STRUCTURAL AND BIOCHEMICAL CHARACTERIZATION OF RHOMBOID INTRAMEMBRANE PROTEASES

By

Jose L Chavez

A dissertation submitted to the Graduate Faculty in Biochemistry in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2011

This manuscript has been read and accepted for the

Graduate Faculty in Biochemistry in satisfaction of the

dissertation requirements for the degree of Doctor in Philosophy.

Dr. Iban Ubarretxena

Date

Chair of Examining Committee

Dr. Edward Kennelly

Date

Executive Officer

Dr. Edward Kennelly

Dr. Lesly Davenport

Dr. K.V. Lakshmi

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

STRUCTURAL AND BIOCHEMICAL CHARACTERIZATION OF RHOMBOID INTRAMEMBRANE PROTEASES

By

Jose L Chavez

Adviser: Dr Iban Ubarretxena

Intramembrane proteases are involved in multiple biological processes including cell growth and development, and apoptosis. There is no conservancy between hydrosoluble and membrane proteases. However, the catalytic residues and surrounding amino acids are absolutely conserved, suggesting that they both protein families share catalytic mechanisms but with two remarkable differences. (1) The ability of intramembrane proteases to cleave their substrates in the hydrophobic interior of the lipid bilayer, and (2) to do so in regions where the substrate displays a-helical conformation. D. melanogaster rhomboid-1 cleaves within the transmembrane domain region of epidermic growth factor receptor (EGFR) ligands Gurken, Keren and Spitz, resulting in their extracellular export. We designed substrate chimeras in which the transmembrane and cytoplasmic regions of Gurken, Keren and Spitz were preserved, while their EGFR ligand ectodomain was replaced by maltose binding protein. In vitro activity assays in detergent using purified components showed that rhomboid-1, H. sapiens RHBDL2, P. aeruginosa PA3086 and E. coli GlpG display comparable activity against these substrate chimeras. Mass spectrometry analysis of the N-terminal reaction product identified a single cleavage site after Ala138 for the Spitz chimeras, after Ala122 for the Keren chimeras, and after Ala245 for the Gurken chimeras that was identical for all rhomboids tested, suggesting a conservation of proteolytic profiles among prokaryotic and eukaryotic rhomboids. The identified cleavage site was located towards

the N-terminal end of the transmembrane domain of each substrate. Positions that were sensitive to alanine scanning were further studied by introducing additional mutations to show that aside of ala in position P1, amino acids with low-helical propensities are necessary in the positions P2 and P1'. Finally, a bulky hydrophobic residue with a high helical propensity is important in P2' position to control the location of cleavage. We also carried out structural work and solved the N-terminal domain of Rhomboid and showed it displays high-affinity for membranes. Our work is put in a more general context by comparison with other intramembrane proteases and future work to unravel the mechanism of substrate binding and unwinding is also discussed.

Acknowledgments

Monica, Valeria, Luis y Pablo.

Table of Contents

Table of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 5

Chapter 1. Intramembrane Proteolysis.

Regulated Intramembrane Proteolysis

Regulated intramembrane proteolysis (RIP) is a prevalent molecular mechanism to activate pools of signaling proteins precursors, which reside in the biological membranes of eukaryotic and prokaryotic cells (1-4). In eukaryotes, RIP is involved in genetic regulation (5), development (6) and apoptosis (7). In prokaryotes, RIP is believed to regulate major biological tasks such as quorum sensing (8, 9) and sporulation (10, 11). As part of their activation, these signaling proteins precursors are specifically proteolyzed within their transmembrane region (3, 12), resulting in the release of cytoplasmic domains (**Fig. 1.1a**) and/or extracellular/luminal (**Fig. 1.1b**). These domains can then diffuse to the nucleus or be secreted outside the cell where they can elicit a biological response. A variety of signaling proteins in eukaryotes are activated in this fashion, including transcriptional activators in the Notch (13) and ErbB-4 cascades (14); cellular growth factors (15, 16); and cholesterol regulatory proteins (17).

The intramembrane proteolysis of the above mentioned precursors is catalyzed by integral membrane proteins termed Intramembrane Cleaving Proteases (i-CLiPs) (18). I-CLiPs have in common the ability to hydrolyze peptide bonds within the transmembrane domain of the substrates, where accessibility to bulk water is restricted due to the non-polar environment of the cell membrane interior. I-CLiPs can be divided into four different groups: S2P metalloproteases, presenilin and signal peptide peptidase (SPP), di-aspartyl proteases, and rhomboid serine proteases (19).

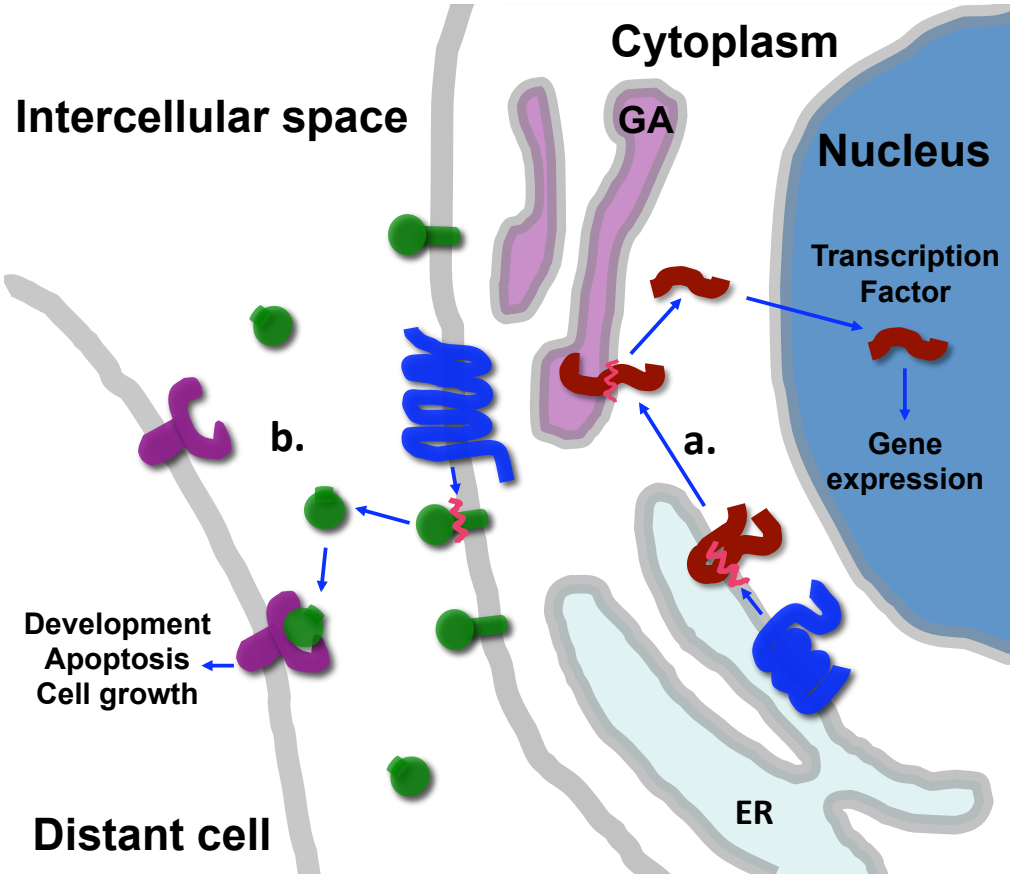

Figure 1.1 Regulated Intramembrane proteolysis. This signal transduction mechanism can occur in two different ways: **a)** the cytoplasmic domain of the substrate is delivered into the cytoplasm and then it acts inside the nucleus as transcriptional regulators of proteins (e.g. HMGCoA reductase). **b)** the extracellular domain of the substrate is secreted outside the cell where it regulates the metabolism and development of distant cells (rhomboid).

An important difference between soluble proteases and intramembrane proteases is not only that in the case of the latter the proteolysis occurs in an environment were water is a limited factor but also that the former require the scissile peptide bond to be located in a region without secondary structure (e.g., thrombin **Fig. 1.2a**), while their intramembrane counterparts are able to cleave a peptide bonds in a region with alpha helical structure (e.g., rhomboid (**Fig. 1.2b**). In this thesis we will focus on the biochemical characterization of rhomboids with emphasis on presenting our efforts to study the substrate specificity and peptide bond selectivity of these intramembrane proteases. In this chapter we will introduce the reader to proteases, and in particular to intramembrane proteases, as well as to membrane proteins. We will also define the specific aims of this thesis.

Cell Membranes

RIP is a ubiquitous signal transduction mechanism, which occurs in the vast majority of cell membranes. Cell membranes define the external boundaries and intracellular organelles of cells. A vast variety of transport processes, signal transduction events, and enzymatic activities occur at these membranes (20) (**Fig. 1.1**). Biological membranes are composed mainly of lipids and proteins. There are three major types of membrane lipids: phospholipids, sterols and glycolipids. Phospholipids, the dominant class of membrane lipids, are constructed from a hydrophobic moiety consisting of two hydrocarbon tails (usually fatty acids), and a hydrophilic moiety consisting of a scaffold to which fatty acids are covalently linked (e.g. glycerol) and a phosphate based polar head group. Because of their cylindrical and amphiphatic nature, the energetically most-favored structure for most phospholipids is to form a continuous bilayer in aqueous environments (21). These bilayers are composed of 2 lipid sheets with the hydrophobic tails of each individual sheet interacting with one another forming a

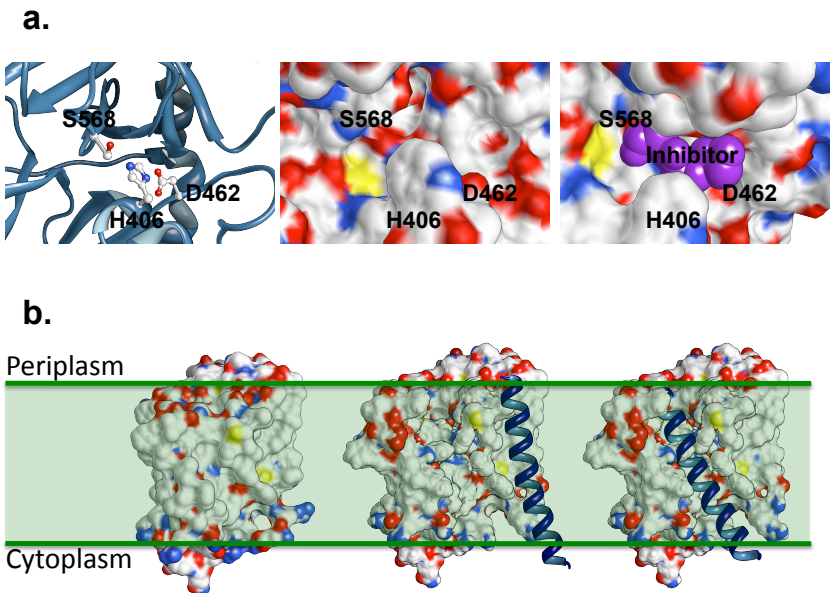

Figure 1.2. **Differences between hydrosoluble and membrane proteases.** The cleavage site of the substrates. **a)** Enzyme substrate complex between thrombin (PDB 1ABJ) and the inhibitor D-Phe-Pro-Arg methylene. **b)** Formation of the complex between E coli GlpG rhomboid and the transmembrane domain of Spitz. Molecular graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR001081).

hydrophobic interior ~30 Å thick (22). This leaves the hydrophilic head group layers, each 10-15 Å thick, to interact with the aqueous media on each side of the bilayer (23).

In membranes, a large fraction of membrane proteins are found in fluid disordered regions. In other regions, phospholipids associate with membrane proteins forming submicroscopic assemblies called lipid rafts (24). Lipids rafts constitute specialized membrane microdomains that engage in signal transduction, membrane trafficking, cytoskeletal organization and motility, polarization and pathogen entry (25, 26). The tightly packed lipids rafts are considered highly dynamic assemblies that float freely in the plane of a surrounding fluid membrane (25).

The fluidity of the membrane depends on the degree of unsaturation (number of double carbon-carbon bonds) and the length of the hydrocarbon tail. The latter defines a fundamental property of synthetic bilayers, made from a single type of phospholipid, i. e. the phase transition temperature (T_c). Close to the T_c the hydrocarbon chains are roughly parallel and have low mobility and close proximity (crystalline state). Above the T_c, a change occurs from the rigid crystalline state to a disordered state (liquid state) characterized by an increased mobility of the hydrocarbon chains, which in turn thins the bilayer (27).

Cholesterol, an essential component for the assembly of lipid rafts, also moderates the fluidity of the membrane. Cholesterol inserts into the bilayer with its long axis perpendicular to the plane of the membrane forming specific complexes with certain classes of phospholipids (28). When present in large amounts in eukaryotic plasma membranes, cholesterol tends to make lipid bilayer less fluid decreasing the permeability of the bilayer to small water-soluble molecules (29). Cholesterol also

inhibits phase transitions by preventing hydrocarbon chains from coming together and crystallizing (30).

The plasma membranes of many mammalian cells are composed predominantly of four types of phospholipids: phosphatidylcholine, phosphatidylethanolamine, phosphatidylserine and sphingomyelin, each type carrying head groups that differ in size, shape, and charge (31). The proportion of each phospholipid varies not only according to the type of membrane (i.e. species, cell type, organelle) but also between the two monolayers of the lipid bilayer. Glycolipids contribute to this asymmetry as they are found exclusively in the noncytosolic monolayer of the lipid bilayer. The asymmetrical distribution observed in the lipid bilayer is functionally important. The hydrophilic head groups of glycolipids are sugar groups that act as specific sites for recognition by carbohydrate-binding proteins. Similarly, binding and functionality of some cytosolic enzymes and peripheral membrane proteins depends on the presence of specific phospholipid head groups.

Membrane Proteins

Membrane proteins are structurally tethered to the biological membranes and are isolated from membrane preparations rendered during cell fractionation of both: prokaryotic and eukaryotic cells. There are three membrane protein classifications: transmembrane, monotopic/peripheral and membrane active peptides (32-34). Transmembrane proteins (also known as integral membrane proteins) have part of their molecule entirely extended through the biological membrane. These transmembrane regions can be composed of alpha helical bundles or of beta-strands arranged into a beta-barrel (**Fig. 1.3a-c**).

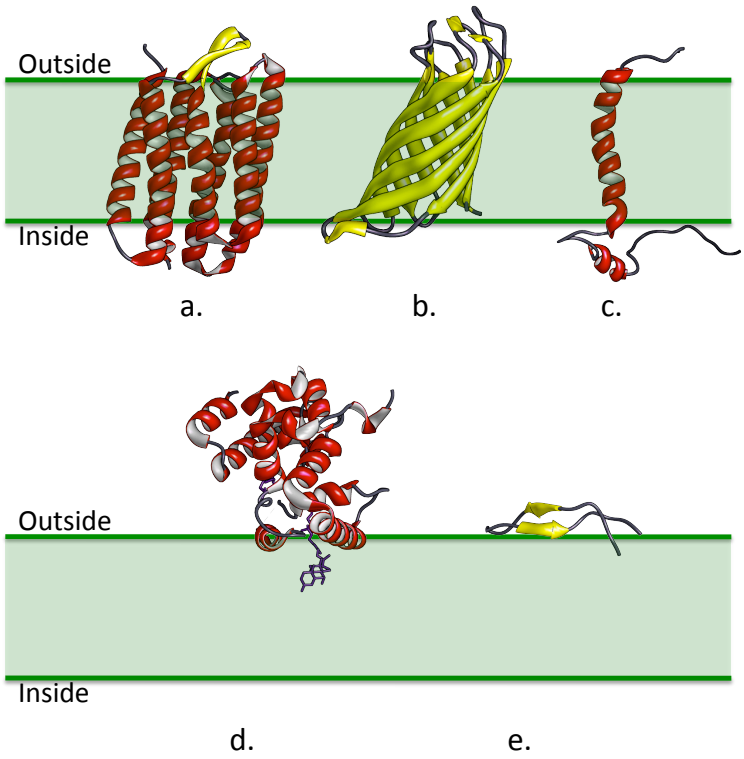

Figure 1.3. Membrane protein classification. a) Alpha helical polytopic: Bacteriorhodopsin (3NS0), **b)** Beta barrel transmembrane: Outer membrane protein A (1QJP), **c)** Alpha helical bitopic: Human CD4 (2KLU), **d)** monotopic Penicillin-binding protein 1A (2oqo), **e)** peripheral: Lactoferricin B (pdb: 1lfc). Molecular graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR001081).

Well-studied examples of transmembrane proteins are bacteriorhodopsin (35), Photosystem I (36), and aquaporin-1 (37). The rhomboid intramembrane proteases, which are the subject of this thesis, are bone-fide alpha-helical integral membrane proteins. The second type of membrane proteins are the monotopic/peripheral membrane proteins that are associated with biological membranes by non-polar interactions with the membrane, with integral membrane proteins, or membrane lipids (**Fig. 1.3d**). Penicillin-binding protein 1A (38), Sphingomyelinase C (39) and Alpha-glycerophosphate oxidase (40) are examples of this type of protein. The last type of membrane proteins is small peptides (less than 40 amino acid residues), which often function as pores or channels (**Fig. 1.3e**). Lactoferrin (41) , Hepcidin-20 (42) and Glucagon (43) belong to this third type of membrane proteins.

Proteases – Biological and medical relevance

According to the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) peptidases are hydrolases that break peptide bonds in proteins and peptides. Proteases are classified mainly according to two criteria: a) the location of the scissile peptide bond in the substrate, and b) the amino acid residue involved in the catalysis of the peptide bond. Exopeptidases cleave terminal peptide bonds located in the N- or C-terminal end of protein or peptides (aminopeptidases and carboxypeptidases, respectively). Endopeptidases hydrolyze peptide bonds other than the carboxy or amino termini of protein or peptides. In terms of the amino acid residue directly involved in catalysis proteases can be classified as serine, threonine, cysteine, aspartic acid, or a metalloproteases.

Table 1.1. Percentage of peptidases in the genome of evolutionarily distant organisms. Peptidases represent around the 5% of the genomes of organisms from humans to bacteria.

Species	Peptidases/ genome (1)	Genes/ genome(1)	%(1)
H. sapiens	1,116	24,194	4.61
M. musculus	932	28,068	3.32
D.melanogaster	772	13,833	5.58
C. elegans	502	20,516	2.45
H. salinarun	67	2749	2.44
E. coli K-12	172	4,243	4.05

(1) Rawlings ND, Morton FR, Kok CY, Kong J, & Barret AJ (2008) Nucleic Acids Res 36, D320-325.

Pepsin, a digestive protease, which was the first ever reported (in 1836 by T. Schwann) animal enzyme, is the founding member of this vast protein family (44). By 1929 Pepsin was among the first proteins to be purified and characterized (45). A few years later, its zymogen pepsinogen, was also characterized and its importance in the regulation of pepsin activity established (46). Since the pioneering work on Pepsin our knowledge of protease function, structure and mechanism has grown tremendously. Evidence of the biological relevance of peptidases is reflected by the number of peptidases genes present in the genomes of different and evolutionarily distant organisms. Peptidases represent, on average, 5% of the total number of protein-encoding genes (47) in *H. sapiens*, *M. musculus*, *D. melanogaster*, *C. elegans*, *H salinarum*, and *E coli* (**Table 1.1**). A similar prevalence is also observed for other important protein families, such as kinases (48) and membrane transporters (49). Peptidases and proteases are not restricted to non-specific functions, such as protein digestion, but also play critical roles in cellular process that require specificity, including apoptosis (50, 51), cell cycle (52), cholesterol metabolism (53), fate and compartmentalization of proteins (54). Proteases have long been useful molecular tools in research and industry. In industry, proteases represent the 40% of the total industrial enzyme market (55, 56) and are commonly used in the detergent industry (57), in the manufacturing of fur and leather (58), and also in baking and cheese elaboration (59). In research, proteases such as trypsin, pepsin, chymotrypsin, are used in protocols for the identification of proteins by MALDI-TOF (60, 61). In addition, thrombin, TEV, enterokinase, are often used to remove tags from over expressed fusion proteins (62). Finally, proteases have long been important medical targets for the pharmaceutical industry.

By the year 2008 the pharmaceutical industry had identified over 14000 proteins targets of medical relevance (63). Among these there were 91 proteases involved in hereditary diseases (64), which represents approximately 10% of the total proteases present in human genome. These hereditary disorders provoked by proteases included Parkinson's disease (DJ-1, UCHL1, HTRA2), renal tubal dysgenesis (ACE), autoimmune lymphoproliferative syndrome (caspase-10), Alzheimer's disease (PSEN-1, PSEN-2), and Gile de Tourette syndrome (IMMP2L)(13, 65-70)

The i-CLiPs are a subfamily of proteases, which share at least five common features: (1) from a functional perspective they are involved in the regulation of cellular physiology; (2) each protease is a polytopic integral MP with catalytic residues located inside the lipid bilayer; (3) they hydrolyze substrates in their transmembrane domains (TMDs); (4) they have short sequence motifs surrounding their catalytic residues that are archetypical of water-soluble proteases of the same mechanistic class; and (5) they occur in large protein families. In the following paragraphs we will introduce the four i-CLiP classes that have been identified to date.

Intramembrane proteases – classes.

GxGD-type aspartyl proteases.

Presenilin and SPPs are involved in signaling events, as well as in the degradation of membrane-bound proteins with no apparent biological functions (71, 72). They span the membrane nine times (**Fig. 1.4a**) and their catalytic motifs YD and GLGD (**Fig. 1.4b**) are located in the middle of adjacent TMD regions (73, 74). A third motif, common to both presenilin and SPP, is a C-terminal PAL sequence (**Fig. 1.4b**) that might contribute to active site conformation (75-77). These di-aspartyl intramembrane proteases also share equivalent pharmacology against transition state analog inhibitors,

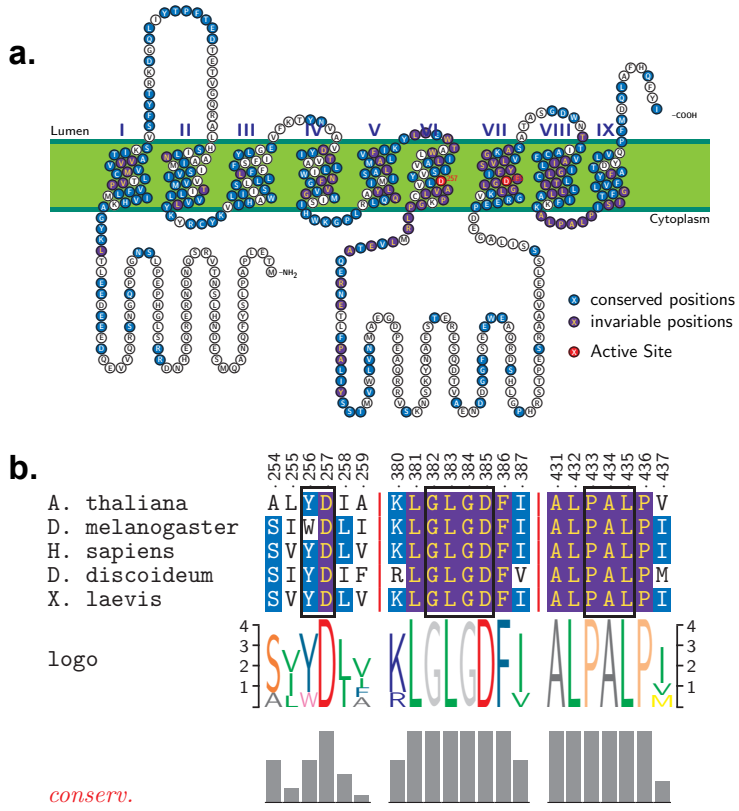

Figure 1.4. Bidimensional topology and conservation of Presenilins. a). Human Presenilin shows nine transmembrane domains and the aspartic residues involved in the catalysis are in consecutive TMD (VI and VII). **b)** Conservation of presenilin from different species (*A. Thaliana*, *D. melanogaster*, H. sapiens, *D. discoideum* and *X. laevis)*. Bidimensional topology was produced using TEXtopo (Eric Beitz (2000), TEXtopo: shaded membrane protein topology plots in LATEX2ε. Bioinformatics 16:1050–1051). Alignments were produced using TEXshade (Eric Beitz (2000), TEXshade: shading and labeling multiple sequence alignments using LATEX2ε. Bioinformatics: 16, 135–139)

non-transition state inhibitors and modulator compounds (78). Aside from these conserved sequence characteristics there are important structural and functional aspects that differentiate these two types of proteases. Whereas presenilin is the catalytic component of a multiprotein complex, SPP acts alone. Presenilin and SPP also display inverted overall topologies (79) that may be significant for substrate specificity of SPP preferentially for type II transmembrane proteins and of presenilin towards type I transmembrane proteins (80, 81). The abolishment of catalytic activity due to mutations of the conserved aspartic acid residues D219 and D265 (within the YD and GLGD motifs respectively) reinforced the identification of SPP as an aspartic protease (82, 83). In addition as described for presenilin, an initial requirement for SPP intramembrane cleavage to occur is the release of part of the extracellular domain from the substrate. Analogous to α- and β-secretase, a signal peptidase liberates the signal peptide from the protein before SPP can cleave within the transmembrane region (83, 84). Moreover, an important requirement for intramembrane cleavage by SPP seems to be the presence of helix-breaking resides to provide flexibility to the transmembrane region and expose the scissile peptide bond (85).

