Charles University

Faculty of Science

Study programme: Microbiology

Mgr. Jakub Began

Rhomboid family intramembrane proteases in prokaryotes: mechanism, substrate repertoires and biological functions in the Gram-positive bacterium *Bacillus subtilis*.

Intramembránové proteasy z rodiny rhomboidů v prokaryotech: mechanismus, repertoáry substrátů a biologické funkce u Gram-positivní bakterie *Bacillus subtilis*.

Doctoral thesis

Supervisor: Ing. Kvido Stříšovský, Ph.D.

Prague, 2019

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně a že jsem řádně uvedl a citoval všechny použité prameny a literaturu. Současně prohlašuji, že tato práce ani její podstatná část nebyla využita k získání jiného nebo stejného akademického titulu.

Declaration:

I hereby declare that I have written this Doctoral Thesis independently and that I have cited all sources and literature. This thesis has not been submitted for any other degree or purposes.

December 13th 2019, Prague

Mgr. Jakub Began

Declaration of co-authors:

I confirm that this Doctoral Thesis was written by Mgr. Jakub Began, and that it is based on work generated during his Ph.D. study in between 2012-2019. I confirm that Jakub's contribution to the reported publications fully corresponds with his specification given in the thesis. In particular, I would like to stress that Jakub has generated most of the data reported in the main body of the thesis and his accepted first-author publication on YqgP in *B. subtilis*.

December 13th, 2019, Prague

Kvido Stříšovský, Ph.D.

ACKNOWLEDGEMENTS

(In Slovak language)

Na tomto mieste by som rád poďakoval všetkým, ktorí so mnou spolupracovali a podporovali ma počas celej doby doktorského štúdia.

V prvom rade ďakujem Dr. Stříšovskému za to, že mi umožnil pracovať pod jeho odborným vedením. Mojim kolegom, či už z našej laborky alebo spolupracovníkom z iných tímov za ich prácu, cenné rady a komentáre a celkové spolunažívanie počas štúdia.

Ďakujem mojej manželke Anne za to, že ma podporovala vo chvíľach, kedy mi to nešlo, za to že ma vždy vypočula, pochválila, rozosmiala a upokojila milým slovom.

Mojej rodine ďakujem za chvíle, kedy som s nimi mohol odpočívať a naberať energiu do ďalšej práce. Predovšetkým mojej mame Márii, ktorá sa o mňa tak dobre starala a vychovala ma.

Mojim priateľom za to, že si vypočuli moje starosti a dokázali ma povzbudiť.

ABSTRACT

Rhomboid proteases are a class of serine intramembrane proteases, a large family of enzymes that catalyze the proteolytic cleavage of membrane proteins within their transmembrane regions, in the hydrophobic environment of cellular lipid membranes. Rhomboid proteases were discovered in 2001 in *Drosophila*. In their pioneering study, Lee *et al*. identified the essential role of Rhomboid-1 protein (Rhom-1), which proteolytically activates the epidermal growth factor (EGF) receptor signaling pathway, in the early stages of fly eye development. Members of the rhomboid superfamily – active proteases (rhomboids) as well as their catalytically-dead counterparts (rhomboid-like proteins, including iRhoms and Derlins) - are widely conserved, implying their biological significance. Rhomboids are present in all kingdoms of life from archea to humans, while proteolytically inactive rhomboid-like proteins are present in eukaryotes only. Rhomboid superfamily proteins play roles in a wide range of processes, as diverse as signaling in metazoan development, mitochondrial biogenesis in yeast, host-cell invasion by protozoan parasites, protein quality control in the endoplasmic reticulum (ER) or bacterial quorum sensing. Rhomboids are the best understood intramembrane proteases from a structural and mechanistic points of view. Most of the work has been done on the rhomboid protease GlpG from the Gram-negative bacterium *Escherichia coli.*

The thesis focuses on the mechanistic characterization of the intramembrane rhomboid protease GlpG from the Gram-negative bacterium *Escherichia coli* and on the identification of biological role of YqgP from the Gram-positive bacterium *Bacillus subtilis.* Based on genetic analyses, GlpG-like and YqgP-like proteases are highly populated among bacterial rhomboid proteins and are also present in several pathogens such as Gram-negative *Salmonella* or *Shigella* and Gram-positive *Listeria* or *Staphylococcus.* To refine our knowledge of the mechanistic principles covering substrate specificity, we have mapped the amino acid preferences of the GlpG rhomboid protease and developed its substrate-derived inhibitors. The tools that we developed for GlpG rhomboid, were subsequently used to characterise the biological function of YqgP. We analysed the degradome and the interactome of YqgP in *Bacillus subtilis in vivo* and identified MgtE, the main magnesium transporter in *Bacillus subtilis*, as the natural substrate of YqgP. Finally, we showed that YqgP cooperates with the membrane embedded AAA+ protease FtsH during membrane protein quality control in *Bacillus subtilis*, representing an ancestral membrane protein degradation system conceptually similar to the eukaryotic ER associated degradation.

ABSTRAKT (Czech)

Proteázy z rodiny rhomboidů patří do rozsáhlé skupiny serinových intramembránových proteáz, které katalyzují proteolytické štěpení membránových proteinů uvnitř jejich transmembránových oblastí, v hydrofobním prostředí lipidických buněčných membrán. Rhomboidové proteázy byly objeveny v roce 2001 v *Drosophila*. V průkopnické studii (Lee *et al.*) byla identifikována zásadní role Rhomboidu-1 (Rhom-1), který v rané fázi vývoje oka octomilky aktivuje signální dráhu receptoru epidermálního růstového faktoru. Rhomboidové proteázy, ať už aktivní proteázy (rhomboidy) či jejich katalyticky neaktivní protějšky (rhomboidové proteiny zahrnující iRhomy a Derliny), jsou silně konzervovány, což naznačuje jejich biologickou významnost. Rhomboidy jsou přítomné napříč živočišnými říšemi, od archea po člověka, zatímco proteolyticky neaktivní rhomboidové proteiny jsou přítomné výlučně v eukaryotních organizmech. Rodina rhomboidových proteinů hraje roli v široké škále rozmanitých biologických procesů, jakými je signalizace ve vývoji metazoí, mitochondriální biogenezi kvasinek, invaze protozoálních parazitů do hostitelských buněk, kvalitativní kontrola proteinů v endoplazmatickém retikulu (ER), či bakteriální quorum sensing. Ze strukturního i mechanistického hlediska jsou rhomboidy nejprostudovanějšími z intramembránových proteáz. Nejvíce práce bylo provedeno na rhomboidu GlpG z Gram-negativní bakterie *Escherichia coli*.

Tato práce se zaměřuje na mechanistickou charakterizaci intramembránové proteázy GlpG z Gram-negativní bakterie *Escherichia coli* a objasnění biologické role YqgP z Gram-pozitivní bakterie *Bacillus subtilis*. Na základě genetických analýz jsou proteázy podobné GlpG a YqgP hojně zastoupeny mezi bakteriálními rhomboidovými proteiny a jsou taktéž přítomné v některých patogenních organizmech, kterými jsou Gram-negativní *Salmonella* a *Shigella* či Gram-pozitivní *Listeria* a *Staphylococcus*. Abychom rozvinuli naše poznatky o mechanistických principech substrátové specifity, zmapovali jsme aminokyselinové sekvence upřednostňované rhomboidem GlpG a vyvinuli také od substrátu odvozené inhibitory. Nástroje, které jsme vyvinuli při studiu rhomboidu GlpG, byly následně použity pro charakterizaci biologické funkce YqgP. Zanalyzovali jsme *in vivo* degradom a interaktom YqgP v *Bacillus subtilis* a identifikovali MgtE, hlavní transporter hořčíku v *Bacillus subtilis*, jako přirozený substrát YqgP. Nakonec popisujeme spolupráci YqgP s AAA+ proteázou FtsH zabudovanou v membráně během kvalitativní kontroly membránových proteinů v *Bacillus subtilis*.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

6xhis	- peptide tag composed of six histidines
AAA+	- a shortcut for "ATPase associated with diverse cellular activities"
AarA	- rhomboid protease from *Providencia stuartii*
ADAM	- a disintegrin and metalloprotease
ATP	- adenosine triphosphate
Bis	- N,N'-Methylenebisacrylamide
BLAST	- in bioinformatics, acronym for "basic local alignment search tool" algorithm
CCD	- charged-coupled device
Cdc48	- cell division protein 48
CMK	- chloromethylketone
CNRS	- Centre national de la recherché scientifique
CorA	- magnesium transporters family ubiquitous in Bacteria and Archea
DDM	- n-Dodecyl β-D-Maltoside detergent
SDS	- Sodim Dodecyl Sulfate
Der	- Derlin
Dfm1	- yeast derlin homolog
DNA	- deoxyribonucleic acid
Doa10	- ubiquitin ligase in yeast, component of ERAD
EDTA	- 2,2′,2″,2‴-(Ethane-1,2-diyldinitrilo)tetraacetic acid
EGF	- epidermal growth factor
EGFR	- epidermal growth factor receptor
EGTA	- ethylene glycol-bis (β-aminoethyl ether)-*N*,*N*,*N*′,*N*′-tetraacetic acid
ER	- Endoplasmic reticulum
ERAD	- Endoplasmic reticulum- associated degradation
ERAD-C	- Endoplasmic reticulum- associated degradation of cytosolic proteins
ERAD-L	- Endoplasmic reticulum- associated degradation of ER-luminal proteins
ERAD-M	- Endoplasmic reticulum- associated degradation of transmembrane proteins
FERM	- domain containing 4.1 protein, ezrin, moesin and radixin
FLAG	- peptide tag composed of residues DYKDDDDK
FPLC	- fast protein liquid chromatography
FRMD8	- FERM domain-containing protein 8
FtsH	- ATP, zinc-dependent metalloprotease
GeLC-MS/MS	- analysis of protein samples from polyacrylamide gel using liquid chromatography coupled to tandem mass spectrometry
GlpG	- rhomboid protease from *Escherichia coli*
gp78	- ubiquitin ligase
HA tag	- peptide tag derived from Human influenza hemagglutinin composed of YPYDVPDYA
HEPES	- 2-[4-(2-hydroxyethyl) piperazin-1-yl]ethanesulfonic acid
Hrd1	- E3 ubiquitin- protein ligase
HRP	- horseradish peroxidase
HSQC	- heteronuclear single quantum coherence spectroscopy
IOCB AS CR	- Institute of Organic Chemistry and Biochemistry, Academy of Sciences of CR
IPTG	- Isopropyl β-D-1-thiogalactopyranoside
LB	- shortcut for "Luria-Bretani"
IRHD	- iRhom homology domain
iRhom	- inactive rhomboid
iTAP	- synonym to FRMD8 protein, iRhom tail-associated protein
ITC	- isothermal titration calorimetry
k_{cat}	- substrate turnover number
K_M	- Michaelis constant
LacY	- lactose permease from *Escherichia coli*
LacYtm2	- second transmembrane domain of LacY
LUCA	- last universal common ancestor

lyso-PC	- 2-acyl-sn-glycero-3-phosphocholines
MBP	- maltose-binding protein from *Escherichia coli*
MdfA	- multidrug transporter in *Escherichia coli*
MES	- 2-morpholin-4-ylethanesulfonic acid
MgtE	- magnesium transporter
mβCD	- methyl-β-cyclodextrin
Ni-NTA	- nickel-charged 2,2',2''-Nitrilotriacetic acid
NMR	- nuclear magnetic resonance
NOE	- Nuclear Overhauser effect
NOESY	- two-dimensional NOE
NTD	- N-terminal domain
OD_{600}	- optical density at 600 nm
p97	- AAA ATPase
PARL	- shortcut for "Presenilins-associated rhomboid-like"
PBS	- phosphate buffer saline
PCR	- polymerase chain reaction
PDB	- protein data bank
PfROM	- rhomboid protease from *Plasmodium falciparum*
PVDF	- polivinylidene difluoride
RHBDD	- rhomboid domain-containing protein
RHBDL	- rhomboid-related protein
RNA	- ribonucleic acid
SP6	- SP6 phage
S2P	- site-2 protease
SDS-PAGE	- sodium dodecyl sulphate polyacrylamide gel electrophoresis
SecDF/YajC	- bacterial protein translocation complex
SecYEG	- bacterial protein translocation complex
SILAC	- stable isotope labelling by amino acids in cell culture
SLC41	- solute carrier family 41
SPP	- signal peptide peptidase
S-tag	- peptide tag derived from pancreatic ribonuclease A composed of KETAAAKFERQHMDS
TACE	- or ADAM17; tumor necrosis factor-alpha converting enzyme
TAE	- Tris-acetate-EDTA buffer
TatA	- subunit A of bacterial protein translocation machine
TBE	- Tris-borate-EDTA
TBS	- Tris buffer saline
TEB4	- or Doa10, E3 ubiquitin protein ligase
TEMED	- N, N, N', N'-tetramethylethane-1,2-diamine
TEV	- tobacco etch virus
TgROM	- rhomboids from *Toxoplasma gondii*
TM	- shortcyt for "tranmembrane"
TMEM115	- transmembrane protein 115 homologous to rhomboid
TOCSY	- total correlated spectroscopy
Tris	- tris(hydroxymethyl)aminomethan
TRPM7	- transient receptor potential cation channel, subfamily 7
Trx	- Thioredoxin from *Escherichia coli*
UBA	- shortcut for "ubiquitin-associated"
UBAC2	- ubiquitin-associated domain-containing protein 2
VBM	- VCP-binding motif
VCP	- Valosin-containing protein
YidC	- bacterial membrane protein insertase
YqgP	- rhomboid protease from *Bacillus subtilis*

1 INTRODUCTION

1.1 Membrane protein homeostasis

Membranes are subcellular compartments with specific physicochemical properties. These include membrane polarity, planarity, and fluidic mosaicity, diffusion control of reaction rates and heterogeneous lipid and protein composition. Thus, the majority of events that take place in any cellular membrane have special mechanistic requirements, as compared to those happening in the cytosol or other solvent accessible environment. Membranes present a mechanical barrier, but also provide the scaffold for and regulate a plethora of fundamental physiological functions as diverse as signal transduction, nutrient uptake or energy metabolism. These are enabled by topologically and functionally diverse membrane proteins, such as receptors, transporters or enzymes.

Protein levels are closely regulated during the whole protein life-time and protein quality is tightly controlled immediately after it leaves the translation machinery. To eliminate prematurely translated proteins or proteins misfolded during protein maturation, and to ensure correct timing of activation or inactivation of particular signal transduction pathways, cells exploit diverse post-translational modifications, including protein degradation. Here, I focus on regulated proteolysis coupled with proteasome or proteasome-like complexes and on intramembrane proteolysis, two distinct mechanisms that, however, may often be interlinked during regulation of protein homeostasis (proteostasis) [1].

1.1.1 Eukaryotic membrane protein quality control

In eukaryotes, most membrane and secretory proteins are generated at the endoplasmic reticulum (ER). The endoplasmic reticulum-associated degradation (ERAD), which is tightly coordinated with protein translation, translocation and folding, is the major and complex mechanism used to control the quality of newly translated transmembrane and secretory proteins. It protects cells from the accumulation of misfolded or non-functional protein variants in the secretory pathway. Transmembrane proteins destined for degradation by ERAD are recognized by incompletely understood machinery, extracted from the membrane (retrotranslocated) by ATP-dependent machinery and tagged by ubiquitin (Ub) for targeting to the ubiquitin-proteasome system in the cytosol.

More specifically, ERAD is initiated by recognition and polyubiquitylation of the misfolded protein portion by membrane E3 ubiquitin ligase (for example Hrd1 and Doa10 ligases in yeast; HRD1-SEL1L, gp78 and a number of poorly described transmembrane E3 ligases in mammals) [2]. Based on the protein part that is recognized, three ERAD branches may be exploited: ERAD-L for luminal, ERAD-

M for membrane and ERAD-C for cytosolic part recognition, respectively [3]. The process continues with the energy-dependent activity of the membrane anchored AAA+ ATPase molecular segregase known as Cdc48 in yeast, or p97/VCP in mammals [4] and its substrate-specific protein adaptors that together bind Ub-tagged protein, pull it out of the ER membrane and direct it to the proteasome complex [5, 6].

1.1.2 Bacterial membrane protein quality control

In bacteria, translation of proteins targeted to the plasma membrane is coupled with membrane insertion mediated by the ubiquitous SecYEG translocon or the YidC insertase (or both). The SecDF/YajC complex is also involved as it ensures that the nascent polypeptide chains are incorporated with correct topology. Depending on the type of protein inserted, these assemblies can act independently or cooperate in dynamic transient higher-order holo-complexes [7-10].

The FtsH protease is a highly conserved transmembrane AAA+ type ATPase/protease, known to be involved in quality control of membrane protein folding (reviewed in [11]). To date, FtsH of *Escherichia coli* is the most well studied. It specifically reduces levels of properly folded active proteins with diverse cellular functions ranging from lipopolysaccharide biosynthesis, heat shock response, or cell division [11]. FtsH is a homohexameric zinc processive metalloprotease with ATP-dependent unfoldase activity. Each monomer has a large globular cytosolic protease domain tethered to the plasma membrane via two transmembrane helices. FtsH recognizes specific unstructured sequence elements termed as degrons at either termini or even at internal site of its substrates, then binds substrate to the pore at the hexamer's interface. This actively disrupts the substrate tertiary structure and dislocates it from the membrane upon ATP-hydrolysis. Adaptor proteins may be required for initial substrate recognition by FtsH and these define the specificity and act as a scaffold for FtsH-substrate interaction. In *E. coli*, FtsH exists as a holo-enzyme, and forms a large megadalton complex with the HflKC hexamer [12]. In *Bacillus subtilis*, a model Gram-positive bacterium, FtsH also exists as part of a higher-order multimeric protein complex, residing in flotillin-rich membrane microdomains [13, 14] and it is involved in cell division, sporulation and biofilm formation [15-18].

1.1.3 Regulated intramembrane proteolysis

The most distinguishing feature of all intramembrane proteases differentiating them from the soluble or membrane-tethered proteases is their ability to catalyse the hydrolysis of the polypeptide chain within the hydrophobic environment of the lipid membrane where water molecules, which are

necessary for the catalysis, are not normally present [19]. Site-specific intramembrane cleavage is usually coupled either with release of physiologically relevant protein ectodomain regulating a signalling pathway, or it leads to controlled degradation of the protein. In either case, intramembrane protease substrate is cut within the transmembrane domain or in its close vicinity - in the juxtamembrane region - and, depending on protease type, is direct or may require pre-cleavage of the substrate by another protease [20] (**Fig. 1**).

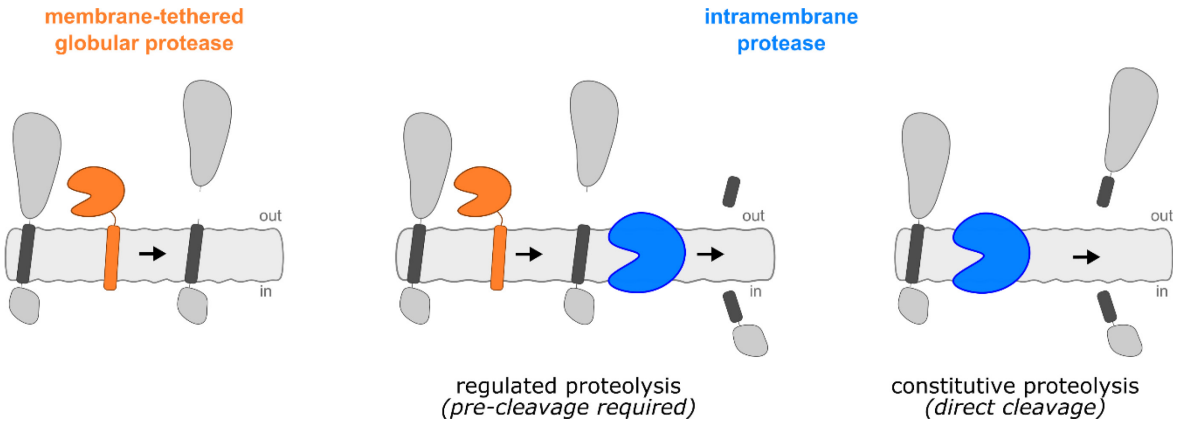


Fig. 1: Proteolytic events on the membrane. Single-pass or polytopic (not shown, for simplicity) transmembrane proteins can be processed in order to release a functional domain, or to direct the protein for degradation. Globular proteases (orange) attached to the membrane cleave their substrates outside the membrane, while intramembrane proteases (blue) shed their substrates within the hydrophobic lipid bilayer either directly or after a preceding proteolytic step performed by a soluble or another intramembrane (not shown) protease.

Intramembrane proteases are nearly ubiquitous and serve as regulators of signal transduction pathways (inside and outside the cell), and play a role in cellular development and differentiation, cell division or cell death. They also control parasitic invasion and pathogen virulence, are part of protein quality control complexes, as well as being involved in pathophysiological processes including a variety of cancer types, inflammation or neurodegenerative diseases (reviewed in [20-24]).

1.1.4 Intramembrane protease families

Based on their mechanisms, intramembrane proteases are traditionally classified into three families. These are i) the metalloproteases, exemplified by Site-2-protease (S2P) [25], which contains a zinc coordinating active site and was the first intramembrane proteolytic enzyme ever discovered. Next are ii) the aspartyl proteases, represented by presenilins and by signal peptide peptidases (SPP) [26-28], and iii) the serine rhomboid proteases [29], which contain an unconventional catalytic Ser-His dyad and which are the main focus of this thesis. More recently, a new class of intramembrane proteases, the glutamate-dependent proteases, has been identified. This is represented by conserved homologs of Rce1 (Ras converting enzyme 1), and has a novel catalytic mechanism employing Glu-His dyad [30].

Rhomboids and presenilins typically cleave type I single-spanning membrane proteins with N-termini facing the extracellular space, while S2Ps and SPPs cleave transmembrane helices of type II topology (**Fig. 2**).

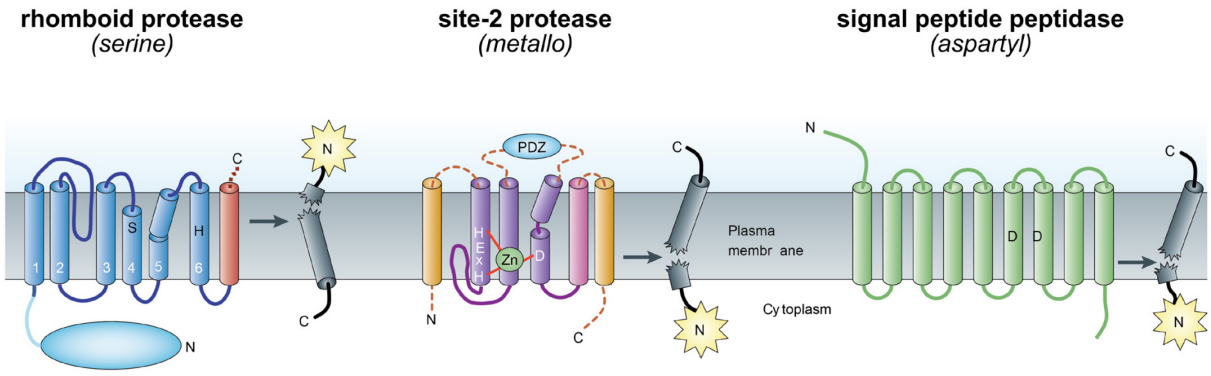


Fig. 2: Catalytic mechanism-based intramembrane protease families. Serine rhomboid proteases contain an unconventional conserved catalytic S-H dyad and cleave membrane proteins with type I topology. Metalloproteases, such as Site-2-protease (S2P) use their conserved HExxH active site motif to coordinate a divalent zinc cation, while signal peptide peptidases (SPP) employ a canonical catalytic DD dyad. Both metalloproteases and aspartyl proteases cleave type II-oriented proteins, with the exception being γ-secretase in the presenilin complex as this cleaves proteins with type I topology (not shown). Figure was adopted from [21] and modified.

Rce1 homologs recognize farnesyl motif within the protein sequence, but little is known about Rce1 substrate topology preferencies [30]. Rhomboid protease substrates do not require pre-cleavage step by other proteases, which indicates distinct mechanism of substrate recognition with implications for the associated biology [31].

1.2 Discovery of the rhomboid protein superfamily

Nearly 30 years ago, classical genetic analyses in *Drosophila melanogaster* identified the product of the *rhomboid* gene as one of the four key protein regulators controlling the early stages of photoreceptor cell differentiation during eye development. During the next ten years, several studies provided evidence that Rhomboid-1, acting together with Star, tightly controls the activation of the epidermal growth factor receptor (EGFR) signaling pathway. It emerged that the EGFR ligand Spitz is recruited by Star and transported from the ER into the Golgi aparatus where it is subsequently processed within its transmembrane region in the presence of Rhomboid-1 [32]. This proteolytically regulated trafficking leads to the effective shedding of the EGF ectodomain from the cell membrane and activation of the pathway. Any remaining doubts were dispelled when Urban, Lee and Freeman provided direct proof for proteolytic activity of Rhomboid-1, thus establishing Rhomboid-1 as the founding member of the serine intramembrane protease family [29, 33].

1.3 Evolution and topology of rhomboid superfamily

Phylogenetic analyses revealed that rhomboid-like genes are distributed throughout all three domains of life [34-39], further implying their biological significance. However, rhomboids share relatively low sequence identity (10-15%) [34], and it was suggested that rhomboids arose in bacteria and through several horizontal gene transfers were aquired by archea and in the early stage of their evolution also by eukaryotes [34]. More recent analysis [35], based on improved BLAST-based analysis combined with structural and functional data of all eukaryotic rhomboid proteins, together with extensive analyses of protozoan [36, 37] and plant rhomboids [38, 39], helped refine our knowledge on the rhomboid phylogeny and classification. It is speculated that rhomboids evolved and spread by primary vertical transfer of the ancestral six-transmembrane helix core-containing active protease from the last universal common ancestor (LUCA), followed by selective pressure-guided gene duplication and subsequent mutation events [35] (**Fig. 3**). There are three major conserved groups identified within the rhomboid protein superfamily (**Fig. 3**), which share a common six-transmembrane helix core (6 TM). These are i) active proteases termed rhomboids; ii) their inactive orthologs, or pseudoproteases, lacking catalytic residues but sharing structure-stabilizing conserved motifs, named iRhoms; and iii) other proteolytically inactive proteins that are homologous to rhomboids but topologically diverse and dissimilar to iRhoms. The latter two groups are jointly termed as rhomboid-like proteins, or rhomboid-like pseudoproteases (**Fig. 3**) [35].

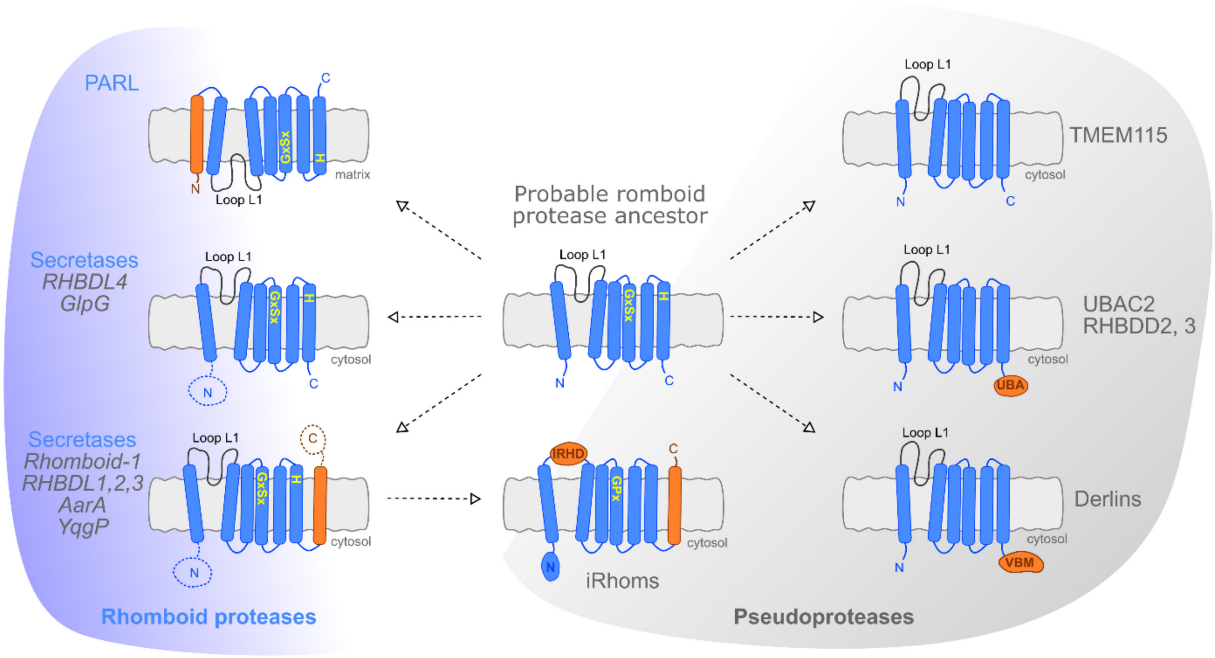


Fig. 3: Evolution and topology of rhomboid superfamily. Rhomboids (active rhomboid proteases) probably evolved from the rhomboid protease core ancestor protease of the last universal common ancestor (LUCA). It is widely accepted that inactive rhomboid proteins (rhomboid-like proteins, pseudoproteases) evolved by gene duplication of rhomboids. The subsequent "activity incompatible" loss-of-function mutations led to the conservation of their pseudoprotease activity. All members share the transmembrane core (in blue). An additional transmembrane domain (orange) is present in some secretase-like and all PARL-like rhomboids, as well as in iRhoms. Some rhomboid secretases also contain N-terminal (dashed blue circle) or C-terminal (dashed orange circle) soluble domains. All iRhoms contain soluble N-terminal domain and large iRhom homology domain (IRHD) inserted in loop L1. Other groups of inactive rhomboids contain characteristic protein-binding domains, such as ubiquitin associated (UBA) in TMEM115, or VCP-binding motif (VBM).

Active rhomboids can be further categorized as PARL type proteases, containing extra N-terminal transmembrane domain (1+6 TM) and residing mostly in mitochondria, and two groups of secretase type proteases that may contain an additional C-terminal transmembrane domain (6 TM and 6+1 TM) [35]. Some rhomboid proteins contain soluble globular domains at one or both termini (**Fig. 3**). Secretase type rhomboid proteases are localized in the secretory pathway. Interestingly, the significance of rhomboid protein conservation is underlined by the fact that multiple paralogs of rhomboid proteins, which probably arose from gene duplication events during evolution, are usually present within one organism, i.e. five rhomboid proteases present in mammals [35], six in *Plasmodium falciparum* [36], seven in *Drosophila*, or even thirteen in *Arabidopsis* [39]. The vast majority of prokaryotes contain one or two active rhomboid proteases.

Intriguingly, rhomboid pseudoprotease genes have been identified in eukaryotes only and are usually present in multiple, compartment-specific copies [35]. How 'inactive rhomboids' evolved is still being extensively discussed. Phylogenetic analyses are often interpreted to mean that in metazoans,

iRhoms emerged by gene duplications of active rhomboids. The proline introduced within the catalytic GxS motif have resulted in a "pseudoactive" site GPx motif, disrupting its architecture. One or the other or both of the catalytic Serine-Histidine dyad residues may have also been lost, though this is not always the case and in some iRhoms, the dyad has even been left unperturbed [35, 40]. Other highly conserved and characteristic topological features that iRhoms have acquired throughout evolution are an additional seventh transmembrane helix, a large N-terminal cytosolic domain and an unusual characteristic globular cysteine-rich domain called the iRhom homology domain (IRHD), connecting transmembrane helices 1 and 2 (**Fig. 3**). Notably, iRhoms are highly prevalent in metazoans, but absent from prokaryotes. A further distinct group of catalytically dead and evolutionarily diverse proteins of the rhomboid superfamily that probably evolved independently and earlier includes Derlin [41], TMEM115 [42], RHBDD2 and RHBDD3 [43, 44] and UBAC2 proteins [45] (**Fig. 3**).

1.4 Rhomboid mechanism and specificity

Any proteolytic processing requires the presence of a catalytic water molecule in the active site of the enzyme, to enable hydrolysis of the peptide bond. This occurs through nucleophilic attack and activation of water, directly (aspartic, glutamic and metalloproteases) or via the formation of a covalent acyl-enzyme intermediate, and its subsequent water-mediated hydrolysis (**Fig. 7**). Mechanistically, rhomboid intramembrane proteases represent quite a unique class of enzymes, since the catalytic cleavage takes place in the hydrophobic environment of the lipid bilayer.

Studies of divergent rhomboids from *Drosophila* (Rhomboid-1)*,* and bacterial species *Escherichia coli* (GlpG)*, Bacillus subtilis* (YqgP) and hyperthermophilic *Aquifex aeolicus* (Aq Rho) [46] suggest that, in contrast to other intramembrane protease classes such as aspartyl γ-secretase [47], rhomboid proteases can act alone, without accessory proteins or non-protein cofactors. Furthermore, rhomboids are not sensitive to metalloprotease, cysteine or aspartyl protease inhibitors, nor to most soluble serine protease inhibitors such as leupeptin or aprotinin, but they are sensitive to dichlorocoumarin moieties [46].

1.4.1 Rhomboid protease architecture

Initial analyses in *Drosophila* identified conserved sequence motives in Rhomboid-1, including the active site S217-H281 catalytic dyad, with serine positioned within the highly conserved $GAS_{217}GG$ region present also in soluble serine proteases, as well as additional residues involved in rhomboid

stability and activity. For the very first time, the intriguing concept of an active site region situated within the plane of the hydrophobic lipid bilayer was discussed [29, 48]. The significance of conserved residues was confirmed using *in vitro* activity assays and structural analyses of wild-type and single-point rhomboid mutants [49, 50].

Structural analyses of GlpG rhomboid orthologs from *Escherichia coli* and *Haemophilus influenzae*, crystallized either in detergent micelles [51-54] or in the presence of lipids (**Fig. 4**) [55] provided valuable information on the rhomboid conformation, underlined the importance of key architectural features and opened the door to detailed mechanistic studies of proteolysis by serine intramembrane proteases. The three-dimensional structures reveal a transmembrane core of GlpG that consists of six α-helices folded into a compact hydrophobic bundle around the central TM4 helix (**Fig. 4**).

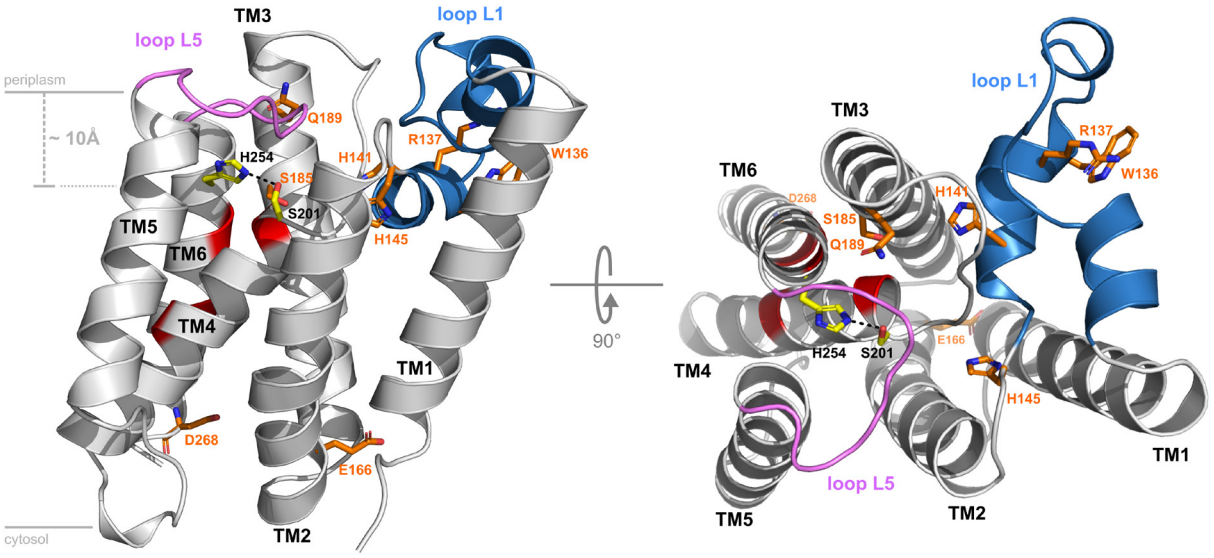


Fig. 4: Crystal structure of GlpG rhomboid from *Escherichia coli* (PDB code: **2xov**). The GlpG transmembrane core is formed of six alpha helices, which are, with the exception of the kinked central helix TM4, almost perpendicular to the lipid bilayer. The base of the active site pocket is buried approximately 10 Å in the plasma membrane. The key architectural features are highlighted using the following color coding: the catalytic S201-H254 dyad is depicted as yellow sticks; the conserved residues that form numerous non-covalent interactions and are essential for GlpG stability are represented by orange sticks; the conserved GxxxG motif, stabilizing helices TM4 and TM6 and positioning the active site cavity in correct conformation is in red; the loop L5 motif, capping the active site, is in magenta; loop L1, a floating device necessary for proper orientation within the membrane, is in blue.

Because the structures of no other rhomboid protease homologs have been solved to date, GlpG has become the main mechanistic and structural model of rhomboid-mediated intramembrane proteolysis. The active site pocket of GlpG is formed by a hydrophilic V-shaped cavity containing conserved residues H150, N154, G199, S201 and H254 of TM helices 2, 4 and 6, as well as residues of loops L3 and L5. The whole molecule is stabilized by numerous weak and several strong non-covalent

interactions. This extensive van der Waals network is orchestrated predominantly by the conserved parts of the molecule, in particular by the two highly conserved GxxxG interaction motifs linking the TM4 and TM6 helices (red in **Fig. 4**). These are crucial for proper orientation of the catalytic dyad. A further contribution is made by residues of the TM4 helix and of loop L3 and by the leucine-rich region of the L1 loop stacking with residues of helix TM3. The TM bundle is also stabilized by several hydrogen-bond rich regions, in particular by residue E166 linking helices TM1, TM2 and TM3 and residue D268 bringing together helices TM3, TM4 and TM6, both on the cytosolic side, and by residue R137 within the L1 loop on the periplasmic side (orange sticks in **Fig. 4**) [50]. The catalytic dyad consists of S201-H254. The serine sits on top of the short and kinked TM4, and is hydrogen-bonded to H254 within the neighboring TM6, and both residues are buried approximately 10 Å below the membrane-periplasm interface. Intriguingly, the active site pocket is accessible to the aqueous solvent at the extramembrane side and is laterally sequestered from the lipid bilayer by the surrounding helices [51-53]. The L1 loop is another important structural feature. It is formed by one α-helix and four 3_{10} helices. Its amphiphilic character allows its partial immersion into the membrane via a conserved W136-R137 motif which is, together with H141 and H145 (blue in **Fig. 4**), required for optimal activity [50, 56]. However, the L1 loop does not interact with the active site pocket and was thus proposed to act as a stabilization element, a floating device that keeps the whole molecule in proper orientation, rather than being directly involved in catalysis [56-58].

