Charles University in Prague

Faculty of Science

Study programme: Biology (N1501)

Branch of study: NIMUN (1511T004)

Bc. Ján Kuzmík

Inhibitors of rhomboid proteases as tools for cell biology

Inhibitory intramembránových proteas z rodiny rhomboidů jako nástroj buněčné biologie

Diploma thesis

Advisor:

Ing. Kvido Stříšovský, Ph.D.

Prague, 2019

Prohlášení:

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

V Praze, 12. 8. 2019

Ján Kuzmík

Acknowledgement

First and foremost, I want to thank my supervisor Kvido Stříšovský for the opportunity to work in his lab on this exciting scientific project. His advice, support and leadership were immensely valuable and I learned a lot while working in the laboratory.

I am happy to acknowledge and thank my colleagues and friends who helped me with the thesis. I am grateful that Anežka Tichá led me in my experimental work, gave me feedback on my thesis, and taught me a lot about good practice of doing research. Jindra Fanfrlík and Martin Lepšík built the computational model and kindly answered many questions about computational modelling and *in silico* drug design. Stancho Stanchev synthesised all of the substrates and inhibitors and was always ready to help. Radko Souček, always in a good mood, carried out the amino acid analysis of the inhibitors. Kubo Began and other colleagues in the lab helped me troubleshoot my experiments, gave me feedback, and were always encouraging.

Abstract

Rhomboid intramembrane serine proteases cleave polypeptide chains within lipid bilayer. Rhomboid proteases were originally discovered in *Drosophila melanogaster* where they regulate ontogenesis of the fly, but they are present in all domains of life. Nowadays, various diseases, such as malaria, amoebiasis, Parkinson's disease, various tumour malignancies, and diabetes, have been linked with rhomboid proteases. However, natural substrates and function of most rhomboids remain elusive. Cell biology tools are needed for unravelling functions of rhomboids, as well as for potential pharmacological applications, and this together fuels the effort to develop specific rhomboid inhibitors. The inhibitors known to date always bear an electrophilic warhead attacking the nucleophilic serine of the atypical serine-histidine catalytic dyad of rhomboid.

From the various developed inhibitors, peptidyl α-ketoamides substituted at the ketoamide nitrogen by hydrophobic groups, discovered in our laboratory, hold the biggest potential. They are potent, reversible, selective, tunable, and are built around a pharmacophore already approved for medical use. Here, I set out to improve peptidyl α-ketoamides by exploring the chemical space in the active site of rhomboid and testing substituents of the ketoamide nitrogen of increasing size, different structure and chemical nature. Branching of the hydrophobic substituents improved inhibitory potency. Of the inhibitors bearing branched substituents, compound **22** was the most potent one. It inhibited endogenous *E. coli* rhomboid GlpG, meaning that it was able to cross the bacterial outer membrane. To understand the reasons for its improved potency in the absence of a co-crystal structure, I analysed a computational model of the complex. It revealed that the branched substituent of compound **22** extensively occupies the prime side of the rhomboid active site forming numerous interactions with the fifth transmembrane helix and loop L5 of the rhomboid. These regions of rhomboid proteases are hypothesized to participate in substrate binding and cleavage. Since rhomboid protease architecture and mechanism is conserved, these findings give impetus to explore the strategy of using branched substituents of ketoamide nitrogen in designing inhibitors of other rhomboids for studying their biological roles and/or for their pharmacological targeting.

Keywords

Rhomboid protease, intramembrane proteolysis, inhibitor design, peptidyl ketoamide, GlpG

Abstrakt

Rhomboidní proteázy jsou intramembránové serinové proteázy. Původně byly objeveny jako proteiny důležité pro zárodečný vývoj *Drosophily melanogaster*, dnes je však známo, že se nachází napříč všemi skupinami organismů a jsou spojovány s řadou významných onemocnění a patogenů. Příkladem je malárie, úplavice, Parkinsonovo onemocnění, řada typů rakoviny či diabetes. Avšak u řady rhomboidních proteáz dosud neznáme jejich přirozené substráty, a tedy ani jejich funkce není často objasněna na molekulární úrovni. Vývoj jejich specifických inhibitorů je důležitý jednak pro jejich využití jako nástrojů buněčné biologie k objasnění role rhomboidních proteáz, ale také pro jejich potenciální využití ve farmakoterapii. Doposud známé inhibitory rhomboidních proteáz jsou charakteristické elektrofilním centrem, které cílí na nukleofilní serin v rámci neobvyklé serin-histidinové katalytické dyády rhomboidů.

Zatím největší potenciál vykazují inhibitory v podobě peptidyl α-ketoamidů s různými substituenty vázanými na dusík ketoamidové skupiny, které byly dříve vyvinuty v naší laboratoři. Jsou velmi účinné, reverzibilní, selektivní a laditelné. Jejich výhodou je také to, že využívají elektrofilní ketoamidovou skupinu, která je součástí i některých již schválených léčiv. Cílem mé práce bylo pokusit se dále zvýšit účinnost těchto inhibitorů, a to na základě analýzy prostorových vlastností aktivního centra rhomboidu a testováním substituentů vázanými na dusík ketoamidové skupiny inhibitoru o různé velikosti a strukturně-chemických vlastnostech. Zvětšující se velikost hydrofobních substituentů a jejich větvení vedlo ke zvýšení účinnosti inhibitoru. Nejúčinnějším větveným inhibitorem byla sloučenina **22**, která byla zároveň schopna překonat vnější bakteriální membránu a inhibovat endogenní rhomboidní proteázu GlpG v bakterii *E. coli*. Zvýšenou účinnost větvených inhibitorů jsem se pokusil objasnit s pomocí počítačového modelu. Zde jsem zjistil, že větvené inhibitory se silně váží na pátý helix a pátou smyčku rhomboidní proteázy – tedy do oblasti označované jako „prime site“, která se podle některých teorií účastní vazby a štěpení substrátů. Význam této studie se dotýká nejen bakteriálních rhomboidních proteáz, ale v důsledku konzervovanosti základní proteinové struktury a mechanismu štěpení je princip větvených inhibitorů potenciálně aplikovatelný i na další rhomboidy za cílem studia jejich funkce či farmakologické intervence.

Klíčová slova

Intramembránová proteolýza, rhomboid, návrh inhibitorů, peptidyl ketoamide, GlpG

List of abbreviations

A. fumigatus	*Aspergillus fumigatus*
A. thaliana	*Arabidopsis thaliana*
AarA	rhomboid protease AarA
Ac	acetyl group
AMA1	apical membrane antigen 1
APP	amyloid precursor protein
APS	ammonium persulfate
BIK1	*Botrytis*–induced kinase 1
CONH	ketoamide
CLEC14A	C-type lectin domain family 14 member
CLPB	caseinolytic peptidase B protein homolog
CTRP	complement component 1q/tumor necrosis factor related protein
D. melanogaster	*Drosophila melanogaster*
DDM	β-D-maltopyranoside
DDR1	discoidin domain receptor family member 1
DMSO	dimethyl sulfoxide
E. coli	*Escherichia coli*
EBL	erythrocyte binding ligand
EDTA	ethylenediaminetetraacetic acid
EGF	epidermal growth factor
EGFR	epidermal growth factor receptor
ERAD	endoplasmic reticulum associated degradation
FLAG	protein tag; sequence DYKDDDDK
FRET	Förster resonance energy transfer
GlpG	rhomboid protease GlpG
Golgi	Golgi apparatus
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HFIP	1,1,1,3,3,3-Hexafluoro-2-propanol HPLC-MS
IC_{50}	half maximal inhibitory concentration
IgG	immunoglobulin G
IL6R	interleukin 6 receptor
IPTG	isopropyl β-D-1-thiogalactopyranoside
iRhom	inactive rhomboid protein
K_i	inhibitory constant
KO	knock-out
KSp35	fluorogenic substrate KSp35
KSp96	fluorogenic substrate KSp96
λ_{ex}	excitation wavelength
λ_{em}	emission wavelength
LacY	lactose permease
LacYTM2	second transmembrane domain of LacY
LB	lysogeny broth

LPS	lipopolysaccharide
LptD	LPS-assembly protein LptD
LUCA	last universal common ancestor
MAEBL	merozoite adhesive erythrocytic binding protein
NMR	nuclear magnetic resonance
OD_{600}	optical density at 600 nm
P. stuartii	*Providencia stuartii*
PARL	PINK1/PGAM5-associated rhomboid like
PBS	phosphate-buffered saline
PD	Parkinson's disease
PDB	protein data bank
PEG8000	Polyethylene glycol 8000
PGAM5	serine/threonine-protein phosphatase PGAM5
PINK1	serine/threonine-protein kinase PINK1
***Plasmodium* spp.**	*Plasmodium* species
pTCRα	α chain of the pre-T cell receptor
PMSF	phenylmethylsulfonyl fluoride
QM	quantum mechanical method
R^2	coefficient of determination
RBL	reticulocyte binding-like protein
RHBDD1-3	rhomboid domain containing 1-3
RHBDL1-4	rhomboid-related protein 1-4
RMSD	root-mean square deviation
ROM	rhomboid
RT	room temperature
SD	standard deviation
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SQM	semiempirical QM method
SPINT1	Kunitz-type protease inhibitor 1
SREBP	sterol regulatory element binding protein
STARD7	StAR-related lipid transfer domain protein 7
T. vaginalis	*Trichomonas vaginalis*
TatA	Sec-independent protein translocase protein
TFE	2,2,2-Trifluoroethanol
TGFα	transforming growth factor α
TMD	transmembrane domain
TMEM115	transmembrane protein 115
TRAP	thyroid hormone receptor-associated protein
TSAP6	endosomal metalloreductase (also STEAP3)
TTC19	tetratricopeptide repeat protein 19
TVAG_166850	putative adhesion of Trichomonas vaginalis
UBAC2	ubiquitin-associated domain-containing protein 2
V. cholera	Vibrio cholerae
VesB	GlyGly-anchored extracellular serine protease VesB

v_i	initial velocity
WT	wild type
YqgP	rhomboid protease YqgP

Table of contents

1. Introduction

Rhomboid proteins (Figure 1) were discovered in a screen for genes affecting the early stages of development of the fruit fly *Drosophila melanogaster*. Mutant fly larvae lacking *rhomboid-1* gene displayed rhomboid-like shape of head skeleton (Mayer and Nüsslein-Volhard, 1988), giving the gene its name. Rhomboid-1 has been later shown to regulate epidermal growth factor (EGF) signalling by cleaving the EGF receptor (EGFR) ligand Spitz (Urban et al., 2001). Bioinformatics analyses revealed that rhomboid-like proteins are widespread and conserved across all domains of life (Koonin et al., 2003; Lemberg and Freeman, 2007). Moreover, since their discovery, rhomboids have been shown to play distinct roles across biology and are nowadays medically relevant proteins. Rhomboids are serine intramembrane proteases (Figure 1). Their active site consists of serine and histidine both forming an unconventional catalytic dyad. It was initially predicted to be buried inside the membrane (Urban et al., 2001) which was subsequently confirmed by a crystal structure (Wang et al., 2006).

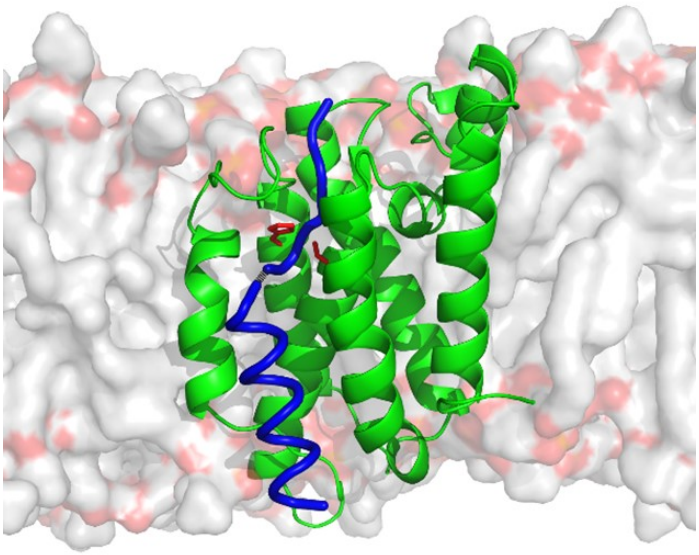

Figure 1. Rhomboid – an intramembrane serine protease. Model of a complex of rhomboid GlpG with its bound substrate based on co-crystal structures, molecular dynamics and modelling. Enzyme is shown in green, catalytic dyad of serine and histidine in red, substrate in blue and membrane lipids in grey. Adopted from (Strisovsky, 2016a).

This thesis focuses on development of rhomboid inhibitors as tools for cell biology. In the literature overview, the evolution of rhomboid-like proteins is briefly discussed (1.1.) followed by a summary of their transmembrane topology (1.1.1.) and biological roles of rhomboid proteases across organisms (1.2.). The focus is on medically relevant rhomboid proteases, their substrates and roles in physiology and pathophysiology. Next, the catalytic mechanism and structure of intramembrane rhomboid proteases are introduced (1.3.). Lastly, rhomboid inhibitors are summarized (1.4.). The focus

of the Results section is on further development of peptidyl α-ketoamides as potent, specific and highly selective rhomboid protease inhibitors.

1.1. Rhomboid superfamily and evolution of rhomboid-like proteins

Rhomboids are present in all domains of life – bacteria, archaea, and eukaryotes (Koonin et al., 2003). It is hypothesised that rhomboid proteases have not diverged from soluble serine proteases, but instead evolved from an ancestral, non-catalytic polytopic rhomboid-like protein that could recognize transmembrane domains (TMDs) of other proteins (Adrain and Freeman, 2012). In the proposed evolution model, the ancient non-catalytic rhomboid-like protein acquired serine protease catalytic residues by converged evolution before formation of the last universal common ancestor (LUCA) (Figure 2) (Adrain and Freeman, 2012). Catalytically inactive rhomboid pseudoproteases have been identified in eukaryotes, but, to some surprise, not in prokaryotes (Koonin et al., 2003). Rhomboid pseudoproteases, together with their active counterparts, form the superfamily of rhomboid-like proteins.

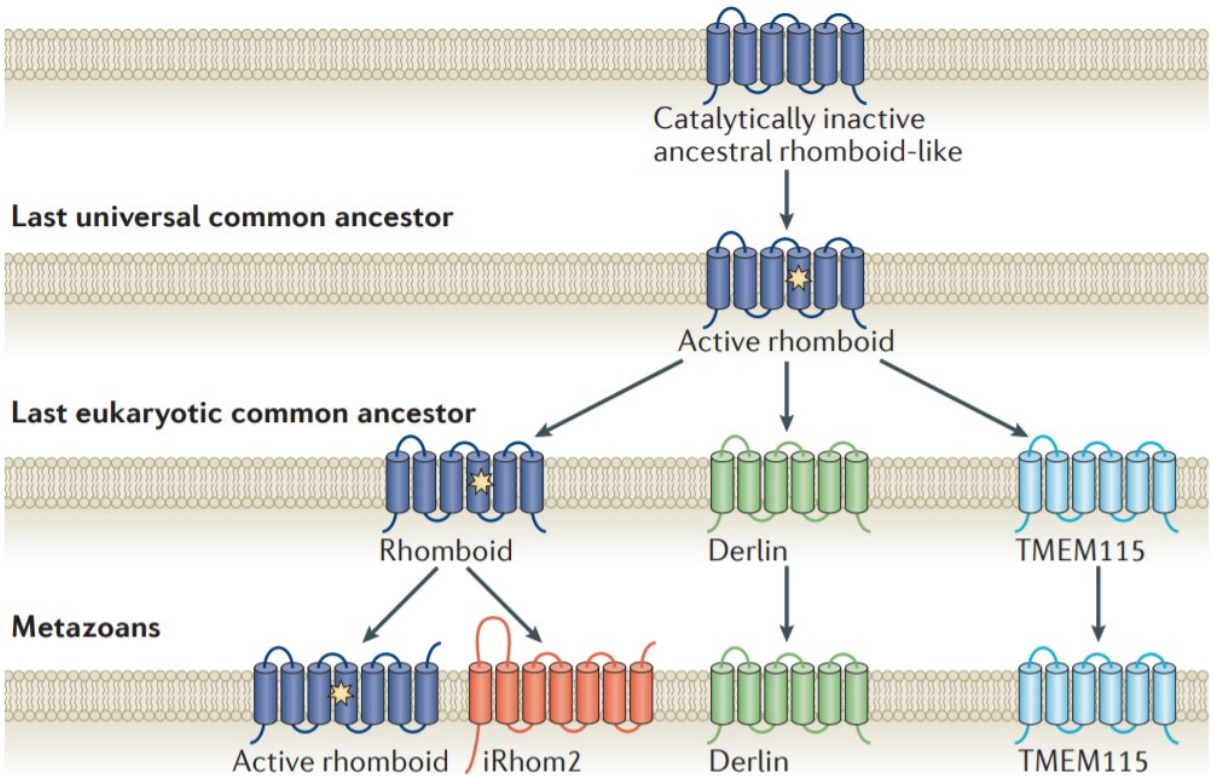

Figure 2. Evolution of the superfamily of rhomboid-like proteins. Proposed model of convergent evolution of rhomboids as serine proteases suggests that rhomboid protease evolved from catalytically inactive ancestor before the formation of last universal common ancestor (LUCA). Evolutionary later, in eukaryotes, two families of pseudoproteases, Derlins and TMEM115, have diverged. Finally, iRhoms have appeared in metazoans. The star represents an active site of rhomboid protease. Adopted from (Adrain and Freeman, 2012).

Rhomboid pseudoproteases, represented by iRhoms, Derlins, UBAC2, RHBDD2, RHBDD3, and TMEM15 proteins, play important roles in protein trafficking regulation, protein degradation, and inflammatory signalling (reviewed in Lemberg and Adrain, 2016). Furthermore, rhomboid pseudoproteases are being associated with many diseases such as cancer, skin diseases, inflammatory diseases, infectious diseases and others (reviewed in Dulloo et al., 2019). The many roles of rhomboid pseudoproteases demonstrate their importance in cell physiology and show that the rhomboid fold can successfully harbour important functions other than proteolysis.

The evolution model (Figure 2)(Adrain and Freeman, 2012) proposes that pseudoproteases have disappeared before the formation of last universal common ancestor (LUCA) and later reappeared in eukaryotes. The presence of rhomboid protease in LUCA and formation of pseudoproteases by divergent evolution is supported by a study based on network based clustering (Kinch and Grishin, 2013). However, this study propose a rhomboid protease duplication in LUCA (Kinch and Grishin, 2013), which, together with proposed counterintuitive disappearance and reappearance of rhomboid pseudoproteases in evolution, leads to an ongoing discussion about evolutionary path of rhomboid-like proteins (Tichá et al., 2018). The main obstacle preventing clear reconstruction of the evolution of rhomboid-like proteins is the low sequence identity of only about 10 percent among the proteins (Koonin et al., 2003).

1.1.1. Transmembrane topology of rhomboid proteases

There are three basic transmembrane topologies of rhomboid proteases (Figure 3). The protein core consists of 6 transmembrane helical domains (TMD1–6) connected by 5 loops L1–L5 (Lemberg and Freeman, 2007). Typically, bacterial and archaeal rhomboids have the basic core topology (termed secretase-type A), whereas eukaryotic rhomboids tend to have an extra loop and TMD either on its C-terminus (termed secretase-type B), or N-terminus (termed PARL-type) (Koonin et al., 2003; Lemberg and Freeman, 2007). The catalytic serine is a part of an evolutionarily conserved GxSx motif on TMD4 (secretase-type A and B), or TMD5 (PARL-type) and catalytic histidine is a part of TMD6 (secretase-type A and B), or TMD7 (PARL-type). The L1 loop contains conserved tryptophan and arginine, is partially immersed in the membrane (Wang et al., 2006), important for enzyme stability (Baker and Urban, 2012; Bondar et al., 2009), and participates in substrate binding (Zoll et al., 2014). The structure of rhomboid proteases is described in section 1.3.2.

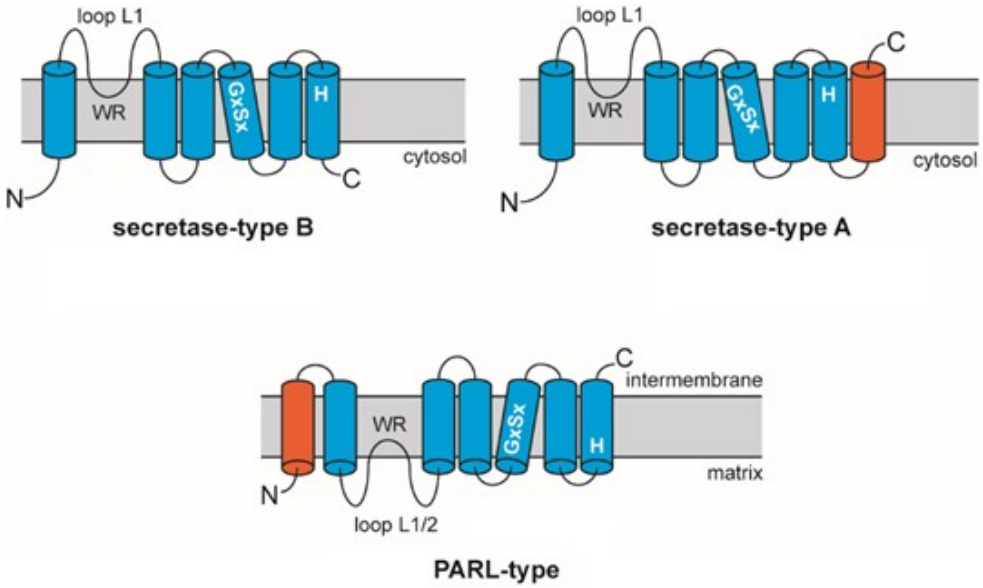

Figure 3. Rhomboid protease topology. Three predicted transmembrane topologies of rhomboids are depicted, with the conserved core, consisting of 6 transmembrane helical domains (TMDs) highlighted in blue, and additional TMDs highlighted in red. Loop L1 is partly protruding into the membrane and contains evolutionarily conserved tryptophan and arginine. Catalytic serine is a part of GxSx motif, which is a part of TMD4 (secretase-type A and B), or TMD5 (PARL-type). Catalytic histidine is a part of TMD6 (secretase-type A and B), or TMD7 (PARL-type). The orientation into cytosol (secretase-type A and B), or into mitochondria matrix (PARL-type) is denoted. Adopted and adjusted from (Tichá et al., 2018).