S-2-P proteases.

The role of S2P was first described in the regulation of cholesterol homeostasis as a response to sterol deprivation in animal cells (17, 53). In this scenario, proteolysis of the sterol regulatory element binding protein (SREBP) releases an N-terminal segment, which then goes into the nucleus to regulate the transcription of the genes of HMGCoA reductase and the low-density lipoprotein receptor (53). SREBP is initially cleaved by site-1 protease within a short hydrophilic loop that links the two transmembrane segments.

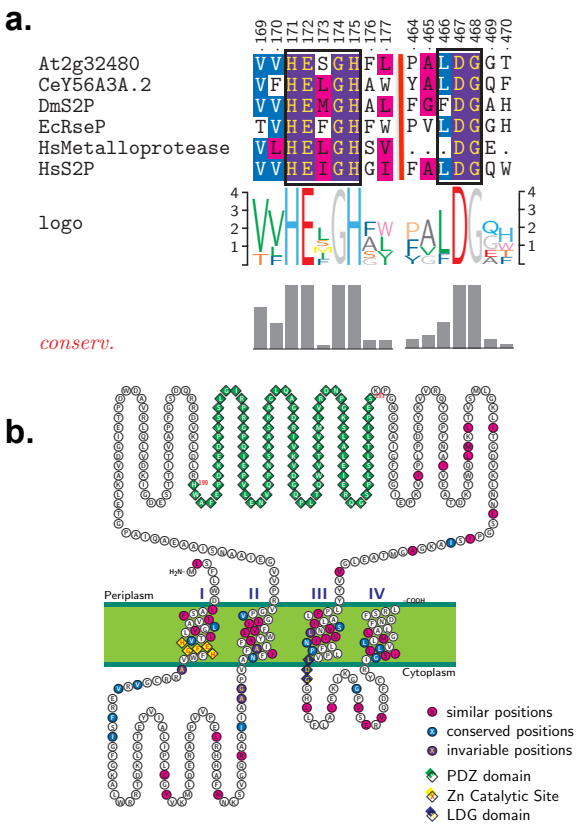

Figure 1.5. Bidimensional topology and conservation of S-2-P metalloproteases. a) Conservation of S-2-P proteases from different species (*A. Thaliana*, *C. elegans*, *D. melanogaster*, *E. coli, H. sapiens*. **b)** Bidimensional topology of RseP, an *E. coli* ortholog of S2P protease. Bidimensional topology were produced using TEXtopo (Eric Beitz (2000), TEXtopo: shaded membrane protein topology plots in LATEX2ε. Bioinformatics 16:1050–1051). Alignments were produced using TEXshade (Eric Beitz (2000), TEXshade: shading and labeling multiple sequence alignments using LATEX2ε. Bioinformatics: 16, 135–139).

S2P proteases show a high conservancy of the motif HExGH, which is the binding site for zinc and the sequence LDG that is the catalytic motif of S2P **(Fig. 1.5a**). These two motifs are located in the cytoplasmic face of the structure of RseP (10), an *E. coli* S2P homolog (**Fig. 1.5b**). Biochemical studies have identified sequence elements on both sides of the membrane required for efficient proteolysis: an arginine residue in the short luminal loop and a cytosolic motif DRSR (86). Substitution of the arginine residue by alanine impeded proteolysis, whereas a conservative substitution by lysine reduced the activity of S2P (86). Later studies revealed that initial sterol-induced cleavage within the luminal loop by site1 protease (S1P), an intramembrane serine protease, occurs in proximity to the conserved arginine residue (53, 87). Following S1P proteolysis, S2P cleaves SREBP substrate intramembranously between a leucine and cysteine (88), located three residues from the C-terminal side of the critical DRSR motif. However, neither leucine nor cysteine was found to be necessary for cleavage, nor were many of the conserved residues within the transmembrane domain, as shown by singly replacing each TM residue with alanine (88).

Rhomboid proteases

Rhomboids were first discovered in *Drosophila melanogaster*. Rhomboid-1 (Rho-1) is expressed in the early stages of development (even before the fecundation) and constitutes an essential factor in the development of the fly (89-91). Impaired function of Rho-1 leads to development abnormalities in eyes and wings (89, 92-94). In *D. melanogaster* Rho-1 has three different substrates, all of them single pass membrane proteins, Gurken, Keren and Spitz. These substrates share a common architecture. They are all bitopic membrane proteins. The N-terminal extracellular domain (27, 13 and 15 kDa for gurken, keren and spitz respectively) is an Epidermal Growth Factor Receptor ligand (EGFR-ligand) and the smaller cytoplasmic C-terminal domain (3.1, 8.3

and 8.4 kDa for gurken, keren and spitz respectively) has no known physiological role (95). Rhomboid-1 cleaves within the TMD region of these substrates to release the EGFR domain into the Golgi Lumen, from where it enters the exocytic pathway to end in the cell's exterior. Once outside the cell the EGFR-ligand can activate EGFRs in distant cells.

The co-localization of both substrate and enzyme in the membrane of the Golgi apparatus (GA) (12, 95, 96) is the only requirement for intramembrane proteolysis. Thus the regulation of Rho-1 activity depends on the differential compartmentalization of rhomboid and substrates. Whereas Rho-1 resides in the membrane of the GA, all the substrates (Gurken, Keren and Spitz) are located in the membrane of the Endoplasmic Reticulum (ER) (97). The membrane protein Star carries out the transportation of substrates from ER membrane to GA membrane. The star-mediated transport of the EGFR-ligand precursors can be modulated by the activity of two other homologs of Rho-1: Rho-2 and Rho-3. These two homologs are present in ER membrane where they cleave Star and Spitz to regulate the amount of Spitz transported to the GA membrane (95, 98).

Rhomboid structure

Although the overall sequence identity among rhomboids is very low (**Fig. 1.6a**), sequence comparison between rhomboids of prokaryotic and eukaryotic origin shows conservation of key catalytic amino acid residues (15) (**Fig. 1.6b**). In addition, all rhomboids share a similar transmembrane topology. Rho-1 was early on established to be a serine protease based on the identification by mutagenesis studies of the catalytic residues serine-217, asparagine-169 and histidine-281 (15, 96, 99, 100). Additionally,

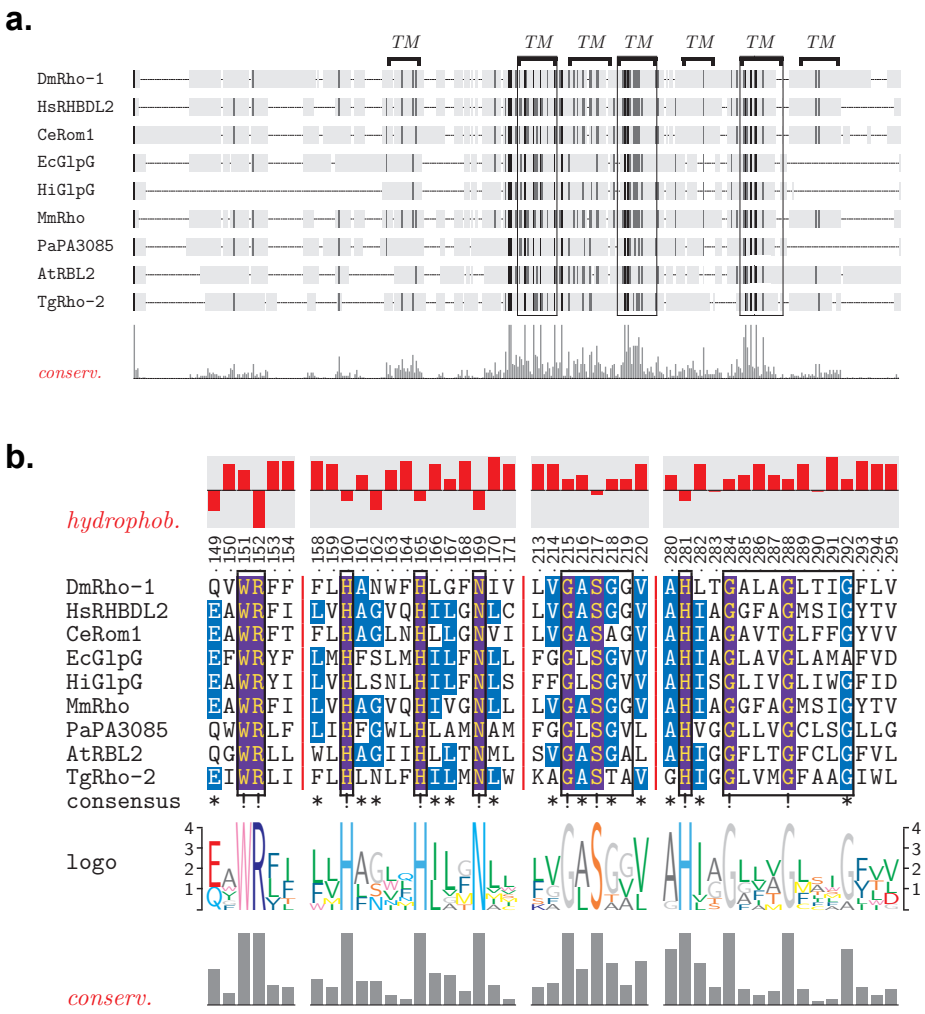

Fig. 1.6. Conservancy of rhomboid proteins. a) Finger print sequence of rhomboid orthologs. Rhomboid proteins does not have a high conservancy in their sequence but they show a structural conservancy and also a high conservancy for the catalytic residues in a their sequence. **b)** Conservancy of the motif $W_{151}R_{152}$ located in the Loop-1, catalytic residues: N_{169}, S_{217}, and H_{281}. Also the motif GxSxG typical of the serine proteases and the residues H160 and h162 involved in the formation of the oxyanion. The alignments were done using TEXShade (Eric Beitz (2000), TEXshade: shading and labeling multiple sequence alignments using LATEX2ε. Bioinformatics: 16, 135–139).

the Rho-1 also displays the catalytic motif GxSxG, which is highly conserved in serine proteases ($G_{215}A_{216}S_{217}G_{218}G_{219}$) (101).

The first rhomboid structure to be solved was that of GlpG (102), an *E coli* rhomboid ortholog (**Fig. 1.7a-b**). The structure of rhomboid hiGlpG (100) (from *Haemophilus influenzae*) was elucidated a few months later. The structure of GlpG shows a serine-201 and a histidine-254 that are part of the catalytic site of GlpG. Asparagine-154, the proposed third component of the catalytic triad and counterpart of an aspartic acid residue present in archetypical serine proteases, is not located within H-bond distance to the histidine-254. Thus eliminating the possibility of a bona-fide serine catalytic triad in this enzyme (102, 103). The idea of a catalytic dyad in rhomboids is supported by the structure of hiGlpG, which lacks an asparagine, or another residue that can stabilize the imidazol ring of histidine, in the active site (100).

However, an N154A GlpG mutant is not able to restore the wild type phenotype in AarA dependent *Providencia stuartii* cells (103) suggesting that this residue is important for rhomboid activity. Furthermore, the alanine substitution of the residue asparagine-169 inactivates Rho-1 in in vivo experiments (15). All this evidence, together with a possible re-positioning of a water molecule in a open-conformation of GlpG after L1 loop removal, indicates a possible role of the residue Asn154 in the conformation of the oxyanion hole (104).

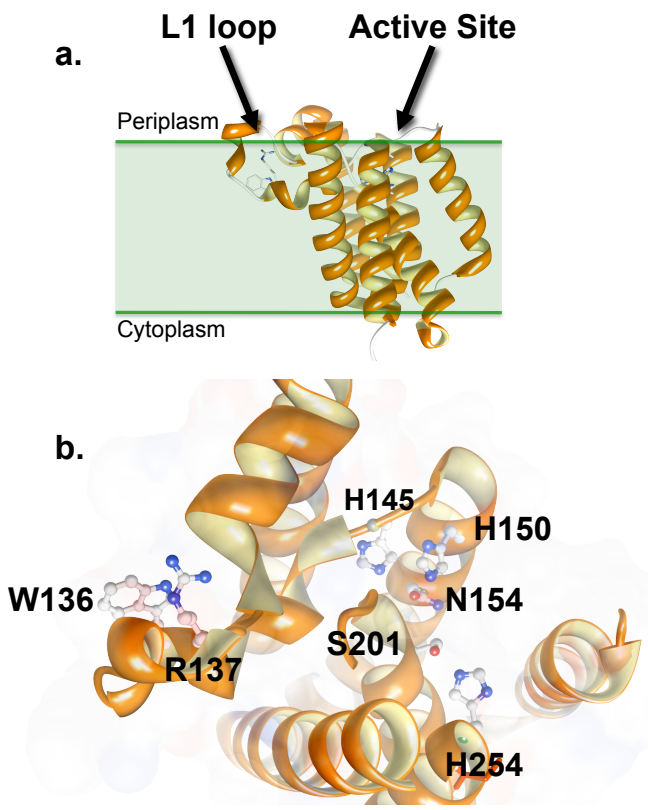

Fig. 1.7. Structure of GlpG. a) *E. coli* rhomboid GlpG has six transmembrane domain and a periplasmic Loop-1. GlpG molecule has a V-cup shaped form opened to the periplasmic face. **b)** The catalytic residues are opened to the periplasmic face of the molecule. Molecular graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR001081).

Biological roles of rhomboids

There is ample evidence that rhomboid homologs from different organisms are involved in important biological processes such as *quorum* sensing (AaaR from *Providencia stuartii*), and apoptosis (RHBDL-2 in humans) (11, 105). However aside from the *D. melanogaster* substrates Gurken, Keren and Spitz, the physiological substrates of other rhomboids are still unknown (106). A very important piece of evidence suggesting that rhomboid activity is conserved is the fact that rhomboid orthologs from different organisms can restore *in vivo* the wild type phenotype in rhomboid knock out cells of *D. melanogaster* (96). To date all studies indicate that several rhomboid orthologs are able to cleave *D. melanogaster* substrates, but it is not known how rhomboids achieve substrate specificity and peptide bond selectivity.

The cleavage site of Rho-1

The ASIASGA sequence, present in the N-terminus quarter of the TMD of Spitz, was found to be necessary for GlpG to recognize and proteolyze this substrate (107) Baker et al, used a substrate C100Spitz-Flag derived from C100Flag which is a recombinant γ-secretase substrate (108). By sequential site directed mutagenesis spitz motif ASIASGA was introduced in the N-terminal portion of the transmembrane domain. After the rhomboid digestion of this substrate the products were analyzed employing mass spectrometry (109) to conclude that DDM solubilized GlpG cleaved C100-Spitz at two adjacent sites, after Ala141 and Gly143, within the substrate motif (110). Maegawa et al. (111, 112) used a different substrate: BLA-LACYTM2. This substrate contains a single transmembrane domain, which is the second transmembrane domain of LacY (lactose permease), flanked by a N-terminal BLA domain (b-lactamase) and a C-terminal

MBP domain (maltose binding protein). Using N-terminal sequence of the MBP adduct product was shown that GlpG rhomboid cleaves the peptide bond a serine and aspartic acid in the N-terminus region of a substrate designed with the TM2 of lactose permease LacY. Strisovsky (113) determined that the cleavage site of AarA occurs in the position Ala138-Ser139 in a substrate designed with the TMD of Spitz and MBP and thioredoxin in the C-terminal and the N-terminal respectively.

Aims of the thesis

Membrane proteases differ from soluble proteases in that soluble proteases require that the cleavage site be located in a region of the substrate with no secondary structure. In the case of membrane proteases, specifically rhomboids, the cleavage site is located in a transmembrane domain. The focus of this thesis is to address the question of how do rhomboids achieve substrate and peptide bond specificity. To this end, we have first determined the cleavage site of different eukaryotic and prokaryotic rhomboids against model substrates derived from the physiological *D. melanogaster* substrates Gurken, Keren and Spitz.

Our experimental approach has been to reconstitute the activity of different rhomboid orthologs, including rhomboid-1 from *D. melanogaster*, in an in vitro assay in detergent. To avoid well-known issues with the expression of wild type Gurken, Keren and Spitz, mainly post-translational modification like glycosylation at the extracellular domain (114) we have designed model substrates in which the extracellular EGFR-ligand domain was substituted by maltose binding protein (MBP). Using these substrates we have demonstrated that all eukaryotic and prokaryotic rhomboids tested display a conserved substrate and peptide bond selectivity.

This thesis will focus on the following questions:

a. Is there a conservancy of the cleavage site between rhomboids from evolutionarily distant organisms?
b. What is the role of the residues located in the substrate motif and how do they affect the catalysis of rhomboid orthologs?
c. What is the role of the soluble domain of GlpG rhomboid in the catalysis of rhomboid?

Bibliography

1. Sik A, Passer BJ, Koonin EV, & Pellegrini L (2004) *J Biol Chem* **279,** 15323-15329.
2. Martys-Zage JL, Kim SH, Berechid B, Bingham SJ, Chu S, Sklar J, Nye J, & Sisodia SS (2000) *J Mol Neurosci* **15,** 189-204.
3. Brown MS, Ye J, Rawson RB, & Goldstein JL (2000) *Cell* **100,** 391-398.
4. Kanaoka MM, Urban S, Freeman M, & Okada K (2005) *FEBS Lett* **579,** 5723-5728.
5. Li T, Wen H, Brayton C, Das P, Smithson LA, Fauq A, Fan X, Crain BJ, Price DL, Golde TE*, et al*. (2007) *J Biol Chem* **282,** 32264-32273.
6. Kim YS, Kim SG, Park JE, Park HY, Lim MH, Chua NH, & Park CM (2006) *Plant Cell* **18,** 3132-3144.
7. Cipolat S, Rudka T, Hartmann D, Costa V, Serneels L, Craessaerts K, Metzger K, Frezza C, Annaert W, D'Adamio L*, et al*. (2006) *Cell* **126,** 163-175.
8. Gallio M, Sturgill G, Rather P, & Kylsten P (2002) *Proc Natl Acad Sci U S A* **99,** 12208-12213.
9. Urban S (2006) *Genes Dev* **20,** 3054-3068.
10. Koide K, Maegawa S, Ito K, & Akiyama Y (2007) *J Biol Chem* **282,** 4553-4560.
11. Stevenson LG, Strisovsky K, Clemmer KM, Bhatt S, Freeman M, & Rather PN (2007) *Proc Natl Acad Sci U S A* **104,** 1003-1008.
12. Urban S, Lee JR, & Freeman M (2002) *EMBO J* **21,** 4277-4286.

13. De Strooper B, Annaert W, Cupers P, Saftig P, Craessaerts K, Mumm JS, Schroeter EH, Schrijvers V, Wolfe MS, Ray WJ, *et al*. (1999) *Nature* **398,** 518-522.

14. Ni CY, Murphy MP, Golde TE, & Carpenter G (2001) *Science* **294,** 2179-2181.

15. Urban S, Lee JR, & Freeman M (2001) *Cell* **107,** 173-182.

16. Denisov IG, Hung SC, Weiss KE, McLean MA, Shiro Y, Park SY, Champion PM, & Sligar SG (2001) *J Inorg Biochem* **87,** 215-226.

17. Rawson RB, Zelenski NG, Nijhawan D, Ye J, Sakai J, Hasan MT, Chang TY, Brown MS, & Goldstein JL (1997) *Mol Cell* **1,** 47-57.

18. Bihel F, Das C, Bowman MJ, & Wolfe MS (2004) *J Med Chem* **47,** 3931-3933.

19. Wolfe MS (2009) *J Biol Chem* **284,** 13969-13973.

20. Lehninger AL, Nelson DL, & Cox MM (2005) *Lehninger principles of biochemistry* (W.H. Freeman, New York).

21. Berg JM, Tymoczko JL, & Stryer L (2002) *Biochemistry* (W.H. Freeman, New York).

22. Mitra K, Ubarretxena-Belandia I, Taguchi T, Warren G, & Engelman DM (2004) *Proc Natl Acad Sci U S A* **101,** 4083-4088.

23. Nagle JF & Tristram-Nagle S (2000) *Biochim Biophys Acta* **1469,** 159-195.

24. Simons K & Vaz WL (2004) *Annu Rev Biophys Biomol Struct* **33,** 269-295.

25. Rajendran L & Simons K (2005) *J Cell Sci* **118,** 1099-1102.

26. Pike LJ (2005) *Biochim Biophys Acta* **1746,** 260-273.

27. Carruthers A & Melchior DL (1988) *Annu Rev Physiol* **50,** 257-271.

28. Rothman JE & Engelman DM (1972) *Nat New Biol* **237,** 42-44.

29. Szabo G (1974) *Nature* **252,** 47-49.

30. Hinz HJ & Sturtevant JM (1972) *J Biol Chem* **247,** 3697-3700.

31. Dowhan W, and Bogdanov, M. (2002) in *Biochemistry of Lipids, Lipoproteins and Membranes*, ed. Vance DEVaJE (Elsevier Science B. V.), p. 35.

32. Lomize AL, Pogozheva ID, Lomize MA, & Mosberg HI (2006) *Protein Sci* **15,** 1318-1333.

33. Lomize AL, Pogozheva ID, Lomize MA, & Mosberg HI (2007) *BMC Struct Biol* **7,** 44.

34. Lomize MA, Lomize AL, Pogozheva ID, & Mosberg HI (2006) *Bioinformatics* **22,** 623-625.

35. Piknova B, Perochon E, & Tocanne JF (1993) *Eur J Biochem* **218,** 385-396.

36. Jordan P, Fromme P, Witt HT, Klukas O, Saenger W, & Krauss N (2001) *Nature* **411,** 909-917.

37. Sui H, Han BG, Lee JK, Walian P, & Jap BK (2001) *Nature* **414,** 872-878.

38. Sung MT, Lai YT, Huang CY, Chou LY, Shih HW, Cheng WC, Wong CH, & Ma C (2009) *Proc Natl Acad Sci U S A* **106,** 8824-8829.

