates 1 and 2 were prepared using oxidation and reduction-alkylation procedures as described (Jaquinod, Trauchessec et al., 2012). The peptides were analysed by online nanoLC-MS/MS (UltiMate 3000 and LTQ-Orbitrap Velos, Thermo Scientific). In particular, peptides were loaded on a 300 µm x 5 mm PepMap C18 precolumn (Thermo Scientific) and separated on a homemade 75 µm x 150 mm C18 column (Gemini C18, Phenomenex). MS and MS/MS data were acquired using Xcalibur (Thermo Scientific). Mascot Distiller (Matrix Science) was used to produce mgf files before identification of peptides and proteins using Mascot (version 2.6) through concomitant searches against Uniprot (*Bacillus subtilis* strain 168 taxonomy, September 2018 version), GFP-tagged YqgP sequences, classical contaminants database (homemade) and the corresponding reversed databases. The Proline software (http://proline.profiproteomics.fr) was used to filter the results (conservation of rank 1 peptides, peptide identification FDR < 1% as calculated on peptide scores by employing the reverse database strategy, minimum peptide score of 25, and minimum of 1 specific peptide per identified protein group) before performing a compilation, grouping and comparison of the protein groups from the different samples. Proteins were considered as potential partners of the bait if they were identified only in the positive co-IPs with a minimum of 3 weighted spectral counts or enriched at least 5 times in positive co-IPs compared to control ones based on weighted spectral counts.

Isothermal titration calorimetry

The binding of divalent cations to the extramembrane domains of YqgP (NTD, CTD) was investigated using titration microcalorimetry. All titration experiments were performed in 50 mM Tris-HCl, pH 7.3, 150 mM NaCl at 25°C using Auto-iTC_{200} instrumentation (MicroCal, Malvern Panalytical Ltd). Prior to the ITC

measurements, protein samples were dialyzed into the same buffer in which divalent cation salt solutions were prepared. Typically, 200 µl of the protein sample was titrated by stepwise 2 µl injections of divalent cation solution. Protein and ligand concentrations were adjusted as follows: 1.5 mM NTD was titrated by 50 mM $MnCl_2$, 50 mM $MgSO_4$, 50 mM $ZnCl_2$ or 50 mM $CaCl_2$, 0.7 mM NTD by 15 mM $CoCl_2$ and 0.25 mM NTD by 7 mM $NiCl_2$ buffered solutions. Control experiments included ligand titration into the buffer only. Concentrations of stock solutions of the analysed proteins were determined by quantitative amino acid analysis, and concentrations of the stock solutions of divalent metal cation salts were determined by elemental analysis. Titration data were processed and analysed using MicroCal Origin 7.0 (MicroCal, Malvern Panalytical Ltd).

Nuclear magnetic resonance spectroscopy

For structure determination of YqgP N-terminal domain ($YqgP_{NTD}$), 350 µl samples of 1 mM $^{15}N/^{13}C$ uniformly labelled protein in 25 mM phosphate buffer pH 6.5 containing 150 mM NaCl and 5% D_2O/ 95% H_2O were used. All NMR spectra for backbone assignments and three-dimensional structure determination were collected on 850 MHz Bruker Avance spectrometer equipped with triple resonance ($^{15}N/^{13}C/^{1}H$) cryoprobe at 37°C. In a series of double and triple resonance spectra (Renshaw, Veverka et al., 2004, Veverka, Lennie et al., 2006) were recorded to determine essentially complete sequence specific resonance backbone and side chain assignments. Constraints for ^{1}H ^{1}H distance required to calculate the structure of $YqgP_{NTD}$ were derived from 3D $^{15}N/^{1}H$ NOESY HSQC and $^{13}C/^{1}H$ NOESY HMQC, which were acquired using a NOE mixing time of 100 ms. The family of converged structures for $YqgP_{NTD}$ was initially calculated using Dyana 2.1 (Herrmann, Guntert et al., 2002). The combined automated NOE assignment and structure determination protocol was used to automatically assign the NOE cross peaks identified in the NOESY spectra and to produce preliminary structures. In addition, backbone torsion angle constraints, generated from assigned chemical shifts using the TALOS+ program (Shen, Delaglio et al., 2009) were included in the calculations. Subsequently, five cycles of simulated annealing combined with redundant dihedral angle constraints were used to produce a set of converged structures with no significant restraint violations (distance and van der Waals violations <0.2 Å and dihedral angle constraint violation <5 °), which was further refined in explicit solvent using the YASARA software with the YASARA forcefield (Harjes, Harjes et al., 2006). The structures with the lowest total energy were selected and analysed using the Protein Structure Validation Software suite (www.nesg.org). Cation binding experiments were performed in 50 mM Tris-HCl, pH 7.3, 150 mM NaCl, 5% D_2O, at 37°C, when the same aliquots of the ligand were added to the protein solution stepwise, after every titration measurement.

Data availability

The mass spectrometry (MS) proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) (Deutsch, Csordas et al., 2017) via the PRIDE (Perez-Riverol, Csordas et al., 2019) partner repository with the dataset identifiers PXD014578

and PXD014566. The NMR structural data have been deposited in the Protein Data Bank (PDB) under accession code 6R0J and in the Biological Magnetic Resonance Data Bank (BMRB) under the accession code 34376.

Results

Quantitative proteomics reveals candidate interactors and substrates of B. subtilis rhomboid protease YqgP.

B. subtilis genome encodes two rhomboid protease genes, *ydcA* and *yqgP* (also known as *gluP* (Mesak et al., 2004)). No proteolytic activity has been detected for YdcA so far (Lemberg, Menendez et al., 2005, Urban, Schlieper et al., 2002b), while YqgP is a commonly used model rhomboid protease cleaving a number of synthetic substrates (Lemberg et al., 2005, Ticha, Stanchev et al., 2017a, Ticha, Stanchev et al., 2017b, Urban & Wolfe, 2005). YqgP has homologs in a number of Gram-positive bacteria, including *Bacilli,* but also *Staphylococci, Listeriae* and others (http://www.ebi.ac.uk/interpro/protein/P54493/similar-proteins), and thus represents an attractive system to explore the cell biology and functions of bacterial rhomboid proteases. To reveal the repertoire of YqgP substrates, we generated a strain deficient in *yqgP* and re-expressed *yqgP* or its catalytically dead mutant S288A (Lemberg et al., 2005) in this background from an ectopic chromosomal locus (strains BS50 and BS51, respectively, see Table EV2). We then conducted quantitative proteomics analysis of membrane fractions of both of these strains employing SILAC labelling and SDS PAGE pre-fractionation (GeLC experiment), where we were seeking peptides belonging to proteins migrating at a lower-than-expected apparent molecular weight selectively in the YqgP-expressing strain relative to the S288A mutant expressing one. This analysis yielded the high-affinity magnesium transporter MgtE as the top candidate substrate of YqgP (Fig. 1A, B, C, Dataset S1, PRIDE dataset PXD014578).

In a complementary approach, we used affinity co-immunopurification and label-free quantitative proteomics to identify proteins associating with YqgP (Fig. 1D,E). We ectopically expressed either a functional YqgP-sfGFP fusion or catalytically dead YqgP.S288A-sfGFP fusion as the sole copy of YqgP in the cell. We solubilised the isolated membranes using the NP-40 detergent, isolated YqgP-sfGFP by anti-GFP affinity pull-down, and analysed the co-isolated proteins by MS-based proteomics. Using this approach we identified several high-confidence interactor candidates including the membrane-anchored protease FtsH, and ATPase subunits A, D, F, and G (Fig.1E, Dataset S2, PRIDE dataset PXD014566). The only high-confidence overlap between the two proteomics datasets was the high-affinity magnesium transporter MgtE (Fig. 1C, E), promoting it to the highest likelihood candidate substrate. MgtE is the main magnesium transporter in *B. subtilis*, it is essential for magnesium homeostasis (Wakeman, Goodson et al., 2014), and we thus examined the possible functional relationship between YqgP and MgtE.

YqgP cleaves the high-affinity magnesium transporter MgtE between its first and second transmembrane helices.

To validate the results of quantitative proteomics and gain more insight into the role of YqgP in the physiology of *B. subtilis*, we first examined the proteolytic status of endogenous MgtE by immunoblotting using an in-house generated antibody recognizing its N-terminal cytosolic domain (MgtE 1-275). Both endogenous and ectopically expressed YqgP cleaved endogenous MgtE, yielding distinct cleavage

products (Fig. 2A). The cleavage was abrogated by a rhomboid-specific peptidyl ketoamide inhibitor (Ticha et al., 2017b) at low nanomolar levels, confirming that it was a rhomboid-specific event (Fig. 2B). Judging by the apparent molecular size of the N-terminal cleavage fragment compared to the *in vitro* translated reference fragments (Lemberg & Martoglio, 2003) of MgtE, we concluded that the cleavage by YqgP occurred within the extracytoplasmic loop between TMH1 and TMH2 of MgtE. This region is close to or within the periplasmic end of TMH2 (Fig. 2C,D), which is consistent with the topology and mechanism of a rhomboid protease (Strisovsky, 2013, Strisovsky, 2016). MgtE transporters function as homodimers with a cytosolic amino-terminal cystathionine-beta-synthase (CBS) domain that senses intracellular Mg^{2+} and a carboxy-terminal five-transmembrane helical pore for Mg^{2+} import (Hattori, Tanaka et al., 2007, Takeda, Hattori et al., 2014). Cleavage of MgtE between TMH1 and TMH2 by YqgP would thus likely inactivate the transporter.

Manganese excess activates cleavage of MgtE by YqgP under conditions of magnesium starvation.

It is known that MgtE homologs can also transport Mn^{2+} (Takeda et al., 2014) under low Mg^{2+} concentration (Mg starvation), and that high concentrations of Mn^{2+} are toxic for bacteria, mainly because Mn^{2+} can mis-metalate Mg^{2+} binding sites (Chandrangsu, Rensing et al., 2017, Hohle & O'Brian, 2014). We therefore hypothesised that YqgP could be involved in the regulation of MgtE under Mn^{2+} stress, and we tested the influence of manganese on the *in vivo* activity of YqgP. While at high extracellular magnesium concentration (1 mM) the effect of 100 µM $MnCl_2$ was not detectable, at low extracellular magnesium concentration (0.01 mM), when endogenous MgtE is upregulated, a shift from 1 to 100 µM $MnCl_2$ activated the YqgP-dependent cleavage of MgtE four-fold (Fig. 3A).

We next tested the importance of YqgP for growth in manganese-stress conditions. During growth in minimal medium at low magnesium concentration (10 µM), adding 75 µM manganese at mid-log phase caused growth arrest and lysis of wild-type *B. subtilis*, which was more severe upon deletion of *yqgP*, and fully rescued by ectopic expression of YqgP (Fig. 3B). This effect was highly reproducible. Interestingly, all strains resumed growth at later time-points, indicating the possible existence of an adaptive mechanism. This did not appear to involve the second *B. subtilis* rhomboid YdcA, because deletion of *ydcA* did not modify the phenotypic behaviour of *B. subtilis* regardless of the presence or absence of *yqgP* (Fig. EV1). Regardless, the *yqgP* deficient strains showed a marked growth disadvantage even in the latter phase of the growth curve, indicating that the role of YqgP for the fitness of *B. subtilis* in these conditions is significant. The toxic effect of 50 µM manganese sulphate was fully prevented by growth in high (5 mM) magnesium sulphate (Fig. 3C), which was consistent with a similar effect on MgtE cleavage by YqgP (Fig. 3A). In addition, increasing concentrations of YqgP inhibitor had a dose-dependent effect on the protective role of YqgP both when it was overexpressed or present at endogenous levels (Fig. 3D), indicating that the pronounced protective effect of ectopically expressed YqgP was a result of its overexpression. Finally, the phenotype manifested also during steady-state growth in rich LB medium supplemented with 75 µM manganese sulphate, where *yqgP* deficient *B. subtilis* ectopically expressing MgtE was delayed in growth compared to the same strain ectopically expressing YqgP (Fig. 3E). Intriguingly, ectopic expression of a proteolytically inactive S288A mutant of

YqgP also rescued the phenotype of the parent strain, which was initially surprising and suggested that YqgP may have a role independent of its protease activity. Together, these results reveal that a physiological function of YqgP is the protection from manganese stress by contributing to the degradation of the main magnesium transporter MgtE in *B. subtilis* (Wakeman et al., 2014).

The N-terminal extramembrane intracellular domain of YqgP mediates the manganese induced activation of MgtE cleavage.

To gain insights into the mechanism of MgtE cleavage by YqgP, we turned our attention to the two extramembrane domains of YqgP, the N-terminal cytosolic domain (NTD), and the C-terminal extracellular domain (CTD). While CTD consists of a predicted tandem array of three tetratricopeptide repeat (TPR) motifs (Zeytuni & Zarivach, 2012) and is thus likely to interact with protein or peptide ligands, NTD did not show significant sequence similarity to any protein of known structure, providing no indication of its molecular function. We first investigated the importance of NTD and CTD for YqgP activity. We created *B. subtilis* strains producing YqgP lacking NTD or CTD and measured the *in vivo* 'specific activities' of these YqgP variants (i.e. substrate conversions at steady state normalised to enzyme expression level) in cleaving MgtE and the heterologous model substrate TatA (Ticha et al., 2017a). While deletion of CTD (Δ388-507, ΔCTD) seems to have led to about 4-6 fold increase in YqgP activity for both substrates, deletion of NTD (Δ1-178, ΔNTD) reduced YqgP activity for MgtE 3-fold, but only marginally so for the model TatA substrate (Fig. 4A). In addition, deletion of NTD also compromised the activation effect of Mn^{2+} on MgtE cleavage by YqgP, while the deletion of CTD had no effect (Fig. 4B, C, D). Taken together, the existence of such a specific regulatory mechanism involving YqgP NTD further substantiated the idea that the manganese-regulated cleavage of MgtE by YqgP is physiologically significant.

The N-terminal intracellular domain of YqgP binds divalent cations and mediates their sensing by YqgP.

Since divalent cations often have related coordination properties, we wondered whether the manganese activation effect on YqgP transmitted via $YqgP_{NTD}$ may be a reflection of a more general divalent metal cation effect. We thus inspected a range of divalent metal cations for activation of MgtE cleavage by YqgP *in vivo* and found that high concentrations (100 µM) of Mn^{2+}, Zn^{2+}, and Co^{2+} activated MgtE cleavage by YqgP significantly in minimal medium with limiting Mg^{2+} conditions (10 µM), while Ca^{2+} and Ni^{2+} did not have appreciable effect (Fig. 5A). Isothermal titration calorimetry (ITC) measurements showed that Mn^{2+}, Zn^{2+}, Co^{2+} and Ni^{2+} bound $YqgP_{NTD}$ with dissociation constants in the submillimolar to millimolar range and approximately 1:1 stoichiometry, while neither Mg^{2+} nor Ca^{2+} showed any calorimetric effect up to 1.5 mM concentrations tested (Fig. 5B). Interestingly, the transition metals Co^{2+}, Mn^{2+} and Zn^{2+} can all be transported by MgtE, albeit with lower efficiency than Mg^{2+}. Overexpression of MgtE confers high sensitivity to Zn^{2+} in *Bacillus firmus* OF4 (Smith, Thompson et al., 1995), and Ni^{2+} cannot be transported but inhibits MgtE transport (Smith et al., 1995). Here we observed that *yqgP*

deficiency made *B. subtilis* more sensitive to Mn^{2+} and Zn^{2+} but not to Co^{2+} or Ni^{2+} stress at low Mg^{2+}, and, consistently, re-expression of YqgP rescued this effect to above wild type levels (Fig. 5C). Collectively, this indicates that $YqgP_{NTD}$ may act as a sensor for a subset of metal cations in order to prevent cation toxicity due to MgtE permissiveness under low Mg^{2+} conditions.