1.2. Biology of rhomboid proteases

Prokaryotes generally encode at least one rhomboid protease in their genome and rhomboid pseudoproteases are generally absent (Koonin et al., 2003). Rhomboid superfamily members have expanded in eukaryotic genomes, probably by gene duplications (Lemberg and Freeman, 2007; Li et al., 2015a), with as many as 14 genes in mammals (Bergbold and Lemberg, 2013) and 25 rhomboid-like genes in soybean (Li et al., 2015a). As mentioned in section 1.1., some of these genes represent inactive rhomboid pseudoproteases, but, since they are not in the spotlight of this thesis, they will not be discussed further.

1.2.1. Rhomboid proteases in microbes

Even though studying **bacterial** rhomboids contributed substantially to biophysical understanding of how intramembrane proteolysis works (1.3.); knowledge of the biological functions of rhomboid proteases in bacteria is limited. A good illustration is ***Escherichia coli*** rhomboid GlpG. *E. coli* is the best studied bacterium and GlpG is the best mechanistically and structurally characterized rhomboid (1.3.2.). GlpG has been shown to mediate gut colonization (Russell et al., 2017) and *glpG* null

allele mutant exhibits slight increase in resistance to antibiotic cefotaxime (Clemmer et al., 2006), but the physiological substrates and functions of GlpG remain unknown.

On the other hand, rhomboid AarA in ***Providencia stuartii*** has been shown to cleave TatA, a component of the twin arginine translocase system activating quorum sensing (Stevenson et al., 2007). Rhomboid mutants in ***Bacillus subtilis*** exhibit decrease in glucose uptake and cell division defects (Mesak et al., 2004); and rhomboid mutants in ***Mycobacterium tuberculosis*** exhibit differences in colony morphology, reduced biofilm formation, and higher susceptibility to antibiotics (Kateete et al., 2012). However, the physiological substrates of rhomboids in both bacteria remain to be discovered.

Cholera, caused by ***Vibrio cholerae***, kills around 91 000 people every year, with the biggest burden in Sub-Saharan Africa (Ali et al., 2015). A recent study shows that *V. cholerae* rhombosortase, a member of a subfamily of rhomboids, cleaves VesB protease in its GlyGly-CTERM domain, facilitating transport of VesB to the outer cell membrane (Gadwal et al., 2018). VesB is able to activate cholera toxin by cleaving its A subunit (Sikora et al., 2011), pointing to rhombosortase's role in virulence of this pathogen. Rhombosortases and GlyGly-CTERM domain containing proteins are found together in many bacteria (Haft and Varghese, 2011), suggesting more substrates and functions of rhombosortases to be discovered (Gadwal et al., 2018).

Unsurprisingly, since **archaea** are difficult to study, not much is known about the role of rhomboid proteases in this domain of life. A rhomboid knock-out mutant in ***Haloferax volcanii*** has mild defects in motility and sensitivity to an antibiotic novobiocin (Parente et al., 2014). Furthermore, a shorter form of N-glycosylation of the S-layer protein has been observed in the rhomboid mutant, but the physiological substrate of the rhomboid is unknown (Parente et al., 2014). An additional proteomic study identified candidate substrates, but these remain to be confirmed (Costa et al., 2018).

Aspergillosis, caused by a **fungus *Aspergillus fumigatus***, infects mostly immunocompromised individuals, and, despite the lack of a thorough epidemiologic study, is estimated to cause about 600 000 deaths annually (Gsaller et al., 2016). *A. fumigatus* encodes two rhomboids and both of them are essential for the fungus growth in pathogen-induced hypoxia and virulence in murine models (Dhingra et al., 2016; Vaknin et al., 2016). Furthermore, one of the rhomboid deletion mutant has increased sensitivity to phagocytic killing by immune system (Vaknin et al., 2016). Both of these phenotypes are linked to processing of the sterol regulatory element binding protein (SREBP) transcription factors, but experimental validation of a rhomboid substrate remains to be carried out. Notably, a rhomboid protease homolog in another fungus, ***Schizosaccharomyces pombe***, have been shown to cleave SREBP (Hwang et al., 2016), supporting the hypothesis of SREBP processing by *A. fumigatus* rhomboids.

1.2.2. Rhomboid proteases in protists

Protist parasites encode rhomboid proteases with proven medical relevance (see below). Malaria, caused by ***Plasmodium* species** (spp.), is responsible for around 435 000 deaths annually (World Health Organization, 2018); amoebiasis, caused by ***Entamoeba histolytica***, is estimated to kill over 100 000 people annually (Kantor et al., 2018); trichomoniasis, caused by ***Trichomonas vaginalis***, has an incidence of around 141 000 (GBD 2015 Disease and Injury Incidence and Prevalence Collaborators, 2016); and toxoplasmosis, caused by ***Toxoplasma gondii***, even though usually asymptomatic, infects 30–50 % of world population (Flegr et al., 2014).

The general pattern of protist parasite rhomboid proteases is the cleavage of adhesin molecules (reviewed in Düsterhöft et al., 2017). Adhesins are surface molecules and are vital for the initial contact of the parasite and host cells, forming an interface between the two cells. However, adhesins often need to be cleaved in the later steps of the invasion/phagocytosis, such as in the example of *Plasmodium* spp. invasion of a red blood cell (Figure 4) (Urban, 2009).

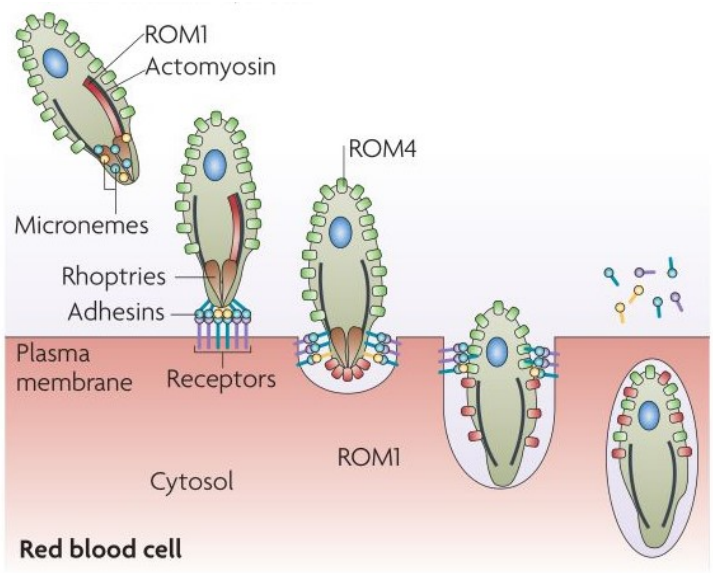

Figure 4. Invasion of *Plasmodium* spp. into a red blood cell. *Plasmodium* spp. organelles micronemes and rhoptries secrete adhesins which bind to the receptors on erythrocyte surface. As the invasion progresses, the adhesins are gradually cleaved by *Plasmodium* spp. rhomboids ROM1 (red) and ROM4 (green). Adopted from (Urban, 2009).

The **apicomplexan** parasite ***Toxoplasma gondii*** expresses 6 rhomboid proteases, TgROM1-6 (Dowse and Soldati, 2005). Two of them, TgROM4, and TgROM5, have been demonstrated to cleave adhesins AMA1 and MICs (Brossier et al., 2005; Buguliskis et al., 2010; Rugarabamu et al., 2015). Shedding of adhesins facilitates host cell invasion, therefore *T. gondii* rhomboids play role in virulence of this parasite (Brossier et al., 2003; Parussini et al., 2012).

***Plasmodium* spp.**, also a member of **apicomplexa** phylum, encode 8 rhomboids (Dowse and Soldati, 2005). They have been associated with parasite's motility, cell growth and host cell invasion. *Plasmodium falciparum* rhomboids PfROM1 and PfROM4 can collectively cleave adhesins AMA1, EBL, RBL, TRAP, CTRP, and MAEBL (Baker et al., 2006; O'Donnell et al., 2006). Adhesin cleavage is important for invasion of *Plasmodium* spp. into host cell (Figure 4). PfROM4 deletions mutants are not viable, pointing further to the essentiality of this protein for the parasite virulence (Lin et al., 2013; O'Donnell et al., 2006).

The **amoebozoa** parasite ***Entamoeba histolytica*** encodes only one rhomboid, EhROM1 (Baxt et al., 2008). EhROM1 has been shown to cleave adhesin *N*-acetyl D-galactosamine-inhibitable lectin (Baxt et al., 2008), and EhROM1 knock-down shows reduced cytotoxicity, hemolytic activity, and altered migration (Rastew et al., 2015). Furthermore, EhROM1 co-localize into vesicles after phagocytosis of erythrocytes, pointing further to EhROM1's role in virulence of this pathogen (Baxt et al., 2008).

The **excavate** parasite ***Trichomonas vaginalis*** encodes 4 rhomboids, TvROM1-4 (Riestra et al., 2015). Two putative adhesins have been demonstrated to be cleaved by rhomboid TvROM1 (Riestra et al., 2015). Furthermore, accumulation of one of the substrates, TVAG_166850, leads to a bigger adherence of parasite to host cells (Riestra et al., 2015). Following the same pattern, over-expression of TvROM1 increased cell cytolysis by parasite (Riestra et al., 2015). Together, these data indicate that rhomboid protease plays important role in the virulence of *T. vaginalis*.

1.2.3. Rhomboid proteases in plants

Plant genomes encode the most rhomboid-like genes, with as many as 25 in soybean (Li et al., 2015a), but, despite their number, very little is known about the role of rhomboids in this domain of life. The best studied plant rhomboids are the ones of ***Arabidopsis thaliana***. *A. thaliana* encodes 13 rhomboid proteases (Tripathi and Sowdhamini, 2006). Their intracellular localizations have been demonstrated and/or predicted (Adamiec et al., 2017). Handful of phenotypes of rhomboid mutants have been observed, but the first natural substrate of a plant rhomboid remains to be identified (reviewed in Adamiec et al., 2017).

1.2.4. Rhomboid proteases in animals

As already mentioned, rhomboids were originally discovered in *Drosophila melanogaster* (Mayer and Nüsslein-Volhard, 1988; Urban et al., 2001). Even though the fly rhomboid proteases are heavily studied, they are not directly medically relevant, and as such are not discussed here in detail (1.2.4.1.).

Some of the human rhomboid proteases are associated with various diseases. These are Parkinson's disease, Alzheimer's disease, various tumour malignancies, and type 2 diabetes. However, none of the connections are clearly established and/or the underlying molecular principles are unclear. Furthermore, there is no unified theme in rhomboid pathophysiology in humans, therefore individual links between human rhomboids and associated pathologies are described on case by case basis in the text below (1.2.4.2.).

1.2.4.1. Rhomboid proteases in *Drosophila melanogaster*

The fruit fly ***Drosophila melanogaster*** encodes 6 rhomboid proteases (Lemberg and Freeman, 2007), and at least 3 of them play pivotal role in embryonic and post-embryonic development of the fly (reviewed in Shilo, 2016). They do so by regulating EGFR signalling. Proteins Spitz, Gurken, Keren, and Star are cleaved by *D. melanogaster* rhomboid proteases (reviewed in Lastun et al., 2016). Spitz, Gurken and Keren are direct ligands of EGFR whereas Star cleavage regulates the amount of mature Spitz (Tsruya et al., 2007).

1.2.4.2. Rhomboid proteases in mammals

Mammals encode 5 catalytically active rhomboid proteases (Table 1) (Koonin et al., 2003), termed RHBDL1-4 and PARL (PINK1/PGAM5-associated rhomboid like). Nothing is known about biological function of **RHBDL1**. Similarly, knowledge about biology of **RHBDL3** is also very scarce, since no natural substrates of RHBDL1 or RHBDL3 have been reported so far. A study focused on the chronological aging of the human brain identified RHBDL3 expression to be the most consistently associated with the brain chronological age (Kumar et al., 2013), but the role of RHBDL3 in brain ageing remains to be uncovered.

Rhomboid protease **RHBDL2** is known to cleave many proteins, namely thrombomodulin, EphrinB2/B3, EGFR, EGF, CLEC14A, IL6R, SPINT1, DDR1, N-cadherin and others. These were identified by proteomics (Johnson et al., 2017) and candidate screens (reviewed in Lastun et al., 2016). Despite the plethora of substrates, biological role of RHBDL2 is less clear. A general pattern seems to be epithelial homeostasis (Johnson et al., 2017). Thrombomodulin cleavage promotes wound healing (Cheng et al., 2011), cleavage of CLEC14A modulates angiogenesis (Noy et al., 2016), and many of the newly identified substrates are expressed and play a role in epithelia (Johnson et al., 2017).

RHBDL2 is linked with tumour malignancies, but its role remains to be further validated. Current evidence consists of the correlation of high RHBDL2 expression with shorter survival of patients with pancreatic ductal adenocarcinoma (Khalid et al., 2019), over-expression of RHBDL2 inducing cell proliferation and resistance to anoikis, which is a type of apoptosis (Cheng et al., 2014), and the already mentioned role of RHBDL2 in angiogenesis (Noy et al., 2016).

Rhomboid protease **RHBDL4** (also known as RHBDD1 - rhomboid domain containing 1), similarly to RHBDL2, is connected with several identified substrates, namely APP, pTCRα, TSAP6, TGFα, BIK1, and others (Fleig et al., 2012; Song et al., 2015; Wan et al., 2012; Wang et al., 2008; Wunderle et al., 2016). RHDL4 plays role in endoplasmic reticulum associated degradation (ERAD), demonstrated on α chain of the pre-T cell receptor (pTCRα) (Fleig et al., 2012). Additionally, cleavage of TSAP6 negatively regulates exosome secretion (Wan et al., 2012). Last but not least, RHBDL4 is able to induce apoptosis by the cleavage of BIK1 (Wang et al., 2008), and various cancer studies further indicate that RHBDL4 plays role in apoptosis and cell growth (Liu et al., 2013; Miao et al., 2017; Song et al., 2015; Wei et al., 2014; Zhang et al., 2018a, 2018b).

Various links between RHBDL4 and different types of tumour malignancies have been made. RHBDL4 is upregulated and correlates with poor patient prognosis in breast cancer (Zhang et al., 2018b) and colorectal cancer (Song et al., 2015). An EGFR substrate TGFα was initially thought to be the substrate of RHBDL4 behind the phenotype in colorectal cancer (Song et al., 2015). However, later study was unable to reproduce the cleavage experiment from Song *et al.* and instead demonstrated secretion of uncleaved TGFα in exosomes as the cause of the phenotype (Wunderle et al., 2016). RHBDL4 depletion leads to apoptosis and reduced cell growth in cell cancer lines derived from colorectal cancer (Miao et al., 2017; Zhang et al., 2018a), breast cancer (Zhang et al., 2018b), glioblastoma (Wei et al., 2014) and hepatocellular carcinoma (Liu et al., 2013), but the precise molecular mechanism is unclear. Last but not least, RHBDL4 can cleave amyloid precursor protein (APP), but the proposed relevance of this for Alzheimer's disease remains hypothetical (Paschkowsky et al., 2016).

Mitochondrial rhomboid protease **PARL** cleaves proteins PINK1, PGAM5, Smac, TTC19, CLPB and STARD7 (Deas et al., 2011; Saita et al., 2017, 2018; Sekine et al., 2012). PARL cleavage of PINK1 is implicated in adipogenesis (Shiau et al., 2017) and cleavage of Smac promotes apoptosis (Saita et al., 2017). Furthermore, PARL regulates various processes of mitochondria homeostasis, namely cleavage of PGAM5 and/or PINK1 regulates mitophagy (Meissner et al., 2011; Yan et al., 2019), and/or activates mitochondria biogenesis (Bernkopf et al., 2018), and cleavage of STARD7 regulates mitochondria membrane lipids homeostasis (Saita et al., 2018).

PARL and its role in mitophagy has been associated with Parkinson's disease (PD). Two PD patients were shown to carry PARL mutation (Shi et al., 2011) and a mutation in PINK1, a substrate of PARL, causes familial PD (Trempe and Fon, 2013). Furthermore, PARL has also been associated with the development of type 2 diabetes (reviewed in Düsterhöft et al., 2017; Spinazzi and De Strooper, 2016) and a role of mitophagy in cancer is slowly being unravelled (reviewed in Vara-Perez et al., 2019).

However, since the current evidence is insufficient, clearer links between PARL and human diseases remain to be drawn (reviewed in Düsterhöft et al., 2017; Spinazzi and De Strooper, 2016).

Table 1: Mammalian rhomboid proteases. Their cellular localization, known or predicted substrates and biological function are summarized. Partly adopted from (Lastun et al., 2016). ERAD – endoplasmic reticulum associated degradation

Rhomboid	Cellular localization	Substrates	Function
RHBDL1	Golgi	unknown	unknown
RHBDL2	plasma membrane	thrombomodulin; CLEC14A; and many others (see the main text)	epithelial homeostasis; wound healing; angiogenesis
RHBDL3	plasma membrane; endosomes; Golgi	unknown	unknown
RHBDL4 (RHBDD1)	endoplasmic reticulum	APP; pTCRα; TSAP6; TGFα; BIK1, and others	ERAD; exosome secretion; cell growth and apoptosis
PARL	inner mitochondrial membrane	PINK1; PGAM5; Smac; TTC19; CLPB and STARD7	mitochondria homeostasis; adipogenesis; apoptosis

1.3. Substrate cleavage and structure of rhomboid proteases

1.3.1. Catalytic mechanism

Rhomboids are intramembrane serine proteases cleaving transmembrane proteins (Urban et al., 2001). Rhomboid proteases have a unique catalytic mechanism, employing the catalytic dyad of serine and histidine, which is distinct from classical serine proteases employing the catalytic triad of serine, histidine, and aspartate (Wang et al., 2006). Catalytic histidine accepts proton from serine hydroxyl group, turning it into nucleophilic oxyanion. Serine's nucleophilicity is often exploited in the development of covalent inhibitors of serine proteases, as is exemplified on rhomboids in 1.4. It has been proposed that, whereas serine in triad is already deprotonated in the ground state of an enzyme, the serine in dyad is being deprotonated simultaneously with ligand docking within the active site of rhomboid protease (Uritsky et al., 2016).

The rhomboid catalytic cycle is depicted in Figure 5. The proton abstraction from serine hydroxyl group by histidine imidazole ring prompts the serine oxyanion for nucleophilic attack of the carbonyl carbon of a peptide bond, resulting in the formation of the first tetrahedral intermediate. The first polypeptide product is cleaved and released. In the next step, a molecule of water hydrolyses the acyl-enzyme complex, resulting in the formation of the second tetrahedral intermediate, release of the second polypeptide product, and restoration of the active site of rhomboid for the next cycle of catalysis (Strisovsky, 2016b). Notably, the active site of a rhomboid protease is buried in the hydrophobic environment of a lipid bilayer, making water accessibility for proteolysis an interesting

topic. The so-called water retention site of rhomboid is hypothesised to provide water acess (further discussed in 1.3.2) (Zhou et al., 2012).

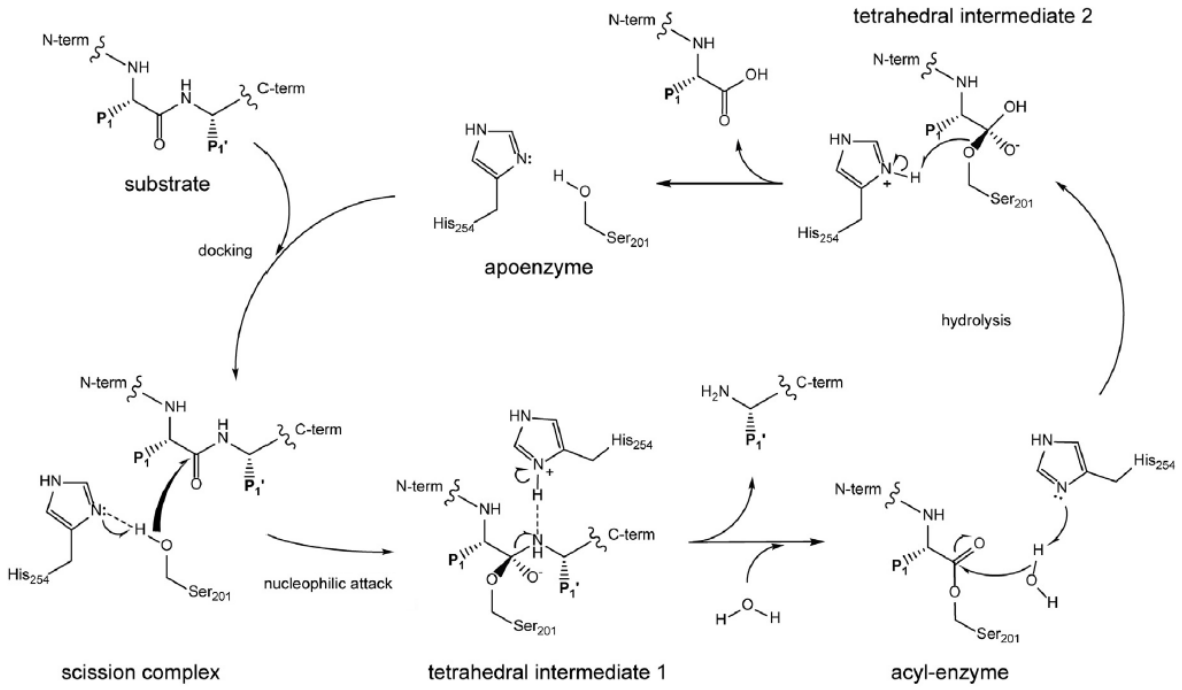

Figure 5. Rhomboid protease reaction mechanism. Rhomboids adopt the classical acid-base mechanism of reaction with a polypeptide substrate being cleaved by the catalytic dyad of serine and histidine. Initially, enzyme-substrate scission complex is formed after docking of a substrate into the active site of an apoenzyme. Catalytic histidine absorbs proton from catalytic serine which is followed by a nucleophilic attack on the carbonyl atom of a peptide bond forming the first tetrahedral intermediate. The first (C-term) polypeptide product is released and acyl-enzyme complex is formed. Nucleophilic addition of a water molecule leads to the formation of the second tetrahedral intermediate, hydrolysis, release of the second (N-term) polypeptide, and regeneration of the catalytic dyad. Please note that the transitory tetrahedral intermediates contain oxyanions. Adopted from (Strisovsky, 2016b).