Molecular dynamics simulation studies proposed retention of water molecules in the hydrophilic cavity formed by residues H141, S181, S185 and Q189, all close neighbours of the catalytic S201, the so called 'water retention site' [59]. Notably, mutating Q189 and S185 residues negatively affected GlpG activity *in vitro* or *in vivo* but had no effect on its stability [60], indicating that this effect relied on affecting the water retention site function.

1.4.2 Interaction of rhomboid with substrate

The key questions that still remain to be answered are i) how substrate approaches the rhomboid active site, ii) what are the key rhomboid and substrate determinants that drive substrate recruitment and last but not least iii) to what extent does the substrate interact with the enzyme's active site. Firstly, I will focus on what has been learned from the structural studies of rhomboid proteases. Substrate features will be discussed afterwards.

The TM5 helix and loop L5, also referred to as an active site "cap" [61], both sterically hinder the active site from the lipid bilayer. In certain GlpG crystal structures [52, 53], the C-terminal half of the TM5 helix and of loop L5 are tilted away from the transmembrane core. The TM5 helix is lifted by 35° in a structure that resembles GlpG in an "open conformation" (**Fig. 5a**). Based on this data, the most probable way for the substrate to access the active site would be through a putative lateral gate formed between transmembrane helices TM2 and TM5. This gating mechanism hypothesis has been supported by an *in vitro* study, carried out in detergent micelles [56], and *in vivo* in *Escherichia coli* and *Drosophila* cells [62], as well as by molecular dynamics simulations [63]. Mutations of residues W236 and F153 at the TM2-TM5 interface, as well as of F245 of the L5 cap, described as "gating enhancers" (**Fig. 5b**), improve rhomboid activity significantly. This agrees with the structural and computational data, and indicates a relatively high degree of conformational flexibility within the TM5-L5 region [52, 64] and its less significant contribution to overall GlpG stability [50]. Crosslinking of residues at positions 236 and 153 after their mutation to cysteines abrogates proteolytic activity of GlpG when using a zero-length crosslinker [56], while it preservers enzymatic activity when using crosslinker of defined linker arm length such that it does not disturb the distance of TM2 and TM5 [65], making it difficult to draw any conclusions.

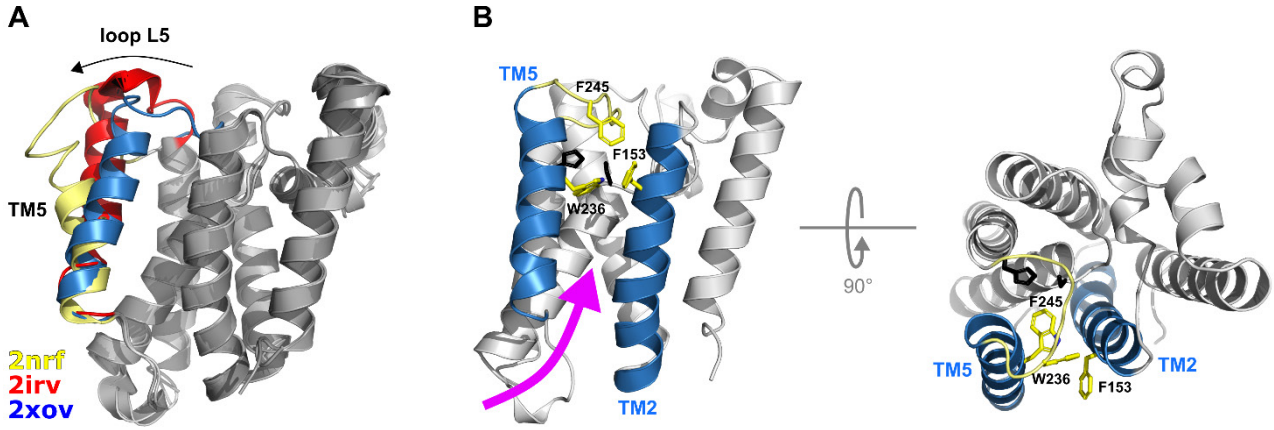


Fig. 5: Structural insights into the mechanism of rhomboid substrate entry. ***(A)*** Structural alignment of three crystal structures of E. coli GlpG (PDB codes: 2nrf, 2irv, 2xov) illustrates the movement of helix TM5 and loop L5 region away from the catalytic centre (uncapping). Each GlpG molecule is coloured in shade of grey with TM5+L5 region highlighted yellow (2nrf), red (2irv) and blue (2xov), respectively. However, extreme kinking of the TM5+L5 region in yellow structure may be crystallization artefact. **(B)** Crystal structure of GlpG (2xov) with region between helices TM2 and TM5 (both in blue), or substrate exosite, where substrate meets rhomboid (curved magenta arrow), prior to proteolysis step. Residues that probably contribute to substrate gating are represented by yellow sticks. Catalytic dyad is depicted in black.

Despite strong evidence for gating within TM2-TM5 region, subsequent structural studies of GlpG-inhibitor complexes readjusted our view of the substrate entry mechanism [65-69]. They

postulate that the TM5 helix, in the extensively lifted conformation, may be an artefact of the crystallization conditions. Currently, it is largely believed that rather than complete immersion by rhomboid via lateral gating, the substrate TM helix instead meets GlpG at its lateral side, referred to as the "exosite" or the "interrogation site" [59, 70, 71]. It is localized probably within the TM5 region and after substrate binds, only the flexible loop L5 lifts away (**Fig. 5**). This enables the locally unwound, β-strand-resembling chain [72] of the substrate to access the active site from the top side of the enzyme [59, 63]. Hence, large TM5 rearrangements are not required, avoiding unfavourable exposure of the active site to the lipid environment.

At this point it is necessary to emphasize that the full resolution of the rhomboid-substrate complex's structure still remains the biggest challenge of the rhomboid mechanistic studies and is the only option for resolving the discrepancies discussed herein.

1.4.3 Specificity determinants in substrate

Soon after the discovery of Rhomboid, several studies shed light on how rhomboid proteases may recognize their substrates. Urban and Schlieper showed that distinct prokaryotic rhomboid genes (Gram-negative, Gram-positive eubacterial and archaeal), code for active rhomboid proteases that can specifically process single-membrane spanning *Drosophila* Rhomboid-1 (Rhom1) substrates, including Spitz, Gurken and Keren, and that rhomboid activity was sufficient to activate the EGFR signalling pathway [48]. Moreover, the expression of Rhomboid-1 also rescues the propagation of the quorum sensing signal in *Providencia stuartii* deficient for rhomboid AarA [73]. These results implied that structurally variable and evolutionarily diverse rhomboids, with very low sequence identity, shared certain key structural features and substrate specificity and that their physiological roles might have been conserved. Follow-up studies then helped understand the sequence and structural motifs that drive substrate recognition by rhomboid. Systematic characterization of artificial substrate-derived proteins cleaved by different eukaryotic and prokaryotic rhomboid proteases *in vivo* [31], or reconstituted *in vitro* [46, 49, 74, 75], led to the identification of the substrate elements that are necessary and sufficient to restore rhomboid-specific cleavage.

Substrates are defined by their transmembrane and juxtamembrane regions, required for hydrolysis by rhomboids, consisting of two separable elements: a linear consecutive sequence motif and a transmembrane domain (**Fig. 6a, d**). Rhomboid-1 recognizes a seven-residue sequence in Spitz transmembrane region, covering the scissile bond. This seven residue sequence contains small (alanine, glycine) and β-branched (threonine, isoleucine) residues, which tend to destabilize the usual α-helical fold, as well as hydrophilic amino acids at the luminal non-prime side of the scissile bond [31]. Similar

observations of GlpG hydrolysis of model transmembrane substrates *in vivo* suggested that the presence of α-helix destabilizers such as glutamine and proline within transmembrane region, or small negatively charged residues at the P1' position improve substrate cleavability (**Fig. 6b, c**) [76]. The most comprehensive analysis of substrate preference was provided through characterization of TatA, the native substrate of AarA rhomboid from *Providencia stuartii*. AarA requires small residues at the P1 position and large and hydrophobic residues at positions P4 and P2' (**Fig. 6c**). Moreover, the site of cleavage in TatA is defined by the sequence motif, which is N-terminal to the transmembrane helix-destabilizing element. This is true even if this sequence motif is distal to the membrane region (**Fig. 6c**) [75].

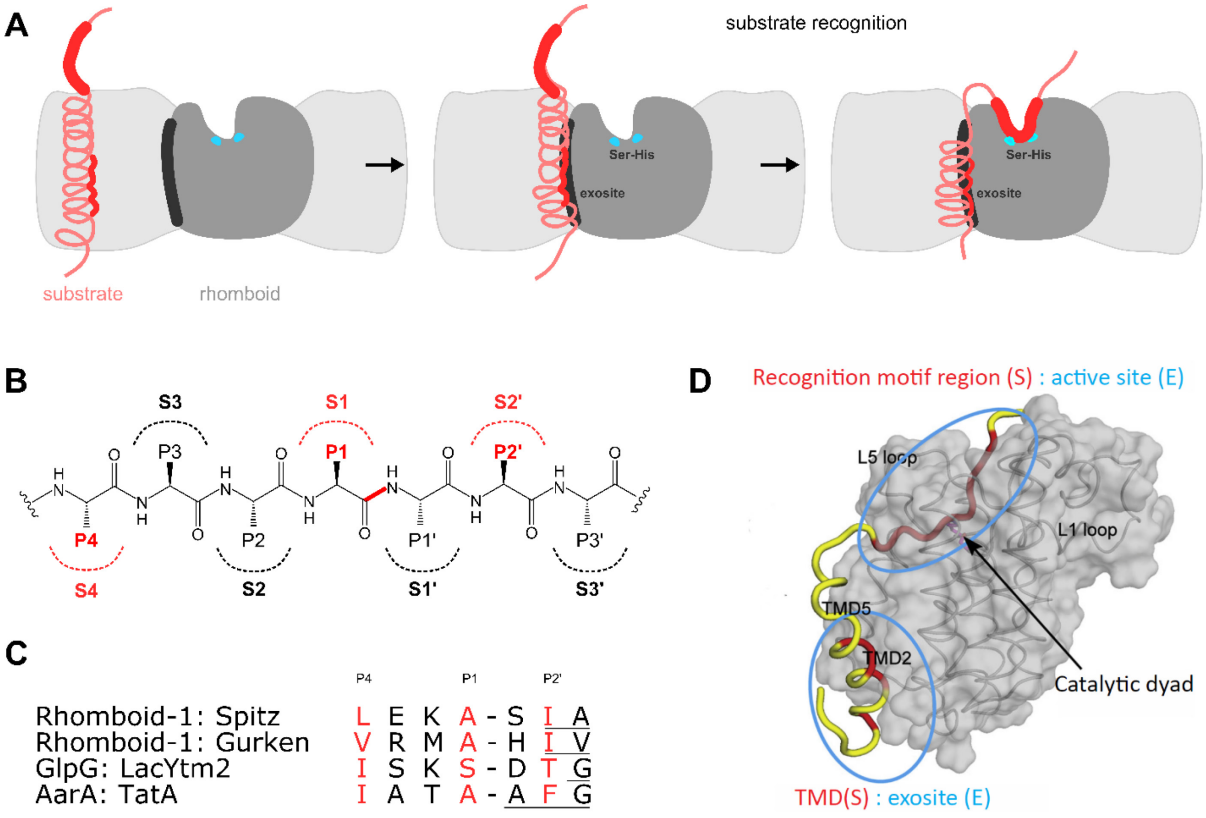


Fig. 6: Substrate recognition by rhomboid. (A) The substrate transmembrane domain (pink) is thought first interact via transmembrane motif (red) with the interrogation site (exosite, highlighted in black), formed by the TM2 and TM5 region of rhomboid. This step allows the substrate to locally unwind, kink and be accommodated in the rhomboid active site (blue dots), where the substrate is subsequently cleaved. **(B)** The residues preceding the scissile bond, in the N-terminal part of the substrate sequence, are marked as non-prime side (P). The residues C-terminal to the scissile bond are marked as prime side (P'). The enzyme subsites (S), which specifically bind substrate chain within the active site pocket are numbered correspondingly [77]. The scissile bond as well as the rhomboid-specificity defining subsite cavities and corresponding substrate residues are highlighted in red. **(C)** The sequence preferences of several rhomboids were characterised experimentally. Rhomboids recognize a seven-residue long substrate region, with small side chains at the P1 position and large and hydrophobic residues at positions P4 and P2' (red). The scissile bond is shown as a black dash and residues within the transmembrane region are underlined [74, 75, 78]. **(D)** A model of possible rhomboid (GlpG, grey) - substrate (Gurken, yellow) complex generated by molecular dynamic simulations[63]. The binding interfaces between both partners are highlighted in blue ovals. It includes 1) the TM region of the substrate (S) (red) interacting with the exosite of enzyme (E) formed by helices TM2 and 5 and loop L5, and 2) the recognition motif region (red) interacting with

the enzyme active site cavity. The catalytic dyad and loop L1 are highlighted. Figure D was adopted and modified from [79].

The transmembrane helix-destabilizing motif, defined as the kinked α-helix region, together with the physicochemical properties of the residues covering the sequence motif, are the two decisive recognition elements that drive the substrate- rhomboid interaction (**Fig. 6d**). Moreover, given the rhomboid and substrate diversity, both motifs can contribute to substrate recognition to a varying extent, depending on their mutual positions within the transmembrane and juxtamembrane regions. In other words, the closer the sequence motif to the TM-destabilizing motif the more significant its destabilizing contribution [31, 75].

Recent biophysical analysis suggests that helical instability is the key feature distinguishing rhomboid substrate from non-substrate and it is indeed shown that introducing a helix-breaking proline turns non-substrate sequences into substrates both *in vivo* and *in vitro* [71]. A combination of enzymological analysis with real-time kinetics measurement in proteoliposomes or in living cells also demonstrated that the affinity (K_M) of GlpG and other diverse rhomboids for any transmembrane helix tested (non-substrate, substrate or mutated substrate) is very low and more or less the same. Thus, the cleavage is driven kinetically, i.e. it relies on the optimal exposure time (k_{cat} parameter) for which substrate is docked within the exosite, until it is processed [70]. The turnover rate is also very slow and depends on the ability of the substrate to be accommodated and retained in the rhomboid exosite (**Fig. 6a, d**).

1.4.4 Understanding rhomboid catalytic mechanism

The development of *in vitro* activity assays and rhomboid-specific substrate-derived inhibitors have advanced the attempts to tease out the mechanistic details of rhomboid-substrate interactions, as reviewed extensively elsewhere [80]. The crystal structures of GlpG with three different classes of mechanism-based serine protease inhibitors shed a light on probable substrate conformational changes that occur upon binding to the active site. Reversible chloro-isocoumarins [69], irreversible diisopropyl-fluorophosphonates and phosphonofluoridates [67, 68], and the slowly reversible monocyclic β-lactams inhibitors [66], all mimic the carbonyl of the substrate scissile bond covalently bound to the catalytic Ser201, albeit each through a slightly different mechanism. Additionally, all of the inhibitors share strong hydrogen bonding with the Ser201 main chain and make several weaker hydrogen bonds with the residues facing the active site.

These key structural analyses confirmed the previous observation that opening of the L5 cap makes the active site more accessible and initiates the formation of the oxyanion hole, typical for all serine proteases [61]. Within the GlpG active site, the analyses also identified the S1 and S2' cavities [66, 67, 69], which accommodate substrate residues and define rhomboid specificity (**Fig. 6**). Notably, the structural analyses closely agree with the results of *in vitro* activity studies that indicate a crucial role for substrate residues P4, P1 and P2' during catalysis (described in section 1.4.3) [75].

It is generally accepted that the rhomboid catalytic cycle (**Fig. 7**) begins after the substrate is docked within the active site pocket. In GlpG, the catalytic S201 is deprotonated and initiates nucleophilic attack of the carbonyl carbon. This leads to the formation of the first oxyanion intermediate, the most characteristic feature of serine protease-driven proteolysis, followed by leaving of the C-terminal (prime-side, P') cleavage product from the acyl-enzyme covalent complex. The intermediate is subsequently hydrolyzed to generate the non-prime side, N-terminal cleavage product. Finally, Ser201 is re-protonated in the presence of His254, which restores the initial apoenzyme state.

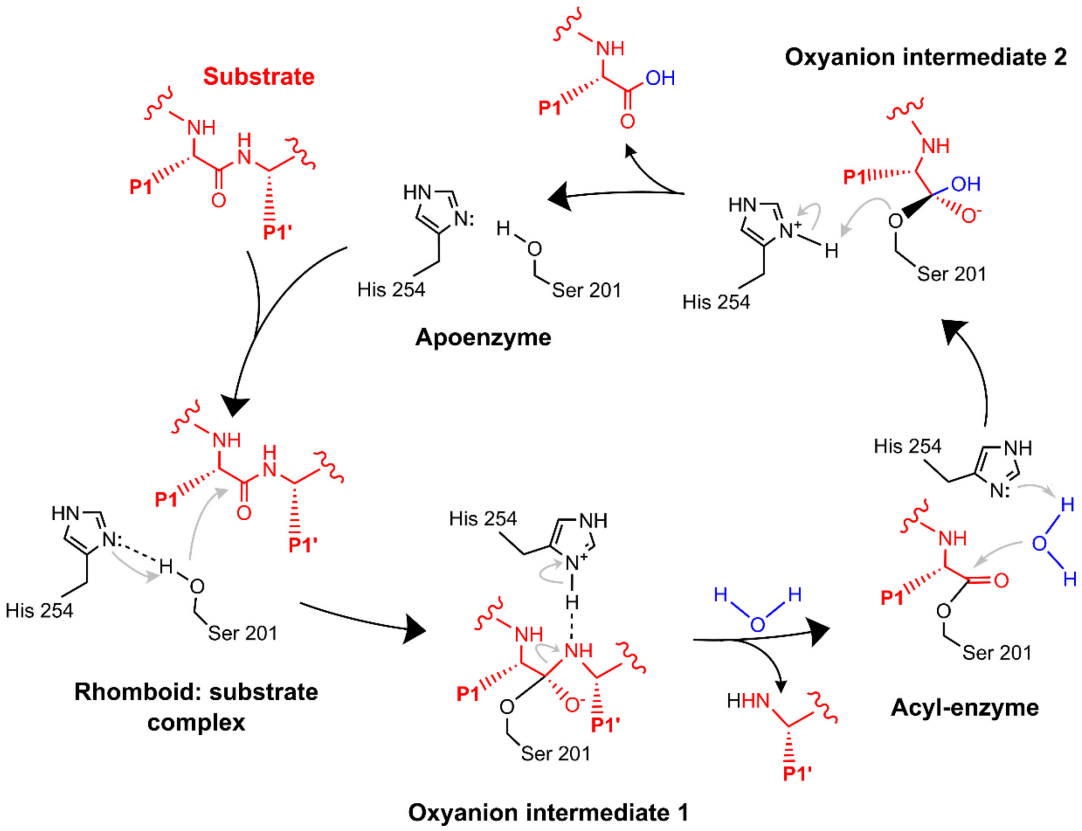


Fig. 7: Reaction cycle of rhomboid-mediated proteolysis. P1-P1' substrate region is highlighted in red, catalytic residues in black, water molecule in blue. Mechanistic details are described in the main text. Adapted from [80] and modified.

To understand the mechanism of rhomboid-mediated catalysis, characterization of all transition states is necessary. Following the reaction coordinate, isocoumarin inhibitor bound to GlpG is reminiscent of the acyl enzyme, while bound phosphonofluiridates resemble the first tetrahedral (oxyanion) intermediate. The binding mode of β-lactam derivatives suggests a putative model for the de-acylation step. Finally, the formation of a true oxyanion with a negatively charged oxygen atom was achieved by inhibiting GlpG with peptidyl aldehydes [59] and peptidyl chlorometylketones (CMKs) (our study [81]) that bind Ser201 reversibly (**Fig. 8**). Furthermore, the GlpG-CMK inhibitor complex also indicates the binding region for the non-prime side part of the substrate within the active site pocket (**Fig. 8**). Surprisingly, peptidyl aldehydes, also bound in non-competitive mode, suggesting that substrate binds to the exosite first [59], and only then the aldehydic moiety modifies the catalytic Serine.

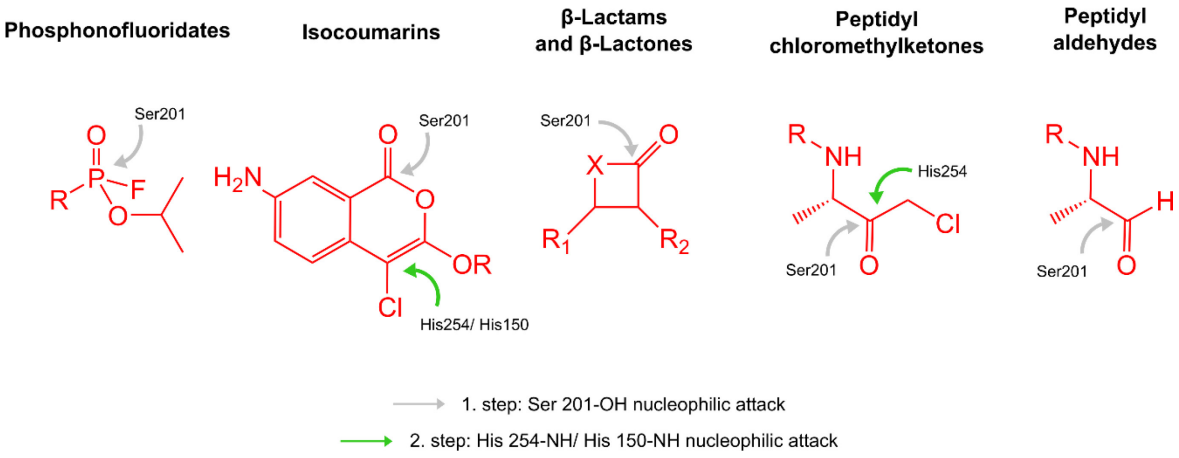


Fig. 8: Mechanism-based rhomboid inhibitors.

The active site pocket of rhomboid protease, which is buried within its transmembrane core domain, is formed by non-canonical Ser-His catalytic dyad, different from the Asp-Ser-His catalytic triad of any soluble serine protease [29, 49, 82]. Presence of the catalytic dyad, together with the substrate sequence determinants, the substrate-to-exosite-binding mechanism, as well as the nature of the environment, where catalysis takes place in, are typical features of rhomboid intramembrane proteolysis. They all contribute to the differences in mechanism and enzyme kinetics between rhomboids and their soluble counterparts.

1.4.5 Rhomboid regulation

Not all proteins that are recognized by rhomboid proteases possess the recognition motif [71], similar to that of AarA substrate TatA. Moreover, model rhomboid substrates are recognized by multiple rhomboids, some with widely different effectivities, documenting that the presence of the recognition motif may not be the only feature defining the substrate [46, 49, 75], as already mentioned above.

1.4.5.1 Interaction of rhomboids with lipids

Interestingly, *in vitro* rhomboid activity can be markedly modulated by the type of detergent and lipid used during reconstitution. In fact, in the case of GlpG of *Escherichia coli* and YqgP of *Bacillus subtilis*, response to the lipid composition is surprisingly completely the opposite [46, 60]. Rhomboid activity can be also affected by membrane-altering agents, such as lyso-PC, or mβCD [83]. Together, this implies a possible rhomboid dependence on the physicochemical properties of the lipid membrane environment, an effect which is common among membrane proteins [84, 85]. To date, several crystal structures clearly show that GlpG interacts with lipids [52, 55, 59, 60]. A biophysical study also shows that the local conformational dynamics of the loop L5 region in GlpG and the transmembrane region encoding the helix-destabilizing motif clearly depend on the microenvironmental composition and differ between detergent micelles and proteoliposomes [58, 71]. Molecular dynamics simulations of GlpG in the lipid environment also hint at local thinning of the bilayer surrounding the rhomboid [58, 86], which is caused by the hydrophobic mismatch, i.e. the difference in the bilayer and rhomboid thickness. This effect may promote local substrate unwinding and initiate binding to the rhomboid core, but may also directly alter rhomboid activity [87]. Very recent membrane biophysics experiments with a diverse set of membrane proteins, including rhomboids, imply that the rate of membrane diffusion is considerably higher for certain rhomboid proteins in thicker lipid membranes [88]. This may be due to the nature of rhomboid structural features such as the position and orientation of the extramembrane domains and the L1 loop, and also the overall irregular shape of the molecule, which together may cause local membrane thinning. Therefore, diffusion rate is an intrinsic feature of any rhomboid protein and greatly affects proteolysis. The effects of membrane composition on rhomboid mechanism are reviewed in [89].

1.4.5.2 Role of rhomboid extramembrane domains

Genomic analyses suggest that many rhomboid proteins include mostly uncharacterised extramembrane N- or C-terminal soluble domains, which are diverse in sequence and possess putative functional motifs [34, 35]. Yeast rhomboid homolog Rbd2 mediates actin assembly during clathrin-mediated endocytosis. Its C-terminal cytosolic domain binds lipid molecules and is sufficient to regulate the process even without rhomboid proteolytic activity [90]. Similarly, mammalian rhomboid RHBDL4, localized to the ER, interacts via its cytosolic C-terminal domain with ubiquitylated substrates having unstable transmembrane helices, and activates their degradation via ERAD (section 1.1.1) [91]. On the other hand, the cytosolic N-terminal domains of GlpG homologs from *Escherichia coli* and *Pseudomonas aeruginosa* appear to have no effect on proteolytic activity, though these measurements were carried out using only a model substrate [64, 92]. Another study indicates a role for the N-terminal soluble domain of mitochondrial rhomboid protease PARL during mitophagy in damaged mitochondria. Here, under mitochondrial stress, PARL undergoes an autocatalytic phosphorylation-dependent cleavage at its N-terminus. When phosphorylation inhibits auto-processing (cleavage), this negatively regulates mitophagy [93, 94]. The cleavage of thrombomodulin by the mammalian rhomboid protease RHBDL2 depends on the interaction between the N-terminal and C-terminal cytosolic domains of rhomboid and substrate, respectively. Interestingly, the thrombomodulin soluble region alone is sufficient to mediate rhomboid processing when part of a model protein containing a typical non-substrate transmembrane sequence. This may imply the existence of diverse and additional mechanisms of substrate recognition [95]. The cytosolic loops of *Drosophila* Rhomboid-4 represent another example of extramembrane regions that directly regulate the proteolytic activity. They contain metal binding residues that can bind divalent calcium, which allosterically potentiates the substrate cleavage [96].

1.4.5.3 Spatial and temporal rhomboid regulation

Rhomboid-mediated shedding, as with any other proteolytic event, is irreversible and thus it is tightly regulated and relies also on the spatial and temporal co-localization of both enzyme and substrate within the membranes along the signalling pathway. Thus, condition for rhomboid and its substrate to be met is the presence in the same membrane compartment.

This level of regulation was first observed in *Drosophila.* Rhomboids-1, 2 and 3 (here, R1-3) all contribute to the activation of the EGFR secretory pathway by cleavage of EGFR ligands. However, their expression is tissue-specific, and R1 and R3 are expressed in the developing eye, while R2 is expressed in the germ line. R3 is also expressed during the embryonal development. Their expression is also

compartment-specific: R1 resides in the plasma membrane only, but both R2 and R3 are localized to the ER membrane, where they fine-regulate the levels of the rhomboid substrates avoiding their further trafficking downstream through the secretory pathway. Moreover, their activities are required at different stages of fly development [97]. In addition, R1 levels are tightly regulated and correspond to the EGFR ligand concentration gradients in intercellular space [98]. Another level of spatial and temporal regulation of rhomboid proteolytic activity is mediated by iRhom pseudoprotease activity and is discussed in section 1.5.2.

1.4.6 Development of biologically relevant rhomboid inhibitors

The second generation of rhomboid-specific inhibitors, represented by monocyclic β- lactams [99] and β-lactones [100, 101], have surpassed the mechanism-based highly reactive isocoumarins, fluorophosphonates and phosphonofluoridates (section 1.4.4) in their selectivity, though they suffer from a rather weak potency. As mammalian and parasite rhomboids are tempting drug targets, there is continued effort to identify new inhibitor scaffolds to overcome both selectivity and potency problems. These have yielded some promising chemical warheads and scaffolds, such as benzoxazinone derivatives [102, 103] or N-methylene-substituted saccharines [104], all synthesized based on computer-based rational drug design.

1.5 Biology of rhomboid proteins

1.5.1 Rhomboid proteases are involved in signalling.

Despite the broad distribution of rhomboids, their physiological substrates are mostly unknown and as a consequence, our knowledge of their biological roles is rather scarce [105-107]. Besides the pioneering work on the biological role of Rhomboid-1 in *Drosophila*, only few other rhomboid proteases have been characterised to date, at least to some extent. The aim here is not to list all of the rhomboid functions described so far. Instead, I will focus on the key active rhomboids and outline their physiological context.

Mammals contain five active rhomboid proteases, RHBDL1-4 residing in the secretory pathway, and the rhomboid PARL in the mitochondrial inner membrane. Strictly considered, PARL is the only member of the family whose physiological substrates have been identified.

RHBDL2 is the best studied secretory protease. It is localized to the plasma membrane and is highly expressed in epithelial cells [108]. RHBDL2 may be involved in wound healing; its absence was

shown to prevent thrombomodulin's physiological processing and decrease cell migration [95]. RHBDL2 may also represent a redundant pathway to shedding of EGF ligands by ADAM10 and ADAM17 proteases in mammals, especially in tissues or physiological conditions with low ADAM levels [109]. A proteomics screen also identifies other membrane proteins as RHBDL2 substrates. These are involved in cell adhesion and migration, and the authors suggest the role for RHBDL2 in epithelial homeostasis [108].

RHBDL4 (or RHBDD1) is the only mammalian rhomboid ortholog residing in the ER membrane [91]. It is topologically distinct from other mammalian rhomboids as its TM core only consists of six helices, it contains the ubiquitin-binding C-terminal domain and its active site faces the ER lumen. Proteolytically active RHBDL4 has been proposed to downregulate exosomal secretion [110, 111], to affect tumor growth in several types of cancer by regulating the levels of TGFα within the secretory pathway [111-113] or to be involved in anti-apoptotic processes [114, 115]. Fleig *et al.* also show that RHBDL4 levels are elevated as a consequence of the unfolded protein response during ER stress. Through its cytosolic C-terminal domain, RHBDL4 can recognize ubiquitinated single-spanning and polytopic membrane proteins with destabilized transmembrane domains. RHBDL4 subsequently cleaves these substrate proteins at multiple sites and triggers their membrane dislocation and subsequent elimination by the ERAD machinery [91]. The C-terminal domain of RHBDL4 interacts with AAA+ ATPase p97 and promotes the trafficking of clients into the proteasome [91, 116]. RHBDL4 thus controls apparently distinct processes in signaling and membrane protein quality control by an as yet to be elucidated mechanism.

Rhomboid proteases of protozoan apicomplexan parasites *Toxoplasma gondii* and *Plasmodium falciparum*, TgROM1-6 and PfROM1, 3, 4, 6-10, respectively, are expressed during different stages of the parasitic life cycle and regulate the host cell adhesion and invasion. The roles of rhomboids TgROM4 and PfROM4 are particularly indispensable as they specifically cleave surface adhesion proteins, such as AMA1, during invasion and allow parasite entry into the host cell [117-119].

The roles of PARL-type rhomboid proteases in signalling are distinct from those of Rhomboid-1 or RHBDL2, since the signal they release initiates mitochondria-related events [120]. PARLs were initially studied in yeasts where Pcp1 and Rbd1 homologues cleave Mgm1, a mitochondrial GTPase, and through this regulate the organelle's morphogenesis [93]. Human PARL rhomboid is autocatalytically processed, with the pattern of processing based on the phosphorylation status of its N-terminal part, which determines its physiological role [94]. PARL is involved in the activation of mitophagy under certain stress conditions, when it cleaves PINK1 kinase in healthy or PGAM5 phosphatase in functionally-impaired mitochondria [121-123]. Its role in apoptosis and necrosis has been also discussed [124].

1.5.2 Rhomboid pseudoproteases mediate membrane protein quality control.

Emerging research implies biologically significant functions for inactive homologs of a variety of enzyme classes such as kinases, phosphatases, nucleases or proteases. It is estimated that these pseudoenzyme genes cover nearly 10% of any metazoan genome [125]. The common feature of any inactive enzyme homologs is the fact that, while remaining highly conserved, they have lost their catalytic activity during evolution and/or employed novel pseudoenzyme specific structural features and physiological roles. Pseudoenzymes mostly act as allosteric regulators of their active counterparts, either by direct binding of and competition for the substrate, thus controlling substrate trafficking and localization, or as scaffold proteins mediating protein-complex assembly [126, 127]. They control a diverse set of processes, ranging from development, immune response, inflammation, protein quality control through to DNA repair [40, 128].

Rhomboid pseudoproteases are represented by iRhoms, which are the most populated within the whole rhomboid superfamily, and by the evolutionarily more distant but homologous subgroups of Derlins, TMEM115, RHBDD2, RHBDD3 and UBAC2 proteins. I will illustrate the molecular aspects of rhomboid-like proteins using the well characterised iRhoms from *Drosophila* and mammals, and Derlins from yeasts and mammals [79, 129]. The (patho) physiological roles of non-catalytic rhomboids are reviewed elsewhere [130] (**Fig. 9**).

In *Drosophila,* iRhom is localized to the ER membrane of neuronal cells of the central nervous system and its expression increases during eye and wing development. At this stage, iRhom restricts the trafficking of Rhomboid-1 protease substrates Spitz and Gurken, from the ER to the Golgi and triggers their degradation in ER-associated degradation (ERAD, for details see Section 1.1.1). Thus, *Drosophila* iRhom down-regulates EGFR signaling by counteracting Rhomboid-1 proteolytic activity in a physiologically relevant manner and specifically prevents Rhomboid-1 interaction with the substrate (**Fig. 9a**). *Drosophila* mutated in iRhom exhibits a sleepy phenotype because of the impaired EGFR signaling in the central nervous system [131].

Mammals contain two rhomboid iRhom pseudoproteases, iRhom1 and 2, both localized to the ER and plasma membrane. While expression of iRhom1 is broad, iRhom2 is only expressed in macrophages. The ortholog iRhom2 is required for the trafficking of ADAM17 metalloprotease (also known as TACE, for tumor necrosis factor α-converting enzyme) from the ER to the Golgi, where ADAM17 is in turn activated by the protease furin, and mature traffics to the plasma membrane (**Fig. 9b**) [132, 133]. Beyond its role in trafficking ADAM17, iRhom2 forms a stable complex with it at the cell surface where it enables rapid activation of ADAM17 by extracellular signals, and determines ADAM17 substrate specificity. ADAM17 is the key activator of transmembrane precursors of cytokine and growth

factors such as TNF, TGFα and other EGF-like ligands. Furthermore, a protein called FRMD8 (or iTAP) stabilizes the iRhom2/ADAM17 complex at the cell surface and prevent its endocytosis and lysosomal degradation [134, 135].

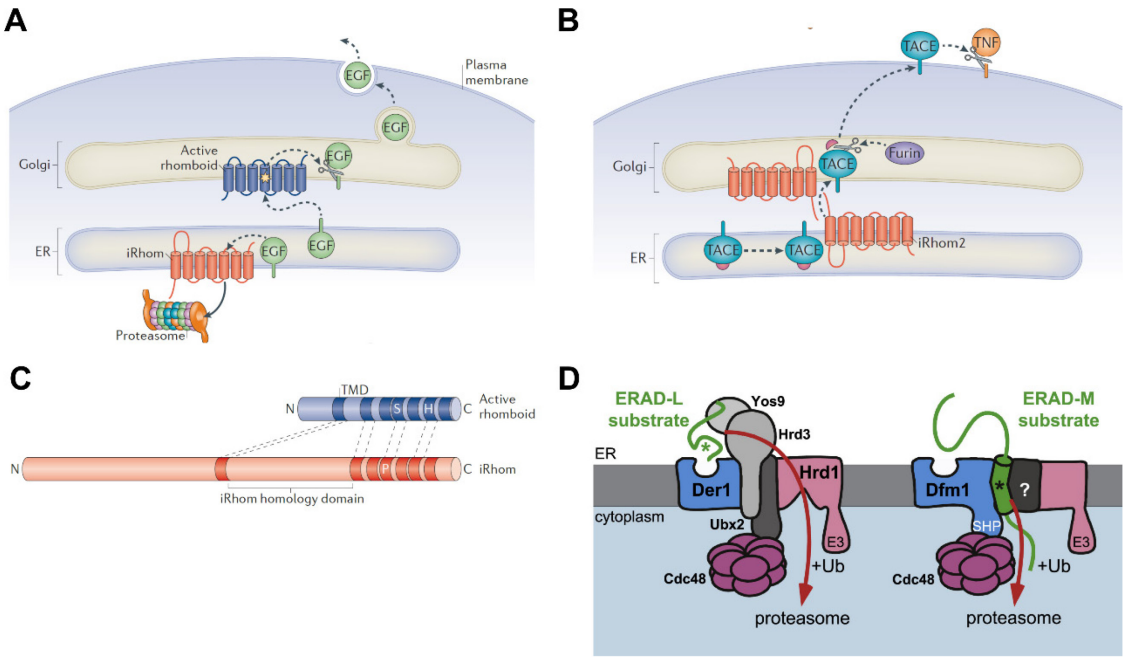


Fig. 9: Physiological mechanisms of rhomboid pseudoproteases: ***(A)*** In *Drosophila*, iRhom resides in the ER membrane, binds full-length EGFR ligands Spitz and Gurken and initiates their protasomal degradation. Thus, it prevents their trafficking into the Golgi and their processing and activation by active Rhomboid-1. **(B)** In mammals, iRhom2 is responsible for the activation of metalloprotease TACE (ADAM17). It acts as a scaffold to assist the maturation and translocation of premature TACE from ER to Golgi. It also remains complexed to the ADAM17, together with FRMD8, on the plasma membrane and regulates ADAM17 activity (not shown). **(C)** Compared to active rhomboid (in blue), iRhom contains distinct structural features such as an N-terminal cytosolic domain and a large IRHD domain replacing the loop L1 region of active rhomboids. **(D)** Derlins Der1 and Dfm1 are rhomboid-like proteins, whose roles are required for the proper function of ERAD of ER-luminal and ER-membrane proteins. Derlins specifically recognize ERAD substrates and launch their unwinding and degradation in the proteasome. Figures A-C are adopted from [*40*] and D from [*136*], respectively.