To understand the basis of the activation of YqgP by divalent cations, we undertook structural analysis of $YqgP_{NTD}$. Domains homologous to $YqgP_{NTD}$ occur only in YqgP orthologs among the soil-dwelling bacteria of the genus *Bacilli* (http://www.ebi.ac.uk/interpro/protein/P54493/similar-proteins), bear little sequence similarity to other proteins and show no hits at ProSite. Thus, to understand the structural basis of manganese-induced activation of YqgP, we first determined the 3D structure of NTD by solution NMR (Fig. 6A, statistics in Table EV3). The overall structural ensemble of $YqgP_{NTD}$ (YqgP 1-177) consists of both α- and β- structural motifs, comprising six β-strand sequences at positions 21-24 (β1), 30-34 (β2), 43-48 (β3), 80-88 (β4), 104-106 (β5) and 109-117 (β6), and five α-helices at positions 3-16 (α1), 54-75 (α2), 96-100 (α3), 121-128 (α4) and 146-168 (α5). The central mixed β-sheet motif is packed in a twisted plane and surrounded by helices α1-4. Notably, negatively charged side-chains are clustered into several solvent-exposed regions (Fig. 6B). Structural similarity search using Dali server (Holm & Laakso, 2016) identified a number of structurally related proteins, with no conserved functional role but enriched in DNA/RNA polymerases and modifying enzymes (Table EV4). In particular, the β-stranded central part of $YqgP_{NTD}$ adopts a fold characteristic for the type II restriction endonucleases (InterPro and Pfam codes IPR018573 and PF09491, respectively) (Kachalova, Rogulin et al., 2008).

Since Mn^{2+} but not Zn^{2+} provided reliable data in ITC titrations, we focused on Mn^{2+} in further structural analyses aiming to understand the structural aspect of divalent cation mediated activation via $YqgP_{NTD}$. NMR titrations of $YqgP_{NTD}$ by the paramagnetic Mn^{2+} induced peak intensity reductions and chemical shift perturbations in a number of residues along NTD. N-H chemical shifts could be caused by direct interaction with the metal ion, or by secondary structural effects, while peak intensity could be also diminished by the paramagnetic effect of Mn^{2+}. The most strongly affected regions were amino acids 27-30, 35-37, 48-63, and 89-98 of the 177 amino acid NTD (Fig. 6C, D, full titration data are available in Dataset S3). To identify potential residues involved in binding Mn^{2+}, we next focused on amino acids typically involved in chelating divalent cations (D, E, H) from each of the regions affected in NMR titrations. We analysed the effect of alanine substitutions in D29, D37, H49, D50, D52, D60 and E90 on the ability of these mutants to elicit Mn-induced activation of YqgP *in vivo* (Fig. 6E). This analysis indicated that the hotspot in which mutations interfered with Mn-induced activation of YqgP was centered around D50, and D50A mutation abrogated Mn-activation completely. Interestingly, several of the residues affected in NMR titrations clustered in one apex of NTD structure, with D50 being central in this region (Fig. 6F), representing the likely Mn-binding region in NTD (Fig. 6G). In summary, deletion of $YqgP_{NTD}$ or mutation of chelating residues from the identified Mn^{2+} binding surface abrogated the Mn^{2+} induced activation of MgtE cleavage by YqgP *in vivo* showing that NTD is the Mn^{2+} sensing device of YqgP.

YqgP works in tandem with the AAA protease FtsH to degrade MgtE in response to the environmental divalent cation concentrations.

To explore more precisely MgtE proteolysis *in vivo*, we decided to construct a *yqgP* mutant strain complemented with a ectopic mutated version of *yqgP* that encodes a catalytically dead YqgP.S288A protein. Surprisingly, we observed that the production of this inactive YqgP.S288A in the *yqgP* deficient background resulted in the production of an alternative band with lower mobility on SDS PAGE, i.e. a higher molecular weight, than that of the cleavage product generated by YqgP (Fig. 7A). Since the S288A variant of YqgP lacks the catalytic serine nucleophile and is thus proteolytically completely inactive (Baker & Urban, 2012, Lemberg et al., 2005, Wang, Zhang et al., 2006), these results suggested that another proteolytic enzyme may participate in MgtE processing under these conditions. Co-immunoprecipitation experiments (Fig. 1E) showed that YqgP interacts with the ATP-dependent, membrane-bound, processive zinc metalloprotease FtsH, and we thus hypothesised that FtsH is the responsible enzyme. Indeed, deletion of *ftsH* led to the disappearance of this alternative cleavage product (Fig. 7A). Experiment with reference fragments of MgtE indicated that the products of proteolysis by FtsH end between residues 340 and 355, which corresponds to the cytosolic side of TMH 3 of MgtE (Fig. 7B, C). This is consistent with the known mechanism of FtsH, which dislocates transmembrane proteins from the membrane and processively proteolyses them (Akiyama, 2003, Chiba, Akiyama et al., 2002, Langklotz, Baumann et al., 2012, Yang, Guo et al., 2018). Interestingly, $YqgP_{NTD}$ is crucial also for this adaptor function of YqgP (Fig. 7D).

To investigate the relationship between YqgP and FtsH in more detail, we performed chase experiments in which translation was inhibited by tetracycline at time-point zero ('translation shut-off'), which enabled us to observe the stability of the MgtE cleavage products generated by YqgP or FtsH (Fig. 7E). This analysis revealed that during growth with ongoing translation, the cleavage products formed by YqgP (lanes 9-12 and 33-36) or FtsH (lanes 17-20) accumulated. In contrast, no cleavage products accumulated after stopping translation. Instead, the YqgP-induced N-terminal cleavage product of MgtE was further degraded by FtsH (compare lanes 13-16 to lanes 37-40) but also by other, unknown proteases (lanes 37-40) (for quantification see Fig. EV2). In contrast, the alternative N-terminal cleavage product of MgtE generated by FtsH in the presence of YqgP.S288A was stable and was not further degraded over the chase period (lanes 21-24). In the absence of YqgP, FtsH was not able to act on MgtE at all (lanes 1-8). Intriguingly, this was mimicked by the expression of active YqgP in the presence of saturating concentrations of its specific peptidyl ketoamide inhibitor (compare Fig. 2B and Fig. 7), which binds into the active site of rhomboid and plugs it (Ticha et al., 2017b). The above data mean that YqgP fulfils dual role, both of a protease and a substrate adaptor for FtsH, and that the formation of MgtE species susceptible to proteolysis by either protease requires ongoing translation. The adaptor function does not require YqgP activity, but it does require an unobstructed active site, and is thus consistent with a 'pseudoprotease' function of YqgP. NTD is essential also for the pseudoprotease role of YqgP (Fig. 7D), and without YqgP protease activity, FtsH cannot degrade MgtE fully (Fig. 7), indicating that YqgP and FtsH closely cooperate in maintaining the proteostasis of MgtE.

In conclusion, the above data yield a model where YqgP responds to divalent metal cation concentrations and cleaves MgtE under low Mg and high Mn^{2+}or Zn^{2+} concentrations, facilitating the dislocation of MgtE cleavage products from the membrane and their degradation by FtsH (Fig. 8). Hence, YgqP serves as a fail-safe mechanism to limit toxicity of metal cations quickly and temporarily. It has a dual role here, because besides cleaving its substrate MgtE, it also presents the cleavage products of MgtE to FtsH, which can then degrade them processively. Interestingly, while the proteolytically inactive YqgP.S288A variant allows formation of an alternative cleavage product by FtsH (i.e. it can still present MgtE to FtsH), this band is not observed when wild-type YqgP is expressed in the presence of a rhomboid protease inhibitor (Fig. 2), suggesting that the active site of YqgP needs to be unobstructed to present MgtE (fragments) to FtsH. This phenomenon is remarkably similar to the roles of the rhomboid-like proteins Derlins, the p97 AAA ATPase and proteasome in ER-associated degradation (ERAD) (Ticha et al., 2018), and may constitute a prokaryotic precursor of ERAD. Furthermore, it provides a plausible mechanistic explanation of how the rhomboid-like pseudoproteases might have evolved from rhomboid proteases (Adrain & Freeman, 2012). In summary, we provide evidence for the role of a bacterial rhomboid protease in the regulation of key membrane protein proteostasis under environmentally restricted conditions in a process similar to the eukaryotic ERAD.

Discussion

Metal ions are essential for life because they serve as cofactors in various cellular processes, but their excess poses a lethal threat. Hence, various mechanisms exist to ensure metal ion homeostasis (Chandrangsu et al., 2017). Soil bacteria, such as *B. subtilis*, frequently encounter fluctuations in metal ion availability and need to possess sophisticated regulatory and feedback mechanisms to specifically respond to a limitation or excess of various metal ions (Chandrangsu et al., 2017). Magnesium (Mg^{2+}) represents the most abundant divalent cation in cells. It is maintained at high free intracellular levels (0.5 to 1 mM) relative to other cations (Froschauer, Kolisek et al., 2004), and it plays vital roles in membrane integrity, charge neutralization of nucleic acids and nucleotides, ribosome structure, and activities of many metabolic enzymes. To take up environmental Mg^{2+}, bacteria usually encode several magnesium transporters of the CorA, MgtA and MgtE families (Groisman, Hollands et al., 2013). Homologs of MgtE are universally conserved up to humans (SLC41 family), indicating their importance. In *B. subtilis*, MgtE is the main magnesium transporter (Wakeman et al., 2014) and is tightly regulated at several levels (Groisman et al., 2013). First, low levels of environmental magnesium trigger upregulation of *mgtE* transcription via a riboswitch mechanism (Dann, Wakeman et al., 2007). Second, the Mg-dependent gating mechanism by the N-terminal cytosolic domain of MgtE ensures metal ion selectivity (Hattori, Iwase et al., 2009). Third, ATP further regulates MgtE (Tomita, Zhang et al., 2017), presumably because the cell needs to maintain a balance in magnesium and ATP concentrations as the two are functionally coupled in a number of physiological processes (Pontes, Sevostyanova et al., 2015, Pontes, Yeom et al., 2016).

Structural principles of ion channel selectivity of MgtE have been studied in detail (Maruyama, Imai et al., 2018, Takeda et al., 2014). At low magnesium concentrations (when Mg^{2+} is not a competitive

inhibitor), MgtE can also transport Co^{2+}, Mn^{2+} and Zn^{2+} to some extent, while Ni^{2+} inhibits MgtE (Smith et al., 1995, Takeda et al., 2014). This implies that at low magnesium concentrations and high concentrations of Co^{2+}, Mn^{2+}, Zn^{2+} or Ni^{2+}, other regulatory mechanisms may be required to maintain the biological function of Mg and/or minimize toxicity of other divalent cations. Indeed, previous observations in other organisms indicate that controlling magnesium transporter selectivity with respect to transition metal ions is a widespread theme. In yeast, excess magnesium prevents manganese toxicity (Blackwell, Tobin et al., 1997), implying that MgtE may transport manganese when magnesium does not compete with it. In *Bradyrhizobium japonicum*, a null mutant in the manganese transporter *mntH* encoding gene does not grow in Mn-limiting medium, but suppressor strains can be selected. One class of suppressors are gain-of function mutants of *mgtE* (mutations affecting gating mechanism) (Hohle & O'Brian, 2014). In this system, when the intracellular Mg^{2+} levels are low, the MgtE mutants transport Mn^{2+} with high affinity, which is not the case at high Mg^{2+} and for the wild-type strain. This is particularly relevant in soil, the natural habitat of *B. subtilis*, where there is a negative correlation between the bio-availability of magnesium and manganese ions. As soil pH decreases, manganese becomes more available and magnesium less available (Gransee & Fuhrs, 2013, Thomas & Sprenger, 2008). Hence, in the commonly found acidic soils, *B. subtilis* has to cope with stress conditions of low magnesium and high manganese, and the regulation of MgtE by YqgP that we describe is important for fitness in similar conditions (Fig. 3, Fig. 5).

The role of YqgP seems to be important during acute changes in cation concentrations. Several hours after manganese shock, even *yqgP* deficient *B. subtilis* starts growing again, albeit at lower rate than wild type or a strain overexpressing YqgP (Fig. 3B). This 'adaptation' may be based on the described genetic mechanisms related to the regulation of Mn metabolism. It is known that *B. subtilis* responds to Mn^{2+} excess in two ways. First, the MntR regulator senses excess intracellular Mn^{2+}. It then represses the transcription of two Mn^{2+}-uptake systems (MntABCD and MntH) and activates the expression of efflux systems (MneP and MneS) (Huang, Shin et al., 2017, Que & Helmann, 2000). In addition, the two manganese efflux pumps, MneP and MneS, are also activated via Mn^{2+}-responsive riboswitches (Dambach, Sandoval et al., 2015, Price, Gaballa et al., 2015). These actions take place at the transcriptional level, hence take some time to efficiently allow the cell to recover from an initial Mn^{2+} shock. As intracellular Mn^{2+} returns to physiological levels, cells resume growth and YqgP may thus lose its importance for fitness in the later stages of adaptation to these toxic conditions.

We find that in *B. subtilis*, rhomboid protease YqgP is required to cleave and inactivate species of MgtE formed at low environmental Mg^{2+} and high Mn^{2+} or Zn^{2+}, which protects *B. subtilis* from toxic Mn^{2+} or Zn^{2+} stress. The identity of the MgtE species susceptible to YqgP cleavage in these conditions is unknown. Since their generation requires ongoing translation (Fig. 7), we hypothesise that these may be mis-metallated forms at some of the metal binding sites throughout the MgtE molecule (Foster, Osman et al., 2014, Maruyama et al., 2018, Takeda et al., 2014) that can form during translation and folding, and which could interfere with the proper channel function of MgtE. YqgP is thus providing a quality control check for properly functioning MgtE. Cleavage of MgtE by YqgP is more pronounced at low Mg^{2+} and high Mn^{2+} or Zn^{2+} concentrations. This may, in part, result from the higher concentration

of susceptible forms of MgtE in these conditions, which is consistent with our translation shut-off chase data (Fig. 7). However, metal sensing and cleavage activation are also directly provided by the intracellular N-terminal extramembrane domain of YqgP. This domain binds divalent metal cations at milimolar or submilimolar affinities and approximately 1:1 stoichiometry (Fig. 5) along a negatively charged surface of the molecule (Fig. 6). How exactly in structural and mechanistic terms Mn^{2+} or Zn^{2+} binding to $YqgP_{NTD}$ activates MgtE cleavage is not clear at the moment. It could be through direct interaction with the membrane embedded MgtE, or by influencing the speed of diffusion of YqgP in the membrane, as recently suggested for several other rhomboid proteases (Kreutzberger, Ji et al., 2019). Interestingly, in addition to the influence on YqgP activity, $YqgP_{NTD}$ is also required for the coupling of YqgP activity to FtsH, the ATP dependent processive protease that associates with YqgP. More generally, there is emerging evidence from this work and the accompanying back-to-back study by Liu et al. that rhomboid proteases can participate in membrane protein quality control in bacteria, where they are very widespread. It remains to be elucidated which structural and dynamic aspects of transmembrane domains rhomboid proteases can recognise in their substrates, to serve as 'detectors' of misfolded or orphan protein species and adaptors of downstream degradative proteases. The variability in substrate specificities and various extramembrane domains of rhomboids may be the basis of their diverse functions in bacteria.