1.3.2. Structure of rhomboid proteases and their interaction with substrates

The only structurally characterized rhomboid protease is GlpG of *Escherichia coli* (Wang et al., 2006; Ben-Shem et al., 2007; Vinothkumar et al., 2010; Vinothkumar, 2011) and of *Haemophilus influenzae* (Lemieux et al., 2007). However, despite the low sequence identity of rhomboid homologs (Koonin et al., 2003), it is probable that rhomboids share many structural features due to the same or very similar topology (1.1.1.). Thus, it is not naïve to suppose that structural features defined for GlpG are shared among other rhomboid proteases (Lemberg and Freeman, 2007).

The active site of GlpG is formed inside the enzyme by surrounding structures (Figure 6) (Wang et al., 2006). TMD4, containing the catalytic Ser201, is encircled by the remaining five TMDs and a

membrane-immersed helical hairpin of L1 loop. The active site is about 10 Å below the surface of the membrane and is accessible to aqueous solvent. A molecule of water is required for the completion of one round of catalysis (1.3.1.). A water retention site, primarily formed by the side chains of polar amino acids His141, Ser181, Ser185, and Gln189, is in a close proximity of the active site, and is predicted to provide water molecules for catalysis (Zhou et al., 2012). In all crystal structures it retains three molecules of water (Wang et al., 2006; Ben-Shem et al., 2007; Lemieux et al., 2007; Vinothkumar et al., 2010; Vinothkumar, 2011; Zhou et al., 2012), and a relay of side chains of polar amino acids connects the water retention site with bulk solvent, possibly facilitating water molecule transfer into the water retention site and active site (Zhou et al., 2012). During substrate cleavage, the charged oxyanion of the first tetrahedral intermediate is stabilized via three hydrogen bonds formed by His150, Asn154 and the main-chain amide of Ser201 in a structure named 'oxyanion hole' (Xue and Ha, 2012). Stabilizing the oxyanion contributes to the catalysis, as can be seen in the limited proteolytic activity of His150Ala and Asn154Ala mutants (Baker et al., 2007).

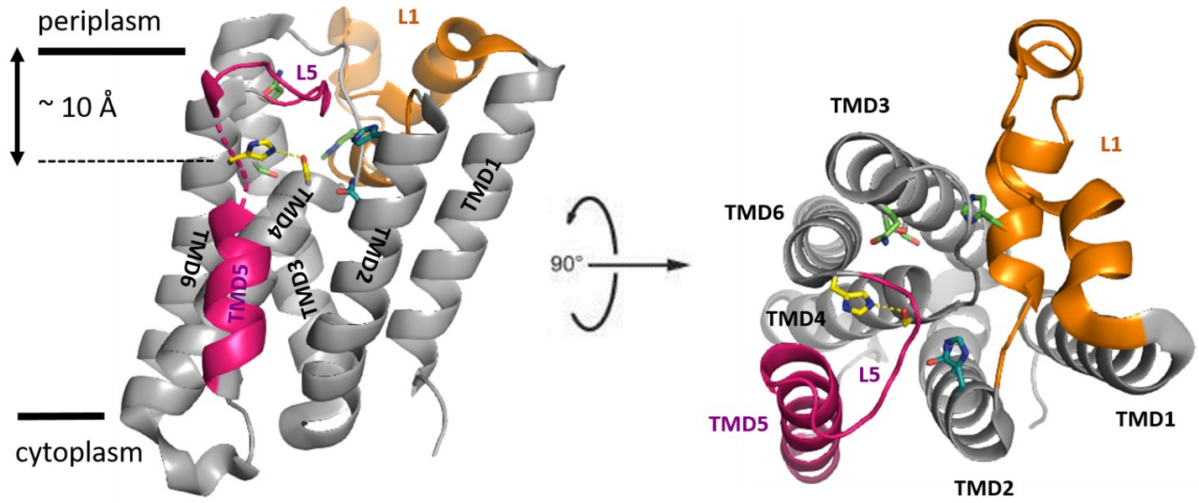

Figure 6. GlpG structure. The side (left) and top (right) view of the bacterial rhomboid GlpG consisting of 6 transmembrane domains (TMD1-TMD6) and 5 connecting loops L1-L5 (PDB code 2XOV) (Vinothkumar et al., 2010). The active site of rhomboid – catalytic dyad of Ser201 and His254 (both yellow) is buried about 10 Å inside the membrane. Catalytic dyad forms a hydrogen bond (yellow dashed line) in the ground state of enzyme. L1 loop (orange) forms a helical hairpin that is immersed in membrane. L5 loop (magenta) forms a cap on top of the active site and together with TMD5 (also magenta) are hypothesized to be mobile elements whose mobility participates in substrate binding. His141, Ser181, Ser185, and Gln189 (all green) form water retention site, providing water for proteolysis. His150 and Asn154 (both cyan) form 'oxyanion hole' during substrate binding. Enzyme is shown in grey cartoon representation and a part of the TMD5 on the side view (left) is for clarity shown as a dashed line.

No structure of a rhomboid-substrate complex is available. It has been proposed that GlpG recognises two elements in its substrates (Figure 7) (Strisovsky et al., 2009) and, accordingly, the

catalysis runs in a spatially and temporally distinct two-phase manner (Cho et al., 2016; Strisovsky, 2016b; Shokhen and Albeck, 2017). The proposed initial step is the binding of a substrate's transmembrane helix to an exosite formed by TMD2 and TMD5 (Strisovsky et al., 2009; Dickey et al., 2013; Shokhen and Albeck, 2017). TMD5 and its attached loop L5 are thought to be flexible elements of GlpG, whose displacement participates in substrate entry into the enzyme. TMD5 is argued to function as a so-called lateral gate for a substrate entry into the active site (Baker et al., 2007; Xue and Ha, 2013; Shokhen and Albeck, 2017), whereas loop L5 forms a cap on top of the active site of rhomboid and is displaced during ligand binding (Wang and Ha, 2007; Vinothkumar et al., 2010; Xue and Ha, 2012; Xue et al., 2012). Once the substrate is docked into the exosite, the next step is partial unwinding of the canonical alpha-helical conformation and binding of approximately 6 amino acids-long recognition motif (Figure 8) into the active site of GlpG (Figure 7) (Strisovsky et al., 2009; Zoll et al., 2014; Brown et al., 2018).

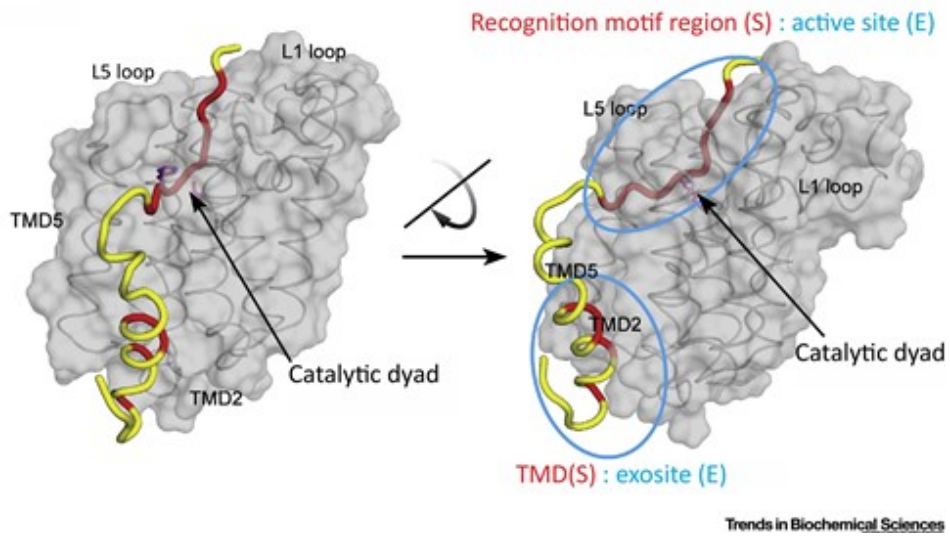

Figure 7. Scission complex of GlpG and its substrate. Computational model of scission complex of GlpG and its artificial substrate Gurken generated by molecular dynamics (Shokhen and Albeck, 2017). The predicted hotspots of interaction (highlighted in red) have the highest partial energy contributions to the stability of the substrate-enzyme complex. Two distinct GlpG sites (highlighted by blue ellipses) recognise two distinct substrate elements. Substrate TMD, in helical conformation, binds into the exosite formed by TMD2 and TMD5. The recognition motif, comprising of about 6 amino acids, binds into the active site. Note that substrate recognition and binding into the two distinct sites of GlpG is thought to be temporally separated (see the main text). Adopted from (Tichá et al., 2018).

GlpG prefers alanine in the P_1 position of the substrate recognition motif (Figure 8), positively charged residues in the P_2 and P_3 positions and large hydrophobic residues in the P_4 and P'_2 positions (Strisovsky et al., 2009; Zoll et al., 2014). The amino acid sequence preferences within the recognition

motif might be different for different rhomboids (Strisovsky, 2016b). Once the recognition motif has been unwound and bound into active site, substrate cleavage follows.

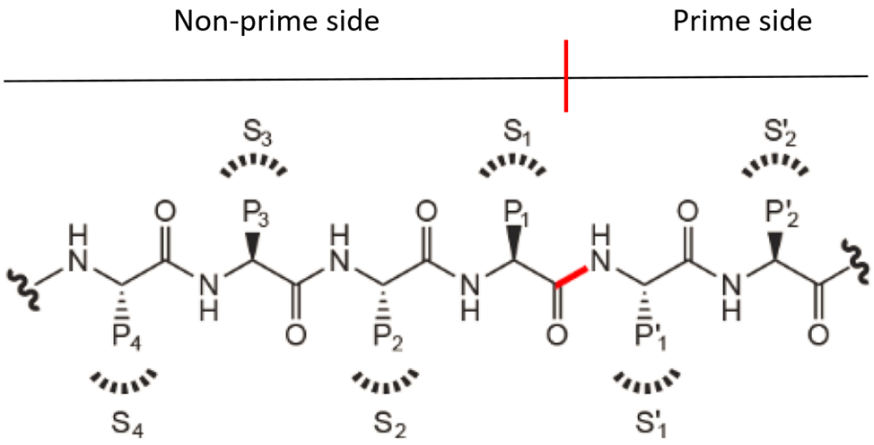

Figure 8. Rhomboid substrate recognition motif. Linear segment of about 6 amino acids of substrate termed P_4 to P'_2 binds into corresponding enzyme subsites S_4 to S'_2 during cleavage. Cleavage point is highlighted in red and non-prime side and prime side of the active site are indicated. Nomenclature is according to (Schechter and Berger, 1967). Adopted and adjusted from (Strisovsky, 2013).

As mentioned in the beginning of this chapter, it remains to be seen how many of the structural features of GlpG are conserved among rhomboid family and the field eagerly awaits structures of rhomboids other than GlpG.

1.4. Rhomboid Inhibitors

With the potential clinical relevance and scarcity of cell biology tools, the development of rhomboid protease inhibitors has been an attractive area of research (Table 2). The two-step model of substrate recognition suggests two different routes in design of rhomboid protease inhibitors (Strisovsky, 2016b). The first one is to exploit the first step – docking of a substrate into exosite, and design TMD helical peptides that would occupy the intramembrane exosite of rhomboid protease. Such inhibitors would prevent formation of docking complex and have been reported for gamma secretase, another intramembrane protease (Das et al., 2003). However, no inhibitor of this nature has been reported for rhomboid yet. All currently known inhibitors exploit the second catalytic step, which mimics the formation of a covalent substrate-enzyme complex. The inhibitors contain electrophilic warhead that binds covalently to the active site of rhomboid, thus preventing covalent binding of a substrate intermediate.

The tosyl phenylalanyl chloromethylketones and dichloroisocoumarin are wide-spectrum inhibitors of serine proteases. The efficiency of theses inhibitors against *Drosophila* Rhomboid-1, albeit

weak, helped to identify the protein as a serine protease (Urban et al., 2001). The obvious drawback of these inhibitors is their limited selectivity. Low selectivity is an issue of all small inhibitors based on highly reactive electrophilic warheads, such as later reported phosphonofluoridates (Xue and Ha, 2012; Xue et al., 2012). Phosphonofluoridates together with isocoumarines were used in crystallization studies that yielded insights into intramembrane proteolysis in atomistic resolution (1.3.2.) (Vinothkumar et al., 2010; Xue and Ha, 2012; Xue et al., 2012). However, due to their lack of selectivity, further development of any of these highly reactive inhibitors as rhomboid-specific reagents has not been pursued.

Various additional scaffolds of inhibitors have been developed and tested in the search for potent, rhomboid-specific inhibitors (Figure 9). These are β-lactams (Pierrat et al., 2011), β-lactones (Wolf et al., 2013), saccharines (Goel et al., 2017), benzoxazinones (Goel et al., 2018), and various peptidyl-based inhibitors, namely peptidyl chloromethylketones (Zoll et al., 2014), peptidyl aldehydes (Cho et al., 2016), and peptidyl α-ketoamides (Tichá et al., 2017a). However, β-lactams show only limited potency *in vivo*, since they are not able to fully inhibit endogenous GlpG (Pierrat et al., 2011). β-lactones and benzoxazinones are not very potent altogether, since the reported apparent *in vitro* half maximal inhibitory concentrations (IC_{50}; concentration of inhibitor required for 50% inhibition) are in micromolar range (Goel et al., 2018; Wolf et al., 2013). Reaction mechanism of inhibition by saccharines is uncertain (Goel et al., 2017). Saccharine inhibitors are expected to crosslink catalytic residues resulting in mobility shift on SDS-PAGE electrophoresis (Vosyka et al., 2013; Zoll et al., 2014). However, the mobility shift was not observed (Goel et al., 2017). Finally, peptidyl chloromethylketones and peptidyl aldehydes are not very potent (Zoll et al., 2014; Cho et al., 2016), but they paved the way for the currently best known class of inhibitors – peptidyl α-ketoamides (Tichá et al., 2017a).

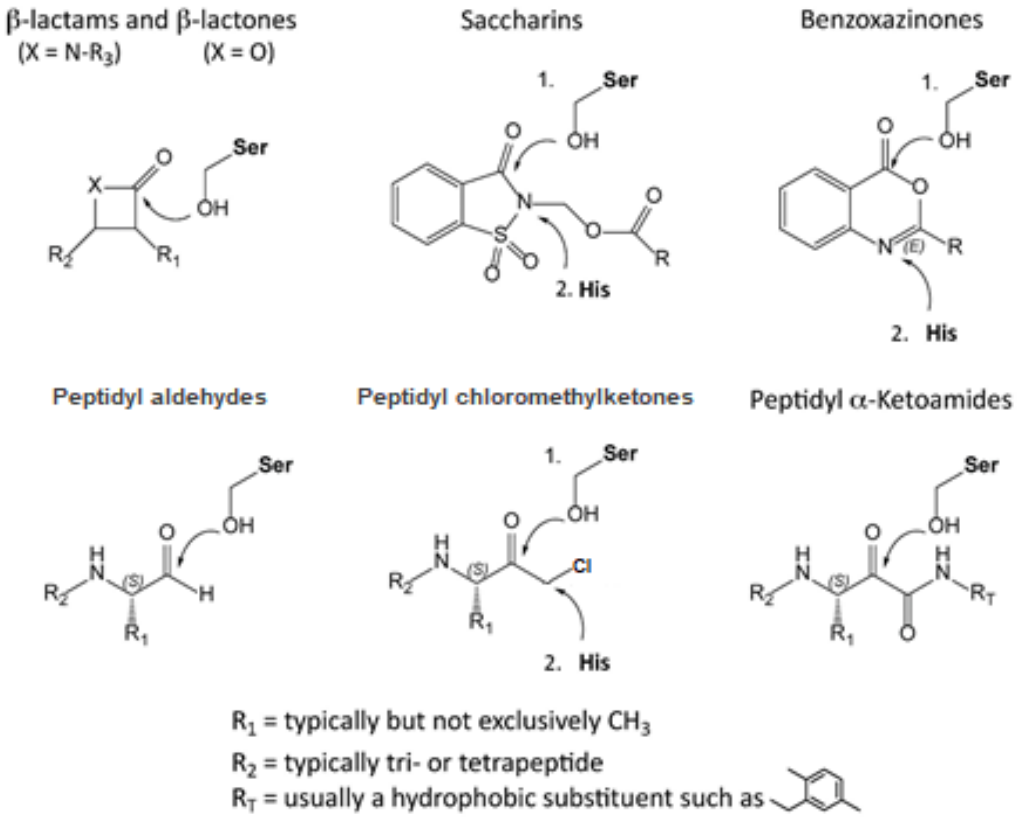

Figure 9. Structure of rhomboid protease inhibitors. A selection of known rhomboid inhibitors is depicted. The point of attack of nucleophilic catalytic serine is indicated by an arrow. In isocoumarines, saccharines, benzoxazinones, and peptidyl chloromethylketones a second covalent bond is formed between the inhibitor and electrophilic catalytic histidine. Adopted and adjusted from (Tichá et al., 2018).

Table 2: Inhibitors of rhomboid proteases. They are ordered chronologically by the year of their discovery and their chemical nature, highest reported *in vitro* potency against *E. coli* rhomboid GlpG, and some further information are listed. Please note that, due to the use of different assays and experimental conditions in different studies, the potency of inhibitors is not strictly comparable. IC_{50} – the half maximal inhibitory concentration; K_i – inhibition constant

Chemical nature	*In vitro* potency (μM)	Note	Year of discovery
tosyl phenylalanyl chloromethylketone	not determined	discovery of rhomboid protease in *Drosophila*	2001 (Urban et al., 2001)
isocoumarines	IC_{50} = 0.74	lack selectivity; turned into fluorescent probes for future inhibitor discovery (Vosyka et al., 2013)	2001 (Urban et al., 2001)
β-lactams	IC_{50} = 0.3	low potency *in vivo*	2011 (Pierrat et al., 2011)
phosphonofluoridates	IC_{50} ~ 50	structural insight into substrate docking into exosite	2012 (Xue et al., 2012)
diisopropyl fluorophoponate	IC_{50} ~ 50	insight into the structure of tetrahedral intermediate	2012 (Xue and Ha, 2012)
β-lactones	IC_{50} = 40	-	2013 (Wolf et al., 2013)
peptidyl chloromethylketones	IC_{50} ~ 100	structural insight into substrate docking into active site	2014 (Zoll et al., 2014)
peptidyl aldehydes	K_i = 20	structural insight into substrate docking into active site	2016 (Cho et al., 2016)
peptidyl α-ketoamides	K_i = 0.045	selective; tunable	2017 (Tichá et al., 2017a)
saccharins	IC_{50} ~ 0.2	uncertain reaction mechanism of inhibition	2017 (Goel et al., 2017)
benzoxazinones	IC_{50} ~ 5	-	2018 (Goel et al., 2018)

Inhibitors of rhomboid proteases are almost exclusively developed and tested against *E. coli* rhomboid GlpG. The reasons are simple – GlpG is the only rhomboid with a solved structure (Wang et al., 2006; Ben-Shem et al., 2007; Vinothkumar et al., 2010; Vinothkumar, 2011), offers a relatively simple biochemical manipulation by being catalytically active in detergent micelles (Lemberg et al., 2005), has a number of known well characterised artificial substrates which can be biochemically modified for cleavage experiments (Tichá et al., 2017b), and *E. coli* background offers well established means of genetic manipulation and protein expression. As a drawback, not much is known about the efficiency with which most rhomboid inhibitors act against other rhomboid proteases, such as potentially clinically relevant human RHBDL2, RHBDL4, PARL, entamoeba EhROM1, and rhomboids of apicomplexan parasites.

Another long standing caveat in rhomboid inhibitor development was their unknown selectivity, often tested only against trypsin or chymotrypsin (Pierrat et al., 2011; Vosyka et al., 2013). However, peptidyl α-ketoamides have addressed this issue. Peptidyl α-ketoamides inhibit some, but not all tested rhomboid proteases and their selectivity for rhomboid proteases over other serine proteases have been demonstrated (Figure 10) (Tichá et al., 2017a). The screen for possible off-targets, 96 human serine hydrolases, resulted in only two hits (prolylcarboxypeptidase and dipeptidylpeptidase), demonstrating selectivity of peptidyl α-ketoamides (Tichá et al., 2017a). Although β-lactams and sacchcarines are also quite selective for rhomboids (Goel et al., 2017; Tichá et al., 2017a), they suffer from other drawbacks, as discussed above. Selectivity of peptidyl α-ketoamides is encouraging in respect of their use as specific inhibitors for studying cell biological roles of rhomboids and/or for their pharmacological targeting. Indeed, peptidyl α-ketoamide is an already approved pharmacophore, since boceprevir, a ketoamide peptidomimetic, has been approved for treatment of hepatitis C virus infection (Njoroge et al., 2008). Tuning of peptidyl α-ketoamides is the topic of the Results section of this thesis, therefore they deserve to be introduced in a greater detail.

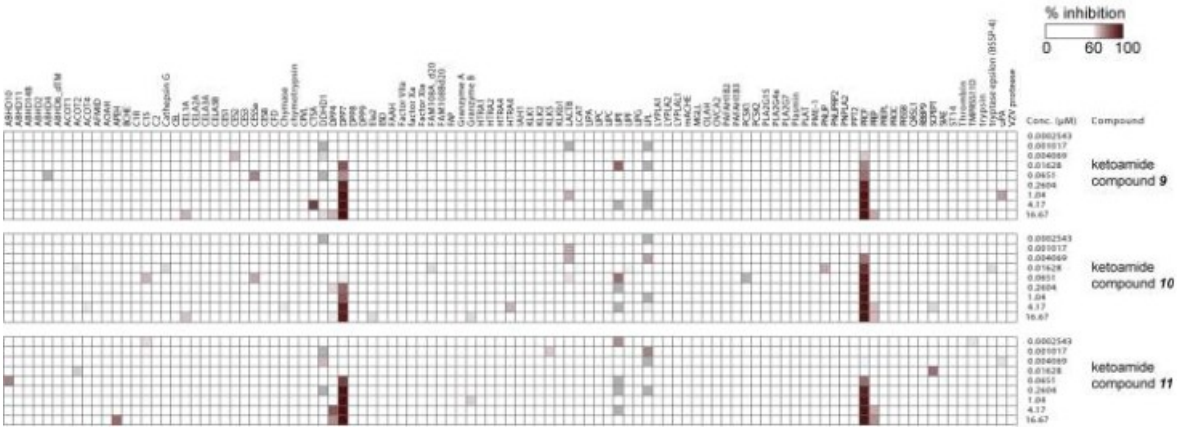

Figure 10. Selectivity of peptidyl α-ketoamides. Selectivity of the three most potent compounds (compound **9**, **10**, and **11**) was tested by EnPlex method (Bachovchin et al., 2014). The method is based on the fluorescence readout of competitive binding between the inhibitor and an activity-based probe. Adopted and adjusted from (Tichá et al., 2017a).