The N-terminal domain and the IRHD domain of iRhom2 are key regulatory elements that are crucial for iRhom2 stability and ADAM17 binding [134], substrate specificity [137] and stimulation of ADAM17 proteolytic activity [137-139] (**Fig. 9c**). Besides its role within the secretory pathway, iRhom2 can also specifically stabilize protein complexes, or conversely, trigger client protein degradation acting as a molecular switch in certain stages of the cellular innate immune response to viral infection [140, 141].

Derlin-like proteins were first studied in yeast, where both paralogs Der1 and Dfm1 are involved in protein dislocation from the ER (retrotranslocation), or ERAD of misfolded soluble luminal (Der1; ERAD-L) and transmembrane proteins (Der1, Dfm1; ERAD-M) [142-144] (**Fig. 9d,** for details see section

1.1.1). Dfm1 recognizes the client transmembrane protein and mediates its ubiquitination (Ub) by the E3 Ub-ligase Hrd1 or Doa10 pathway [145]. The Ub-tagged protein is extracted from the ER membrane into the cytosol, via the activity of the AAA+ ATPase Cdc48 (known as p97/VCP in mammals), and directed to the proteasome for degradation. Dfm1 is crucial for proper function of the retrotranslocation machinery by scaffolding the ERAD components. Importantly, its C-terminal cytosolic SHP domain was shown to interact with Cdc48 ATPase, which provides the energy for the entire process [143, 146] (**Fig. 9d**). In the other case, the Dfm1/Doa10 complex cooperates with Ypf1, a signal peptide peptidase SPP ortholog from yeast, to dislocate and degrade the zinc transporter Zrt1, thus controlling its levels in a zinc-dependent manner [147].

Der1 homologs are conserved in mammals where they also perform significant roles in ERAD [148, 149]. However, the precise mechanism of derlin action is still unclear. Derlins share a common six TM core with the rhomboid proteases of the secretory pathway such as RHBDL4 and bacterial rhomboid GlpG. This may suggest a similar binding mode shared by the rhomboid protease-substrate interaction and the derlin rhomboid pseudoprotease-client protein interaction. Given the common fold and other structural features, it has been speculated that derlins may bind their client proteins through a mechanism resembling the rhomboid recognizing the unwound transmembrane region of its substrate followed by subsequent recognition by the active site [136].

1.5.3 Bacterial rhomboid proteases

1.5.3.1 AarA in Providencia stuartii

The only substrate-related function of a bacterial rhomboid protease has been described in the Gram-negative bacterium *Providencia stuartii* (*Ps*). The rhomboid AarA was identified in screens for genes controlling peptidoglycan and aminoglycoside modifications [150, 151]. The monocistronic *aarA* null mutant is unable to promote the secretion of an extracellular signal for quorum sensing and is defective in cell division [152]. This inability can be rescued by overexpression of *Proteus mirabilis (Pm)* or *Escherichia coli (Ec)* homologs of the TatA single-spanning membrane protein, a component of twin-arginine translocase (Tat) system [78, 153]. In contrast to *Pm* and *Ec* TatA, the *Ps* TatA contains an additional N-terminal sequence and its specific shedding by AarA is an essential step required for the assembly of the Tat translocon, for the activation of Tat-mediated secretion of folded proteins and for the propagation of diverse Tat-dependent signals in intercellular communication [78, 154].

1.5.3.2 GlpG in *Escherichia coli*

The rhomboid protease GlpG from *Escherichia coli*, and its homolog from *Haemophilus influenzae,* are currently the best characterised intramembrane proteases. GlpG has become the main mechanistic and structural model for intramembrane proteolysis as discussed in more detail in the previous sections. Paradoxically, little is known about its physiological role. Genetic ablation of *glpG* does not lead to any remarkable phenotype [155] and no natural substrates have been identified to date. GlpG cleaves model substrates containing a type I-orientated TM helix from the multi-spanning proteins LacY, a lactose permease [74], TatA of *Providencia* AarA [48], the artificially truncated multi-spanning multidrug transporter MdfA [156], and Spitz, Gurken and Keren, natural substrates of *Drosophila* Rhomboid-1. Transposon-sequencing screens of extra-intestinal pathogenic *Escherichia coli* [157] identified *glpG* as one of the genes responsible for bacterium propagation in the intestinal mucus layer, a natural niche of intestinal *E. coli* strains [158]. In *in vitro* experiments, a *glpG* loss-of-function mutant exhibits an impaired growth phenotype, which was actually related to the polar effect of the mutation disrupting also the *glpR* gene, located downstream in the *glpEGR* operon. In contrast, *in vivo* assays performed in mouse gut revealed that GlpG itself is responsible for effective colonization of the gut by *E. coli*. However, the mechanism of GlpG activity in this context is elusive, since no physiologically relevant substrate has been identified [158].

1.5.3.3 YqgP in Bacillus subtilis

YqgP (also known as GluP) and YdcA are two rhomboid-like genes present in the model Gram-positive eubacterium *Bacillus subtilis*. While YqgP was characterised as an active rhomboid protease both *in vivo* and *in vitro*, no proteolytic activity of YdcA has been ever detected [48, 49]. YqgP is part of a predicted glucokinase operon [159] and its molecule consists of a seven TM core and additional N-terminal cytosolic and C-terminal periplasmic globular extramembrane domains (UniProtKB-P54493). To date, only one study was dedicated to the biological role of YqgP. The authors suggested that YqgP might be involved in glucose transport and also in cell division, since rhomboid deletion results in decreased glucose uptake and filamentous growth, respectively. It should be noted that the impact of possible polar effects of the *yqgP* deletion was not ruled out by the study [160], rescue experiments have not been performed, and the specificity of the reported phenoptypes is thus unclear. YqgP-like rhomboid proteins are highly populated in bacteria and their homologs are found in several pathogenic strains. The main aim of this thesis is to understand the function of YqgP in *B. subtilis* via identification of its physiological substrates and possibly regulatory proteins.

1.6 Metal ion homeostasis and toxicity

Metal ions act as essential structural or catalytic cofactors of functionally-diverse proteins, which play key roles in, amongst others, replication and transcription, protein biosynthesis, respiration, stress response, or in the case of pathogens, also in virulence and other (patho) physiological processes. Metal ion homeostasis is tightly regulated. Cells use active uptake and efflux mechanisms, as well as sensitive metal-sensing systems that buffer the metal pools within physiologically relevant concentrations or rapidly adapt to the fluctuating metal availability in the environment. The metal dysbalance, caused either by its limitation or surplus, leads to adverse non-physiological effects. This situation may be particularly relevant for transition metals, such as manganese, iron, zinc, cobalt, nickel or copper, normally occurring in trace but essential amounts. Transition metals are usually more reactive than other elements with a similar coordination chemistry, such as magnesium. Increased levels of transition metals may thus outcompete (mismetallate) magnesium at in its natural binding sites, causing toxicity by blocking or inhibiting magnesium-dependent processes [161-164].

1.6.1 Magnesium homeostasis

Magnesium is the most abundant and important divalent cation in any living cell. It is present at milimolar concentrations and its role as a cofactor is essential during DNA replication, protein biosynthesis or energy metabolism. Magnesium mostly exists in complex with proteins or phosphonucleotides (Mg^{2+}-ATP complex, mostly) and the intracellular availability of free Mg^{2+} depends on the physiological status of the cell. Due to its significance, magnesium homeostasis is controlled on multiple levels, both in the cytosol and extracellular space, and is executed by the activity of transporters, channels, carriers or sensors (extensively reviewed elsewhere [165]).

In bacteria, three structurally and mechanistically distinct magnesium transporter classes are present - MgtE, CorA and MgtA. They are usually present in multiple copies and each class is active in certain physiological conditions. Magnesium regulatory systems are best understood in *Salmonella enterica* serovar Typhimurium [165]. Notably, MgtE and CorA homologs are widely distributed in organisms ranging from bacteria to humans [166-168].

1.6.2 Magnesium transporter MgtE

MgtE is highly conserved and abundant magnesium transporter present in both Gram-negative and Gram-positive bacteria. Based on phylogenetic analyses, it is encoded by up to 50% of all sequenced prokaryotes and is assumed to be the primary magnesium import mechanism [169, 170]. In *Borrelia burgdorferi,* the *mgtE* gene product was identified as a putative virulence factor [171] and in the opportunistic pathogen *Pseudomonas aeruginosa*, its function appears to be essential for inhibition of the type III secretion system (T3SS), which is involved in virulence [172]. In addition, *Bacillus subtilis* MgtE can functionally substitute for its homolog TRPM7 of the SLC41 family in mammals [173, 174]. Magnesium transport in an *E. coli* mutant lacking essential Mg^{2+} transporters can be restored by heterologous expression of *Thermus thermophilus* MgtE [175].

MgtE is one of four magnesium transporters encoded in *Bacillus subtilis* and it is the main system for Mg^{2+} import that this bacterium naturally utilizes. It shares 34% identity and 54% similarity to the structural and physiological model transporter MgtE from *Thermus thermophilus*. Chromosomal deletion of *mgtE* results in a strain that is unable to grow without a large excess of extracellular Mg^{2+}, at which point other transporters are probably activated [176].

1.6.2.1 MgtE structure and regulation

Crystal structures of MgtE from *Thermus thermophilus (Tt* MgtE*)* reveal its homodimeric nature, with each monomer having a 5-helical TM domain linked by a connecting helix (also termed the plug helix) to a soluble cytosolic domain, which consists of an N-domain and two repeats of a cystathione-β-synthase (CBS) subunit (**Fig. 10**). The Mg^{2+} conducting pore is formed at the dimer interface, by the TM2 and TM5 helices of both monomers. Inside the pore, there is a broadly conserved pair of negatively charged aspartates, D432, one from each monomer, which coordinate Mg^{2+} cation. Full-length MgtE contains seven Mg^{2+} binding sites (Mg1-Mg7) with varying metal binding affinities (**Fig. 10**). The Mg1 site is made up of two conserved aspartates within the pore, while the Mg2-7 sites are located either within the plug helix (Mg2+3), or within the cytosolic domain (Mg4-7). The Mg-binding sites are occupied sequentially, in a conformation-dependent manner and cooperatively stabilize the dimeric complex [175, 177, 178]. Depending on its magnesium occupancy, the cytosolic domain serves as an intracellular Mg^{2+} sensing element [179-182].

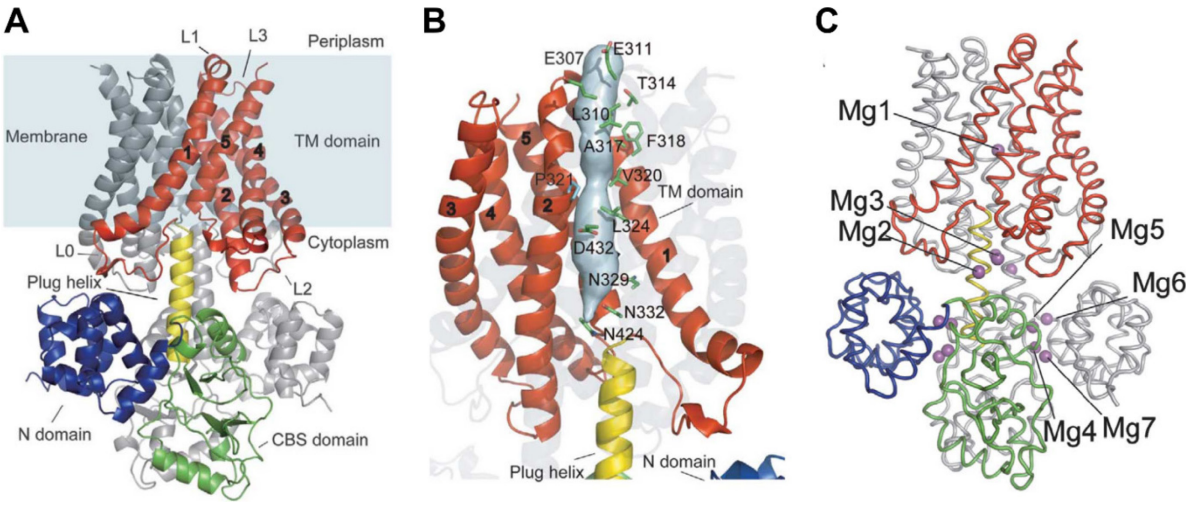
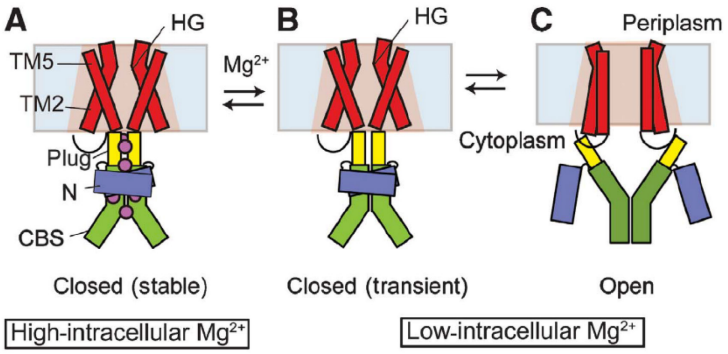


Fig. 10: Structural features of magnesium transporter MgtE from *Thermus thermophilus.* **(A)** Crystal structure of full length MgtE homodimer. For illustration, subunits of one monomer are highlighted: N domain in blue, two repeats of CBS domain in green, plug helix in yellow and TM domain in red. TM helices 1-5 and loops L0-L3 are indicated. **(B)** Surface as well as the residues forming of the ion-conducting channel are shown. **(C)** MgtE contains seven magnesium-binding sites, Mg1 at the dimer interface, Mg2 and Mg3 in the plug helix and Mg5-7 within the cytosolic part. Figure adapted from [175].

Detailed structural, biophysical and molecular dynamics analyses revealed how MgtE activity is regulated (reviewed in [182]). Structures of either full-length MgtE, or transmembrane or cytosolic MgtE domains alone in the presence or absence of Mg^{2+} suggest the following transport mechanism [175, 178, 179, 183]. When intracellular Mg^{2+} levels are low, the Mg-binding sites that are directly involved in dimerization are unoccupied. This leads to the repulsion of the negatively charged residues in the cytosolic regions of both monomers and to a specific rotation of the plug helices (**Fig. 11**). This in turn initiates a movement of the cytosolic domain monomers apart from each other, leading to the unplugging of the pore, which is now ready for Mg^{2+} transport [175, 177, 178, 181, 184, 185]. Moreover, structural and calorimetric analysis indicates that binding of ATP within the CBS domain stabilizes MgtE in a compact closed dimer conformation and increases its affinity for magnesium [185].



(Figure legend text continues on the next page)

Fig. 11: Gating mechanism of MgtE. (A-C) MgtE is proposed to open the ion-conducting pore when intracellular concentration of $Mg^{2+}$is low and transiently closes in order to prevent the cation influx, when necessary. Figure adapted from [*175*] and modified.

MgtE activity is also regulated transcriptionally, at RNA level, by an Mg^{2+} sensing RNA riboswitch. The *mgtE* mRNA contains an Mg^{2+}-sensing element the 5' untranslated region (termed the M-box). When cytosolic Mg^{2+} levels increase, this leader sequence forms a compact Mg^{2+}-bound tertiary structure referred to as a transcription terminator. Its presence results in lower RNA abundance and, in turn, in lower MgtE translation [186].

1.6.2.2 MgtE mismetallation by transition metals

If a bacterium encounters conditions with elevated levels of transition metal ions, such as Co^{2+}, Zn^{2+}, Ni^{2+} or Mn^{2+}, the magnesium transporting activity may be blocked or corrupted. Cation selectivity and mismetallation propensity differ between magnesium transporter classes. A possible explanation rationalizing why organisms employ multiple mechanistically diverse transporters may be that they can thus better avoid the diverse effects caused by changes in their environment.

MgtE can selectively transport Co^{2+} in *Bacillus firmus* OF4 *in vivo* [170], or *in vitro* [175], whereas Zn^{2+} and Ni^{2+} block MgtE magnesium transporting activity, but are not themselves transported [170]. In an *E. coli* strain lacking magnesium transporting activity, overexpression of MgtE leads to hypersensitivity to Co^{2+} and Ni^{2+} [175]. In a *Bradyrhizobium japonicum* mutant lacking its high affinity manganese import system, a suppressory gain-of-function mutation of MgtE, which leads to an impaired gating mechanism, compensates for the Mn^{2+} uptake system loss by importing Mn^{2+} cations. Due to this, however, in magnesium limiting conditions, when the MgtE channel is open, Mn^{2+} can enter the cell, which may ultimately result in manganese accumulation and toxicity [187, 188].

Structural and electrophysiological study of *T. thermophilus* MgtE, reconstituted into proteoliposomes, shows that Mg^{2+} is transported with high selectivity, which is governed by the size of the ion conducting pore as well as by the unique Mg^{2+} geometry and hydration chemistry [175, 183]. Magnesium transporting activity *in vitro* is, however, strongly inhibited by Mn^{2+}, which binds to the additional cation binding sites at the periplasmic side and stabilizes the closed state conformation, thus preventing the formation of the conducting pore [183].

2 AIMS

- To analyse the substrate preferences of GlpG rhomboid protease from *Escherichia coli*.
- To analyse the contributions of S4 subsite residues to the substrate specificity of GlpG.
- To design optimal peptide sequence for development of potent rhomboid inhibitors.
- To test the designed peptidyl ketoamide compounds as inhibitors of endogenous YqgP rhomboid protease in *Bacillus subtilis.*
- To analyse the biological role of YqgP in *Bacillus subtilis*.

3 MATERIALS AND METHODS

3.1 Consumables

3.1.1 Chemicals

$^{13}C_6$ $^{15}N_2$ L-Lysine HCl	(Silantes #211604102)
2 % bis solution	(Bio-Rad, #1610142)
21 L-Amino acids + Glycine	(Merck, #09416-1EA)
2-mercaptoethanol	(Merck, #M6250)
^{35}S- Methionine	(MGP, #KSB-01)
40 % Acrylamide solution	(Bio-Rad, #1610140)
99 % Acetic Acid	(Penta, #19970)
Acetone	(Penta, #10050)
Agarose for DNA electrophoresis	(Serva, #11404)
Ammonium chloride (^{15}N, 99%)	(Cambridge Isotope Laboratories, #NLM-467-10)
Ammonium chloride	(Merck, #A9434)
Ammonium persulfate	(Merck, #248614)
Ammonium sulfate	(Merck, #A4418)
Ampicillin	(Biotika, #1808005)
BIS-TRIS	(Merck, #14879)
Blocker™ Casein in TBS	(ThermoFisher Scientific, #37532)
Boric Acid	(Penta, #18710)
Brilliant Blue G	(Merck, #B0770)
Bromophenol Blue sodium salt	(Merck, #B5525)
Calcium chloride, dihydrate	(Merck, #7902)
Chloramphenicol	(Merck, #C0378)
Cobalt (II) chloride	(Merck, #232696)
cOmplete™ EDTA-free Protease Inhibitor Cocktail	(Merck, #11836170001)
D-(+)-Glucose anhydrous	(Merck, #G7021)
D-(+)-Maltose monohydrate	(Merck, #M5885)
D-(+)-Xylose	(Merck, #X3877)
Deuterium oxide (D2O)	(Merck, #151882)
D-Glucose (^{13}C, 99%)	(Cambridge Isotope Laboratories, #CLM-1396-1)
Dimethyl sulfoxide	(Merck, #D2650)
Disodium phosphate dihydrate	(Penta, #12330)
DL-Dithiothreitol	(Merck, #D9779)
EDTA	(Merck, #E5134)
EGTA	(Merck, #E4378)
Erythromycin	(Merck, #E5389)
Ethanol, 96%	(Penta, #70390)
G(5')ppp(5')G RNA Cap Structure analogue	(NEB, #S1407S)
GelRed stain	(Biotium, #41003)
Glutaraldehyde solution, 25%	(Merck, #G5882)
Glycerol, unhydrous	(Penta, #14550)
Glycine	(Merck, #G8898)
HEPES	(Merck, #H3375)
Hydrochloric acid 35%	(Penta, #19350)

Imidazole	(Merck, #56750)
IPTG	(BioSynth, #I8000)
Iron (II) sulfate heptahydrate	(Merck, #215422)
Isopropanol	(Penta, #17500)
Kanamycin sulfate	(Merck, #K4000)
L/Glutathione reduced	(Merck, #G4251
LB Agar	(Merck, L2897)
LB Broth with agar, powder	(Merck, #L2897)
Lincomycin hydrochloride	(Merck, #62143)
Lithium chloride, 8M solution	(Merck, #L7026)
L-Lysine hydrochloride	(Merck, #L9037)
Magnesium sulfate heptahydrate	(Merck, #63138)
Magnesium sulfate, 2.5M solution	(Merck, #83266)
Manganese (II) chloride tetrahydrate	(Merck, #221279)
MES hydrate	(Merck, #2933)
Methanol	(Penta, #21190)
MS-SAFE Protease and Phosphatase Inhibitor	(Merck, #MSSAFE)
N, N, N‘, N‘-tetramethylethylendiamine (TEMED)	(Sigma-Aldrich, #T9281)
Na_2CO_3	(Penta, #28010)
NaCl	(Penta, #16610)
n-Dodecyl-b-D-Maltopyranoside (DDM)	(Anatrace, #D310)
Neomycin	(Merck, #N6386)
Nickel (II) chloride	(Merck, #339350)
Orange G sodium salt	(Merck, #O3756)
Phenylmethylsulphonyl fluorid (PMSF)	(Merck, #78830)
Potassium phosphate dibasic (K_2HPO_4)	(Merck, #P8281)
Potassium phosphate monobasic (KH_2PO_4)	(Merck, #P5655)
Ribonucleoside Triphospoate Set	(Roche, #11277057001)
RNasin® Ribonuclease Inhibitor, Recombinant	(Promega, #N2511)
Sodium carbonate	(Penta, #28000)
Sodium Dodecyl Sulfate (SDS)	(BioRad, #1610301)
Sodium hydroxide, solid	(Penta, #15740)
Sodium phosphate monobasic	(Merck, #S8282)
Spermidine	(Merck, #85558)
Streptomycin	(Merck, #S6501)
Tetracycline	(Merck, #T7660)
Trichloroacetic acid	(Merck, #T4885)
Triton X-100	(Merck, #T8787)
Trizma base	(Merck, #T1503)
Trizma hydrochloride	(Merck, #T5941)
Tryptone	(VWR, #84610)
Tween® 20	(Merck, #P7949)
Yeast extract	(VWR, #84601)
Zinc sulfate heptahydrate	(Merck, #Z0251)

3.1.2 Enzymes

Restriction endonucleases	(all purchased from NEB)
Lysozyme from chicken egg white	(Merck, #L6876)
Pierce™ Universal Nuclease for Cell Lysis	(ThermoFisher Scientific, #88700)
Q5® High-Fidelity 2x Master Mix, containing Q5 DNA polymerase	(NEB, #M0492L)
PPP Master Mix, containing Taq DNA polymerase	(Top-Bio, #P125)
Antarctic Phosphatase	(NEB, #M0289L)
Pyrophosphatase, Inorganic	(Yeast) (NEB, #M2403S)
SP6 RNA polymerase	(NEB, #M0207S)

3.1.3 Antibodies

Anti MgtE2-275	(Agro-Bio, for our purposes)
Anti $YqgP_{CTD}$	(Agro-Bio, for our purposes)
Anti $YqgP_{NTD}$	(Agro-Bio, for our purposes)
Anti-Thioredoxin antibody produced in rabbit	(Merck, #T0803-.2ML)
Monoclonal ANTI-FLAG® M2 antibody produced in mouse	(Merck, #F1804)
Penta-His Tag Monoclonal Antibody, mouse	(ThermoFisher Scientific, #P21315)
Donkey anti-Mouse IgG (H+L) Secondary Antibody, DyLight 800	(Invitrogen, #SA5-10172)
Donkey anti-Rabbit IgG (H+L) Secondary Antibody, DyLight 800	(Invitrogen, #SA5-10044)
Goat anti-mouse IgG-HRP	(Santa Cruz, #sc-2005)
Goat anti-rabbit IgG-HRP	(Santa Cruz, #sc-2004)

3.1.4 Other consumables

- Commercial kits and solutions:

2% Bis Solution	(BIO-RAD, # 1610142);
40% Acrylamide Solution	(BIO-RAD, # 1610140);
4–20% Mini-PROTEAN® TGX™ Precast Protein Gels	(BIO-RAD, #4561096, #4561096);
Blocker™ Casein in TBS	(Thermo, #37532);
CutSmart® buffer	(NEB, #B7204S);
GenElute™ Bacterial Genomic DNA Kits	(Merck, #NA2110-1KT);
InstantBlue™ Protein Stain	(Expedeon, #ISB1L)
Luminata™ Crescendo Western HRP Substrate	(Merck, #161245);
Luminata™ Forte Western HRP substrate	(Merck, #151555);
NucleoBond® Xtra Midi	(Macherey-Nagel #740410);
Pierce™ 660nm Protein Assay Reagent	(ThermoFisher Scientific, #22660);
QIAprep Spin Miniprep Kit	(QIAGEN, #27106);
QIAquick Gel Extraction Kit	(QIAGEN, #28706);
QIAquick PCR Purification Kit	(QIAGEN, # 28106);
REVERT**™** Total Protein Stain Kit	(LI-COR, #926-11010);
RNaseZap®	(Merck, #R2020);
Wheat Germ Extract	(Promega, #L4380)
SingleQuant Assay Kit	(Serva, #39226.1)

- Standards:

Color Prestained protein Standard, Broad Range (11-245 kDa)	(NEB, #P7712);
HyperLadder™50bp	(Bioline, BIO33040);
HyperLadder™1kb	(Bioline, BIO33026)

- Blotting:

NC2 Nitrocellulose Membrane	(SERVA, #71224.01);
Immobilon-P PVDF Membrane	(Merck, #IPVH00010)

- Resins & columns:

Ni-NTA Agarose	(QIAGEN, #30210);
Amylose Resin High Flow	(NEB, #E8022S)
Glutathione Sepharose ® 4 Fast Flow	(GE Healtcare Life Sciences, #17513201)
Superdex™ 75 10/300 GL	(GE Healthcare Life Sciences, #10074694);
Superdex™ 200 10/300 GL	(GE Healthcare Life Sciences, #10079727);
HiLoad™ 16/600 Superdex™ 75 pg,	(GE Healthcare Life Sciences, #10076221);
HiLoad™ 16/60 Superdex™ 200 prep grade	(GE Healthcare Life Sciences, #10033316);
Desalting columns	(GE Healthcare Life Sciences, #28918007, #28918008, #17085101)
Pierce™ Disposable Columns	(ThermoFisher Scientific, #29920, #29922, #29924)

- Concentrating: Centrifugal concentrators Vivaspin 2 and Vivaspin 20 (Sartorius)

- Dialysis: Dialysis membranes Spectra/Por, (Specrrum Laboratories)
  Slide-A-Lyzer® Dialysis cassettes G2, (ThermoFisher Scientific)

- Filtration: Nalgene® Reusable Filter Holder with Reciever (ThermoFisher Scientific, #300-4050)
  MF-Millipore™ Membrane Filter, 0.22 µm pore size (Millipore, #GSWP01300)
  Rotilabo®-syringe filters, sterile, 0.22 µm pore size (Carl Roth, #P668.1)

3.2 Instruments

- Centrifuges: 5424; 5424R (Eppendorf)
  Megafuge 2.0R (Hereaus Instruments)
  Alegra X-15R; Avanti J-30I; Optima MAX-XP; Optima L-100 XP (Beckman Coulter)
  Evolution RC (Sorvall)

- Incubators: New Brunswick™ Innova® 44 Stackable Incubator Shaker (Eppendorf);
  Peltier-cooled incubator IPP 400 (Memmert)

- Cell disruption: EmulsiFlex®-C3 High Pressure Homogeniser (Avestin)
  Ultrasonic homogeniser Sonopuls HD2200 (Bandelin);

- Spectrometers: Eppendorf Biospectrometer® kinetic (Eppendorf)
  NanoDrop™ One Microvolume UV-Vis Spectrophotometer (ThermoFisher Scientific)
  Spectrometer Bruker Advance III HD, 850 MHz Cryo (Bruker BioSpin GmbH)

- Electrophoresis: Electrophoresis Power Supply Consort EV202, EV231, EV243 (BIO-RAD)
  Owl™ EasyCast B-Series Horizontal Gel System Apparatus (ThermoFisher Scientific)
  Mini-PROTEAN Tetra Cell module (BIO-RAD)

- Blotting: Mini Trans-blot cell module (BIO-RAD)

- Shaking: IKA® HS260 basic and IKA® KS260 basic shakers (IKA)
  DRS-12 Digital Rocking Shaker (ELMI)

WT16 Rocking Platform (Biometra)
SRT6D, SRT9D Digital tube rollers (Stuart)
Multi Bio RS-24 Programmable Rotator (Biosan)

- Mixing: Thermomixer C; Thermomixer comfort (Eppendorf)
- Chromatography: FPLC ÄKTA Explorer, (GE Healthcare/ Amersham Biosciences)
- Thermocycler: TProfessional TRIO PCR Thermocycler (Biometra)
- Visualisation: LAS 3000 CCD camera (Fujifilm Corporation)
  LI-COR® Odyssey CLx Infra-Red Fluorescence Imaging System (LI-COR Biosciences)
  UV Transluminator Quantum ST4-1100/26MX (Vilber Lourmat)
  Typhoon 9410 (GE Healthcare)
  Amersham Autoradiography Exposure Cassette (Amersham biosciences)
- Microcalorimetry: Auto-iTC200 isothermal titration calorimeter
- Other: Laboratory autoclave MLS-3020 (Sanyo);
  Magnetic stirrer Variomag Mono (Merck);
  Magnetic Hotplate Stirrer IKA® RCT basic (IKA);
  pH meter EUTECH pH50 SET (ThermoFisher Scientific)
  Water bath myBAtH™ (Benchmark Scientific)
  Vortex Genie 2 (Scientific Industries)
  Analytical Balance Adventurer™ PRO AV313C (Ohaus)
  Analytical Balance XA110/X (RADWAG)
  Balance EK-400H (AND)
  Balance PLS 4000-2 (KERN)
  Slab Gel Dryer GD2000 (Hoefer Inc.)

3.3 Software

- Signal quantification: Image Studio Lite ver 5.2 (LI-COR Biosciences)
  ImageQuant 8.0 (GE Healthcare)
- Data visualization: PyMOL™ 2.2.3 (Schrodinger)
  Inkscape 0.92.3 (general public licence GNU)
  GraphPad Prism 7.01 (GraphPad)
  Microsoft Office 2013
- NMR data acquisition: Topspin 3.2 pl 6 (Bruker)
- NMR protein structure determination: NMRFAM Sparky [189]
- NMR protein structure refinement: YASARA forcefield (www.yasara.org)
- Bibliography: EndNote X8 (Clarivative Analytics)
- Microcalorimetry data analysis: MicroCal Origin 7.0 (MicroCal, GE Healthcare)
- Quantitative protemics data analysis: MaxQuant LFQ v1.5.2.8 [190]
  QARIP [191]
- Protein topology prediction: Phobius [192]

3.4 Primers, constructs and bacterial strains

Table 1: List of primers

name	Sequence	Description
oJB230	AGATAGAGGTACCGATGGAATCATGGATTGCAAC	pJB91 cloning
oJB231	ATAGAGATCTAGAACCCTGCTCTTTATTTTTGC	reverse primer forpJB91-98,100, 109-112
oJB232	AGATAGAGGTACCGATGGAATCAACTTCCGC	pJB92 cloning
oJB233	AGATAGAGGTACCGATGGAATCAACTGGTGC	pJB93 cloning
oJB234	AGATAGA GGTACC G ATGGAATCAACTATTGATACGG	pJB94 cloning
oJB235	AGATAGA GGTACC G ATGGAATCAACTATTTGGACG	pJB95 cloning
oJB236	AGATAGA GGTACC G ATGGAATCAACTATTGCAACG	pJB96,97,100 cloning
oJB237	AGATAGA GGTACC G ATGGAATCACGAGTTCGACATGCA GCTTTTGGTAGCCCTTG	pJB98
oJB240	AGATAGA GGTACC G CGAGTTCGACATGCA GATACGGGTATTATTTTTGCC	pJB99 cloning
oJB241	ATAGAGA TCTAGA TTTGCGCAGCCC	pJB99 cloning
oJB244	AGATAGA GGTACC G ATGGAATCACGTATTGCAAC	pJB109 cloning
oJB245	AGATAGA GGTACC G ATGGAATCAACTGTTGCAAC	pJB110 cloning
oJB246	AGATAGA GGTACC G ATGGAATCAACTATTCGAACG	pJB111 cloning
oJB247	AGATAGA GGTACC G ATGGAATCAACTATTGCACATGC	pJB112 cloning
oJB371	TTAAGAAGGAGATATACAT ATGTTTTTGCTGGAGTATAC	pJB171 cloning
oJB372	AAATACAAGTTTTCTCTCTCGAG TTTTCCGTTTTGAAAAACC	pJB171 cloning
oJB391	AACCTGTATTTTCAGGGCGC CATGG GA TCGCCTTCACACCAG	pJB179 cloning
oJB392	CGAAAGCGGCCGCGACTAG T TTATTA TGATTCTTTATTGTTTTCAATTTGC	pJB179 cloning
oPR209	ACGATGACAAGGTACCGCGAATCAGCAAAAGTGATAC	RISKS LacYtm2 cloning
oPR210	ATTTTATCGCTCATTCTAG	reverse primer for LacYtm2
oPR211	ACGATGACAAGGTACCGCATGTTAGCAAAAGTGATACGGG	HVSKS LacYtm2 cloning
oPR212	ACGATGACAAGGTACCGCATATCCGAAAAAGTGATACGGGTAT	HIRKS LacYtm2 cloning
oPR213	ACGATGACAAGGTACCGCATATCAGCCATAGTGATACGGGTATTAT	HISHS LacYtm2 cloning
oPR225	ACGATGACAAGGTACCGCATATCAGCAAAGCAGATACGGGTATTATTTTT	HISKA LacYtm2 cloning
oPR293	TTAGCTTGCATGCGGCTAGCTTATGATTCTTTATTGT	reverse primer for YqgP mutants
oPR294	ACCTGCAGGCATGCAAGCTTACATAAGGAGGAACTACTATGTTTACATATTTATTTATTG	pPR155 cloning
oPR296	ACCTGCAGGCATGCAAGCTTACATAAGGAGGAACTACTATGTTTTTGCTGGAGTAT	pPR157 cloning
oPR297	TTAGCTTGCATGCGGCTAGCTTATGAAGGCGAATGCAATC	pPR157 cloning
oPR355	TCGAATTGCCCGGGGTCGACTTATCAAGCGTAATCTGGAACATCGTATGGGTAGGCCAGGTTAGCGTCGA	pPR200, pPR290 cloning
oPR359	CTAGAAATGTTTGCAAAACGATTCAAAACCTCTTTACTGCCGTTATTCGCTGGATTTTTATTGCTGTTTCATTTGGTTCTGGCAGGA	pPR200, pPR290 cloning
oPR360	ATGAAACAGCAATAAAAATCCAGCGAATAACGGCAGTAAAGAGGTTTTGAATCGTTTTGCAAACATTTCTAGAAGATCCATGATTGT	pPR200, pPR290 cloning
oPR361	TTGGTTCTGGCAGGAGCCAAAATCGAAGAAGG	pPR200, pPR290 cloning
QN213	AAGAAGCTACATCATCATACCCTGCTCTTTATTTTTGCTCTC	*in vitro* translation of: *in vitro* reverse primer
QN481	CGATTAGGTGACACTATAGAATACCATGTCAGGTTCAGGTTCAGGTTCAGGTGAATCAACTATTGCAACG	Pst TatA A8G,T,V,R,P
QN484	CGATTAGGTGACACTATAGAATACCATGTCAGGTTCAGGTTCAGGTTCAGGTGAATCATTCATTGCAACGGCC	Pst TatA T4F
QN486	CGATTAGGTGACACTATAGAATACCATGTCAGGTTCAGGTTCAGGTTCAGGTGAATCAACTATTTTCACGGCC	Pst TatA A6F
QN488	CGATTAGGTGACACTATAGAATACCATGTCAGGTTCAGGTTCAGGTTCAGGTGAATCAACTATTTTCACGGCC	Pst TatA T4G
QN520	CGATTAGGTGACACTATAGAATACCATGTCAGGTTCAGGTTCAGGTTCAGGTGAATCAGCTATTGCAACGGCC	Pst TatA T4A
QN579	CGATTAGGTGACACTATAGAATACCATGTCAGGTTCAGGTTCAGGTTCAGGTGAATCAACTATTGGCACGGCC	Pst TatA A6G
QN619	CGATTAGGTGACACTATAGAATACCATGTCAGGTTCAGGTTCAGGTTCAGGTGAATCAACTATTCCAACGGCCG	Pst TatA A6P
QN633	CGATTAGGTGACACTATAGAATACCATGTCAGGTTCAGGTTCAGGTTCAGGTGAATCAACTATTACAACGGCCG	Pst TatA A6T
QN634	CGATTAGGTGACACTATAGAATACCATGTCAGGTTCAGGTTCAGGTTCAGGTGAATCAACTATTCACACGGCCG	Pst TatA A6H
QN635	CGATTAGGTGACACTATAGAATACCATGTCAGGTTCAGGTTCAGGTTCAGGTGAATCAACTATTAAAACGGCCG	Pst TatA A6K
QN636	CGATTAGGTGACACTATAGAATACCATGTCAGGTTCAGGTTCAGGTTCAGGTGAATCAACTATTGAAACGGCCG	Pst TatA A6E
QN637	CGATTAGGTGACACTATAGAATACCATGTCAGGTTCAGGTTCAGGTTCAGGTGAATCACCTATTGCAACGGCCG	Pst TatA T4P
QN638	CGATTAGGTGACACTATAGAATACCATGTCAGGTTCAGGTTCAGGTTCAGGTGAATCACACATTGCAACGGCCG	Pst TatA T4H
QN639	CGATTAGGTGACACTATAGAATACCATGTCAGGTTCAGGTTCAGGTTCAGGTGAATCAAAGATTGCAACGGCCG	Pst TatA T4K
QN640	CGATTAGGTGACACTATAGAATACCATGTCAGGTTCAGGTTCAGGTTCAGGTGAATCAGAGATTGCAACGGCCG	Pst TatA T4E
QN709	CGATTAGGTGACACTATAGAATACCATGTCAGGTTCAGGTTCAGGTTCAGGTGAATCAGTCATTGCAACGGCCG	Pst TatA T4V
QN710	CGATTAGGTGACACTATAGAATACCATGTCAGGTTCAGGTTCAGGTTCAGGTGAATCACTCATTGCAACGGCCG	Pst TatA T4L
QN711	CGATTAGGTGACACTATAGAATACCATGTCAGGTTCAGGTTCAGGTTCAGGTGAATCAACTATTGTCACGGCCG	Pst TatA A6V
QN712	CGATTAGGTGACACTATAGAATACCATGTCAGGTTCAGGTTCAGGTTCAGGTGAATCAACTATTCTCACGGCCG	Pst TatA A6L
QN999	CGATTAGGTGACACTATAGAATACCATGTCAGGTTCAGGTTCAGGTTC	Pst TatA T4I,W,S,C,M,D,N,Q,R,Y,P; A6I,W,S,C,M,D,N,Q,R,Y,P; A8I,N,Q,Y