The specialised role that YqgP plays in the quality control of an essential transporter protein may seem unexpected, but conceptually, transporters are indeed regulated at a number of levels and very finely using specialised mechanisms. For example, amino acid permeases of the major facilitator superfamily in yeast are down-regulated from the cell surface under high nutrient conditions by an increase in ubiquitin E3 ligase activity of Rsp5, which initiates their trafficking from the plasma membrane to multi-vesicular bodies (Lauwers, Erpapazoglou et al., 2010). Other member of the same transporter family in yeast requires a specialised membrane chaperone Shr3, which assists the folding of the transporters and prevents their efficient degradation by ERAD (Kota, Gilstring et al., 2007).

Intriguingly, in addition to the role of YqgP in regulating MgtE, we suspect it can be engaged in other functions. In co-immunoprecipitation experiments with YqgP we identified a number of ATP synthase subunits with high confidence and abundance by mass spectrometry (Fig. 1E). Since we have not identified any of these subunits in our substrate screen (Fig. 1C), we did not follow them up. However, this finding raises the possibility that further to its role in MgtE proteostasis, YqgP may be involved in the relationship between magnesium homeostasis and ATP synthesis, which would be consistent with a growing body of evidence supporting a regulatory relationship between magnesium and ATP levels. First, it was reported that a magnesium transporter-defective strain of *B. subtilis* can be rescued by inactivation of ATP synthase encoding genes (Wakeman et al., 2014), suggesting that lowering ATP levels allows the cell to cope with magnesium deprivation. This is consistent with findings from Groisman and colleagues in *Salmonella enterica,* where low levels of Mg^{2+} in the cell lead to ATP synthase inhibition and reduced ribosome biogenesis. As a result, the pool of magnesium that is not ATP-bound anymore can be employed to support low levels of ribosome assembly, a priority process that requires large amounts of magnesium (Pontes et al., 2016). Finally, in the same bacterium, ATP

synthase activity can be directly inhibited by the MgtC protein, whose expression is upregulated during magnesium deprivation (Lee, Pontes et al., 2013). Strikingly, MgtC downregulation in the membrane of *S. enterica* is somewhat reminiscent of how MgtE is controlled by YqgP and FtsH in *B. subtilis*. MgtR, a small hydrophobic peptide, is thought to destabilize MgtC prior to degradation by FtsH (Alix & Blanc-Potard, 2008). Hence, during magnesium deprivation and cation toxicity in *B. subtilis*, it is tempting to speculate that YqgP may also inhibit the activity of ATP synthase to preserve the free Mg^{2+} pool for translation purposes while assisting FtsH for the degradation of MgtE.

A number of native and model rhomboid substrates are single-pass membrane proteins (Johnson, Brezinova et al., 2017, Riestra, Gandhi et al., 2015, Saita, Nolte et al., 2017, Strisovsky et al., 2009, Urban, Lee et al., 2002a). The cleavage of the transmembrane core of MgtE by YqgP described here adds to growing evidence that rhomboid proteases do engage in cleaving polytopic membrane proteins under some conditions (Erez & Bibi, 2009), although the features of transmembrane domains that they recognise still need to be properly structurally understood. Our results show that degradation of the susceptible MgtE species requires both YqgP and FtsH, with YqgP cleavage of the transmembrane core of MgtE being required for full degradation by FtsH. In addition, YqgP presents MgtE to FtsH, for which YqgP activity is dispensable, but its unoccupied active site is essential, thus behaving partially also as a pseudoprotease. In summary, here we uncover an ancestral membrane-associated degradation mechanism including a substrate receptor/protease and processive degradative dislocase and protease. This is conceptually remarkably similar to the process of ERAD in which rhomboid pseudoproteases Derlins co-operate with ubiquitin E3 ligases such as Hrd1, the AAA ATPase p97/VCP, and the proteasome, in dislocating transmembrane proteins from the membranes of endoplasmic reticulum and degrading them (Neal, Jaeger et al., 2018, Ticha et al., 2018). Our work emphasises that even the active rhomboid proteases can have pseudoprotease functions, and suggests that the balance between their protease and pseudoprotease roles may vary. In this context it is interesting to note that the second rhomboid protease encoded in *B. subtilis* genome, YdcA, lacks any detectable protease activity despite having all sequence hallmarks of an active rhomboid protease (Lemberg et al., 2005, Urban et al., 2002b). Similarly, mammalian RHBDL1 and RHBDL3 failed to cleave any known model substrates of rhomboid proteases (Johnson et al., 2017), and it ought to be borne in mind that in these or other cases, the pseudoprotease roles may be key to the physiological functions of these rhomboid proteins. More generally, other intramembrane proteases may have physiological roles that do not depend on their catalytic activity, such as the role of SPP in membrane protein dislocation from the ER (Loureiro, Lilley et al., 2006).

A number of bacteria encode YqgP homologs (http://www.ebi.ac.uk/interpro/protein/P54493/similar-proteins), and the function of YqgP in magnesium homeostasis control that we describe here might be conserved more widely, because MgtE homologs are also very widespread (https://www.ebi.ac.uk/interpro/protein/O34442/similar-proteins), although it is possible that YqgP homologs will have also other functions. This is indeed suggested by two recent reports. First, it was shown that the *Staphylococcus aureus* YqgP homolog Rbd is synthetic lethal with the nucleoid occlusion protein Noc and is involved in the initiation of DNA replication (Pang, Wang et

al., 2017). Second, in a recent archived report, the same protein (although termed ActH therein) has been identified as a regulatory component of a complex with peptidoglycan amidase LytH and is required for LytH activation (Do, Schaefer et al., 2019). In neither case was it established whether the described function of *S. aureus* YqgP homolog Rbd/ActH depends on its protease activity. It is intriguing to speculate whether these functions will rely on a similar protease or pseudoprotease roles within a degradation machinery such as the one we describe here. It is fully conceivable that YqgP homologs can fulfil a range of functions in diverse bacteria depending on the client protein but using analogous general mechanism as we outline here.

In a generalised summary, this work implicates bacterial rhomboid proteases in the quality control of polytopic membrane proteins in cooperation with other processive proteases, which could be regarded as an evolutionary functional ancestor of the eukaryotic ER-associated degradation (ERAD). In our system, rhomboid plays the role of a protease that facilitates dislocation of the membrane protein by the ATP dependent AAA protease FtsH, but also the role of a substrate adaptor handing over the substrates to FtsH. It is conceivable that with a more efficient and specialised dislocation machinery in higher organisms, the requirement for protease activity of the rhomboid like protein (Derlin) could become less important, while the adaptor function becoming key. This logic would explain the particularly high occurrence of pseudoproteases in the rhomboid superfamily, which includes Derlins that are well-known essential components of the eukaryotic substrate recognition and dislocation machinery in ERAD (Neal et al., 2018).

Author contributions

KS, TD, BC and Jakub Began (JB) designed research, JB, BC, TD, VV, RH, PS, MK, JD, MB, YC and PR performed research and analysed data, and KS and JB wrote the manuscript with contributions from TD, Jana Brezinova, BC, VV, AG, RH, PS, YC and MK.

Conflict of interest

The authors declare no conflict of interest with the content of this article.

Acknowledgements

We thank Libor Krásný for critical reading of the manuscript. TD thanks David Rudner and Eric Cascales for their support and guidance. We thank Radko Souček for amino acid analysis, and Stanislava Matějková for elemental analysis. KS and MK acknowledge financial support from the Ministry of Education, Youth and Sports of the Czech Republic (project no. LO1302), European Regional Development Fund (ERDF/ESF) project No. CZ.02.1.01/0.0/0.0/16_019/0000729, Gilead Sciences & IOCB Research Centre, and the National Subvention for Development of Research Organisations (RVO: 61388963) to the Institute of Organic Chemistry and Biochemistry. This work was also supported

by the CNRS, Aix-Marseille University (AMU) and a Marie-Curie International Reintegration Grant (PIRG08-GA-2010-276750) to TD. BC and JD were supported by fellowships from AMU, and JD also by the Fondation pour la Recherche Médicale (FRM, FDT20160435133). Proteomics experiments were partly supported by the Agence Nationale de la Recherche (ProFI grant ANR-10-INBS-08-01). JB and Jana Březinová acknowledge support from the Grant Agency of Charles University (GA UK) in Prague (PhD grant project no. 170214). Authors state no conflicts of interest with the content of this article.

Supplementary information is available at The EMBO Journal Online.

References

Adrain C, Freeman M (2012) New lives for old: evolution of pseudoenzyme function illustrated by iRhoms. *Nat Rev Mol Cell Biol* 13: 489-98

Akiyama Y (2003) Reconstitution of Membrane Proteolysis by FtsH. *J Biol Chem* 278: 18146-18153

Alix E, Blanc-Potard AB (2008) Peptide-assisted degradation of the Salmonella MgtC virulence factor. *EMBO J* 27: 546-57

Baker RP, Urban S (2012) Architectural and thermodynamic principles underlying intramembrane protease function. *Nat Chem Biol* 8: 759-68

Blackwell KJ, Tobin JM, Avery SV (1997) Manganese uptake and toxicity in magnesium-supplemented and unsupplemented Saccharomyces cerevisiae. *Appl Microbiol Biotechnol* 47: 180-184

Campo N, Rudner DZ (2006) A Branched Pathway Governing the Activation of a Developmental Transcription Factor by Regulated Intramembrane Proteolysis. *Mol Cell* 23: 25-35

Casabona MG, Vandenbrouck Y, Attree I, Coute Y (2013) Proteomic characterization of *Pseudomonas aeruginosa* PAO1 inner membrane. *Proteomics* 13: 2419-23

Chandrangsu P, Rensing C, Helmann JD (2017) Metal homeostasis and resistance in bacteria. *Nat Rev Microbiol* 15: 338-350

Chiba S, Akiyama Y, Ito K (2002) Membrane Protein Degradation by FtsH Can Be Initiated from Either End. *J Bacteriol* 184: 4775-4782

Costa MI, Cerletti M, Paggi RA, Trotschel C, De Castro RE, Poetsch A, Gimenez MI (2018) Haloferax volcanii Proteome Response to Deletion of a Rhomboid Protease Gene. *J Proteome Res* 17: 961-977

Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26: 1367-1372

Dambach M, Sandoval M, Updegrove TB, Anantharaman V, Aravind L, Waters LS, Storz G (2015) The ubiquitous yybP-ykoY riboswitch is a manganese-responsive regulatory element. *Mol Cell* 57: 1099-1109

Dann CE, 3rd, Wakeman CA, Sieling CL, Baker SC, Irnov I, Winkler WC (2007) Structure and mechanism of a metal-sensing regulatory RNA. *Cell* 130: 878-92

Deutsch EW, Csordas A, Sun Z, Jarnuczak A, Perez-Riverol Y, Ternent T, Campbell DS, Bernal-Llinares M, Okuda S, Kawano S, Moritz RL, Carver JJ, Wang MX, Ishihama Y, Bandeira N, Hermjakob H, Vizcaino JA (2017) The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition. *Nucleic Acids Res* 45: D1100-D1106

Do T, Schaefer K, Santiago AG, Coe KA, Fernandes PB, Kahne D, Pinho MG, Walker S (2019) The cell cycle in *Staphylococcus aureus* is regulated by an amidase that controls peptidoglycan synthesis. *BioRxiv* 634089; doi: https://doi.org/10.1101/634089

Doan T, Morlot C, Meisner J, Serrano M, Henriques AO, Moran CP, Jr., Rudner DZ (2009) Novel secretion apparatus maintains spore integrity and developmental gene expression in *Bacillus subtilis*. *PLoS Genet* 5: e1000566

Erez E, Bibi E (2009) Cleavage of a multispanning membrane protein by an intramembrane serine protease. *Biochemistry* 48: 12314-22

Foster AW, Osman D, Robinson NJ (2014) Metal preferences and metallation. *J Biol Chem* 289: 28095-103

Freeman M (2014) The rhomboid-like superfamily: molecular mechanisms and biological roles. *Annu Rev Cell Dev Biol* 30: 235-54

Froschauer EM, Kolisek M, Dieterich F, Schweigel M, Schweyen RJ (2004) Fluorescence measurements of free [Mg2+] by use of mag-fura 2 in *Salmonella enterica*. *FEMS Microbiol Lett* 237: 49-55

Gibson DG (2011) Enzymatic assembly of overlapping DNA fragments. *Methods Enzymol* 498: 349-61

Gransee A, Fuhrs H (2013) Magnesium mobility in soils as a challenge for soil and plant analysis, magnesium fertilization and root uptake under adverse growth conditions. *Plant Soil* 368: 5-21

Groisman EA, Hollands K, Kriner MA, Lee EJ, Park SY, Pontes MH (2013) Bacterial Mg^{2+} homeostasis, transport, and virulence. *Annu Rev Genet* 47: 625-46

Harjes E, Harjes S, Wohlgemuth S, Muller KH, Krieger E, Herrmann C, Bayer P (2006) GTP-Ras disrupts the intramolecular complex of C1 and RA domains of Nore1. *Structure* 14: 881-8

Harwood CR, Cutting SM (1990) *Molecular biological methods for Bacillus*. Wiley, Chichester ; New York

Hattori M, Iwase N, Furuya N, Tanaka Y, Tsukazaki T, Ishitani R, Maguire ME, Ito K, Maturana A, Nureki O (2009) Mg2+-dependent gating of bacterial MgtE channel underlies Mg^{2+} homeostasis. *EMBO J* 28: 3602-3612

Hattori M, Tanaka Y, Fukai S, Ishitani R, Nureki O (2007) Crystal structure of the MgtE Mg^{2+} transporter. *Nature* 448: 1072-5

Herrmann T, Guntert P, Wuthrich K (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol* 319: 209-27

Hohle TH, O'Brian MR (2014) Magnesium-dependent processes are targets of bacterial manganese toxicity. *Mol Microbiol* 93: 736-47

Holm L, Laakso LM (2016) Dali server update. *Nucleic Acids Res* 44: W351-5

Huang X, Shin JH, Pinochet-Barros A, Su TT, Helmann JD (2017) *Bacillus subtilis* MntR coordinates the transcriptional regulation of manganese uptake and efflux systems. *Mol Microbiol* 103: 253-268

Ivankov DN, Bogatyreva NS, Honigschmid P, Dislich B, Hogl S, Kuhn PH, Frishman D, Lichtenthaler SF (2013) QARIP: a web server for quantitative proteomic analysis of regulated intramembrane proteolysis. *Nucleic Acids Res* 41: W459-64

Jaquinod M, Trauchessec M, Huillet C, Louwagie M, Lebert D, Picard G, Adrait A, Dupuis A, Garin J, Brun V, Bruley C (2012) Mass spectrometry-based absolute protein quantification: PSAQ strategy makes use of "noncanonical" proteotypic peptides. *Proteomics* 12: 1217-21