1.4.1. Peptidyl α-ketoamides as specific rhomboid protease inhibitors

Peptidyl α-ketoamides with various 'tail' hydrophobic substituents of ketoamide nitrogen (Figure 9 and Figure 11), developed in our laboratory, were demonstrated to efficiently inhibit rhomboid proteases in both, *in vitro* and *in vivo*, conditions (Tichá et al., 2017a). Additionally, Tichá *et al.* demonstrated their selectivity (Figure 10; discussed above in 1.4.), reversible and non-competitive mode of action, and propose ways of their further improvement (discussed below).

The major step in development of peptidyl inhibitors was identification of GlpG preference for amino acids in the P_1-P_5 positions of the recognition motif (Figure 8) of the GlpG artificial substrate TatA (Zoll et al., 2014). The best cleaved amino acid sequence, RVRHA, was coupled with various electrophilic warheads and the apparent *in vitro* IC_{50} values were measured (Tichá et al., 2017a). Ac-RVRHA-ketoamide was not the most potent candidate, with both, Ac-RVRHA-chloromethylketone and Ac-RVRHA-boronate exhibiting greater potency in this initial screen (Tichá et al., 2017a). However, unlike chloromethylketones and boronates, α-ketoamides offered clear path for further modifications by substituting the ketoamide nitrogen and extension into the prime side of rhomboid active site (Liu et al., 2004).

The significance of the prime and non-prime side of the inhibitor was explored (Figure 11) (Tichá et al., 2017a). First, various tail substituents were attached to the ketoamide nitrogen of Ac-RVRHA-ketoamide-tail inhibitor and apparent *in vitro* IC_{50} values of these compounds were measured (Figure 11A). The potency of inhibitors clearly correlates with the size of a tail substituent on the non-prime side. Inhibition constant (K_i) of compound **11**, the most potent reported rhomboid inhibitor, is (45 ± 8) nM. Next, the importance of the peptidyl in the prime side was assessed (Figure 11B). Truncating of the peptidyl part of compound **9** has great effect on the inhibitor potency, except for the terminal arginine in P_5 position, which seems to play smaller role in inhibitor binding to the enzyme.

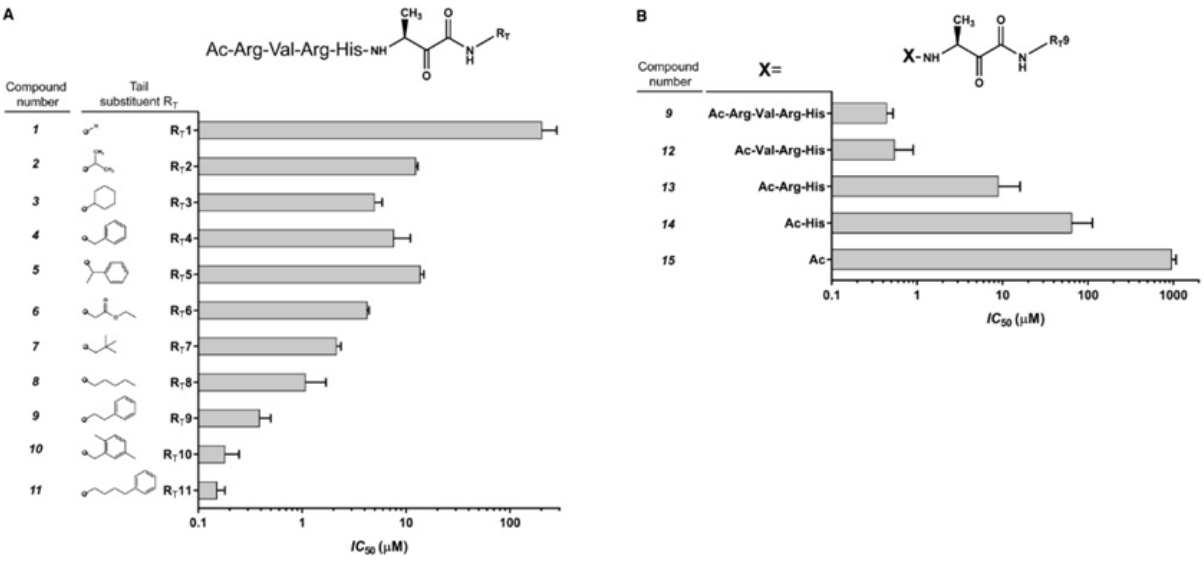

Figure 11. *In vitro* inhibition potency of peptidyl α-ketoamides. The apparent *in vitro* IC_{50} of various peptidyl α-ketoamides inhibitors of GlpG were measured in 0.05 % (w/v) DDM and 10 μM KSp35 substrate (3.4.) (Tichá et al., 2017b). **A)** Peptidyl α-ketoamide tail substituents (R_T) of Ac-RVRHA-$CONH_2$-R_T inhibitor core (depicted on top) were tested. **B)** Truncating of peptidyl (X) of compound **9** from A) was tested. Adopted from (Tichá et al., 2017a).

The ability of peptidyl α-ketoamides to inhibit GlpG and YqgP *in vivo* has been demonstrated. The apparent *in vivo* IC_{50} of compound **11** against GlpG is (2.7 ± 0.1) nM and against YqgP ~ 5-10 nM (Tichá et al., 2017a). The outer membrane of gram-negative bacterium *E. coli* was genetically permeabilized by a mutation in *lptD* gene (Ruiz et al., 2005). LptD is responsible for lipopolysaccharide (LPS) assembly at the outer membrane of bacterium (Li et al., 2015b), and LPS contributes to the integrity of outer membrane (reviewed in Nikaido, 2003). Therefore, *lptD* mutant has greater membrane permeability in general and the mutation improves inhibitor access to the rhomboid GlpG, which is located in the inner membrane of the bacterium. Indeed, all of the studies reporting *in vivo* inhibition use the *lptD* mutant bacterium, suggesting low membrane permeability of these compounds (Pierrat et al., 2011; Cho et al., 2016; Tichá et al., 2017a; Goel et al., 2017, 2018). However, peptidyl α-ketoamides are able to cross the outer membrane of *E. coli* in *lptD* wild-type (WT) strain, since compound **11** has an apparent *in vivo* IC_{50} of ~ 110 nM in *E. coli* MC4100 (A. Tichá, not reported).

The structure models of compounds **9** and **10** soaked into crystals of GlpG were solved at 2.16 Å (compound **9**; PDB code 5MT6) and 1.78 Å (compound **10**; PDB code 5MTF) resolution, respectively, providing an atomistic resolution of binding of inhibitor to GlpG (Figure 12) (Tichá et al., 2017a). The inhibitors do not precipitate and probably binds directly from the solution into the upper part of the intramembrane protease as an extended β strand (Figure 12A). The ketoamide warhead binds covalently to the catalytic residue Ser201 and is further stabilized by hydrogen bonds with catalytic residues Ser201 and His254, and oxyanion stabilising residues His150 and Asn154 (Figure 12B). The hydrophobic tail substituent extensively occupies the hypothesized S'_2 subsite of the enzyme (Figure 12C). β-lactam L29 and isocoumarin S016 occupies larger space at the prime side of the active site than the tested peptidyl α-ketoamides (Figure 12D). Based on this inhibitor structural alignment, Tichá *et al.* propose further development of peptidyl α-ketoamides by testing larger or branched tail substituentst of ketoamide nitrogen.

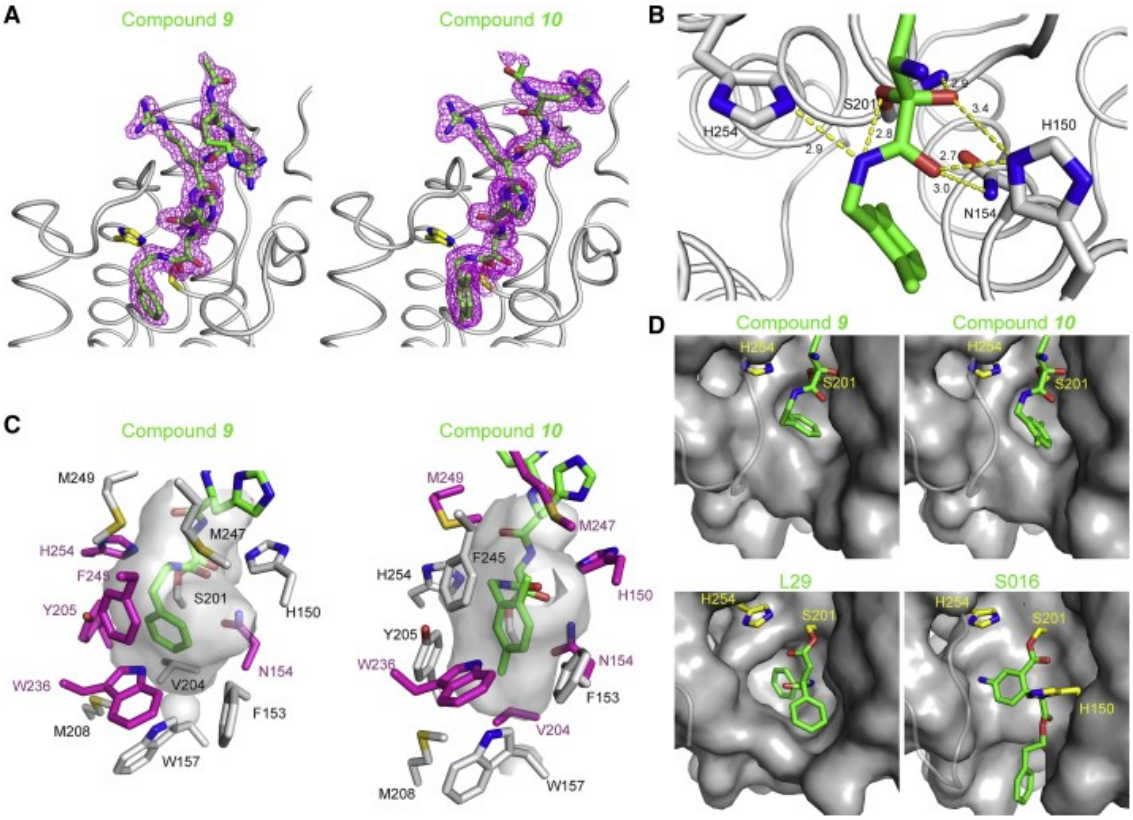

Figure 12. Binding of peptidyl α-ketoamides to rhomboids. Two co-crystal structures of compounds **9** (PDB code 5MT6) and **10** (PDB code 5MTF) and GlpG were solved at 2.16 Å and 1.78 Å resolution, respectively. **A)** Inhibitors, shown in green and surrounded by electron density map, were soaked into GlpG crystals (grey). Catalytic dyad is shown in yellow. **B)** Ketoamide warhead, shown in green, binds covalently into the rhomboid (grey) active site. Six hydrogen bonds, showed as yellow dashed lines, stabilise the binding. **C)** Peptidyl α-ketoamide tail substituents of compounds **9** and **10** interact with the prime side of GlpG active site. The inhibitor is shown in green. The cavity surrounding the tails is shown as inversed surface. The amino acid side chains forming van der Waals contacts with the inhibitor are shown as magenta sticks. The remaining amino acid side chains forming the cavity are shown as grey sticks. **D)** Structure comparison of binding modes of ketoamide tail substituents (compounds **9** and **10**) and β-lactam L29 (PDB code 3ZMI) and isocoumarin S016 (PDB code 3ZEB). Inhibitors are shown in green, protein in grey, and catalytic residues in yellow. Adopted from (Tichá et al., 2017a).

2. Aims of the study

This study aims to improve N-substituted peptidyl α-ketoamides as specific rhomboid protease inhibitors for studying their biological roles and/or for their pharmacological targeting. For testing *in vitro* inhibitory potency, model rhomboid GlpG from *E. coli* needs to be expressed, solubilised into micelles, and purified. Next, to explore chemical space of the prime side of rhomboid active site in inhibitor design, apparent *in vitro* and *in vivo* IC_{50} of peptidyl α-ketoamides bearing various tail substituents are assessed in an attempt to improve inhibitor potency. Since this attempt was successful, importance of peptidyl part and outer membrane permeability of the newly developed inhibitor are tested *in vitro* and in *E. coli* strain MC4100, respectively. Last, but not least, to explain the improved potency, the interactions of the tail substituent of the newly developed inhibitor with GlpG are analysed in a computational model of the complex.

3. Material

3.1. Instruments and tools

96-well plate	Greiner Bio-One, Austria
Analytical weights XA 110/X	Radwag, Poland
Autoclave MLS-3020U	Sanyo, Japan
Centrifugal concentrators Vivaspin® 20 (cutoff 50 kDa)	Sartorius Stedim Biotech, Germany
Centrifuge Allegra® X-15R	Beckman Coulter, USA
Centrifuge Avanti® J-30I	Beckman Coulter, USA
Centrifuge Eppendorf® 5424	Eppendorf, Germany
(Ultra-) centrifuge Optima™ L-90K	Beckman Coulter, USA
Column for desalting PD-10	GE Healthcare, Great Britain
Dounce homogenizer	Sigma-Aldrich, USA
Electrophoresis/Western blot apparatus	Bio-Rad, USA
Filter paper	Macherey-Nagel, Germany
Filters 0.45 µm	Carl Roth, Germany
Freezer Comfort (-20 °C)	Liebherr, Germany
Freezer Ultra-Low (-80 °C)	Sanyo, Japan
Homogenizer EmulsiFlex®-C3	Avestin, Canada
Microplate reader TECAN Infinite® M1000	TECAN, Austria
Power supply Consort EV202	Sigma-Aldrich, USA
Scanner EPSON perfection V37	EPSON, Japan
TALON™ Metal Affinity Resin	Clontech Laboratories, USA
Thermomixer	Eppendorf, Germany
Imager Odyssey® CLx	LI-COR Biosciences, USA
Incubator Innova® 44/44R	New Brunswick Scientific, USA
Incubator IPP 400	Memmert, Germany
Spectrophotometer BioSpectrometer® kinetic	Eppendorf, Germany
Spectrophotometer NanoDrop ND-1000	Thermo Scientific, USA
Water bath	Benchmark Scientific, USA

3.2. Chemicals and consumables

2 % Bis-acrylamide	BioRad #1610142
40 % Acrylamide	BioRad #1610140
4–20 % acrylamide precast gel	BioRad #4561094
$(NH_4)_2SO_4$	Penta #25240-31000
α-lactose	Sigma #L3625
β-D-maltopyranoside (DDM)	Antrace #69227-93-6
β-mercaptoethanol	Sigma #M3148
Antibody, rabbit anti-FLAG IgG	Cell Signaling Technology #2368
Antibody, IRDye® 800CW donkey anti-Rabbit IgG	Invitrogen #SA5-10044
Agarose	Serva #120274

Ammonium persulfate (APS)	Sigma #248614
Ampicillin	Biotika #1808005
Blocker™ Casein in TBS	Thermo #37532
$CaCl_2$	Penta #16750-31000
Cobalamin	Sigma #V2876
$CoCl_2$	Sigma #60818
Color protein standard	NEB #P7712
Coomassie Brilliant Blue G-250	Sigma #B-0770
cOmplete^{TM} protease inhibitors	Roche #33576900
$CuCl_2$	Sigma #307483
Dimethyl sulfoxide (DMSO)	Sigma #D8418
Dimethyl sulfoxide (DMSO), non-aqueous	Invitrogen #12345
E. coli strain C41 (DE3)	Lucigen #60452-1
E. coli strain NR698	(Ruiz et al., 2005)
E. coli strain NR698 *glpG::tet*	(Pierrat et al., 2011)
E. coli strain MC4100	(Akiyama, 2002)
E. coli strain MC4100 *glpG::tet*	This thesis
$FeCl_3$	Sigma #157740
Ethylenediaminetetraacetic acid (EDTA)	Sigma # E5134
Glycerol	Penta #56-81-5
Glucose	Penta #12020-31000
H_3BO_3	Sigma #B6768
HEPES	Sigma #H3375
Imidazole	Sigma #56750
Immobilon®-FL PVDF	Sigma #IPFL00010
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Sigma #I6758
Kanamycin	Sigma #60615
L-Alanine	Sigma #A7627
L-Arginine	Sigma #A5006
L-Asparagine	Sigma #A0884
L-Aspartic acid	Sigma #A9256
L-Glutamic acid	Sigma #G1251
L-Glutamine	Sigma #G3126
L-Glycine	Sigma #G8898
L-Histidine	Sigma #H8000
L-Isoleucine	Sigma #I2752
L-Leucine	Sigma #L8000
L-Lysine	Sigma #L5501
L-Methionine	Sigma #M9625
L-Phenylalanine	Sigma #P2126
L-Proline	Sigma #P0380
L-Rhamnose	Sigma #W373011
L-Serine	Sigma #84959
L-Threonine	Sigma #T8625
L-Tryptophan	Sigma #T0254
L-Valine	Sigma #W527718

Lysogeny broth (LB)	Sigma #L30
Methanol	Lach-Ner #20038-AT0
$MgCl_2$	Affymetrix #78641
$MgSO_4$	Penta #43180-31000
$MnCl_2$	Sigma #221279
Na_2HPO_4	Sigma #S3264
Na_2MoO_4	Sigma #243655
NaCl	Penta #16610-31000
$NiCl_2$	Sigma #339350
Pierce™ 660nm Protein Assay	Thermo #22660
Pierce™ universal nuclease	Thermo #88701
Plasmid pET-25b+	(Lemberg et al., 2005)
Plasmid pPR61	(Tichá et al., 2017a)
Phenylmethylsulfonyl fluoride (PMSF)	Sigma #78441B
Polyethylene glycol 8000 (PEG8000)	Affymetrix #19966
Potassium phosphate dibasic	Sigma # P8281
Potassium phosphate monobasic	Sigma # P5655
Sodium dodecyl sulphate (SDS)	BioRad #1610301
Tetramethylethylenediamine (TEMED)	Roth #2367.1
Tris	Sigma #T1503
TWEEN® 20	Sigma #P7949
$ZnSO_4$	Sigma #Z0251

Material used in experiments not carried out by the author of this thesis (inhibitor and substrate synthesis and characterization) is not listed.

3.3. Buffers and media

LB medium: 20 g lysogeny broth dissolved in 1 l of sterile water

Modified PASM-5052 growth medium: 50 mM Na_2HPO_4; 50 mM KH_2PO_4; 25 mM $(NH_4)_2SO_4$; 1 mM $MgSO_4$; 0.1 mM $FeCl_3$; 1 mM $CaCl_2$; 1 mM $MnCl_2$; 1 mM $ZnSO_4$; 0.2 mM $CoCl_2$; 0.1 mM $CuCl_2$; 0.2 mM $NiCl_2$; 0.1 mM Na_2MoO_4; 0.1 mM Na_2SeO_3; 0.1 mM H_3BO_3; 50 µM HCl; 0.5 % (v/v) glycerol; 0.05 % (w/v) glucose; 0.2 % (w/v) α-lactose; 220 µg/ml each of 17 amino acids (cysteine, methionine, and tyrosine not included); 135 µg/ml L-Methionine; 100 nM cobalamin; 100 µg/ml ampicillin

Purification buffer A: 20 mM HEPES (pH 8.0); 10 % (v/v) glycerol; 100 mM NaCl; 5 mM $MgCl_2$

Purification buffer B: 20 mM HEPES (pH 8.0); 20 % (v/v) glycerol; 300 mM NaCl; 5 mM $MgCl_2$; 10 mM imidazole

Equilibration buffer: 25 mM HEPES (pH 8.0); 10 % (v/v) glycerol; 500 mM NaCl; 10 mM imidazole; 10 mM $MgCl_2$; 5 mM β-mercaptoethanol; 0.05 % (w/v) DDM

Washing buffer A: 20 mM HEPES (pH 7.4); 10 % (v/v) glycerol; 300 mM NaCl; 25 mM imidazole; 0.05 % (w/v) DDM

Washing buffer B: 20 mM HEPES (pH 7.4); 10 % (v/v) glycerol; 300 mM NaCl; 50 mM imidazole; 0.05 % (w/v) DDM

Elution buffer A: 20 mM HEPES (pH 7.4); 10 % (v/v) glycerol; 300 mM NaCl; 250 mM imidazole; 0.05 % (w/v) DDM

Elution buffer B: 20 mM HEPES (pH 7.4); 10 % (v/v) glycerol; 300 mM NaCl; 350 mM imidazole; 0.05 % (w/v) DDM

Final buffer: 20 mM HEPES (pH 7.4); 10 % (v/v) glycerol; 150 mM NaCl; 0.05 % (w/v) DDM

Cleavage buffer A: 20 mM HEPES (pH 7.4); 150 mM NaCl; 0.05 % (w/v) DDM; 10 % (v/v) DMSO

Cleavage buffer B: 50 mM potassium phosphate (pH 7.4); 20% (v/v) glycerol; 150 mM NaCl; 0.05 % (w/v) DDM; 0.05 % (w/v) PEG8000; 10 % (v/v) DMSO

6× SDS sample buffer: 350 mM Tris (pH 6.8); 10 % (w/v) SDS; 50 % (v/v) glycerol; 10 % (v/v) β-mercaptoethanol; 0.05 % (w/v) bromophenol blue

SDS-PAGE running buffer: 25 mM Tris; 192 mM glycine; 0.1 % (w/v) SDS

Western blot running buffer: 12.5 mM Tris; 96 mM glycine; 5 % (v/v) methanol; 0.005 % (w/v) SDS

4 %/12 % polyacrylamide gel: 2 ml of stacking 4 % gel (0.19 ml 40 % acrylamide; 0.1 ml 2 % bis-acrylamide; 0.5 ml 1 M Tris (pH 6.8); 20 µl 10 % (w/v) SDS; 1.17 ml water; 1.6 µl TEMED; 20 µl 10 % (w/v) APS); 5 ml of resolving 12 % gel (1.45 ml 40 % acrylamide; 1 ml 2 % bis-acrylamide; 1.25 ml 1.5 M Tris (pH 8.8); 50 µl 10 % (w/v) SDS; 1.2 ml water; 2 µl TEMED; 50 µl 10 % (w/v) APS)

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

All media have been autoclaved. All buffers except for 6× SDS sample buffer and SDS-PAGE and Western blot running buffers have been filtered through 0.45 µm pores (Carl Roth).