- "oPR" primers were prepared in our laboratory, by Mgr. Petra Rampírová
- "QN" primers were prepared by Dr. Kvido Stříšovský

Table 2: List of DNA constructs

used in	Description	Source
In vitro translation templates:		
		Dr. Stříšovský [75]
pKS150	Pst TatA T4G-His 6	Dr. Stříšovský [75]
pKS182	Pst TatA T4A	Dr. Stříšovský [75]
pKS278	Pst TatA T4V	Dr. Stříšovský [75]
pKS421	Pst TatA T4I	Dr. Stříšovský [75]
pKS141	Pst TatA T4F-His 6	Dr. Stříšovský [75]
pKS244	Pst TatA T4H	Dr. Stříšovský [75]
pKS245	Pst TatA T4K	Dr. Stříšovský [75]
pKS246	Pst TatA T4E	Dr. Stříšovský [75]
pKS386	Pst TatA T4W	Dr. Stříšovský [75]
pKS425	Pst TatA T4S	Dr. Stříšovský [75]
pKS387	Pst TatA T4C	Dr. Stříšovský [75]
pKS420	Pst TatA T4M	Dr. Stříšovský [75]
pKS70	Pst TatA T4L	Dr. Stříšovský [75]
pKS422	Pst TatA T4D	Dr. Stříšovský [75]
pKS389	Pst TatA T4N	Dr. Stříšovský [75]
pKS424	Pst TatA T4Q	Dr. Stříšovský [75]
pKS388	Pst TatA T4R	Dr. Stříšovský [75]
pKS423	Pst TatA T4Y	Dr. Stříšovský [75]
pKS243	Pst TatA T4P	Dr. Stříšovský [75]
pKS212	Pst TatA A6G	Dr. Stříšovský [75]
pKS239	Pst TatA A6T	Dr. Stříšovský [75]
pKS280	Pst TatA A6V	Dr. Stříšovský [75]
pKS427	Pst TatA A6I	Dr. Stříšovský [75]
pKS143	Pst TatA A6F-His 6	Dr. Stříšovský [75]
pKS240	Pst TatA A6H	Dr. Stříšovský [75]
pKS241	Pst TatA A6K	Dr. Stříšovský [75]
pKS242	Pst TatA A6E	Dr. Stříšovský [75]
pKS390	Pst TatA A6W	Dr. Stříšovský [75]
pKS431	Pst TatA A6S	Dr. Stříšovský [75]
pKS391	Pst TatA A6C	Dr. Stříšovský [75]
pKS426	Pst TatA A6M	Dr. Stříšovský [75]
pKS281	Pst TatA A6L	Dr. Stříšovský [75]
pKS428	Pst TatA A6D	Dr. Stříšovský [75]
pKS393	Pst TatA A6N	Dr. Stříšovský [75]
pKS430	Pst TatA A6Q	Dr. Stříšovský [75]
pKS392	Pst TatA A6R	Dr. Stříšovský [75]
pKS429	Pst TatA A6Y	Dr. Stříšovský [75]
pKS231	Pst TatA A6P	Dr. Stříšovský [75]
pKS119	Pst TatA A8G	Dr. Stříšovský [75]
pKS124	Pst TatA A8T - His6	Dr. Stříšovský [75]
pKS95	Pst TatA A8V	Dr. Stříšovský [75]
pKS377	Pst TatA A8I	Dr. Stříšovský [75]
pKS374	Pst TatA A8N	Dr. Stříšovský [75]
pKS375	Pst TatA A8Q	Dr. Stříšovský [75]
pKS125	Pst TatA A8R - His6	Dr. Stříšovský [75]
pKS373	Pst TatA A8Y	Dr. Stříšovský [75]
pKS62	Pst TatA A8P	
E. coli:		
pET+G	pET25bM + GlpG-6xHis	Dr. Lemberg [49]
pET+G-F146A	pET25bM + GlpG F146A-6xHis	This work
pET+G-F146I	pET25bM + GlpG F146I-6xHis	This work
pET+G-S201A	pET25bM + GlpG S201A-6xHis	This work
pET+Y	pET25bM + YqgP-His6	Dr. Lemberg [49]
pJB100	*pMALp2E + MBP-3xFLAG-pstatA-Trx-6xhis-S*	This work
pJB109	*pMALp2E + MBP-3xFLAG-pstatA T4R-Trx-6xhis-S*	This work
pJB110	*pMALp2E + MBP-3xFLAG-pstatA I5V-Trx-6xhis-S*	This work
pJB111	*pMALp2E + MBP-3xFLAG-pstatA A6R-Trx-6xhis-S*	This work
pJB112	*pMALp2E + MBP-3xFLAG-pstatA T7H-Trx-6xhis-S*	This work
pJB171	*yqgP 1-177-TEV-6xhis in pET25b*	This work
pJB179	*6xhis-TEV-yqgP 384-507 in pHIS2*	This work
pJB184	*GST-6xhis-TEV-mgtE 2-275 in pGEX6P1*	This work
pJB91	*pMALp2E + MBP-3xFLAG-pstatA T4W-Trx-6xhis-S*	This work
pJB92	*pMALp2E + MBP-3xFLAG-pstatA I5S-Trx-6xhis-S*	This work

pJB93	*pMALp2E + MBP-3xFLAG-pstatA I5G-Trx-6xhis-S*	This work
pJB94	*pMALp2E + MBP-3xFLAG-pstatA A6D-Trx-6xhis-S*	This work
pJB95	*pMALp2E + MBP-3xFLAG-pstatA A6W-Trx-6xhis-S*	This work
pJB96	*pMALp2E + MBP-3xFLAG-pstatA A8G-Trx-6xhis-S*	This work
pJB97	*pMALp2E + MBP-3xFLAG-pstatA A8V-Trx-6xhis-S*	This work
pJB98	*pMALp2E + MBP-3xFLAG-pstatA RVRHA-Trx-6xhis-S*	This work
pJB99	*pMALp2E + MBP-3xFLAG-lacytm2 RVRHA-Trx-6xhis-S*	This work
pKS506	*pMALp2E + MBP-3xFLAG-lacytm2-Trx-6xhis-S*	This work
pPR117	*pMALp2E + MBP-3xFLAG-lacytm2 RISKS-Trx-6xhis-S*	This work
pPR118	*pMALp2E + MBP-3xFLAG-lacytm2 HVSKS-Trx-6xhis-S*	This work
pPR119	*pMALp2E + MBP-3xFLAG-lacytm2 HIRKS-Trx-6xhis-S*	This work
pPR120	*pMALp2E + MBP-3xFLAG-lacytm2 HISHS-Trx-6xhis-S*	This work
pPR121	*pMALp2E + MBP-3xFLAG-lacytm2 HISKA-Trx-6xhis-S*	This work
B. subtilis:		
pGP886	*xkdE::Pxyl-N-YFP (Erm)*	Gift from Dr. Krásný + [193]
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
pJB220	*amyE::Pxyl-yqgP L51A (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
pPR155	*amyE::Pxyl-yqgP Δ1-178 (Spec)*	This work
pPR157	*amyE::Pxyl-yqgP Δ388-507 (Spec)*	This work
pPR274	*amyE::Pxyl-yqgP Δ1-178 S288A (Spec)*	This work
pPR275	*amyE::Pxyl-yqgP Δ388-507, S288A(Spec)*	This work
pPR290	*xkdE::Pxyl-MBP-FLAG-pstatA I5G,I10G-Trx-HA (Erm, Lin)*	This work
pTM13	*amyE::Phyperspank-yqgP (Spec)*	This work
pTM31	*amyE::Phyperspank-yqgP S288A(Spec)*	This work
pTM89	*ycgO::Phyperspank-FLAG-mgtE (Erm)*	This work
pTM93	*amyE::Pxyl-yqgP (Spec)*	This work
pTM94	*amyE::Pxyl-yqgP, S288A (Spec)*	This work

- "pTM" constructs were prepared in the laboratory of Dr. Thierry Doan, at CNRS, France

- "pPR" constructs were prepared in our laboratory by Mgr. Petra Rampírová

- "pKS" constructs were prepared by Dr. Kvido Stříšovský

Table 3: List of bacterial strains

Strain	Information	source
E. coli:		
DH5α	*commercial strain/* for DNA preparation	
C41(DE3)	*commercial strain/* for overexpression of membrane proteins	Gift: Dr. Vinothkumar
BL21Star™(DE3)	*commercial strain/* for overexpression of soluble proteins	Invitrogen
Lemo21(DE3)	*commercial strain/* for overexpression of soluble proteins	NEB
MC4100	*commercial strain/* for GlpG *in vivo* assay	Gift: Dr. Akiyama
KS52	MC4100 *ΔglpG::tet/ for* GlpG *in vivo* assay	Our laboratory
B. subtilis:		
BTM2*	*commercial strain/ Wild type 168 trpC+*	BGSC
BTM78*	*168 trpC+ ΔyqgP::cat*	This work
BTM677*	*168 trpC+ ΔyqgP*	This work
BKE23380	*commercial strain/ 168 trpC2 ΔlysA::erm*	BGSC
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
BTM501*	*168 trpC+ ΔyqgP::cat amyE::Phyp-yqgP (Spec)*	This work
BTM502*	*168 trpC+ ΔyqgP::cat amyE::Phyp-yqgP S288A (Spec)*	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
BS188	*168 trpC+ ΔyqgP amyE::Pxyl-yqgP S288A (Spec) xkdE::Pxyl-MBP-FLAG-psTatA I5G,I10G-Trx-HA (Erm, Lin)*	This work
BTM462*	*168 trpC+ ftsH::erm*	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
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
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
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
BS200	*168 trpC+ ΔyqgP amyE::Pxyl-yqgP L51A (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
BTM843*	*commercial strain/ 168 trpC+ ywlD::erm*	BSFA collection
BTM844*	*168 trpC+ ΔywlD::erm ΔyqgP::cat*	This work
BTM845*	*168 trpC+ ΔywlD::erm ΔyqgP::cat amyE::Phyperspank-yqgP (Spec)*	This work

* BTM strains were prepared in the laboratory of Dr. Thierry Doan, at CNRS, Marseille, France

3.5 Buffers, solutions and media

1x M9 minimal medium (for *Bacillus* assays)**:** 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 1mM $MgSO_4$ (high Mg^{2+}), 1 µM $MnCl_2$, 100 µM $CaCl_2$, 10 µM $FeSO_4$ and 1x 18aa mix solution, 50 µg/ml L-Lysine, 20 µg/ml L-Tyrosine.

1x M9 minimal medium for NMR (protein overexpression in *E. coli*): 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).

1x M9 minimal medium for SILAC: 1x M9 minimal medium + 50 µg/ml ($^{12}C_6{}^{14}N_2$)-L-Lysine (light, naturally occurring isotope) or 50 µg/ml ($^{13}C_6{}^{15}N_2$)-L-lysine ("heavy" isotope, (Silantes GmbH, Germany).

3.5x BisTris buffer: 26.2 g/l BisTris, pH 6.5.

4x Orange G loading dye: 25% (w/v) Orange G sodium salt, 20% (w/v) Sucrose.

4x Rhomboid cleavage buffer (for *in vitro* activity assays)**:** 200 mM HEPES-NaOH pH 7.4, 40% (v/v) glycerol, 200 mM EDTA.

5x M9 salts: 34 g/l $Na_2HPO_4.2H_2O$, 12 g/l KH_2PO_4, pH 7.4, 2.5 g/l NaCl, 5 g/l NH4Cl.

6x SDS-PAGE reducing sample buffer: 0.35M Tris-HCl, pH 6.8, 30% (v/v) glycerol, 10% (w/v) SDS, 60 µl/ml β-mercaptoethanol, 0.12 mg/ml bromphenol blue.

10x *Bacillus* T-base: 183 g/l K_2HPO4, 60 g/l KH_2PO_4, 20 g/l $(NH_4)_2SO_4$, 10 g/l sodium citrate.

12% BisTris polyacrylamide gel: 1.2 ml of 40 % Acrylamide, 1.6 ml of 2 % bis solution, 1.2 ml of 3.5x BisTris buffer, 2 µl TEMED, 20 µl ammonium persulfate, 0.2 ml MilliQ water.

100x 18aa mix solution (-Lys, -Tyr)**:** Recipe for amino acid mix solution was adopted from Table 2.4 in [194] and modified.

Amylose elution buffer: Buffer C + 10 mM D-(+)-maltose.

***Bacillus* SDS-PAGE lysis buffer:** 20 mM Tris-HCl pH 7.5, 10 mM EDTA, 1 mM $MgCl_2$.

***Bacillus* MS lysis buffer:** Buffer A, 0.4x MS-SAFE inhibitors cocktail, 0.05 mg/ml lysozyme, 250U Pierce Universal Nuclease, 1mM EDTA

BisTris-MES running buffer: 10.6 g/l MES, 6 g/l Tris base, 0.372 g/l EDTA, 1 g/l SDS.

Buffer A: 20 mM HEPES-NaOH pH 7.4, 10% (v/v) glycerol, 0.1M NaCl.

Buffer B: 20 mM HEPES-NaOH pH 7.4, 10% (v/v) glycerol, 0.3M NaCl.

Buffer C: Buffer B + 0.05% (w/v) DDM.

Fixing solution: 24.2 g/l boric acid, 44.5 g/l, $Na_2HPO_4.2H_2O$, 0.25% (v/v) Glutaraldehyde, pH 6.2.

Glutathione elution buffer: Buffer C + 10 mM reduced Glutathione.

LB rich medium: 10 g Tryptone, 5 g Yeast extract, 10 g NaCl, with 1M NaOH to pH 7.

MS Buffer: 20 mM HEPES-NaOH pH 7.4, 0.3 M NaCl.

NiNTA elution buffer: Buffer C + 10, 25, 50 or 250 mM imidazole.

NMR buffer: 7.5 mM K_2HPO4, 17.5 mM KH_2PO_4, pH 6.5 containing 150 mM NaCl and 5% D_2O/ 95% H_2O.

Razor Blue Stain: 50mg/l Coomassie Brilliant Blue G-250, 3ml/l acetic acid.

SpC medium: 1x Bacillus 10x T-base, 0.5% (w/v) D-Glucose, 0.2% (w/v) Yeast extract, 1x 18aa mix solution, 50 µg/ml L-Lysine, 20 µg/ml L-Tyrosine, 0.9 mM $MgSO_4$.

SpII + EGTA medium: SpII medium + 2 mM EGTA pH 8.

SpII medium: 1x Bacillus 10x T-base, 0.5% (w/v) D-Glucose, 0.1% (w/v) Yeast extract, 1x 18aa mix solution, 50 µg/ml L-Lysine, 20 µg/ml L-Tyrosine, 5 mM $MgSO_4$, 0.5 mM $CaCl_2$.

TAE buffer: 4.84 g/l Tris base, 0.372 g/l EDTA, 11.4% (v/v) acetic acid, pH 8.4.

TBE buffer: 10.8 g/l Tris base, 0.744 g/l EDTA, 55 g/l boric acid, pH 8.3.

TEV buffer: 5 mM Tris-HCl, pH 8, 1 mM EDTA, 1 mM β-mercaptoethanol.

Titration Buffer: 50 mM Tris, pH 7.3, 100 mM NaCl.

Transcription buffer: 120 mM HEPES-NaOH pH 7.5, 40 mM DTT, 2 mM spermidine.

Tris-Glycine running buffer: 5 g/l Tris base, 18.8 g/l Glycine, 1 g/l SDS, pH 8.8.

3.6 DNA manipulation

3.6.1 DNA cloning

Genes were PCR amplified by Q5® High-fidelity DNA polymerase (NEB) using gene- and vector-specific primers (synthetized by Sigma-Aldrich, listed in **Table 1**) in PCR thermocycler (Biometra). Typically, 50 µl PCR reaction contained 0.5 mM forward and reverse primer; 25 µl of Q5® High-Fidelity 2x Master Mix and template DNA: 10 ng of plasmid and 100 ng of genomic DNA, respectively. PCR program was set up as follows: 1st step: 98°C/30 s, 30x (2nd step: 98°C/10 s, 3rd step: 56°C/ 20 s, 4th step: 72°C/ 50 s) and final 5th step: 72°C/ 2 min. If necessarry, PCR product were further purified using QIAquick PCR Purification Kit (section 3.1.4). The quality and quantity of PCR products was analysed using horizontal agarose gel electrophorsesis (section 3.6.2) and measurement of absorbances A_{280} and A_{260}, respectively, by NanoDrop™ One Microvolume UV-Vis Spectrophotometer (ThermoFisher Scientific).

DNA plasmids were double digested with restriction endonucleases (NEB), in reaction typically containing 20U of each enzyme in conditions recommended by manufacturer at 37°C, overnight. Digested vectors were dephosphorylated, using 25U of antarctic phosphatase (NEB, #M0289) at 37°C, for 1 hour, separated in agarose gel electrophoresis and purified using QIAquick Gel Extraction Kit (QIAGEN, #28115).

PCR-amplified inserts were ligated by isothermal overlap assembly [195] into particular expression or integration vector, using Gibson assembly® Master Mix (NEB, #E2611), following manufacturer's recommendations and ligated constructs were transformed into *E. coli* DH5α (section 3.6.3). All DNA constructs (**Table 2**) were verified by Sanger sequencing (Eurofins Genomics).

3.6.2 Horizontal agarose gel electrophoresis

The 1% (w/v) TAE-buffered agarose gel, supplemented with 10000x GelRed dye (Biotium) was used for separation and visualisation of all DNA products. Prior to loading on the gel, DNA sample was mixed with 4x Orange G loading dye. Electrophoresis proceeded in horizontal orientation at constant 120V for 45 min. All *in vitro* transcribed RNA products were separated on 1.3% (w/v) TBE-buffered agarose gel, supplemented with Ethidium Bromide, at constant 125V for 45 min. DNA was detected by UV Transluminator Quantum ST4 (Vilber Lourmat). To estimate the size of DNA or RNA product, commercially available DNA markers- Hyper Ladder™50bp or 1kb (Bioline), respectively, were used.

3.6.3 DNA transformation in *Escherichia coli*

Competent *Escherichia coli* were transformed with plasmid DNA (assembly mixture or purified vector for DNA preparation) following either heat shock, or electroporation protocol. In heat shock protocol, an aliquot of chemically competent cells was mixed with sufficient amount of plasmid DNA, incubated for 20 min on ice, then for 1 min at 42°C and heat shocked for 2 min on ice. Highly competent electrocompetent cells were mixed with plasmid DNA, electroporated by 1700V 5ms pulse, using Multiporator® (Eppendorf). Cell suspension was then diluted 10x with sterile LB medium, shaked for 1 hour at 37°C, streaked onto sterile solid LB agar plate, supplemented with an apropriate antibiotic(s) and incubated at 37°C, overnight.

3.6.4 DNA transformation in *Bacillus subtilis*

Competent *Bacillus subtilis* were prepared as described previously [194], with minor modifications. Briefly, particular strain was streaked from -80°C glycerol stock culture onto fresh solid LB agar plate, supplemented with appropriate antibiotics, grown overnight. Single colony was then inoculated into fresh liquid SpC medium (section 3.5) and grown for 5 hours at 37°C, with vigorous aeration. Culture was then diluted 10x into 20 ml of fresh SpII medium (section 3.5) and incubated for additional 90 min. Culture was centrifuged at 2000 xg, for 10 min, at 25°C and pellet resuspended back in 2 ml of the supernatant. Competent cells were diluted 2x with SpII+ EGTA medium, mixed with linear DNA (ScaI-linearized plasmid DNA [Table 2] or isolated bacilli chromosomal DNA [Table 3]), incubated at 37°C for 30 min, diluted 2x with 1x T-base + 0,5% (w/v) glucose solution, incubated for another 30 min. Finally, cell suspension was streaked onto fresh solid antibiotic-supplemented LB agar plate and incuibated at 37°C, overnight.

3.6.5 Isolation of DNA

Plasmid DNA was produced in *Escherichia coli* DH5α and prepared by QIAprep Spin Miniprep Kit (QIAGEN) or NucleoBond® Xtra Midi kit (Macherey-Nagel). Both, *E. coli* and *B. subtilis* chromosomal DNAs were isolated by GenElute™ Bacterial Genomic DNA Kits (Merck). Restriction-digested linear plasmid DNA was purified by horizontal agarose gel electrophoresis and gel-extracted, using QIAquick Gel Extraction Kit (QIAGEN). All isolation procedures followed manufacturers' recommendations.

3.7 Bacterial growth conditions

3.7.1 *Escherichia coli* growth

Routinely, *Escherichia coli* strains were cultivated in LB medium (for composition, see section 3.5) or in minimal M9 medium (section 3.5). Several colonies of overnight LB agar plate culture were inoculated into fresh LB medium, grown @ 37°C, overnight with gentle shaking. Culture was diluted to starting OD_{600} around 0.05 into fresh growth medium and incubated @ 37°C until OD_{600} reached value 0.6- 0.8. For induction of protein overexpression from *lac* promoter, 0.5 mM IPTG (SciTech) was added at this time point, and cultivation continued at protein-specific temperature and time conditions.

For unlabeled protein production, *E. coli* cells were cultivated in LB rich medium. For ^{15}N or $^{15}N^{13}C$ uniformly labeled protein production, cells were cultivated in minimal M9 medium (section 3.5), supplemented with 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).

3.7.2 *Bacillus subtilis* growth

Routinely, *Bacillus subtilis* strains (section 3.4, Table 3) were cultivated, as follows: single colony from overnight LB agar plate was inoculated into fresh LB medium and pre-culture was grown at 37°C, with vigorous aeration, until it reached OD_{600} around 1.0. Pre-culture was then diluted to starting OD_{600} of 0.025, into fresh LB or minimal M9 medium, respectively and cultivation continued until the OD_{600} reached values ranging from 0.5 to 2, at the exponential growth stage. In case of protein overexpression, proteins were expressed from ectopic sites *amyE* or *xkdE*, from xylose (P_{XYL}) or hyperspank ($P_{HYPERSPANK}$) promoters. Expression was induced by 1% (w/v) D-(+)-xylose (Merck), or 0.1 mM IPTG (SciTech).

For SILAC-based quantitative proteomics, *Bacilli* strains auxotrophic for Lysine (section 3.4, Table 3) were grown in modified minimal M9 medium, supplemented with natural "light" ($^{12}C_6{}^{14}N_2$) or unnatural "heavy" ($^{13}C_6{}^{15}N_2$) isotopic variants of L-lysine (Silantes GmbH, Germany). Exponentially growing cells were harvested when OD_{600} reached ~ 1.

In divalent cation stress assays, *Bacilli* strains were cultivated in minimal M9 medium, in the presence of increasing concentrations of divalent salts of magnesium, manganese, zinc, cobalt, nickel and calcium, respectively, added to the culture at different stages of growth, either at the beginning of cultivation or during exponential phase, as it is specified in Results section (4.2.3, 4.2.7).

In vivo rhomboid activity assays were performed either in LB rich or M9 minimal medium with defined concentrations of divalent cations. *Bacilli* LB pre-cultures were washed three times in 1x M9 salts solution and diluted into the minimal medium. Cultures were grown until exponential phase, which usually took 1.5 to 2 hours in LB and 4 to 5 hours in M9 medium, respectively, at 37°C with vigorous aeration and samples were taken for immunoblot analysis.

In translation shut-off assay, *Bacilli* strains were cultivated at 30°C, with vigorous shaking in rich LB medium, until OD_{600} reached 0.5. YqgP expression from $P_{hyperspank}$ was then induced by 0.1 mM IPTG, for 30 min. After that, tetracycline was added to final concentration 20 µg/ml, what inhibit protein translation and samples were taken at defined time points.

3.8 Protein expression and isolation

3.8.1 *In vitro* protein production

^{35}S-Methionine radiolabelled rhomboid substrate corresponding to *Providencia stuartii* TatA 2-97 sequence, its single-point mutants in positions T4 to A8, as well as unlabelled *Bacillus subtilis* MgtE-derived fragments corresponding to first 300, 315, 330, 340, 355 and 370 residues, were generated by *in vitro* translation using Wheat Germ Extract (Promega) as described previously [75, 196].

3.8.1.1 *PCR and in vitro transcription*

PCR products coding for particular gene were amplified with pair of specific primers containing SP6 phage promoter sequence, Kozak sequence and start codon in forward and stop codon in the reverse primer (section 3.4, Table 1). To resolve substrate from product in gel-based *in vitro* activity assay (section 3.11.1), additional GSGSGSGS linker sequence was introduced after start methionine. To increase radiolabelling effectivity, four additional methionine residues were introduced at the 3' terminus of the construct. Typical 40-µl PCR reaction was composed of 20 µl of PPP Master Mix (Top-Bio), 2.4 µl of 10 µM forward and reverse primers and 10 ng of template plasmid DNA. PCR program was set up as follows: 1. step: 94°C/1 min, 35x (2. step: 94°C/1 min, 3. step: 52°C/ 30 s, 4. step: 72°C/ 30 s) and final 5. step: 72°C/ 10 min. PCR products were used for *in vitro* transcription, directly.

In vitro transcription reaction contained 8 µl of 5x Transcription buffer, 6 µl of 20 mM NTP solution, 8 µl of 100 mM magnesium acetate, 4 µl of 5mM G(5')ppp(5')G RNA Cap Structure analogue (NEB), 0.1U Yeast inorganic pyrophosphatase (NEB), 20U of SP6 RNA polymerase (NEB), 2.5U of RNasin

(Promega) and 40 ng/µl of specific PCR product in total volume of 40 µl. Transcription reaction was incubated at 42°C/ 2 hours. To purify and concentrate transcription product, reaction was mixed with 18 µl of 8M $LiCl_2$ and incubated at -20°C for 30 min. The pellet of precipitated mRNA was washed in 70% (v/v) ethanol and dissolved in RNAse free water. RNA purity and concentration were determined by horizontal agarose gel electrophoresis (section 3.6.2) and by measuring absorbances A_{280} and A_{260} on NanoDrop™ One Microvolume UV-Vis Spectrophotometer (ThermoFisher Scientific).

3.8.1.2 *In vitro translation*

In vitro translation reaction contained 12.5 µl of Wheat Germ Extract (Promega), 2.5 µl of 1M potassium acetate (Promega), 2 µl of 1 mM Amino Acid Mixture Minus Methionine (Promega), 1µl of 25mM magnesium acetate, 1.25 µl of (^{35}S) L-Methionine (MGP) or 1 µl of 200 µM natural L-Methionine (Merck) and 100-200 ng/ µl of purified mRNA, in total volume of 25 µl. Translation reaction proceeded at 25°C for 2 hours. Translated radiolabeled rhomboid substrates were used in *in vitro* rhomboid activity. Translated MgtE fragments were used as molecular weight references in MgtE cleavage sites mapping.

3.8.2 Heterologous expression of membrane proteins in *Escherichia coli*

Recombinant active full-length bacterial rhomboids GlpG and YqgP or catalytic double mutant GlpG S201A, H254A and single mutants GlpG F146A and GlpG F146I, respectively, were expressed as C-terminally six-histidine tagged constructs from pET25-b vector, in *Escherichia coli* C41(DE3)- strain optimized for overexpression of membrane proteins [197]. Expression was induced with 0.5 mM IPTG at culture OD_{600} ~ 0.6 and cultivation continued at 16°C, with gentle shaking, overnight. Overexpressed proteins were isolated as described in sections 3.8.4 and 3.8.5.

Recombinant rhomboid artificial transmembrane substrates, derived from 2nd transmembrane domain of *Escherichia coli* LacY protein (LacYtm2) [74] and from *Providencia stuartii* TatA protein [78], and their transmembrane region mutants were cloned as MBP-3xFLAG-*substrateTM*-Trx-Stag-6xHis fusion proteins containing MBP (*E. coli* Maltose-binding protein) and Trx (*E. coli* Thioredoxin) mobility tags and 3xFLAG, S-tag and 6x Histidine affinity and detection tags. Fusion constructs were expressed from pMALp2E in *E. coli* strain JB17, derived from MC4100 (section 3.4, **Table 3**). Expression of the substrate variants was induced with 1 mM IPTG at culture OD_{600} ~ 0.6 and cultivation continued at 25°C, with gentle shaking, overnight. Overexpressed proteins were isolated as described in sections 3.8.4 and 3.8.5.

Overexpression of artificial transmembrane substrates in wt *E. coli* MC4100 and MC4100 *glpG::tet* strains, for *in vivo rhomboid* activity assay (section 3.11.2), was induced with 1mM IPTG at culture OD_{600} ~ 0.5 at 37°C for 3 hours. Samples were analysed by SDS-PAGE and Immunoblotting with quantification of fluorescence signal (sections 3.9.2, 3.9.4 and 3.9.6).

3.8.3 Heterologous expression of soluble proteins in *Escherichia coli*

Extramembrane domains of YqgP and MgtE were prepared for biophysical characterization and antibody production as follows: The N-terminal cytosolic domain of YqgP ($YqgP_{NTD}$; YqgP1-177; NTD) was expressed as construct containing C-terminal six-histidine-tag and TEV site (NTD-TEVsite-6xHis) from pET25b in *E. coli* BL21Star™(DE3). The C-terminal periplasmic domain of YqgP ($YqgP_{CTD}$; YqgP384-507; CTD) was expressed as N-terminally six-histidine-tagged and TEV site containing construct (6xhis-TEVsite-CTD) from pHIS-2 vector in *E. coli* Lemo21(DE3). The N-terminal cytosolic domain of MgtE (MgtE2-275) was expressed as GST-6xHistidine-TEVsite-MgtE2-275 construct from pGEX6P1 vector in *E. coli* Lemo21(DE3). In all cases, the overexpression was induced by addition of 0.5 mM IPTG at OD_{600} ~0.8 at 25°C with gentle shaking, overnight. Overexpressed proteins were isolated as described below.

3.8.4 Subcellular fractionation of *Escherichia coli*

Protein-overexpressing *E. coli* cultures were harvested by centrifugation at 6000 xg/ 15 min/ 4°C. Cell pelets were resuspended in buffer A, supplemented with 1 mM EDTA and 1 mM PMSF (section 3.5) and disrupted by passing three times through EmulsiFlex®-C3 High Pressure Homogeniser (Avestin). The suspension was cleared from cell debris by low speed centrifugation at 15000 xg/ 30 min/ 4°C and supernatant, containing cytosolic and membrane fraction, was used. To isolate overexpressed soluble affinity-tagged protein, supernatant was filtered through paper filter and used for affinity chromatography, directly. To isolate membrane protein fraction, low speed supernatant was further fractionated by ultracentrifugation at 100 000 xg/ 2 hours/ 4°C. Isolated membranes-enriched pellet was washed once in fresh 0.1 M Na_2CO_3 and once in 1M NaCl to get rid-off peripheral membrane proteins and resuspended in buffer B, supplemented with 1x Complete™ EDTA-free Protease Inhibitor Cocktail (section 3.5). To solubilize membrane proteins, washed pellet was mixed with 1.5% (w/v) n-dodecyl-β-D-maltoside (DDM, Avanti) detergent and incubated for 30 min at 4°C with rolling, followed with short ultracentrifugation at 100 000 xg/ 30 min/ 4°C to separate unsolubilized fraction. Cleared soluble membrane protein fraction was further purified by affinity chromatography.

3.8.5 Purification of proteins for biophysical and kinetic experiments

3.8.5.1 *Ni-NTA affinity chromatography*

Fractions containing six-histidine-tagged soluble ($YqgP_{NTD}$, $YqgP_{CTD}$ and MgtE2-275) or membrane proteins (GlpG and its mutants, YqgP, TatA-derived and LacYtm2-derived chimeras) were loaded onto Ni-NTA Agarose (Qiagen), equilibrated in buffer C + 10 mM imidazole, in gravity flow arrangement. His-tagged proteins were specifically eluted by increasing concentration of imidazole present in Ni-NTA elution buffer, ranging from 25 mM to 250 mM. Each fraction was analysed by SDS-PAGE, for presence of isolated protein. Specific for membrane proteins purification, 0.05 % (w/w) DDM was present in all equilibration and elution steps.

3.8.5.2 *Amylose affinity chromatography*

Second purification step was introduced to purify MBP-tagged TatA- and LacYtm2-derived chimeric rhomboid substrates to homogeneity. Ni-NTA elution fractions containing desired protein were joined, buffer-exchanged using PD desalting column (GE Healthcare) into buffer C, loaded onto Amylose Resin (NEB) packed into gravity flow column, equilibrated in the same buffer and specifically eluted by 10 mM maltose, present in Amylose elution buffer. Course of the purification was monitored by SDS-PAGE.

3.8.5.3 *Glutathione sepharose affinity chromatography*

GST-6xHis-tagged MgtE was purified to homogeneity in second step using Glutathione Sepharose® 4 Fast Flow (GE Healthcare Life Sciences). Ni-NTA elution fractions containing MgtE domain were joined, buffer exchanged into buffer C, loaded onto Glutathione Sepharose column and specifically eluted by 10 mM reduced Glutathione present in the Glutathione elution buffer. Course of the purification was monitored by SDS-PAGE.

3.8.5.4 Affinity tag removal

6x-Histidine and GST-6x-Histidine tags were removed from purified proteins by incubation with 6x-Histidine-tagged TEV protease. Briefly, TEV protease was added to protein solution in 200:1 protein-to-protease ratio (w/w), in TEV buffer, for 16h at 25°C. Protein solution was then applied on PD desalting

column, buffer-exchanged with buffer C + 10 mM imidazole and loaded onto Ni-NTA Agarose. Course of the tag removal was detected by SDS-PAGE and immunoblotting. Tag-free protein, which was present in flow-through fraction, was collected, concentrated by ultrafiltration (Vivaspin concentrating units) and used for further applications.

3.8.5.5 *Size exclusion chromatography*

Eventually, proteins coming from affinity purification were additionally purified by size exclusion chromatography. Respecting its molecular weight, concentrated protein solution was applied either on Superdex 75 or Superdex 200 column connected to FPLC system. Course of the chromatography was monitored by SDS-PAGE.

3.8.5.6 *Purification of polyclonal antibodies*

Anti-MgtE, anti $YqgP_{NTD}$ and anti-$YqgP_{CTD}$ polyclonal antibodies were produced in rabbit (Agro-Bio). Tag-free MgtE2-275, YqgP1-177 and YqgP384-507 proteins were isolated, as described above, and used for rabbit immunization. Antibodies were then purified from the blood sera using Affigel 10 activated affinity medium (Bio-Rad) coupled via primary amines to MgtE2-275 protein, and Affigel 15 activated affinity medium (Bio-Rad) coupled either to YqgP1-177 or YqgP384-507, respectively. Purification was performed as recommended by manufacturer.

3.8.5.7 *Protein demetallation*

To remove any contaminating divalent metals from protein preparations that could affect the analysis in downstream applications, both $YqgP_{NTD}$ and $YqgP_{CTD}$ proteins were incubated with 20 mM EDTA at 25°C, overnight. Subsequently, EDTA was removed by ultrafiltration [198], using Vivaspin concentrators (Sartorius). In the final step, protein solution was loaded onto PD-10 desalting column (GE Healthcare) and buffer was exchanged for Titration buffer (section 3.5).

3.8.6 Isolation of *B. subtilis* membrane subproteome for SILAC-based quantitative proteomics

Bacilli cultures were grown in SILAC minimal M9 medium (section 3.5). In experiment 1, heavy lysine-labelled and light lysine-labelled exponentially growing Bacillus strains BS50 and BS51 (Table 3)

expressing YqgP or catalytic mutant YqgP S288A, respectively, were mixed in OD_{600} ratio 1:1. In control experiment 2, cell cultures were SILAC-labelled in reverse orientation: light BS50 and heavy BS51.

The heavy+ light cell suspensions were harvested by centrifugation at 4000 xg/ 15 min/ 25°C. Cell pelets were resuspended in *Bacillus* lysis buffer (section 3.5) and incubated first at 37°C for 15 min, then on ice for 15 min, with episodic gentle tube converting. Protoplasts were then disrupted by sonication, using sonicator Sonopuls HD2200 (Bandelin) set to 4×30 s pulse cycle and 30% power. Cell debris was separated by centrifugation at 15 000 xg/ 30 min/4°C and cleared supernatant ultracentrifuged at 100 000 xg/ 1.5 hour/4°C. Pellets, containing crude membranes, were washed with 0.1 M Na_2CO_3 and 1 M NaCl and resuspended in MS buffer (section 3.5).

3.9 Protein analysis

3.9.1 Protein concentration determination

Protein concentration was estimated by colorimetric Pierce™ 660 nm Protein assay (ThermoFisher Scientific), as recommendet by the manufacturer. Precise protein concentration was determined by measurement of absorbance at 280 nm using extinction coefficient derived from amino acid analysis (provided by core facility at IOCB AS CR). Final detergent concentration in membrane protein sample was determined by phenol and sulfuric acid-based colorimetric assay, as described elsewhere [199].