Johnson N, Brezinova 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: 7283

Kachalova GS, Rogulin EA, Yunusova AK, Artyukh RI, Perevyazova TA, Matvienko NI, Zheleznaya LA, Bartunik HD (2008) Structural analysis of the heterodimeric type IIS restriction endonuclease R.BspD6I acting as a complex between a monomeric site-specific nickase and a catalytic subunit. *J Mol Biol* 384: 489-502

Käll L, Krogh A, Sonnhammer ELL (2004) A Combined Transmembrane Topology and Signal Peptide Prediction Method. *J Mol Biol* 338: 1027-1036

Kinch LN, Grishin NV (2013) Bioinformatics perspective on rhomboid intramembrane protease evolution and function. *Biochim Biophys Acta* 1828: 2937-43

Koo BM, Kritikos G, Farelli JD, Todor H, Tong K, Kimsey H, Wapinski I, Galardini M, Cabal A, Peters JM, Hachmann AB, Rudner DZ, Allen KN, Typas A, Gross CA (2017) Construction and Analysis of Two Genome-Scale Deletion Libraries for *Bacillus subtilis*. *Cell Syst* 4: 291-305 e7

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

Kota J, Gilstring CF, Ljungdahl PO (2007) Membrane chaperone Shr3 assists in folding amino acid permeases preventing precocious ERAD. *J Cell Biol* 176: 617-28

Kreutzberger AJB, Ji M, Aaron J, Mihaljevic L, Urban S (2019) Rhomboid distorts lipids to break the viscosity-imposed speed limit of membrane diffusion. *Science* 363

Langklotz S, Baumann U, Narberhaus F (2012) Structure and function of the bacterial AAA protease FtsH. *Biochim Biophys Acta* 1823: 40-8

Lauwers E, Erpapazoglou Z, Haguenauer-Tsapis R, Andre B (2010) The ubiquitin code of yeast permease trafficking. *Trends Cell Biol* 20: 196-204

Lee EJ, Pontes MH, Groisman EA (2013) A bacterial virulence protein promotes pathogenicity by inhibiting the bacterium's own F1Fo ATP synthase. *Cell* 154: 146-56

Lemberg MK, Martoglio B (2003) Analysis of polypeptides by sodium dodecyl sulfate–polyacrylamide gel electrophoresis alongside in vitro-generated reference peptides. *Anal Biochem* 319: 327-331

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: 464-72

Loureiro J, Lilley BN, Spooner E, Noriega V, Tortorella D, Ploegh HL (2006) Signal peptide peptidase is required for dislocation from the endoplasmic reticulum. *Nature* 441: 894-7

Maruyama T, Imai S, Kusakizako T, Hattori M, Ishitani R, Nureki O, Ito K, Maturana AD, Shimada I, Osawa M (2018) Functional roles of Mg(2+) binding sites in ion-dependent gating of a Mg(2+) channel, MgtE, revealed by solution NMR. *eLife* 7

Mesak LR, Mesak FM, Dahl MK (2004) Expression of a novel gene, *gluP*, is essential for normal *Bacillus subtilis* cell division and contributes to glucose export. *BMC Microbiol* 4: 13

Neal S, Jaeger PA, Duttke SH, Benner C, C KG, Ideker T, Hampton RY (2018) The Dfm1 Derlin is required for ERAD retrotranslocation of integral membrane proteins. *Mol Cell* 69: 306-320 e4

Nicolas P, Mader U, Dervyn E, Rochat T, Leduc A, Pigeonneau N, Bidnenko E, Marchadier E, Hoebeke M, Aymerich S, Becher D, Bisicchia P, Botella E, Delumeau O, Doherty G, Denham EL, Fogg MJ, Fromion V, Goelzer A, Hansen A et al. (2012) Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. *Science* 335: 1103-6

Pang T, Wang X, Lim HC, Bernhardt TG, Rudner DZ (2017) The nucleoid occlusion factor Noc controls DNA replication initiation in *Staphylococcus aureus*. *PLoS Genet* 13: e1006908

Parente J, Casabuono A, Ferrari MC, Paggi RA, De Castro RE, Couto AS, Gimenez MI (2014) A rhomboid protease gene deletion affects a novel oligosaccharide N-linked to the S-layer glycoprotein of *Haloferax volcanii*. *J Biol Chem* 289: 11304-17

Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J, Mayer G, Eisenacher M, Perez E, Uszkoreit J, Pfeuffer J, Sachsenberg T, Yilmaz S, Tiwary S, Cox J, Audain E, Walzer M, Jarnuczak AF et al. (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* 47: D442-D450

Pontes MH, Sevostyanova A, Groisman EA (2015) When too much ATP is bad for protein synthesis. *J Mol Biol* 427: 2586-2594

Pontes MH, Yeom J, Groisman EA (2016) Reducing ribosome biosynthesis promotes translation during low Mg(2+) stress. *Mol Cell* 64: 480-492

Price IR, Gaballa A, Ding F, Helmann JD, Ke A (2015) Mn(2+)-sensing mechanisms of *yybP-ykoY* orphan riboswitches. *Mol Cell* 57: 1110-1123

Que Q, Helmann JD (2000) Manganese homeostasis in *Bacillus subtilis* is regulated by MntR, a bifunctional regulator related to the diphtheria toxin repressor family of proteins. *Mol Microbiol* 35: 1454-68

Renshaw PS, Veverka V, Kelly G, Frenkiel TA, Williamson RA, Gordon SV, Hewinson RG, Carr MD (2004) Sequence-specific assignment and secondary structure determination of the 195-residue complex formed by the *Mycobacterium tuberculosis* proteins CFP-10 and ESAT-6. *J Biomol NMR* 30: 225-6

Riestra AM, Gandhi S, Sweredoski MJ, Moradian A, Hess S, Urban S, Johnson PJ (2015) A *Trichomonas vaginalis* rhomboid protease and its substrate modulate parasite attachment and cytolysis of host cells. *PLoS Pathog* 11: e1005294

Russell CW, Richards AC, Chang AS, Mulvey MA (2017) The rhomboid protease GlpG promotes the persistence of extraintestinal pathogenic *Escherichia coli* within the gut. *Infect Immun* 85

Saita S, Nolte H, Fiedler KU, Kashkar H, Venne AS, Zahedi RP, Kruger M, Langer T (2017) PARL mediates Smac proteolytic maturation in mitochondria to promote apoptosis. *Nat Cell Biol* 19: 318-328

Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* 44: 213-23

Smith RL, Thompson LJ, Maguire ME (1995) Cloning and characterization of MgtE, a putative new class of Mg2+ transporter from *Bacillus firmus* OF4. *J Bacteriol* 177: 1233-8

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: 1003-8

Strisovsky K (2013) Structural and mechanistic principles of intramembrane proteolysis--lessons from rhomboids. *FEBS J* 280: 1579-603

Strisovsky K (2016) Why cells need intramembrane proteases - a mechanistic perspective. *FEBS J* 283: 1837-45

Strisovsky K, Sharpe HJ, Freeman M (2009) Sequence-specific intramembrane proteolysis: identification of a recognition motif in rhomboid substrates. *Mol Cell* 36: 1048-59

Takeda H, Hattori M, Nishizawa T, Yamashita K, Shah ST, Caffrey M, Maturana AD, Ishitani R, Nureki O (2014) Structural basis for ion selectivity revealed by high-resolution crystal structure of Mg^{2+} channel MgtE. *Nat Commun* 5: 5374

Thomas FM, Sprenger S (2008) Responses of two closely related oak species, Quercus robur and Q. petraea, to excess manganese concentrations in the rooting medium. *Tree Physiol* 28: 343-53

Ticha A, Collis B, Strisovsky K (2018) The rhomboid superfamily: structural mechanisms and chemical biology opportunities. *Trends Biochem Sci* 43: 726-739

Ticha A, Stanchev S, Skerle J, Began J, Ingr M, Svehlova K, Polovinkin L, Ruzicka M, Bednarova L, Hadravova R, Polachova E, Rampirova P, Brezinova J, Kasicka V, Majer P, Strisovsky K (2017a) Sensitive Versatile Fluorogenic Transmembrane Peptide Substrates for Rhomboid Intramembrane Proteases. *J Biol Chem* 292: 2703-2713

Ticha A, Stanchev S, Vinothkumar KR, Mikles DC, Pachl P, Began J, Skerle J, Svehlova K, Nguyen MTN, Verhelst SHL, Johnson DC, Bachovchin DA, Lepsik M, Majer P, Strisovsky K (2017b) General and Modular Strategy for Designing Potent, Selective, and Pharmacologically Compliant Inhibitors of Rhomboid Proteases. *Cell Chem Biol* 24: 1523-1536 e4

Tomita A, Zhang M, Jin F, Zhuang W, Takeda H, Maruyama T, Osawa M, Hashimoto KI, Kawasaki H, Ito K, Dohmae N, Ishitani R, Shimada I, Yan Z, Hattori M, Nureki O (2017) ATP-dependent modulation of MgtE in Mg^{2+} homeostasis. *Nat Commun* 8: 148

Urban S (2016) A guide to the rhomboid protein superfamily in development and disease. *Semin Cell Dev Biol* 60: 1-4

Urban S, Lee JR, Freeman M (2002a) A family of Rhomboid intramembrane proteases activates all Drosophila membrane-tethered EGF ligands. *EMBO J* 21: 4277-86

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

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: 1883-8

Veverka V, Lennie G, Crabbe T, Bird I, Taylor RJ, Carr MD (2006) NMR assignment of the mTOR domain responsible for rapamycin binding. *J Biomol NMR* 36 Suppl 1: 3

Wagner S, Klepsch MM, Schlegel S, Appel A, Draheim R, Tarry M, Hogbom M, van Wijk KJ, Slotboom DJ, Persson JO, de Gier JW (2008) Tuning *Escherichia coli* for membrane protein overexpression. *Proc Natl Acad Sci U S A* 105: 14371-6

Wakeman CA, Goodson JR, Zacharia VM, Winkler WC (2014) Assessment of the requirements for magnesium transporters in *Bacillus subtilis*. *J Bacteriol* 196: 1206-14

Wang Y, Zhang Y, Ha Y (2006) Crystal structure of a rhomboid family intramembrane protease. *Nature* 444: 179-80

Yang Y, Guo R, Gaffney K, Kim M, Muhammednazaar S, Tian W, Wang B, Liang J, Hong H (2018) Folding-Degradation Relationship of a Membrane Protein Mediated by the Universally Conserved ATP-Dependent Protease FtsH. *J Am Chem Soc* 140: 4656-4665

Zeytuni N, Zarivach R (2012) Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module. *Structure* 20: 397-405

Zoll S, Stanchev S, Began J, Skerle J, Lepsik M, Peclinovska L, Majer P, Strisovsky K (2014) Substrate binding and specificity of rhomboid intramembrane protease revealed by substrate-peptide complex structures. *EMBO J* 33: 2408-21

Figures legend and Figures.

Fig. 1: Quantitative proteomics reveals candidate interactors and substrates of rhomboid protease YqgP from *B. subtilis*.

(A) Schematic representation of the SILAC-based quantitative proteomics experiment in *B. subtilis*. Cells overexpressing active YqgP (BS50, Table EV2) or its catalytically dead mutant YqgP.S288A (BS51), both auxotrophic for lysine, were grown in parallel in "heavy" (containing $^{13}C_6{}^{15}N_2$- Lysine isotope) or "light" (containing stable $^{12}C_6{}^{14}N_2$- Lysine) M9 minimal medium, respectively. After mixing the cell cultures in the 1:1 ratio (based on OD_{600}), cell suspension was lysed and the fraction enriched for transmembrane proteome was analysed in a GeLC-MS/MS experiment. Using bioinformatic analysis of the MS data, highest-scoring candidate substrates were further evaluated.

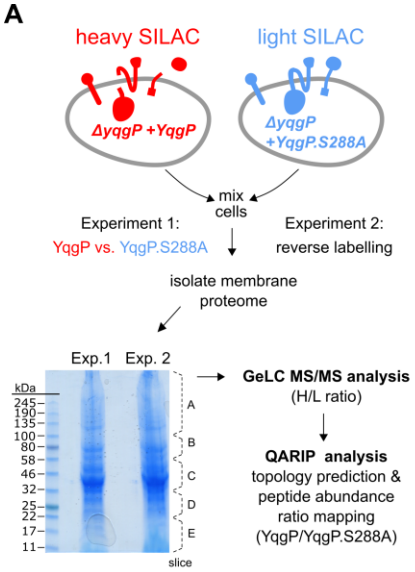
(B) Diagrammatic representation of a result of GeLC-MS/MS analysis of excised gel regions described in (A). Each coloured block depicts a quantified peptide at its amino acid position according to the horizontal axis. Possible ongoing proteolysis was identified by high abundance ratio of YqgP/YqgP.S288A for a given protein and lower apparent molecular weight than expected for the corresponding full-length protein, as estimated from the position of the respective gel slice relative to the molecular weight marker.

(C) Table of best substrate candidates of two GeLC experiments. In order to assess possible cleavage sites in combination with topology information, the QARIP software (Ivankov et al., 2013) was used to summarize results from experiments 1 and 2. Abundance ratios (YqgP/YqgP.S288A) for the intracellular parts of proteins were calculated for each gel slice separately by QARIP from peptide ratios computed by MaxQuant (Cox & Mann, 2008). Substrate candidates highlighted in red were identified in both experiments.

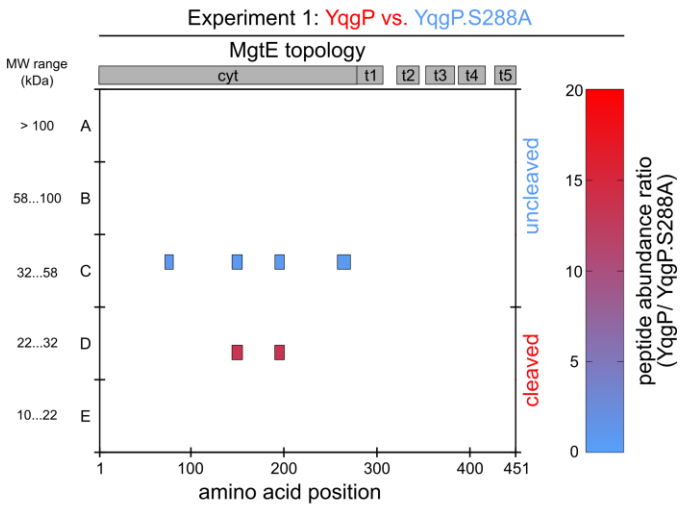
(D) Schematic representation of the affinity co-purification experiment in *B. subtilis* to identify YqgP interactors.

(E) Results of MS analyses of affinity co-purification experiments in wild-type *B. subtilis* control (strain BTM2, Table EV2) and *B. subtilis* deficient in endogenous YqgP expressing the wild-type YqgP-sfGFP bait (strain BTM84, Table EV2) or the proteolytically inactive YqgP.S288A-sfGFP bait (strain BBM1, Table EV2). Proteins were considered as potential interactors of YqgP if they were identified only in both positive co-purifications with a minimum of 3 weighted spectral counts or enriched at least 5 times in positive bait samples compared to control ones based on weighted spectral counts. The protein highlighted in red was the only overlapping hit between the two proteomics approaches.