39. Openshaw AE, Race PR, Monzo HJ, Vazquez-Boland JA, & Banfield MJ (2005) *J Biol Chem* **280,** 35011-35017.

40. Colussi T, Parsonage D, Boles W, Matsuoka T, Mallett TC, Karplus PA, & Claiborne A (2008) *Biochemistry* **47,** 965-977.

41. Hwang PM, Zhou N, Shan X, Arrowsmith CH, & Vogel HJ (1998) *Biochemistry* **37,** 4288-4298.

42. Hunter HN, Fulton DB, Ganz T, & Vogel HJ (2002) *J Biol Chem* **277,** 37597-37603.

43. Sasaki K, Dockerill S, Adamiak DA, Tickle IJ, & Blundell T (1975) *Nature* **257,** 751-757.

44. Florkin M (1957) *Rev Med Liege* **12,** 139-144.

45. Northrop JH (1929) *Science* **69,** 580.

46. Herriott RM & Northrop JH (1936) *Science* **83,** 469-470.

47. Rawlings ND, Barrett AJ, & Bateman A (2010) *Nucleic Acids Res* **38,** D227-233.

48. Manning G, Whyte DB, Martinez R, Hunter T, & Sudarsanam S (2002) *Science* **298,** 1912-1934.

49. Ren Q & Paulsen IT (2007) *J Mol Microbiol Biotechnol* **12,** 165-179.

50. Pitzer F, Dantes A, Fuchs T, Baumeister W, & Amsterdam A (1996) *FEBS Lett* **394,** 47-50.

51. Delic J, Morange M, & Magdelenat H (1993) *Mol Cell Biol* **13,** 4875-4883.

52. Havens CG, Ho A, Yoshioka N, & Dowdy SF (2006) *Mol Cell Biol* **26,** 4701-4711.

53. Sakai J, Duncan EA, Rawson RB, Hua X, Brown MS, & Goldstein JL (1996) *Cell* **85,** 1037-1046.

54. Lyko F, Martoglio B, Jungnickel B, Rapoport TA, & Dobberstein B (1995) *J Biol Chem* **270,** 19873-19878.

55. Arbige MV & Pitcher WH (1989) *Trends in Biotechnology* **7,** 330-335.

56. Gupta R, Beg QK, & Lorenz P (2002) *Appl Microbiol Biotechnol* **59,** 15-32.

57. Maurer KH (2004) *Curr Opin Biotechnol* **15,** 330-334.

58. Gupta R & Ramnani P (2006) *Appl Microbiol Biotechnol* **70,** 21-33.

59. Singh A, Ghosh VK, & Ghosh P (1994) *Letters in Applied Microbiology* **18,** 177-180.

60. James P, Quadroni M, Carafoli E, & Gonnet G (1993) *Biochem Biophys Res Commun* **195,** 58-64.

61. Yates JR, 3rd, Speicher S, Griffin PR, & Hunkapiller T (1993) *Anal Biochem* **214,** 397-408.

62. Chatterjee S, Schoepe J, Lohmer S, & Schomburg D (2005) *Protein Expr Purif* **39,** 137-143.

63. Zhu F, Han B, Kumar P, Liu X, Ma X, Wei X, Huang L, Guo Y, Han L, Zheng C*, et al*. (2010) *Nucleic Acids Res* **38,** D787-791.

64. Quesada V, Ordonez GR, Sanchez LM, Puente XS, & Lopez-Otin C (2009) *Nucleic Acids Res* **37,** D239-243.

65. Alcalay RN, Caccappolo E, Mejia-Santana H, Tang MX, Rosado L, Ross BM, Verbitsky M, Kisselev S, Louis ED, Comella C*, et al*. *Arch Neurol* **67,** 1116-1122.

66. Gribouval O, Gonzales M, Neuhaus T, Aziza J, Bieth E, Laurent N, Bouton JM, Feuillet F, Makni S, Ben Amar H*, et al*. (2005) *Nat Genet* **37,** 964-968.

67. Zhu S, Hsu AP, Vacek MM, Zheng L, Schaffer AA, Dale JK, Davis J, Fischer RE, Straus SE, Boruchov D*, et al*. (2006) *Hum Genet* **119,** 284-294.

68. De Strooper B (2003) *Neuron* **38,** 9-12.

69. De Strooper B & Annaert W (2001) *J Cell Biol* **152,** F17-20.

70. Petek E, Windpassinger C, Vincent JB, Cheung J, Boright AP, Scherer SW, Kroisel PM, & Wagner K (2001) *Am J Hum Genet* **68,** 848-858.

71. Kopan R & Ilagan MX (2004) *Nat Rev Mol Cell Biol* **5,** 499-504.

72. Weihofen A, Lemberg MK, Ploegh HL, Bogyo M, & Martoglio B (2000) *J Biol Chem* **275,** 30951-30956.

73. Haass C & Steiner H (2002) *Trends Cell Biol* **12,** 556-562.

74. Steiner H, Kostka M, Romig H, Basset G, Pesold B, Hardy J, Capell A, Meyn L, Grim ML, Baumeister R, *et al*. (2000) *Nat Cell Biol* **2,** 848-851.

75. Sato C, Takagi S, Tomita T, & Iwatsubo T (2008) *J Neurosci* **28,** 6264-6271.

76. Wang J, Brunkan AL, Hecimovic S, Walker E, & Goate A (2004) *Neurobiol Dis* **15,** 654-666.

77. Wang J, Beher D, Nyborg AC, Shearman MS, Golde TE, & Goate A (2006) *J Neurochem* **96,** 218-227.

78. Iben LG, Olson RE, Balanda LA, Jayachandra S, Robertson BJ, Hay V, Corradi J, Prasad CV, Zaczek R, Albright CF, *et al*. (2007) *J Biol Chem* **282,** 36829-36836.

79. Friedmann E, Lemberg MK, Weihofen A, Dev KK, Dengler U, Rovelli G, & Martoglio B (2004) *J Biol Chem* **279,** 50790-50798.

80. Martoglio B & Golde TE (2003) *Hum Mol Genet* **12 Spec No 2,** R201-206.

81. Xia W & Wolfe MS (2003) *J Cell Sci* **116,** 2839-2844.

82. Okamoto K, Moriishi K, Miyamura T, & Matsuura Y (2004) *J Virol* **78,** 6370-6380.

83. Weihofen A, Binns K, Lemberg MK, Ashman K, & Martoglio B (2002) *Science* **296,** 2215-2218.

84. Weihofen A & Martoglio B (2003) *Trends Cell Biol* **13,** 71-78.

85. Lemberg MK & Martoglio B (2002) *Mol Cell* **10,** 735-744.

86. Hua X, Sakai J, Brown MS, & Goldstein JL (1996) *J Biol Chem* **271,** 10379-10384.

87. Sakai J, Rawson RB, Espenshade PJ, Cheng D, Seegmiller AC, Goldstein JL, & Brown MS (1998) *Mol Cell* **2,** 505-514.

88. Duncan EA, Dave UP, Sakai J, Goldstein JL, & Brown MS (1998) *J Biol Chem* **273,** 17801-17809.

89. Baonza A, Casci T, & Freeman M (2001) *Curr Biol* **11,** 396-404.

90. Wasserman JD & Freeman M (1998) *Cell* **95,** 355-364.

91. Freeman M (1994) *Mech Dev* **48,** 25-33.

92. Wasserman JD, Urban S, & Freeman M (2000) *Genes Dev* **14,** 1651-1663.

93. Urban S, Brown G, & Freeman M (2004) *Development* **131,** 1835-1845.

94. Fuss B & Hoch M (2002) *Curr Biol* **12,** 171-179.

95. Tsruya R, Wojtalla A, Carmon S, Yogev S, Reich A, Bibi E, Merdes G, Schejter E, & Shilo BZ (2007) *EMBO J* **26,** 1211-1220.

96. Urban S, Schlieper D, & Freeman M (2002) *Curr Biol* **12,** 1507-1512.

97. Klambt C (2002) *Curr Biol* **12,** R21-23.

98. Yogev S, Schejter ED, & Shilo BZ (2008) *EMBO J* **27,** 1219-1230.

99. Koonin EV, Makarova KS, Rogozin IB, Davidovic L, Letellier MC, & Pellegrini L (2003) *Genome Biol* **4,** R19.

100. Lemieux MJ, Fischer SJ, Cherney MM, Bateman KS, & James MN (2007) *Proc Natl Acad Sci U S A* **104,** 750-754.

101. Brenner S (1988) *Nature* **334,** 528-530.

102. Wang Y, Zhang Y, & Ha Y (2006) *Nature* **444,** 179-180.

103. Clemmer KM, Sturgill GM, Veenstra A, & Rather PN (2006) *J Bacteriol* **188,** 3415-3419.

104. Wang Y & Ha Y (2007) *Proc Natl Acad Sci U S A* **104,** 2098-2102.

105. Nakagawa T, Guichard A, Castro CP, Xiao Y, Rizen M, Zhang HZ, Hu D, Bang A, Helms J, Bier E*, et al*. (2005) *Dev Dyn* **233,** 1315-1331.

106. Pascall JC & Brown KD (2004) *Biochem Biophys Res Commun* **317,** 244-252.

107. Urban S & Freeman M (2003) *Mol Cell* **11,** 1425-1434.

108. Li YM, Lai MT, Xu M, Huang Q, DiMuzio-Mower J, Sardana MK, Shi XP, Yin KC, Shafer JA, & Gardell SJ (2000) *Proc Natl Acad Sci U S A* **97,** 6138-6143.

109. Wu Z, Yan N, Feng L, Oberstein A, Yan H, Baker RP, Gu L, Jeffrey PD, Urban S, & Shi Y (2006) *Nat Struct Mol Biol* **13,** 1084-1091.

110. Baker RP, Young K, Feng L, Shi Y, & Urban S (2007) *Proc Natl Acad Sci U S A* **104,** 8257-8262.

111. Maegawa S, Ito K, & Akiyama Y (2005) *Biochemistry* **44,** 13543-13552.

112. Maegawa S, Koide K, Ito K, & Akiyama Y (2007) *Mol Microbiol* **64,** 435-447.

113. Strisovsky K, Sharpe HJ, & Freeman M (2009) *Mol Cell* 36, 1048-1059.

114. Lee JR, Urban S, Garvey CF, & Freeman M (2001) Cell 107, 161-171.

Chapter 2: Conservation of proteolytic profiles among prokaryotic and eukaryotic rhomboids.

Introduction

Rhomboids constitute a nearly ubiquitous subfamily of I-CLiPs believed to play a role in intercellular signaling (1-10). In *D. melanogaster,* rhomboid-1 resides in the Golgi membrane and has three known substrates - Gurken, Keren and Spitz - all of which are precursors of epidermic growth factor receptor (EGFR) ligands (11). These substrates share a similar architecture (**Fig. 2.1a**) with a large N-terminal extracellular/luminal domain (~500 amino acid residues long) that constitutes the EGFR ligand, and a single transmembrane domain (TMD) followed by a short (~40-80 amino acid residues long) cytoplasmic region. These ligands remain inactive while anchored in the ER membrane and their degradation by rhomboid-1 is triggered only after they are transported to the Golgi membrane by the membrane protein Star (12-14). Rhomboid-1 activity results in the release of the EGFR ligand domain into the Golgi lumen from where it is exported to the cell exterior to activate EGFRs in a neighboring cell (11, 12, 14-16). More recently, rhomboid orthologues have been found to affect mitochondrial membrane remodeling in *S. cerevizae* (17) and in *D. melanogaster* (18), while in mammals, the precise function and physiological significance of rhomboids RHBDL2 (19) and PARL (17) have yet to be delineated. In prokaryotes, the rhomboid AarA from the gram-negative bacteria *P. stuarti* (20) has been found to be implicated in the generation of quorum-sensing signals (10, 21, 22). Remarkably, it was also demonstrated that AarA could trigger EGFR signaling when expressed in flies and that rhomboid-1 could rescue the bacterial phenotype caused by AarA deletion (21). Although the overall sequence homology among rhomboids is low, they all share a common architecture with multiple TMDs and an

absolute conservation of catalytic Ser and His amino acid residues (10). Recent X-ray structures of *E. coli* GlpG (23-25) and *H. influenzae* (26) rhomboids confirmed the role of the Ser-His dyad in catalysis and showed that prokaryotic rhomboids are built from six transmembrane helices with an active site located in a V-shaped cup that is accessible to the extracellular medium.

Rhomboids across species appear to share substrate specificity. Initially, Urban et al employed cleavage assays in cultured cells to show that several prokaryotic rhomboids, including those from *E. coli*, *P. aeruginosa* and *B. subtilis*, could cleave the *D. melanogaster* substrates Gurken, Keren and Spitz (10). However, the precise proteolytic profiles of rhomboid-1 and even of key prokaryotic rhomboids against these ligand precursors are unknown. So far, the GlpG cleavage sites against Spitz and Gurken have just been inferred from experiments employing a minimum substrate motif or indirect cleavage site determination assays. Using gel mobility assays to analyze rhomboid-1 reaction products the proteolytic site in Spitz was roughly mapped to a small N-terminal region (ASIASGA) within its predicted TMD (16). Baker et al. later employed mass spectrometry and activity assays in detergent, using *E. coli* GlpG and an artificial substrate containing just the ASIASGA motif of Spitz to show cleavage at two adjacent sites (after Ala141 and Gly143, respectively) within this motif (27). Regarding Gurken, Maegawa et al. (28) constructed a Gurken based chimeric substrate and used cysteine labeling experiments in *E. coli* membranes to map GlpG cleavage to a juxtamembrane region (after His246).

To compare the proteolytic profiles between eukaryotic and prokaryotic rhomboids we have designed substrate chimeras derived from Gurken, Keren and Spitz, in which their transmembrane and cytoplasmic regions were preserved, while their

epidermic growth factor receptor ligand ectodomain was replaced by maltose binding protein (MBP). In vitro activity assays using purified substrates and enzymes solubilized in detergent showed that rhomboid-1, RHBDL2 from *H. sapiens*, PA3086 from *P. aeruginosa* and GlpG from *E. coli* displayed proteolytic activity against all the substrate chimeras. Mass spectrometry analysis of the reaction products indicated that all eukaryotic and prokaryotic rhomboids tested displayed identical proteolytic profiles against each of the substrate chimeras.

Material and Methods

Chemicals.

Protein concentration was determined with Coomassie Protein Assay Reagent from Thermo. Anti-MBP (E8032S) was purchased from New England BioLab and anti-poly histidine antiserum (H1029) was bought from Sigma. Secondary anti-mouse IgG HRP conjugated antibody was obtained from Bethyl. ECL western blotting substrate (3210G) was purchased from Pierce. All detergents used were from Anatrace.

Design and production of rhomboid substrates.

We designed protein chimeras (**Fig. 2.1**) in which the highly glycosylated EGFR ligand ectodomain of *D. melanogaster* Gurken, Keren and Spitz was substituted by MBP to produce two versions: one containing the region of the TMD, flanked by approximately twenty N-terminal amino acid residues and the entire C-terminal region of the wild-type sequences; and the other in which the C-terminal region was missing. cDNAs for Gurken, Keren and Spitz were obtained from the Drosophila Genomics Resource Center and MBP (without a signal sequence) was amplified from plasmid H-MBP-3C (29). The chimeras also included a unique thrombin cleavage site (LVPRGS) to aid in the ensuing

mass spectrometry experiments and a C-terminal His_6-tag for purification (**Fig. 2.1b**). These chimeras termed, Gurken-TMD/CT, Keren-TMD/CT, Spitz-TMD/CT, Gurken-TMD, Keren-TMD and Spitz-TMD were subcloned into pET-29b plasmids. The MBP domain was cloned in between NdeI and KpnI restriction sites. The sequences corresponding to the cytoplasmic domain, transmembrane domain and 20 amino acids upstream from the transmembrane domain of each substrate were cloned in between NheI and BamHI restriction sites.

All the substrate chimeras were expressed in CE43 (DE3) (OverExpress) *E. coli* cells. Freshly transformed cells were grown at 37°C in enriched media (composition per liter: 10 g Tryptone, 5 g yeast extract, 8.5 g Na_2HPO_4, 3.0 g KH_2PO_4, 0.5 g NaCl, 1.0 g NH_4CL, 10 mg thiamine, 6.0 g glucose) containing 34 μg/mL of kanamycin and were induced OD_{600} = 0.7–0.8 with 0.4 μM IPTG, and grown for 4 h at 37°C. Following incubation at 37°C for 4 h the cells from 1000 ml cultures were harvested by centrifugation at 4000xg and the pellet was resuspended in 40 mL of lysis buffer (50 mM tris-HCl, pH 7.4, 300 mM NaCl, 10 mM imidazol and 10% glycerol) containing a protease inhibitor cocktail (100 μM PMSF, 1 μg/ml leupeptin, 1 μg/ml pepstatin and 1μg/ml of aprotinin). This suspension was incubated for 30 min on ice after the addition of 100 μg each of RNase, DNase and lysozyme, followed by cell lysis using a microfluidizer (microfluidics M-110P) at 1500 psi. Cell debris was removed by centrifugation at 40,000xg for 20 min, and cytoplasmic *E. coli* membranes were isolated by ultracentrifugation at 100,000xg for 40 min at 4°C. Membranes were resuspended in lysis buffer (36 mL of buffer per 2 grams of inner membranes). The membranes were then solubilized by adding 20% Triton X-100 up to a final concentration of 2%. This suspension was rocked for 4 h at 4°C. Insoluble material was removed by ultracentrifugation at 100,000 x g for 40 min at 4°C. 0.5 ml of Ni-NTA beads (Qiagen)

were added to the supernatant and the suspension was rocked for 4 h at 4°C. The Ni-NTA beads were then washed with 10-bead volumes of lysis buffer, containing 0.2% Triton X-100, washed again with 10-bead volumes of lysis buffer, containing 70 mM imidazole and 0.1% n-Dodecyl-b-D-maltopyranoside (DDM), and finally eluted with 5-bead volumes of lysis buffer containing 250 mM imidazole and 0.1% DDM. This eluate was then dialyzed against PBS buffer at pH 7.4 and stored at -80°C. The identity of each substrate was confirmed by in-gel chymotrypsin digestion and mass spectrometry.

Cloning and production of rhomboids

cDNAs for PA3086 and GlpG were a kind gift from Dr. Matthew Freeman. Codon optimized rhomboid-1 and RHBDL2 synthetic genes were purchased from Biomatik (Wilmington, USA). The rhomboids PA3086 and GlpG were cloned into pET-15b plasmids for inducible expression as N-terminal His_6-tag fusion proteins and expressed in CE43 (DE3) (OverExpress) *E. coli* cells. The rhomboids rhomboid-1 and RHBDL2 were cloned into pET-28b plasmids for inducible expression as N-terminal His_6-tag fusion proteins and expressed in BL21-CodonPlus® (DE3)-RIPL *E. coli* cells.

Fresh transformants were grown in 1000 ml cultures at 30°C in LB broth containing either 100 µg/ml Ampicillin or 34 µg/ml Kanamycin until their induction (OD_{600}= 0.4) with 0.4 µM IPTG, at which point they were incubated for another 12 h at 16°C before harvesting. Following harvesting by centrifugation at 4000xg, the pelleted cells were resuspended in 20 ml of lysis buffer (50 mM tris-HCl, pH 7.4, 700 mM NaCl, and 10% glycerol) containing protease inhibitor cocktail (100µM PMSF, 1µg/ml leupeptin, 1 µg/ml pepstatin and 1µg/ml of aprotinin). The resuspension was incubated for 30 min on ice with 100 µg each of RNase, DNase and lysozyme and the cells were lysed with a microfluidizer (microfluidics M-110P) at 1500 psi. Cell debris was removed

by centrifugation at 40,000x*g* for 20 min, and cytoplasmic *E. coli* membranes were isolated from the supernatant by ultracentrifugation at 100,000x*g* for 1 hour at 4°C. Membranes were resuspended in lysis buffer containing 10 mM imidazole and solubilized in 2% TX-100 for 4 h at 4°C. Insoluble material was removed by ultracentrifugation at 100,000xg for 40 min at 4°C and to the supernatant 0.5 ml of Ni-NTA beads (Qiagen) were added, followed by rocking for 4 h at 4°C. The Ni-NTA beads were then washed with 10-bead volumes of lysis buffer, containing 30 mM imidazole and 0.2% Triton X-100, washed again with 10-bead volumes of lysis buffer, containing 80 mM imidazole and 0.1% DDM, and finally eluted with 5-bead volumes of lysis buffer containing 250 mM imidazole and 0.1% DDM. In the case of rhomboid-1 and RHBDL2 the eluate from the Ni-NTA beads was dialyzed against buffer (50 mM tris-HCl, pH 7.4, 700 mM NaCl, 10% glycerol, 0.1% DDM) and stored at -80°C. The rhomboids PA3086 and GlpG were additionally purified by size exclusion chromatography using a 16/60 HR200 Superdex column (GE Healthcare) run in 50 mM tris-HCl, pH 7.4, 700 mM NaCl, 10% glycerol and 0.1% DDM. The identity of each rhomboid was confirmed by in-gel chymotrypsin digestion and mass spectrometry. Site directed mutagenesis (QuikChange site-directed mutagenesis kit, Stratagene) was employed to obtain the active site GlpG, rhomboid-1 and RHBDL2 variants. Each protein variant was purified as described above.

The membrane domain of PA3086 was expressed and purified according to the protocol described above for the full-length rhomboid with minor modifications. The membrane domain of GlpG was obtained by the digestion of full-length GlpG with trypsin (1/70 w/w ratio) for 48 hours at 4°C, according to a published protocol (23). The resulting fragment was purified by size exclusion chromatography.

Rhomboid In vitro Activity Assay.

Routinely 0.2 μg/ml of purified substrate and 0.05-0.1 μg/ml of purified enzyme were incubated overnight at 37°C in buffer (50 mM tris-HCl, pH 7.4, 700 mM NaCl, 10% glycerol) containing 0.1% DDM, protease inhibitor cocktail (100μM PMSF, 1μg/ml leupeptin, 1 μg/ml pepstatin and 1μg/ml of aprotinin) and 2 mM β-mercaptoethanol. The reaction was quenched by the addition of 5x SDS-PAGE sample loading buffer. The reactions with prokaryotic rhomboids were evaluated using 12% SDS-PAGE (2 μg of protein loaded) gels stained with coomassie blue. For anti-MBP western blotting of the reactions with eukaryotic rhomboids, 2 μg of protein was resolved in a 12% SDS-PAGE gel, blotted onto PVDF membrane and immunoreacted with anti-MBP (1:3000 dilution) primary antibody followed by reaction with secondary antibody (1:10000 dilution) and visualization using chemiluminescence (Pierce).

Cleavage Site Determination by mass spectrometry.