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

Samples for SDS-PAGE were prepared as follows: *E. coli* cell culture samples were centrifuged at 5000 xg/ 10 min and pellets resuspended in 1x SDS-PAGE reducing sample buffer. Purification fractions were mixed with 6x SDS-PAGE reducing sample buffer. *B. subtilis* cell culture samples were centrifuged at 5000 xg/ 10 min and pellets were resuspended in *Bacillus* SDS-PAGE lysis buffer, containing 1 mg/ml Lysozyme (Merck) and 0.25 U of Universal Nuclease and incubated at 37°C for 10 min. Cell lysate was then mixed with 6x SDS-PAGE reducing sample buffer. To wash concentrate the radiolabelled protein sample prior to SDS-PAGE, 20% (v/v) trichloroacetic acid (TCA) solution was added to 1:1 volume ratio. After 30 min incubation at -20°C, mixture was centrifuged at 20 000 xg/10 min/ 4°C. Protein-containing TCA precipitate was washed twice in acetone and resuspended in 1x SDS-PAGE sample buffer. All SDS-PAGE samples were heated at 60°C for 10 min.

To compare protein content in between different conditions, equivalent amounts of material were loaded within particular experiment, as estimated by OD_{600} read-outs or determined by

spectrophotometric measurement. Depending on the protein molecular weight, samples were resolved either on 4-20% Tris-glycine SDS-PAGE system (Bio-Rad) or on 12% BisTris-MES SDS-PAGE system. Electrophoresis was terminated after samples' front reached the gel end. Color Prestained protein Standard, Broad Range (11-245 kDa) (NEB) was used as molecular weight standard in all SDS-PAGE experiments.

3.9.3 Coomassie staining of proteins in polyacrylamide gel

For routine protein in-gel visualisation, polyacrylamide gel was immersed in demi water and heated in microwave oven until boiling. Gel was then rinsed twice in fresh demi water, immersed in Razor Blue Stain solution, containing Coomassie Blue G-250 (section 3.5), heated until boiling and rocked for 10 min at room temperature. Last step was then repeated with demi water, for background Coomassie destaining. For gel-based *in vitro* activity staining, Coomassie-based InstantBlue™ Protein Stain (Expedeon) was used, following protocol recommended by manufacturer.

3.9.4 Immunoblot analysis

To detect protein expression, stability or analyse rhomboid-mediated proteolysis in samples of cell cultures, purification fractions or rhomboid activity assays were electrotransferred onto NC2 Nitrocellulose Membrane (SERVA) or Immobilon-P PVDF Membrane (Merck) by Mini Trans-blot cell module (BIO-RAD). Constant voltage of 100V was applied for 2 hours long transfer. Blots with transferred proteins were quantified for total protein content using REVERT™ Total Protein Stain Kit (LI-COR) and fluorescence detection. Afterwards, blot membrane was blocked by Casein in Blocker™ Casein in TBS for 1 hour at room temperature. Blots were incubated with primary antibodies (section 3.1.3) for 1 hour at room temperature or overnight at 4°C, depending on their sensitivity. For signal amplification, blot was incubated either with HRP-conugated or fluorescent secondary antibodies, both diluted in Blocker™ Casein in TBS solution, at room temperature, for 1 hour. Chemiluminescent signal was generated by incubation of blot with Luminata™ Crescendo or Forte Western HRP Substrates (Merck) and subsequently detected by CCD camera LAS-3000 (Fujifilm). Fluorescence blots were visualised on the Odyssey® CLx Imaging System (LI-COR®).

3.9.5 Autoradiography

In vitro translated (^{35}S)-labelled rhomboid substrate-containing samples were resolved on 12% BisTris-MES SDS-PAGE system. After that, polyacrylamide gel was fixed in Fixing solution (section 3.5) and equilibrated in 40% methanol with 10% acetic acid, for 15 min. Gels were dried under vacuum at 80°C, for 40 min, using Slab Gel Dryer GD2000 (Hoefer Inc.). ^{35}S radiaton signal was developed using Autoradiography Exposure Cassette (GE Amersham) and visualized by Typhoon 9410 Imager (GE Amersham).

3.9.6 Protein signal quantification

Protein abundance signals resulting from chemiluminiscence, radioactivity or fluorescence detection were quantified by densitometry, using ImageQuant 8.0 (GE Healthcare) or Image Studio Lite ver 5.2 (LI-COR Biosciences), respectively. Data were used for substrate conversion calculation.

3.9.7 SILAC-based quantitative proteomics

Bacillus subtilis transmembrane protein-enriched membrane fractions (for isolation details, see section 3.8.6) were resolved on 4-20% Tris-glycine SDS-PAGE (Bio-Rad), in separate gel lanes, which were the sliced to fractions A1-E1 for experiment 1 and A2-E2 for experiment 2 (section 3.8.6). Each slice was then analysed in standard GeLC-MS/MS experiment, by our collaborator Jana Březinová at IOCB AS CR. She also performed all quantitative MS data analyses, using MaxQuant software (section 3.3), as well as the database search and bioinformatic validation of candidate substrates, using Uniprot/ Swissprot *B. subtilis* database (downloaded 17/05/15) and QARIP software [191], respectively.

3.10 Protein biophysical characterization

3.10.1 Protein structure determination by solution NMR

NMR spectra for solution structure determination of YqgP N- and C-terminal domains were acquired from 350 µl samples of 1 mM ^{15}N/^{13}C uniformly labelled $YqgP_{NTD}$ and 0.7 mM ^{15}N/^{13}C uniformly labelled $YqgP_{CTD}$. All 1D, 2D and 3D spectra were measured in 25 mM phosphate buffer, pH 6.5 containing 150 mM NaCl and 5% D_2O/ 95% H_2O and collected on 850 MHz Bruker Avance spectrometer (Bruker BioSpin GmbH) equipped with triple resonance (^{15}N, ^{13}C, ^{1}H) cryoprobe, at 37°C.

Sequence-specific backbone resonances were assigned from collected spectra $^{15}N/^{1}H$ HSQC, HNCO, HNCACB and CACB(CO)NH. Resonances of aliphatic side-chain carbons and protons were assigned from HCCH-TOCSY, ^{15}N-edited NOESY, ^{13}C-edited NOESY and ^{15}N-edited TOCSY spectra. Resonances of aromatic ring protons were assigned from 2D-TOCSY and 2D-NOESY spectra. ^{1}H-^{1}H distance constraints were calculated using 3D-$^{15}N/^{1}H$ NOESY-HSQC and $^{13}C/^{1}H$ NOESY-HMQC spectra. All NOESY spectra were collected with NOE mixing time of 100 ms. TOCSY spectra were collected with mixing time of 60 ms. Raw data were processed by Dr. Václav Veverka at IOCB AS CR. Resonances from spectra collected for $YqgP_{NTD}$ were assigned manually by Rozálie Hexnerová (IOCB AS CR) and for $YqgP_{CTD}$ by myself, with guidanance from Dr. Veverka.

Protein structure calculations for both $YqgP_{NTD}$ and $YqgP_{CTD}$ were performed by Dr. Veverka. Using programs Cyana 2.1 for innitial families of converged structures were calculated. Another calculations included generation of backbone torsion angle constraints, using program TALOS+, and additional hydrogen bond constraints. Structures were subsequently refined in explicit solvent wity YASARA forcefield.

3.10.2 Isothermal titration calorimetry

The binding affinity of divalent cations to tag-free (section 3.8.5.4) demetallated (section 3.8.5.7) $YqgP_{NTD}$ domain was determined by titration microcalorimetry using using Auto-iTC_{200} instrumentation (MicroCal, GE Healtcare Life Sciences). Typically, 200 µl of $YqgP_{NTD}$ was titrated stepwise by 2 µl injections of divalent metal cation solution, until saturation by ligand was reached. The protein and cation concentrations were adjusted: 1.5 mM $YqgP_{NTD}$ was titrated by 50 mM $MgSO_4$, $MnCl_2$, $ZnCl_2$ or 50 mM $CaCl_2$, 0.7 mM $YqgP_{NTD}$ by 15 mM $CoCl_2$ and 0.25 mM NTD by 7 mM $NiCl_2$ solutions. Control titrations were performed by adding ligand solution to the buffer. Prior to titration, $YqgP_{NTD}$ sample was dialysed into the same Titration buffer (section 3.5), in which ligand solutions were prepared. Precise protein and ligand concentrations were determined by amino acid analysis and elemental analysis, performed by core facilities at IOCB AS CR. Titration data were processed using MicroCal Origin 7.0 (MicroCal, GE Healthcare Life Sciences). Analyses were performed with help of Dr. Milan Kožíšek, at IOCB.

3.10.3 NMR titration experiments

$^{15}N/^{1}H$ HSQC 2D spectra were acquired for ^{15}N-labelled 400 µM tag-free (section 3.8.5.4) demetallated (section 3.8.5.7) $YqgP_{NTD}$ domain, titrated stepwise with divalent metal salts: 10, 20 and

40 µM $MnCl_2$; 100, 200, 400 and 800 µM $CoCl_2$ and 100, 200, 4000 and 800 µM $NiCl_2$. All titration experiments were performed in Titration buffer (section 3.5). All metal solutions were prepared in this buffer, as well. To map the metal interaction regions in $YqgP_{NTD}$ domain, chemical shifts and intensities changes of backbone signals were assigned individually, comparing each titration double resonance spectrum with control double resonance spectrum of metal-free $YqgP_{NTD}$. Chemical shift perturbation plots (CSP) and intensity plots (I/I_0) were calculated for each titration condition, using program NMRFAM Sparky [189]. Titration data analyses were carried out by Dr. Pavel Srb, at IOCB.

3.11 Rhomboid activity assays

Purified rhomboid variant and either (^{35}S)-radiolabelled *in vitro* translated substrates or chimeric substrate-derived fusion poteins were used, to analyse the substrate preferences and kinetic parameters of GlpG, respectively. To elucidate rhomboid inhibitors' properties, chimeric substrate-derived fusion proteins were overexpressed in cells expressing endogenous GlpG or YqgP proteases.

3.11.1 In vitro assays

Amino-acid sequence preferences of GlpG were analysed as follows: Purified full-length GlpG variant (wt, F146I or F146A and inactive S201A, H254A as control) was reconstituted in 0.05% (w/v) DDM detergent micelles and mixed with (^{35}S)-Met radiolabelled TatA substrate bearing point mutation at one of positions T4 to A8 (TIATA mutated to **X**IATA, T**X**ATA, Ti**X**TA, TIA**X**A, TIAT**X**, where „**X**" stands for one of 19 remaining naturally occuring amino acids).

Normally, the 16-µl protoelytic reaction contained 20 (µg/ml) GlpG, 4 µl of 4x Rhomboid cleavage buffer (section 3.5), 4 µl of 2M NaCl, 1.6 µl of 0.5%(w/v) DDM and 2 µl of *in vitro* translated substrate, which was incubated for 40 min at 37°C. Cleavage was terminated by rapid freezing in liquid nitrogen. Cleavage product were resolved on BisTris-MES SDS-PAGE system (section 3.9.2) and protein signal quantified by autoradiography and densitometry (sections 3.9.5 and 3.9.6). Substrate conversion (α) was the calculated from protein band intensities as the ratio $\alpha = 4*I_P/(4*I_P + 5*I_S)$, where "$I_P$" stands for C-terminal product and "I_S" for full-length substrate gel band intensites, each multiplied (*) by factor characterizing number of radioactive methionines present (5 in FL substrate and 4 in product).

Gel-based assays were also used to determine *in vitro* molar catalytic activity of GlpG towards recombinant purified LacYtm2-derived MBP-3xFLAG-*LacYtm2*-Trx-S-6xHis fusion proteins (section 3.8.2) bearing mutations in recognition motif (HISKS to **R**ISKS, H**V**SKS, HI**R**KS, HIS**H**S, HISK**A**, **RVRHA**).

Isolated recombinant substrates were exposed to full-length GlpG in reaction containing 1x rhomboid cleavage buffer (section 3.5) and GlpG and substrate concentrations were adjusted for each enzyme-substrate pair, to ensure linearity condition during measurement of initial reaction rate. In particular, 100-µl reaction always contained 1.47 µM substrate and 0.8 µM GlpG for wild-type (HISKS), **R**ISKS, H**V**SKS, HIS**H**S, 0.08 µM GlpG for HISK**A** or 0.016 µM GlpG for HI**R**KS and **RVRHA** variants. In 10-minutes intervals, 10-µl samples were taken and freezed rapidly in liquid nitrogen to terminate cleavage reaction. Samples were resolved by 4-20% Tris-Glycine SDS-PAGE system (section 3.9.2). Gels were stained by InstantBlue™ Protein Stain (Expedeon), bands were quantified by densitometry, using program ImageQuant (sections 3.9.3 and 3.9.6) and substrate conversions (α) were defined as ratio $\alpha = MWF*I_P/(MWF*I_P + I_S)$, where $MWF = MW_S/MW_P$ and stands for molecular weight factor that normalizes the quantified signal to molecular weight of full-length substrate (MW_S) and cleavage product (MW_P). Molar catalytic activity was derived from substrate conversion as number of substrate molecules cleaved by one molecule of enzyme per one minute.

3.11.2 In vivo assay in *Escherichia coli*

Recombinant TatA- and LacYtm2-derived MBP-3xFLAG-*LacYtm2*-Trx-S-6xHis fusion proteins were overexpressed in wild type *E. coli* MC4100 expressing endogenous GlpG or in its *glpG::tet* mutant, as described above (section 3.8.2). Signal from fluorescence immunoblot, detecting rhomboid substrate at its C-terminus by primary anti-5his and secondary fluorescent antibodies (section 3.1.3), in cell lysates was quantified (sections 3.9.4 and 3.9.6) and steady-state substrate conversion (α) was calculated as $\alpha = I_P/(I_P + I_S)$, where I_S and I_P are fluorescence intensities of bands corresponding to full-length substrate and C-terminal cleavage product, respectively.

3.11.3 In vivo assay in *Bacillus subtilis*

To test inhibitory potency of peptidyl RVRHA-phenylbutyl ketoamide compound, *Bacillus subtilis* 168 strain BS87 expressing endogenous YqgP and ectopically overexpressing artificial LacYtm2-derived substrate (*ydcA::neo xkdE::AmyE$_{SP}$-MBP-3xFLAG-lacYtm2-Trx-HA)* and its *yqgP::tet* mutant were cultivated in the presence of increasing concentration of inhibitor compound 11, ranging from 0.6 to 50 nM (**Table 3**). Compound 11 was synthesised by our collaborator Dr. Stancho Stanchev, at IOCB AS CR. Full-length substrate and N-terminal cleavage product were detected by immunoblotting using primary anti-FLAG and secondary fluorescent antibodies (section 3.1.3 and 3.9.4) and fluorescent signals were quantified using program Image Studio Lite ver 5.2 (LI-COR Biosciences). Steady state substrate conversions were calculated as described in previous section 3.11.2.

To analyse contributions of soluble domains and manganese binding region on YqgP activity, *Bacilli* strains ectopically expressing YqgP and its domain-truncated variants or single-point mutants and either expressing enogenous MgtE or ectopically overexpressing TatA substrate I5G.I10G as fusion protein, were cultivated (previous paragraph; **Table 3**; section 3.7.2). The primary anti MgtE2-275 and anti-Trx (Merck) and corresponding secondary fluorescent antibodies (section 3.1.3) were used to visualise N-terminal MgtE region and C-terminal TatA fusion protein region, respectively. Steady state substrate conversion was calculated as described in previous section 3.11.2. Specific activity was derived from steady state conversion as the conversion per one molecule of enzyme, which was also quantified from the fluorescence signal.

4 RESULTS

This section outlines the results that I contributed to three publications, and the results of my first author publication. I have divided the Results section into two parts. In the first part, I present data which I contributed to the comprehensive analysis of substrate specificity and mechanism of rhomboid protease GlpG from *Escherichia coli.*

In the second part, I present published results from quantitative proteomics screening for substrates and interactors of rhomboid protease YqgP from *Bacillus subtilis* that helped us decipher the biological role of YqgP in membrane protein quality control and its mechanism.

4.1 Results I: Specificity and mechanism of GlpG from *Escherichia coli*

The scope of **Publication_1** was to analyse the mechanism of substrate binding in the active site and the specificity of rhomboid protease GlpG. We employed structural, enzymological and comuputational approaches to study the substrate preferences of GlpG rhomboid protease from *Escherichia coli*. We developed a series of novel substrate-derived mechanism-based irreversible peptidyl chloromethylketone (CMKs) inhibitors and were able to improve their inhibitory potency by modulating their peptidyl part, which corresponds to the non-prime side sequence of the substrate (p. 14, **Fig. 6**). The structures of the complexes revealed that CMKs interact with GlpG in a substrate-like manner and for the very first time we identified the interactions that occur in the S1 to S4 subsites of the enzyme in detail. We confirmed the importance of the S1 subsite, which protrudes into the so-called „water-retention site" as well as of the S4 subsite, which is formed by residues of the L1 loop and contributes to the enzyme specificity. Structural analysis, supported by *in vitro* activity assays, underlined the role of the F146 residue within the S4 subsite of GlpG, indicating that it interacts with the side-chain of the substrate P4 residue.

I mapped the substrate preferences of GlpG by mutating single amino acids at positions T4-A8 of a model substrate derived from *Providencia stuartii* TatA protein (these amino acids correspond to non-prime side positions P5-P1 according to the Schechter and Berger nomenclature [77]), and I characterised the amino acids highly preferred in the substrate at individual positions (presented in sections 4.1.1 and 4.1.2). These residues were introduced into the peptidyl part of the CMK inhibitors, resulting in dramatic improvement of their inhibitory potency. Lastly, these findings were used to model the likely interaction mode of the substrate-derived heptapeptide spanning the P4-P3' positions within the active site cleft of GlpG.

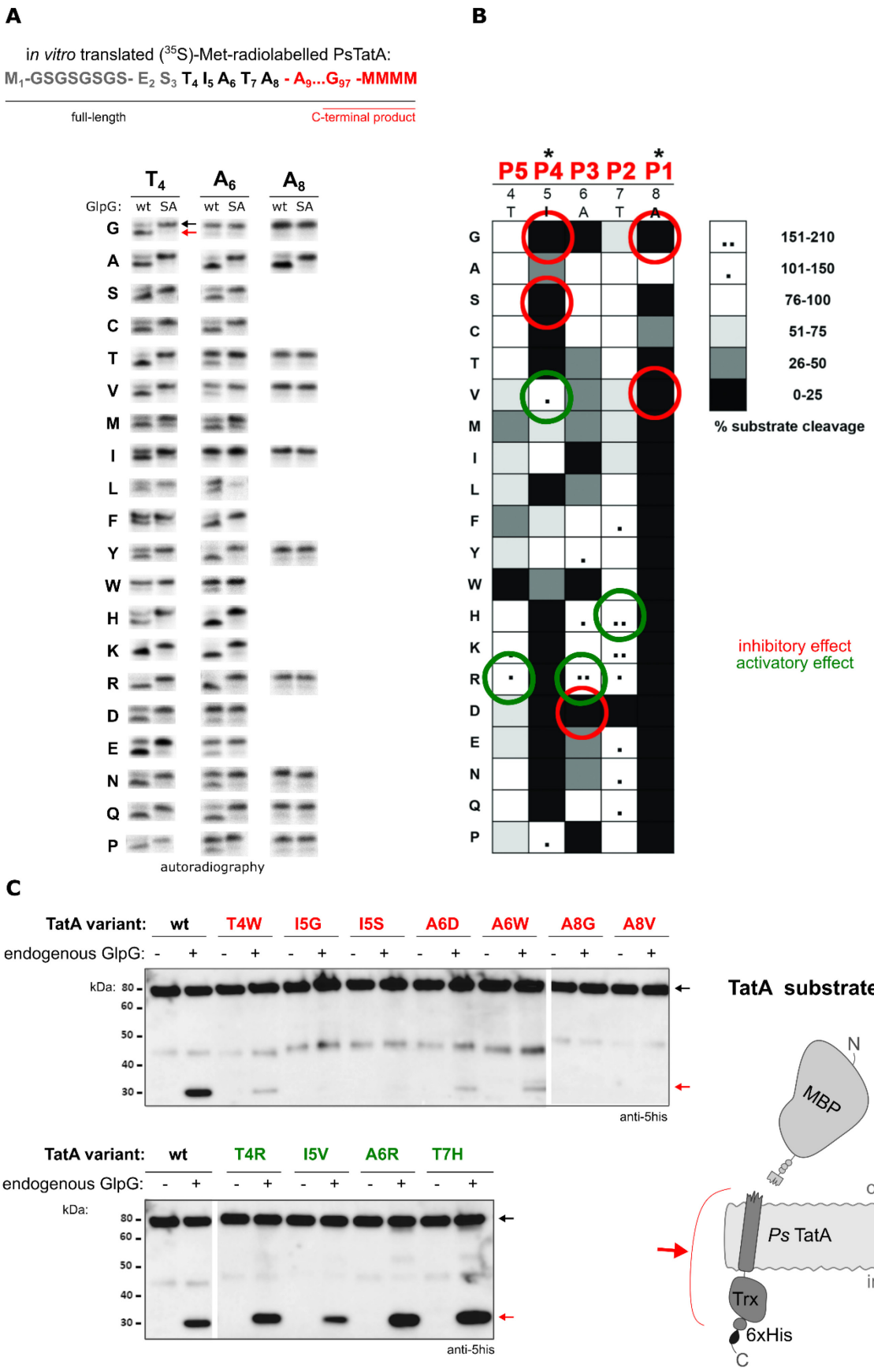
The main goal of **Publication_2** was to develop versatile substrates that could be used in studies of rhomboid kinetics and mechanism, and in inhibitor development. To do this, based on our findings in **Publication_1** on GlpG substrate preferences, we designed fluorogenic transmembrane peptides, with an optimised non-prime side region, some of which were efficiently cleaved by diverse rhomboid proteases while others being specific for GlpG (presented in section 4.1.3). Furthermore, we developed a robust and sensitive *in vitro* rhomboid activity assay, which works in detergent or liposomes.

In **Publication_3** we reported on a general strategy for the development of potent and selective-rhomboid inhibitors, based on the ketoamide warhead moieties with modifiable C-terminal substituents at the amide nitrogen, which cover both the non-prime and prime sides of the rhomboid substrates. Based on our previous results, we were able to significantly enhance the inhibitory potency of tetra- and pentapeptidyl ketoamides by introducing amino acids which are highly preferred by GlpG in the P5

to P1 positions. Hydrophobic substitutents of the amine group of peptidyl-ketoamides resulted in even greater inhibition, and such compounds were highly potent also *in vivo*, inhibiting endogenous GlpG and YqgP rhomboid proteases in living *Escherichia coli* and *Bacillus subtilis*, respectively, at nanomolar concentrations (presented in section 4.1.4). The structures of GlpG in complex with ketoamide inhibitors helped us to explore the S4 to S2‘ active site cavities, and propose avenues for further improvement of the inhibitors.

4.1.1 GlpG shows amino acid preferences in substrate positions P4 and P1.

We performed a comprehensive positional mutagenesis of each of the T4, I5, A6, T7 and A8 residues of the model full length substrate TatA, corresponding to the P5 to P1 non-prime side positions of a generic substrate. We generated a panel of *in vitro* translated radiolabelled TatA single-point mutants, of which I prepared all twenty variants at each of the positions T4, A6 and A8. Each TatA variant was then incubated with purified active GlpG, reconstituted in detergent micelles. The substrate cleavage was analysed by SDS-PAGE (**Fig. 12a**) and autoradiography and quantified by densitometry **Fig. 12b**). Positions P1 and P4, represented by the A8 and I5 residues of TatA, respectively, were the most sensitive to amino acid substitutions. In the case of A8, any mutation except for A8C resulted in complete abortion of rhomboid *in vitro* activity, indicating that GlpG strictly prefers a small non-branched side chain in the P1 position, such as alanine or, to some extent, also cysteine. Mutagenesis of residue I5 of TatA at position P4 had a less dramatic effect than that of residue A8, but was the second most significant compared to the remaining positions. Only mutations affecting the hydrophobic nature of the side chain (such as I5K, I5R, I5D, I5E, I5N or I5Q) resulted in almost complete cleavage inhibition. On the other hand, positions P2, P3 and P5 were not so restrictive and tolerated most of the mutations with no significant effect on GlpG activity. Certain mutations were not tolerated along most of the tested region and almost completely inhibited the cleavage, such as glycine in P1 and P4, tryptophan in P1, P3 and P5 and aspartate in all but the P5 position (**Fig. 12b**, positions highlighted in red circles). These mutations, namely T4W, I5G, I5S, A6D, A8G and A8V had adverse effects also on the cleavage, by endogenous GlpG, of TatA variants expressed as fusion proteins in living *E. coli* (**Fig. 12c**), emphasising the consistency of data acquired in detergent-based *in vitro* assays and in *in vivo* asays in biological membranes. Interestingly, mutations in P5-P1 positions to asparagine, lysine, histidine or valine, such as T4R, I5V, A6R or T7H actually improved the cleavability of the TatA substrate 1.5 to 2-fold *in vitro* (**Fig. 12b**, positions highlighted by green circles) and in living *E. coli* cells (**Fig. 12c**).



(Figure legend text continues on the next page)

Fig. 12: Substrate specificity of rhomboid protease GlpG. (A) Radiolabelled full-length *Providencia stuartii* TatA proteins with single mutations in the T4 to A8 residues were digested by purified active GlpG, or by its catalytic mutant. The products of each cleavage reaction were resolved by SDS-PAGE and visualised by autoradiography. **(B)** A specificity matrix showing GlpG preferences in each position. Substrate conversions are shown in shades of grey: preferred amino acids are in white, non-preferred ones in black. Residues, which improved cleavage, when compared to wild-type TatA, are indicated by black dots. Non-tolerated and tolerated positions, which were further verified *in vivo*, are shown by red or green circles, respectively. **(C)** The TatA-derived MBP-3xFLAG-TatA-Thx-6xHis-S fusion proteins bearing single inhibitory (red) or activatory (green) mutations of the TatA T4 to A8 positions were overexpressed in wt *Escherichia coli* MC4100 or its *glpG::tet* mutant. Cleavage was analysed by immunoblotting and visualised using the anti-5his antibody, which allows detection of the C-terminal cleavage product. In (A) and (C) full length TatA proteins are indicated by black arrows and C-terminal cleavage product by red arrows.

4.1.2 Subsite S4 of GlpG contributes to substrate recognition.

The crystal structure of GlpG in complex with tetrapeptidyl Ac-IATA-CMK, the peptidyl tail of which mimics the P4-P1 non-prime side region of the TatA substrate, provided us with detailed information on the non-covalent interaction network between the peptide and the active site, indicating features which contribute to effective interaction in a substrate-resembling manner (for details see **Publication_1)**. Of the GlpG residues that were involved in the IATA tetrapeptide binding, we focused on F146, located in loop L1 (p.10, **Fig. 4**), which was previously shown to contribute to GlpG activity [50]. In our crystal structure (**Publication_1**), the F146 residue of GlpG interacted with the isoleucine side chain of Ac-IATA-CMK inhibitor, which corresponds to isoleucine I5 in the native sequence of the substrate.

We thus compared the *in vitro* activity of purified wild-type GlpG to that of its F146 mutants, where hydrophobic residues of varying volume were introduced, specifically alanine (F146A) and isoleucine (F146I). The *in vitro* translated radiolabelled TatA variants, with isoleucine I5 mutated to all remaining amino acids, were used as substrates (**Fig. 13**). Both F146A and F146I GlpG variants cleaved substrate mutants with smaller side chains, such as TatA I5A, I5C or I5P, with lower efficiency than wild-type GlpG, as a cost of losing the S4 subsite interaction provided by F146. Conversely, the "complementary" TatA substrate mutation I5F as well as mutations I5M and I5W, all introducing larger hydrophobic side chains, led to a significant increase in substrate processing by both rhomboid mutant variants but not by wild-type GlpG. This was probably due to the restoration of the interaction between the S4 subsite residue of the F146-mutated enzyme with the P4 residue of the mutated substrate (**Fig. 13**). The role of the S4 subsite in substrate binding was further supported by structural analyses of wt GlpG and its F146I mutant in complex with the Ac-FATA-CMK inhibitor, which were performed by the first author of **Publication_1** Dr. Sebastian Zoll and are discussed to some extent also in the Discussion section.

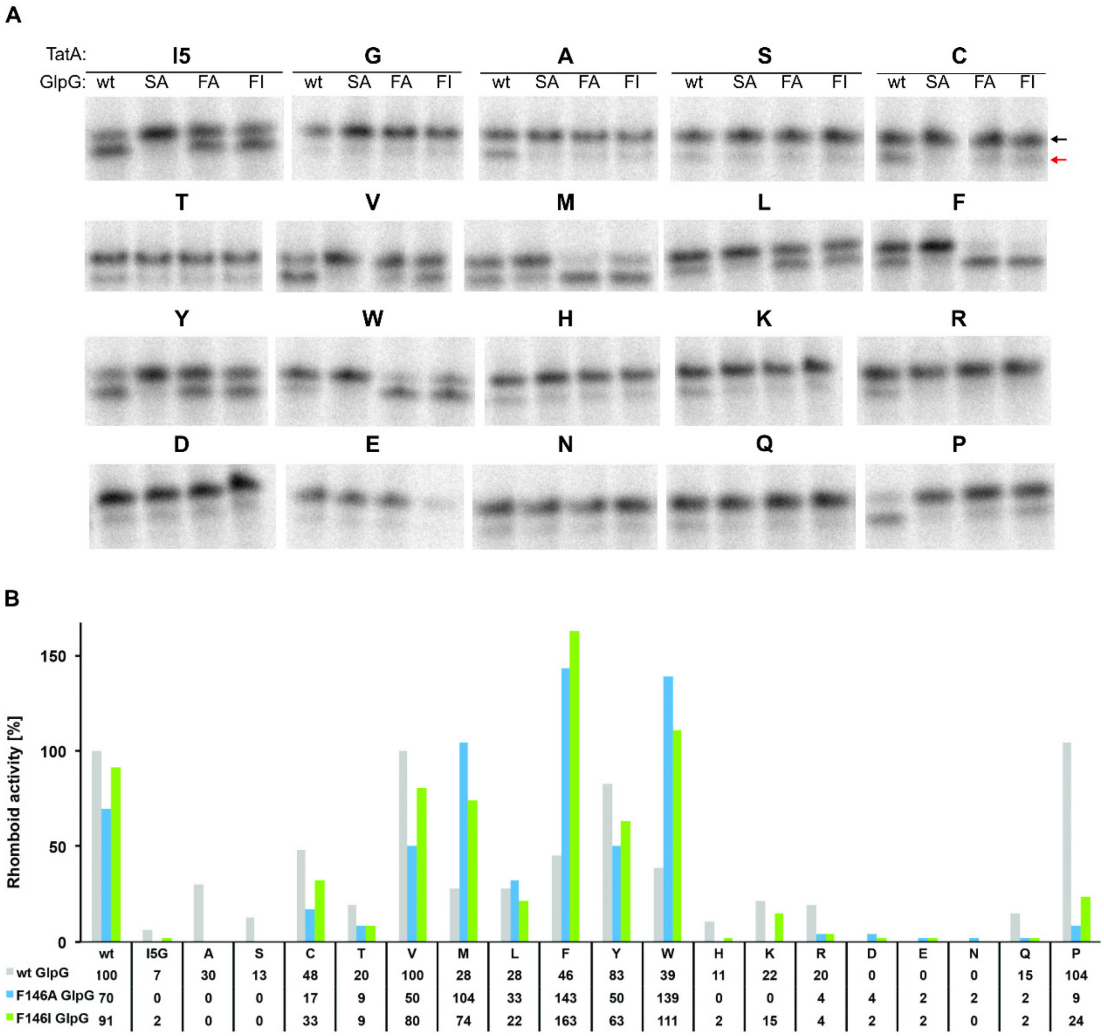


Fig. 13: Comparison of the activity of wild-type GlpG with that of the S4 subsite mutants GlpG F146I and F146A.
(A) Autoradiographic detection of the cleavage of TatA I5 mutants by wt, F146A and F146I GlpG rhomboids *in vitro*, in DDM detergent micelles. Black and red arrows indicate the full-length TatA protein and the GlpG-specific C-terminal cleavage product, respectively. **(B)** Quantification of substrate conversion. The rhomboid activity values represent the conversion of a single TatA I5 mutant by the wild-type (in grey), F146A (in blue) or F146I (in green) GlpG variant. The values are relative to the activity of wt GlpG towards wt TatA, which is set to be 100%. Changing the bulky phenylalanine to the smaller alanine or isoleucine in the F146A or F146I GlpG mutants reduces its activity towards TatA mutants with smaller side chains in the P4 position (I5A, I5C, or I5P). This negative effect can be compensated for by introducing a bulky residue in the TatA substrate, such as in I5M, I5F or I5W. Representative values from one of three independent experiments are shown.

4.1.3 Optimisation of the non-prime side region of the substrate improves cleavability.

Mutations T4R, I5V, A6R, and T7H in the P1-P5 positions of the TatA substrate, which were shown to boost cleavage by GlpG significantly (**Fig. 12**), were introduced into the equivalent positions of the LacYtm2 substrate derived from *Escherichia coli* lactose permease. LacYtm2 has been widely used in rhomboid enzymological studies as a canonical model substrate for diverse rhomboid proteases. Residues HISKS, at positions P5 to P1 of wild-type LacYtm2 substrate were mutated individually to make the RISKS, HVSKS, HIRKS, HISHS and HISKA variants, as well as in combination to make the RVRHA variant. These proteins were then expressed as fusion proteins MBP-LacYtm2-Trx and either isolated and incubated *in vitro* with active GlpG or co-expressed with endogenous GlpG in living *E. coli* cells. The time courses of individual *in vitro* reactions were analysed by SDS-PAGE (**Fig. 14a**), and quantification of substrate conversion was used to calculate initial rates of the reaction (**Fig. 14b**) and subsequently to determine the molar catalytic activity of GlpG on single LacYtm2 mutant substrates (**Fig. 14c**). Although mutations H/R, I/V and K/H in positions P5, P4 and P2 (HISKS to RISKS, HVSKS, HISHS), respectively, did not enhance substrate cleavage, amino acid changes introducing arginine into the P3 (S/R) position in the variant HIRKS and alanine into the P1 (S/A) position in the variant HISKA or the combination of all five mutations resulting in variant RVRHA significantly improved GlpG-dependent processing. Specifically, the activity of GlpG increased 16-fold for HIRKS, 7-fold for HISKA and 64-fold for the RVRHA substrate, when compared with cleavage of the wild-type LacYtm2 substrate (**Fig. 14c**). This additive effect of activatory substitutions in the RVRHA substrate could not be meausred when substrate fusions were assayed *in vivo*, since the steady-state conversion of wild-type substrate was already almost complete by the time the samples were taken (**Fig. 14d**).

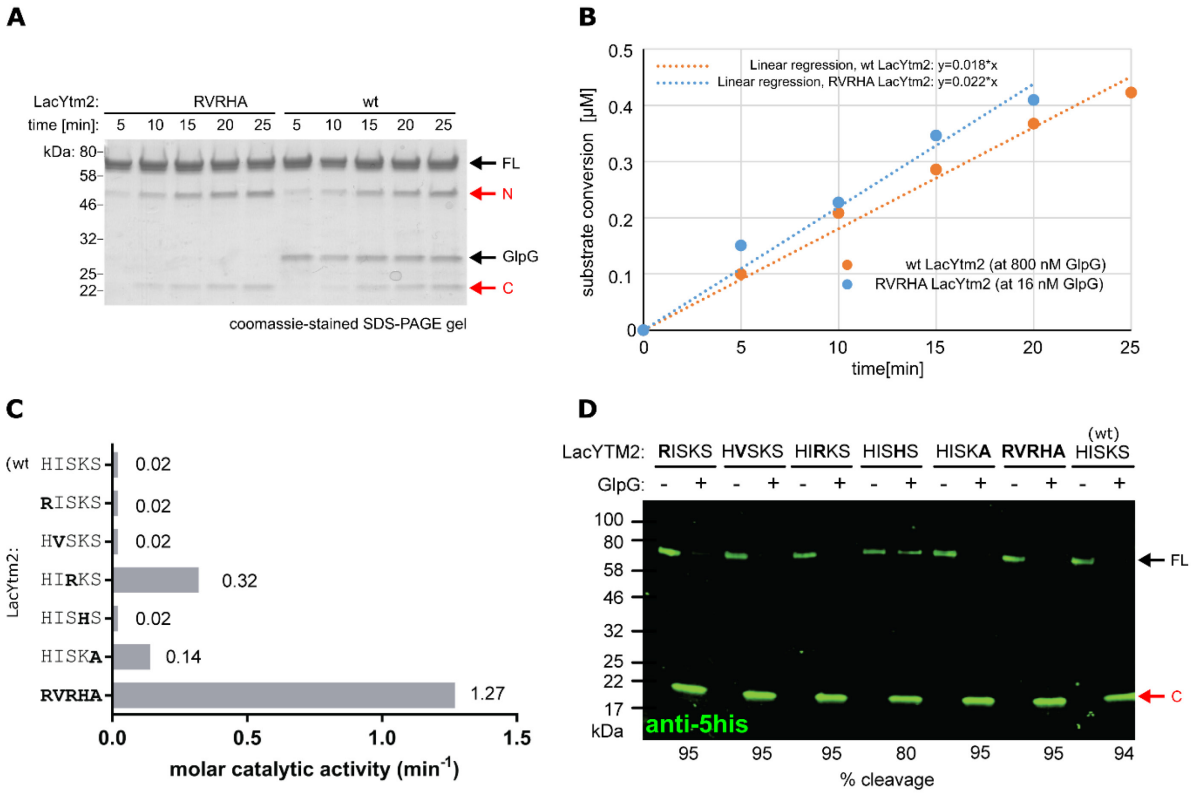


Fig. 14: Mutations in the non-prime side region of the LacYtm2 substrate improve the kinetic parameters of wild-type GlpG. (A) Representative *in vitro* gel-based assay showing the time course of the cleavage by active GlpG of the RVRHA and wild-type variants of the MBP-LacYtm2-Trx fusion protein. **(B)** The cleavage reactions, shown in panel A, were quantified by densitometry and values for RVRHA and wild-type substrate conversions were plotted for each time point. The slope of the linear regression curve served to calculate the reaction initial rate. **(C)** Molar catalytic activity of GlpG variants, defined as the number of substrate molecules cleaved by one molecule of enzyme per one minute. **(D)** Cleavage profiles of individual LacYtm2 mutants *in vivo*. Substrates were overexpressed as MBP-LacYtm2-Trx fusion proteins in wild-type *E. coli* MC4100, expressing endogenous GlpG (+) or in its *glpG::tet* mutant (-) and analysed by immunoblotting using fluorescence detection by anti-5his, which enabled cleavage to be quantified (% of cleavage, values are indicated below the blot). Full length substrate (FL) is indicated with a black arrow, the N-terminal (N) and C-terminal (C) cleavage products are indicated with red arrows. Representative results from one of three independent experiments are depicted.