A



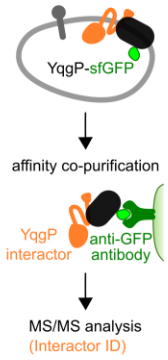
B



C

Gel slice	MW range (kDa)	Uniprot accesion	Gene name	MW theor. (kDa)	Protein	Topology prediction (Phobius)	TM domains	Identified peptides	Abundance ratio for the intracellular part
Exp2_D	22 to 32	O34442	mgtE	50.8	Magnesium transporter MgtE	Nin-Cout	5	2	17.94
Exp1_C	32 to 58	O31707	yknU	66.2	Uncharacterized ABC transporter ATP-binding protein YknU	Nin-Cin	6	1	17.49
Exp1_D	22 to 32	O34442	mgtE	50.8	Magnesium transporter MgtE	Nin-Cout	5	2	13.45
Exp2_C	32 to 58	O07549	yheH	76.3	Probable ABC transporter ATP-binding protein YheH	Nout-Cout	4	1	2.16
Exp1_C	32 to 58	P20166	ptsG	75.5	Glucose-specific phosphotransferase enzyme IIA component	Nout-Cin	11	1	2.07
Exp2_E	10 to 22	P54467	yqfB	15.7	Uncharacterized protein YqfB	Nout-Cin	1	1	2

D



E

Uniprot accession	Gene name	Protein	Weighted spectral counts			
			Experiment 1		Experiment 2	
			WT ctrl (BTM2)	ΔyqgP + YqgP-sfGFP (BTM84)	WT ctrl (BTM2)	ΔyqgP + YqgP.S288A-sfGFP (BBM1)
P54493	yqgP-sfGFP	Rhomboid protease YqgP-sfGFP fusion		143		67
P37809	atpD	ATP synthase subunit beta		93	2	62
P37808	atpA	ATP synthase subunit alpha		78		27
P37476	ftsH	ATP-dependent zinc metalloprotease FtsH		30		18
P37814	atpF	ATP synthase subunit b		15		20
O31645	manP	PTS system mannose-specific EIIBCA component	1	12		13
O32047	secDF	Protein translocase subunit SecDF		5		13
P37810	atpG	ATP synthase gamma chain		8		10
P17631	dnaJ	Chaperone protein DnaJ		7		11
O34442	mgtE	Magnesium transporter MgtE		5		9
O34358	htrA	Serine protease Do-like HtrA		7		9
P17820	dnaK	Chaperone protein DnaK		3		9
O31501	swrC	Swarming motility protein SwrC		3		6
P37811	atpH	ATP synthase subunit delta		4		6
Q01465	mreB	Rod shape-determining protein MreB		5	2	13
P39794	treP	PTS system trehalose-specific EIIBC component		5		4
P37812	atpC	ATP synthase epsilon chain		8		4
P96712	bmr3	Multidrug resistance protein 3		4		4
O32023	yqzC	Uncharacterized protein YqzC		3		4
P54420	asnB	Asparagine synthetase [glutamine-hydrolyzing] 1		3		3
P36948	rbsC	Ribose import permease protein RbsC		4		4

Fig. 2: YqgP cleaves the high-affinity magnesium transporter MgtE between its first and second transmembrane helices.

(A) Steady-state cleavage profile of endogenous MgtE processed by endogenous and ectopically overexpressed YqgP from the inducible $P_{hyperspank}$ promoter in living *B. subtilis* (BTM2 and BTM501, respectively, Table EV2), in minimal medium at low magnesium concentration (10 µM). Strain lacking YqgP (Δ*yqgP*, BTM78, Table EV2) was used as a control. Proteins were detected by immunoblotting with chemiluminescence detection.

(B) The cleavage is efficiently inhibited by 1 µM STS736, a specific peptidyl ketoamide rhomboid inhibitor (Ticha et al., 2017b).

(C) To map the cleavage site region within endogenous MgtE (second lane from the left, from strain BTM501), MgtE-derived reference fragments encoding first 300, 315, 330 amino acids as well as full-length MgtE (black arrow), were *in vitro* transcribed and translated. Mobility of the N-terminal cleavage product (red arrow) of MgtE on SDS PAGE was compared to the mobilities of the translated reference fragments.

(D) Diagrammatic display of the mapping shows that YqgP cleaves MgtE in a periplasmic region between transmembrane helices 1 and 2 (red arrow).

Data information: In all panels, endogenous full-length MgtE is indicated by a black arrow, and N-terminal cleavage product by red arrows. Endogenous MgtE was visualised by anti-MgtE(2-275) (α-MgtE) and ectopic YqgP by anti-YqgP antibodies.

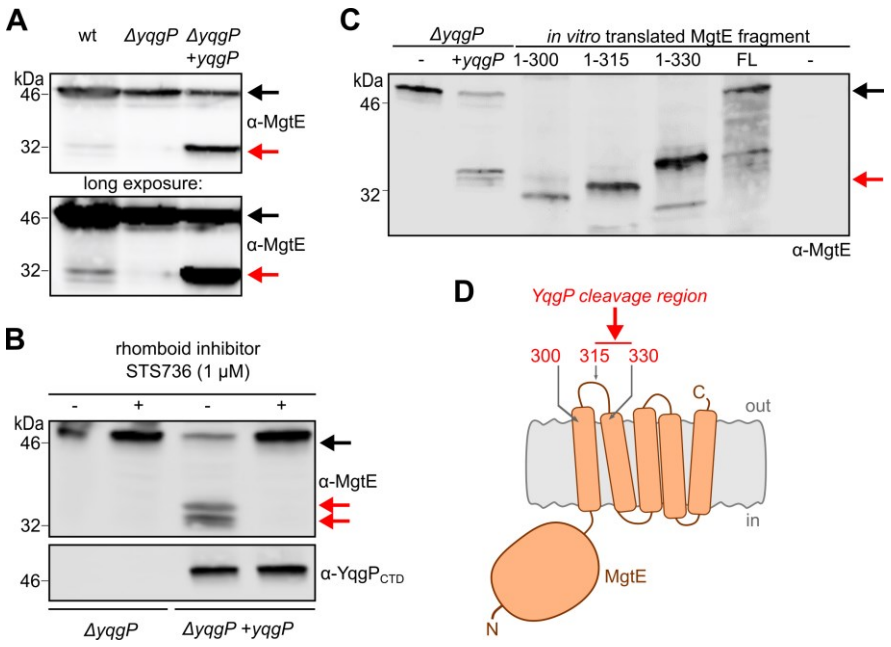


Fig. 3: Magnesium starvation and manganese excess activate cleavage of MgtE by YqgP, which is beneficial in manganese stress conditions.

(A) Detection and quantification of the cleavage of endogenous MgtE by YqgP in living *B. subtilis* cells (BS72, Table EV2) depending on the concentrations of magnesium and manganese ions. Cells were cultivated in glucose M9 minimal medium with limiting (0.01 mM) or high (1 mM) concentration of $MgSO_4$, in the presence or absence of 100 µM $MnCl_2$, and analysed by western blotting with near-infrared detection (upper panel). Black arrow denotes full-length MgtE, and red arrow denotes its N-terminal cleavage product formed by YqgP. The corresponding fluorescence signals were quantified by densitometry, and are displayed as relative specific activity, which is substrate conversion normalized to enzyme expression level (lower panel).

(B) Growth curves of wild-type (BTM843, Table EV2), *yqgP* deficient (BTM844, Table EV2) and rescue (BTM845, Table EV2) strains of *B. subtilis* in M9 minimal medium with limiting magnesium (0.01 mM $MgSO_4$), exposed to manganese stress elicited by adding 75 µM $MnSO_4$ in mid-exponential phase (stress-phase denoted by blueish background). All strains further contain a deletion in the putative manganese efflux pump MntP (*ΔywlD,* Table EV2*).* Bottom panel shows that manganese is more toxic in the *yqgP* deficient strain than in the wild type strain, and that reintroduction of YqgP rescues fitness during manganese stress to above-wild-type level. Top panel shows no difference between the strains in the absence of manganese stress. Data are shown as individual datapoints from three independent experiments overlaid with dashed line connecting average values from each time-point, which illustrates the reproducibility of the assay.

(C) Top panel: growth curves of wild-type (BTM843), *yqgP* deficient (BTM844) and rescue (BTM845) strains of *B. subtilis* in M9 minimal medium with limiting magnesium (0.01 mM $MgSO_4$), exposed to manganese shock elicited by adding 75 µM $MnSO_4$ in mid-exponential phase. Bottom panel: manganese toxicity is prevented by further adding 5 mM magnesium ($MgSO_4$) in otherwise identical conditions.

(D) Inhibition of YqgP by a rhomboid-specific peptidyl ketoamide inhibitor (STS736, i.e. compound ***9*** from (Ticha et al., 2017b)) abolishes the YqgP induced fitness of *B. subtilis* under manganese stress, in a dose-dependent manner, both with endogenous YqgP (top panel) and overexpressed YqgP (bottom panel). Media and growth conditions were identical to those used in panel (B).

(E) Overexpression of heterologous MgtE inhibits growth of *yqgP* deficient strain (BTM610, Table EV2) in rich LB medium supplemented with 75 µM $MnSO_4$. Cell fitness is improved by overexpression of YqgP (BTM611, Table EV2) or its catalytically dead mutant YqgP.S288A (BTM612, Table EV2). For clarity, for panels C – E, representative experiments of 2-3 independent biological replicates are shown unless stated otherwise.

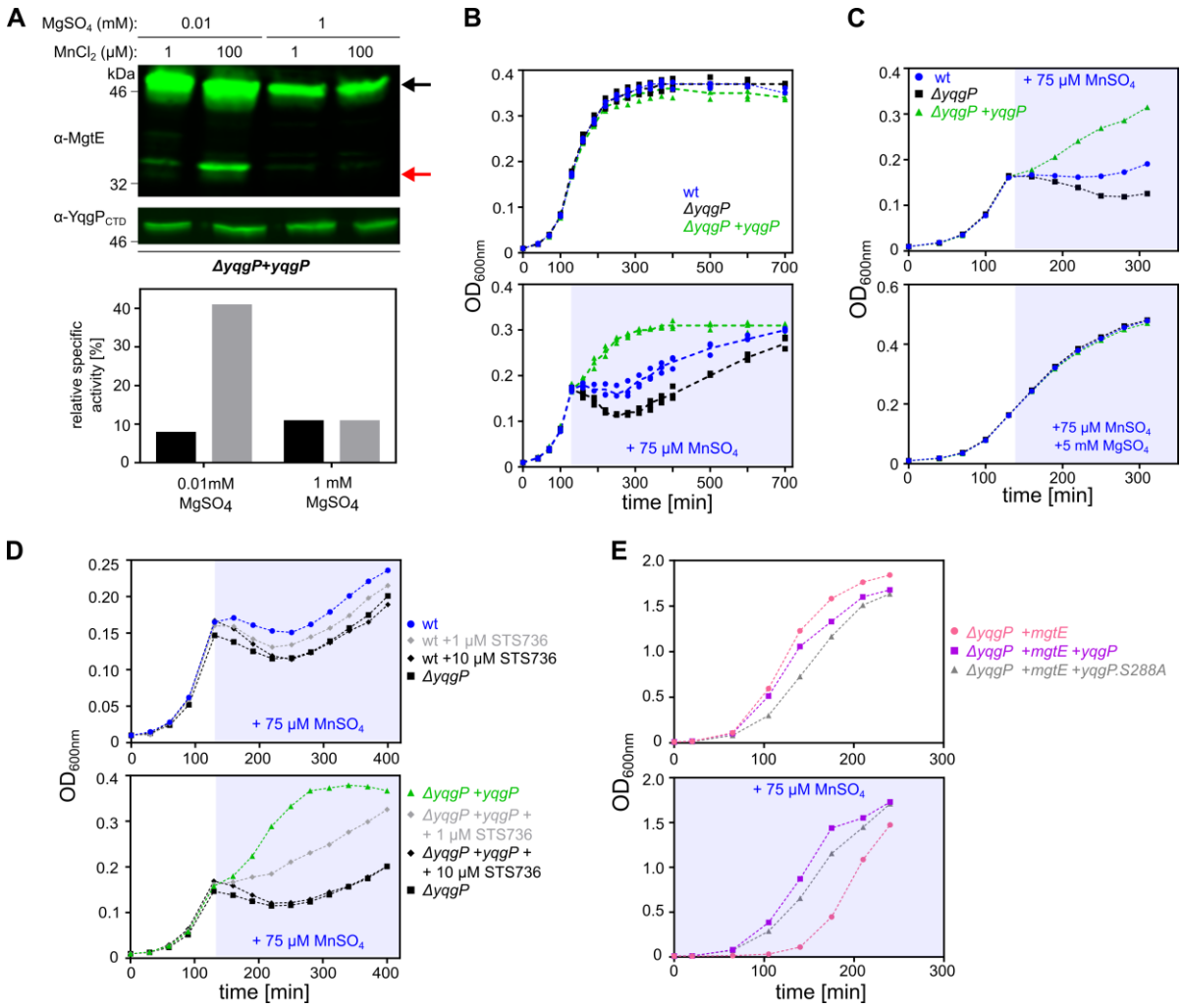

A
MgSO4 (mM): 0.01 1
MnCl2 (μM): 1 100 1 100
kDa
46
α-MgtE
32
α-YqgPCTD
46
ΔyqgP+yqgP
relative specific activity [%]
0.01mM MgSO4
1 mM MgSO4
B
wt
ΔyqgP
ΔyqgP +yqgP
+ 75 μM MnSO4
OD600nm
time [min]
C
wt
ΔyqgP
ΔyqgP +yqgP
+ 75 μM MnSO4
+75 μM MnSO4
+5 mM MgSO4
OD600nm
time [min]
D
wt
wt +1 μM STS736
wt +10 μM STS736
ΔyqgP
ΔyqgP +yqgP
ΔyqgP +yqgP +
+ 1 μM STS736
ΔyqgP +yqgP +
+ 10 μM STS736
+ 75 μM MnSO4
OD600nm
time [min]
E
ΔyqgP +mgtE
ΔyqgP +mgtE +yqgP
ΔyqgP +mgtE +yqgP.S288A
+ 75 μM MnSO4
OD600nm
time [min]

Fig. 4: The N-terminal extramembrane intracellular domain of YqgP mediates the manganese induced activation of MgtE cleavage.

(A) Schematic of YqgP domain architecture. Transmembrane helices are shown as grey boxes, and the position of the catalytic dyad of S288 and H339 within the transmembrane rhomboid core is displayed.

(B) Detection (left panel) and quantification (right panel) of steady-state conversion of endogenous MgtE and model chimeric substrate derived from *Providencia stuartii* TatA (Stevenson et al., 2007, Ticha et al., 2017a, Ticha et al., 2017b) (schemes in the middle) by ectopically expressed YqgP and its N- and C-terminally truncated variants (strains BS184-187, Table EV2) in living *B. subtilis* grown in rich LB medium. Black arrows indicate full-length substrates and red arrows indicate cleavage products.

(C) Similar analysis of the same strains but grown in minimal M9 medium (with 10 µM $MgSO_4$) supplemented with increasing concentrations of $MnCl_2$. Black arrows denote full-length substrate (MgtE), red arrows denote the cleavage products generated by YqgP.