To increase the yield of product for its analysis by mass spectrometry, a total of 200 μg of chimeric substrate were mixed with 50 μg of rhomboid in a 1.0 ml reaction buffer (50 mM tris-HCl, pH 7.4, 700 mM NaCl, 10% glycerol) containing 0.1% DDM, protease inhibitor cocktail (100μM PMSF, 1μg/ml leupeptin, 1μg/ml pepstatin and 1μg/ml of aprotinin) and 2 mM β-mercaptoethanol. The reaction was then incubated at 37°C for 8 hr in a shaker. Following digestion a 0.25 ml aliquot of Ni-NTA beads were added to the mixture and incubated by rocking for 1 hr at 4°C. The Ni-NTA beads containing bound substrate, C-terminal product and enzyme were removed by filtration, and the supernatant containing the N-terminal reaction product was concentrated (Amicon Ultra, 30 kDa cut of) to a final volume of 50μl (**Fig. 2**5A). The product was then precipitated by the addition of 1 ml of pre-chilled 10% trichloroacetic acid (TCA) in acetone and

recovered by centrifugation at 17,000xg for 10 min using a microcentrifuge. The resulting pellet was washed once with 1 ml of pre-chilled acetone, dried at room temperature for 10 min and resuspended in 1 ml of 8 M urea (in 50 mM tris-HCl at pH 7.4). Following solubilization, 3 steps of concentration/dilution using 50 mM ammonium bicarbonate at pH 7.4 removed the urea. Finally, the N-terminal product was concentrated (Amicon Ultra, 30 kDa cut of) to a final volume of 50 μl. At this stage, the protein ranged routinely in between 0.5-1.5 mg/ml and the purity of the isolated reaction product was evaluated by SDS-PAGE. The sample was then divided into two 20 μl aliquots. To one aliquot 5 μL of 5% trifluoroacetic acid (TFA) in acetone was added and stored at 4°C. To the second aliquot one Unit of human alpha thrombin (Enzyme Research Laboratory, USA) was added and following incubation for 1 hour at 4°C the reaction was stopped by adding 5% TFA in acetone. Both aliquots were then analyzed by MALDI-TOF (matrix-assisted laser desorption/ionization time-of-flight) mass spectrometry. For MALDI-TOF analysis, a modified thin-layer method of sample preparation was implemented (30, 31). Saturated α-cyano-4-hydroxycinnamic acid (4HCCA) matrix was prepared in 2 parts acetonitrile, 1 part water, and 0.1% final TFA solvent mixture and further diluted 4-fold with isopropanol. About 40 ml of diluted matrix solution was applied over the entire gold surface of the sample plate. The organic solvents were allowed to dry in ambient air. The matrix was then gently wiped with a tissue, leaving behind a faint layer of 4HCCA only visible as a yellowish reflection. We will refer to this preparation as the "ultra thin layer". Each protein sample was diluted with the matrix solution to a concentration of 1 mM. A small aliquot (1 ml) of protein-matrix solution was spotted onto the sample plate pretreated to form the ultra thin layer. The spots were then washed for a few seconds with 0.1% aqueous TFA to help remove salts. The 0.1% TFA was subsequently aspirated with a vacuum line. Myoglobin (Equus caballus) was used as an external

calibrant. The samples and the calibrant were analyzed in a MALDI-TOF mass spectrometer (Voyager DE-STR, Applied Biosystems) operating in linear delayed extraction mode. This instrument uses a nitrogen laser that delivers pulses of ultraviolet light (λ = 337 nm) at 3 Hz to the matrix spots. Five hundred individual scans were averaged into a single spectrum. The spectra were smoothed and further analyzed using the software M-over-Z.

Results

Design and production of substrate chimeras derived from rhomboid-1 physiological substrates Gurken, Keren and Spitz.

Our goal was to develop a novel in vitro assay that would allow us to determine the proteolytic profiles of rhomboids against substrates resembling the *D. melanogaster* EGFR ligand precursors Gurken, Keren and Spitz. These substrates are membrane proteins containing a single transmembrane domain, an extra cellular domain and a cytoplasmic domain (**Fig. 2.1a**). The extracellular domain is EGFR ligand that is liberated by the rhomboid **(12, 32-35)**. These physiological substrates have been employed previously to test the activity of rhomboids from several species in cultured cells **(10)**. However, their use for in vitro biochemical assays is precluded by low expression levels **(36)**. To circumvent expression issues we designed protein chimeras in which the highly glycosylated EGFR ligand ectodomain of the physiological substrates (**Fig. 2.1b**) was substituted by MBP (MW=42 kDa). The resulting chimeras contained MBP at the N-terminus followed by the TMD (including ~20 N-terminal juxtamembrane amino acid residues) and the soluble cytoplasmic domain of each physiological substrate (**Fig. 2.1c**). These chimeras termed G (Gurken-TMD/CT), K (Keren-TMD/CT) and S (Spitz-TMD/CT) also included a unique thrombin cleavage site (LVPRGS) to aid in

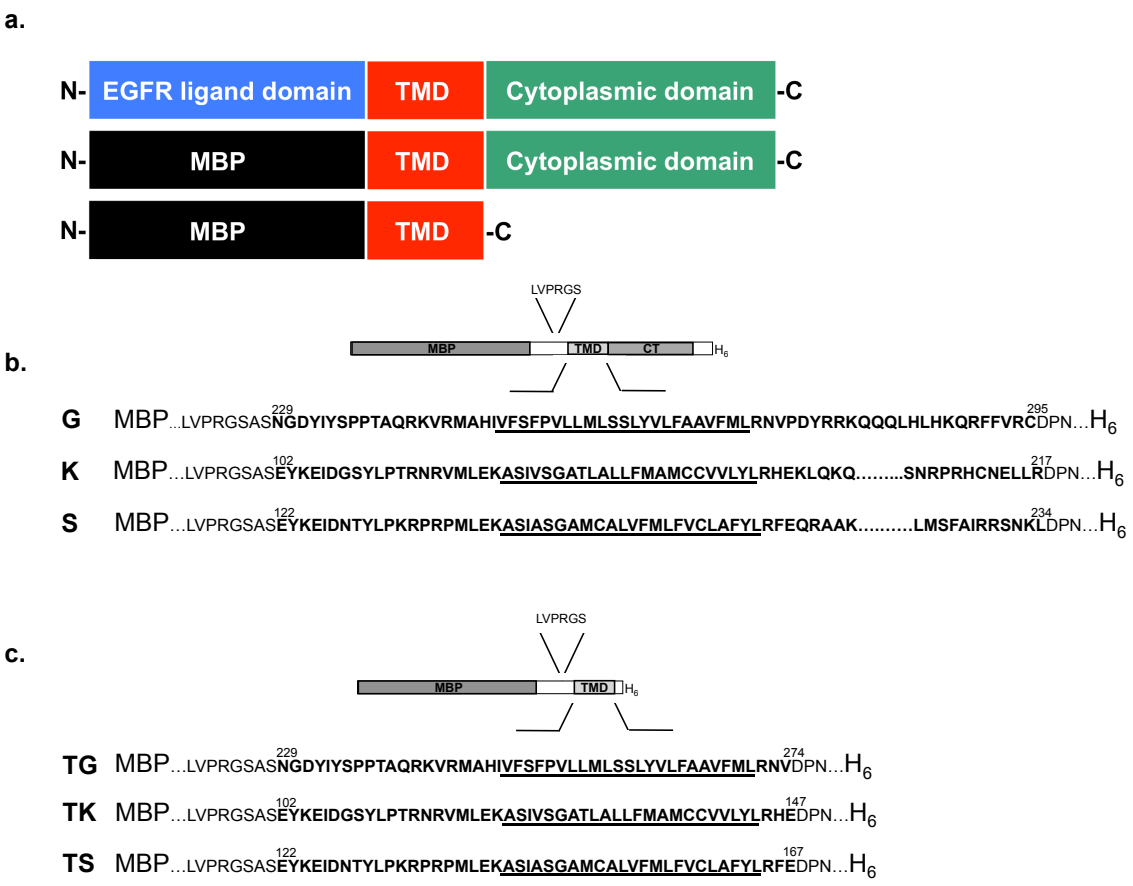

Figure 2.1. Design of substrate chimeras derived from Gurken, Keren and Spitz. a). Schematic displaying the domain structure of the *D. melanogaster* rhomboid-1 substrates Gurken, Keren and Spitz and the chimeric substrates used in this study. The EGFR ligand domain of the physiological substrates is represented in blue, the TMD in red and the cytoplasmic domain in green. In our chimeras the EGFR domain was substituted by maltose binding protein (MBP; in black) to produce two versions: one containing the juxtamembrane, transmembrane and cytosolic regions; and the other containing only the juxtamembrane and transmembrane regions. **b)** Sequence detail of the chimeric substrates Gurken-TMD/CT, Keren-TMD/CT and Spitz-TMD/CT. A thrombin cleavage site (discontinuous underline) separates MBP from the amino acid sequence corresponding to the natural substrates (depicted in bold) and the C-terminus includes a His_6tag for purification. The predicted TMD is underlined. The location where the sequences where truncated to produce the substrates Gurken-TMD, Keren-TMD and Spitz-TMD that lack the cytosolic C-terminal domain is given. Specific amino acid residues are numbered for clarity.

the ensuing mass spectrometry experiments and a C-terminal His_6-tag for purification. Urban et al. had shown that the TMD of Spitz **(16)** contains the sequence determinants to trigger cleavage. However, studies of thrombomodulin cleavage by RHBDL2 have raised the possibility that the C-terminal region of the substrates might also play a role in substrate recognition **(19)**. Thus, to compare the effect of the cytoplasmic region on activity chimeras that lacked this fragment were also made, termed TG (Gurken-TMD), TK (Keren-TMD) and TS (Spitz-TMD). These chimeras could be overexpressed in high yield in *E. coli* (10-20 mg of protein per liter of *E. coli* culture), extracted from the cytoplasmic membrane fraction with 2% TX-100, and detergent exchanged to 0.1% DDM and purified using Ni-NTA chromatography (**Fig. 2.2a**). The purified chimeras displayed good stability and could be stored either at 4^{o}C for over a month or at 37^{o}C for over a week without detectable degradation. Incubation of the chimeras with thrombin resulted in their efficient degradation (**Fig. 2.2b**) into a band with an apparent molecular weight of ~43 kDa. Mass spectrometry analysis of this band showed it corresponds to MBP plus a short linker region, produced by thrombin cleavage in between the amino acid residues Arg and Gly within the engineered thrombin cleavage site (not shown). The C-terminal product containing the selected region of Gurken, Keren or Spitz migrated in the front of the gel (not shown). Since the thrombin cleavage site is ~20 amino acid residues away from the TM region, we argued that proteolytic cleavage by rhomboid would yield a ~43 kDa band corresponding to MBP plus a short C-terminal tail.

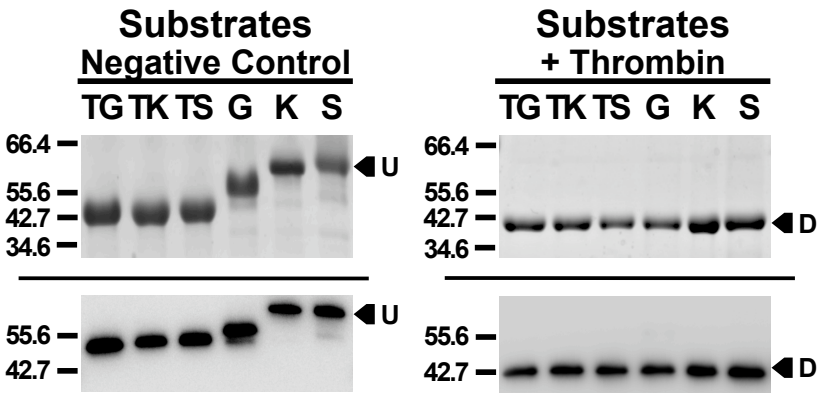

Fig. 2.2. Purification of the substrate chimeras. 12% SDS-PAGE analysis of coomassie stained chimeric substrates following purification (**left panel**) and after digestion with thrombin (**right panel**). TG, TK, TS G, K, and S, abbreviate the substrates: Gurken-TMD, Keren-TMD, Spitz-TMD, Gurken-TMD/CT, Keren-TMD/CT and Spitz-TMD/CT respectively. Protein bands corresponding to the undigested (**U**) and digested (**D**) substrates are indicated with an arrow on the right and the molecular weight marker positions are shown on the left. Identities of the bands were confirmed by western blot using anti-MBP as primary antibody. Western are shown below the SDS-PAGE.

Production of prokaryotic and eukaryotic rhomboids.

For our assays to establish the proteolytic profiles of prokaryotic and eukaryotic rhomboids we selected PA3086 from *P. aeruginosa*, GlpG from *E. coli*, rho-1 from *D. melanogaster* and RHBDL2 from *H. sapiens*. All these rhomboids were selected because they have shown to display proteolytic activity in in vitro assays (5, 23, 24, 26). In addition, structural information for the transmembrane region of GlpG (23) and the cytoplasmic region of PA3086 (37) exists. Moreover, rho-1 from *Drosophila* is the only rhomboid from which the physiological substrates and the biological function are well known. Finally, RHBDL2 rhomboid provides context into human biology (19, 38). We overexpressed the rhomboids GlpG and PA3086 from containing N-terminal His_6-tags in *E. coli*. Induction with IPTG resulted in the overexpression of both GlpG and PA3086 (**Fig. 2.3a**, lane 2) with apparent molecular weights of 28 and 30 kDa, respectively, in agreement with the expected molecular weight of 32 kDa for His_6-GlpG and 33 kDa for His_6-PA3086. Membrane extraction using 2% Triton X-100 resulted in further enrichment of the target proteins (**Fig. 2.3a**, lane 3). A single Ni-NTA chromatography step allowed us to exchange detergent to 0.1% DDM and final purification was achieved using gel filtration (**Fig. 2.3a**, lanes 4-5). Typical yields obtained ranged from 3-5 mg of purified protein per liter of *E. coli* culture. Anti-His_6 western blotting (**Fig. 2.3b**) and peptide mapping using mass spectrometry (not shown) confirmed the identities of the purified proteins. This purification protocol was applied for the purification of PA3086 with similar results (**Fig. 2.3b**). The membrane domain of GlpG should include a chymotrypsin-removing step, which consists of a batch-wise equilibration in *p*-aminobenzamidine-sepharose after digestion with chymotrypsin. This step is important in order that avoid proteolysis activity from other enzyme than rhomboid. Then the protein was purified using 16/60 HiLoad HR Superdex. Membrane domain of PA3086 (ΔNT-PA3086) was

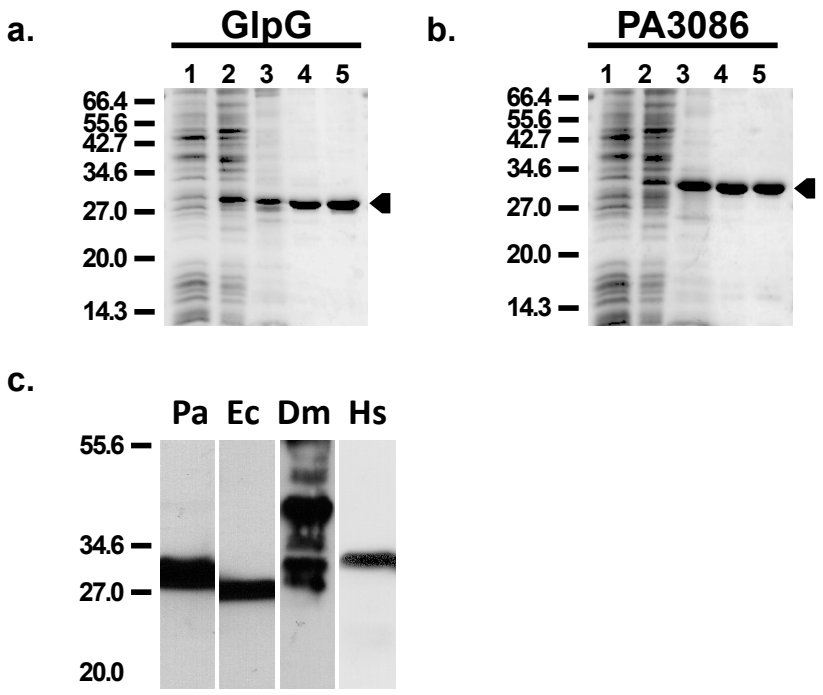

Fig. 2.3. Purification of prokaryotic and eukaryotic rhomboids. a). 15% SDS-PAGE analysis of coomassie stained *E. coli* GlpG rhomboid at various stages during purification. Lane 1, *E. coli* lysate before IPTG induction; lane 2, *E. coli* lysate after harvesting; lane 3, *E. coli* cytoplasmic membranes solubilized in TX-100; lane 4, after Ni-NTA chromatography purification in DDM; lane 5, after size exclusion chromatography over a HR200 superdex column. **b).** As in **figure 3a,** but for *P. aeruginosa* PA3086 rhomboid. **c).** Anti-His_6-tag western blot analysis of purified rhomboids. Ec, Pa, Dm and Hs abbreviate the rhomboids *E. coli* GlpG, *P. aeruginosa* PA3086, *D. melanogaster* rhomboid-1 and *H. sapiens* RHBDL2, respectively. The molecular weight marker positions are shown on the left.

purified by a single Nickel-bead IMA purification step. The protein was then purified by size-exclusion chromatography employing 16/60 HiLoad HR Superdex.

Rho-1 and RHBDL2 have been previously expressed in *E. coli* and used, in crude membrane solubilizated form, for activity assays (39). We improved the expression levels of rhomboid-1 and RHBDL2 in *E. coli* using codon-optimized synthetic genes. Following membrane solubilization and Ni-NTA chromatography these eukaryotic rhomboids could be purified, albeit with lower yield and quality, compared to their prokaryotic counterparts (**Fig. 2.3b**). For rhomboid-1 two main anti-His_6 immunoreactive bands could be observed, one band migrated with an apparent molecular weigh of ~42 kDa, in good agreement with the expected molecular weight of 41.5 kDa for the His_6-tagged protein. A second band at ~30 kDa most likely corresponds to a N-terminal proteolytic fragment of the protein. For RHBDL2 a single band was observed with an apparent molecular weight of ~33 kDa in agreement with the expected molecular weight of 34.9 kDa for the His_6-tagged protein.

Prokaryotic and eukaryotic rhomboids display proteolytic activity against the substrate chimeras.

As evaluated by SDS-PAGE, incubation of purified GlpG and PA3086 with the purified substrate chimeras G, K and S in 0.1% DDM at 37^{o}C for 8 hr resulted in their degradation into a ~43 kDa species (**Fig. 2.4**, top panels). Anti-MBP western blotting indicated that this band corresponded to the MBP-containing N-terminal product (**Fig. 2.4**, lower panels). Densitometry analysis showed that GlpG and PA3086 could cleave these substrates efficiently. GlpG displayed a slight preference for G, whereas PA3086 could also cleave K with 100% efficiency. Both enzymes displayed lower activity against S. In these assays, the cytoplasmic regions of the physiological substrates Gurken,

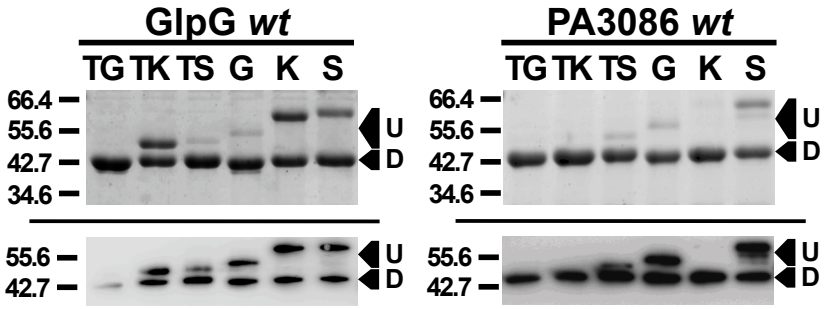

Fig. 2.4. Digestion of chimeric substrates by prokaryotic and eukaryotic rhomboids. **Left panel:** 12% SDS-PAGE analysis of coomassie stained digestion products of substrates Gurken-TMD, Keren-TMD, Spitz-TM, Gurken-TMD/CT, Keren-TMD/CT and Spitz-TMD/CT (TG, TK, TS, G, K and S, respectively) were incubated with purified GlpG at 37^{o}C for 8 hr in 0.1% DDM. Below the gel is shown the resulting western blot using anti-MBP as primary antibody. **Right panel.** As in the left panel but for *P. aeruginosa* rhomboid PA3086. Protein bands corresponding to the undigested (**U**) and digested (**D**) substrates are indicated with an arrow on the right and the molecular weight marker positions are shown on the left.

Keren and Spitz had no effect on activity, as opposed to previous studies that raised the possibility that the C-terminal region of the substrates might play a role in substrate recognition (19). The observed proteolytic activity was enzyme specific since alanine substitution of either of the catalytic residues Ser201 or His254 in GlpG, or Ser207 or His257 in PA3086 abolished substrate degradation completely (**Fig. 2.5**). Notably, mutation of Asn154 in GlpG - predicted to play a secondary role in catalysis (10) – almost abolished degradation of the Keren and Spitz chimeras, but not of the Gurken chimera. Notably, mutation of or Asn157 in PA3086 reduced but did not abolish activity against any of the substrates, suggesting this position is less sensitive in this enzyme.

Recent X-ray structures of *E. coli* (23-25) rhomboids corresponded to a catalytically active truncated form of the protein consisting only of the transmembrane region. A recent solution NMR structure of PA3086 from our laboratory (37) suggested that the N-terminus of rhomboids could interact with the cellular membrane and raised the possibility that this domain interacts with the cytoplasmic domain of the substrates (37). To test this possibility, we generated GlpG and PA3086 (ΔNT-GlpG and ΔNT-PA3086, respectively) variants lacking the N-terminal cytoplasmic region and tested their activity against the chimeras Gurken-TMD, Keren-TMD and Spitz-TMD. As shown in figure 6 the truncated enzymes behaved as full-length versions displaying comparable substrate specificity, suggesting that the N-terminal region of these rhomboids does not play a role in our in vitro assays.