4.1.4 Potent peptidyl ketoamide rhomboid inhibitor effectively blocks endogenous YqgP.

The RVRHA pentapeptide, which represents a sequence preferred by GlpG in positions P5 to P1 of the substrate, was added to the N-terminal side of the ketoamide warhead, C-terminally substituted with the phenylbutyl moiety (**Fig. 15a**). The resulting compound **11** inhibited cleavage of a model substrate by the YqgP rhomboid. *Bacillus subtilis* cells, expressing endogenous YqgP and overexpressing the model LacYtm2 substrate as fusion protein MBP-3xFLAG-LacYtm2-Trx-HA, were incubated with increasing concentrations of ketoamide inhibitor compound **11** (from 0.6 nM to 50 nM, **Fig. 15b**). The apparent IC_{50} values were estimated from the quantification of *in vivo* steady-state conversion products and endogenous YqgP was inhibited by 50% when approximately 5-10 nM inhibitor was used.

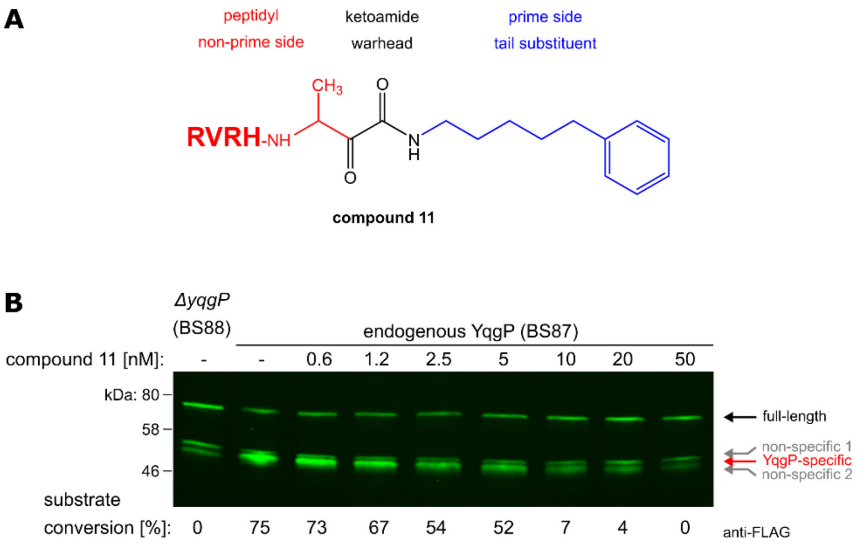


Fig. 15: Endogenously expressed YqgP is inhibited by ketoamide at low nanomolar concentrations. (A) Schematic representation of compound 11. The RVRHA pentapeptide (red) and the phenylbutyl moiety (blue) dramatically improve the inhibitory potency of the ketoamide inhibitor (black). **(B)** Steady-state conversion of the MBP-3xFLAG-LacYtm2-Trx-HA fusion protein by endogenous YqgP in living *Bacillus subtilis* cells BS87. The cells expressing the substrate were incubated with increasing concentrations of compound 11. The apparent IC_{50} was estimated to be 5-10 nM. The substrate was visualised by fluorescence immunoblotting, detecting the 3xFLAG tag N-terminal to the rhomboid cleavage site. Substrate conversion for individual inhibitor concentration points was quantified and is shown below the blot. Representative data from one of three independent experiments are shown.

4.1.5 My contributions to the co-authored publications

In **Publication_1**, I contributed:

- Cloning and *in vitro* production of the (^{35}S-Met) radiolabeled TatA substrates mutated in residues T4, A6 and A8
- Overexpression and isolation of recombinant wild-type GlpG, the GlpG.S201A catalytic mutant, and the GlpG.F146I and GlpG.F146A subsite S4 mutants
- Design, cloning and expression of recombinant MBP-3xFLAG-TatA-Trx-6xhis-S fusion proteins with T4W, I5G, I5S, A6D, A6W, A8G, A8V, T4R, I5V, A6R or T7H mutations in TatA sequence
- Expression and isolation of recombinant MBP-3xFLAG-TatA-Trx-6xhis-S fusion protein with wild-type TatA sequence
- Gel-based *in vitro* rhomboid activity assays with radiolabeled TatA T4, A6 and A8 substrate mutants and densitometric quantification and evaluation of substrate conversions
- Gel-based *in vitro* rhomboid activity assays with purified wild-type, the S201A, F146I and F146A GlpG variants and radiolabelled TatA I5 mutants and densitometric quantification and evaluation of substrate conversions
- Gel-based *in vivo* rhomboid activity assays with recombinant TatA fusion protein variants and evaluation of their steady-state conversions

In **Publication_2**, I contributed:

- Expression and isolation of recombinant wt, RISKS, HVSKS, HIRKS, HISHS, HISKA and RVRHA LacYtm2 variants
- Gel-based *in vitro* assays with LacYtm2 purified substrates, determination of reaction initial rates and calculation of molar catalytic activities
- Gel-based *in vivo* rhomboid assay with LacYtm2 variants expressed in wt *Escherichia coli* MC4100 and its *glpG::tet* mutant and evaluation of their steady-state conversions

In **Publication_3**, I contributed:

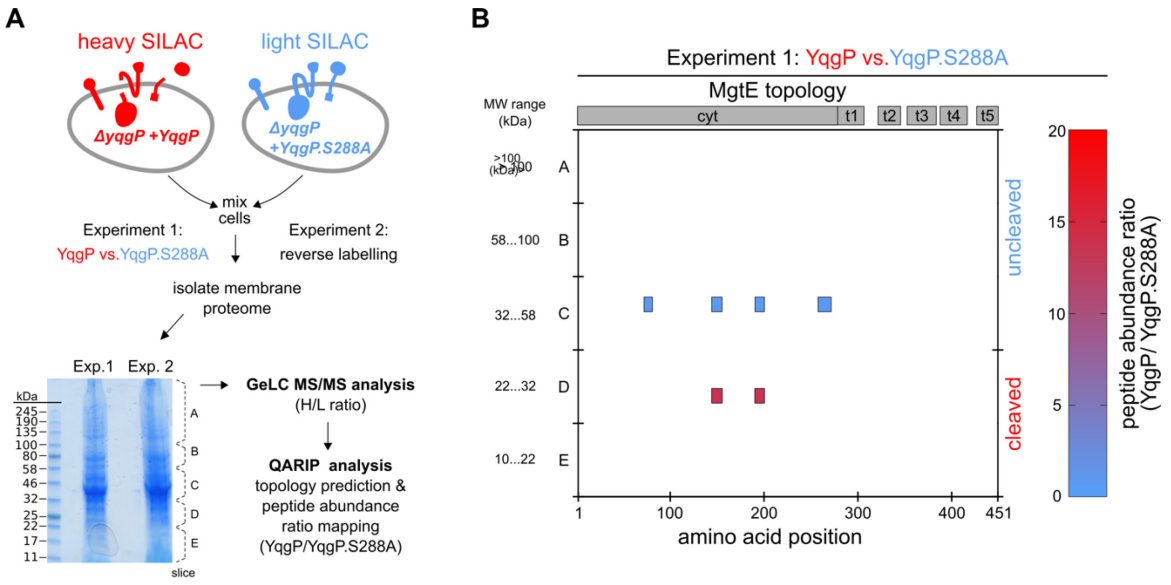
- Design of the model substrate $AmyE_{sp}$-MBP-3xFLAG-LacYtm2-Trx-HA fusion protein, suitable for expression in *Bacillus subtilis*
- Gel-based *in vivo* YqgP activity assay with this substrate and analysis of the inhibitory potency of compound **11** against endogenous YqgP and determination of its apparent IC_{50}

4.2 Results II: Unravelling the biological role of rhomboid protease YqgP in *Bacillus subtilis*

In **Publication_4**, our main goal was to understand the biology of YqgP rhomboid in *Bacillus subtilis*. According to our unpublished genomic analysis, YqgP-like seven-transmembrane-helices-core-containing rhomboid proteases are among the most frequently occuring rhomboid protease type in bacteria. Moreover, *Bacillus subtilis* is the main model Gram-positive bacterium used to study competence, cell division, biofilm formation and sporulation. The genomic, transcriptomic and proteomic analyses provide us with valuable tools to study YqgP-related biology. We searched the *B. subtilis* proteome for putative substrates and interactors of the YqgP rhomboid by proteomics methods, and identified MgtE, the major magnesium transporter of this bacterium, as the natural substrate. We showed that MgtE cleavage is upregulated during manganese stress and investigated this phenomenon in molecular detail. Structural and biophysical analyses revealed that YqgP senses elevated manganese concetrations via its N-terminal domain ($YqgP_{NTD}$), which also specifically recognizes MgtE. Moreover, we showed that YqgP cooperates with membrane protein quality control protease FtsH to control MgtE proteostasis.

4.2.1 SILAC-based quantitative proteomic screen reveals YqgP substrates.

We compared the membrane proteome of *Bacillus subtilis* expressing wild-type active YqgP with that of the strain expressing the YqgP variant bearing the inactivating mutation S288A. Exponentially growing "WT" or "S288A" cells, cultured in medium supplemented with heavy (H) or light (L) lysine isotope, respectively, were mixed in equal amounts based on OD_{600} read-out. The H+L cell suspension was then harvested, sub-fractionated to isolate the membrane protein fraction and analyzed in a GeLC-MS/MS experiment (**Fig. 16a, b**). To be considered a putative substrate of YqgP, a membrane protein had to fulfil the following parameters: i) it met the criteria for reliable MS identification, ii) it was present in both labelling experiments (Exp. 1: H to L, Exp. 2: L to H), iii) it was present in that gel slice, the molecular weight range of which corresponded with the probable size of the cleavage product, and iv) identified peptides had a high peptide abundance ratio (YqgP/YqgP 288A). The MgtE magnesium transporter was identified as the highest-scoring putative substrate of YqgP (**Fig. 16c**).



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

Fig. 16: SILAC-based proteomic analysis of YqgP degradome in *Bacillus subtilis*: (A) *Bacillus subtilis* cells expressing heterologous wild-type YqgP or its catalytic mutant YqgP.S288A were cultured in SILAC minimal medium, supplemented with either heavy (H) or light (L) lysine isotope, respectively. Exponentially growing cultures were mixed in OD_{600} ratio 1:1, H+L membrane proteins were isolated and analysed in a GeLC-MS/MS experiment. **(B)** Representative three-dimensional peptidogram for the MgtE substrate. MS-identified peptides were plotted on the peptidogram, considering i) peptide position in the gel, ii) peptide position in the full-length sequence and iii) peptide abundance ratio (YqgP/YqgP S288A) having H/L and L/H ratios included and expressed as the overall abundance of YqgP-related peptides to YqgP S288A-related peptides. **(C)** Table with the highest scoring YqgP substrate candidates, sorted by the total protein abudance ratio.

MgtE was also identified in a concurrent pull-down screen for YqgP interactors, which was performed in the collaborating laboratory of Dr. Thierry Doan at CNRS (Marseille, France), and the results of which are outlined in **Publication_4**. Based on these consistent data, we decided to explore the role of YqgP in MgtE physiology and regulation in more detail.

4.2.2 YqgP cleaves magnesium transporter MgtE.

To be able to detect MgtE cleavage, we first generated a rabbit polyclonal anti-MgtE antibody, which recognises the N-terminal cytosolic domain of MgtE (MgtE2-275). In a gel-based assay, we then analysed cell lysates from wild-type, *yqgP* knock-out and YqgP-overexpressing cultures and showed that YqgP specifically cleaves MgtE, generating N-terminal cleavage products that migrate similarly to the 32

kDa MW standard (**Fig. 17a**). Also, the rhomboid ketoamide inhibitor STS736 (compound **11**, for details see section 5.1.4) completely aborted MgtE processing by YqgP (**Fig. 17b**). We then mapped the putative YqgP cleavage region in MgtE. *In vitro* translated MgtE reference fragments, corresponding to the first 300, 315 and 330 residues of MgtE, or to full-length MgtE (FL) were analysed by immunoblot in parallel with the cell lysates containing MgtE cleaved in the presence of active YqgP (*ΔyqgP +yqgP*). The rhomboid-specific cleavage product migrated to a similar extent as the MgtE1-315 and MgtE1-330 fragments (**Fig. 17c**), suggesting that MgtE is cleaved in a topologically relevant region of the loop between its transmembrane helices TM1 and TM2 and that the cleavage probably leads to MgtE inactivation (**Fig. 17d**).

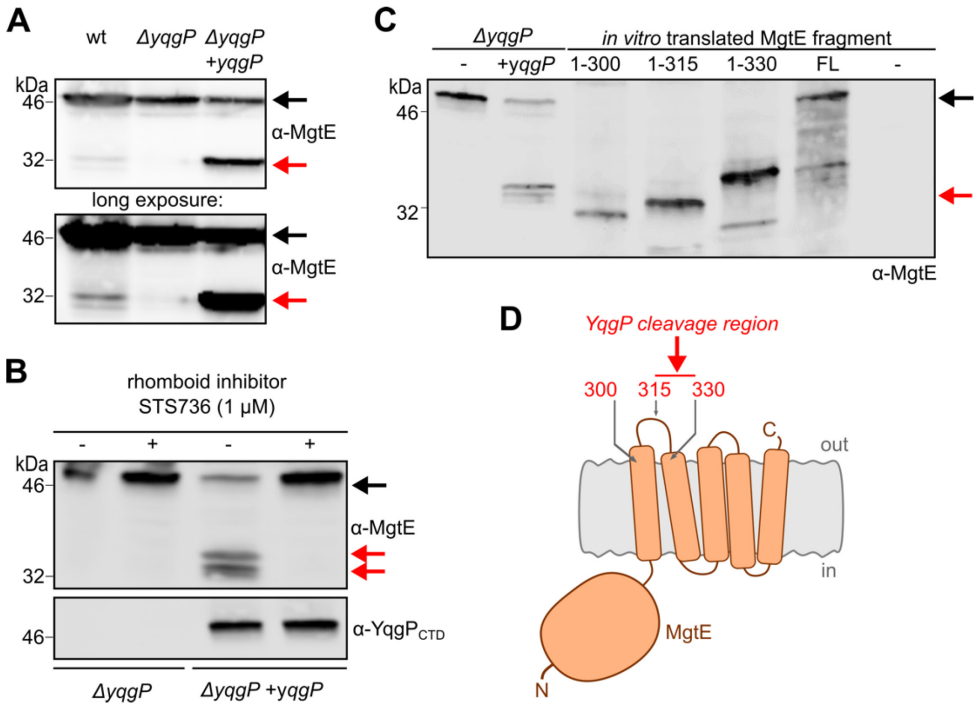


Fig. 17: YqgP cleaves magnesium transporter MgtE. (A) Detection of YqgP-specific cleavage of endogenous MgtE in wild-type (BTM2, **Table 3**) and YqgP-overexpressing *Bacillus subtilis* (BTM501, **Table 3**), using specific anti-MgtE2-275 rabbit polyclonal antibody. The *ΔyqgP* strain (BTM78, **Table 3**) was used as control. Black arrows indicate full-length MgtE and red arrows the MgtE N-terminal cleavage product. **(B)** YqgP-specific cleavage of MgtE was completely inhibited by STS736 (compound 11), a N-pentapeptidyl-C-phenylbutyl-ketoamide compound developed in our laboratory. YqgP was visualised using an anti-YqgP$_{CTD}$ antibody, recognising the YqgP 384-507 epitope. **(C)** The position of YqgP-induced cleavage site in MgtE was mapped by immunoblotting, using a set of *in vitro* translated MgtE protein fragments corresponding to MgtE1-300, 1-315, 1-330 and full-length (FL) MgtE. **(D)** Based on the migration pattern, the cleavage site was estimated to be in the loop region located between MgtE transmembrane helices 1 and 2.

These results indicated that YqgP may regulate MgtE function. This could be true especially when the intracellular magnesium concentration levels are low and the MgtE channel opens in order to import the missing Mg^{2+} to meet the cell requirements. It was previously shown that MgtE transporting activity can by corrupted or inhibited by elevated divalent transition metal cations during magnesium

depletion, which in turn might be toxic for the cell (as described in section 1.6.2.2, p. 29). Therefore, we investigated the role of YqgP in MgtE degradation, under conditions of increased Mn^{2+}, Zn^{2+}, Co^{2+} and Ni^{2+} divalent transition metal ion concentrations in the culture medium.

4.2.3 MgtE cleavage is enhanced during manganese stress.

First, we asked to what extent divalent manganese affects MgtE processing by YqgP. *Bacillus subtilis* cells expressing heterologous wild-type YqgP from the weak P_{xyl} promoter were cultured in the presence of either low (0.01 mM) or high (1 mM) $MgSO_4$ concentration and supplemented with either low (1 µM) or high (100 µM) $MnCl_2$. Immunoblot analysis of the cell lysates and steady-state conversion quantification revealed that MgtE processing is increased 4-fold, when cells were exposed to high manganese concentrations, but this activation occured only under low magnesium conditions (**Fig. 18**).

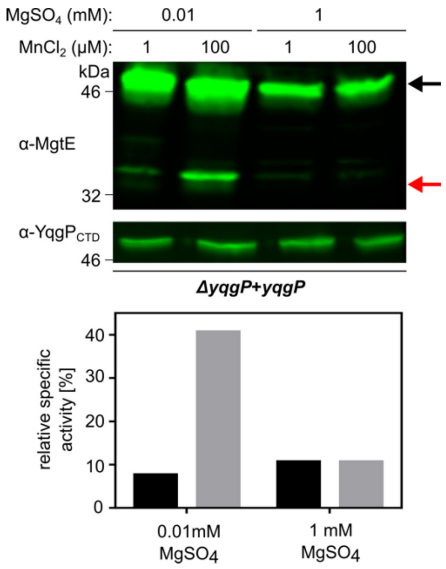


Fig. 18: Elevated manganese levels activate MgtE degradation, during magnesium starvation. Fluorescent immunoblot detection (upper panel) and quantification of cleavage events (lower panel) in cells expressing ectopic YqgP. MgtE processing is specifically increased upon presence of high (100 µM) $MnCl_2$, but only when *B. subtilis* suffer from limited magnesium conditions. In this situation, the MgtE metal ion conducting pore is permanently open and may transport Mn^{2+} cations that out-titrate Mg^{2+} cations. MgtE levels were visualised by anti-MgtE2-275 and YqgP levels by anti-$YqgP_{CTD}$ antibodies. Specific activities were calculated as described in the methods section 3.11.3 (p. 55). Full-length MgtE and MgtE N-terminal cleavage product are indicated with black and red arrows, respectively. The representative results from one of three experiments are displayed.

Secondly, we asked whether YqgP also contributes to overall cell fitness during increased manganese stress. The growth of wild-type *Bacillus subtilis* (wt), its rhomboid knock-out mutant (*ΔyqgP*) and the rescue mutant, ectopically expressing wild-type YqgP (*ΔyqgP +yqgP*), in response to manganese shock induced by addition of 75 µM Mn^{2+} salt during mid-exponential phase, was examined. Cells lacking YqgP recovered significantly more slowly than the wild-type or rescue strains. Moreover, the growth

curve of the rescue mutant remained mostly unaffected and more or less resembled the growth of the untreated cells, indicating that YqgP activity can specifically rescue the adverse growth effects caused by manganese toxicity (as shown in **Publication_4**).

4.2.4 YqgP extramembrane domains have opposing effects on YqgP activity.

The extramembrane domains of YqgP represented putative regulatory elements of YqgP function. To assess the contributions of either domain to rhomboid activity, we generated truncated variants YqgPΔNTD and YqgPΔCTD, lacking the first 177 N-terminal or the last 123 of C-terminal residues, respectively (**Fig. 19a**), and analysed their activities against the natural substrate MgtE and the model *Providencia stuartii* TatA-derived substrate. Quantification of steady-state substrate conversions by full-length and both truncated variants in living *Bacillus subtilis* showed that the N-terminal domain of YqgP plays a role in substrate recognition, since its deletion led to a 3-fold decrease in MgtE processing, while cleavage of the TatA substrate was mostly unaffected (**Fig. 19b**).

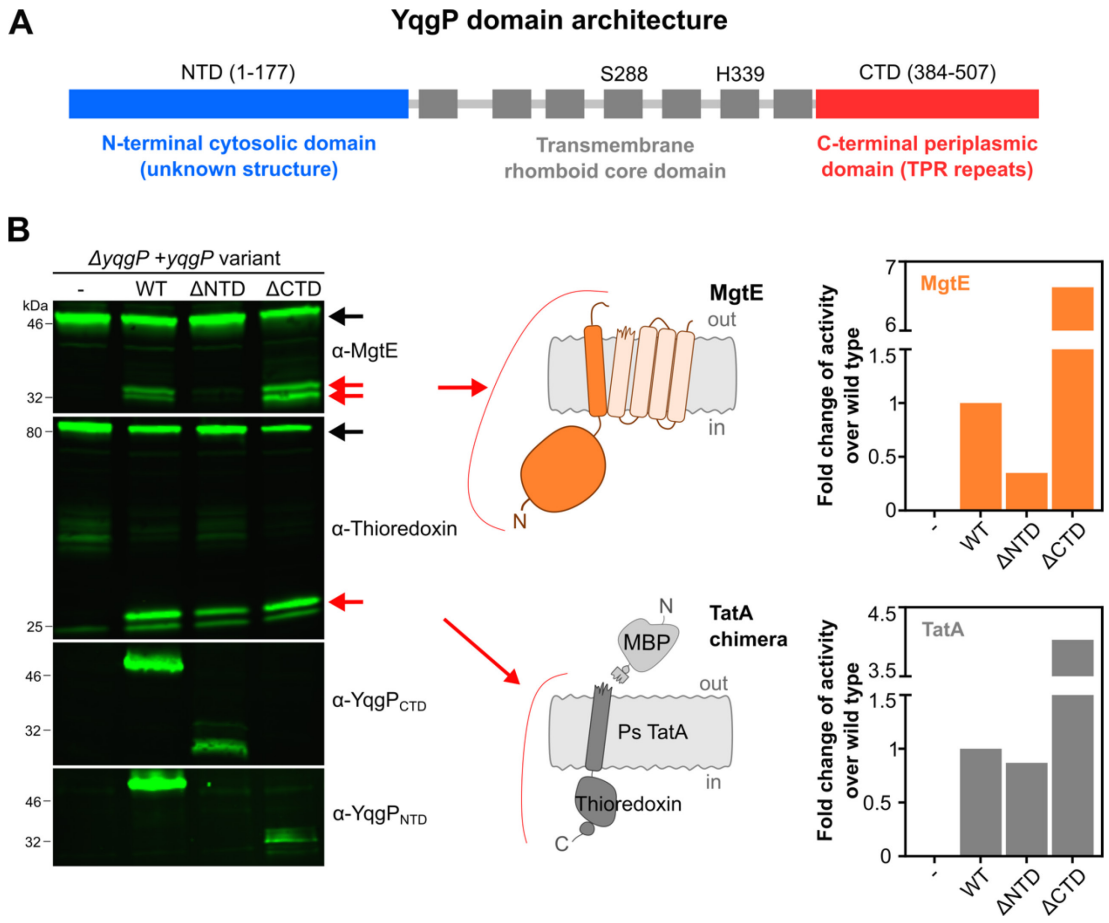


Fig. 19: Extramembrane domains of YqgP play opposite roles in MgtE processing. (A) Schematic representation of YqgP domain architecture. N-terminal and C-terminal extramembrane domains are in blue and red, respectively, and transmembrane core in grey. **(B)** Analysis of steady-state conversions of

MgtE and TatA substrates in living *Bacillus subtilis,* expressing full-length YqgP or its truncated variants lacking either the N-terminal (YqgPΔNTD) or the C-terminal soluble domains (YqgPΔCTD), using fluorescent immunoblotting. The N-terminal domain is necessary for proper rhomboid activity against MgtE, while the C-terminal domain is inhibitory for both the native and model substrates. Full-length substrates are indicated by black arrows. Cleavage products are indicated by red arrows, as highlighted also in their schematic illustrations in the midle part of the panel. MgtE was detected by the anti-MgtE2-275 antibody, the thioredoxin tag in the C-terminal part of the overexpressed fusion construct of the model TatA substrate by the anti-Trx antibody and ectopically overexpressed YqgP variants either by the anti-$YqgP_{NTD}$ or the anti-$YqgP_{CTD}$ antibodies. Representative immunoblots (left-hand panel) and the signal quantification (right-hand panel) are shown. MBP is maltose binding protein from *Escherichia coli.*

Intriguingly, the steady-state activity of the YqgPΔCTD variant against MgtE was almost 4-fold higher and against TatA more than 6-fold higher than the activity of full-length rhomboid. This suggested that the C-terminal domain may represent a general autoinhibitory subunit naturally regulating YqgP activity.

4.2.5 YqgP N-terminal domain senses increased manganese during MgtE cleavage

Since the effect of YqgP N-terminal domain deletion was specific for MgtE cleavage, we tested whether the domain can also be involved in the regulation of manganese-dependent MgtE processing. *Bacillus subtilis* cells expressing either full-length (FL), ΔNTD or ΔCTD YqgP variants were titrated with increasing amounts of divalent Mn^{2+} and the steady-state conversion of substrate to product was quantified (**Fig. 20**). The activity of the C-terminal truncation mutant, similarly to full-length YqgP, was enhanced by addition of 100 µM $MnCl_2$. In contrast, the N-terminal mutant was not capable of manganese sensing and its activity remained the same at all manganese concentrations tested.

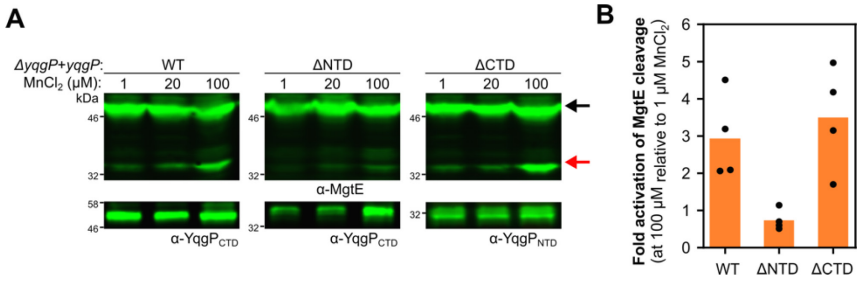


Fig. 20 : The N-terminal domain of YqgP senses manganese levels at toxic concentrations. (A) Fluorescent immunoblot analysis of *Bacillus subtilis* strains ectopically expressing full-length YqgP or its truncated variants lacking either the N-terminal (YqgPΔNTD) or the C-terminal soluble domains (YqgPΔCTD). Cells were incubated with 1, 20 or 100 µM concentrations of $MnCl_2$ and MgtE cleavage in cell lysates was detected using the anti-MgtE2-275 antibody. YqgP variants were detected by $YqgP_{NTD}$ or $YqgP_{CTD}$ antibodies. **(B)** Quantification of steady state MgtE conversion in presence of low and high manganese by YqgP truncated variants from four independent experiments, one of which is shown in panel A. MgtE conversion was quantified from the fluorescent immunoblot as described in section 3.9.6. The cleavage is displayed as the fold activation in 100 µM $MnCl_2$ relative to 1 µM $MnCl_2$, individually for each YqgP variant.

4.2.6 Structural characterisation of the YqgP extramembrane domains

To delineate the mechanistic principles driving the regulation of YqgP activity, we solved the three-dimensional structures of the individual YqgP soluble domains, the N-terminal cytosolic domain $YqgP_{NTD}$ and the C-terminal periplasmic domain $YqgP_{CTD}$, using solution NMR (**Fig. 21**). $^{15}N/^{13}C$ uniformly labelled $YqgP_{NTD}$ and $YqgP_{CTD}$ were expressed as C-terminally 6x-his tagged and N-terminally 6x-his tagged constructs, respectively. The high stabilities of both proteins at milimolar concentrations, using 1 mM $YqgP_{NTD}$ and 0.7 mM $YqgP_{CTD}$ protein solutions, enabled us to measure 2D and 3D NMR spectra with almost all ^{15}N, ^{13}C and 1H resonances assigned and to calculate high-quality structural models (**Fig. 21a**). We obtained clusters of 30 converged structures for each of YqgP domains (**Fig. 21b, d**), which were used to calculate the mean structures. The NMR constraints and structural statistics for both domains' clusters are summarised in **Table 4**.

The overall $YqgP_{NTD}$ structure comprises of five α-helices at positions 3-16 (α1), 54-75 (α2), 96-100 (α3), 121-128 (α4), 146-168 (α5) and six β-strand motives at positions 21-24 (β1), 30-34 (β2), 43-48 (β3), 80-88 (β4), 104-106 (β5), 109-117 (β6) and forms a mixed α+β bundle, in which central β -sheet is twisted and surrounded by helices 1-4 (**Fig. 21a, b**). It also contains several negative side chain-clustering regions on its surface suggesting putative ion binding motifs (**Fig. 21c**). Moreover, a structural similarity search using the Dali server identified functionally unrelated proteins capable of metal cation coordination adopting fold characteristic for type II restriction endonuclease, which binds the Mg^{2+} cofactor in its active site (InterPro and Pfam codes: IPR018573 and PF09491 respectively).

The overall $YqgP_{CTD}$ structure is formed of seven α-helices, almost perpendicular to each other, at positions 7-22 (α1), 26-37 (α2), 44-57 (α3), 60-73 (α4), 78-90 (α5), 94-107 (α6) and 112-125 (α7). As predicted, it contains three tetratricopeptide repeats (TPR), wich are well characterised protein-, peptide- and glycan-interacting motifs [200]. Interestingly, while the negatively-charged residues are clustered into a continuous region on the surface facing the N-terminus, the exposed region on the opposite C-terminal-facing side of the molecule is mostly positively-charged (**Fig. 21e**).

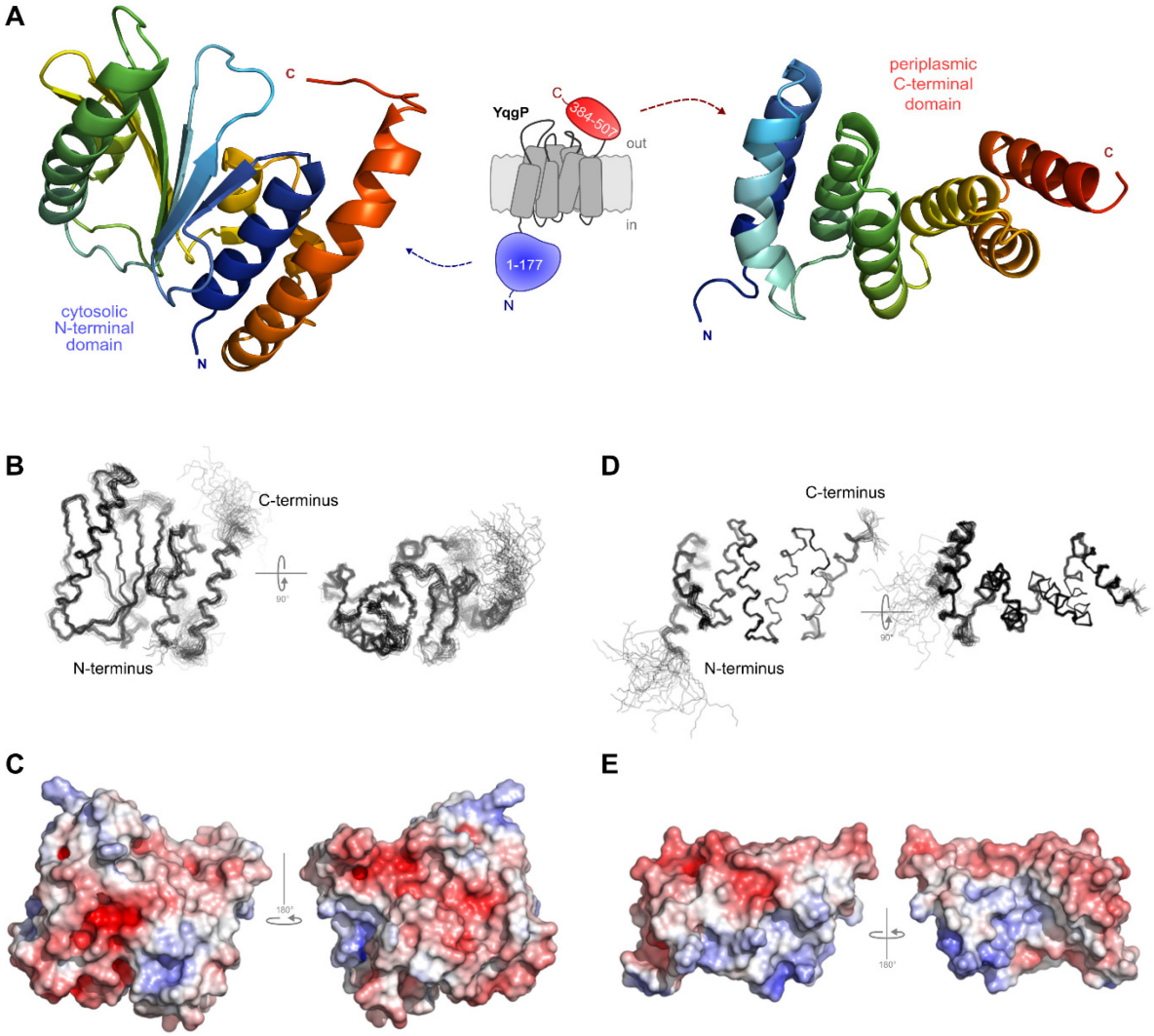


Fig. 21: Structural characterisation of YqgP soluble domains. (A) The N-terminal cytosolic domain (PDB entry: 6R0J) and the C-terminal periplasmic domain (PDB entry: 6R0O) structures determined by solution NMR. The overall structures were calculated from 30 converged structures of **(B)** $YqgP_{NTD}$ and **(D)** $YqgP_{CTD}$. The protein surface charges were calculated for **(C)** $YqgP_{NTD}$ and **(E)** $YqgP_{CTD}$. The PyMOL™ 2.2.3 software was used for visualization of all structures.

Table 4: NMR constraints and statistics for the YqgP soluble domain structures.

YqgP soluble domain:	NTD	CTD
Non-redundant distance and angle constrains		
Total number of NOE restraints	2517	2950
Short-range NOEs		
Intra-residue (i = j)	290	700
Sequential (| i - j | = 1)	690	676
Medium-range NOEs (1 < | i - j | < 5)	542	789
Long-range NOEs (| i - j | ≥ 5)	995	785
Torsion angles	300	216
Hydrogen bond constraints	-	-
Total number of restricting constraints	2817	3166
Total number of restricting constraints per restrained residue	5.4	25.5
Residual constraint violations		
Distance violations per structure		
0.1 – 0.2 Å	3.63	4.77
0.2 – 0.5 Å	1.63	0.97
> 0.5 Å	0	0
r.m.s. of distance violation per constraint	0.01 Å	0.01 Å
Maximum distance violation	0.49 Å	0.50 Å
Dihedral angle viol. per structure		
1 – 10 °	2.93	1.8
> 10 °	0	0
r.m.s. of dihedral violations per constraint	0.33°	0.35°
Maximum dihedral angle viol.	5.0°	5.0°
Ramachandran plot summary		
Most favoured regions	94.6 %	94.4 %
Additionally allowed regions	4.9 %	5.6 %
Generously allowed regions	0.4 %	0.0 %
Disallowed regions	0.0 %	0.0 %
r.m.s.d. to the mean structure	*all/ordered*[1]	*all/ordered*[1]
All backbone atoms	3.0/1.0 Å	1.5/0.3 Å
All heavy atoms	3.5/1.5 Å	1.5/0.8 Å
PDB entry	**6R0J**	**6R0O**
BMRB accession code	**34376**	**34377**

[1]Residues with sum of phi and psi order parameters > 1.8

4.2.7 Activation of MgtE cleavage by YqgP is induced by transition metal cations.

The fact that MgtE magnesium-transporting activity can be either hijacked or inhibited by divalent transition metal cations such as manganese, zinc, cobalt or nickel (as discussed in section 1.6.2.2, p. 29), especially if these are more abundant than magnesium itself, led us to test their possible activatory effect on YqgP. *Bacillus subtilis* cells overexpressing YqgP were titrated individually with increasing concentrations of Zn^{2+}, Co^{2+}, Ni^{2+} and Ca^{2+} and Mn^{2+} salts and MgtE conversion was quantified (**Fig. 22**). All titrations were performed in defined medium containing low levels (10 µM) of $MgSO_4$. Besides activation by manganese, only the addition of 100 µM $ZnCl_2$ or 100 µM $CoCl_2$ improved MgtE processing 6-fold or 2-fold, respectively, while Ni^{2+} and Ca^{2+} had no significant effect.

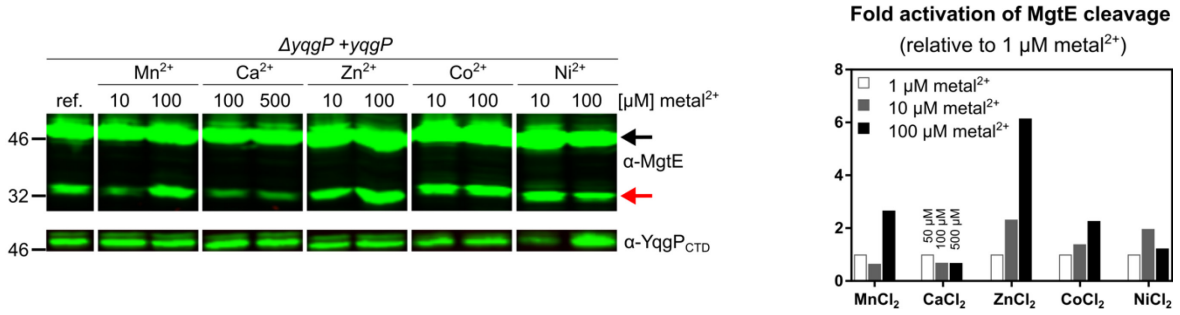


Fig. 22: Transition metal ions effect on MgtE processing by YqgP. Fluorescent immunoblot detection (left-hand panel) and steady-state conversion analysis (right-hand panel) of MgtE cleavage in the presence of diverse divalent metal cations. High concentrations of manganese, zinc and cobalt divalent salts support MgtE processing by YqgP, while increased nickel concentration has no effect. Calcium was used as control. Blots were visualised by the anti-MgtE2-275 and anti-$YqgP_{CTD}$ antibodies and fluorescent secondary antibodies. The black arrow indicates full-length MgtE and the red arrow indicates the MgtE N-terminal cleavage product.