3.4. Peptide substrates

Fluorogenic peptides KSp35 and KSp96 (Table 3) have been synthetized and reconstituted into micelles in our laboratory by Dr. Stancho Stanchev as previously described (Rath and Deber, 2013; Tichá et al., 2017b). Specifically, peptide was initially dissolved in HFIP which was subsequently evaporated and peptide was redissolved in TFE and diluted in 0.05 % (w/v) n-dodecyl β-D-maltopyranoside (DDM). Aliquots of 500 µl 50 µM peptides were lyophilized at -70 °C until dry. **KSp35** sequence is based on the second transmembrane domain of LacY (LacYTM2), which is an artificial GlpG substrate (Maegawa et al., 2005). Förster resonance energy transfer (FRET) pair is connected to the peptide through amino acids in the P_5 and P'_4 positions (Figure 8); the fluorophore EDANS is conjugated to the glutamic acid in the P_5 position and the quencher DABCYL is conjugated to the lysine in the P'_4 position. The P_5 and P'_4 positions in the substrates are only loosely important for substrate recognition as GlpG easily cleaves substrates with various substituents in these positions (Strisovsky et al., 2009; Zoll et al., 2014; Tichá et al., 2017b). **KSp96** sequence is also based on LacYTM2 (Maegawa et al., 2005), with the P_5-P_3 and P_1 amino acids of LacYTM2 being substituted for more

cleavage-preferred ones (Zoll et al., 2014; Tichá et al., 2017b). The fluorophore EDANS is conjugated to the glutamic acid in the P_2 position (instead of the P_5 in KSp35 in order to increase FRET efficiency) and the quencher DABCYL is conjugated to the same residue as in KSp35 – lysine in the P'_4 position. Similarly to the P_5, P_2 position in the substrate has only minor effect on GlpG cleavage and can accommodate various substitutions (Strisovsky et al., 2009; Zoll et al., 2014). Additionally, KSp96 contains substitution in the P_8 position (Arg for Asp) and an additional N-terminal Arg that should increase peptide solubility (Rath and Deber, 2013), both without significant effect on cleavage efficiency. Biophysical characterization of fluorogenic GlpG substrates has already been carried out previously (Tichá et al., 2017b).

Table 3: The sequences of FRET-based fluorogenic peptide substrates, KSp35 and KSp96. Cleavage site is highlighted by asterisk.

Peptide	Sequence
KSp35	KRHDIN(E-EDANS)ISKS*DTG(K-DABCYL)IFAAISLFSLLFQPLFGLSKK
KSp96	KRHRINRVR(E-EDANS)A*DTG(K-DABCYL)IFAAISLFSLLFQPLFGLSKKR

3.5. Inhibitors

Compounds **16** – **25** (Table 4) have been synthetized and characterised by mass spectrometry and NMR in our laboratory by Dr. Stancho Stanchev as described previously (Tichá et al., 2017a). Compounds **16** – **18**, and **21** – **25** displayed characteristic double peak on HPLC-MS, corresponding to the same molecular mass, which suggest that these inhibitors are a mixture of two diastereomers. The racemisation most probably occurs at the histidine in the P_2 position (Figure 8). Compound **21** contains majority of one diastereomer whereas compound **22** is made of majority of the second diastereomer. The absolute diastereomerisation of compounds **21** and **22** is unknown.

Inhibitors were dissolved in non-aqueous DMSO and concentrations were determined by quantitative amino acid analysis performed by Ing. Radko Souček at the Institute of Organic Chemistry and Biochemistry of Czech Academy of Sciences.

Table 4: Tested peptidyl α-ketoamides. Tested compounds are a mixture of two diastereomers with chiral centre probably at the alpha carbon of His in the P_2 position. Compounds **21** and **22** are partially separated fractions with majority of one or the other diastereomer.

Compound	Identifier	Structure
16	STS1002	Ac-VRHA-CONH-phenylethyl
17	STS1074a	Ac-VRHA-CONH-phenylpropyl
18	STS1058b	Ac-VRHA-CONH-benzyloxyethyl
19	STS898b	Ac-VRHA-CONH-phenyl(pyridin-4-yl)methyl
20	STS1026	Ac-VRHA-CONH-2-(diphenylamino)ethyl
21	STS1034a	Ac-VRHA-CONH-1-(benzyloxy)-3-phenylpropan-2-yl
22	STS1034b	Ac-VRHA-CONH-1-(benzyloxy)-3-phenylpropan-2-yl
23	STS1252	Ac-RHA-CONH-1-(benzyloxy)-3-phenylpropan-2-yl
24	STS1259	Ac-HA-CONH-1-(benzyloxy)-3-phenylpropan-2-yl
25	STS1265	Ac-A-CONH-1-(benzyloxy)-3-phenylpropan-2-yl

3.6. Computational model

The model of an enzyme-inhibitor complex of GlpG and compound **21/22** was built by Dr. Jindřich Fanfrlík at the Institute of Organic Chemistry and Biochemistry of Czech Academy of Sciences. The parent model (PDB code 5MT6) (Tichá et al., 2017a) of Ac-RVRHA-CONH-phenylethyl (compound **9**) soaked into crystals of GlpG was used as the starting point of modelling. Compounds **21** and **22** are diastereomers with a chiral centre probably at the alpha carbon of His in the P_2 position (3.5)**.** The stereoconfiguration of compounds **21** and **22** is unknown (3.5.), but this information was not necessary for this computational model. The goal of the computational model is to analyse the uniform tail substituent of compounds **21** and **22.** The configuration of His in P_2 position from the parent model (PDB code 5MT6) was kept intact in the modelled complex.

The terminal arginine in the P_5 position is not present in the modelled compound **21/22**, thus it was deleted and acetyl group was added to terminal valine in the P_4 position. Waters were discarded. Hydrogens were added and optimised in Amber (Salomon-Ferrer et al., 2013) by Dr. Martin Lepšík at the Institute of Organic Chemistry and Biochemistry of Czech Academy of Sciences. The branching point of the modelled branched tail substituent is an asymmetric carbon, suggesting two possible stereoisomers of the tail (see also Figure 15). However, single peak on HPLC-MS read-out (obtained by Dr. Stancho Stanchev in our laboratory) of acetyl ketoamide conjugated with the modelled branched tail substituent suggest only one isomeric product of the synthesis of these compounds. The chirality of the branching atom was proposed (Dr. Stancho Stanchev) to be identical with the one in the parent crystal (PDB code 5MT6). Thus, the second branch of the tail was simply built with the Builder function in PyMOL, starting from the corresponding hydrogen of the asymmetric carbon. Different rotational conformations of the tail were considered, but, due to the spherical clashes, only one of them (the one

corresponding to the parent crystal; PDB code 5MT6) was chosen for further work. The energy of the enzyme-inhibitor complex was optimised by hybrid QM/SQM method. The QM part comprised the inhibitor and nearby residues – His150, Asn154, Ser201, Met247, Met249, His254. The QM part was treated at the DFT-D3/BLYP/DZVP level (Hostaš and Řezáč, 2017). The rest of the system was treated at the PM6-D3H4X level (Stewart, 2007; Řezáč and Hobza, 2012). The environment was described by the COSMO implicit solvent model (Klamt and Schüürmann, 1993). The coupling between QM and SQM was done by program Cuby4 (Řezáč, 2016), which calls Turbomole 7.0 (Ahlrichs et al., 1989) and Mopac (Stewart, 2004) for QM and SQM, respectively. Residues further than 10 Å of the inhibitor and Ser201 in the crystal structure were frozen during the optimisation. In order to find the correct binding mode of the new branch, molecular dynamics simulation (only residues Phe153, Trp157, Trp236, Asp243-Ala250 and the inhibitor were free while the rest of the complex was frozen) proceeded in 100 000 iterations (time step of 0.001 ps, temperature 300 K, Berendsen termostat and igb7 solvent model). Snapshots were taken every 10 000 iterations. The energies of 10 snapshots were again optimised by the QM/SQM method described above. The one with the lowest energy was chosen for further analysis.

3.7. Software

ChemBioDraw version 14.0.0.117	PerkinElmer Inc.
GraphPad Prism versions 7.00, 7.03 and 8.00 for Windows	GraphPad Software Inc.
Image Studio version 5.2.5	LI-COR, Inc.
LigPlot+ version 2.1	(Laskowski and Swindells, 2011)
The PyMOL Molecular Graphics System, versions 1.8.6 and 2.0.0	Schrödinger, LLC.

4. Methods

4.1. Protein expression and purification

Bacterial rhomboid protease GlpG was expressed in *E. coli* strain C41 (DE3) (Miroux and Walker, 1996) transfected with plasmid pET-25b+ which contains *glpG* under the lac1 promotor (Lemberg et al., 2005). After expression, GlpG was purified and its catalytic activity was confirmed.

4.1.1. GlpG expression

E. coli cells, from 10 % (v/v) glycerol stock stored at -80 °C (Ultra Low; Sanyo), were incubated in **LB medium** overnight at 37 °C, 220 rpm (Innova 44/44R; New Brunswick Scientific). Afterwards, 1 liter of **Modified PASM-5052 growth medium** (Studier, 2005) was inoculated with 10 ml of the overnight culture and incubated for 2.5 h at 37 °C, 220 rpm (Innova 44/44R; New Brunswick Scientific). GlpG expression from lac1 promotor was induced by 0.5 ml of 1 M isopropyl β-D-1-thiogalactopyranoside (IPTG) when the optical density at 600 nm (OD_{600}) was 0.6 (measured by BioSpectrometer® kinetic; Eppendorf). The cells were cultured overnight at 16 °C, 220 rpm (Innova 44/44R; New Brunswick Scientific). On the next day, OD_{600} of the culture was 2. Cells were harvested by centrifugation for 15 min at 4 °C, 6 000 × g (Avanti® J-30I; rotor JLA-9.1000; Beckman Coulter). Supernatant was disposed of, whereas cell pellet containing the protein was flash frozen in liquid nitrogen and stored at -80 °C (Ultra Low; Sanyo) until protein isolation.

4.1.2. GlpG purification

Cell pellet was resuspended in 30 ml of **Purification buffer A**. Inhibitors of proteases, EDTA and PMSF, were added to 1 mM final concentration to prevent proteolytic protein degradation. Suspension was homogenized with a Dounce homogenizer and plasma membranes were disrupted by high pressure homogenization (EmulsiFlex® C3; Avestin). Afterwards, cell debris was removed by centrifugation for 30 min at 4 °C, 10 000 × g (Avanti® J-30I; rotor JLA-9.1000; Beckman Coulter). Supernatant, containing cell membranes, was further ultracentrifuged for 2 h at 4 °C, 130 000 × g (Optima™ L-90K; rotor 45Ti; Beckman Coulter). Pellet was resuspended in **Purification buffer B** with cOmplete™ protease inhibitors (Roche) and protein concentration was estimated by Pierce™ 660nm Protein Assay to be 24.3 mg/ml (measured according to manufacturer protocol; Thermo Scientific). GlpG was diluted in **Purification buffer B** to a concentration of 4.9 mg/ml. Detergent n-dodecyl β-D-maltopyranoside (DDM) was added to a final 1.5 % (w/v) concentration to solubilize membrane proteins. Suspension was incubated for 1 h at room temperature (RT) and subsequently ultracentrifuged (1 h, 4 °C, 130 000 × g; Optima™ L-90K; rotor 45Ti; Beckman Coulter) to isolate solubilized membrane proteins.

GlpG was purified on TALON™ metal affinity resin (Clontech Laboratories). 1 ml of 50 % (v/v) of the resin was centrifuged twice for 2 minutes at 700 × g (Eppendorf® 5424; Eppendorf), supernatant was decanted, and the pellet containing resin was mixed with solubilised proteins and incubated for 1 h at 4 °C. The resin-enzyme mixture was loaded onto the column and washed twice with 50 ml of **Washing buffer A** and twice with 50 ml of **Washing buffer B**. GlpG was eluted four times with 2.5 ml of **Elution buffer A** and four times with 2.5 ml of **Elution buffer B**. To get rid of imidazole, which increases the risk of protein aggregation, the elution buffer was exchanged by gel filtration (column PD-10; GE Healthcare). The column was equilibrated with **Final buffer,** individual fractions (Elution buffer A fractions 1-4 and Elution buffer B fractions 1-4) were separately loaded on the individual columns, and the **Final buffer** was used for protein elution. The washing and elution fractions were analysed by SDS-PAGE electrophoresis (45 min, 200 V; Consort EV202; Sigma-Aldrich) on 4 %/12 % polyacrylamide gel. The gel was stained with **Razor Blue Stain** according to the manufacturer protocol. First five elution fractions (Elution buffer A fraction 1-4 and Elution buffer B fraction 1), which contained high amounts of GlpG, were pooled together and concentrated by centrifugation (4000 × g, 4 °C; Allegra X-15R; Beckman Coulter) in Vivaspin® 20 concentrator (cut off 50 kDa; Sartorius Stedim Biotech) from 107 µg/ml to 538 µg/ml (measured at NanoDrop ND-1000; Thermo Scientific; extinction coefficient 87 890; molecular weight 33.7 kDa). Enzyme was aliquoted, flash frozen in liquid nitrogen, and kept at -80 °C (Ultra Low; Sanyo) until further use.

4.1.3. *In vitro* GlpG activity

GlpG was diluted in **Cleavage buffer A** to 400 nM concentration. A FRET based fluorogenic substrate KSp35 (3.4.), previously reconstituted into micelles (0.05 % (w/v) DDM), was dissolved to a concentration of 50 µM in 20 mM HEPES (pH 7.4) containing 150 mM NaCl. 40 µl of GlpG were mixed with 10 µl of KSp35 in a 96-well plate. Substrate cleavage was measured by fluorescence microplate reader TECAN Infinite® M1000 (TECAN) for 1 hour at 37 ° as an increase in emission at 493 nm wavelength after excitation by UV light (335 nm). Cleavage data were analysed as described in 4.4.1.

4.2. *In vitro* inhibition assay

Inhibitors were dissolved in non-aqueous DMSO in order to obtain 20 mM stocks and further diluted in **Cleavage buffer B** to desired concentrations. 20 µl of each inhibitor concentration were mixed with 20 µl of 12 nM GlpG and incubated in thermomixer at 37 °C, 1 000 rpm for 1 h (Eppendorf). A FRET based fluorogenic substrate KSp96 (3.4.), previously reconstituted into micelles (0.05 % (w/v) DDM), was dissolved to 50 µM concentration in **Cleavage buffer B without DDM and DMSO**. Afterwards, GlpG activity was measured similarly as in 4.1.3: 25 µl of GlpG with inhibitor were mixed with 25 µl of KSp96 in a 96-well plate. Substrate cleavage was measured on fluorescence microplate

reader TECAN Infinite® M1000 (TECAN) for 1 hour at 37 °C as an increase in fluorescence (emission at 493 nm wavelength after excitation by UV light at 335 nm). Cleavage data were analysed as described in 4.4.1.

4.3. *In vivo* inhibition assay

In vivo assays were either performed in *E. coli* strain NR698 with genetically permeabilised outer membrane (Ruiz et al., 2005), or in wild type cells of the same genetic background, strain MC4100. Cells were transformed with pPR61 plasmid (Tichá et al., 2017a) coding MBP-FLAG-LacYTM2-Trx expression construct (Strisovsky et al., 2009). NR698 and MC4100 *glpG* knock-out mutants were used as controls.

Bacterial cells stored at -80 °C (Ultra Low; Sanyo) were streaked on an agar plate containing kanamycin (50 µg/ml) and incubated overnight at 37 °C (IPP 400; Memmert). On the next day, 10 ml of **LB medium** with 50 µg/ml kanamycin were inoculated with a single bacterial colony. Cells were incubated overnight at 37 °C, 220 rpm (Innova 44/44R; New Brunswick Scientific) and the overnight culture was passaged into fresh 10 ml of **LB medium** with 50 µg/ml kanamycin. Cells were grown to OD_{600} of 0.6 (measured by BioSpectrometer® kinetic; Eppendorf). In the meantime, dilution series of inhibitor in DMSO was prepared. 3 µl of inhibitor were mixed with 147 µl of the cell culture and pre-incubated 15 min at room temperature in order to form the enzyme-inhibitor complex. Expression of a tagged substrate MBP-Flag-LacYTM2-Trx-His (Strisovsky et al., 2009) was induced by 16.7 µl of 10 mM L-rhamnose. Substrate was expressed and cleaved for 4 hours at 25 °C, 1 000 rpm (Thermomixer; Eppendorf). OD_{600} of samples was measured by microplate reader TECAN Infinite® M1000 (TECAN) and cells were harvested by centrifugation (10 min, 6 000 × g; Eppendorf® 5424). Supernatant was disposed of and cell pellet was resuspended to the same concentration of biomass (calculation based on previously measured OD_{600}) in **SDS sample buffer** with 20 mM $MgCl_2$ and Pierce universal nuclease (1 µl/ml; Thermo). Samples were loaded on a 4-20 % polyacrylamide gel and electrophoresis ran for 45 min at 220 V (Consort EV202; Sigma-Aldrich). Proteins were transferred onto methanol-activated Immobilon®-FL PVDF membrane (Thermo) by western blotting during 2 h at 100 V (45 min, 200 V; Consort EV202; Sigma-Aldrich). The blot was washed with PBS and blocked by Casein Blocker (Thermo). Substrate was labelled overnight at 4 °C by a rabbit antibody against FLAG tag (α-DYKDDDDK; Cell Signaling Technology) diluted 1:4000 in 0.1 % (v/v) TWEEN-20. On the next day, the blot was washed with PBS containing 0.1 % (v/v) TWEEN-20 and incubated with a secondary antibody, IRDye® 800CW donkey anti-Rabbit IgG (Invitrogen), which was diluted 1:10 000 in Casein Blocker (Thermo) and 0.1 % (v/v) TWEEN® 20 (Sigma). Blot was again washed with PBS containing 0.1 % (v/v) TWEEN® 20 (Sigma) and pure PBS. Dry membrane was scanned on Odyssey® CLx (LI-COR) to measure fluorescence emission at 700/800 nm. Cleavage data were analysed as described in 4.4.2.

4.4. Data analysis

4.4.1. *In vitro* assays

Substrate cleavage data were analysed using GraphPad Prism®. An average of 2-3 blank values of a substrate alone was subtracted at each time point. Initial velocity (v_i) of a substrate cleavage was estimated as a slope of a linear regression fitted to the initial linear part of the cleavage data. For the inhibition assays, an average out of 2-3 v_i values of uninhibited reaction was considered to represent 100 % of the reaction velocity and relative rates of inhibited reactions were calculated. Initial reaction rates plotted against inhibitor concentrations were fitted with four parametric equation:

$$Relative\ Initial\ Reaction\ Rate = Bottom + \frac{Top - Bottom}{1 + \frac{concentration^{Hill\ Slope}}{IC_{50}{}^{Hill\ Slope}}}$$

– where *Bottom* and *Top* represent the bottom and top plateau of the curve; *Hill Slope* represents the slope of the curve. Outliers (Motulsky and Brown, 2006) were excluded and IC_{50} was re-calculated.

4.4.2. *In vivo* assays

Fluorescence of individual bands on western blot was analysed using Image Studio Lite. Individual bands were hand-marked, close surrounding area was considered as a background and an average value of its fluorescence was subtracted from the signal. Substrate cleavage was determined as the ratio of a fluorescence of a cleaved fragment to a sum of fluorescence of both, cleaved and uncleaved forms of the substrate. DMSO treated control was considered as 100 % cleavage control, and relative substrate cleavage for each inhibitor concentrations was calculated accordingly. IC_{50} was calculated in GraphPad Prism 7. The data were fitted with the same parametric curve as in 4.4.1. Data didn't contain outliers (Motulsky and Brown, 2006).

4.5. Analysis of the computational model

The computational model (3.6.) of compound **21**/**22** in a complex with GlpG was analysed by calculation of the root-mean-square deviation (RMSD) of inhibitors and determination of inhibitor-enzyme interactions.

Parent inhibitor (compound **9;** PDB code 5MT6) and the modelled one (compound **21**/**22**) were aligned in PyMOL and RMSD of atomic positions of all (46) corresponding non-hydrogen atoms was calculated. Atoms present only in one of the inhibitors (the second branch of the tail substituent of compound **21**/**22** and arginine in the P_5 position of 5MT6 inhibitor) were omitted from the analysis.

Ligplot+ was used in the analysis of inhibitor-enzyme interactions. The distance cut offs between hydrogen-acceptor and donor-acceptor of hydrogen bonds were 2.7 Å and 3.35 Å,

respectively; the minimum and maximum contact distances for van der Waals interactions were 2.9 Å and 3.9 Å, respectively; and the analyses was performed for any atomic contacts.

5. Results

5.1. Purified GlpG is catalytically active and cleaves artificial substrate *in vitro*.

In preparation for testing *in vitro* inhibition (5.2. and 5.3.), recombinant GlpG from *E. coli* had to be expressed and solubilised into micelles. The process is well established in our laboratory (Tichá et al., 2017a, 2017b). A new batch of GlpG was successfully expressed, purified and solubilized which is demonstrated by the enzyme's catalytic activity *in vitro* (Figure 13). Initially, the reaction proceeds linearly for several minutes with a high coefficient of determination (R^2 = 0.9996). The linear character of the initial reaction rate is utilized in estimating IC_{50} of inhibitors (4.4.1.; Figure 14).

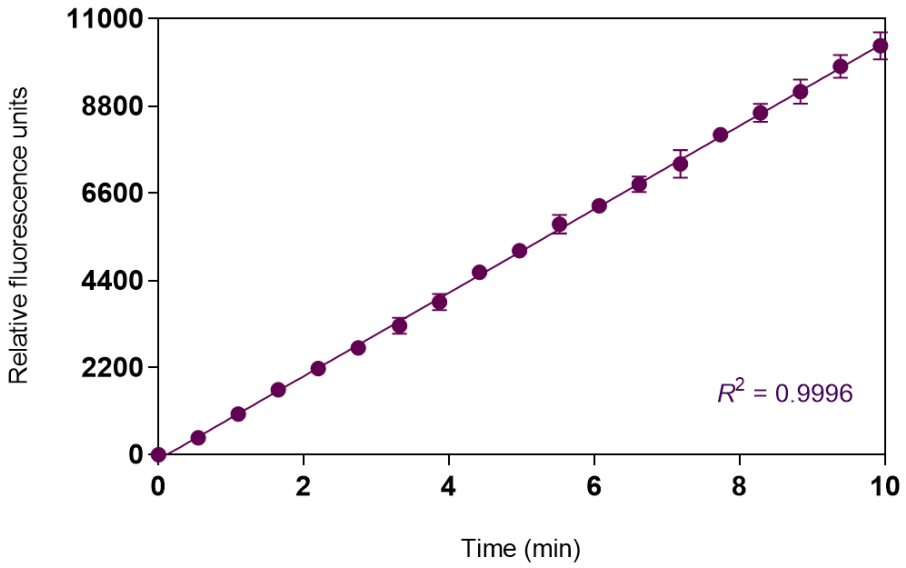

Figure 13. Newly expressed and purified GlpG is catalytically active. Cleavage of a FRET-based fluorogenic substrate KSp35 (10 µM; 3.4.) by GlpG (320 nM) in detergent micelles (0.05 % (w/v) DDM) was measured as described in 4.1.3. The data are fitted with linear regression (R^2 = 0.9996). The experiment was performed many times (>10) and a representative result is shown. The data were measured in triplicates, represented here as an average with standard deviations (SD).