Due to the lower purity and yield obtained for rhomboid-1 and RHBDL2, we employed anti-MBP western blotting, instead of SDS-PAGE, to monitor their activity against the substrate chimeras (**Fig. 2.6**). We note that due to the significant substrate loaded more than one MBP inmunoreactive band appear in the substrate-alone control

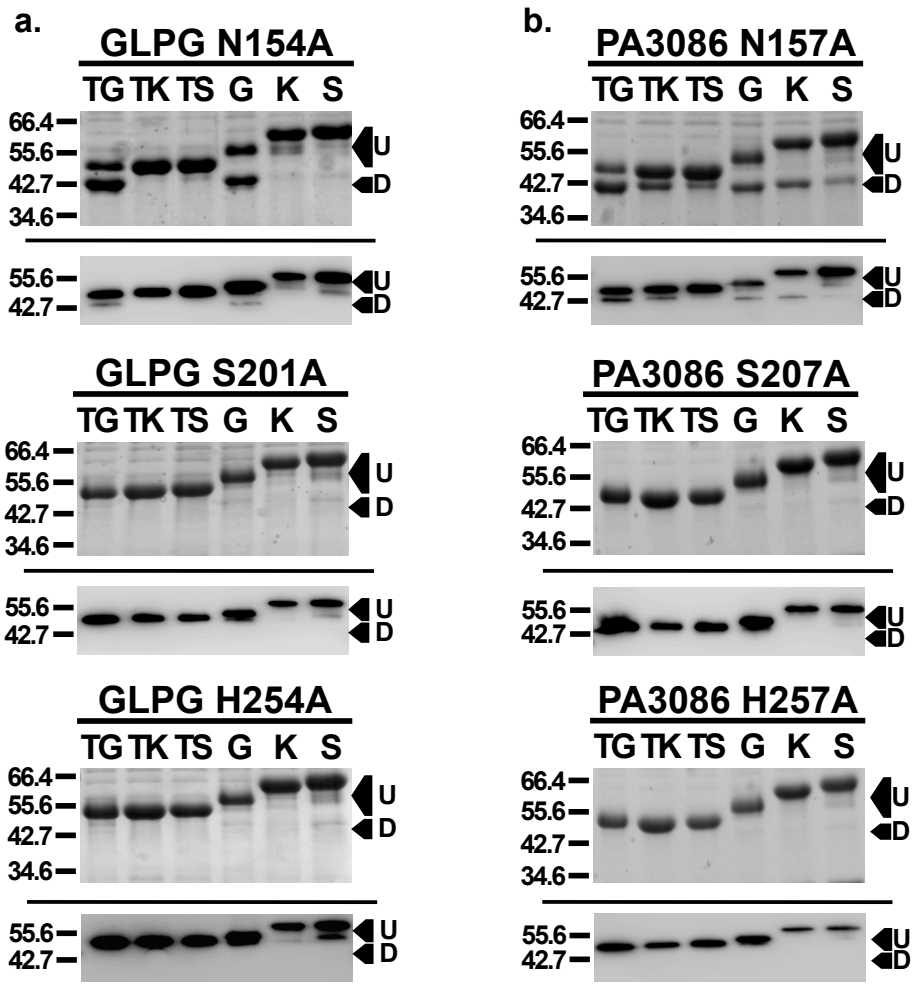

Figure 2.5. Activity of active site rhomboid mutants. **a).** 12% SDS-PAGE analysis of coomassie stained digestion products of substrates Gurken-TMD, Keren-TMD, Spitz-TM, Gurken-TMD/CT, Keren-TMD/CT and Spitz-TMD/CT (TG, TK, TS, G, K and S, respectively). Substrates were incubated with purified GlpG mutants N154A (Upper), S201A (middle) and H254A (lower) at 37oC for 8 hr in 0.1% DDM. Below the gels is shown the resulting western blot using anti-MBP as primary antibody. **b).** As in **a.** but for *P. aeruginosa* rhomboid PA3086 mutants N157A, S207A and H257A. Protein bands corresponding to the undigested (**U**) and digested (**D**) substrates are indicated with an arrow on the right and the molecular weight marker positions are shown on the left.

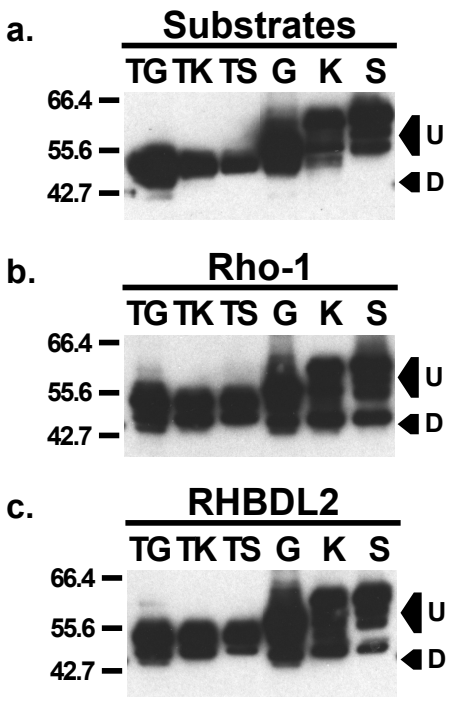

Figure 2.6. Digestion of the chimeric substrates by eukaryotic rhomboids. Western blot analysis of the digestion of substrates: Gurken-TMD, Keren-TMD, Spitz-TM, Gurken-TMD/CT, Keren-TMD/CT and Spitz-TMD/CT (TG, TK, TS, G, K and S, respectively) by eukaryotic rhomboids. The substrates were incubated with purified *D. melanogaster* Rho-1 or *Homo sapiens* RHBDL2 at 37˚C for 8 hr in 0.1% DDM. **a).** Substrate controls, Substrates were incubated without enzyme. **b).** Substrates digested by *D. melanogater* Rho-1. **c).** Substrates digested by RHBDL2**.** Protein bands corresponding to the undigested (**U**) and digested (**D**) substrates are indicated with an arrow on the right and the molecular weight marker positions are shown on the left.

(**Fig. 2.6a**, top panel). Clearly upon incubation with rhomboid-1 or RHBDL2, a new MBP immunoreactive band at ~43 kDa could be identified, indicating that both eukaryotic rhomboids displayed activity against all substrates.

Eukaryotic and prokaryotic rhomboids share cleavage site specificity.

To accurately map the substrate cleavage site profiles of our eukaryotic and prokaryotic rhomboids we employed matrix-assisted laser desorption/ionization – time of flight (MALDI-TOF) mass spectrometry. Rhomboid cleavage of our chimeric substrates yields a N-terminal and a C-terminal product (**Fig. 2.7a**). The C-terminal product contains most of the TMD sequence and thus requires detergent for solubilization. In general, mass spectrometry approaches for characterization of detergent solubilized membrane proteins involve extraction from organic solvents to strip out bound detergent and enable spectra of unfolded proteins to be recorded (40). Since complete detergent removal is seldom achieved, accurate membrane protein mass determination by mass spectrometry can be challenging. Therefore, we focused on the N-terminal product, which is essentially MBP followed by a hydrophilic C-terminal linker segment, which is water-soluble and thus more suitable for mass spectrometry analysis than the detergent bound C-terminal product. Figure 8B-F shows the cleavage site determination procedure for GlpG and RHBDL2 in their reaction with the substrate TS. Following incubation, using a concentration of substrate and enzyme that maximizes the yield of the reaction, an efficient degradation of the substrate can be achieved. To this reaction mixture Ni-NTA resin was added to bind away uncleaved substrate (**Fig. 2.7b**), C-terminal product and enzyme (all containing His_6-tags), yielding a purified N-terminal product. TCA precipitation of this product followed by MALDI-TOF mass spectrometry analysis (**Fig. 2.7.c**) yielded a mass spectrum showing two main peaks at *m/z* values, corresponding to

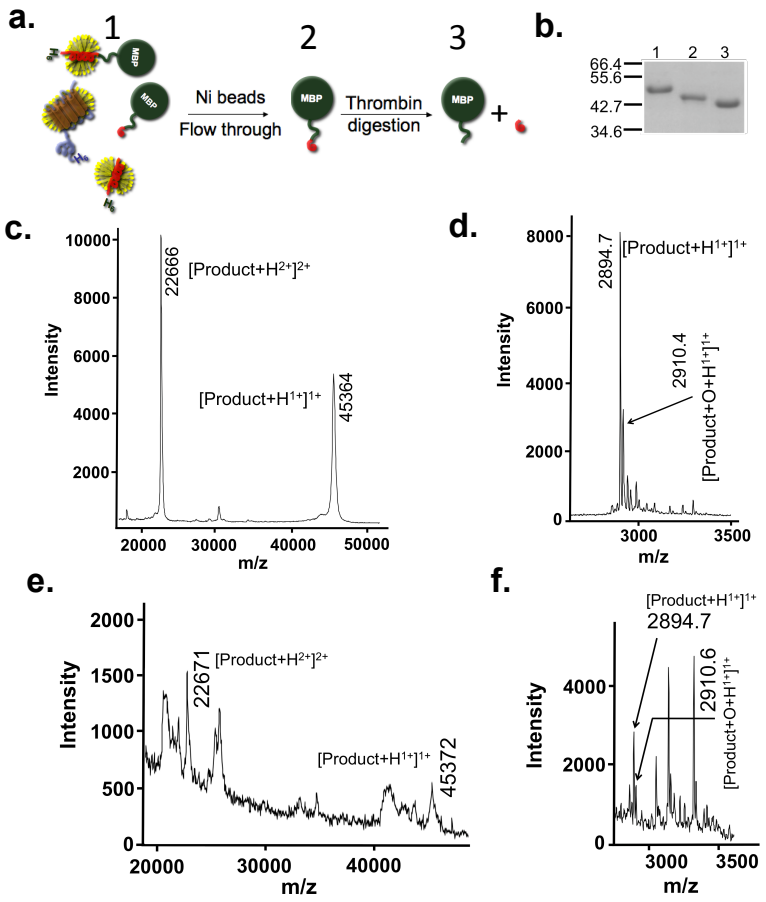

Figure 2.7. Cleavage site determination by MALDI-TOF mass spectrometry. **a).** Schematic showing the different steps in our mass spectrometry analysis: **1)** Following incubation of substrate with enzyme the sample contains uncleaved substrate, C-terminal product and enzyme (all containing His_6-tags), in addition to untagged N-terminal product. **2)** Addition of Ni-NTA resin to this reaction mixture yields a purified N-terminal product, which after TCA precipitation, can be analyzed by MALDI-TOF mass spectrometry. The large mass of this product (~45 kDa) can prevent an accurate determination of the cleavage site. 3) To increase the accuracy of the mass measurements a fraction of this product was further incubated with thrombin to yield two species: one, corresponding to MBP (in green) with a linker region cleaved at the engineered thrombin site (**LVPRGS**), and the other, corresponding to a short water-soluble peptide (in red) flanked by the thrombin cleavage site and the rhomboid cleavage site. The entire reaction mixture can be analyzed by MALDI-TOF mass spectrometry and the mass of the short peptide accurately determined to yield an unambiguous cleavage site determination. The detergent solubilized components are shown surrounded by a detergent micelle (in yellow). **b).** A 12% SDS-PAGE

analysis of the digestion of the substrate Spitz-TMD by GlpG. Lane 1, Spitz-TMD before digestion; lane 2, N-terminal product of the digestion with GlpG after purification through Ni NTA beads and TCA precipitation; lane 3, purified N-terminal product (as in lane 2) following incubation with thrombin for 1 hour at 4°C. The molecular weight marker positions are shown on the left. **c).** Mass spectrum recorded in positive ionization mode of the N-terminal product of the digestion of Spitz-TMD by GlpG (this sample corresponds to that shown in lane 2 of **b**. The observed *m/z* for selected peaks is given. **d).** Mass spectrum recorded in positive ionization mode of the peptide between the thrombin cleavage site and the GlpG cleavage site. The observed *m/z* for the oxidized (at methionine) peptide is also given. **e).** As in **c**, but for the reaction of Spitz-TMD with RHBDL2. **f).** As in **d).**, but for the reaction of Spitz-TMD with RHBDL2.

single and double protonated species with a mass of 45364 ± 70 and 45332 (2 x 22666) ± 70 Da, respectively. Comparison with the calculated masses of possible N-terminal reaction products suggested that GlpG cleaved Spitz-TMD after amino acid residue Ala138 (**Table 2.1**). The relative high molecular weight of the N-terminal MBP-containing product can produce a mass determination error up to ± 70 Da that could lead to inaccuracies in the identification of the exact cleavage site. To solve this issue, the purified N-terminal MBP-containing product was further treated with thrombin. Thrombin digestion of this product yields two species: one corresponding to the MBP plus a linker region cleaved at the engineered thrombin site (LVPRGS) in between the amino acid residues Arg and Gly and the other corresponding to a short water-soluble peptide flanked by the thrombin cleavage site and the rhomboid cleavage site. Mass spectrometry analysis of this peptide (**Fig. 2.7d**) produced a main peak with a mass of 2,894.7 ± 2 Da. The accuracy of the mass measurement for such a small peptide, in combination with the results obtained in **figure 2.7c**, confirmed the single GlpG cleavage site after Ala138 within TS. Due to the significantly lower N-terminal product yields obtained with RHBDL2 the intensity of the obtained mass spectra were considerably lower (**Fig. 2.7e-f**), but accurate enough to map the proteolytic site in TS to the same Ala138. This cleavage site analysis was carried out for all four rhomboids in their reaction with TG, TK, TS, G, K and S. As summarized in Table 1, a single cleavage site after Ala138 for the Spitz chimeras, after Ala122 for the Keren chimeras, and after Ala245 for the Gurken chimeras, which was identical for *D. melanogaster*, *H. sapiens*, *P. aeruginosa* and *E. coli* rhomboids was observed. These results are in excellent agreement with the proteolytic profiles for the prokaryotic rhomboids AarA, GlpG and YqgP published very recently by Strisovsky et al. (41).

Table 2.1. Observed and calculated masses of rhomboid cleavage products.

Substrate	*Calculated mass of products (Da)*		*Observed mass of products (Da)*			
			Enzyme			
	Product sequence	Calc.mass	GlpG	PA3086	Rho-1	RHBDL-2
Spitz-TMD	**MBP**-PRGS--EKA_{138}	45385.3	45364[a]	45309.0	45337.0	45372.0
	GS--EKA_{138}	2895.3	2894.7[b]	2892.8	2894.2	2894.7
Spitz-TMD/CT	**MBP**-PRGS--EKA_{138}	45385.3	45377.0	45284.0	ND	ND
	GS--RMA_{138}	2370.6	2365.8	2369.3	2367.5	2371.9
Keren-TMD	**MBP**-PRGS--EKA_{122}	45306.2	45230.0	45253.0	45252.0	45289.0
	GS--EKA_{122}	2816.2	2812.6	2815.2	2812.7	2815.2
Keren-TMD/CT	**MBP**-PRGS--EKA_{122}	45306.2	45284	45307	ND	ND
	GS--EKA_{122}	2816.2	2813.2	2817.4	2811.4	2812.3
Gurken-TMD	**MBP**-PRGS--RMA_{245}	44860.6	44860	44854	44822	44829
	GS--RMA_{245}	2370.6	2371.1	2370.5	2370.1	2369.4
Gurken-TMD/CT	**MBP**-PRGS--RMA_{245}	44860.6	44845	44880	ND	ND
	GS--RMA_{245}	2370.6	2365.8	2369.3	2367.5	2371.9

(a) Observed mass of the N-terminal product of the reaction between rhomboids and the substrate chimeras. The error in this measurement is ±70Da.

(b) Observed mass of peptide product of MBP adducted product by thrombin. The error in this measurement is ±2 Da.

Cleavage site specificity in rhomboids.

Sequence alignment (**Fig. 2.8a**) between non-redundant Gurken, Keren and Spitz orthologues showed the amino acid residue Alanine to be absolutely conserved at the P1 site among all analyzed sequences. In addition the P1' site, corresponding to either Serine or Histidine, and the P2' sites, corresponding to Isoleucine were also highly conserved. To test the influence of these conserved amino acid positions on scissile bond selectivity we carried out Alanine scanning mutagenesis (except for Ala245 which was mutated to Phenylalanine) of a region of Spitz-TMD flanking the scissile bond and incubated each substrate variant with GlpG (**Fig. 2.8b**). Compared to the wild type substrate sequence substitution of any of the amino acid residues Leu-Glu-Lys (positions P4-P2) by alanine markedly increased substrate degradation.

Discussion

The ability to cleave peptide bonds within the plane of the membrane is the defining property of i-CLiPs. The proteolytic profiles of representative members of this family, including gamma-secretase, signal peptide peptidase and S2P have been determined (reviewed in (42)). In contrast, the proteolytic profiles of key eukaryotic and prokaryotic rhomboids, including rhomboid-1 and GlpG, have not been fully established. Here we have employed in vitro activity assays, coupled to MALDI-TOF mass spectrometry measurements, to determine the proteolytic profiles of *D. melanogaster* rhomboid-1, *H. sapiens* RHBDL2*, E. coli* GlpG and *P. aeruginosa* PA3086 against chimeric substrates derived from the physiological rho-1 substrates Gurken, Keren and Spitz. These wild-type substrates have been employed previously to assay the activity of rhomboids in cultured cells (10); however, poor expression levels precluded their use for

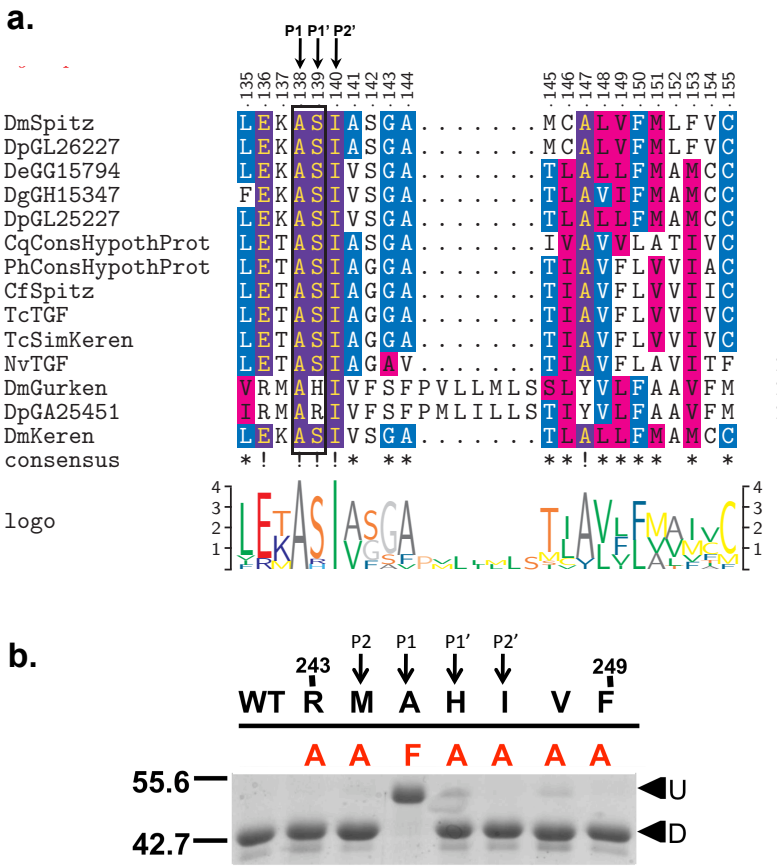

Figure 2.8. Effect of Gurken-TMD mutations on GlpG activity. a). A multiple sequence alignment (T-Coffee) of Gurken, Keren and Spitz orthologs sequences around their TMD region showing the conservation of amino acid residues at P1, P1' and P2' sites (using the Schechter and Berger nomenclature (58). Redundant sequences were eliminated and the color intensity increases as a function of identity. **b).** A 12% SDS-PAGE analysis of the digestion of single Gurken-TMD Alanine variants by GlpG. Each position around the cleavage site was mutated into Alanine (or to Phenylalanine in the case of Ala245). The wild-type sequence is shown in black with mutations in red. Protein bands corresponding to the undigested (**U**) and digested (**D**) substrates are indicated with an arrow on the right and the molecular weight marker positions are shown on the left.

in vitro assays. As an alternative we replaced their EGFR ligand ectodomain by MBP, while maintaining their N-terminal juxtamembrane, transmembrane and the C-terminal cytosolic regions intact. Our in vitro activity assays using chimeric substrates and enzymes purified in the detergent DDM recapitulate key biochemical data obtained in COS cells expressing wild-type substrates (10, 16, 19). Moreover, mutation of the catalytic residues Ser201 and His254 in GlpG, and Ser207 and His257 in PA3086 abolished substrate degradation, indicating the specificity of our assay. Mutation of Asn154 into Ala reduced GlpG activity against TG and G in half, supporting initial mutagenesis experiments in cultured cells (10) and in vitro (39, 43), which suggest Asparagine is not an obligate member of the catalytic triad. Notably under comparable conditions, this mutation abolished TK, TS, K and S proteolysis. This notion is supported by the fact that mutation of Asn157 in PA3086 did not abolish activity against any of the substrates. Indeed, a similar case was reported for *E. coli* Outer Membrane Phospholipase A (OMPLA), a serine hydrolases containing a catalytic dyad Ser-His, where Asn plays an indirect role in catalysis (44). Site directed mutagenesis of these residues show that Ser and His are essential for the catalysis and that Asn contributes to catalysis but is not essential for enzymatic activity. Structural studies have confirmed that Asn154 is beyond hydrogen bond contact to His254 and thus, a possible role for this amino acid residue in oxyanion hole stabilization has been suggested (25, 26, 45, 46). The results raise the possibility that the influence of Asn in proteolytic activity might be substrate and enzyme dependent.

Mass spectrometry analysis of the amino terminal product of the reaction between rhomboids and our chimeric substrates identified a single cleavage site after Ala138 in the case of the Spitz chimeras, after Ala122 for the Keren chimeras, and after Ala245 for the Gurken chimeras. Compared to previous efforts to map the proteolytic

profiles of rhomboids our results showed significant differences (**Fig. 2.9**). Regarding Spitz, Urban et al. (16) had initially employed activity assays in cultured cells, followed by gel mobility analysis, to map the proteolytic profile of rhomboid-1 to a short region (ASIASGA) in the amino terminus of the predicted TMD of Spitz. More recently Baker et al. introduced this minimum motif into C100Flag, a popular recombinant γ-secretase substrate (47), and employed mass spectrometry to conclude that DDM solubilized GlpG cleaved C100-Spitz at two adjacent sites, after Ala141 and Gly143, within the motif (27). We found that indeed all rhomboids tested, including GlpG, cleaved our Spitz chimeras within the same motif, but at a single site three amino acid residues before. Consistent with our findings, the chimeras derived from the Spitz homologue Keren were cleaved at an equivalent site. As the minimum substrate employed by Baker et al. (27) lacks the sequence context present in our chimeric substrates, the outlined differences are likely the result of the two markedly divergent substrate designs. In the case of Gurken Lemberg et al. first employed SDS-PAGE analysis to compare reference peptides with Gurken products generated by TX-100 solubilized *B. subtilisis* YqgP, to roughly map the cleavage to a region around the His246-Ile247-Val248 sequence (39). Later, Maegawa et al. employed cysteine accessibility assays in *E. coli* cells over expressing GlpG and a substrate chimera Bla-GknTMD-MBP (composed of the periplasmic Bla (β-lactamase) domain and the TMD of Gurken followed by MBP) to interpret a single cleavage site after His246 (28). In contrast, we found the cleavage to occur one amino acid residue before, i.e. after Ala245. At this point we cannot rule out the possibility that the observed differences with Gurken arise from the fact that our experiments were carried out in detergent, or due to the different cleavage site determination procedures employed.