(D) Western blots of four independent experiments shown in panel (C) were quantified by near-infrared fluorescence detection, quantified by densitometry and the results are displayed as fold activation of MgtE cleavage at 100 µM $MnCl_2$ relative to 1 µM $MnCl_2$ activity. Average values and all fours datapoints are plotted for each indicated YqgP variant.

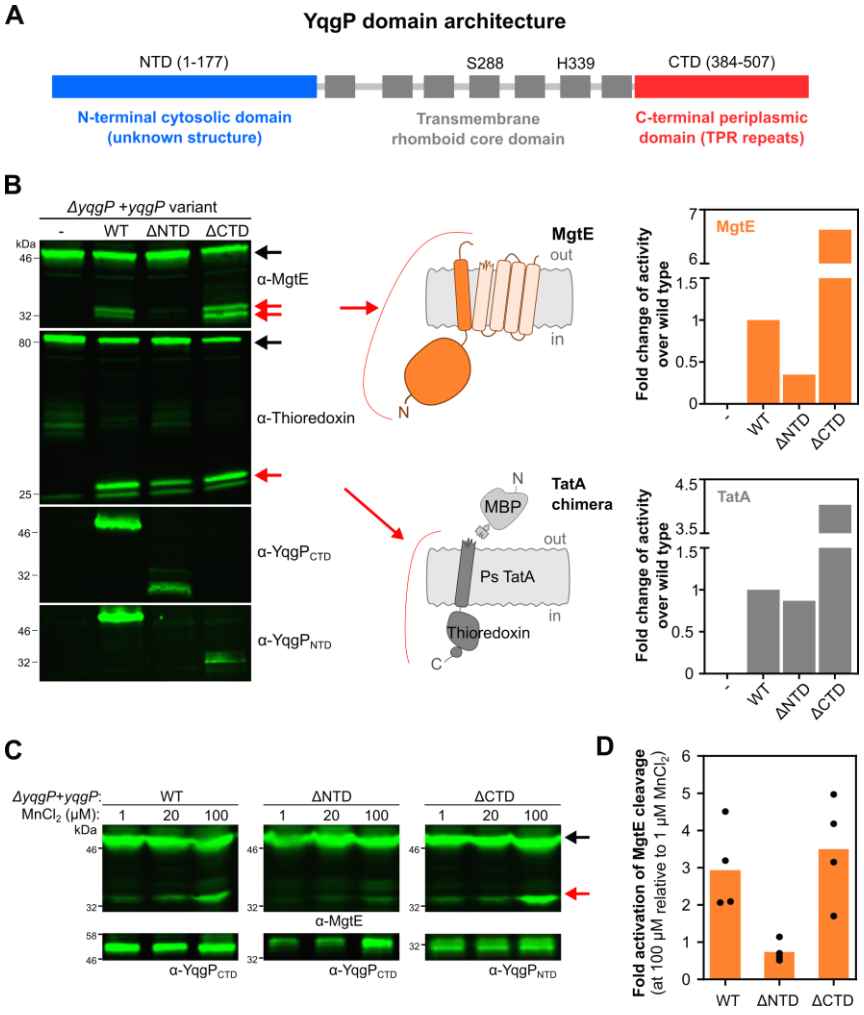

A
YqgP domain architecture
NTD (1-177)
S288
H339
CTD (384-507)
N-terminal cytosolic domain (unknown structure)
Transmembrane rhomboid core domain
C-terminal periplasmic domain (TPR repeats)
B
ΔyqgP +yqgP variant
-
WT
ΔNTD
ΔCTD
kDa
α-MgtE
α-Thioredoxin
α-YqgP$_{CTD}$
α-YqgP$_{NTD}$
MgtE
out
in
N
TatA chimera
MBP
Ps TatA
Thioredoxin
C
Fold change of activity over wild type
TatA
C
ΔyqgP+yqgP:
MnCl$_2$ (μM):
1
20
100
D
Fold activation of MgtE cleavage (at 100 μM relative to 1 μM MnCl$_2$)

Fig. 5: The N-terminal intracellular globular domain of YqgP binds divalent cations.

(A) Detection (left panel) and quantification (right panel) of steady-state conversion of endogenous MgtE by YqgP (strain BS72, Table EV2) grown in minimal M9 medium containing 10 µM $MgSO_4$ and 1 µM each of $MnCl_2$, $ZnCl_2$, $CoCl_2$, $NiCl_2$ and $CaCl_2$ (reference conditions) with the additions of 10 µM or 100 µM of a given divalent cation salt solution ($MnCl_2$, $ZnCl_2$, $CoCl_2$, or $NiCl_2$), or 50 µM and 500 µM for $CaCl_2$. Western blots were quantified by near-infrared fluorescence detection (left), quantified by densitometry and displayed as relative specific activity (graph on the right), which is substrate conversion normalized to enzyme expression level. Black arrow denotes full-length substrate (MgtE), and red arrow denotes the N-terminal cleavage product(s) generated by YqgP. All bands originate from the same western blot and identical treatment series. Irrelevant lanes have been cropped out for clarity, and the source blot is available online as Source Data.

(B) Isothermal titration calorimetry of purified recombinant $YqgP_{NTD}$ and selected divalent cations.

(C) Cation toxicity assays and their relationship to YqgP activity. Wild type *B. subtilis* (BTM2, Table EV2), its variant lacking YqgP (*ΔyqgP,* BTM78, Table EV2), and the rescue strain ectopically expressing YqgP (BTM501, Table EV2) were cultivated in minimal M9 medium containing 10 µM $MgSO_4$ and 1 µM each of $MnCl_2$, $ZnCl_2$, $CoCl_2$, $NiCl_2$ and $CaCl_2$ (i.e. reference conditions), and in mid-exponential phase were stressed by the addition of either 70 µM $MnCl_2$, 500 µM $ZnCl_2$, 25 µM $CoSO_4$ or 400 µM $NiCl_2$ (pale blue area). YqgP activity specifically improved cell fitness during manganese and zinc stress, while it had no effect on growth of cells cultivated in the presence of high cobalt and nickel concentrations. Representative experiments of 2-3 independent replicates are shown.

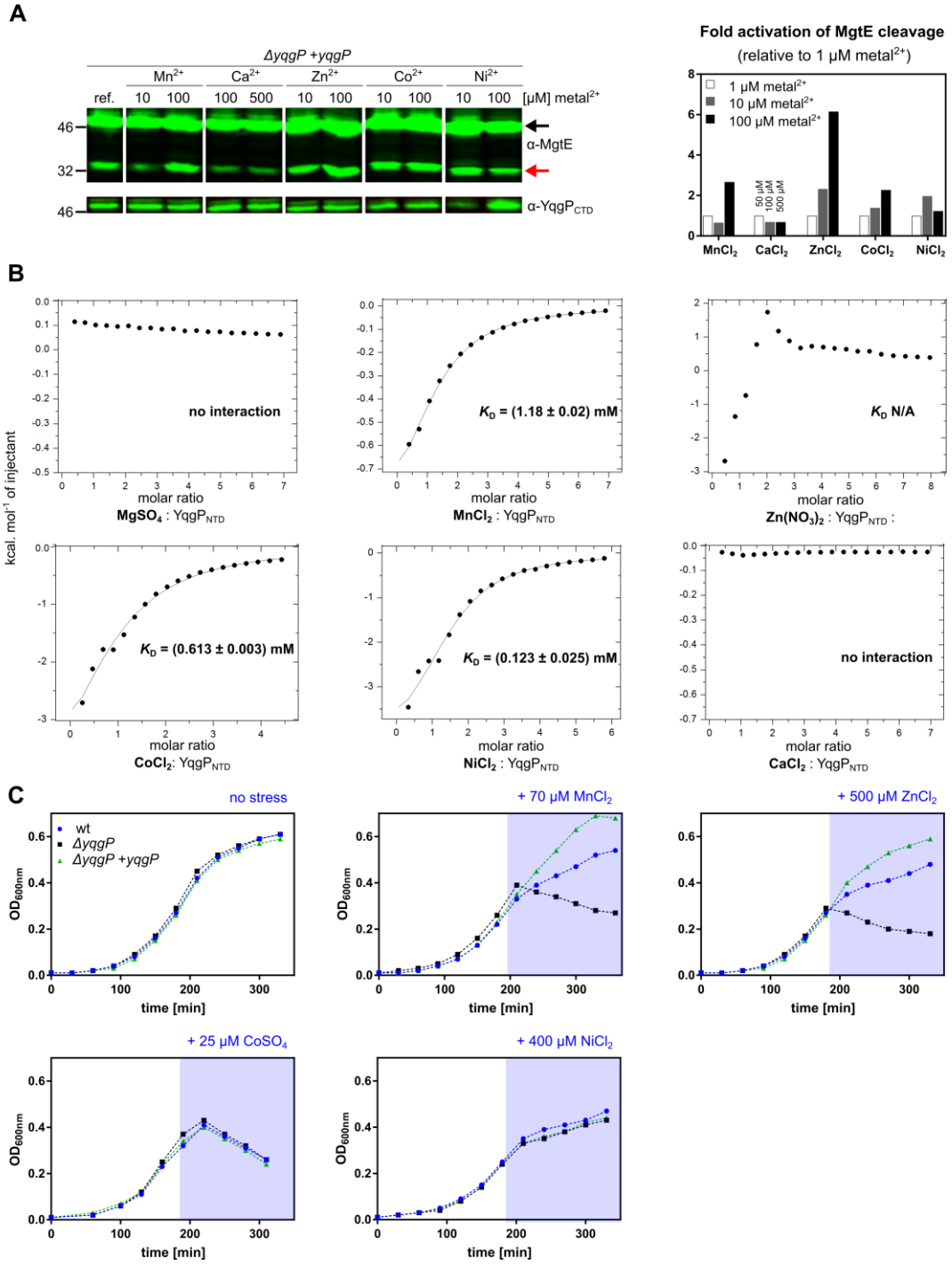
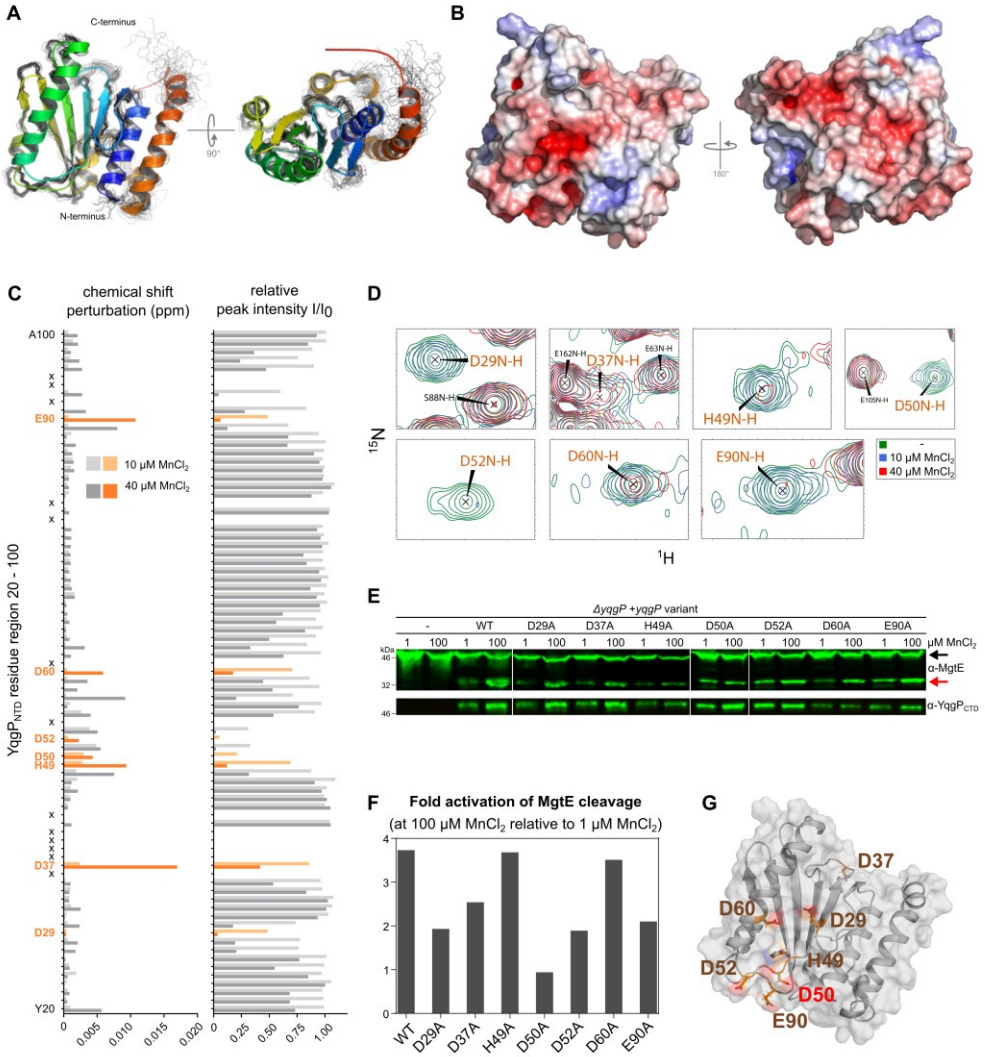

A
ΔyqgP +yqgP
Mn2+
Ca2+
Zn2+
Co2+
Ni2+
ref. 10 100 100 500 10 100 10 100 10 100 [μM] metal2+
46
32
46
α-MgtE
α-YqgPCTD
Fold activation of MgtE cleavage
(relative to 1 μM metal2+)
1 μM metal2+
10 μM metal2+
100 μM metal2+
50 μM
100 μM
500 μM
MnCl2 CaCl2 ZnCl2 CoCl2 NiCl2
B
kcal. mol-1 of injectant
no interaction
molar ratio
MgSO4 : YqgPNTD
KD = (1.18 ± 0.02) mM
MnCl2 : YqgPNTD
KD N/A
Zn(NO3)2 : YqgPNTD :
KD = (0.613 ± 0.003) mM
CoCl2: YqgPNTD
KD = (0.123 ± 0.025) mM
NiCl2 : YqgPNTD
no interaction
CaCl2 : YqgPNTD
C
no stress
wt
ΔyqgP
ΔyqgP +yqgP
OD600nm
time [min]
+ 70 μM MnCl2
+ 500 μM ZnCl2
+ 25 μM CoSO4
+ 400 μM NiCl2

Fig. 6: Solution NMR analysis reveals manganese-binding region in the N-terminal domain of YqgP.