5.2. Further growth of the tail substituents of peptidyl α-ketoamide inhibitors increases potency *in vitro* and *in vivo*.

The structure alignment of the inhibitor-GlpG co-crystals (Figure 12) revealed that the prime side of the active site of rhomboid offers further potential for a growth of the tail substituents of peptidyl α-ketoamide inhibitors (Tichá et al., 2017a). To explore this opportunity, the peptidyl α-ketoamides containing VRHA sequence and various tail substituents (R_T) with increasing size, branching and differing chemical nature (compounds **16 – 22;** 3.5.) were screened for *in vitro* inhibitory

potency (Figure 15). The reported apparent *in vitro* IC_{50} values were calculated from the relative initial rates of reactions in the presence of increasing concentrations of inhibitors, such as presented in Figure 14.

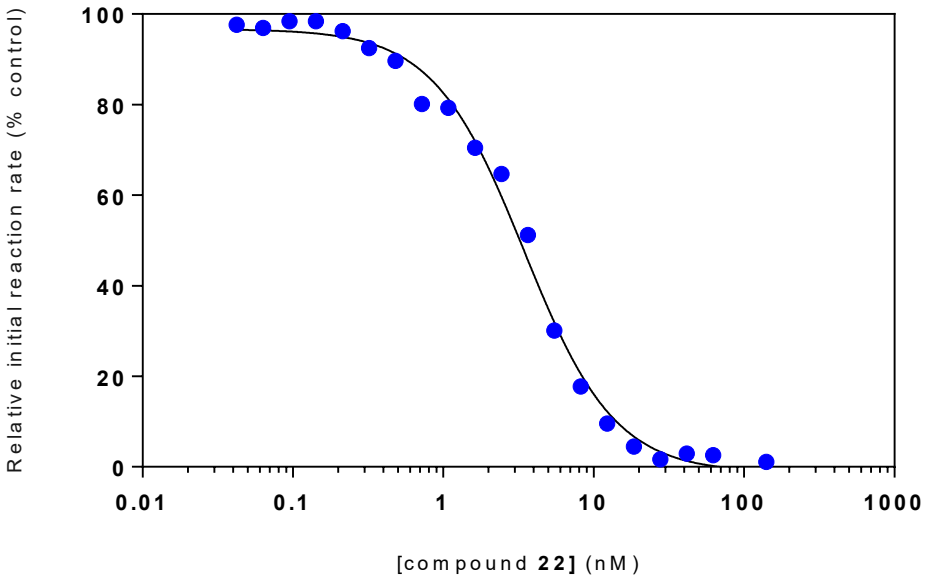

Figure 14. Measurement of apparent *in vitro* IC_{50}. Inhibitor potency was determined based on the change of the reaction rate of cleavage of a FRET-based fluorogenic substrate KSp96 (25 µM; 3.4.) by GlpG (3 nM) upon treatment with α-ketoamide inhibitors as described in 4.2. The relative initial reaction rates (v_i; % of non-inhibited reaction) were plotted against the inhibitor concentration. The half maximal inhibitory concentration (IC_{50}) was calculated as the middle point of top and bottom plateaus of initial reaction rate as described in 4.4.1. Measurements were performed in 50 mM potassium phosphate (pH 7.4), 20 % (v/v) glycerol, 150 mM NaCl, 0.05 % (w/v) DDM, 0.05 % (w/v) PEG8000 and 10 % (v/v) DMSO. The substrate cleavage was followed as an increase of fluorescence (λ_{ex} = 335 nm, λ_{em} = 493 nm).

Compound **16** has an apparent *in vitro* IC_{50} of (66 ± 9.5) nM based on three measurements (Figure 15). Lengthening the tail substituents slightly increases inhibitory potency (compound **17**, IC_{50} = (56 ± 4) nM). Introduction of oxygen into the tail substituent decreases inhibitory potency (compound **18**, IC_{50} = (120 ± 13) nM). Branching of the tail substituent initially decreases inhibitory potency (compound **19**, IC_{50} = (2778 ± 251) nM), but lengthening the branches gradually leads to improvement of the potency (compound **20**; IC_{50} = (33 ± 2.1) nM); compound **21**, IC_{50} = (11 ± 0.27) nM); compound **22**, IC_{50} = (3.9 ± 0.75) nM).

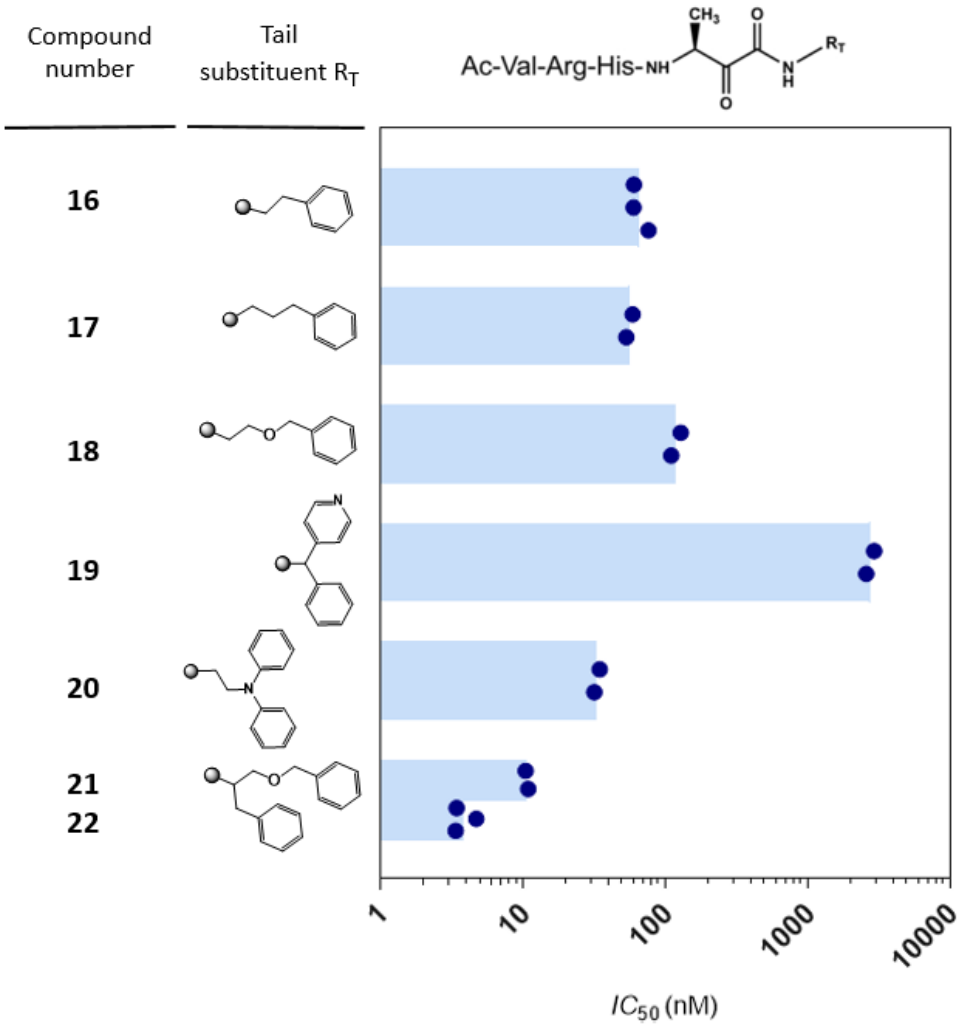

Figure 15. Screen of tail substituents (R_T) of peptidyl α-ketoamides. Peptidyl α-ketoamides based on VRHA sequence were substituted at the amide nitrogen of the ketoamide group with various R_T substituents (top). Apparent *in vitro* IC_{50} values of inhibitors were assessed as relative v_i of cleavage of a FRET-based fluorogenic substrate Ksp96 (25 µM; 3.4.) by GlpG (3 nM) as described in 4.2. Circles represent individual experiments, columns show average of these data. Compounds **21** and **22** are diastereomers (3.5.).

Next, compounds **16**, **18**, and **22** were tested in the live *E. coli* strain NR698 with genetically permeabilised membrane (Ruiz et al., 2005), such as performed previously (1.4.1.) (Tichá et al., 2017a). In the experiment, endogenous GlpG cleaved chimeric substrate MBP-Flag-LacYTM2-Trx-His (Strisovsky et al., 2009) in the presence of increasing concentrations of inhibitor. The apparent *in vivo* IC_{50} was determined as the ratio of cleaved and total substrate in cell lysates compared to a DMSO control (Figure 16). The best compound from the *in vitro* screen (Figure 15), compound **22**, was chosen for testing (Figure 16A). Additionally, to explore the effect of branching, compounds **16** and **18** were also tested (Figure 16B and C). The tail substituents of compounds **16** and **18** represent individual branches of the tail substituent of compound **22**.

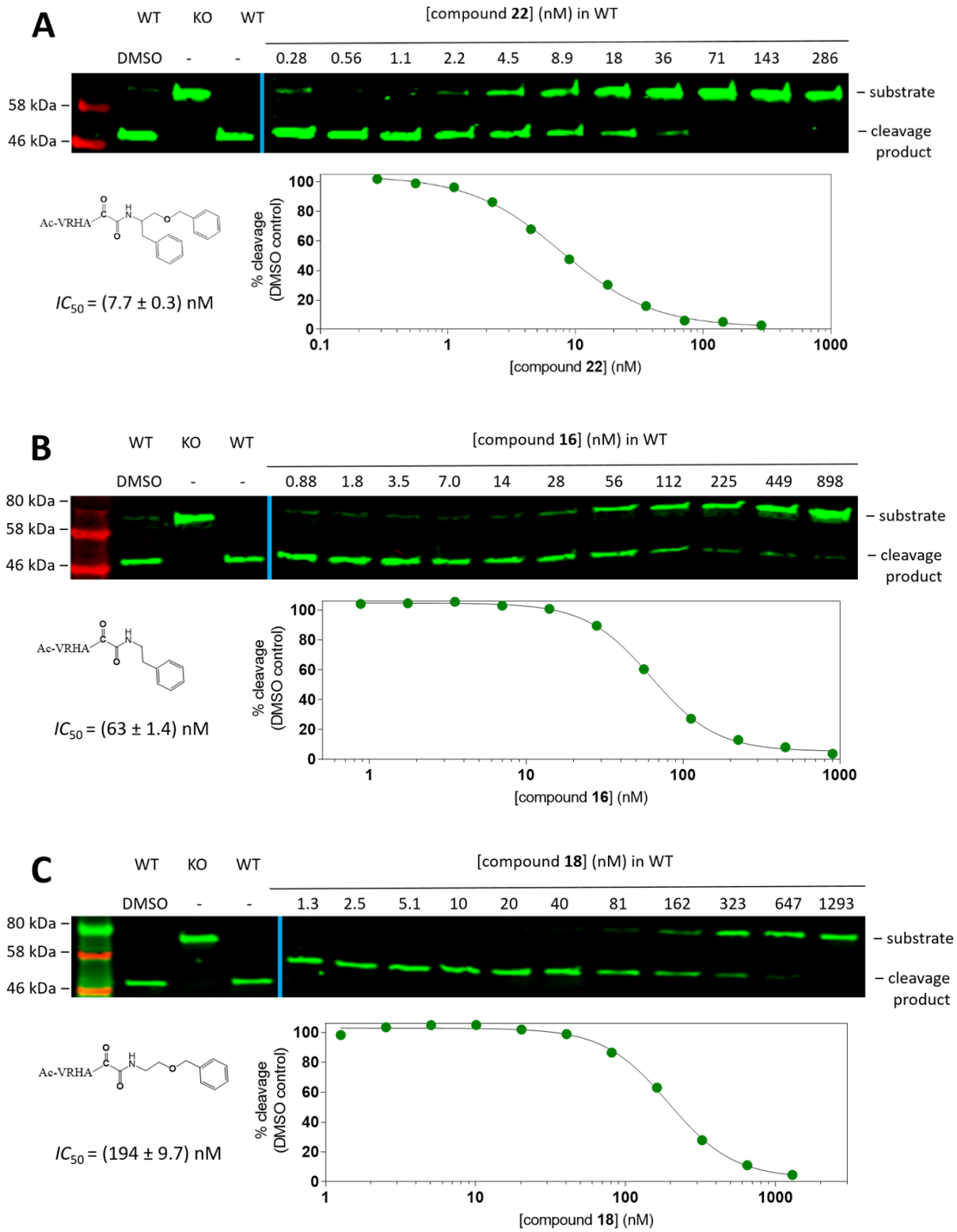

Figure 16. Inhibitors are potent *in vivo*. Compounds **22** (**A**), **16** (**B**), and **18** (**C**) are potent inhibitors of *E. coli* rhomboid GlpG. The chimeric substrate MBP-Flag-LacYTM2-Trx-His (Strisovsky et al., 2009) was expressed in live *E. coli* strain NR698 (Ruiz et al., 2005). Substrate cleavage in cell lysates after 4 h of co-incubation with increasing concentrations of inhibitor was assessed by immunoblotting for Flag tag (4.3.) and quantified as described in 4.4.2. Continued on the next page.

Figure 16. Continuation.
DMSO control was considered the 100 % cleavage control, relative cleavage in the samples was plotted against inhibitor concentration and IC_{50} was calculated as described in 4.4.2. Experiments were carried out 2 – 3 times per each compound and representative results are shown. Parts of the western blots were horizontally flipped (interface highlighted by blue line), so that the inhibitor concentration in samples is in ascending order, thus matching the graphs below the western blots. WT – GlpG wild-type; KO – GlpG knock-out.

5.3. Truncating the peptidyl part of inhibitor decreases potency.

The effect of the amino acids in the $P_4 – P_2$ positions (Figure 8) of peptidyl α-ketoamides on inhibitor potency was explored. The peptidyl part of the most effective compounds in *in vitro* (Figure 15), compounds **21** (IC_{50} = (11 ± 0.27) nM) and **22** (IC_{50}= (3.9 ± 0.75) nM), was gradually truncated, yielding compounds **23** – **25**. Compounds **21** and **22** are diastereomers (probably in the His in the P_2 position) bearing the same tail substituent, whereas compounds **23** – **25** are mixtures of diastereomers (3.5.).

Employing the same *in vitro* inhibition assay as in section 5.2., apparent *in vitro* IC_{50} (Figure 14) of these compounds were measured. Truncation of the peptidyl part leads to a dramatic decrease in inhibitor potency. Deletion of the Val in the P_4 position causes only a mild effect on inhibitor potency (compound **23**, IC_{50} = (42 ± 15) nM), whereas the effect of omitting the Arg in the P_3 position is more pronounced (compound **24**, IC_{50} = (782 ± 11) nM). Omitting the His in the P_2 position increases IC_{50} from the nanomolar to micromolar range (compound **25**, IC_{50} = (368 501 ± 5 650) nM) (Figure 17).

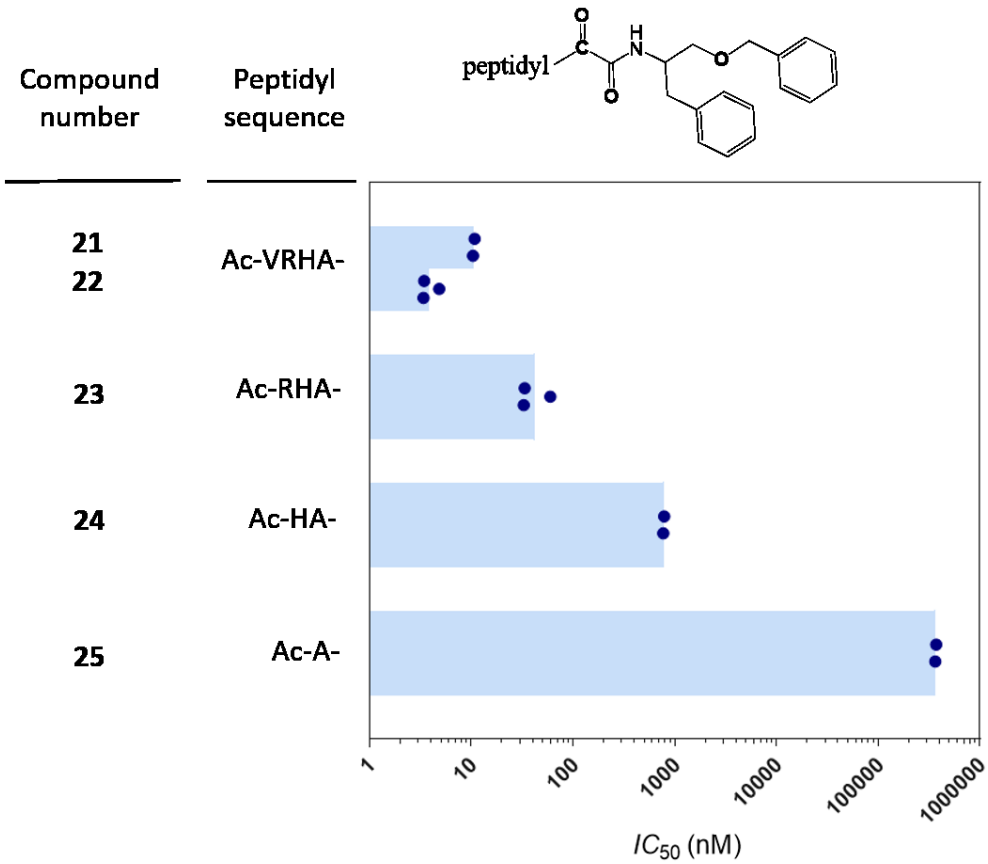

Figure 17. Effect of truncation of peptidyl part of inhibitors. Apparent *in vitro* IC_{50} of inhibitors was assessed as relative v_i of cleavage of a FRET-based fluorogenic substrate KSp96 (25 µM; 3.4.) by GlpG (3 nM) as described in 4.2. Circles represent results of individual experiments, the columns show their average. Precise concentration of compound **25** could not be measured precisely, as there are no residues left for quantitative amino acid analysis (its IC_{50} is probably lower). Compounds **21** and **22** are diastereomers (3.5.). The experiments were carried out in 50 mM potassium phosphate (pH 7.4), 20 % (v/v) glycerol, 150 mM NaCl, 0.05 % (w/v) DDM, 0.05 % (w/v) PEG8000 and 10 % (v/v) DMSO. The reaction progress was followed as an increase of fluorescence upon substrate cleavage (λex = 335 nm, λem = 493 nm).

5.4. Inhibitors penetrate the outer membrane of *E. coli* and inhibit endogenous GlpG.

The addition of a bigger hydrophobic branched tail substituent and omission of the polar arginine in the P_5 position might improve cell membrane permeability of compound **22** in comparison to compound **11**. *E. coli* MC4100 cells with intact outer membranes were co-incubated with increasing inhibitor concentrations and cleavage of MBP-FLAG-LacYTM2-Trx chimeric substrate was monitored (Strisovsky et al., 2009) as in 5.2. Compound **22** inhibited endogenous GlpG in *E. coli* MC4100 (Figure 18).

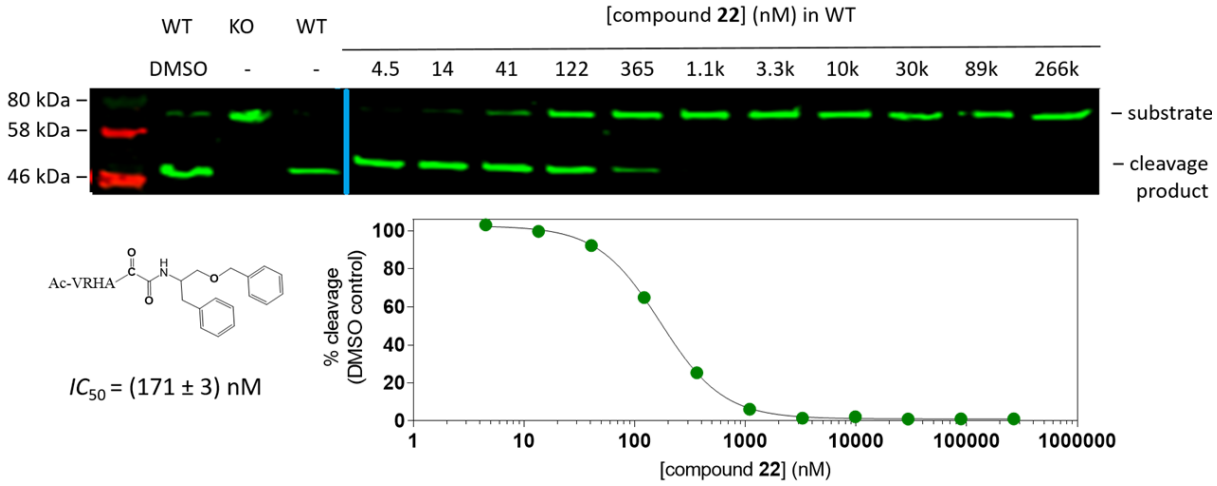

Figure 18. Compound 22 crosses the outer *E. coli* membrane and inhibits endogenous GlpG. The chimeric substrate MBP-Flag-LacYTM2-Trx-His (Strisovsky et al., 2009) was expressed in live *E. coli* strain MC4100. Substrate cleavage in cell lysates, after 4 h of co-incubation in the presence of increasing concentration of inhibitor, was assessed by immunoblotting for Flag tag and fluorescence quantification as described in 4.3. DMSO control was considered a 100 % cleavage control, relative cleavage in the samples was plotted against inhibitor concentration and IC_{50} was calculated as described in 4.4.2. Experiment was carried out twice and a representative result is shown. Part of the western blot was horizontally flipped (interface highlighted by blue line), so that the inhibitor concentration in samples is in ascending order, thus matching the graph below the western blots. WT – GlpG wild-type; KO – GlpG knock-out.