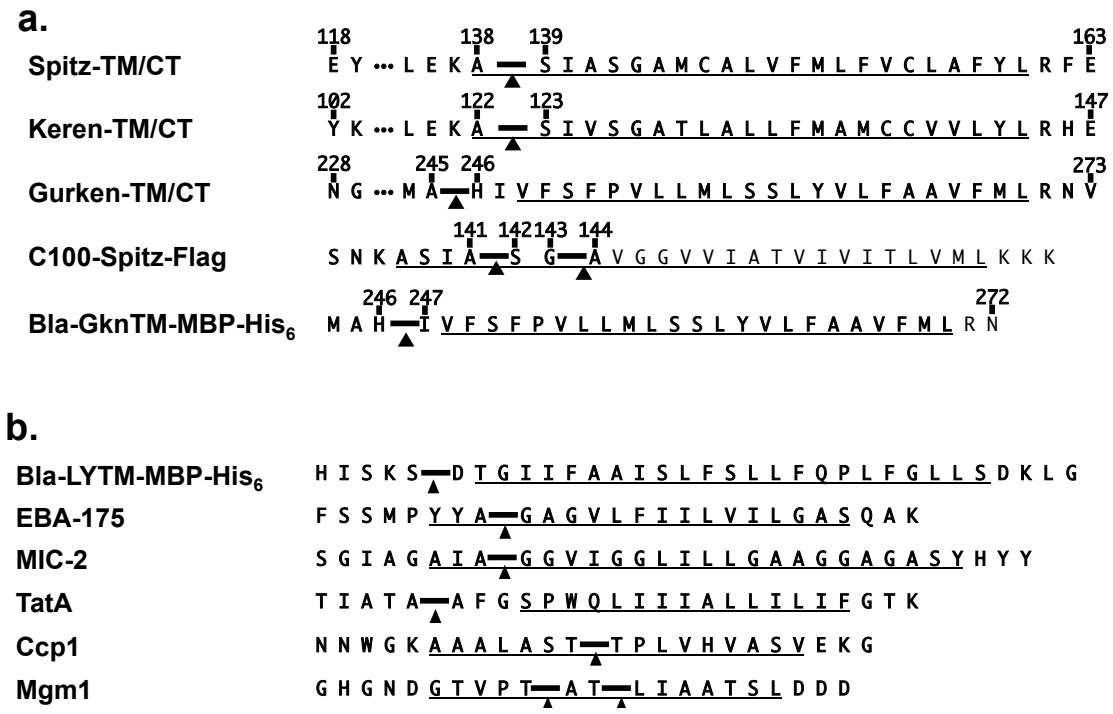

Figure 2.9. Comparison of our proteolytic profiles with previously published results. The sequence regions corresponding to the physiological substrates Gurken, Keren and Spitz are shown in bold. The TMD regions of the substrates are underlined. A black arrow marks the cleavage sites. **a).** Cleavage sites on substrates derived from Gurken, Keren and Spitz. The results obtained for eukaryotic and prokaryotic rhomboids against our chimeras are shown on top denoted by Spitz-TMD/CT, Keren-TMD/CT and Gurken-TMD/CT. The cleavage sites in C100-Spitz (27) and Bla-GknTMD-MBP (28) were obtained with GlpG. **b).** Cleavage sites produced by prokaryotic rhomboids against other substrates: Bla-LY2TMD-MBP (28), EBA-175 (48), MIC-2 (46) and TatA (22).

Experiments showing that *P. stuarti* AarA rhomboid could trigger EGFR signaling when expressed in flies and that rhomboid-1 could rescue the bacterial phenotype caused by AarA deletion (21), suggested a conservation of substrate specificity between rhomboids across species. Biochemical data obtained with in cell activity assays further reinforced this possibility (10, 48). Our data showing that all prokaryotic and eukaryotic rhomboids tested displayed identical proteolytic profiles against each of the substrate chimeras, with a single cleavage site after Alanine at the P1 position, extends the conservation to their peptide bond specificity(22, 49-52). This conservation across species is remarkable, specially when taking into account the low (~6%) sequence identity among prokaryotic and eukaryotic rhomboids (10), notwithstanding the fact that they all share a common architecture with multiple TMDs and an absolute conservation of catalytic Ser and His amino acid residues. A preference for Alanine at the P1 site has been hinted before from studies suggesting a preference for a small side chain amino acid residue at this position (**Fig. 2.9**). Indeed, substitution of Ala245 into Phenylalanine abolished GlpG activity against Gurken-TMD, whereas alanine scanning of flanking residues, including His247, did not affect cleavage. Such a drastic effect elicited by a single point mutation suggests a marked specificity at the P1 site, in contrast with more promiscuous I-CLiPs such as gamma-secretase (reviewed in (42)) known to cleave at multiple sites without a clear specificity. One possible explanation for the conservation of peptide bond specificity, in the context of low sequence identity among the proteases, is that it is governed at the level of the substrate, in agreement with previous studies by Urban et al. (53) suggesting that specificity is governed by residues in the substrate TMD. Understanding the sequence determinants governing rhomboid substrate and peptide bond specificity will require more extensive and systematic mutagenesis

experiments. The activity assay and cleavage site determination procedure that we have developed is likely to be a powerful tool for such studies.

We are aware of the fact that our assays were carried out in detergent micelles and thus comparison to the situation in biological membranes should be made with precaution. With this caveat in mind, but considering that our substrate chimeras contain significant sequence identity with the physiological substrates, we consider a likely possibility that the proteolytic profiles we observed for rhomboid-1 reflect those occurring in the Golgi membrane against the EGFR ligands Gurken, Keren and Spitz. We also note that the cleavage site in all the substrates locates to the N-terminal end of their predicted TMDs. We have previously measured the phosphate-to-phosphate distance of Golgi membranes from rat hepatocytes to be ~39.5 Å (54). Taking into account a translation per amino acid residue in helical conformation of 1.5 Å, the cleavage sites in the physiological substrates Gurken, Keren and Spitz would locate to a region away from the hydrophobic core of the membrane, in proximity to the interface between the membrane and the surrounding water. Indeed, the observation that for both the Keren and Spitz chimeras the P2 and P3 sites correspond to the charged amino acid residues (Lys and Glu, respectively) supports the idea that the cleavage site in vivo locates to a juxtamembrane position, in agreement with studies by Maegawa et al. that first raised the possibility that rhomboids can cleave juxtamembrane regions of their substrates (28). The crystal structure of *E. coli* GlpG was refined to 1.7 Å (55) and led to the estimation that the catalytic Ser201 is positioned only 3 Å below the anticipated surface of the bilayer at an optimal position for cleavage of a juxtamembrane bond. We hypothesize that for the substrates Gurken, Keren and Spitz the juxtamembrane cleavage site might already be solvent exposed before it transfers to the active site for

hydrolysis. In such a case, transfer of this region of the substrate to the aqueous environment of the active site could pose a low energetic penalty.

Bibliography

1. Wolfe MS (2009) *J Biol Chem* **284,** 13969-13973.

2. Pascall JC & Brown KD (1998) *FEBS Lett* **429,** 337-340.

3. Gottlieb E (2006) *Cell* **126,** 27-29.

4. O'Donnell RA & Blackman MJ (2005) *Curr Opin Microbiol* **8,** 422-427.

5. Kanaoka MM, Urban S, Freeman M, & Okada K (2005) *FEBS Lett* **579,** 5723-5728.

6. Dowse TJ, Pascall JC, Brown KD, & Soldati D (2005) *Int J Parasitol* **35,** 747-756.

7. Dowse TJ & Soldati D (2005) *Trends Parasitol* **21,** 254-258.

8. Howell SA, Hackett F, Jongco AM, Withers-Martinez C, Kim K, Carruthers VB, & Blackman MJ (2005) *Mol Microbiol* **57,** 1342-1356.

9. Koonin EV, Makarova KS, Rogozin IB, Davidovic L, Letellier MC, & Pellegrini L (2003) *Genome Biol* **4,** R19.

10. Urban S, Schlieper D, & Freeman M (2002) *Curr Biol* **12,** 1507-1512.

11. Urban S, Lee JR, & Freeman M (2002) *EMBO J* **21,** 4277-4286.

12. Lee JR, Urban S, Garvey CF, & Freeman M (2001) *Cell* **107,** 161-171.

13. Pascall JC, Luck JE, & Brown KD (2002) *Biochem J* **363,** 347-352.

14. Tsruya R, Schlesinger A, Reich A, Gabay L, Sapir A, & Shilo BZ (2002) *Genes Dev* **16,** 222-234.

15. Freeman DM (2004) *Nature Reviews* **5,** 188-197.

16. Urban S, Lee JR, & Freeman M (2001) *Cell* **107,** 173-182.

17. McQuibban GA, Saurya S, & Freeman M (2003) *Nature* **423,** 537-541.

18. McQuibban GA, Lee JR, Zheng L, Juusola M, & Freeman M (2006) *Curr Biol* **16,** 982-989.

19. Lohi O, Urban S, & Freeman M (2004) *Curr Biol* **14,** 236-241.

20. Rather PN, Ding X, Baca-DeLancey RR, & Siddiqui S (1999) *J Bacteriol* **181,** 7185-7191.

21. Gallio M, Sturgill G, Rather P, & Kylsten P (2002) *Proc Natl Acad Sci U S A* **99,** 12208-12213.

22. Stevenson LG, Strisovsky K, Clemmer KM, Bhatt S, Freeman M, & Rather PN (2007) *Proc Natl Acad Sci U S A* **104,** 1003-1008.

23. Wang Y, Zhang Y, & Ha Y (2006) *Nature* **444,** 179-180.

24. Wu Z, Yan N, Feng L, Oberstein A, Yan H, Baker RP, Gu L, Jeffrey PD, Urban S, & Shi Y (2006) *Nat Struct Mol Biol* **13,** 1084-1091.

25. Ben-Shem A, Fass D, & Bibi E (2007) *Proc Natl Acad Sci U S A* **104,** 462-466.

26. Lemieux MJ, Fischer SJ, Cherney MM, Bateman KS, & James MN (2007) *Proc Natl Acad Sci U S A* **104,** 750-754.

27. Baker RP, Young K, Feng L, Shi Y, & Urban S (2007) *Proc Natl Acad Sci U S A* **104,** 8257-8262.

28. Maegawa S, Koide K, Ito K, & Akiyama Y (2007) *Mol Microbiol* **64,** 435-447.

29. Alexandrov A, Dutta K, & Pascal SM (2001) *Biotechniques* **30,** 1194-1198.

30. Cadene M & Chait BT (2000) *Anal Chem* **72,** 5655-5658.

31. Fenyo D, Wang Q, DeGrasse JA, Padovan JC, Cadene M, & Chait BT (2007) *J Vis Exp***,** 192.

32. Wasserman JD, Urban S, & Freeman M (2000) *Genes Dev* **14,** 1651-1663.

33. Wasserman JD & Freeman M (1998) *Cell* **95,** 355-364.

34. Freeman M (1994) *Mech Dev* **48,** 25-33.

35. Freeman M, Kimmel BE, & Rubin GM (1992) *Development* **116,** 335-346.

36. Urban S & Wolfe MS (2005) *Proc Natl Acad Sci U S A* **102,** 1883-1888.

37. Del Rio A, Dutta K, Chavez J, Ubarretxena-Belandia I, & Ghose R (2007) *J Mol Biol* **365,** 109-122.

38. Pascall JC & Brown KD (2004) *Biochem Biophys Res Commun* **317,** 244-252.

39. Lemberg MK, Menendez J, Misik A, Garcia M, Koth CM, & Freeman M (2005) *Embo J* **24,** 464-472.

40. le Coutre J, Whitelegge JP, Gross A, Turk E, Wright EM, Kaback HR, & Faull KF (2000) *Biochemistry* **39,** 4237-4242.

41. Strisovsky K, Sharpe HJ, & Freeman M (2009) *Mol Cell* **36,** 1048-1059.

42. Beel AJ & Sanders CR (2008) *Cell Mol Life Sci* **65,** 1311-1334.

43. Lei X, Ahn K, Zhu L, Ubarretxena-Belandia I, & Li Y (2008) *Biochemistry*.

44. Snijder HJ, Ubarretxena-Belandia I, Blaauw M, Kalk KH, Verheij HM, Egmond MR, Dekker N, & Dijkstra BW (1999) *Nature* **401,** 717-721.

45. Wang Y & Ha Y (2007) *Proc Natl Acad Sci U S A* **104,** 2098-2102.

46. Meissner M, Reiss M, Viebig N, Carruthers VB, Toursel C, Tomavo S, Ajioka JW, & Soldati D (2002) *J Cell Sci* **115,** 563-574.

47. Li YM, Lai MT, Xu M, Huang Q, DiMuzio-Mower J, Sardana MK, Shi XP, Yin KC, Shafer JA, & Gardell SJ (2000) *Proc Natl Acad Sci U S A* **97,** 6138-6143.

48. Urban S & Freeman M (2003) *Mol Cell* **11,** 1425-1434.

49. O'Donnell RA, Hackett F, Howell SA, Treeck M, Struck N, Krnajski Z, Withers-Martinez C, Gilberger TW, & Blackman MJ (2006) *J Cell Biol* **174,** 1023-1033.

50. Maegawa S, Ito K, & Akiyama Y (2005) *Biochemistry* **44,** 13543-13552.

51. Opitz C, Di Cristina M, Reiss M, Ruppert T, Crisanti A, & Soldati D (2002) *EMBO J* **21,** 1577-1585.

52. Tatsuta T, Augustin S, Nolden M, Friedrichs B, & Langer T (2007) *EMBO J* **26,** 325-335.

53. Urban S (2003) *Molecular Cell* **11,** 1425-1434.

54. Mitra K, Ubarretxena-Belandia I, Taguchi T, Warren G, & Engelman DM (2004) *Proc Natl Acad Sci U S A* **101,** 4083-4088.

55. Wang Y, Maegawa S, Akiyama Y, & Ha Y (2007) *J Mol Biol* **374,** 1104-1113.

Chapter 3: Peptide bond selectivity in Rhomboid

Introduction

Early evidence of substrate specificity in rhomboids was provided by experiments showing that Rho-1 activity, in COS cells, is dependent on the N-terminal sequence, ASIASGA, of Spitz (1). Mutagenesis studies of this sequence suggested an enhancement of Rho-1 activity when canonical helix-destabilizing residues were introduced one at the time (2). The opposite effect was observed when phenylalanine, considered to be a helix-stabilizing residue, was introduced (2). These early experiments raised the hypothesis that the loss of helical conformation in the N-terminal region of the TMD of Spitz, where the motif ASIASGA is located, facilitated rhomboid activity by exposing the scissile peptide bond.

The recent development of in vitro rhomboid activity assays (3-5) has shed new light into rhomboid specificity. An artificial substrate (C100Spitz-Flag), based on the γ-secretase substrate C100-Flag, was generated by introducing the motif ASIASGA into the N-terminal region of the TMD of C100-Flag (5). Using this substrate in combination with mass spectrometry, *E. coli* rhomboid GlpG was shown to cleave within the ASIASGA sequence at two sites in between Ala^{141}-Ser^{142} and Gly^{143}-Ala^{144} (6). At the same time Maegawa et al (4), using a substrate containing the second TMD of lactose permease (LacY) flanked by beta-lactamase (BLA) and maltose binding protein (MBP), showed that GlpG cleaved in between the serine and aspartic acid residues outside the membrane, just before the N-terminus of the predicted TMD. Further evidence to support the idea that rhomboids can cleave at the membrane/water interface was provided by the same group, showing that a similar chimeric substrate containing the TMD of Gurken was cleaved in between His^{246}-Ile^{247}. Ha et al. (7) reviewed the literature

regarding the proteolytic profiles of rhomboids orthologs from different species to propose that rhomboids have a preference to cleave after an amino acid with a small side chain. Some evidence for this hypothesis was provided by Erez et al (8) showing that GlpG can cleave after alanine and glycine even in the soluble region of a polytopic membrane protein

In Chapter 2 we employed substrate chimeras in which the transmembrane and cytoplasmic regions of Gurken, Keren and Spitz were preserved, to assay the *in vitro* activity of rhomboid-1, *H. sapiens* RHBDL2, *P. aeruginosa* PA3086 and *E. coli* GlpG. Mass spectrometry analysis of the N-terminal reaction product identified a single cleavage site after Ala138 for the Spitz chimeras, after Ala122 for the Keren chimeras, and after Ala245 for the Gurken chimeras that was identical for all these prokaryotic and eukaryotic rhomboids. Our proteolytic mapping was later confirmed by an independent study published by Strisovsky et al. (9) and showed significant differences with previous studies (4, 6). Our initial mutagenesis experiments showed a preference for Alanine at the P1 site and prompted us to extend our studies with the goal of addressing the following questions: How do rhomboids achieve peptide bond specificity? What is the role of substrate helicity in cleavage?

Material and methods

Design of Substrates and site directed mutagenesis

The substrates used in this study were derived from the natural substrates of Rho-1: Gurken and Spitz, both of them from *Drosophila melanogaster*. These substrates were designed in order to maintain the basic structure and molecular characteristics of Spitz and Gurken in the region where occurs the molecular events that lead to the selection of the cleavage site and proteolysis of these substrates by

rhomboids. These artificial substrates have a maltose binding protein (MBP) domain (excluding the signal peptide) substituting the N-terminal extracellular domain of the original substrates, and a his_6tag in the C-terminus. After the MBP domain there is the sequence encoding 20 amino acid residues upstream the N-terminal end of the transmembrane domain of each substrate. A thrombin cleavage site (LVPRGS) was introduced at the end of the MBP domain to facilitate the determination of the cleavage site assay by MALDI-TOF. The substrate with the sequences of Gurken and Spitz were called as **TG** and **TS** respectively. The cytoplasmic domain of the original substrates was not used in this study because we determined previously that the absence of this domain does not affect the enzymatic activity or the cleavage site selectivity. MBP domain sequence was obtained by PCR from plasmid H-MBP-3C (10) and sequences from Spitz and Gurken were obtained from Drosophila Genomics Resource Center.

Expression of substrates and rhomboid proteins

Rhomboid substrates TG, TS and their different mutants were expressed according to the protocol described in the material methods section of the chapter 2. In the same way, rhomboid proteins GlpG, PA3086, Rho-1 and RHBDL2 were expressed in the same way as described in the chapter 2.

Mutagenesis studies.

Alanine screening mutagenesis was performed in the transmembrane domain of the residues located in the transmembrane domain and in the surroundings of the N-terminal of the transmembrane domain of each substrate. These mutants were obtained by site directed mutagenesis (QuikChange site-directed mutagenesis kit, Stratagene). Dead mutants: GlpG(S201A), PA3086(S207A), Rho-1(S217A) and RHBDL2(S176A) were expressed as indicated in the same section of chapter 2.

Rhomboid activity assay

The activity of prokaryotic rhomboids were determined by densitometric analysis of the bands obtained after digestion of the different substrates and the resolution of the products in 15% SDS-PAGE. The proteolysis reaction was carry out mixing 5 μg of purified rhomboid and 10 μg of purified substrate in 100 μL of RB and 2mM β-mercaptoethanol plus inhibitors cocktail and 0.1% DDM. This reaction mixture was rocked at 37°C for 16 hours. After the digestion the reaction is stopped adding 25 μL of 5x SLB. 15% SDS-PAGE gel was used to verify the reaction and then the gel was analyzed by densitometry. The Digestion Index was determined by the ratio of digested substrate and the total amount of substrate. For the mutants study this digestion index was normalized using the digestion index obtained for the wild type substrate. For eukaryotic rhomboids due to the low level of expression of these rhomboids in *E. coli* cells the activity was determined by western blot using anti-MBP primary antibody. In the case of these rhomboid proteins 0.5 mL of the reaction mixtures were loaded in 15% SDS-PAGE gels and the western blot was analyzed by densitometry using analysis of images. The level of digestion was only verified qualitatively due to the low level of digestion.

Cleavage site determination assay

In order to determine the cleavage site of the different mutants we centered our attention on the purification of the MBP product rendered after the rhomboid proteolysis of the different substrates and mutants. A customary reaction was conducted mixing 200 μg of purified substrate and 200 μg of purified rhomboid in 500 μL of RB buffer plus inhibitor cocktail, 0.1% DDM and 2 mM β-mercaptoethanol. After incubation for 16 hours 0.5 mL of Ni-beads previously equilibrated with the same reaction buffer and rocked by 2

hours. The flow through was obtained by filtration of the reaction mixture using a 2 mL disposable PP column (Thermo scientific). The purified protein was precipitated adding 1 mL of 10% TCA in acetone chilled in dry ice and kept overnight at -20. Precipitated protein was then centrifuged at 13000xg for 10 minutes. After the protein pellet was washed once with 1 mL pre-chilled acetone it was solubilized in 1 mL of 8M urea buffered with 50mM Tris (pH 7.4). This protein solution was then concentrated using amicon ultra-4. The urea was removed from this sample by 3 steps of concentration/dilution steps using the same concentrator. The product of digestion was verified by SDS-Page. This solution was centrifuged at 17000 x g in order to eliminate any insoluble material discarding the pellet and divided the supernatant in two aliquots of 20 μL. Five μL of 5% Trifluoroacetic acid (TFA) in acetone was added to the first aliquot, which contains the MBP-adducted product, and then it was sent for mass spectra analysis (MALDI-TOF). The second 20 μL aliquot was digested with 0.5 units of human alpha thrombin (Enzyme Research Laboratories) for 1 hour at 4°C in order to deliver the peptide between the thrombin cleavage site and the rhomboid cleavage site and the reaction was stopped adding 5 μL of 5% Trifluoroacetic acid in acetone. This sample was sent to MALDI-TOF analysis. Regions between 0.8 to 4 kD and between 20 to 60 kD were examined in both samples.

Results

Rhomboid activity can be modulated by alanine scanning mutagenesis of the substrate TMD.

To further investigate the substrate specificity of rhomboids we carried out an alanine scanning mutagenesis of the TMD region of Spitz and Gurken and tested each substrate variant for their effect on the activity of the rhomboids GlpG and PA3086. For

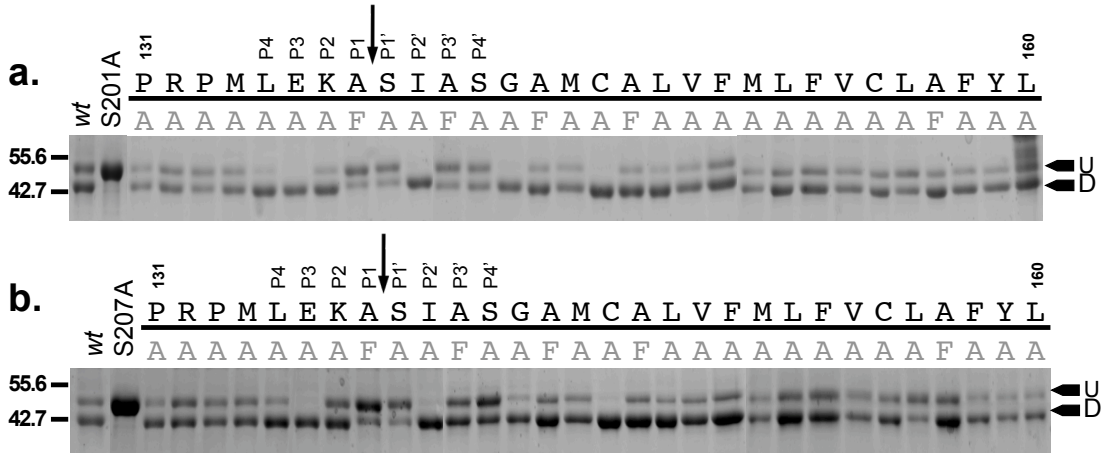

Figure 3.1. Alanine screening of Spitz substrate. Activity assay of GlpG a) and PA3086. b) rhomboid proteins with the different alanine screening mutants of Spitz. Each panel (a and b) show the spitz native sequence (bold) and the mutation for each position (gray). The black arrow indicates the cleavage site position for the native substrate. Activity of the catalytic mutant of GlpG (S201A panel a) and PA3086 (S207A panel b) are shown as a control. The bands that correspond to the undigested (U) and digested (D) substrate are indicated at the right of each panel by a black arrow.

these initial studies the chimera of Spitz (TS), which lacks the cytoplasmic region, was employed. In TS all the amino acid residues, from Pro131 to Leu160 were mutated, one at the time, into alanine (alanine positions in the native sequence were mutated to phenylalanine). We chose alanine because it has been often used to study protease specificity, and since it is commonly found in the hydrophobic region of transmembrane helices (11, 12), we did not anticipate expression or folding problems. Each substrate variant was purified in detergent using metal affinity chromatography and the activity assays were carried out essentially as described in Chapter 2 of this thesis. To facilitatethe interpretation of the activity data (**Fig. 3.1**) we quantified the activity displayed in the SDS-PAGE using densitometry (**Fig. 3.2**). The activity values for each mutant were normalized relative to the wild-type substrate, normalized digestion index (ndi) of 1. In this assay the GlpG catalytically dead mutant S201A and the PA3086 catalytically dead mutant S207A had an ndi of 0.