4.2.8 Biophysical analyses reveal YqgP NTD binding to transition metal ions.

The results described in the previous sections made us wonder whether $YqgP_{NTD}$ binds divalent cations. We performed ITC and NMR titration analyses with purified $YqgP_{NTD}$ and a panel of divalent metal cations salts. ITC revealed that the domain specifically bound Mn^{2+}, Co^{2+} and Ni^{2+} with submillimolar to milimolar affinity with dissociation constants K_d around 1.2 mM, 0.6 mM and 0.1 mM, respectively, while no binding was observed for Mg^{2+}, Ca^{2+} and no dissociation constant could be measured for Zn^{2+} (**Fig. 23**).

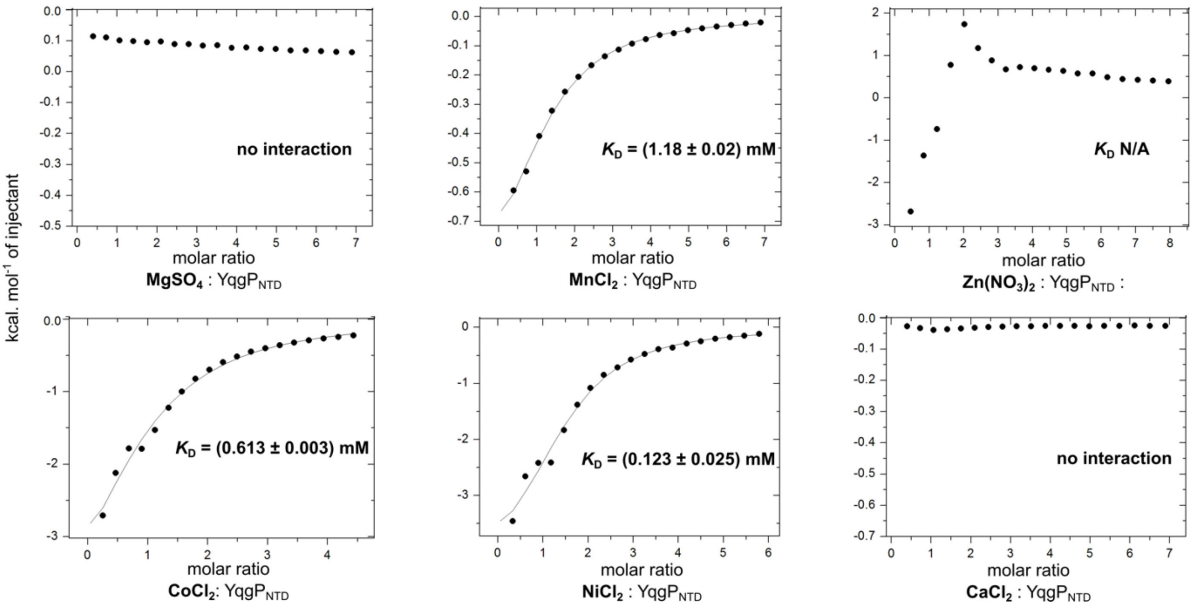


Fig. 23: Isothermal titration analysis showed that recombinant $YqgP_{NTD}$ binds divalent cations. Purified YqgP N-terminal domain was titrated with stepwise injections of divalent metal stock solutions until saturation by ligand was reached (for details see section 3.10.2).

We next focused on the mapping of the manganese binding region of $YqgP_{NTD}$ using solution NMR titration experiments. Since the Mn^{2+} cation is paramagnetic, and thus active in the magnetic field applied during NMR measurements, we inspected the 2D $^{1}H/^{15}N$ NMR spectra of 400 µM $YqgP_{NTD}$, titrated stepwise with 10, 20 and 40 µM $MnCl_2$, not only for perurbations in chemical shifts (CSP), but also for changes in intensities of peak signals (I/I_0), caused by Mn^{2+} paramagnetic effects. These changes jointly corresponded to residues directly or indirectly involved in manganese interaction (**Fig. 24a, b**). The domain regions most affected included residues 27-30, 35-37, 48-63 and 89-98. Of these, residues putatively chelating divalent cations, such as aspartate, glutamate or histidine, were chosen for further *in vivo* validation. Single-point full-length YqgP variants mutated to alanine in positions D29, D37, H49, D50, D52, D60 and E90, which are located on the solvent-accessible surface of $YqgP_{NTD}$ (**Fig. 24e**), were tested *in vivo* for their ability to enhance MgtE cleavage (**Fig. 24c**). Steady-state conversion determination revealed that mutation D50A completely abrogates the Mn-activation effect, suggesting that the manganese-binding site is likely located in this region of $YqgP_{NTD}$ (**Fig. 24d, e**).

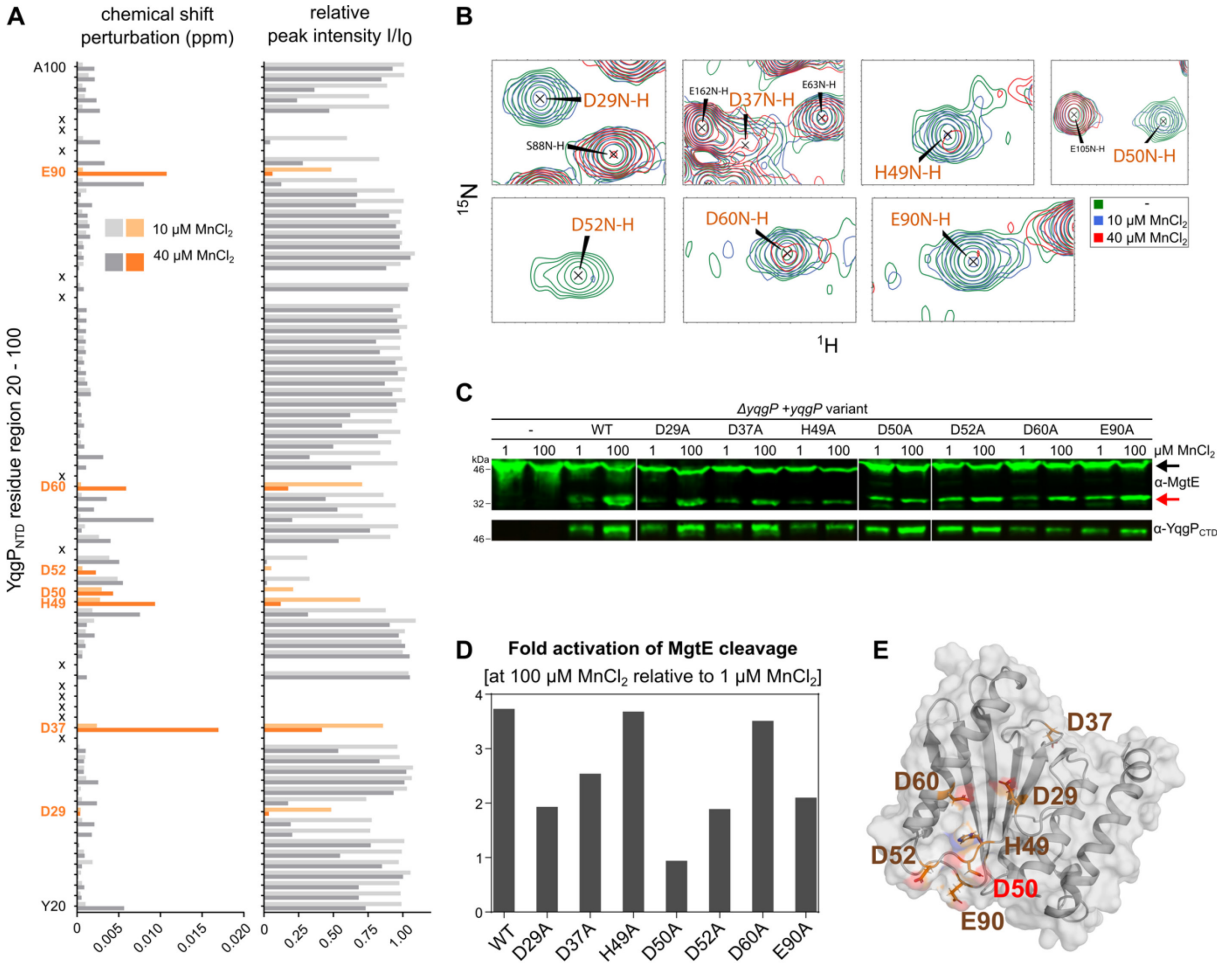


Fig. 24: Solution NMR titration revealed the manganese-binding site in YqgP$_{NTD}$. (A) Chemical shift perturbation (CSP, left-hand panel) and relative peak intensity (I/I_0, right-hand panel) plots were calculated from the 2D $^{1}H/^{15}N$ spectrum of 400 µM (^{15}N) YqgP$_{NTD}$ titrated with 10 and 40 µM $MnCl_2$. Residue regions, which were most affected by addition of manganese (columns in shades of orange) were shifted significantly and their intensities decreased at the same time. Residues marked as (x) were not involved in CSP calculations, due to the low quality of the respective resonance signals. **(B)** Detailed views of merged 2D $^{1}H/^{15}N$ spectra of YqgP$_{NTD}$ with buffer (green), or 10 µM (blue) or 40 µM $MnCl_2$ (red) showing residues, which were further verified using a **(C)** gel-based *in vivo* activity assay. **(D)** Steady-state conversions for alanine mutants of YqgP in these positions, quantified from (C), revealed that D50 is likely the key residue involved in manganese-binding. **(E)** Residues that were identified in the NMR titration experiment form a negatively-charged region concentrated on the protein surface.

4.2.9 YqgP and FtsH jointly regulate MgtE proteostasis.

In the experiment with the inactive mutant YqgP.S288A, we observed that MgtE was alternatively processed, generating an N-terminal cleavage product of higher molecular weight (**Fig. 25a**- lane 3, **Fig. 25b**- lane 2) than that produced by wild-type YqgP (**Fig. 25a**- lane 1, **Fig. 17**). The presence of this alternative cleavage product was clearly dependent on the presence of YqgP.S288A variant. We assumed that another protease must be involved in this cleavage. In our CoIP-based proteomic srcreen for YqgP interacting partners (data are not shown here, for reference, see **Publication_4**), we identified

the FtsH protein, an ATP-dependent zinc-metalloprotease, which is the major membrane protein quality control protease conserved in bacteria. We assumed that, with highest likelihood, FtsH could be the putative protease responsible for the alternative cleavage. In agreement with this idea, no alternative MgtE cleavage was induced in the presence of YqgP.S288A and absence of FtsH (**Fig. 25a**- lane 4), similarly to the pattern observed with the strain lacking YqgP at all (**Fig. 25b**- lane 1). To estimate the alternative cleavage site in MgtE, we compared the mobility of the respective product in *Bacillus subtilis* lysate with *in vitro* translated reference fragments corresponding to C-terminally truncated MgtE1-340, 1-355, 1-370 and full-length variants. The alternative product migrated to a level similar to the MgtE1-340 and 1-355 fragments, indicating that the cleavage occurs on the cytosolic side of MgtE in loop L2, between helices TM2 and TM3 (**Fig. 25b, c**).

We also asked whether either of the two YqgP soluble domains is involved in YqgP cooperation with FtsH. The truncated variants of YqgP.S288A (YqgP.S288A-ΔNTD and YqgP.S288A-ΔCTD) were tested for cleavage of endogenous MgtE in the presence or absence of endogenous FtsH. Immunoblot detection of MgtE cleavage showed that YqgP lacking its N-terminal domain was not capable of alternative FtsH-specific processing of MgtE. This suggested that the NTD is likely involved in MgtE presentation to FtsH (**Fig. 25d**).

To resolve the molecular aspects of YqgP/FtsH-mediated proteolysis in more detail, we performed translation shut-off chase experiments by adding tetracycline. We then detected the temporal stability (chase) of the cleavage products of MgtE in cells expressing wild-type or YqgP.S288A variants and possessing or lacking endogenous FtsH (**Fig. 25e**). Once the protein translation was terminated, the N-terminal cleavage product of MgtE generated by wild-type YqgP was unstable and was further shedded by FtsH, but also by other unknown protease(s) (**Fig. 25e**- lanes 13-16 and 37-40). Intriguingly, after translation shut-off, the alternative FtsH-specific N-terminal cleavage product remained stable over the chase period (**Fig. 25e**- lanes 21-24). Additionally, no cleavage was detected in the absence of YqgP, suggesting that YqgP is the primary protease (**Fig. 25e**- lanes 1-8).

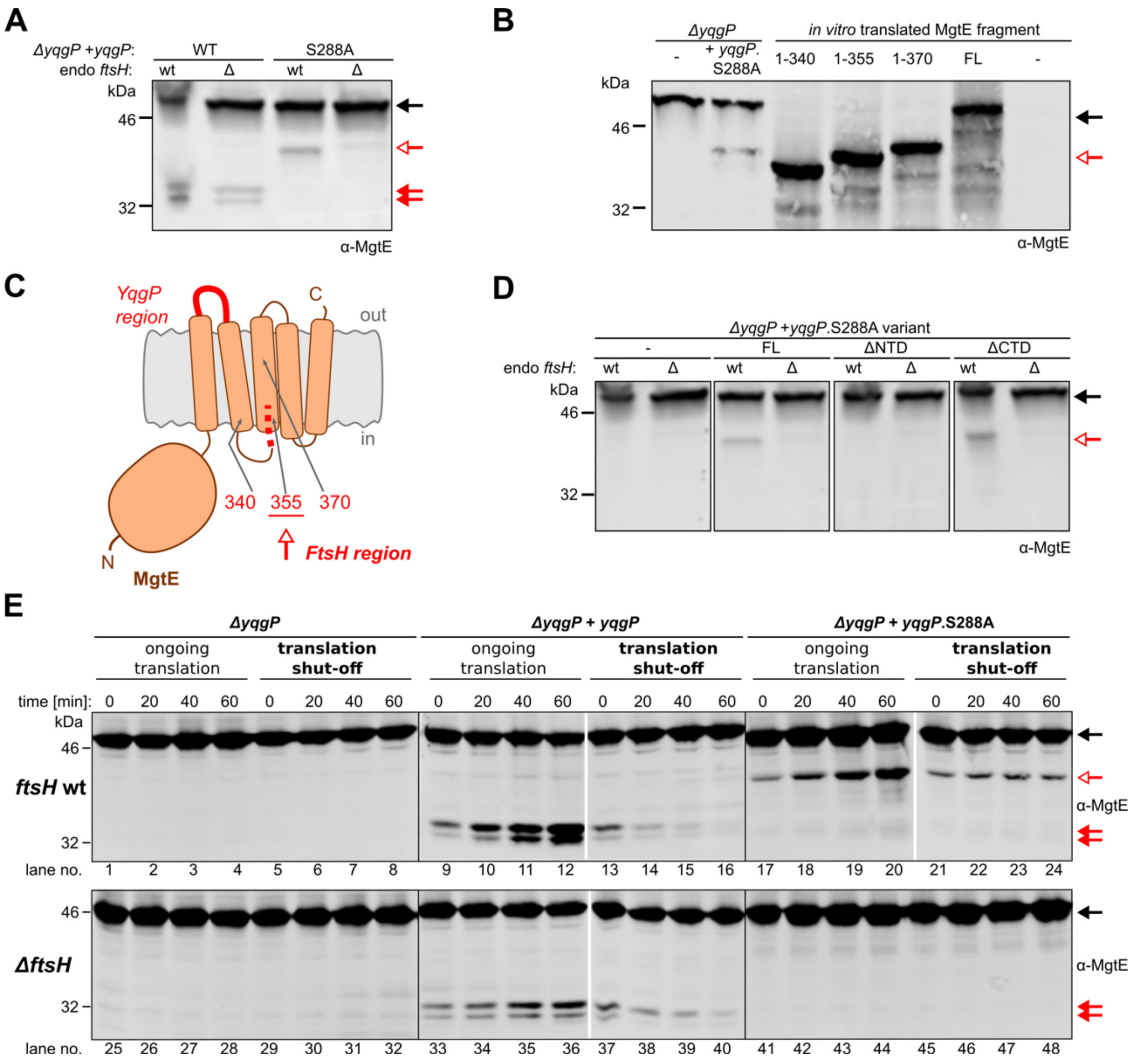


Fig. 25: YqgP is the primary protease, which cooperates with FtsH on MgtE processing, during translation. (A) Immunoblot detection of MgtE in cells expressing wild-type YqgP or inactive YqgP.S288A in the presence (lanes 1, 3) or absence of endogenous FtsH (lanes 2, 4) showing that deletion of *ftsH* led to the disappearance of the alternative cleavage product of MgtE (red arrow with empty head). Conversely, the wild-type YqgP-specific cleavage (red arrow) of MgtE was not affected by FtsH. **(B)** Mapping of the FtsH specific region shown on immunoblot. The first two lanes correspond to MgtE signal in *yqgP* knockout cells and YqgP.S288A-expressing cells and remaining lanes to signal *in vitro* generated MgtE fragments truncated from MgtE C-terminus. **(C)** Scheme of MgtE monomer indicating that FtsH region is likely located in between TM helices 2 and 3, at the cytosolic side. **(D)** The YqgP N-terminal domain is necessary for FtsH activity, since its truncation did not lead to alternative processing of endogenous MgtE by FtsH. **(E)** Temporal resolution of MgtE cleavage. The translation shut-off was performed by addition of 20 µg/ml tetracycline to exponentially growing cells ectopically expressing YqgP or YqgP.S288A mutant, in the presence or absence of endogenous FtsH. Samples for the immunoblot were taken at time points 0, 20, 40 and 60 min, after tetracycline addition. The wild-type YqgP-dependent MgtE cleavage product (red arrow) was not stable and was further degraded by FtsH and/or by other unknown protease(s). YqgP was active only during MgtE translation. Moreover, the alternative FtsH-dependent N-terminal cleavage product was stable during the chase period. In all immunoblots in this figure, endogenous MgtE was visualised using rabbit anti-MgtE2-275 (MgtE N-terminal cytosolic domain), overexpressed YqgP using the rabbit anti-YqgP$_{NTD}$ and anti-YqgP$_{CTD}$ and endogenous FtsH using the rabbit anti *Bacillus subtilis* FtsH antibodies.

5 DISCUSSION

The identification of the natural substrates and regulatory partners and the understanding of the mechanism of action are indispensable steps in deciphering the physiological role of an enzyme. Our knowledge on the biology and regulation of rhomboid proteases, an evolutionarily widespread and conserved superfamily of serine intramembrane proteases, as discussed in the introduction, is rather scarce, for several reasons. First, rhomboids and their substrates are membrane proteins, naturally of low abundance in the membrane and are generally quite difficult to isolate in amounts sufficient for structural and mechanistic characterisation, ii) proteolytic events within the lipid environment are naturally slow and the protein-protein interactions weak and transient, thus iii) the mechanisms that rhomboids utilise for substrate recognition and processing have been practically studied only for the GlpG rhomboid protease of *Escherichia coli*, and last but not least iv) general and sensitive methods for the identification of natural substrates and interactors and general enzymological tools have been missing.

First, I will discuss three studies dedicated to the mechanism of substrate recognition, all of which I have co-authored. To give a more comprehensive view on what we learnt, I will discuss not only the results that I have contributed, but also work that was performed by the co-authors.

In **Publication_1**, which I have reported on in sections 4.1.1 and 4.1.2, we provide mechanistic and structural details of the substrate-rhomboid interaction within the active site pockets of GlpG. Looking at the panel of individually mutated *Providencia stuartii* TatA proteins (in each of the non-prime side P1-P5 positions) that were used as GlpG substrates, the P1 and P4 positions are most sensitive to amino acid substitutions. GlpG strictly prefers small unbranched residues in the P1 position and large hydrophobic residues in the P4 position. These findings are consistent with previously published observations on the AarA rhomboid and on Rhomboid-1 of *Drosophila melanogaster*, which I have commented on in the Introduction (for comparison, see **Fig. 6c** and **Fig. 12a, b**). Thus, the results most likely imply conserved mechanistic roles of the S4 and S1 subsites in GlpG-related rhomboids.

The mechanism-based peptidyl chloromethylketone (CMK) inhibitors covalently bind both active site residues S201 and H254 of GlpG (substrates bind covalently only to the serine) (**Fig. 7**), and are thus less suitable models of the tetrahedral intermediate of the reaction. Nevertheless, they are good models for structural studies because they are irreversible. Addition of tetrapeptidyl substituent comprising the sequence naturally encoded in the TatA substrate to the CMK warhead, resulting in Ac-IATA-CMK, dramatically increases the inhibitory potency of the compound, and introduction of substitutions in the P1-P4 positions of the inhibitor's peptidic part has a similar effect as these substitutions have in a substrate. For example, when non-tolerated P4 site mutations in IATA, such as GATA, SATA, IATG, IATV and IDTA are introduced into the inhibitor, this reduces the inhibitory potency. This implies that the

peptidic part of the inhibitor is accomodated in the active site of GlpG in a substrate-like manner. The crystal structure of GlpG in complex with Ac-IATA-CMK thus reveals the interactions of the substrate-like tetrapeptide with rhomboid intramembrane protease. The P1 and P4 substrate residues are tightly bound in the respective subsites. It also shows that the S2 and S3 subsites of the enzyme are not well defined and allow accommodation of voluminous substrate side chains. The structure of the rhomboid-inhibitor complex also reveals significant movement of the L5 loop as a result of binding of the peptidyl part of the CMK inhibitor. However, structural evidence rationalising the binding of the substrate's prime side region (C-terminal to the scissile bond) to the rhomboid is still lacking. Therefore, we performed molecular dynamics experiment, which simulated the binding of substrate transmembrane residues covering the P1' to P3' positions. Contradictory to the hypothesis that the TM5 helix lateraly moves and opens an imaginary gate (reviewed in section 1.4.2, p. 11), we observed no such movement of the TM5 helix and the structural data made us conclude that large changes within the whole TM5 region are not required for substrate access.

Inspection of the S1 subsite in GlpG, occupied by the P1 residue of the inhibitor reveals a continuous deeper hydrophilic cavity, which contains three catalytic water molecules. This explains why only small residues such as alanine, or cysteine and serine naturally occur in the P1 position and why larger or negatively charged side chains are excluded- these would either occupy and disturb the water-retention site or be repulsed due to their electrostatic character, respectively (for details, see section 1.4.1, p. 9). We then show that the S4 subsite, which discriminates small and charged residues at P4 position and tolerates hydrophobic side chains of the substrate, is a solvent-exposed groove formed by the hydrophopic residues M120, M144 and F146 of the L1 loop. Most specifically, F146 interacts with the P4 isoleucine of substrate. Substitution of F146 residue for smaller amino acid, such as alanine or isoleucine abrogates the interaction with the substrate *in vitro*. Indeed, mutations in the substrate P4 position, compensating for the loss of the Van der Waals' interaction of the isoleucine side chain within the mutated S4 subsite, such as I5F, I5M and I5W, restore the cleavage of the substrate to wild type levels.

In the follow up study, condensed in **Publication_2**, we describe versatile rhomboid substrate, which is cleaved by a range of rhomboids. These substrates are compatible with detailed and precision enzyme kinetics as well as high-throughput screening. Since the LacYtm2 substrate is the only one of four tested substrates that is cleaved by all of the tested rhomboids (*Escherichia coli* GlpG, *Bacillus subtilis* YqgP, *Providencia stuartii* AarA and *Bacteroides thetaiotaomicron* Btio3), we have chosen it for further sequence optimisation. We present the synthetic peptide KSp31, derived from the transmembrane and juxtamembrane regions of LacYtm2, as well as its fluorescent UV and Red-shifted variants KSp35 and KSp76, generated by introducing (Lys)DABCYL-(Glu)EDANS and (Lys)TAMRA-

(Cys)QXL610 fluorophore-quencher pairs at positions P5 and P4' of the substrate, respectively. All peptide substrate variants are efficiently cleaved by GlpG, AarA and BtioR3 and poorly by YqgP, in the detergent-based *in vitro* assay, at the same site as the parental chimeric LacYtm2 substrate. The modest activity of YqgP may be attributed to the absence of lipids in the reaction, as it was shown previously that GlpG and YqgP prefer, and in case of YqgP also depend on the presence of lipid molecules during *in vitro* catalysis [46]. Moreover, it was previously shown that membrane lipids that positively regulate YqgP activity *in vitro* had the opposite effect on the activity of GlpG [46]. In order to tune and optimise the substrate cleavability, we first introduced the activatory mutations, based on GlpG preference screen, into the recombinant LacYtm2-derived proteins. We show that the HISKA (P1 position) and RVRHA (P1-5 positions) mutants are cleaved significantly better than the parental HISKS and that multiple mutations have got additive effect of these changes on substrate cleavability. The same effect of RVRHA was observed when the sequence is introduced into the wild-type LacYtm2-derived fluorescent peptide substrate KSp76, generating the KSp64 molecule. Interestingly, the KSp64 substrate is more selective then KSp76, as it is barely cleaved by AarA or YqgP proteases when compared with GlpG.

Inhibition of biologically relavant rhomboids represents another tempting goal in this field. To get closer to this goal, we show the strategy based on development of modifiable specific and highly selective compounds. In **Publication_3**, we describe peptidyl ketoamides as new rhomboid protease inhibitors. A series of inhibitors containing previously optimised RVRHA pentapeptide, with substituted amine group of the ketoamide (**Fig. 15a**) exhibit an impressive increase of inhibitory potency. The knowledge of rhomboid subsite preferences proves to be crucial, since the binding of the peptidic part of the inhibitor resembles substrate interaction. Our coumpounds bind covalently and reversibly with a slow binding mechanism. Thet act in a non-competitive mode, as described previously [59] which means that they interact either with free or exosite-bound, but active site-unoccupied rhomboid (**Fig. 6a, d**) and approach the catalytic center from the surrounding solvent.

Interestingly, we observe that the RVRHA-based ketoamide inhibitor effectively blocks *Bacillus subtilis* rhomboid YqgP *in vivo* activity, while the introduction of the RVRHA sequence does not improve the *in vitro* cleavage of the peptide substrate by YqgP. This effect is not observed for GlpG, since the RVRHA substrate is cleaved *in vitro* significantly better than wild-type substrate and the RVRHA ketoamide inhibits GlpG *in vivo* similarly to YqgP, at nanomolar range. We attribute this discrepancy in substrate cleavage to the poor activity of YqgP in the detergent micelles. The high potency of ketoamides against rhomboids is partly due to the presence of the hydrophobic substituent of the amidic nitrogen (phenylbutyl in this case) that interacts both with GlpG and, given by the active site residue conservation, probably also with YqgP prime site subsites S1' and S2'. Possibly, this could be

examined by determining the YqgP substrate preferences by substrate positional scanning screen, as was performed for GlpG, as described in **Publication_1**. Another possibility could be to analyse the contributions of RVRHA or wild-type peptide sequence such as HISKS of LacYtm2 or TIATA of TatA in ketoamides lacking or having the N-phenylbutyl substituent. Importantly, peptidyl ketoamides have been used clinically for inhibitioin of hepatitis C virus protease and treatment of hepatitis C virus infection, implying that our peptidyl ketoamides will likely be compatible with possible future pharmacological use.

My first author publication, listed here as **Publication_4**, discusses the deciphering of the biological function of YqgP rhomboid protease from *Bacillus subtilis.* To analyse the degradome and the interactome of YqgP, we wanted to use a robust and highly sensitive approach that would enable the identification of, most likely, low abundant rhomboid-related cleavage products, while keeping the conditions as close to those found in the native lipid membrane as possible. In quantitative proteomics experiments based either on SILAC labelling or label-free co-immunoprecipitation and pull-down techniques, we identify MgtE, the main magnesium transporter in *Bacillus subtilis*, as the natural substrate of YqgP rhomboid. Endogenous MgtE is processed by endogenous and overexpressed YqgP and cleavage is inhibited specifically by low micromolar peptidyl-ketoamide inhibitor. The approximate YqgP cleavage site is located within the L1 loop region between transmembrane helices TM1 and TM2 at the periplasmic side of MgtE, which is in agreement with the topology and orientation of rhomboid active site. This region of cleavage also means that YqgP cleavage inactivates the transporter. Based on what we have learnt about the biology of MgtE from the relevant literature, which was reviewed in section 0, we tested whether YqgP degrades actively transporting MgtE during elevated concentrations of divalent transition metals such as Mn^{2+}, Zn^{2+}, Co^{2+} or Ni^{2+}, which were shown to impair MgtE function either by hijacking the transport itself or by blocking the conducting pore, especially in low-magnesium conditions, when the MgtE homodimer is in its open conformation. It was not naïve to think that in such a situation, which is likely to occur in the native niche that *Bacillus subtilis* inhabits, unregulated import of transition metal cations, the intracellular homeostasis of which is strictly controlled, would lead to potentially toxic mismetallation of magnesium-dependent processes.

Addition of manganese during magnesium depletion indeed enhances YqgP activity towards MgtE. Such conditions affect overall fitness of both wild-type and Δ*yqgP Bacillus subtilis* strains, but the growth arrest of the mutant lacking YqgP is significantly more pronounced, and overexpression of YqgP overcompensates the fitness loss, mimicking the situation when cells grew without manganese stress. We observe that also zinc and to some extent also cobalt divalent cations enhance the activity of YqgP towards MgtE, indicating a putative role of the rhomboid in degradation of mismetallated MgtE molecules. Interestingly, ectopic steady state overexpression of MgtE in unstressed cells leads to culture

growth inhibition. This effect is rescued by the overexpression of wild-type YqgP, or strikingly even by the overexpression of its YqgP.S288A catalytic mutant.

We investigated YqgP regulation in more detail. Intriguingly, deletion of YqgP extramembrane domains affects the MgtE cleavage in opposite ways. While the YqgPΔNTD activity is decreased with respect to Mgte (but not TatA), the YqgPΔCTD variant cleaves both MgtE as well as model substrate TatA with more or less the same efficiency and significantly better than does wild-type enzyme. We show that NTD contributes to manganese stress response *in vivo.* It also serves as transient metal sensor and it binds different transient cations with low millimolar affinity *in vitro*. Although these interactions are considered rather weak, one would expect that, thinking of the metal sensing nature of the interaction, the affinities may be physiologically relevant. Focusing on the solution NMR structure of YqgP N-terminal domain, based on the Dali server search for structurally similar proteins, I found that its domain fold of YqgP NTD resembles that of a type II restriction endonuclease, which binds the Mg^{2+} cation as a cofactor in its active site. The NMR titration also detects the stepwise changes in chemical shifts and peak intensities in 2D NMR spectra of the ^{15}N- labelled NTD caused by manganese addition. We show probable interaction site for manganese formed by D50 and D52 residues, which are located in the negatively-charged pocket at the surface of the NTD.

Finally, we confirm our pull-down data that YqgP associates with FtsH, which is the main membrane protein quality control protease in *Bacillus subtilis.* Both proteases cooperate on the cleavage of MgtE (**Fig. 26**), and we show that YqgP-NTD is responsible for the cooperation.

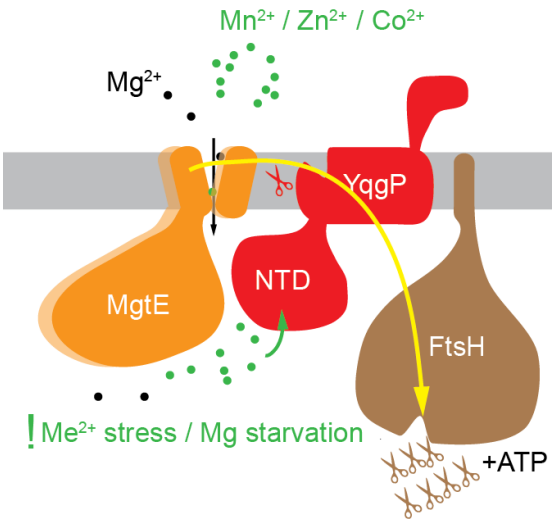


Fig. 26: The role of YqgP in magnesium transport homeostasis: YqgP recognizes MgtE during elevated Mn^{2+}, Zn^{2+} Co^{2+} conditions, which are sensed by rhomboid's N-terminal cytolic domain. The primary cleavage by YqgP results in the cleavage products being presented to the processive AAA+ zinc-dependent metalloprotease FtsH, which finishes their degradation and removes them from the membrane.

The effective degradation of MgtE by FtsH requires the active site of YqgP to be unobstructed, which indicates that YqgP recognizes MgtE first, cleaves it, and only then presents it to FtsH. The YqgP-dependent degradation of MgtE requires ongoing translation. Most likely, YqgP recognizes partially unfolded or prematurely terminated MgtE. The folding stress may be induced in specific conditions, such as transition metal cation stress. However, we were not able to link this phenomenon to FtsH activity, mainly due to the technical difficulties. Such assay requires the cultivation of the thermosensitive *ΔftsH* strain in minimal medium with low magnesium and high manganese content. Since *ftsH* mutants are already sick, this experiment is practically impossible to perform because the *ftsH* mutants just do not grow well enough in the additionally stressing conditions.

The wild-type YqgP-dependent degradation product, but not the alternative, YqgP.S288A-dependent product, is cleared off the membrane by FtsH and possibly also by other protease/s, as shown in Fig. 25e. Deletion of *ftsH* leads to a moderate stabilization of the wild-type YqgP-dependent cleavage product, while the alternative cleavage product dependent on YqgP.S288A is stable over the chase period. Therefore, we suggest that for effective MgtE processing and subsequent shedding of the degradation products, the activity of wild-type YqgP is crucial and only correctly cleaved MgtE can be presented to the proteolytic hub that FtsH is part of, and that this proteolytic hub does not recognise full-length MgtE. The fact that even catalytic mutant YqgP.288A is able to present the substrate to FtsH implies that YqgP can act also as the pseudoprotease, indicating a straightforward evolutionary strategy towards the emergence of multiple rhomboid pseudoproteases observed in eukaryotic proteomes.

6 CONCLUSIONS

- Amino acid preferences of GlpG in the P1 to P5 subsites were identified. GlpG prefers amino acids with small unbranched side chains at the P1 and with large hydrophobic side chains at the P4 position of the substrate non-prime side. Structural evidence for these observations was provided.

- The knowledge of the GlpG substrate preferences was used as the basis that serves for the design of versatile fluorogenic substrate and potent inhibitors, to study rhomboid structure, enzymology and biology.

- Two classes of substrate-derived mechanism based inhibitors- peptidyl chloromethylketones and N-substituted peptidyl ketoamides were developed in our laboratory.

- The N-substituted pentapeptidyl ketoamides with the peptidyl part resembling the optimal substrate sequence in the P1-P5 positions, are highly selective and potent rhomboid protease inhibitors active against endogenous bacterial rhomboids GlpG and YqgP of *Escherichia coli* and *Bacillus subtilis*, respectively.

- Magnesium transporter MgtE was identified as the natural substrate of *Bacillus subtilis* rhomboid YqgP.

- YqgP activity improves fitness of *Bacillus subtilis* during transition metal cation stress.

- The N-terminal cytoplasmic domain of YqgP is crucial for its activity, it binds Mn^{2+} *in vitro* and acts as the manganese toxicity sensor *in vivo.*

- YqgP cooperates with FtsH, the main membrane protein quality control protease in *Bacillus subtilis*, in maintaining proteostasis of MgtE. YqgP also acts as a substrate adaptor for FtsH.

7 LIST OF PUBLICATIONS

Publication 1:

Zoll, S., Stanchev, S., **Began, J.**, Skerle, J., Lepsik, M., Peclinovska, L., Majer, P. and Strisovsky, K., *Substrate binding and specificity of rhomboid intramembrane protease revealed by substrate-peptide complex structures.* **EMBO J**, 2014. 33(20): 2408-21.

Publication 2:

Tichá, A., Stanchev, S., Škerle, J., **Began, J.**, Ingr, M., Švehlová, K., Polovinkin, L., Růžička, M., Bednárová, L., Hadravová, R., Poláchová, E., Rampírová, P., Březinová, J., Kašička, V., Majer, P., and Strisovsky, K., *Sensitive Versatile Fluorogenic Transmembrane Peptide Substrates for Rhomboid Intramembrane Proteases*, **J Biol Chem**, 2017. 292(7): 2703-2713.

Publication 3:

Tichá, A., Stanchev, S., Vinothkumar, K. R., Mikles, D. C., Pachl, P., **Began, J.,** Škerle, J., Švehlová, K., Nguyen, M. T. N., Verhelst, S. H. L., Johnson, D. C., Bachovchin, D. A., Lepšík, M., Majer, P., and Strisovsky, K., *General and Modular Strategy for Designing Potent, Selective, and Pharmacologically Compliant Inhibitors of Rhomboid Proteases,* **Cell Chem Biol**, 2017. 24(12): 1523-1536.e4.