5.5. Branched tail substituent of compound 21/22 binds into the prime side of GlpG active site.

In order to explain the potency of compounds **21** and **22,** a model of GlpG with bound compound **21/22** was built based on the co-crystal model of GlpG with compound **9** (PDB code 5MT6) (Tichá et al., 2017a) as described in 3.6. Compounds **21** and **22** are diastereomers of unknown configuration, probably differing in the configuration of histidine in P_2 position (3.5.). In the model, the isomerisation of P_2 histidine is the same as in the parent crystal (PDB code 5MT6). Compound **21/22** is similar to compound **9**, differing only in the terminal arginine in the P_5 position and one branch of the tail substituent (Figure 11 and Figure 15). Accordingly, the arginine was deleted, the additional branch modelled, and molecular dynamics simulation and energy optimisation of the enzyme-inhibitor complex were carried out (3.6.). The conformer with the lowest energy (Figure 19) was chosen for further analysis.

The branched tail substituent of compound **21/22** binds into the prime side of GlpG active site. Alignment of all (46) shared non-hydrogen atoms of inhibitors between the model and parent inhibitor

(PDB code 5MT6) resulted in a RMSD of only 0.624 Å. . This means that, in this model, compound **9** and its diastereomeric equivalent of compound **21/22** bind in virtually the same pose (Figure 19A).

Next, Ligplot+ was used to analyse interactions between GlpG and the tail substituent of compound **21/22** (4.5.). Modelled inhibitor binds into the same cavity as the parent one (Figure 19b and Figure 12C), which is a proposed S'_2 subsite (Vinothkumar et al., 2013). There is a slight shift of the original phenylethyl branch of the inhibitor which newly engages in van der Waals interactions with residues Trp157 from TMD2, Val204 from TMD4, and no longer interacts with residue Phe245 from the L5 loop. Residue Phe245 is now in contact with the newly modelled benzyloxy branch. The van der Waals interactions between the original phenylethyl branch of the inhibitor with residues Asn154 from TMD2, Tyr205 from TMD4, Trp236 from TMD5, and residue His254 from TMD6 are intact. The newly modelled branch engages in van der Waals interactions with residues Met149, His150, Phe153 from TMD2 and residues Phe245, Met247, Met249 from L5 loop. No hydrogen bond is formed by the oxygen from the newly modelled benzyloxy branch. The weak N-π bond between Trp236 and phenyl ring of the inhibitor is intact and no π... π stacking is observed.

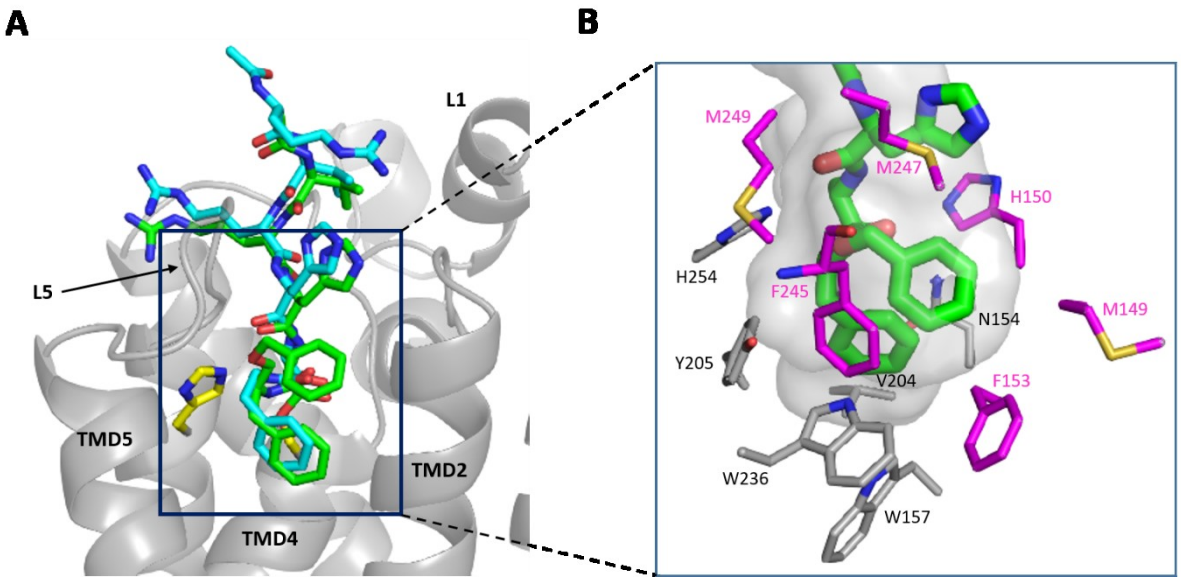

Figure 19. Model of GlpG-compound 21/22 complex. A) Structure alignment of the modelled (3.6.) and parent complex (PDB code 5MT6). Compound **21/22** (green sticks) binds in a similar way to the compound **9** (cyan sticks). The new benzyloxy branch of the tail substituent of compound **22** binds in-between TMD2, TMD5, L5 loop, and the lipid environment of a membrane. Enzyme of the modelled complex is shown as grey cartoon representation and catalytic dyad as yellow sticks. **B)** Van der Waals contacts between the enzyme and inhibitor were identified by Ligplot+ (4.5.). The enzyme cavity is shown as inversed surface, inhibitor as green sticks and side chains of residues engaged in van der Waals interactions as grey (interacting with the original branch of the inhibitor tail) and magenta (interacting with the new branch) sticks, respectively.

6. Discussion

Rhomboids are potential therapeutic targets for many diseases, such as malaria, amoebiasis, Parkinson's disease, various tumour malignancies and diabetes. However, more cell biology tools are needed for unravelling functions of rhomboid proteases (reviewed in 1.2.). Peptidyl α-ketoamides are the most potent and promising class of rhomboid inhibitors (reviewed in 1.4.). They are selective, based on a pharmacologically complaint chemotype (Njoroge et al., 2008), and two co-crystal structures of peptidyl α-ketoamide soaked into crystals of *E. coli* rhomboid GlpG were solved (1.4.1.) (Tichá et al., 2017a). In my thesis I focused on further development of peptidyl α-ketoamides as inhibitors of rhomboid proteases.

First, *E. coli* rhomboid GlpG was successfully expressed, solubilised, and purified as demonstrated by its catalytic activity (Figure 13). GlpG is a membrane protein and in all the *in vitro* experiments presented (Figure 13, Figure 14, Figure 15, and Figure 17) it is solubilised in micelles formed by DDM. The micellar concentration, effective concentrations of reactants, and initial reaction rate all depend on the detergent concentration (Tichá et al., 2017b). Therefore, for better comparison of the results, the detergent concentration was set uniformly for all experiments to be 0.05 % (w/v) DDM. The same concentration of detergent was already used previously, in the initial discovery of peptidyl α-ketoamides as rhomboid inhibitors (Tichá et al., 2017a).

The crystal structures of β-lactam L29 (PDB code 3ZMI), isocoumarin S016 (PDB code 3ZEB), and compounds **9** and **10** (PDB codes 5MT6 and 5MTF, respectively) revealed, that the prime side of rhomboid active site is relatively malleable and able to accommodate inhibitors of various sizes and chemical nature (Figure 12) (Tichá et al., 2017a). β-lactam and isocoumarin occupy more space at the prime side of the active site than tail substituents of compound **9** and **10**, suggesting that a bigger or branched tail substituent of the ketoamide nitrogen could be accommodated into the enzyme. Moreover, the whole transmembrane domain of a substrate is important for cleavage, suggesting that rhomboid forms stabilizing interactions extending relatively far into the prime side of the active site (Tichá et al., 2017b), which was also predicted by a computational simulation (Shokhen and Albeck, 2017). This together led to an *in vitro* screen for potency of peptidyl α-ketoamides based on VRHA sequence and tail substituents of increasing size, branching, and different chemical nature (Figure 15). The previous screen revealed that GlpG prefers hydrophobic tail substituents (Figure 11) (Tichá et al., 2017a), therefore all tested compounds (**16**-**22**) bore tail substituents of a hydrophobic nature and at least one phenyl ring. Elongating the tail (compounds **16** and **17**) slightly increases the potency, which is in agreement with previously published data (Figure 11) (Tichá et al., 2017a). However, the introduction of oxygen into the tail of compound **18** had a negative effect on inhibitor potency.

An oxygen or a nitrogen was also introduced into compounds **19** – **22**, but usually into varying positions. Overall, this screen was too small to draw any other conclusion about the effect of heteroatoms in the tail substituents of peptidyl α-ketoamides. Next, the effect of branching was explored. The compound **19** contains short branches and has relatively low potency. Its low potency could be explained by hypothetical sterical clashes of the newly introduced phenyl ring with catalytic residue His254 (hypothesis based on the crystal structures 5MT6 and 5MTF; (Tichá et al., 2017a)). Further extensions of the branches of compounds **20** – **22** improved the inhibitory potency and resulted in the highest *in vitro* inhibitory potency reported to date – apparent *in vitro* IC_{50} of compound **22** is (3.9 ± 0.8) nM. The two most potent inhibitors, compound **22** (IC_{50} = (3.9 ± 0.8) nM) and compound **21** (IC_{50} = (11 ± 0.3) nM) are diastereomers whose absolute configurations have not been characterised at the moment, but it is most likely affecting the His in the P2 position. Additional experiments, such as chiral column chromatography or structural studies, should be carried out to determine the stereoconfiguration of these inhibitors. Intriguingly, the tail substituent of compounds **21** and **22** is virtually built of the tail substituents of compounds **16** and **18**, pointing to a positive and additive effect of tail combination and branching. Additionally, the tail substituent of compounds **21** and **22** contains the same oxygen in its tail substituent as compound **18**. The introduction of oxygen in compound **18** had a negative effect on inhibitor potency, which suggests that substituting oxygen for carbon in the tail substituent of compounds **21** and **22** might improve the potency of these compounds even further.

Previously reported apparent *in vitro* IC_{50} (Figure 11) and IC_{50} values presented in this thesis (Figure 15 and Figure 17) are not directly comparable, since different substrates, substrate concentration and enzyme concentration were used in this thesis compared to previously reported inhibitors (Tichá et al., 2017a). Enzyme concentration inherently plays a role in experimental determination of IC_{50} (Cheng and Prusoff, 1973), therefore GlpG concentration was deliberately low in this thesis (3 nM) compared to the previous study (400 nM) (Tichá et al., 2017a) in order to characterise apparent IC_{50} of very potent inhibitors. In order to obtain high quality data even with lower enzyme concentration, KSp96 as a better GlpG substrate (Tichá et al., 2017a) was used to follow the reaction progress. Taken together, this means that compounds **1**-**15** would probably perform better in experimental conditions used in this thesis in comparison to what is reported in Figure 11. However, Tichá *et al.* also measured true inhibition constant (K_i) of their best inhibitor, compound **11,** which was (45 ± 8) nM. K_i is, unlike IC_{50}, a thermodynamic constant and since $K_i \leq IC_{50}$ (Cheng and Prusoff, 1973), compound **22** (IC_{50} = (3.9 ± 0.8) nM) is *in vitro* more potent than compound **11**.

Peptide character of an inhibitor is generally undesired in drug design, because peptide bonds represent low stability against proteolysis (Gentilucci et al., 2010). Therefore, the effect of truncating the Ac-VRHA peptidyl part of peptidyl α-ketoamides on inhibitor potency was explored as the next

step. Apparent *in vitro* IC_{50} values of inhibitors with a gradually truncated peptidyl part were measured (Figure 17). The residues in the P2-P4 positions play an important role in the inhibitor binding affinity (Figure 17), such as is in agreement with previously reported data (Tichá et al., 2017a; Zoll et al., 2014). Excluding the valine in the P_4 (compound **23**) and arginine in the P_3 (compound **24**) position, respectively, increased the IC_{50} by an approximately one order of magnitude in each case. The histidine in P_2 position contributes more, as omitting this residue decreased the IC_{50} ~500-fold (compound **25**). However, it should be noted that the concentration of the Ac-A-core inhibitor (compound **25**) is not precise, as it could not be measured by routine quantitative amino acid analysis. True concentration and related IC_{50} of this inhibitor is probably lower, making the significance of the histidine in P_2 position also smaller. In conclusion, the residues in P_2-P_4 positions play a substantial role in the inhibitor binding affinity to GlpG, with the residues closer to the cleavage site being gradually more important.

The significance of the peptidyl part is relatively in agreement with data reported previously (Figure 11B) (Tichá et al., 2017a). Excluding the valine in the P_4 position (compound **13**) and the arginine in the P_3 position (compound **14**), respectively, also increased the IC_{50} by approximately one order of magnitude each, which is in agreement with the data reported here (Figure 17). The effect of omitting the histidine in the P_2 position is ambiguous between the studies. Previously, omitting this residue increased the IC_{50} ~16-fold (compound **15**, Figure 11B), whereas here it was ~500-fold (compound **25**, Figure 17). However, as mentioned above, the true significance of the histidine in P_2 position reported here is most probably smaller. Moreover, the different concentrations of enzyme (400 nM vs 3 nM) need to be taken into account when comparing the results with Tichá *et al.* Enzyme concentration affects the IC_{50} differently in different ranges of inhibitor potency (Shoichet, 2006). The lower concentration of the enzyme results in more precise measurements and higher contrast in-between the potency of the inhibitors, which could partially cause the observed discrepancy of omitting the histidine in the P_2 position. Last but not least, Tichá *et al.* report relatively big SD in some of their measurements, thus the relative importance of omitting individual residues cannot be compared strictly, but more like a general trend.

In order to measure inhibitor potency in the natural environment of rhomboid – a membrane of a living cell – compounds **16**, **18**, and **22** were tested for inhibition in *E. coli* (Figure 16). Out of the tested inhibitors, compound **22** is the best *in vitro* inhibitor, whereas compounds **16** and **18** are individual virtual constituents of the tail substituent of compound **22**. The relative potency of compounds **16**, **18**, and **22** in *in vivo* setting follows the same trend as in *in vitro* one (Figure 15). *In vivo* experiments were carried out in the same way as reported previously (compounds **9** – **11**) (Tichá et al., 2017a), using a mutant *E. coli* strain with a genetically permeabilised outer membrane, making the results directly comparable. Compounds **16**, **18**, and **22** are potent nanomolar inhibitors of GlpG

(Figure 16). The apparent *in vivo* IC_{50} value of compound **22** is in the same range as for compounds **9** – **11** (Tichá et al., 2017a). Compounds **16** and **18** are worse *in vivo* inhibitors than previously tested compounds **9** – **11** (Tichá et al., 2017a)**.** The discrepancy of compound **22** between the improved potency *in vitro* but similar potency *in vivo* in comparison to compounds **9**-**11** could be explained by at least two phenomena. First, the inhibitor potency could have reached the concentration of GlpG in the experiment making virtually all of the inhibitor molecules bind to the enzyme in a stoichiometric manner, making determination of further improvement of inhibitory potency in this assay unfeasible. Experiments in cells with GlpG overexpression could be carried out to test this hypothesis. Secondly, compound **22**, in comparison to compounds **9** – **11**, does not bear a terminal arginine in the P_5 position. Arginine in the P_5 position was shown to be not important for inhibitor potency of peptidyl α-ketoamides *in vitro* (Figure 11B) (Tichá et al., 2017a). However, it has been shown to play a role in binding of peptidyl aldehydes, where it improves the inhibitory potency 5× (Cho et al., 2016). Additionally, the polar residue might behave differently in micelles and natural environment of lipid bilayer, for example interact with the polar heads of lipids and/or affect outer membrane permeability and enzyme accessibility for the inhibitor (Li et al., 2013).

The effect of addition of a bigger hydrophobic branched tail substituent and omitting the polar arginine in P_5 position on cell membrane permeability was investigated in wild type *E. coli* MC4100. Compound **22** inhibited endogenous GlpG in *E. coli* MC4100 with apparent IC_{50} of (179 ± 6) nM and (175 ± 6) nM, respectively in two independent experiments (Figure 18), whereas compound **11** has a lower apparent *in vivo* IC_{50} of ~ 110 nM (A. Tichá, not reported). Therefore, further optimisations are still needed in order to obtain an inhibitor that would efficiently cross the outer membrane of Gram-negative bacteria.

In an attempt to explain the potency of compounds **21** and **22,** Ligplot+ (Laskowski and Swindells, 2011) analyses of van der Waals contacts between the tail substituent of compounds and GlpG in a computational model (3.6.) were carried out (Figure 19). Analyses revealed that the new branch of the tail substituent of compound **21**/**22** engages in van der Waals interactions with residues from TMD2 and L5 loop (Figure 19B). TMD2 and L5 loop are proposed to play a role in substrate binding, specifically, L5 loop is to be displaced during substrate docking and TMD2 is proposed to play a role in initial substrate recognition (1.3.2.), indicating that the new branch possibly interacts with some of the residues important for substrate binding. Indeed, TMD2 is proposed to be a part of the rhomboid exosite and the computational model suggests that the new branch of the tail substituents of compound **21**/**22** is located on top of this site (Figure 19A). Moreover, rhomboid proteases have the same or very similar topology (Figure 3), and it is hypothesised that rhomboids recognise and cleave their substrates in a similar manner (1.3.2.). Therefore, the structure, spatial arrangement, and role of

TMD2 and loop L5 in substrate recognition might be conserved among the rhomboid proteases. This suggests that the binding mode of the new branch of compound **21**/**22** could be translatable to other rhomboids. Last, but not least, the hydrophobic phenyl ring of the new branch of compound **21**/**22** slightly protrudes out of the enzyme (Figure 19B) into the lipid environment, suggesting possible interactions with membrane lipids. These would be enzyme-independent/unspecific which could also be potentially used for future inhibitor design of other rhomboid proteases.

Alignment of inhibitors between the computational model (compound **21**/**22**) and parent inhibitor (compound **9**) (PDB code 5MT6) resulted in a small RMSD of only 0.624 Å (Figure 19A), suggesting that inhibitors bind in the same pose. However, the small RMSD is not surprising, since most of the modelled complex was frozen during the energy optimisation and molecular dynamics simulation (3.6.). Further structural studies (e.g. crystallization) would determine the accuracy of the computational modelling and might provide an answer to compound **21** and **22** isomerisation (3.5.).

Peptidyl α-ketoamides are selective rhomboid inhibitors (Tichá et al., 2017a), likely suitable for cell biological assays. An attractive direction in their further development is to test them against potentially medically relevant rhomboids such as human RHBDL2, RHBDL4, PARL, entamoeba EhROM1, or rhomboids of apicomplexan parasites. At least one substrate of each one of these rhomboid proteases is known (reviewed in 1.2.), which opens up the possibility to base the peptidyl part of the inhibitor on the known sequence of the substrate. Indeed, rhomboids exhibit some specificity in their recognition motif (Strisovsky, 2016b). Another option is to design the tail substituent of ketoamide nitrogen. Composing a branched hydrophobic tail substituent, investigated here, might offer a new way of designing medically or scientifically attractive rhomboid inhibitors.

7. Conclusions

This study aimed to improve potent and selective rhomboid protease inhibitors - peptidyl α-ketoamides - as tools for studying cell biology of these enzymes. *E. coli* rhomboid GlpG was successfully purified and used for *in vitro* screening for potency of tail substituents accommodated at the amide nitrogen of peptidyl α-ketoamides. The screen resulted in the discovery of inhibitors with branched tail substituent as a new, highly potent class of peptidyl α-ketoamides. Potency of compound **22** has been confirmed *in vivo*, but the outer membrane permeabilisation of Gram-negative bacterium was not improved. The importance of peptidyl part of peptidyl α-ketoamides has been confirmed and a computational model was analysed in an attempt to explain the effect of branching on inhibition potency. Analysis of the computational model revealed that the new branch of the tail substituent of compound **22** interacts with TMD2 and loop L5 of the enzyme, and the lipid environment of a membrane. TMD2 and loop L5 are hypothesized to be conserved regions of rhomboid participating in substrate binding, thus the binding pose of the branched tail substituent might be conserved among rhomboids. This rationalizes composing of the branched tail substituent of peptidyl α-ketoamides as a possible way of development of peptidyl α-ketoamide inhibitors for rhomboids other than GlpG.

8. Literature

Adamiec, M., Ciesielska, M., Zalaś, P., and Luciński, R. (2017). Arabidopsis thaliana intramembrane proteases. Acta Physiol. Plant. *39*, 146.

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

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

Akiyama, Y. (2002). Proton-motive force stimulates the proteolytic activity of FtsH, a membrane-bound ATP- dependent protease in Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. *99*, 8066–8071.

Ali, M., Nelson, A.R., Lopez, A.L., and Sack, D.A. (2015). Updated Global Burden of Cholera in Endemic Countries. PLoS Negl. Trop. Dis. *9*.

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

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

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

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

Baxt, L.A., Baker, R.P., Singh, U., and Urban, S. (2008). An Entamoeba histolytica rhomboid protease with atypical specificity cleaves a surface lectin involved in phagocytosis and immune evasion. Genes Dev. *22*, 1636–1646.

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

Bergbold, N., and Lemberg, M.K. (2013). Emerging role of rhomboid family proteins in mammalian biology and disease. Biochim. Biophys. Acta *1828*, 2840–2848.

Bernkopf, D.B., Jalal, K., Brückner, M., Knaup, K.X., Gentzel, M., Schambony, A., and Behrens, J. (2018). Pgam5 released from damaged mitochondria induces mitochondrial biogenesis via Wnt signaling. J. Cell Biol. *217*, 1383–1394.

Bondar, A.-N., del Val, C., and White, S.H. (2009). Rhomboid protease dynamics and lipid interactions. Struct. Lond. Engl. 1993 *17*, 395–405.

Brossier, F., Jewett, T.J., Lovett, J.L., and Sibley, L.D. (2003). C-terminal processing of the toxoplasma protein MIC2 is essential for invasion into host cells. J. Biol. Chem. *278*, 6229–6234.

Brossier, F., Jewett, T.J., Sibley, L.D., and Urban, S. (2005). A spatially localized rhomboid protease cleaves cell surface adhesins essential for invasion by Toxoplasma. Proc. Natl. Acad. Sci. U. S. A. *102*, 4146–4151.

Brown, M.C., Abdine, A., Chavez, J., Schaffner, A., Torres-Arancivia, C., Lada, B., JiJi, R.D., Osman, R., Cooley, J.W., and Ubarretxena-Belandia, I. (2018). Unwinding of the Substrate Transmembrane Helix in Intramembrane Proteolysis. Biophys. J. *114*, 1579–1589.