The substrate amino acid sequence dependence of GlpG activity can be separated into three groups: those amino acid positions where substitution to alanine resulted in a significant reduction of activity (ndi =0.5-0.8; positions: 138, 139, 141, 142, and 156); those positions in which mutation to alanine had no significant effect (ndi =0.9-1.2; positions: 131-134, 145, 150-155, 157-160); and those positions in which the activity increased significantly (ndi =1.3-1.6; positions: 135-137, 140, 143-144, 146-149). The rhomboid PA3086 displayed a very similar activity dependence compared to GlpG. The largest reduction in activity occurred for both GlpG and PA3086 rhomboids when the amino acids A138, S139 (residues at positions P1 and P1') were substituted by to alanine and phenylalanine, respectively. This is perhaps not surprising since the P1 and P1' residues that flank the scissile bond are often the most critical for classical serine

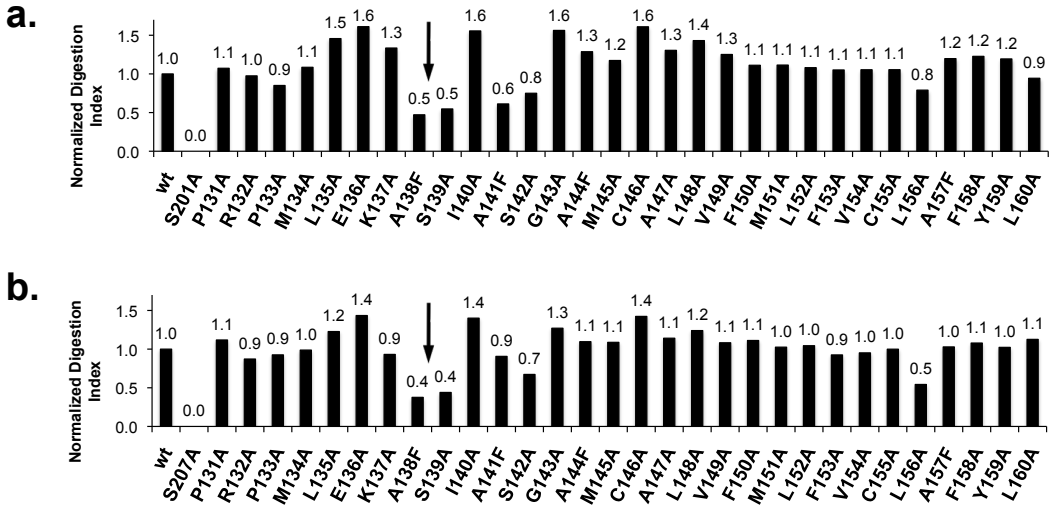

Figure 3.2. Normalized Digestion index determination of GlpG and PA3086 with the alanine screening Spitz substrates mutants. The Normalized Digestion Activity was determined from for the gels in Fig 3.1. Each panel (a and b) show the different values of the NDI for GlpG (a) and PA3086 (b) rhomboid proteins with the different mutants of Spitz. Activity were normalized with the corresponding wild type rhomboid.

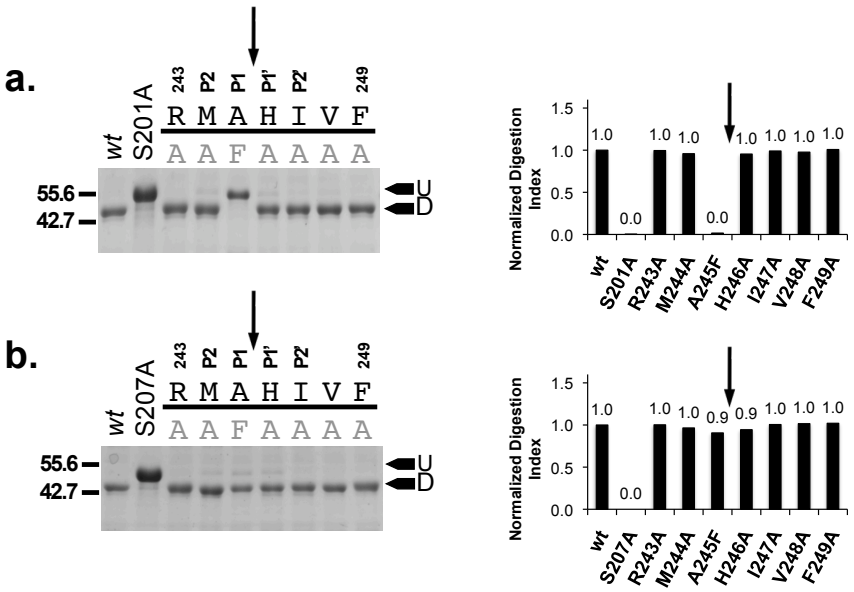

Figure 3.3. Normalized Digestion index determination of GlpG and PA3086 with the alanine screening Gurken substrates mutants. The Normalized digestion activity was determined for GlpG (**a** and **c**) and PA3086 (**a** and **b** right panel) rhomboid proteins with the different mutants of Gurken. The bands that correspond to the undigested (**U**) and digested (**D**) substrate are indicated at the right of each gel by a arrow. Activity of the catalytic mutant of GlpG (S201A panel **a**) and PA3086 (S207A panel **b**) are shown as a control. The left panel of the panels **a** and **b** show the different values of the NDI obtained with the alanine screening mutants of Gurken. In all the figures black arrows indicate the position of the wild type substrate.

protease recognition (13) and GlpG has been previously shown to be sensitive to these two positions using a different model substrate (3). More surprisingly, rhomboid activity was also strongly inhibited when the same mutations were introduced in the same residue tandem (A141 and S142) at the P3' and P4' sites. We note also that activity of both rhomboids was inhibited when alanine was introduced far away from the cleavage site at the P18' position. Finally, rhomboid activity increased markedly when alanines were introduced upstream of the cleavage site at P2, P3, and P4 sites, as well as downstream at P2', P5', and P8' sites.

The observed activity modulation of either GlpG or PA3086A upon alanine substitution at key positions relative to the cleavage site was a clear indication of the substrate specificity by rhomboids. To further support this idea we carried out similar experiments (**Fig. 3.3**) using a shorter region (from Arg243 to Phe249) of the Gurken based chimera, TG. The activity of GlpG rhomboid was abolished when the alanine at P1 site was mutated to phenylalanine, whereas PA3086 rhomboid activity was reduced only by ~10%.

The proteolytic profile of rhomboid can be shifted by mutagenesis.

To better understand the sequence-dependent modulation of rhomboid activity we examined the cleavage site profiles against the TS and TG alanine mutants. As described in chapter 2, our substrate chimeras allow the accurate determination of the cleavage sites by mass spectrometry. To evaluate the mutants we followed a similar procedure as for the WT substrates. **Figure 3.4** depicts the activity and the cleavage profile of both GlpG and PA3086 rhomboids against the TS alanine mutants. The

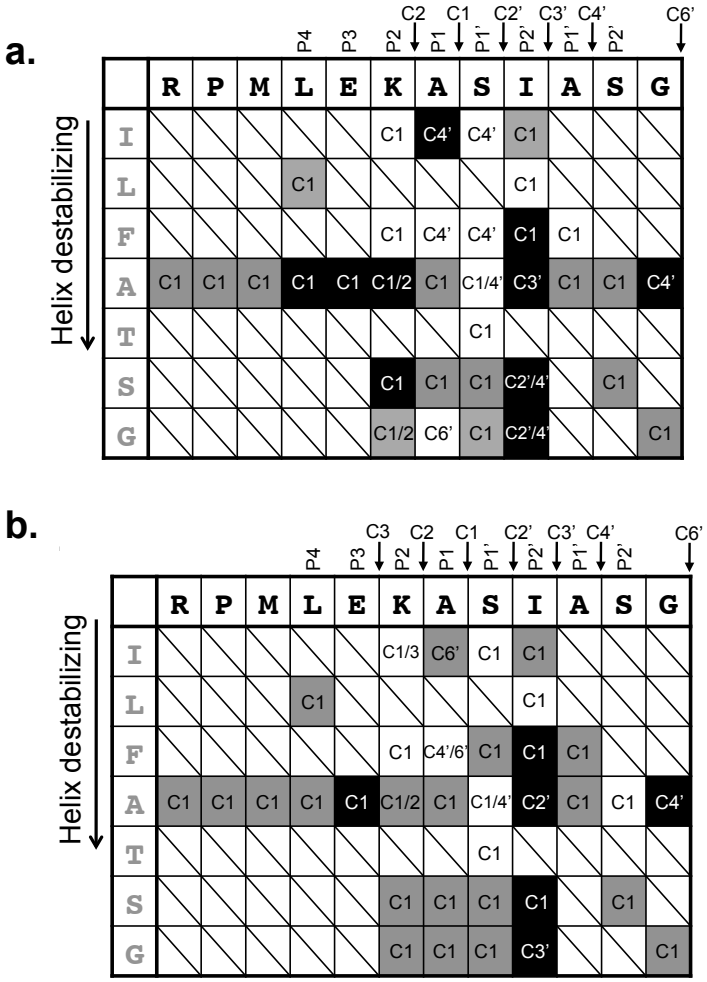

Figure 3.4. Summary of the effect of the mutations in the region of the substrate motif. a) GlpG results of determination of cleavage site. The cleavage site are indicated in as C2, C1, C2', C3', C4', and C6'. Each position indicates a particular mutant. The color code is the next: gray for no effect on activity, white for inhibition effect and black for increasing in the activity. In the same way each position indicates the cleavage site determined by MS analysis of the digestion products. **b)** Analysis of the cleavage site of the different spitz substrate mutants by PA3086. The color and cleavage site are the same as indicated above for GlpG rhomboid.

substitution by alanines of the amino acid residues at positions P7-P3 does not alter the cleavage profile and the cleavage occurs in between Ala138-Ser139 (termed C1). An alanine instead of a lysine at the P2 position, however, introduces a secondary cleavage site (termed C2) in between Ala137-Ala138. Mutation of the alanine at the P1 site into phenylalanine abolishes the cleavage at C1 and creates a secondary identical site at C4' in between Ala140-Ser141. An alanine residue, instead of a serine, at the P1' position creates two sites at C1 and C4'. Substitution by alanine of the isoleucine at position P2' introduces yet another new cleavage site at C3' in between Ala140-Ala141. Introducing alanines at positions P3' and P4' does not alter the cleavage profile, whereas substitution of the glycine at P5' shifted the cleavage site to C4'. Almost identical profiles were observed for GlpG and PA3086 rhomboids against the TS alanine mutants. The picture that emerges is relatively complex. A single amino acid substitution at positions P2-P2' can change the cleavage profile significantly with the creation of a secondary site in between the second Ala-Ser sequence. A preference for cleavage in between two alanines is evident. In contrast for Gurken (**Fig. 3.5**), alanine mutagenesis did not alter the proteolytic profile. Notably even the presence on two consecutive alanines in the M244A mutant did not introduce a secondary site, suggesting significant differences between Spitz and Gurken.

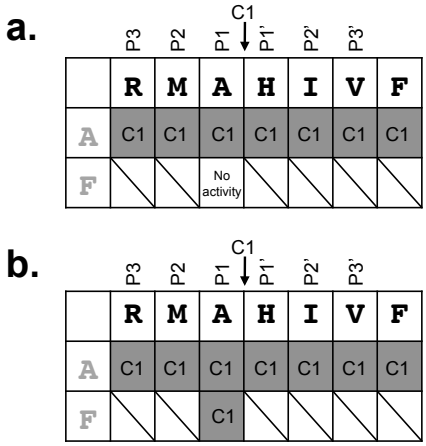

Figure 3.5. Summary of the effect of the alanine screening mutations in the region of TG motif. a). Analysis of the cleavage site of the different gurken substrate mutants by GlpG. The cleavage site are indicated in as C1. Each position indicates a particular mutant. The color code is the next: gray for no effect on activity, white for inhibition effect and black for increasing in the activity. In the same way each position indicates the cleavage site determined by MS analysis of the digestion products. **b**). As in a. but for PA3086. The color and cleavage site are the same as indicated above for GlpG rhomboid.

Rhomboid specificity against Spitz is determined by the P2-P2' positions

We next investigated in more detail the chemistry modulating rhomboid cleavage at positions P2-P2' by measuring the effect in rhomboid activity (**Fig. 3.6**) and in the proteolytic profile (**Fig. 3.5**) of additional amino acid substitutions. As with the alanine scanning mutagenesis of TS the rhomboids GlpG and PA3086 displayed very similar activity dependence. Substitution of the lysine at the P2 position by large aliphatic residues had an inhibitory effect on activity, whereas the introduction of small amino acid side chains had a negligible effect. Remarkably, substitution by glycine introduced the same new C2 site as alanine. As expected substitutions at the P1 position had a large effect on the proteolytic profile, with isoleucine shifting cleavage to the secondary C4' site. Substitution by glycine shifted cleavage all the way to the C6' site, presumably to a region deep within the hydrophobic core of the membrane. At the P1' position, substitution by amino acids with side-chains of with smaller volume had no effect, whereas side chains of similar or larger volume shifted cleavage to the C4' site. All substitutions at the P2' position had a marked effect in increasing activity. Large side-chains, with comparable volumes to isoleucine, had no effect on the profiles, whereas smaller side chains gave rise to secondary cleavage sites at C2', C3' and C4' positions. Interestingly there appeared to be some correlation between cleavage profile and amino acid helix-propensity in some positions and between cleavage profile and size in others. These data suggest that if helix-destabilizing residues are substituted into the P2-P2' positions the cleavage site profile and activity are not affected. On the other hand helix-stabilizing shift the cleavage to secondary cleavage sites: C2', C3', C4' and even C6'.

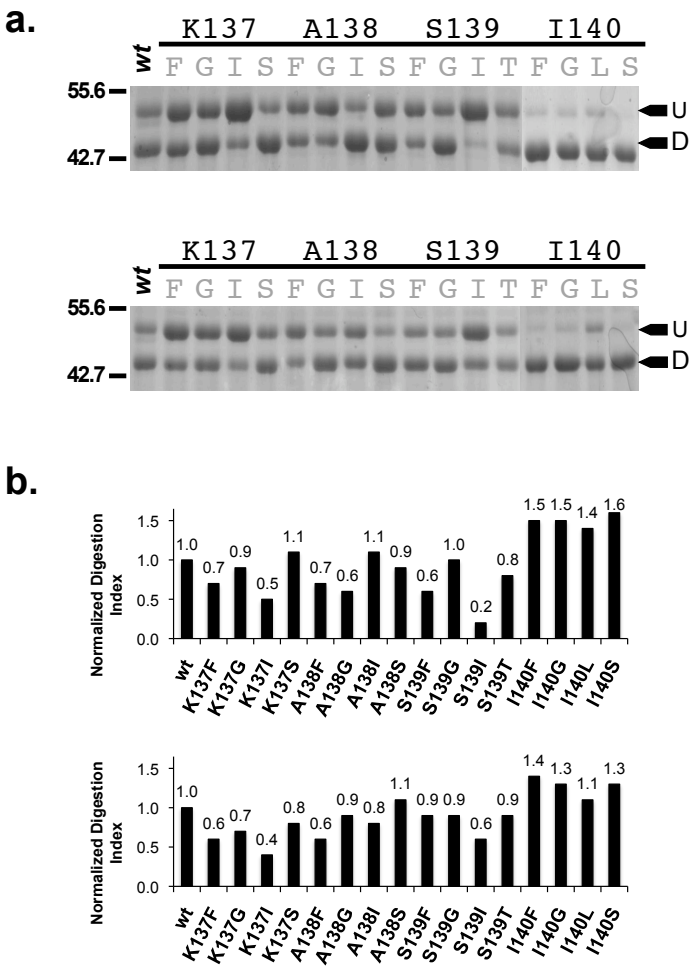

Figure 3.6. Substrate motif analysis by site directed mutagenesis. Activity assay of GlpG and PA3086 a) rhomboid proteins with the different substrate motif mutants of Spitz. Each panel of this figure show the wild type control and the catalytic mutants of the corresponding rhomboid. The black arrow indicates the cleavage site position for the native substrate. The bands that correspond to the undigested (U) and digested (D) substrate are indicated at the right of each panel by a black arrow. **b**). Normalized Digestion Index determination of GlpG (above) and PA3086 (below) with the site directed mutagenesis study of Spitz KASI motif mutants. The Normalized Digestion Activity was determined for GlpG (above) and PA3086 (PA3086) rhomboid proteins with the different mutants of the KASI Spitz motif.

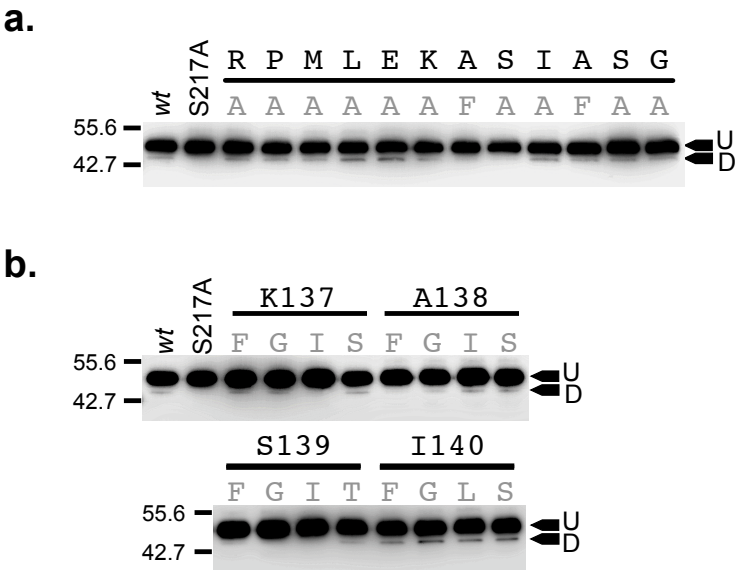

Figure 3.7. Alanine screening mutagenesis and substrate motif analysis mutagenesis studies of human Rho-1 rhomboid. Activity assay of Rho-1 with Spitz alanine screening mutants **(a),** and spitz KASI mutagenesis study mutants (**b)**. Each panel (**a** and **b**) shows the activity of the wild type control and the catalytic Rho-1 mutant S217A. The black arrow indicates the cleavage site position for the native substrate. The bands that correspond to the undigested (**U**) and digested (**D**) substrate are indicated at the right of each panel by a black arrow.

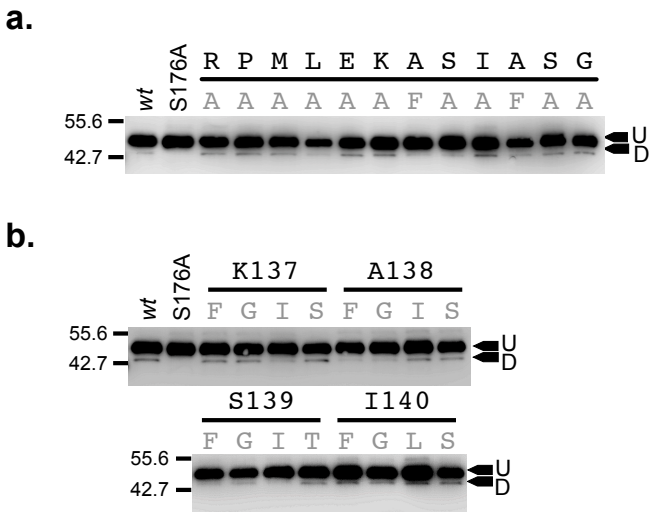

Figure 3.8. Alanine screening mutagenesis and substrate motif analysis mutagenesis studies of human RHBDL2 rhomboid. Activity assay of RHBDL2 with Spitz alanine screening mutants (**a**) and spitz KASI mutagenesis study mutants (**b**). Each panel (**a** and **b**) show the activity of the wild type control and the catalytic Rho-1 mutant S176A. The black arrow indicates the cleavage site position for the native substrate. The bands that correspond to the undigested (**U**) and digested (**D**) substrate are indicated at the right of each panel by a black arrow.

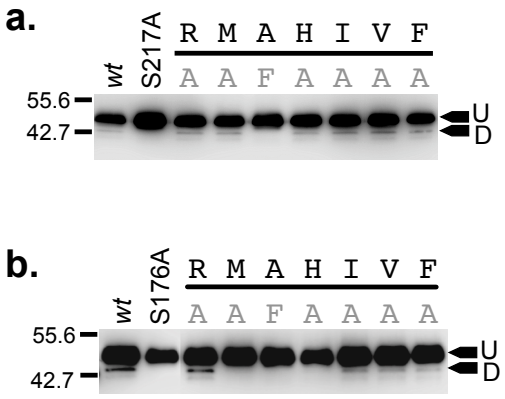

Figure 3.9. Activity assay of the alanine screening mutant of TG substrate mutants with Rho-1 and RHBDL-2 eukaryotic rhomboids. Activity assay of Rho-1 (**a**) and RHBDL-2 (**b**) of alanine screening mutants of TG substrate are shown. Each panel show a wild type control and the corresponding serine mutant of Rho1 and RHBDL-2 rhomboid (Rho-1 (S217A) and RHBDL-2 (S176A) respectively). The black arrow indicates the cleavage site position for the native substrate. The bands that correspond to the undigested (**U**) and digested (**D**) substrate are indicated at the right of each panel by a black arrow.

Eukaryotic rhomboids Rho-1 and RHBDL2 display comparable substrate specificity to their prokaryotic homologs.

Eukaryotic rhomboids Rho-1 and RHBDL2 display considerably lower activity against our substrate chimeras, likely due to the lower purity of these enzymes (Chapter 2). As a result, cleavage site determination is challenging and prone to errors. Thus, here we just focused on measuring the activity of these two enzymes against the TS (**Figs. 3.7** & **3.8**) and TG (**Fig. 3.9**) mutants. Mutation of the amino acid residues at the P1 and P1' positions reduced the activity of both rhomboids, whereas the remaining positions had no effect. Substitution by helix-stabilizing amino acid residues reduced the activity of Rho-1 against TS. RHBDL2 activity is apparently inhibited when the positions P2 and P1' are mutated to alanine and to phenylalanine in the case of P1. In the case of Rho-1 there was no detectable activity when Ala-245 was mutated to phenylalanine. These data suggest a conservation of substrate specificity between eukaryotic and prokaryotic rhomboids.

Discussion.

In this research we study what are the role of the amino acids in the transmembrane domain Spitz using an artificial Spitz chimera that includes the sequence of the complete expected TMD and 20 amino acids beyond the N-terminal end. In our alanine screening mutagenesis experiments we found the residues between P2 to P2' are the most sensitive to mutations. Extended mutagenesis studies of this positions, using conservative and non-conservative mutations, shown that rhomboids have a strong preference for alanine in the position P1, a helix-breaking residue in the position P1' and a bulky hydrophobic residue in the position P2'.

Position P2 is structurally important.

The residue in the position P2: Lys137 is known to play a role in the delimitation of the border of the transmembrane domain because it is positive charged and a strong helix-breaking residue (14-16). When this position is changed to amino acids with a higher hydrophobicity the activity of both GlpG and PA3086 is reduced (**Fig. 3.10a**) with the exception of the mutant K137A. This mutant, K137A) increase the activity . It is possible that the increasing in the activity, when K137A is used, is consequence of two different events of the enzymatic mechanism: The destabilization of the helical structure in the substrate and the introduction of a new cleavage site in the substrate. Strisovsky et al (9) found that the when threonine at position P2 in TatA is mutated to an amino acid with a higher hydrophobicity like phenylalanine, tyrosine or isoleucine (16, 17) the activity of AarA can be reduced to 50% or lower. It is possible to observe that there is a higher activity when P2 position is occupied by alanine. This higher activity with alanine can be explained taking in consideration that rhomboid has a high preference for alanine in the position P1 and the introduction of another alanine residue can increase the affinity of the rhomboid for Spitz.