Publication 4:

Began, J., Cordier, B., Březinová, J., Delisle, J., Hexnerová, R., Srb, P., Rampírová, P., Kožíšek, M., Couté, Y., Galinier, A., Veverka, V., Doan, T., Strisovsky, K., *Rhomboid intramembrane protease YqgP licences bacterial membrane protein quality control as adaptor of FtsH AAA protease.* **EMBO J**, 2019, accepted for publishing (12.12.2019)

8 REFERENCES

1. Dalbey, R.E., P. Wang, and J.M. van Dijl, *Membrane proteases in the bacterial protein secretion and quality control pathway.* Microbiol Mol Biol Rev, 2012. **76**(2): p. 311-30.
2. Preston, G.M. and J.L. Brodsky, *The evolving role of ubiquitin modification in endoplasmic reticulum-associated degradation.* Biochem J, 2017. **474**(4): p. 445-469.
3. Zattas, D. and M. Hochstrasser, *Ubiquitin-dependent protein degradation at the yeast endoplasmic reticulum and nuclear envelope.* Crit Rev Biochem Mol Biol, 2015. **50**(1): p. 1-17.
4. Yedidi, R.S., P. Wendler, and C. Enenkel, *AAA-ATPases in Protein Degradation.* Front Mol Biosci, 2017. **4**: p. 42.
5. Bodnar, N.O. and T.A. Rapoport, *Molecular Mechanism of Substrate Processing by the Cdc48 ATPase Complex.* Cell, 2017. **169**(4): p. 722-735 e9.
6. Bodnar, N. and T. Rapoport, *Toward an understanding of the Cdc48/p97 ATPase.* F1000Res, 2017. **6**: p. 1318.
7. Kato, Y., K. Nishiyama, and H. Tokuda, *Depletion of SecDF-YajC causes a decrease in the level of SecG: implication for their functional interaction.* FEBS Lett, 2003. **550**(1-3): p. 114-8.
8. Komar, J., et al., *Membrane protein insertion and assembly by the bacterial holo-translocon SecYEG-SecDF-YajC-YidC.* Biochem J, 2016. **473**(19): p. 3341-54.
9. Kuhn, A., H.G. Koch, and R.E. Dalbey, *Targeting and Insertion of Membrane Proteins.* EcoSal Plus, 2017. **7**(2).
10. Dalbey, R.E., P. Wang, and A. Kuhn, *Assembly of Bacterial Inner Membrane Proteins.* Annual Review of Biochemistry, 2011. **80**(1): p. 161-187.
11. Bittner, L.M., J. Arends, and F. Narberhaus, *When, how and why? Regulated proteolysis by the essential FtsH protease in Escherichia coli.* Biol Chem, 2017. **398**(5-6): p. 625-635.
12. Saikawa, N., Y. Akiyama, and K. Ito, *FtsH exists as an exceptionally large complex containing HflKC in the plasma membrane of Escherichia coli.* J Struct Biol, 2004. **146**(1-2): p. 123-9.
13. Lopez, D. and R. Kolter, *Functional microdomains in bacterial membranes.* Genes Dev, 2010. **24**(17): p. 1893-902.
14. Bach, J.N. and M. Bramkamp, *Flotillins functionally organize the bacterial membrane.* Mol Microbiol, 2013. **88**(6): p. 1205-17.
15. Wehrl, W., M. Niederweis, and W. Schumann, *The FtsH protein accumulates at the septum of Bacillus subtilis during cell division and sporulation.* J Bacteriol, 2000. **182**(13): p. 3870-3.
16. Yepes, A., et al., *The biofilm formation defect of a Bacillus subtilis flotillin-defective mutant involves the protease FtsH.* Mol Microbiol, 2012. **86**(2): p. 457-71.
17. Thi Nguyen, H.B. and W. Schumann, *The sporulation control gene spo0M of Bacillus subtilis is a target of the FtsH metalloprotease.* Res Microbiol, 2012. **163**(2): p. 114-8.
18. Deuerling, E., et al., *The ftsH gene of Bacillus subtilis is involved in major cellular processes such as sporulation, stress adaptation and secretion.* Mol Microbiol, 1997. **23**(5): p. 921-33.
19. Erez, E., D. Fass, and E. Bibi, *How intramembrane proteases bury hydrolytic reactions in the membrane.* Nature, 2009. **459**(7245): p. 371-8.
20. Lichtenthaler, S.F., M.K. Lemberg, and R. Fluhrer, *Proteolytic ectodomain shedding of membrane proteins in mammals-hardware, concepts, and recent developments.* EMBO J, 2018.
21. Urban, S., *Making the cut: central roles of intramembrane proteolysis in pathogenic microorganisms.* Nature Reviews Microbiology, 2009. **7**(6): p. 411-23.
22. Avci, D. and M.K. Lemberg, *Clipping or Extracting: Two Ways to Membrane Protein Degradation.* Trends Cell Biol, 2015. **25**(10): p. 611-622.
23. Strisovsky, K., *Why cells need intramembrane proteases - a mechanistic perspective.* The FEBS Journal, 2016. **283**(10): p. 1837-1845.
24. Wolfe, M.S., *Intramembrane proteolysis.* Chem Rev, 2009. **109**(4): p. 1599-612.
25. Kinch, L.N., K. Ginalski, and N.V. Grishin, *Site-2 protease regulated intramembrane proteolysis: sequence homologs suggest an ancient signaling cascade.* Protein Science, 2006. **15**(1): p. 84-93.
26. Ponting, C.P., et al., *Identification of a novel family of presenilin homologues.* Human Molecular Genetics, 2002. **11**(9): p. 1037-44.
27. Weihofen, A., et al., *Identification of signal peptide peptidase, a presenilin-type aspartic protease.* Science, 2002. **296**(5576): p. 2215-8.

28. De Strooper, B., et al., *Deficiency of presenilin-1 inhibits the normal cleavage of amyloid precursor protein.* Nature, 1998. **391**(6665): p. 387-90.
29. Urban, S., J.R. Lee, and M. Freeman, *Drosophila rhomboid-1 defines a family of putative intramembrane serine proteases.* Cell, 2001. **107**(2): p. 173-82.
30. Manolaridis, I., et al., *Mechanism of farnesylated CAAX protein processing by the intramembrane protease Rce1.* Nature, 2013. **504**(7479): p. 301-5.
31. Urban, S. and M. Freeman, *Substrate specificity of rhomboid intramembrane proteases is governed by helix-breaking residues in the substrate transmembrane domain.* Mol Cell, 2003. **11**(6): p. 1425-34.
32. Lee, J.R., et al., *Regulated intracellular ligand transport and proteolysis control EGF signal activation in Drosophila.* Cell, 2001. **107**(2): p. 161-71.
33. Wasserman, J.D., S. Urban, and M. Freeman, *A family of rhomboid-like genes: Drosophila rhomboid-1 and roughoid/rhomboid-3 cooperate to activate EGF receptor signaling.* Genes Dev, 2000. **14**(13): p. 1651-63.
34. Koonin, E.V., et al., *The rhomboids: a nearly ubiquitous family of intramembrane serine proteases that probably evolved by multiple ancient horizontal gene transfers.* Genome Biol, 2003. **4**(3): p. R19.
35. Lemberg, M.K. and M. Freeman, *Functional and evolutionary implications of enhanced genomic analysis of rhomboid intramembrane proteases.* Genome Res, 2007. **17**(11): p. 1634-46.
36. Dowse, T.J. and D. Soldati, *Rhomboid-like proteins in Apicomplexa: phylogeny and nomenclature.* Trends Parasitol, 2005. **21**(6): p. 254-8.
37. Santos, J.M., A. Graindorge, and D. Soldati-Favre, *New insights into parasite rhomboid proteases.* Mol Biochem Parasitol, 2012. **182**(1-2): p. 27-36.
38. Li, Q., et al., *Differential evolution of members of the rhomboid gene family with conservative and divergent patterns.* New Phytol, 2014.
39. Tripathi, L.P. and R. Sowdhamini, *Cross genome comparisons of serine proteases in Arabidopsis and rice.* BMC Genomics, 2006. **7**: p. 200.
40. Adrain, C. and M. Freeman, *New lives for old: evolution of pseudoenzyme function illustrated by iRhoms.* Nat Rev Mol Cell Biol, 2012. **13**(8): p. 489-98.
41. Greenblatt, E.J., J.A. Olzmann, and R.R. Kopito, *Derlin-1 is a rhomboid pseudoprotease required for the dislocation of mutant alpha-1 antitrypsin from the endoplasmic reticulum.* Nat Struct Mol Biol, 2011. **18**(10): p. 1147-52.
42. Ong, Y.S., et al., *TMEM115 is an integral membrane protein of the Golgi complex involved in retrograde transport.* J Cell Sci, 2014. **127**(Pt 13): p. 2825-39.
43. Abba, M.C., et al., *Rhomboid domain containing 2 (RHBDD2): a novel cancer-related gene over-expressed in breast cancer.* Biochim Biophys Acta, 2009. **1792**(10): p. 988-97.
44. Liu, J., et al., *Rhomboid domain-containing protein 3 is a negative regulator of TLR3-triggered natural killer cell activation.* Proc Natl Acad Sci U S A, 2013. **110**(19): p. 7814-9.
45. Christianson, J.C., et al., *Defining human ERAD networks through an integrative mapping strategy.* Nat Cell Biol, 2011. **14**(1): p. 93-105.
46. Urban, S. and M.S. Wolfe, *Reconstitution of intramembrane proteolysis in vitro reveals that pure rhomboid is sufficient for catalysis and specificity.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(6): p. 1883-8.
47. Edbauer, D., et al., *Reconstitution of gamma-secretase activity.* Nat Cell Biol, 2003. **5**(5): p. 486-8.
48. Urban, S., D. Schlieper, and M. Freeman, *Conservation of intramembrane proteolytic activity and substrate specificity in prokaryotic and eukaryotic rhomboids.* Curr Biol, 2002. **12**(17): p. 1507-12.
49. Lemberg, M.K., et al., *Mechanism of intramembrane proteolysis investigated with purified rhomboid proteases.* EMBO J, 2005. **24**(3): p. 464-72.
50. Baker, R.P. and S. Urban, *Architectural and thermodynamic principles underlying intramembrane protease function.* Nat Chem Biol, 2012. **8**(9): p. 759-68.
51. Wang, Y., Y. Zhang, and Y. Ha, *Crystal structure of a rhomboid family intramembrane protease.* Nature, 2006. **444**(7116): p. 179-80.
52. Ben-Shem, A., D. Fass, and E. Bibi, *Structural basis for intramembrane proteolysis by rhomboid serine proteases.* Proc Natl Acad Sci U S A, 2007. **104**(2): p. 462-6.
53. Wu, Z., et al., *Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry.* Nat Struct Mol Biol, 2006. **13**(12): p. 1084-91.
54. Lemieux, M.J., et al., *The crystal structure of the rhomboid peptidase from Haemophilus influenzae provides insight into intramembrane proteolysis.* Proc Natl Acad Sci U S A, 2007. **104**(3): p. 750-4.

55. Vinothkumar, K.R., *Structure of rhomboid protease in a lipid environment.* J Mol Biol, 2011. **407**(2): p. 232-47.
56. Baker, R.P., et al., *Enzymatic analysis of a rhomboid intramembrane protease implicates transmembrane helix 5 as the lateral substrate gate.* Proc Natl Acad Sci U S A, 2007. **104**(20): p. 8257-62.
57. Wang, Y., et al., *The role of L1 loop in the mechanism of rhomboid intramembrane protease GlpG.* Journal of Molecular Biology, 2007. **374**(4): p. 1104-13.
58. Bondar, A.N., C. del Val, and S.H. White, *Rhomboid protease dynamics and lipid interactions.* Structure, 2009. **17**(3): p. 395-405.
59. Cho, S., S.W. Dickey, and S. Urban, *Crystal Structures and Inhibition Kinetics Reveal a Two-Stage Catalytic Mechanism with Drug Design Implications for Rhomboid Proteolysis.* Mol Cell, 2016. **61**(3): p. 329-340.
60. Zhou, Y., et al., *An internal water-retention site in the rhomboid intramembrane protease GlpG ensures catalytic efficiency.* Structure, 2012. **20**(7): p. 1255-63.
61. Wang, Y. and Y. Ha, *Open-cap conformation of intramembrane protease GlpG.* "Proceedings of the National Academy of Sciences, USA", 2007. **104**(7): p. 2098-102.
62. Urban, S. and R.P. Baker, *In vivo analysis reveals substrate-gating mutants of a rhomboid intramembrane protease display increased activity in living cells.* Biological Chemistry, 2008. **389**(8): p. 1107-15.
63. Shokhen, M. and A. Albeck, *How does the exosite of rhomboid protease affect substrate processing and inhibition?* Protein Sci, 2017. **26**(12): p. 2355-2366.
64. Brooks, C.L., et al., *Insights into substrate gating in H. influenzae rhomboid.* Journal of Molecular Biology, 2011. **407**(5): p. 687-97.
65. Xue, Y. and Y. Ha, *Large lateral movement of transmembrane helix S5 is not required for substrate access to the active site of rhomboid intramembrane protease.* J Biol Chem, 2013. **288**(23): p. 16645-54.
66. Vinothkumar, K.R., et al., *Structure of rhomboid protease in complex with beta-lactam inhibitors defines the S2' cavity.* Structure, 2013. **21**(6): p. 1051-8.
67. Xue, Y. and Y. Ha, *Catalytic mechanism of rhomboid protease GlpG probed by 3,4-dichloroisocoumarin and diisopropyl fluorophosphonate.* J Biol Chem, 2012. **287**(5): p. 3099-107.
68. Xue, Y., et al., *Conformational change in rhomboid protease GlpG induced by inhibitor binding to its S' subsites.* Biochemistry, 2012. **51**(18): p. 3723-31.
69. Vinothkumar, K.R., et al., *The structural basis for catalysis and substrate specificity of a rhomboid protease.* EMBO J, 2010. **29**(22): p. 3797-809.
70. Dickey, S.W., et al., *Proteolysis inside the membrane is a rate-governed reaction not driven by substrate affinity.* Cell, 2013. **155**(6): p. 1270-81.
71. Moin, S.M. and S. Urban, *Membrane immersion allows rhomboid proteases to achieve specificity by reading transmembrane segment dynamics.* Elife, 2012. **1**: p. e00173.
72. Tyndall, J.D., T. Nall, and D.P. Fairlie, *Proteases universally recognize beta strands in their active sites.* Chemical Reviews, 2005. **105**(3): p. 973-99.
73. Gallio, M., et al., *A conserved mechanism for extracellular signaling in eukaryotes and prokaryotes.* Proc Natl Acad Sci U S A, 2002. **99**(19): p. 12208-13.
74. Maegawa, S., K. Ito, and Y. Akiyama, *Proteolytic action of GlpG, a rhomboid protease in the Escherichia coli cytoplasmic membrane.* Biochemistry, 2005. **44**(41): p. 13543-52.
75. Strisovsky, K., H.J. Sharpe, and M. Freeman, *Sequence-specific intramembrane proteolysis: identification of a recognition motif in rhomboid substrates.* Mol Cell, 2009. **36**(6): p. 1048-59.
76. Akiyama, Y. and S. Maegawa, *Sequence features of substrates required for cleavage by GlpG, an Escherichia coli rhomboid protease.* Mol Microbiol, 2007. **64**(4): p. 1028-37.
77. Schechter, I. and A. Berger, *On the size of the active site in proteases. I. Papain.* Biochemical and Biophysical Research Communications, 1967. **27**(2): p. 157-62.
78. Stevenson, L.G., et al., *Rhomboid protease AarA mediates quorum-sensing in Providencia stuartii by activating TatA of the twin-arginine translocase.* Proc Natl Acad Sci U S A, 2007. **104**(3): p. 1003-8.
79. Tichá, A., B. Collis, and K. Strisovsky, *The Rhomboid Superfamily: Structural Mechanisms and Chemical Biology Opportunities.* Trends in Biochemical Sciences, 2018.
80. Strisovsky, K., *Mechanism and Inhibition of Rhomboid Proteases.* Methods Enzymol, 2017. **584**: p. 279-293.
81. Zoll, S., et al., *Substrate binding and specificity of rhomboid intramembrane protease revealed by substrate-peptide complex structures.* EMBO J, 2014. **33**(20): p. 2408-21.

82. Ekici, O.D., M. Paetzel, and R.E. Dalbey, *Unconventional serine proteases: variations on the catalytic Ser/His/Asp triad configuration.* Protein science : a publication of the Protein Society, 2008. **17**(12): p. 2023-2037.
83. Urban, S. and S.M. Moin, *A subset of membrane-altering agents and gamma-secretase modulators provoke nonsubstrate cleavage by rhomboid proteases.* Cell Rep, 2014. **8**(5): p. 1241-7.
84. Lee, A.G., *How lipids affect the activities of integral membrane proteins.* Biochimica et Biophysica Acta, 2004. **1666**(1-2): p. 62-87.
85. Sanders, C.R. and J.M. Hutchison, *Membrane properties that shape the evolution of membrane enzymes.* Curr Opin Struct Biol, 2018. **51**: p. 80-91.
86. Reddy, T. and J.K. Rainey, *Multifaceted substrate capture scheme of a rhomboid protease.* J Phys Chem B, 2012. **116**(30): p. 8942-54.
87. Foo, A.C., et al., *Influence of hydrophobic mismatch on the catalytic activity of E. coli GlpG rhomboid protease.* Protein Sci, 2014.
88. Kreutzberger, A.J.B., et al., *Rhomboid distorts lipids to break the viscosity-imposed speed limit of membrane diffusion.* Science, 2019. **363**(6426).
89. Paschkowsky, S., F. Oestereich, and L.M. Munter, *Embedded in the Membrane: How Lipids Confer Activity and Specificity to Intramembrane Proteases.* J Membr Biol, 2018. **251**(3): p. 369-378.
90. Cortesio, C.L., E.B. Lewellyn, and D.G. Drubin, *Control of lipid organization and actin assembly during clathrin-mediated endocytosis by the cytoplasmic tail of the rhomboid protein Rbd2.* Mol Biol Cell, 2015. **26**(8): p. 1509-22.
91. Fleig, L., et al., *Ubiquitin-Dependent Intramembrane Rhomboid Protease Promotes ERAD of Membrane Proteins.* Molecular Cell, 2012. **47**(4): p. 558-69.
92. Sherratt, A.R., et al., *Activity-Based Protein Profiling of the Escherichia coli GlpG Rhomboid Protein Delineates the Catalytic Core.* Biochemistry, 2012. **51**(39): p. 7794-803.
93. McQuibban, G.A., S. Saurya, and M. Freeman, *Mitochondrial membrane remodelling regulated by a conserved rhomboid protease.* Nature, 2003. **423**(6939): p. 537-41.
94. Shi, G. and G.A. McQuibban, *The Mitochondrial Rhomboid Protease PARL Is Regulated by PDK2 to Integrate Mitochondrial Quality Control and Metabolism.* Cell Rep, 2017. **18**(6): p. 1458-1472.
95. Lohi, O., S. Urban, and M. Freeman, *Diverse substrate recognition mechanisms for rhomboids; thrombomodulin is cleaved by Mammalian rhomboids.* Curr Biol, 2004. **14**(3): p. 236-41.
96. Paslawski, W., et al., *Cooperative folding of a polytopic alpha-helical membrane protein involves a compact N-terminal nucleus and nonnative loops.* Proc Natl Acad Sci U S A, 2015. **112**(26): p. 7978-83.
97. Yogev, S., E.D. Schejter, and B.Z. Shilo, *Drosophila EGFR signalling is modulated by differential compartmentalization of Rhomboid intramembrane proteases.* EMBO J, 2008. **27**(8): p. 1219-30.
98. Gabay, L., R. Seger, and B.Z. Shilo, *MAP kinase in situ activation atlas during Drosophila embryogenesis.* Development, 1997. **124**(18): p. 3535-41.
99. Pierrat, O.A., et al., *Monocyclic beta-lactams are selective, mechanism-based inhibitors of rhomboid intramembrane proteases.* ACS Chem Biol, 2011. **6**(4): p. 325-35.
100. Wolf, E.V., et al., *A new class of rhomboid protease inhibitors discovered by activity-based fluorescence polarization.* PLoS One, 2013. **8**(8): p. e72307.
101. Wolf, E.V., A. Zeissler, and S.H. Verhelst, *Inhibitor Fingerprinting of Rhomboid Proteases by Activity-Based Protein Profiling Reveals Inhibitor Selectivity and Rhomboid Autoprocessing.* ACS Chem Biol, 2015. **10**(10): p. 2325-33.
102. Yang, J., et al., *Benzoxazin-4-ones as novel, easily accessible inhibitors for rhomboid proteases.* Bioorg Med Chem Lett, 2018. **28**(8): p. 1423-1427.
103. Goel, P., et al., *Discovery and validation of 2-styryl substituted benzoxazin-4-ones as a novel scaffold for rhomboid protease inhibitors.* Bioorganic & Medicinal Chemistry Letters, 2018. **28**(8): p. 1417-1422.
104. Goel, P., et al., *Discovery and Biological Evaluation of Potent and Selective N-Methylene Saccharin-Derived Inhibitors for Rhomboid Intramembrane Proteases.* Biochemistry, 2017. **56**(51): p. 6713-6725.
105. Dusterhoft, S., U. Kunzel, and M. Freeman, *Rhomboid proteases in human disease: Mechanisms and future prospects.* Biochim Biophys Acta, 2017. **1864**(11 Pt B): p. 2200-2209.
106. Lastun, V.L., A.G. Grieve, and M. Freeman, *Substrates and physiological functions of secretase rhomboid proteases.* Semin Cell Dev Biol, 2016. **60**: p. 10-18.
107. Freeman, M., *Rhomboids, signalling and cell biology.* Biochem Soc Trans, 2016. **44**(3): p. 945-50.
108. Johnson, N., et al., *Quantitative proteomics screen identifies a substrate repertoire of rhomboid protease RHBDL2 in human cells and implicates it in epithelial homeostasis.* Sci Rep, 2017. **7**(1): p. 7283.

109. Adrain, C., et al., *Mammalian EGF receptor activation by the rhomboid protease RHBDL2.* EMBO Rep, 2011. **12**(5): p. 421-7.
110. Wan, C., et al., *Exosome-Related Multi-Pass Transmembrane Protein TSAP6 Is a Target of Rhomboid Protease RHBDD1-Induced Proteolysis.* PLoS One, 2012. **7**(5): p. e37452.
111. Wunderle, L., et al., *Rhomboid intramembrane protease RHBDL4 triggers ER-export and non-canonical secretion of membrane-anchored TGFalpha.* Sci Rep, 2016. **6**: p. 27342.
112. Song, W., et al., *Rhomboid domain containing 1 promotes colorectal cancer growth through activation of the EGFR signalling pathway.* Nat Commun, 2015. **6**: p. 8022.
113. Miao, F., et al., *RHBDD1 upregulates EGFR via the AP-1 pathway in colorectal cancer.* Oncotarget, 2017. **8**(15): p. 25251-25260.
114. Han, J., et al., *Lentivirus-mediated knockdown of rhomboid domain containing 1 inhibits colorectal cancer cell growth.* Mol Med Rep, 2015.
115. Ren, X., et al., *Rhomboid domain containing 1 inhibits cell apoptosis by upregulating AP-1 activity and its downstream target Bcl-3.* FEBS Lett, 2013. **587**(12): p. 1793-8.
116. Lim, J.J., et al., *Structural insights into the interaction of p97 N-terminus domain and VBM in rhomboid protease, RHBDL4.* Biochem J, 2016. **473**(18): p. 2863-80.
117. Brossier, F., et al., *A spatially localized rhomboid protease cleaves cell surface adhesins essential for invasion by Toxoplasma.* "Proceedings of the National Academy of Sciences, USA", 2005. **102**(11): p. 4146-51.
118. O'Donnell, R.A., et al., *Intramembrane proteolysis mediates shedding of a key adhesin during erythrocyte invasion by the malaria parasite.* J Cell Biol, 2006. **174**(7): p. 1023-33.
119. Buguliskis, J.S., et al., *Rhomboid 4 (ROM4) affects the processing of surface adhesins and facilitates host cell invasion by Toxoplasma gondii.* PLoS Pathog, 2010. **6**(4): p. e1000858.
120. Spinazzi, M. and B. De Strooper, *PARL: The mitochondrial rhomboid protease.* Semin Cell Dev Biol, 2016.
121. Meissner, C., et al., *The mitochondrial intramembrane protease PARL cleaves human Pink1 to regulate Pink1 trafficking.* Journal of Neurochemistry, 2011. **117**(5): p. 856-67.
122. Meissner, C., et al., *Intramembrane protease PARL defines a negative regulator of PINK1- and PARK2/Parkin-dependent mitophagy.* Autophagy, 2015. **11**(9): p. 1484-98.
123. Sekine, S., et al., *Rhomboid protease PARL mediates the mitochondrial membrane potential loss-induced cleavage of PGAM5.* J Biol Chem, 2012. **287**(41): p. 34635-45.
124. Spinazzi, M., et al., *PARL deficiency in mouse causes Complex III defects, coenzyme Q depletion, and Leigh-like syndrome.* Proc Natl Acad Sci U S A, 2019. **116**(1): p. 277-286.
125. Pils, B. and J. Schultz, *Inactive Enzyme-homologues Find New Function in Regulatory Processes.* Journal of Molecular Biology, 2004. **340**(3): p. 399-404.
126. Eyers, P.A. and J.M. Murphy, *The evolving world of pseudoenzymes: proteins, prejudice and zombies.* BMC Biol, 2016. **14**(1): p. 98.
127. Murphy, J.M., P.D. Mace, and P.A. Eyers, *Live and let die: insights into pseudoenzyme mechanisms from structure.* Current Opinion in Structural Biology, 2017. **47**: p. 95-104.
128. Murphy, J.M., H. Farhan, and P.A. Eyers, *Bio-Zombie: the rise of pseudoenzymes in biology.* Biochem Soc Trans, 2017. **45**(2): p. 537-544.
129. Dulloo, I., S. Muliyil, and M. Freeman, *The molecular, cellular and pathophysiological roles of iRhom pseudoproteases.* Open Biol, 2019. **9**(3): p. 190003.
130. Lemberg, M.K. and C. Adrain, *Inactive rhomboid proteins: New mechanisms with implications in health and disease.* Semin Cell Dev Biol, 2016. **60**: p. 29-37.
131. Zettl, M., et al., *Rhomboid family pseudoproteases use the ER quality control machinery to regulate intercellular signaling.* Cell, 2011. **145**(1): p. 79-91.
132. Adrain, C., et al., *Tumor necrosis factor signaling requires iRhom2 to promote trafficking and activation of TACE.* Science, 2012. **335**(6065): p. 225-8.
133. McIlwain, D.R., et al., *iRhom2 regulation of TACE controls TNF-mediated protection against Listeria and responses to LPS.* Science, 2012. **335**(6065): p. 229-32.
134. Oikonomidi, I., et al., *iTAP, a novel iRhom interactor, controls TNF secretion by policing the stability of iRhom/TACE.* Elife, 2018. **7**: p. e35032.
135. Kunzel, U., et al., *FRMD8 promotes inflammatory and growth factor signalling by stabilising the iRhom/ADAM17 sheddase complex.* Elife, 2018. **7**: p. e35012.
136. Avci, D. and M.K. Lemberg, *Membrane Protein Dislocation by the Rhomboid Pseudoprotease Dfm1: No Pore Needed?* Mol Cell, 2018. **69**(2): p. 161-162.

137. Maretzky, T., et al., *iRhom2 controls the substrate selectivity of stimulated ADAM17-dependent ectodomain shedding.* Proc Natl Acad Sci U S A, 2013. **110**(28): p. 11433-8.
138. Cavadas, M., et al., *Phosphorylation of iRhom2 Controls Stimulated Proteolytic Shedding by the Metalloprotease ADAM17/TACE.* Cell Rep, 2017. **21**(3): p. 745-757.
139. Grieve, A.G., et al., *Phosphorylation of iRhom2 at the plasma membrane controls mammalian TACE-dependent inflammatory and growth factor signalling.* Elife, 2017. **6**.
140. Luo, W.W., et al., *iRhom2 is essential for innate immunity to DNA viruses by mediating trafficking and stability of the adaptor STING.* Nat Immunol, 2016. **17**(9): p. 1057-66.
141. Luo, W.W., et al., *iRhom2 is essential for innate immunity to RNA virus by antagonizing ER- and mitochondria-associated degradation of VISA.* PLoS Pathog, 2017. **13**(11): p. e1006693.
142. Knop, M., et al., *Der1, a novel protein specifically required for endoplasmic reticulum degradation in yeast.* EMBO J, 1996. **15**(4): p. 753-63.
143. Sato, B.K. and R.Y. Hampton, *Yeast Derlin Dfm1 interacts with Cdc48 and functions in ER homeostasis.* Yeast, 2006. **23**(14-15): p. 1053-64.
144. Goder, V., P. Carvalho, and T.A. Rapoport, *The ER-associated degradation component Der1p and its homolog Dfm1p are contained in complexes with distinct cofactors of the ATPase Cdc48p.* FEBS Lett, 2008. **582**(11): p. 1575-80.
145. Stolz, A., et al., *Dfm1 forms distinct complexes with Cdc48 and the ER ubiquitin ligases and is required for ERAD.* Traffic, 2010. **11**(10): p. 1363-9.
146. Neal, S., et al., *The Dfm1 Derlin Is Required for ERAD Retrotranslocation of Integral Membrane Proteins.* Mol Cell, 2018. **69**(2): p. 306-320 e4.
147. Avci, D., et al., *The Yeast ER-Intramembrane Protease Ypf1 Refines Nutrient Sensing by Regulating Transporter Abundance.* Mol Cell, 2014.
148. Ye, Y., et al., *A membrane protein complex mediates retro-translocation from the ER lumen into the cytosol.* Nature, 2004. **429**(6994): p. 841-7.
149. Lilley, B.N. and H.L. Ploegh, *A membrane protein required for dislocation of misfolded proteins from the ER.* Nature, 2004. **429**(6994): p. 834-40.
150. Rather, P.N., et al., *Providencia stuartii genes activated by cell-to-cell signaling and identification of a gene required for production or activity of an extracellular factor.* J Bacteriol, 1999. **181**(23): p. 7185-91.
151. Rather, P.N., M.M. Parojcic, and M.R. Paradise, *An extracellular factor regulating expression of the chromosomal aminoglycoside 2'-N-acetyltransferase of Providencia stuartii.* Antimicrob Agents Chemother, 1997. **41**(8): p. 1749-54.
152. Rather, P.N. and E. Orosz, *Characterization of aarA, a pleiotrophic negative regulator of the 2'-N-acetyltransferase in Providencia stuartii.* J Bacteriol, 1994. **176**(16): p. 5140-4.
153. Goosens, V.J. and J.M. van Dijl, *Twin-Arginine Protein Translocation.* Curr Top Microbiol Immunol, 2017. **404**: p. 69-94.
154. Fritsch, M.J., et al., *Processing by rhomboid protease is required for Providencia stuartii TatA to interact with TatC and to form functional homo-oligomeric complexes.* Mol Microbiol, 2012. **84**(6): p. 1108-23.
155. Clemmer, K.M., et al., *Functional Characterization of Escherichia coli GlpG and Additional Rhomboid Proteins Using an aarA Mutant of Providencia stuartii.* Journal of Bacteriology, 2006. **188**(9): p. 3415-3419.
156. Erez, E. and E. Bibi, *Cleavage of a multispanning membrane protein by an intramembrane serine protease.* Biochemistry, 2009. **48**(51): p. 12314-22.
157. Vila, J., et al., *Escherichia coli: an old friend with new tidings.* FEMS Microbiol Rev, 2016. **40**(4): p. 437-463.
158. Russell, C.W., et al., *The Rhomboid Protease GlpG Promotes the Persistence of Extraintestinal Pathogenic Escherichia coli within the Gut.* Infect Immun, 2017. **85**(6).
159. Skarlatos, P. and M.K. Dahl, *The glucose kinase of Bacillus subtilis.* J Bacteriol, 1998. **180**(12): p. 3222-6.
160. Mesak, L.R., F.M. Mesak, and M.K. Dahl, *Expression of a novel gene, gluP, is essential for normal Bacillus subtilis cell division and contributes to glucose export.* BMC Microbiol, 2004. **4**: p. 13.
161. Singh, J., et al., *Structure-based design of a potent, selective, and irreversible inhibitor of the catalytic domain of the erbB receptor subfamily of protein tyrosine kinases.* J Med Chem, 1997. **40**(7): p. 1130-5.
162. Jakubovics, N.S. and H.F. Jenkinson, *Out of the iron age: new insights into the critical role of manganese homeostasis in bacteria.* Microbiology, 2001. **147**(Pt 7): p. 1709-18.
163. Barwinska-Sendra, A. and K.J. Waldron, *The Role of Intermetal Competition and Mis-Metalation in Metal Toxicity.* Adv Microb Physiol, 2017. **70**: p. 315-379.

164. Capdevila, D.A., K.A. Edmonds, and D.P. Giedroc, *Metallochaperones and metalloregulation in bacteria.* Essays Biochem, 2017. **61**(2): p. 177-200.
165. Groisman, E.A., et al., *Bacterial Mg2+ homeostasis, transport, and virulence.* Annu Rev Genet, 2013. **47**: p. 625-46.
166. Moomaw, A.S. and M.E. Maguire, *The unique nature of mg2+ channels.* Physiology (Bethesda), 2008. **23**: p. 275-85.
167. Papp-Wallace, K.M. and M.E. Maguire, *Magnesium Transport and Magnesium Homeostasis.* EcoSal Plus, 2008. **3**(1).
168. Quamme, G.A., *Molecular identification of ancient and modern mammalian magnesium transporters.* Am J Physiol Cell Physiol, 2010. **298**(3): p. C407-29.
169. Townsend, D.E., et al., *Cloning of the mgtE Mg2+ transporter from Providencia stuartii and the distribution of mgtE in gram-negative and gram-positive bacteria.* J Bacteriol, 1995. **177**(18): p. 5350-4.
170. Smith, R.L., L.J. Thompson, and M.E. Maguire, *Cloning and characterization of MgtE, a putative new class of Mg2+ transporter from Bacillus firmus OF4.* J Bacteriol, 1995. **177**(5): p. 1233-8.
171. Aron, L., et al., *Identification and mapping of a chromosomal gene cluster of Borrelia burgdorferi containing genes expressed in vivo.* FEMS Microbiol Lett, 1996. **145**(3): p. 309-14.
172. Chakravarty, S., et al., *Pseudomonas aeruginosa Magnesium Transporter MgtE Inhibits Type III Secretion System Gene Expression by Stimulating rsmYZ Transcription.* J Bacteriol, 2017. **199**(23).
173. Sahni, J., Y. Song, and A.M. Scharenberg, *The B. subtilis MgtE magnesium transporter can functionally compensate TRPM7-deficiency in vertebrate B-cells.* PLoS One, 2012. **7**(9): p. e44452.
174. Sahni, J. and A.M. Scharenberg, *The SLC41 family of MgtE-like magnesium transporters.* Mol Aspects Med, 2013. **34**(2-3): p. 620-8.
175. Hattori, M., et al., *Mg(2+)-dependent gating of bacterial MgtE channel underlies Mg(2+) homeostasis.* EMBO J, 2009. **28**(22): p. 3602-12.
176. Wakeman, C.A., et al., *Assessment of the requirements for magnesium transporters in Bacillus subtilis.* J Bacteriol, 2014. **196**(6): p. 1206-14.
177. Ishitani, R., et al., *Mg2+-sensing mechanism of Mg2+ transporter MgtE probed by molecular dynamics study.* Proc Natl Acad Sci U S A, 2008. **105**(40): p. 15393-8.
178. Maruyama, T., et al., *Functional roles of Mg(2+) binding sites in ion-dependent gating of a Mg(2+) channel, MgtE, revealed by solution NMR.* Elife, 2018. **7**.
179. Hattori, M., et al., *Crystal structure of the MgtE Mg2+ transporter.* Nature, 2007. **448**(7157): p. 1072-5.
180. Ragumani, S., et al., *Structural studies on cytosolic domain of magnesium transporter MgtE from Enterococcus faecalis.* Proteins, 2010. **78**(2): p. 487-91.
181. Imai, S., et al., *Spatial distribution of cytoplasmic domains of the Mg(2+)-transporter MgtE, in a solution lacking Mg(2+), revealed by paramagnetic relaxation enhancement.* Biochim Biophys Acta, 2012. **1824**(10): p. 1129-35.
182. Payandeh, J., R. Pfoh, and E.F. Pai, *The structure and regulation of magnesium selective ion channels.* Biochim Biophys Acta, 2013. **1828**(11): p. 2778-92.
183. Takeda, H., et al., *Structural basis for ion selectivity revealed by high-resolution crystal structure of Mg2+ channel MgtE.* Nat Commun, 2014. **5**: p. 5374.
184. Corral-Rodriguez, M.A., et al., *Nucleotide binding triggers a conformational change of the CBS module of the magnesium transporter CNNM2 from a twisted towards a flat structure.* Biochem J, 2014. **464**(1): p. 23-34.
185. Tomita, A., et al., *ATP-dependent modulation of MgtE in Mg(2+) homeostasis.* Nat Commun, 2017. **8**(1): p. 148.
186. Dann, C.E., 3rd, et al., *Structure and mechanism of a metal-sensing regulatory RNA.* Cell, 2007. **130**(5): p. 878-92.
187. Hohle, T.H. and M.R. O'Brian, *The mntH gene encodes the major Mn(2+) transporter in Bradyrhizobium japonicum and is regulated by manganese via the Fur protein.* Mol Microbiol, 2009. **72**(2): p. 399-409.
188. Hohle, T.H. and M.R. O'Brian, *Magnesium-dependent processes are targets of bacterial manganese toxicity.* Mol Microbiol, 2014. **93**(4): p. 736-47.
189. Lee, W., M. Tonelli, and J.L. Markley, *NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy.* Bioinformatics, 2015. **31**(8): p. 1325-7.
190. Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.* Nat Biotechnol, 2008. **26**(12): p. 1367-72.
191. Ivankov, D.N., et al., *QARIP: a web server for quantitative proteomic analysis of regulated intramembrane proteolysis.* Nucleic Acids Res, 2013. **41**(Web Server issue): p. W459-64.

192. Käll, L., A. Krogh, and E.L.L. Sonnhammer, *A Combined Transmembrane Topology and Signal Peptide Prediction Method.* Journal of Molecular Biology, 2004. **338**(5): p. 1027-1036.
193. Gerwig, J., et al., *The protein tyrosine kinases EpsB and PtkA differentially affect biofilm formation in Bacillus subtilis.* Microbiology, 2014. **160**(Pt 4): p. 682-91.
194. Harwood, C.R. and S.M. Cutting, *Molecular Biological Methods for Bacillus*. 1991: Wiley.
195. Gibson, D.G., *Enzymatic assembly of overlapping DNA fragments.* Methods Enzymol, 2011. **498**: p. 349-61.
196. Lemberg, M.K. and B. Martoglio, *Analysis of polypeptides by sodium dodecyl sulfate–polyacrylamide gel electrophoresis alongside in vitro-generated reference peptides.* Analytical Biochemistry, 2003. **319**(2): p. 327-331.
197. Miroux, B. and J.E. Walker, *Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels.* J Mol Biol, 1996. **260**(3): p. 289-98.
198. Monico, A., et al., *Drawbacks of Dialysis Procedures for Removal of EDTA.* PLoS One, 2017. **12**(1): p. e0169843.
199. Urbani, A. and T. Warne, *A colorimetric determination for glycosidic and bile salt-based detergents: applications in membrane protein research.* Anal Biochem, 2005. **336**(1): p. 117-24.
200. Zeytuni, N. and R. Zarivach, *Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module.* Structure, 2012. **20**(3): p. 397-405.

9 APPENDICES