Buguliskis, J.S., Brossier, F., Shuman, J., and Sibley, L.D. (2010). Rhomboid 4 (ROM4) affects the processing of surface adhesins and facilitates host cell invasion by Toxoplasma gondii. PLoS Pathog. *6*, e1000858.

Cheng, Y., and Prusoff, W.H. (1973). Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. Biochem. Pharmacol. *22*, 3099–3108.

Cheng, T.-L., Wu, Y.-T., Lin, H.-Y., Hsu, F.-C., Liu, S.-K., Chang, B.-I., Chen, W.-S., Lai, C.-H., Shi, G.-Y., and Wu, H.-L. (2011). Functions of Rhomboid Family Protease RHBDL2 and Thrombomodulin in Wound Healing. J. Invest. Dermatol. *131*, 2486–2494.

Cheng, T.-L., Lai, C.-H., Jiang, S.-J., Hung, J.-H., Liu, S.-K., Chang, B.-I., Shi, G.-Y., and Wu, H.-L. (2014). RHBDL2 Is a Critical Membrane Protease for Anoikis Resistance in Human Malignant Epithelial Cells. Sci. World J. 902987.

Cho, S., Dickey, S.W., and Urban, S. (2016). Crystal Structures and Inhibition Kinetics Reveal a Two-Stage Catalytic Mechanism with Drug Design Implications for Rhomboid Proteolysis. Mol. Cell *61*, 329–340.

Clemmer, K.M., Sturgill, G.M., Veenstra, A., and Rather, P.N. (2006). Functional Characterization of Escherichia coli GlpG and Additional Rhomboid Proteins Using an aarA Mutant of Providencia stuartii. J. Bacteriol. *188*, 3415–3419.

Costa, M.I., Cerletti, M., Paggi, R.A., Troetschel, C., De Castro, R.E., Poetsch, A., and Gimenez, M.I. (2018). Haloferax volcanii Proteome Response to Deletion of a Rhomboid Protease Gene. J. Proteome Res. *17*, 961–977.

Das, C., Berezovska, O., Diehl, T.S., Genet, C., Buldyrev, I., Tsai, J.-Y., Hyman, B.T., and Wolfe, M.S. (2003). Designed Helical Peptides Inhibit an Intramembrane Protease. J. Am. Chem. Soc. *125*, 11794–11795.

Deas, E., Plun-Favreau, H., Gandhi, S., Desmond, H., Kjaer, S., Loh, S.H.Y., Renton, A.E.M., Harvey, R.J., Whitworth, A.J., Martins, L.M., et al. (2011). PINK1 cleavage at position A103 by the mitochondrial protease PARL. Hum. Mol. Genet. *20*, 867–879.

Dhingra, S., Kowlaski, C.H., Thammahong, A., Beattie, S.R., Bultman, K.M., and Cramer, R.A. (2016). RbdB, a Rhomboid Protease Critical for SREBP Activation and Virulence in Aspergillus fumigatus. MSphere *1*.

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

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

Dulloo, I., Muliyil, S., and Freeman, M. (2019). The molecular, cellular and pathophysiological roles of iRhom pseudoproteases. Open Biol. *9*.

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

Flegr, J., Prandota, J., Sovičková, M., and Israili, Z.H. (2014). Toxoplasmosis--a global threat. Correlation of latent toxoplasmosis with specific disease burden in a set of 88 countries. PloS One *9*, e90203.

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

Gadwal, S., Johnson, T.L., Remmer, H., and Sandkvist, M. (2018). C-terminal processing of GlyGly-CTERM containing proteins by rhombosortase in Vibrio cholerae. PLoS Pathog. *14*, e1007341.

GBD 2015 Disease and Injury Incidence and Prevalence Collaborators (2016). Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet Lond. Engl. *388*, 1545–1602.

Gentilucci, L., De Marco, R., and Cerisoli, L. (2010). Chemical Modifications Designed to Improve Peptide Stability: Incorporation of Non-Natural Amino Acids, Pseudo-Peptide Bonds, and Cyclization. Curr. Pharm. Des. *16*, 3185–3203.

Goel, P., Jumpertz, T., Mikles, D.C., Tichá, A., Nguyen, M.T.N., Verhelst, S., Hubalek, M., Johnson, D.C., Bachovchin, D.A., Ogorek, I., et al. (2017). Discovery and Biological Evaluation of Potent and Selective N-Methylene Saccharin-Derived Inhibitors for Rhomboid Intramembrane Proteases. Biochemistry *56*, 6713–6725.

Goel, P., Jumpertz, T., Tichá, A., Ogorek, I., Mikles, D.C., Hubalek, M., Pietrzik, C.U., Strisovsky, K., Schmidt, B., and Weggen, S. (2018). Discovery and validation of 2-styryl substituted benzoxazin-4-ones as a novel scaffold for rhomboid protease inhibitors. Bioorg. Med. Chem. Lett. *28*, 1417–1422.

Gsaller, F., Hortschansky, P., Furukawa, T., Carr, P.D., Rash, B., Capilla, J., Müller, C., Bracher, F., Bowyer, P., Haas, H., et al. (2016). Sterol Biosynthesis and Azole Tolerance Is Governed by the Opposing Actions of SrbA and the CCAAT Binding Complex. PLoS Pathog. *12*, e1005775.

Haft, D.H., and Varghese, N. (2011). GlyGly-CTERM and Rhombosortase: A C-Terminal Protein Processing Signal in a Many-to-One Pairing with a Rhomboid Family Intramembrane Serine Protease. PLoS ONE *6*.

Hostaš, J., and Řezáč, J. (2017). Accurate DFT-D3 Calculations in a Small Basis Set. J. Chem. Theory Comput. *13*, 3575–3585.

Hwang, J., Ribbens, D., Raychaudhuri, S., Cairns, L., Gu, H., Frost, A., Urban, S., and Espenshade, P.J. (2016). A Golgi rhomboid protease Rbd2 recruits Cdc48 to cleave yeast SREBP. EMBO J. *35*, 2332–2349.

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

Kantor, M., Abrantes, A., Estevez, A., Schiller, A., Torrent, J., Gascon, J., Hernandez, R., and Ochner, C. (2018). Entamoeba Histolytica: Updates in Clinical Manifestation, Pathogenesis, and Vaccine Development. Can. J. Gastroenterol. Hepatol. *2018*.

Kateete, D.P., Katabazi, F.A., Okeng, A., Okee, M., Musinguzi, C., Asiimwe, B.B., Kyobe, S., Asiimwe, J., Boom, W.H., and Joloba, M.L. (2012). Rhomboids of Mycobacteria: characterization using an aarA mutant of Providencia stuartii and gene deletion in Mycobacterium smegmatis. PloS One *7*, e45741.

Khalid, M., Idichi, T., Seki, N., Wada, M., Yamada, Y., Fukuhisa, H., Toda, H., Kita, Y., Kawasaki, Y., Tanoue, K., et al. (2019). Gene Regulation by Antitumor miR-204-5p in Pancreatic Ductal Adenocarcinoma: The Clinical Significance of Direct RACGAP1 Regulation. Cancers *11*, 327.

Kinch, L.N., and Grishin, N.V. (2013). Bioinformatics perspective on rhomboid intramembrane protease evolution and function. Biochim. Biophys. Acta BBA - Biomembr. *1828*, 2937–2943.

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

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

Kumar, A., Gibbs, J.R., Beilina, A., Dillman, A., Kumaran, R., Trabzuni, D., Ryten, M., Walker, R., Smith, C., Traynor, B.J., et al. (2013). Age-associated changes in gene expression in human brain and isolated neurons. Neurobiol. Aging *34*, 1199–1209.

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

Lastun, V.L., Grieve, A.G., and Freeman, M. (2016). Substrates and physiological functions of secretase rhomboid proteases. Semin. Cell Dev. Biol. *60*, 10–18.

Lemberg, M.K., and Adrain, C. (2016). Inactive rhomboid proteins: New mechanisms with implications in health and disease. Semin. Cell Dev. Biol. *60*, 29–37.

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

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

Lemieux, M.J., Fischer, S.J., Cherney, M.M., Bateman, K.S., and James, M.N.G. (2007). The crystal structure of the rhomboid peptidase from Haemophilus influenzae provides insight into intramembrane proteolysis. Proc. Natl. Acad. Sci. U. S. A. *104*, 750–754.

Li, L., Vorobyov, I., and Allen, T.W. (2013). The Different Interactions of Lysine and Arginine Side Chains with Lipid Membranes. J. Phys. Chem. B *117*, 11906–11920.

Li, Q., Zhang, N., Zhang, L., and Ma, H. (2015a). Differential evolution of members of the rhomboid gene family with conservative and divergent patterns. New Phytol. *206*, 368–380.

Li, X., Gu, Y., Dong, H., Wang, W., and Dong, C. (2015b). Trapped lipopolysaccharide and LptD intermediates reveal lipopolysaccharide translocation steps across the *Escherichia coli* outer membrane. Sci. Rep. *5*, 11883.

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

Liu, X.-N., Tang, Z.-H., Zhang, Y., Pan, Q.-C., Chen, X.-H., Yu, Y.-S., and Zang, G.-Q. (2013). Lentivirus-mediated silencing of rhomboid domain containing 1 suppresses tumor growth and induces apoptosis in hepatoma HepG2 cells. Asian Pac. J. Cancer Prev. APJCP *14*, 5–9.

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

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

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

Meissner, C., Lorenz, H., Weihofen, A., Selkoe, D.J., and Lemberg, M.K. (2011). The mitochondrial intramembrane protease PARL cleaves human Pink1 to regulate Pink1 trafficking. J. Neurochem. *117*, 856–867.

Mesak, L.R., Mesak, F.M., and Dahl, M.K. (2004). Expression of a novel gene, gluP, is essential for normal Bacillus subtilis cell division and contributes to glucose export. BMC Microbiol. *4*, 13.

Miao, F., Zhang, M., Zhao, Y., Li, X., Yao, R., Wu, F., Huang, R., Li, K., Miao, S., Ma, C., et al. (2017). RHBDD1 upregulates EGFR via the AP-1 pathway in colorectal cancer. Oncotarget *8*, 25251–25260.

Miroux, B., and Walker, J.E. (1996). Over-production of Proteins inEscherichia coli: Mutant Hosts that Allow Synthesis of some Membrane Proteins and Globular Proteins at High Levels. J. Mol. Biol. *260*, 289–298.

Motulsky, H.J., and Brown, R.E. (2006). Detecting outliers when fitting data with nonlinear regression - a new method based on robust nonlinear regression and the false discovery rate. BMC Bioinformatics *7*, 123.

Nikaido, H. (2003). Molecular Basis of Bacterial Outer Membrane Permeability Revisited. Microbiol. Mol. Biol. Rev. *67*, 593–656.

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

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

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

Parente, J., Casabuono, A., Ferrari, M.C., Paggi, R.A., De Castro, R.E., Couto, A.S., and Giménez, M.I. (2014). A Rhomboid Protease Gene Deletion Affects a Novel Oligosaccharide N-Linked to the S-layer Glycoprotein of Haloferax volcanii. J. Biol. Chem. *289*, 11304–11317.

Parussini, F., Tang, Q., Moin, S.M., Mital, J., Urban, S., and Ward, G.E. (2012). Intramembrane proteolysis of Toxoplasma apical membrane antigen 1 facilitates host-cell invasion but is dispensable for replication. Proc. Natl. Acad. Sci. U. S. A. *109*, 7463–7468.

Paschkowsky, S., Hamze, M., Oestereich, F., and Munter, L.M. (2016). Alternative Processing of the Amyloid Precursor Protein Family by Rhomboid Protease RHBDL4. J. Biol. Chem. *291*, 21903–21912.

Pierrat, O.A., Strisovsky, K., Christova, Y., Large, J., Ansell, K., Bouloc, N., Smiljanic, E., and Freeman, M. (2011). Monocyclic β-Lactams Are Selective, Mechanism-Based Inhibitors of Rhomboid Intramembrane Proteases. ACS Chem. Biol. *6*, 325–335.

Rastew, E., Morf, L., and Singh, U. (2015). Entamoeba histolytica rhomboid protease 1 has a role in migration and motility as validated by two independent genetic approaches. Exp. Parasitol. *154*, 33–42.

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

Řezáč, J. (2016). Cuby: An integrative framework for computational chemistry. J. Comput. Chem. *37*, 1230–1237.

Řezáč, J., and Hobza, P. (2012). Advanced Corrections of Hydrogen Bonding and Dispersion for Semiempirical Quantum Mechanical Methods. J. Chem. Theory Comput. *8*, 141–151.

Riestra, A.M., Gandhi, S., Sweredoski, M.J., Moradian, A., Hess, S., Urban, S., and Johnson, P.J. (2015). A Trichomonas vaginalis Rhomboid Protease and Its Substrate Modulate Parasite Attachment and Cytolysis of Host Cells. PLoS Pathog. *11*.

Rugarabamu, G., Marq, J.-B., Guérin, A., Lebrun, M., and Soldati-Favre, D. (2015). Distinct contribution of Toxoplasma gondii rhomboid proteases 4 and 5 to micronemal protein protease 1 activity during invasion. Mol. Microbiol. *97*, 244–262.

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

Russell, C.W., Richards, A.C., Chang, A.S., and Mulvey, M.A. (2017). The Rhomboid Protease GlpG Promotes the Persistence of Extraintestinal Pathogenic Escherichia coli within the Gut. Infect. Immun. *85*, e00866-16.

Saita, S., Nolte, H., Fiedler, K.U., Kashkar, H., Venne, A.S., Zahedi, R.P., Krüger, M., and Langer, T. (2017). PARL mediates Smac proteolytic maturation in mitochondria to promote apoptosis. Nat. Cell Biol. *19*, 318.

Saita, S., Tatsuta, T., Lampe, P.A., König, T., Ohba, Y., and Langer, T. (2018). PARL partitions the lipid transfer protein STARD7 between the cytosol and mitochondria. EMBO J. *37*.

Salomon-Ferrer, R., Case, D.A., and Walker, R.C. (2013). An overview of the Amber biomolecular simulation package. Wiley Interdiscip. Rev. Comput. Mol. Sci. *3*, 198–210.

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

Sekine, S., Kanamaru, Y., Koike, M., Nishihara, A., Okada, M., Kinoshita, H., Kamiyama, M., Maruyama, J., Uchiyama, Y., Ishihara, N., et al. (2012). Rhomboid Protease PARL Mediates the Mitochondrial Membrane Potential Loss-induced Cleavage of PGAM5. J. Biol. Chem. *287*, 34635–34645.

Shi, G., Lee, J.R., Grimes, D.A., Racacho, L., Ye, D., Yang, H., Ross, O.A., Farrer, M., McQuibban, G.A., and Bulman, D.E. (2011). Functional alteration of PARL contributes to mitochondrial dysregulation in Parkinson's disease. Hum. Mol. Genet. *20*, 1966–1974.

Shiau, M.-Y., Lee, P.-S., Huang, Y.-J., Yang, C.-P., Hsiao, C.-W., Chang, K.-Y., Chen, H.-W., and Chang, Y.-H. (2017). Role of PARL-PINK1-Parkin pathway in adipocyte differentiation. Metabolism. *72*, 1–17.

Shilo, B.-Z. (2016). Developmental roles of Rhomboid proteases. Semin. Cell Dev. Biol. *60*, 5–9.

Shoichet, B.K. (2006). Interpreting Steep Dose-Response Curves in Early Inhibitor Discovery. J. Med. Chem. *49*, 7274–7277.

Shokhen, M., and Albeck, A. (2017). How does the exosite of rhomboid protease affect substrate processing and inhibition? Protein Sci. *26*, 2355–2366.

Sikora, A.E., Zielke, R.A., Lawrence, D.A., Andrews, P.C., and Sandkvist, M. (2011). Proteomic analysis of the Vibrio cholerae type II secretome reveals new proteins, including three related serine proteases. J. Biol. Chem. *286*, 16555–16566.

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

Spinazzi, M., and De Strooper, B. (2016). PARL: The mitochondrial rhomboid protease. Semin. Cell Dev. Biol. *60*, 19–28.

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

Stewart, J.J.P. (2004). Optimization of parameters for semiempirical methods IV: extension of MNDO, AM1, and PM3 to more main group elements. J. Mol. Model. *10*, 155–164.

Stewart, J.J.P. (2007). Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. J. Mol. Model. *13*, 1173–1213.

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

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

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

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

Studier, F.W. (2005). Protein production by auto-induction in high-density shaking cultures. Protein Expr. Purif. *41*, 207–234.

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., et al. (2017a). General and Modular Strategy for Designing Potent, Selective, and Pharmacologically Compliant Inhibitors of Rhomboid Proteases. Cell Chem. Biol. *24*, 1523.

Tichá, A., Stanchev, S., Škerle, J., Began, J., Ingr, M., Švehlová, K., Polovinkin, L., Růžička, M., Bednárová, L., Hadravová, R., et al. (2017b). Sensitive Versatile Fluorogenic Transmembrane Peptide Substrates for Rhomboid Intramembrane Proteases. J. Biol. Chem. *292*, 2703–2713.

Tichá, A., Collis, B., and Strisovsky, K. (2018). The Rhomboid Superfamily: Structural Mechanisms and Chemical Biology Opportunities. Trends Biochem. Sci. *43*, 726–739.

Trempe, J.-F., and Fon, E.A. (2013). Structure and Function of Parkin, PINK1, and DJ-1, the Three Musketeers of Neuroprotection. Front. Neurol. *4*, 38.

Tripathi, L.P., and Sowdhamini, R. (2006). Cross genome comparisons of serine proteases in Arabidopsis and rice. BMC Genomics *7*, 200.

Tsruya, R., Wojtalla, A., Carmon, S., Yogev, S., Reich, A., Bibi, E., Merdes, G., Schejter, E., and Shilo, B.-Z. (2007). Rhomboid cleaves Star to regulate the levels of secreted Spitz. EMBO J. *26*, 1211–1220.

Urban, S. (2009). Making the cut: central roles of intramembrane proteolysis in pathogenic microorganisms. Nat. Rev. Microbiol. *7*, 411–423.

Urban, S., Lee, J.R., and Freeman, M. (2001). Drosophila Rhomboid-1 Defines a Family of Putative Intramembrane Serine Proteases. Cell *107*, 173–182.

Uritsky, N., Shokhen, M., and Albeck, A. (2016). Stepwise Versus Concerted Mechanisms in General-Base Catalysis by Serine Proteases. Angew. Chem. Int. Ed. *55*, 1680–1684.

Vaknin, Y., Hillmann, F., Iannitti, R., Ben Baruch, N., Sandovsky-Losica, H., Shadkchan, Y., Romani, L., Brakhage, A., Kniemeyer, O., and Osherov, N. (2016). Identification and Characterization of a Novel Aspergillus fumigatus Rhomboid Family Putative Protease, RbdA, Involved in Hypoxia Sensing and Virulence. Infect. Immun. *84*, 1866–1878.

Vara-Perez, M., Felipe-Abrio, B., and Agostinis, P. (2019). Mitophagy in Cancer: A Tale of Adaptation. Cells *8*, 493.

Vinothkumar, K.R. (2011). Structure of Rhomboid Protease in a Lipid Environment. J. Mol. Biol. *407*, 232–247.

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

Vinothkumar, K.R., Pierrat, O.A., Large, J.M., and Freeman, M. (2013). Structure of rhomboid protease in complex with β-lactam inhibitors defines the S2' cavity. Struct. Lond. Engl. 1993 *21*, 1051–1058.

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

Wan, C., Fu, J., Wang, Y., Miao, S., Song, W., and Wang, L. (2012). Exosome-Related Multi-Pass Transmembrane Protein TSAP6 Is a Target of Rhomboid Protease RHBDD1-Induced Proteolysis. PLoS ONE *7*.

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

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

Wang, Y., Guan, X., Fok, K.L., Li, S., Zhang, X., Miao, S., Zong, S., Koide, S.S., Chan, H.C., and Wang, L. (2008). A novel member of the Rhomboid family, RHBDD1, regulates BIK-mediated apoptosis. Cell. Mol. Life Sci. CMLS *65*, 3822–3829.

Wei, X., Lv, T., Chen, D., and Guan, J. (2014). Lentiviral vector mediated delivery of RHBDD1 shRNA down regulated the proliferation of human glioblastoma cells. Technol. Cancer Res. Treat. *13*, 87–93.

Wolf, E.V., Zeißler, A., Vosyka, O., Zeiler, E., Sieber, S., and Verhelst, S.H.L. (2013). A New Class of Rhomboid Protease Inhibitors Discovered by Activity-Based Fluorescence Polarization. PLoS ONE *8*.

Wunderle, L., Knopf, J.D., Kühnle, N., Morlé, A., Hehn, B., Adrain, C., Strisovsky, K., Freeman, M., and Lemberg, M.K. (2016). Rhomboid intramembrane protease RHBDL4 triggers ER-export and non-canonical secretion of membrane-anchored TGFα. Sci. Rep. *6*, 27342.

Xue, Y., and Ha, Y. (2012). Catalytic Mechanism of Rhomboid Protease GlpG Probed by 3,4-Dichloroisocoumarin and Diisopropyl Fluorophosphonate. J. Biol. Chem. *287*, 3099–3107.

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

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

Yan, C., Gong, L., Chen, L., Xu, M., Abou-Hamdan, H., Tang, M., Désaubry, L., and Song, Z. (2019). PHB2 (prohibitin 2) promotes PINK1-PRKN/Parkin-dependent mitophagy by the PARL-PGAM5-PINK1 axis. Autophagy 1–16.

Zhang, M., Miao, F., Huang, R., Liu, W., Zhao, Y., Jiao, T., Lu, Y., Wu, F., Wang, X., Wang, H., et al. (2018a). RHBDD1 promotes colorectal cancer metastasis through the Wnt signaling pathway and its downstream target ZEB1 (vol 37, 22, 2018). J. Exp. Clin. Cancer Res. *37*, 60.

Zhang, X., Zhao, Y., Wang, C., Ju, H., Liu, W., Zhang, X., Miao, S., Wang, L., Sun, Q., and Song, W. (2018b). Rhomboid domain-containing protein 1 promotes breast cancer progression by regulating the p-Akt and CDK2 levels. Cell Commun. Signal. *16*, 65.

Zhou, Y., Moin, S.M., Urban, S., and Zhang, Y. (2012). An internal water-retention site in the rhomboid intramembrane protease GlpG ensures catalytic efficiency. Struct. Lond. Engl. 1993 *20*, 1255–1263.

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

Documents

World Malaria Report 2018. Geneva: World Health Organization; 2018. Licence: CC BY-NC-SA 3.0 IGO.