(A) Solution NMR structure of cytosolic N-terminal domain of YqgP ($YqgP_{NTD}$, amino acids 1-177), showing ensemble of 30 structures with lowest energy, and the structure with minimum energy displayed as cartoon.
(B) Electrostatic surface of $YqgP_{NTD}$ structure reveals highly negatively charged areas. Left view is oriented as the left view in panel (A).
(C) Chemical shift perturbation values and relative peak intensities for backbone resonances (^{15}N and ^{1}H) of $YqgP_{NTD}$ calculated for complexes of 400 µM $YqgP_{NTD}$ with 10 µM (light grey, light orange) or 40 µM $MnCl_2$ (dark grey, dark orange bars). Only residues 20-100 of YqgP are displayed for simplicity. Residues identified as potential manganese-binding region based on their chemical shifts and intensity changes upon NMR titrations are depicted in orange. The X sign marks residues whose chemical shifts were not calculated due to the lack of resonance signal.
(D) Details of 2D $^{15}N/^{1}H$ HSQC spectra of free (green) and Mn-titrated $YqgP_{NTD}$ (10 µM $MnCl_2$ in red and 40 µM $MnCl_2$ in blue) for the residues marked in orange in panel (C). As a reference, residues marked in black do not display spectral changes upon titration, e.g. peak intensity does not change for S88N-H upon Mn titration (overlay of green, blue and red contours), but peak intensity is severely depleted for D29N-H upon Mn titration (absence of red contours). Full titration data are available in Dataset S3.
(E, F) Detection (E) and quantification (F) of steady-state conversions of endogenous MgtE by YqgP variants bearing single-point mutations in putative Mn-binding region (BS184; 187; 196-203, Table EV2), cultivated in modified M9 minimal medium supplemented by low (1 µM) or high (100 µM) $MnCl_2$. Black arrow marks full-length MgtE, and red arrow marks its N-terminal cleavage product by YqgP.
(G) Overview of residues affected by Mn^{2+} binding to $YqgP_{NTD}$ mapped onto its solution NMR structure. The view is oriented as the left view in panels (A and B).


A
C-terminus
N-terminus
90°
B
180°
C
chemical shift perturbation (ppm)
relative peak intensity I/I0
YqgP NTD residue region 20 - 100
A100
E90
D60
D52
D50
H49
D37
D29
Y20
10 μM MnCl2
40 μM MnCl2
0
0.005
0.010
0.015
0.020
0.25
0.50
0.75
1.00
D
D29N-H
S88N-H
E162N-H
D37N-H
E63N-H
H49N-H
E105N-H
D50N-H
D52N-H
D60N-H
E90N-H
15N
1H
10 μM MnCl2
40 μM MnCl2
E
ΔyqgP +yqgP variant
- WT D29A D37A H49A D50A D52A D60A E90A
1 100 1 100 1 100 1 100 1 100 1 100 1 100 1 100 1 100 μM MnCl2
kDa
46
32
46
α-MgtE
α-YqgPCTD
F
Fold activation of MgtE cleavage
(at 100 μM MnCl2 relative to 1 μM MnCl2)
4
3
2
1
0
WT
D29A
D37A
H49A
D50A
D52A
D60A
E90A
G
D37
D60
D29
H49
D52
D50
E90

1

2

Fig. 7: The YqgP rhomboid functions as both a protease and a pseudoprotease adaptor of the AAA-protease FtsH to maintain proteostasis of MgtE.

(A) The proteolytically dead mutant YqgP.S288A behanes as a pseudoprotease and acts as a substrate adaptor of FtsH. Pattern of C-terminally truncated forms of endogenous MgtE formed *in vivo* in the presence of ectopically expressed active YqgP (strain BS72, Table EV2) or its inactive S288A mutant (BS73, Table EV2) in rich LB medium in the presence or absence of endogenous FtsH. Black arrow denotes the full-length MgtE substrate, and red arrows denote its cleavage products generated by YqgP. While the presence of the faster migrating YqgP-dependent products (full red arrows) do not depend on the presence of FtsH, the slower migrating product formed in the presence of YqgP.S288A (open red arrow) does, indicating that FtsH protease is responsible for the alternative cleavage of MgtE. The same arrow symbolics is valid throughout this figure.
(B) Mapping of the boundary of the FtsH generated proteolytic product of MgtE. Immunoblot comparing the SDS PAGE mobility of the N-terminal fragment of endogenous MgtE generated by FtsH in the presence of the pseudoprotease version of YqgP (catalytically dead YqgP.S288A, strains BTM78, BTM502, Table EV2) to the mobilities of *in vitro* translated reference fragments corresponding to MgtE 1-340, -355 and -370 shows that YqgP.S288A acts as an FtsH adaptor, and that FtsH processing of MgtE in the presence of YqgP.S288A stops at around amino acid 355, which is near or at the cytoplasmic end of TMH 3 of MgtE.
(C) Schematic depiction of the regions in MgtE defining its remaining cleavage products formed by FtsH and YqgP.
(D) *In vivo* cleavage of endogenous MgtE in the presence of ectopically expressed full-length YqgP.S288A (strain BS73), YqgPΔNTD.S288A (BS185, Table EV2) or YqgPΔCTD.S288A (BS186, Table EV2), in rich LB medium shows that $YqgP_{NTD}$ is necessary for the FtsH-dependent processing of MgtE.
(E) Analysis of the *in vivo* kinetics of formation and fate of MgtE cleavage products. *B. subtilis* strains expressing YqgP or its S288A mutant (strains BTM78; 501 and 502, Table EV2) in the presence or absence of endogenous FtsH protease (strains BTM795, BTM796 and BTM797, Table EV2) were cultured in LB medium and at early exponential phase, after having been expressing YqgP variants for 30 min, were either left grown untreated (as a control with ongoing translation) or treated with 20 µg/mL tetracycline to stop proteosynthesis (translation shut-off). At given time intervals afterwards, all cultures were analysed by α-MgtE western blotting with near infrared detection. Equal cell number (judged by OD_{600}) was loaded into each lane. Black arrow marks full-length MgtE, red full arrows mark YqgP-dependent and red open arrows mark YqgP.S288A-dependent cleavage product of MgtE.

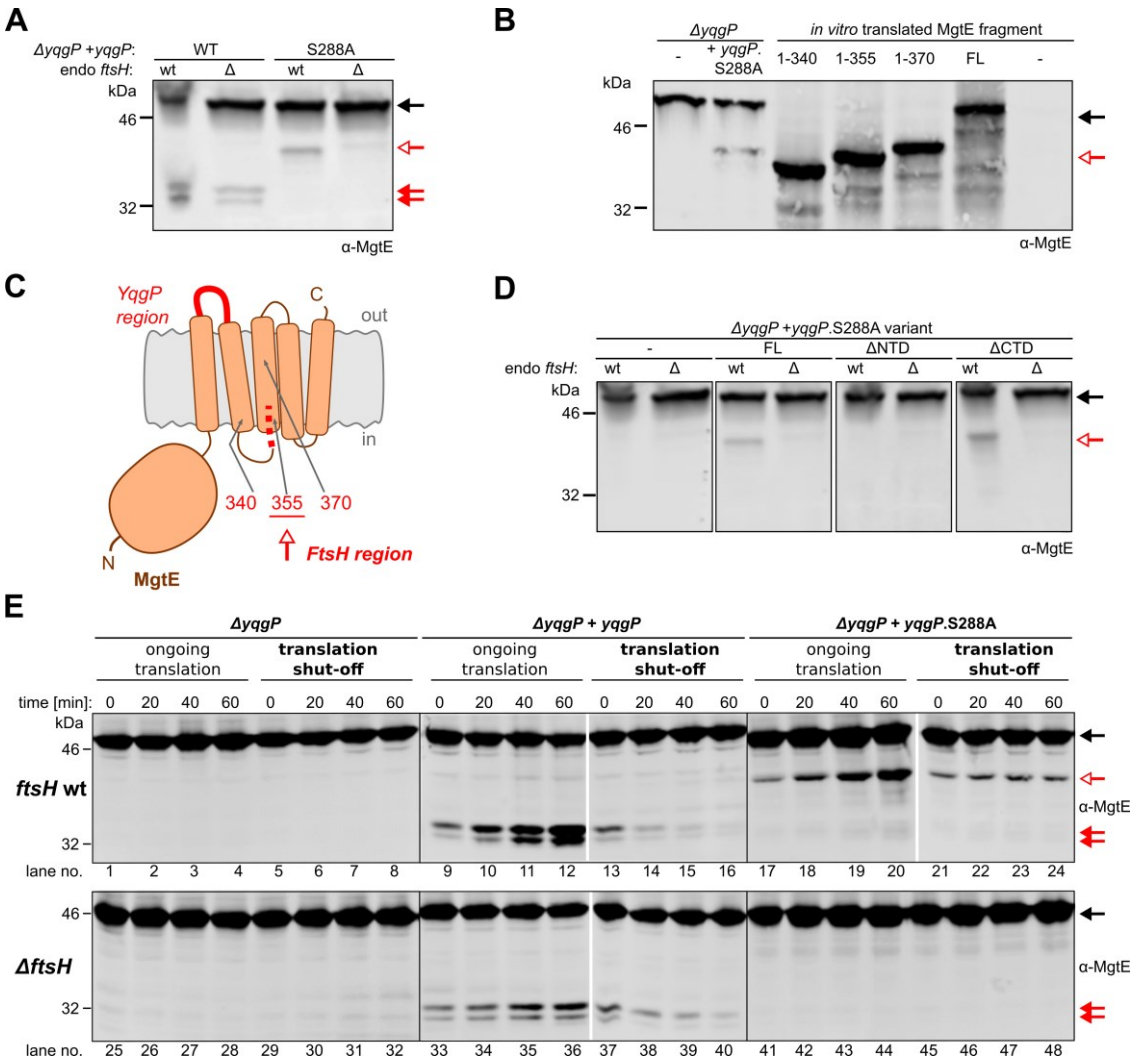

A
ΔyqgP +yqgP:
endo ftsH:
WT
S288A
wt
Δ
wt
Δ
kDa
46
32
α-MgtE
B
ΔyqgP
in vitro translated MgtE fragment
-
+ yqgP. S288A
1-340
1-355
1-370
FL
-
kDa
46
32
α-MgtE
C
YqgP region
C
out
in
340
355
370
FtsH region
N
MgtE
D
ΔyqgP +yqgP.S288A variant
-
FL
ΔNTD
ΔCTD
endo ftsH:
wt
Δ
wt
Δ
wt
Δ
wt
Δ
kDa
46
32
α-MgtE
E
ΔyqgP
ΔyqgP + yqgP
ΔyqgP + yqgP.S288A
ongoing translation
translation shut-off
time [min]:
0
20
40
60
kDa
46
ftsH wt
32
α-MgtE
lane no.
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
46
ΔftsH
32
α-MgtE
lane no.
25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48

Fig. 8: Summary of the role of YqgP in MgtE proteostasis

Under conditions of magnesium starvation, MgtE is upregulated and open, poised to transport Mg^{2+} inside the cell. If other divalent cations, such as Mn^{2+} or Zn^{2+} are present in the environment at relatively high concentrations, MgtE can transport them or be inhibited by them, which causes toxicity. Under these conditions, the N-terminal cytosolic domain of YqgP binds Mn^{2+} or Zn^{2+}, and activates YqgP for cleavage of MgtE between its transmembrane helices 1 and 2. YqgP also interacts with FtsH, and presents MgtE or its cleavage product(s) for processive proteolysis by FtsH. The cleavage of MgtE by YqgP is required for its full dislocation and degradation by FtsH. The presentation/adaptor function of YqgP is not dependent on its catalytic activity, but it requires its active site unoccupied. YqgP thus behaves also as a pseudoprotease which are common in the rhomboid superfamily. The connection of a rhomboid-like protease/pseudoprotease (YqgP) with a processive degradative membrane-bound protease/dislocase (FtsH) and its role in membrane protein quality control in *B. subtilis* that we identify here represent a striking analogy to the eukaryotic role of rhomboid-like pseudoproteases Derlins in ER-associated degradation.

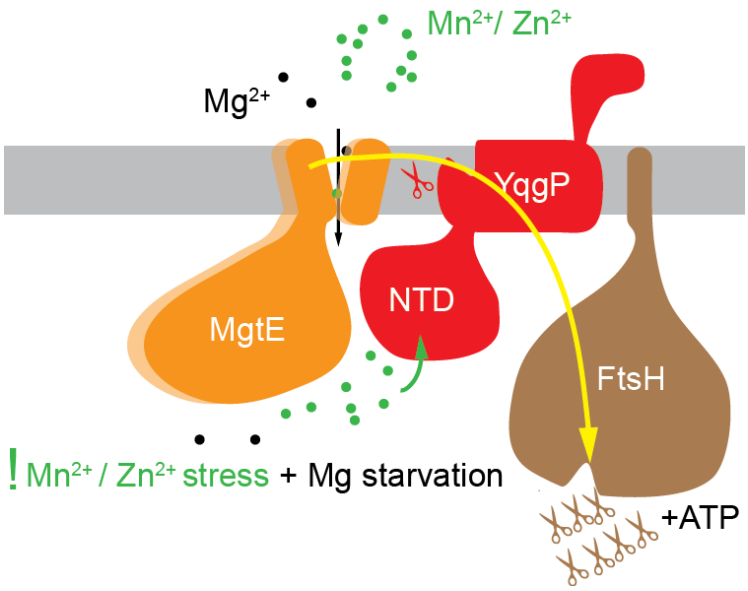


Figure Legend for EV Figures

Fig. EV1: Analysis of the role of *ydcA* in the phenotypic behaviour of *B. subtilis* during Mn stress.

Growth curves of wild-type (BTM843, Table EV2), *yqgP* deficient (BTM844, Table EV2), *ydcA* deficient (BTM1001, Table EV2), *yqgP ydcA* deficient (BTM1003, Table EV2) and YqgP rescue (BTM845 and BTM1005, Table EV2) strains of *B. subtilis* in M9 minimal medium with limiting magnesium (0.01 mM $MgSO_4$), exposed to manganese stress elicited by adding 75 µM $MnSO_4$ in mid-exponential phase (stress-phase denoted by blueish background, in bottom panel). All strains further contain a deletion in the putative manganese efflux pump MntP (*ΔywlD,* Table EV2*).* Bottom panel shows that manganese is more toxic in both *yqgP* deficient strains (black squares or open black circles) than in the wild type or *ydcA* rhomboid deficient strains. The overexpression of YqgP in both *ΔyqgP* and *ΔyqgPΔydcA* strains rescues fitness during manganese stress to above-wild-type level. Top panel shows no difference between the strains in the absence of manganese stress.

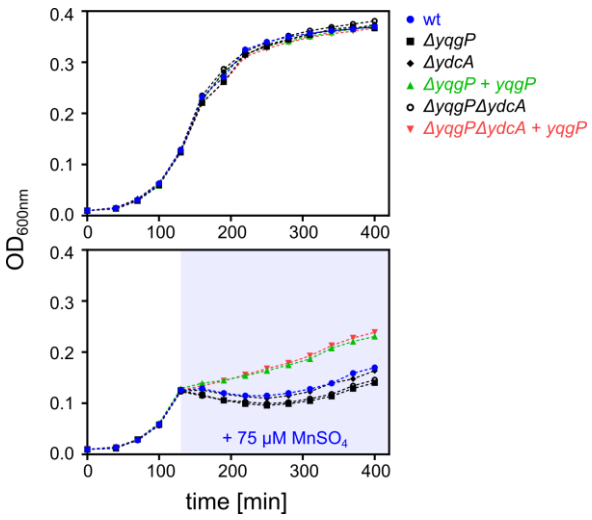


Fig. EV2: Stability of MgtE cleavage products depending on the presence of FtsH

Translational stop chase experiment described in Fig. 7 (lanes 13-16 and 37-40) was performed in four independent replicates and quantified by near-infrared western blotting using α-MgtE antibody as described in Methods. The conversion of MgtE was plotted against time as average values ± SD, indicating that in the absence of endogenous FtsH, MgtE cleavage products are more stable.

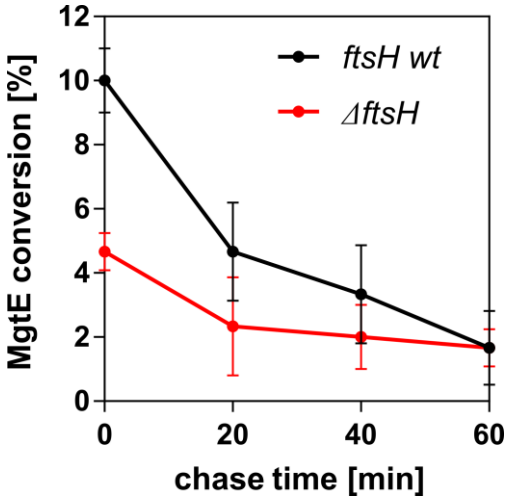


Table EV1: List of DNA constructs used in this work.

Plasmid constructs used in this work and referred to in the main text are listed below.

Plasmid name	Description	Source
For *B. subtilis*		
pTM13	*amyE::Phyperspank-yqgP (spec)*	This work
pTM26	*amyE::Phyperspank-yqgP-sfGFP (spec)*	This work
pTM31	*amyE::Phyperspank-yqgP S288A(spec)*	This work
pTM79	*ycgO::PmgtE-FLAG-mgtE (kan)*	This work
pTM89	*ycgO::Phyperspank-FLAG-mgtE (erm)*	This work
pTM92	*amyE::Pxyl (spec)*	This work
pTM93	*amyE::Pxyl-yqgP (spec)*	This work
pTM94	*amyE::Pxyl-yqgP S288A (spec)*	This work
pPR155	*amyE::Pxyl-yqgP Δ1-178 (spec)*	This work
pPR157	*amyE::Pxyl-yqgP Δ388-507 (spec)*	This work
pTM89	*ycgO::Phyperspank-FLAG-mgtE (erm)*	This work
pPR290	*xkdE::Pxyl-MBP-FLAG-pstatA I5G,I10G-Trx-HA (erm, lin)*	This work
pJB216	*amyE::Pxyl-yqgP D29A (spec)*	This work
pJB217	*amyE::Pxyl-yqgP D37A (spec)*	This work
pJB218	*amyE::Pxyl-yqgP H49A (spec)*	This work
pJB219	*amyE::Pxyl-yqgP D50A (spec)*	This work
pJB221	*amyE::Pxyl-yqgP D52A (spec)*	This work
pJB222	*amyE::Pxyl-yqgP D60A (spec)*	This work
pJB223	*amyE::Pxyl-yqgP E90A (spec)*	This work
pWX467	*loxP-erm-loxP casette*	Gift from D.Z. Rudner
pGP886	*xkdE::Pxyl-YFP-NTD (erm, lin)*	Gift from L. Krásný
For *E. coli*		
pJB171	*yqgP 1-177-TEV-6×His in pET25b*	This work
pJB179	*6×His-TEV-yqgP 384-507 in pHIS2*	This work
pJB184	*GST-6×His-TEV-mgtE 2-275 in pGEX6P1*	This work
pBM9	*6×His-yqgP 1-170 in pRSET-A (ThermoFisher)*	This work
pBM11	*6×His-yqgP 385-507 in pRSET-A*	This work

Table EV2: List of *Bacillus subtilis* strains used in this work

Bacillus subtilis strains that were used in this work and referred to in the main text are listed below. BGSC, Bacillus Genetic Stock Center at www.bgsc.org; NBRP, National BioResource Project at www.nbrp.jp;

Strain ID	Genotype	source
BTM2	*Wild type 168 trpC+*	BGSC, (Nicolas et al., 2012)
BTM70	*168 trpC+ amyE::Phyperspank-yqgP-sfGFP (spec)*	This work
BTM78	*168 trpC+ ΔyqgP::cat*	This work
BTM84	*168 trpC+ ΔyqgP::cat, amyE::Phyperspank-yqgP-sfGFP (spec)*	This work
BBM1	*168 trpC+ ΔyqgP::cat, amyE::Phyperspank-yqgP.S288A-sfGFP (spec)*	This work
BTM462	*168 trpC+ ftsH::erm*	BGSC, (Koo, Kritikos et al., 2017)
BTM501	*168 trpC+ ΔyqgP::cat amyE::Phyp-yqgP (spec)*	This work
BTM502	*168 trpC+ ΔyqgP::cat amyE::Phyp-yqgP S288A (spec)*	This work
BTM610	*168 trpC+ ΔyqgP::cat (spec) ycgO::Phyp-FLAG-mgtE (erm)*	This work
BTM611	*168 trpC+ ΔyqgP::cat amyE::Phyp-yqgP (spec) ycgO::Phyp-FLAG-mgtE (erm)*	This work
BTM612	*168 trpC+ ΔyqgP::cat amyE::Phyp-yqgP S288A (spec) ycgO::Phyp-FLAG-mgtE (erm)*	This work
BTM659	*168 trpC+ ΔyqgP::erm*	
BTM677	*168 trpC+ ΔyqgP*	This work
BTM799	*168 trpC+ ΔyqgP ftsH::erm*	This work
BTM795	*168 trpC+ ΔyqgP::cat ftsH::erm*	This work
BTM796	*168 trpC+ ΔyqgP::cat amyE::Phyperspank-yqgP (spec) ftsH::erm*	This work
BTM797	*168 trpC+ ΔyqgP::cat amyE::Phyperspank-yqgP S288A (spec) ftsH::erm*	This work
BTM843	*168 trpC+ ywlD::erm*	This work
BTM844	*168 trpC+ ΔywlD::erm ΔyqgP::cat*	This work
BTM845	*168 trpC+ ΔywlD::erm ΔyqgP::cat amyE::Phyperspank-yqgP (spec)*	This work
BTM872	*168 trpC+ ΔyqgP amyE::Pxyl-yqgP Δ1-178 (spec) ftsH::erm*	This work
BTM873	*168 trpC+ ΔyqgP amyE::Pxyl-yqgP Δ388-507 (spec) ftsH::erm*	This work
BTM874	*168 trpC+ ΔyqgP amyE::Pxyl-yqgP (spec) ftsH::erm*	This work
BTM875	*168 trpC+ ΔyqgP amyE::Pxyl-yqgP S288A (spec) ftsH::erm*	This work
BTM886	*168 trpC+ ΔyqgP amyE::Pxyl-yqgP Δ1-178, S288A (spec) ftsH::erm*	This work
BTM888	*168 trpC+ ΔyqgP amyE::Pxyl-yqgP Δ388-507, S288A (spec) ftsH::erm*	This work
BTM1001	168 trpC+ *ΔywlD::erm ydcA::kan*	This work
BTM1003	168 trpC+ *ΔywlD::erm ydcA::kan yqgP::cat*	This work
BTM1005	168 trpC+ *ΔywlD::erm ydcA::kan yqgP::cat amyE::Phyperspank-yqgP (spec)*	This work
BS50	*168 trpC+ ΔyqgP::cat amyE::Phyp-yqgP (spec) ΔlysA::erm*	This work
BS51	*168 trpC+ ΔyqgP::cat amyE::Phyp-yqgP S288A (spec) ΔlysA::erm*	This work
BS72	*168 trpC+ ΔyqgP amyE::Pxyl-yqgP (spec)*	This work
BS73	*168 trpC+ ΔyqgP amyE::Pxyl-yqgP S288A (spec)*	This work
BS55	*168 trpC+ ΔyqgP amyE::Pxyl-yqgP Δ1-178 (spec)*	This work
BS57	*168 trpC+ ΔyqgP amyE::Pxyl-yqgP Δ388-507 (spec)*	This work
BS184	*168 trpC+ ΔyqgP xkdE::Pxyl-MBP-FLAG-TatA I5G,I10G-Trx-HA (erm, lin)*	This work
BS185	*168 trpC+ ΔyqgP amyE::Pxyl-yqgP d1-178 (spec) xkdE::Pxyl-MBP-FLAG-psTatA I5G,I10G-Trx-HA (erm, lin, lin)*	This work
BS186	*168 trpC+ ΔyqgP amyE::Pxyl-yqgP d388-507 (spec) xkdE::Pxyl-MBP-FLAG-psTatA I5G,I10G-Trx-HA (erm, lin)*	This work
BS187	*168 trpC+ ΔyqgP amyE::Pxyl-yqgP (spec) xkdE::Pxyl-MBP-FLAG-psTatA I5G,I10G-Trx-HA (erm, lin)*	This work
BS196	*168 trpC+ ΔyqgP amyE::Pxyl-yqgP D29A (spec) xkdE::Pxyl-MBP-FLAG-psTatA I5G,I10G-Trx-HA (erm, lin)*	This work
BS197	*168 trpC+ ΔyqgP amyE::Pxyl-yqgP D37A (spec) xkdE::Pxyl-MBP-FLAG-psTatA I5G,I10G-Trx-HA (erm, lin)*	This work
BS198	*168 trpC+ ΔyqgP amyE::Pxyl-yqgP H49A (spec) xkdE::Pxyl-MBP-FLAG-psTatA I5G,I10G-Trx-HA (erm, lin)*	This work
BS199	*168 trpC+ ΔyqgP amyE::Pxyl-yqgP D50A (spec) xkdE::Pxyl-MBP-FLAG-psTatA I5G,I10G-Trx-HA (erm, lin)*	This work
BS201	*168 trpC+ ΔyqgP amyE::Pxyl-yqgP D52A (spec) xkdE::Pxyl-MBP-FLAG-psTatA I5G,I10G-Trx-HA (erm, lin)*	This work
BS202	*168 trpC+ ΔyqgP amyE::Pxyl-yqgP D60A (spec) xkdE::Pxyl-MBP-FLAG-psTatA I5G,I10G-Trx-HA (erm, lin)*	This work
BS203	*168 trpC+ ΔyqgP amyE::Pxyl-yqgP E90A (spec) xkdE::Pxyl-MBP-FLAG-psTatA I5G,I10G-Trx-HA (erm, lin)*	This work
BKE23380	*168 trpC2 ΔlysA::erm*	BGSC

Table EV3: NMR restraints and structural statistics for the final water-refined set of YqgPNTD domain structures.

Parameters were generated by the Protein Structure Validation Software suite, as specified in the Methods section.

Non-redundant distance and angle constrains	
Total number of NOE restraints	2517
Short-range NOEs	
Intra-residue (i = j)	290
Sequential (| i - j | = 1)	690
Medium-range NOEs (1 < | i - j | < 5)	542
Long-range NOEs (| i - j | ≥ 5)	995
Torsion angles	300
Hydrogen bond constraints	-
Total number of restricting constraints	2817
Total number of restricting constraints per restrained residue	5.4
Residual constraint violations	
Distance violations per structure	
0.1 – 0.2 Å	3.63
0.2 – 0.5 Å	1.63
> 0.5 Å	0
r.m.s. of distance violation per constraint	0.01 Å
Maximum distance violation	0.49 Å
Dihedral angle viol. per structure	
1 – 10 °	2.93
> 10 °	0
r.m.s. of dihedral violations per constraint	0.33°
Maximum dihedral angle viol.	5.0°
Ramachandran plot summary	
Most favoured regions	94.6 %
Additionally allowed regions	4.9 %
Generously allowed regions	0.4 %
Disallowed regions	0.0 %
r.m.s.d. to the mean structure	*all/ordered*[1]
All backbone atoms	3.0/1.0 Å
All heavy atoms	3.5/1.5 Å
PDB entry	**6R0J**
BMRB accession code	**34376**

Table EV4: List of proteins structurally similar to $YqgP_{NTD}$.

The top 20 PDB matches from a Dali server search are displayed and sorted by the Z-score, which indicates a significant similarity match when higher than 2. The quality of the alignment is further measured by RMSD (Root-Mean-Square deviation), L_{ali} (alignment length, or number of structurally equivalent residues), nres (number of aligned protein residues), and %id (sequence identity).

No	Name	Z-score	RMSD	L_{ali}	nres	%id	PDB entry
1	Nt.BspD6I nicking endonuclease from *Bacillus*	6.8	4.2	122	594	11	4wl5
2	DNA-directed RNA polymerase II- largest subunit from *Saccharomyces cerevisiae*	6.5	3.5	111	215	8	1i3q
3	Human DNA-directed RNA polymerase II- subunit RPB1	6.1	3.5	107	210	9	5iy6
4	Restriction endonuclease FokI from *Planomicrobium okeanokoites*	5.8	3.8	116	568	7	1fok
6	Sporulation specific protein 16 from *Saccharomyces cerevisiae*	5.5	3.5	86	197	9	6bzf
5	Zinc-dependent Dnase from *Escherichia coli* EC869	5.5	2.9	95	213	5	4g6u
7	EndoMS endonuclease from *Thermococcus kodakarensis* (strain ATCC BAA-918)	5.4	3.4	91	239	13	5gke
8	Hypothetical protein AF1548 from *Archaeoglobus fulgidus*	5.4	3.1	104	184	6	1y88
9	Glycosyl transferase family 8 from *Anaerococcus prevotii*	5.3	4.8	101	234	8	3tzt
10	Protein VC1899 from *Vibrio cholerae*	5.2	3.3	106	380	7	1xmx
11	Cas4 nuclease SSO0001 from *Sulfolobus solfataricus*	5.2	9.9	84	206	6	4ic1
12	R.BspD6I-S subunit type IIS restriction endonuclease from *Bacillus sp.* D6	4.9	3.9	114	186	11	2p14
13	Restriction endonuclease SdaI from *Streptomyces diastaticus*	4.8	4	98	319	12	2ixs
15	Putative Aromatic Acid Decarboxylase from *P. aeruginosa*	4.8	4.1	97	502	9	4ip2
16	Putative decarboxylase FDC1 from *Aspergillus niger* (strain CBS513.88)	4.8	4.4	95	499	6	4za4
14	Restriction endonuclease BamHI from *Bacillus amyloliquefaciens*	4.8	4.5	106	212	5	1esg
17	3-octaprenyl-4-hydroxybenzoate decarboxylase (UbiD) from Escherichia coli	4.7	4.1	91	474	9	2idb
20	Cu,Mo-CO Dehydrogenase (CODH) from *Oligotropha carboxidovorans*	4.6	6	91	805	10	1n63
18	Uncharacterized protein YqeQ from *Escherichia coli*	4.6	4.9	99	180	9	3c0u
21	Human double-strand break repair protein Mre11A	4.6	3.5	78	386	9	3t1i

Supplementary material legends

Dataset S1: Complete protein list of the quantitative proteomics experiment displayed in Fig. 1A-C.

The MS data are available from the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) (Deutsch et al., 2017) via the PRIDE (Perez-Riverol et al., 2019) partner repository with the dataset identifier PXD014578.

Dataset S2: Complete protein list of the co-immunopurification experiment displayed in Fig. 1D-E.

The MS data are available from the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) (Deutsch et al., 2017) via the PRIDE (Perez-Riverol et al., 2019) partner repository with the dataset identifier PXD014566.

Dataset S3: Full NMR titration data of $YqgP_{NTD}$ by Mn^{2+}, excerpt of which is displayed in Fig. 6C.

Figures

Fig. 1

Fig. 2

Fig. 3

Fig. 4

Fig. 5

Fig. 6

Fig. 7

Fig. 8

Fig. EV1

Fig. EV2