Rhomboid has a preference for alanine at position P1

The role of the position P1 it is hard to be study in the Spitz and TatA (P. stuartii rhomboid substrate) because these substrates contains multiple alanine residues close to the position P1-P1'. Less difficult to study is Gurken derived chimera substrate because the only alanine residue present in the cleavage region of the substrate is in the position P1. When Gurken-A245F is used GlpG activity is completely annulled because is evident that the enzyme does not have a chance to act on another peptide bond like when Spitz is using as substrate. However when PA3086 is used the substrate is

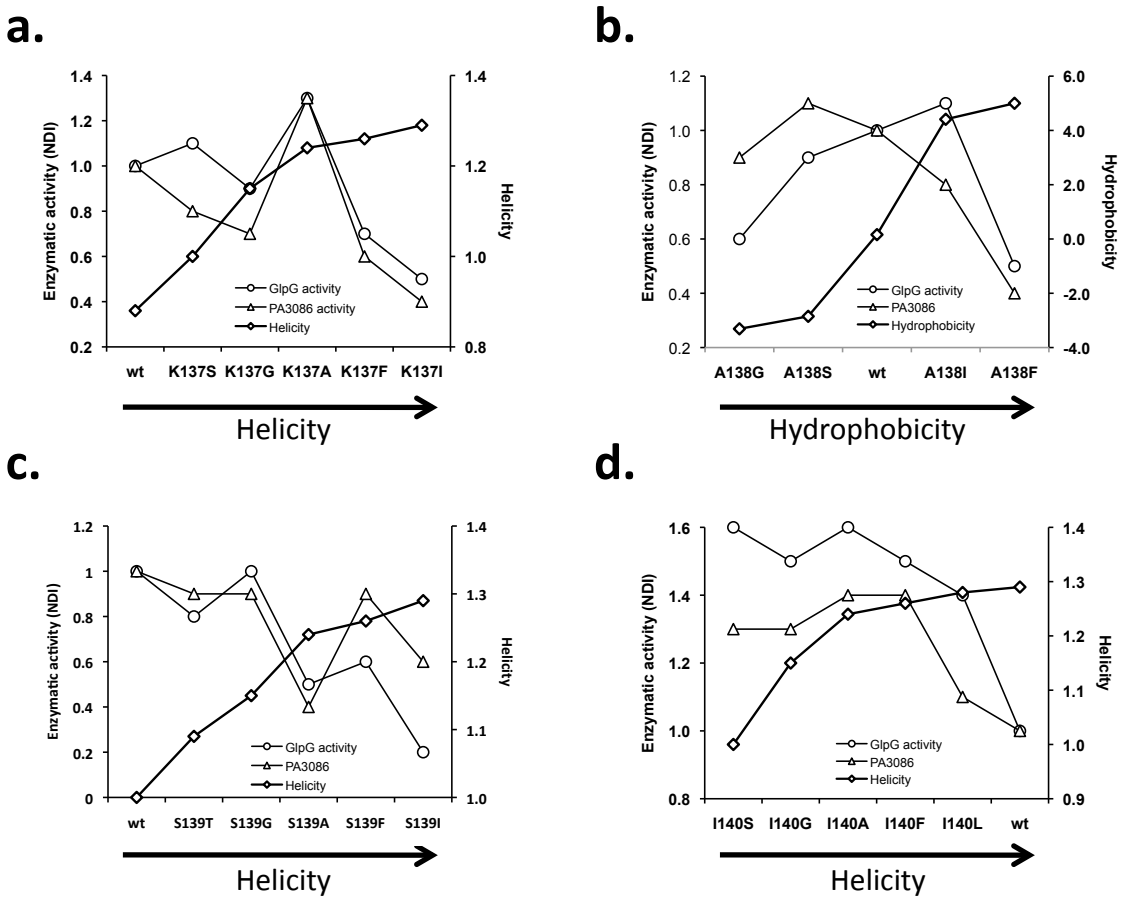

Fig. 3.10. Physical properties vs. activity of the amino acids used in the mutagenesis studies of substrate motif at the positions P2, P1, P1' and P2'. a). Helicity of the amino acids used in the mutagenesis studies at the position P2. NDI was plotted versus hydrophobicity. **b**). Same as **a.** but with the amino acids used in the position P1. **c**). Same as **a.** but with the amino acids used in the position P1'. **b**). Helicity of the amino acids used in the mutagenesis studies at the position P2'. NDI was plotted versus helicity.

cleaved in the same position C1 even if P1 is phenylalanine. The effect of the hydrophobicity in the position P1 is shown in the figure **3.10b**. In general it is possible according to the consensus ((3, 8, 9, 18-21) that it is necessary a small size residue in the position P1. Nevertheless, we found that the mutant Spitz-A138G with GlpG has a lower activity and when the mutants Spitz-A138S and Spitz-A138I the activity were not that much different than the wild type Spitz. One factor in the specificity can be the methylene group (C(α)-CH2-R) in the side chain of the amino acid located in the position P1. This property plus the small side chain amino acid can account for the correct activity of rhomboids. Strisovsky (9) found that in the position P1, the activity is conserved only if the amino acids used in the mutations are alanine or cysteine, all the other ones inhibit the enzyme.

Rhomboid has a preference for alanine at position P1'

Our study confirmed that the role of P1' is determinant in the rhomboid activity and the selection of the cleavage site. We found that the activity of rhomboid is inhibited progressively if the helicity in this position increases. There is a strong similarity with the cleavage site of signal peptide peptidase. This enzyme has a preference for residues with small side chains like alanine, valine and glycine(22, 23) but also phenylalanine and threonine. Gamma secretase complex also show a preference for this kind of amino acids: glycine, valine and leucine and threonine(24). In the Figure **3.10c** we can observe that the higher activities are observed with the amino acids that has a lower helicity but also with a smaller size chain

P2'

We found that that isoleucine could be acting as a frame for the recognition of the cleavage site and further digestion in the position P1. At the same time our studies shown that every single mutation performed in this position increases the activity of rhomboid but at the same time the cleavage site was modified in the case of the mutations: I140A, I140S and I149G when GlpG was used. In the case of PA3086 a similar effect was observed indicating that in both enzymes isoleucine plays a similar role in the mechanism of selection of the scissile peptide. In the figure **3.10d** it is possible to observe that the helicity of this position plays a role in the activity and also in the selection of the scissile peptide bond. However, this role has to pay a penalty giving that isoleucine increases the helicity. This can have a important biological relevance because it can be mean that rhomboids has a specificity for the scissile peptide bond.

Substrate motif

Once the cleavage sites of rhomboids site of rhomboid from evolutionarily distant organisms was determined the remained question is: how do rhomboids achieve peptide bond specificity? All rhomboids included in this research displayed a preference for an alanine residue at P1 position when wild type substrates are used. However there were 3 consecutive alanine residues in the substrates motif: K**A**SI**A**SG**A** in Spitz. This characteristic in the sequence of Spitz make difficult to study the role of alanine in position P1 because rhomboid has the possibility to shift the scissile peptide bond. How the enzyme recognizes with a high level of specificity the Ala138-Ser139 peptide bond in this case? The answer to this question will explain the differences obtained by Baker, *et al* (6) who determined that the cleavage site of rhomboid were two: Ala141/Ser142 and Gly143/Ala144 instead Ala138/Ser139. In this last research C100Spi-Flag as was used

in the determination of the cleavage site of rhomboid. According to this study, introduction of helix stabilizing residues, such as phenylalanine and tyrosine, reduces the digestion of the chimera Spitz substrate in insect cells that co-express rhomboid proteins. In the same study the digestion of the Spitz substrate was increased when helix-destabilizing residues (small side chain and beta branched side chain residues) were introduced in the substrate motif. This same effect of destabilizing residues in the activity of proteases was previously observed in the activity of signal peptide peptidase (25).

Finally we found for both enzymes, GlpG and PA3086, that the mutation G143A increases the activity of digestion for both rhomboids and also change in the position of the cleavage site to C4' (Gly143-Ala144). This aspect shows that rhomboids have a mechanism that is more ample and that the recognition in the position P1 and those extensive interactions are occurring in the selection of the substrate and cleavage site. These characteristics were observed in eukaryote rhomboid Rho-1 and RHBDL2 but unfortunately it was not possible to analyze the cleavage site for all these mutants.

As a consequence of our mutagenesis study we can propose that there is a motif in the substrate that the enzyme that define the substrate recognition by rhomboids and also specified the selectivity for the scissile peptide bond:

Table 3.1. Proposed Substrate motif for rhomboid intramembrane proteases

Position	Characteristic	Examples
P2	(Structural function) Strong helix destabilizing	Lysine
P1	Helix destabilizing Small side chain and, Cα-(CH2)-R	Alanine, glycine, Valine (?), Isoleucine.
P1'	Helix destabilizing residue.	Alanine, serine, glycine, cysteine.
P2'	Helix stabilizer Beta-branched residue	Isoleucine, threonine (?)

Bibliography

1. Urban S, Lee JR, & Freeman M (2001) *Cell* **107,** 173-182.
2. Urban S & Freeman M (2003) *Mol Cell* **11,** 1425-1434.
3. Akiyama Y & Maegawa S (2007) *Mol Microbiol* **64,** 1028-1037.
4. Maegawa S, Ito K, & Akiyama Y (2005) *Biochemistry* **44,** 13543-13552.
5. Urban S & Wolfe MS (2005) *Proc Natl Acad Sci U S A* **102,** 1883-1888.
6. Baker RP, Young K, Feng L, Shi Y, & Urban S (2007) *Proc Natl Acad Sci U S A* **104,** 8257-8262.
7. Ha Y (2008) *Semin Cell Dev Biol*.
8. Erez E & Bibi E (2009) *Biochemistry* **48,** 12314-12322.
9. Strisovsky K, Sharpe HJ, & Freeman M (2009) *Mol Cell* **36,** 1048-1059.
10. Alexandrov A, Dutta K, & Pascal SM (2001) *Biotechniques* **30,** 1194-1198.

11. Landolt-Marticorena C, Williams KA, Deber CM, & Reitchmeier RAF (1993) *J Mol Biol* **229,** 602-608.

12. Arkin IT & Brunger AT (1998) *Biochim Biophys Acta* **1429,** 113-128.

13. Hedstrom L (2002) *Chem Rev* **102,** 4501-4524.

14. Tulumello DV & Deber CM (2009) *Biochemistry* **48,** 12096-12103.

15. Tang YC & Deber CM (2002) *Biopolymers* **65,** 254-262.

16. Liu LP & Deber CM (1998) *J Biol Chem* **273,** 23645-23648.

17. Deber CM & Li SC (1995) *Biopolymers* **37,** 295-318.

18. Fluman N, Cohen-Karni D, Weiss T, & Bibi E (2009) *J Biol Chem* **284,** 32296-32304.

19. Erez E, Fass D, & Bibi E (2009) *Nature* **459,** 371-378.

20. Lemberg MK, Menendez J, Misik A, Garcia M, Koth CM, & Freeman M (2005) *EMBO J* **24,** 464-472.

21. Maegawa S, Koide K, Ito K, & Akiyama Y (2007) *Mol Microbiol* **64,** 435-447.

22. Weihofen A, Binns K, Lemberg MK, Ashman K, & Martoglio B (2002) *Science* **296,** 2215-2218.

23. Heimann M, Roman-Sosa G, Martoglio B, Thiel HJ, & Rumenapf T (2006) *J Virol* **80,** 1915-1921.

24. Bohm C, Seibel NM, Henkel B, Steiner H, Haass C, & Hampe W (2006) *J Biol Chem* **281,** 14547-14553.

25. Lemberg MK & Martoglio B (2002) *Mol Cell* **10,** 735-744.

Chapter 4.

Solution Structure and Dynamics of the N-terminal Cytosolic Domain of Rhomboid Intramembrane Protease from *Pseudomonas aeruginosa*: Insights into a Functional Role in Intramembrane Proteolysis.

Introduction.

In regulated intramembrane proteolysis (RIP), membrane proteins are specifically proteolyzed within their transmembrane region(1, 2) resulting in the release of soluble fragments that relocate within the intracellular or extracellular medium, where they can elicit a biological response. The intramembrane-cleaving proteases (ICLiPs) constitute a novel class of integral membrane proteins that are divided into aspartic proteases, metalloproteases and serine proteases based on their putative catalytic residues(3). Among aspartic intramembrane proteases the most prominent are the presenilins(4). Human presenilins release transcriptional activators in the Notch(5) and ErbB-4(6) cascades, and catalyze the final proteolytic step in the production of the neuropathogenic β-amyloid peptide(7). Metalloproteases are represented by the site 2 (S2P) zinc metalloproteases that regulate cholesterol biosynthesis(8). The only serine proteases discovered so far are the rhomboids(9) . In *Drosophila melanogaster,* Rhomboid cleaves the membrane bound epidermal growth factor receptor (EGFR) ligands, Spitz, Keren and Gurken, leading to their extracellular release and activation of the single EGF receptor pathway in a neighboring cell(2, 10-12). Though first discovered in *Drosophila* and shown to be key in every aspect of its development and homeostasis, Rhomboid homologs have been found in organisms ranging from bacteria

to humans and shown to play a role in intercellular signaling as well(13). In the Gram-negative bacterium *Providencia stuarti,* the rhomboid AarA(14) has been implicated in the generation of quorum-sensing signals(15, 16). In *Saccharomyces cerevisae*, which lacks the EGFR signaling pathway, the Rhomboid homolog RBD1 has been found to affect mitochondrial membrane remodeling(17). In mammals, the Rhomboid orthologs RHBDL2(18) and PARL(17) have been discovered, though their exact function and physiological significance are yet to be delineated. Recently, rhomboids have also been discovered in *Arabidopsis thaliana(19)*.

Sequence comparison between prokaryotic and eukaryotic rhomboids reveals low sequence homology overall, except for a serine protease-type catalytic triad (Asn, His and Ser) and their flanking regions, which are highly conserved, including the conservation of an overall topology consisting of six to seven transmembrane domains (TMDs)(2). Mutation of these putative catalytic residues both in prokaryotic and eukaryotic rhomboids abolishes protease activity. In addition, several prokaryotic(15), human(20), and plant(19) rhomboids have been shown to cleave *Drosophila* substrates, indicating strong mechanistic similarities across kingdoms.

Among rhomboids shown to display protease activity the presence of a relatively large (about 80–100 amino acid residues) N-terminal cytoplasmic region before the first TMD(12) with no significant sequence homology to other proteins, is evident. Although the catalytic elements of rhomboids appear to be wholly contained within the conserved TMDs (11), it has been suggested that this region plays a role in the regulation of intramembrane proteolysis (13). A variant of Rhomboid from *Drosophila*, that lacked this N-terminal domain, retained the ability to cleave substrate and displayed a phenotype comparable to the wild-type protein, suggesting that the catalytic elements of rhomboids were contained within the region harboring the multiple TMDs(11). Notably however, the

activity displayed by this Rhomboid N-terminal variant was greatly reduced compared to the wild-type protein(11), suggesting that the N-terminal region plays a role in the regulation of the catalytic activity of rhomboids. Based on sequence length, this region, or a significant part thereof, was expected to form a folded domain. We have determined the solution structure of the N-terminal domain of the *Pseudomonas aeruginosa* Rhomboid, herein referred to as NRho, using multidimensional NMR methods. The domain consists of a novel α/β fold and provides structural insight into its functional role in intramembrane proteolysis.

Materials and Methods.

Overexpression and purification of NRho.

A DNA construct encoding the N-terminal domain (amino acid residues 1–87) of *P. aeruginosa* rhomboid (NRho) was cloned into the pET-15b expression vector (Novagen), for inducible overexpression as an N-terminal His-tag fusion in *E. coli* BL21 (DE3) codon plus cells (Stratagene). Cells were grown at 37 °C until an A_{600} of 0.7 in media containing $^{15}NH_4Cl$ and/or [$^{13}C_6$] glucose, and induced by addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.4 mM. Cells were harvested after 4 h of induction, washed with 0.9% (w/v) NaCl, resuspended in the lysis buffer (50 mM Tris-HCl (pH 8.0), 300 mM NaCl, 10 mM imidazole) containing a protease inhibitor cocktail (Roche) and lysed by sonication. The cell debris was removed by centrifugation and the lysate was incubated with Ni-NTA agarose beads (Qiagen) for 4 h at 4 °C. After incubation, beads were washed with lysis buffer followed by PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM sodium phosphate dibasic, 2 mM potassium phosphate, pH 7.4), and incubated with thrombin (Novagen, 50 units per liter of cell culture) in PBS containing 2.5 mM of $CaCl_2$ for 16 h at room temperature. The cleaved protein was eluted and dialyzed against buffer 20 mM MES (pH 5.5), 5 mM NaCl and

purified using a HiTrap-SP (Amersham-Pharmacia Biotech) column. The purified protein was dialyzed against the NMR buffer (20 mM Bis-Tris (pH 6.5), 50 mM NaCl), and concentrated to 1 mM. 10 mg of pure $^{15}N,^{13}C$-labeled protein could be obtained from 1 L of culture. Purity of the sample was assessed by SDS-PAGE and its identity confirmed by mass spectroscopy.

Overexpression and purification of membrane domains of PA3086 and GlpG rhomboids.

Sequences encoding the predicted membrane domain of PA3086 (ΔNT-PA3086) (residues 87-286) genes were cloned into the pET-15b for inducible expression as N-terminal His_6tag fusion proteins and expressed as described previously. The membrane domain of GlpG was obtained from purified full-length GlpG which was digested with trypsin (1/70 ratio) at 4°C for 48 hours (21). After the digestion with trypsin was finished the obtained protein preparation was passed through a HiTrap™ Benzamidine FF (high sub) (GE Healthcare) in order to eliminate the remnants of trypsin and avoid it interferes with the rhomboid digestion reaction. The resulting protein was finally purified by size exclusion chromatography using a 16/60 HR200 Superdex column.

NMR spectroscopy.

Resonance assignment.

NMR spectra were recorded using ^{15}N and $^{15}N/^{13}C$ labeled NRho on either a Varian Inova spectrometer operating at ^{1}H frequency of 600 MHz or Bruker Avance spectrometers equipped with CryoProbes operating at ^{1}H frequencies of 700 and 800 MHz. All spectrometers were equipped with triple-resonance probes capable of applying pulsed field gradients along the *z*-axis. All experiments were performed at 25 °C, processed using the NMRPipe suite(22) and analyzed using NMRView(23). A

standard strategy based on heteronuclear triple-resonance NMR was used for resonance assignment(24). The HNCACB/CBCA(CO)NH pair was used for sequential assignment. Complete (excluding the last six residues 82–87) side-chain ^{13}C assignments were obtained using (H)C(CO)NH and ^{1}H assignments using HBHA(CBCACO)NH and H(CCO)NH experiments. Aromatic assignments were obtained using a combination of a constant time ^{13}C-^{1}H HSQC optimized for aromatic residues, and $H^{\delta}/H^{\varepsilon}$-$C^{\beta}$ correlation experiments developed by Yamazaki *et al*(25).

Angular and H-bond constraints.

$^{3}J_{NH}$ values were obtained from an HNHA(26) experiment and converted into backbone angle constraints. Additional backbone dihedral angle constraints were obtained from backbone chemical shifts using the TALOS(27) software suite. Hydrogen bond constraints were obtained from an HNCO experiment optimized for the through-H-bond NC′ coupling designed by Cordier *et al(25).*

Distance constraints.

The NOE-derived distance restraints were obtained from 3D ^{15}N-edited NOESY (150 ms mixing time), and 3D ^{13}C-edited NOESY (150 ms mixing time, separately optimized for aliphatics and aromatics). 3D NOESY cross-peak volumes/intensities were obtained using NMRView and converted into distance restraints using the ADR (ambiguous distance restraints) protocol within the Aria(28) suite. The manual NOE assignments were supplemented with those assigned automatically using the Aria scheme as described(28).

Structure calculation.

Manually assigned NOEs were combined with dihedral and hydrogen-bonding constraints into a standard assignment/structure-calculation protocol using Aria 2.0(28). This protocol yielded additional assignments that were incorporated in subsequent runs. Structure calculations were performed using a Cartesian dynamics simulated annealing protocol using: (i) a high temperature dynamics at 2000 K (10,000 steps), (ii) 4000 steps of refinement, (iii) a Cartesian dynamics cooling stage from 2000 to 1000 K (6000 steps) and (iv) a second Cartesian dynamics cooling stage from 1000 K to 50 K (4000 steps). Force constants of 5, 25 and 200 kcal mol^{-1} $Å^2$ for the dihedral constraints and 10, 10 and 50 kcal mol^{-1} $Å^2$ for the distance constraints (ambiguous, unambiguous and hydrogen-bond) were used during the three temperature stages of the Cartesian dynamics protocol. Prochiral valine and leucine methyl groups were treated using the floating chirality method(28). In the final refinement step, the lowest energy structures were solvated in a 7.5 Å box of water and refined with the full Lenard-Jones non-bonded potential and electrostatic interactions from the OPLS force-field using a protocol described in detail previously(29). The resulting NMR structures were evaluated with MOLMOL(30), DeepView and PROCHECK(31). In the final run, 1500 structures were calculated and the 100 lowest energy structures were used in the water-refinement protocol. 20 of the lowest energy water-refined structures with no violations of experimental constraints (no violations >0.5 Å for distance restraints and >5.0 ° for dihedral restraints) were chosen to represent the final structural ensemble of NRho.

Measurement and interpretation of relaxation rates.

Backbone relaxation rates.

A complete set of R_1, R_2, and $^1H^N$–{^{15}N}NOE measurements(32) for NRho were made at 600 and 800 MHz. Recycle delays of 1.5 s were used in the R_1 and R_2 relaxation measurements. The following relaxation delays were used to measure the R_2

values at 600 MHz, 10, 30 (×2), 70, 90, 110, 130, and 150 ms; and at 800 MHz, 4 (×2), 12, 22, 32, 42, 72, 102 (×2), and 148 ms. For the R_1 measurements, the following variable relaxation delays were used at 600 MHz, 0, 50 (×2), 90, 170, 330, 510 (×2), and 750 ms; and at 800 MHz, 10, 40 (×2), 80, 160, 320, 500 (×2), and 750 ms. All data were processed using the NMRpipe suite(22) and relaxation rates were obtained using in-house software that utilized the ODRPACK(33) library. The following average relaxation rates (residues 2–81) were obtained: R_1 (600 MHz) = 2.16(±0.3) s^{-1}, R_1 (800 MHz) = 1.62(±0.17) s^{-1}, R_2 (600 MHz) = 7.7(±4.29) s^{-1}, R_2 (800 MHz) = 10.09(±3.35) s^{-1}. Steady-state $^1H^N$–{^{15}N}NOE values (0.70 ± 0.35 at 800 MHz, 0.68 ± 0.16 at 600 MHz) were obtained by recording two spectra with and without a 3.0 s period of proton saturation. Errors in NOE values were obtained from duplicate datasets.

Analysis of hydrodynamic properties.

The hydrodynamic properties of NRho were determined using the program DIFFTENS v2.0 utilizing previously described methods(34). Errors (68.3% confidence limits) in the principal values and orientation of the rotational diffusion tensor were obtained from the analytically determined inverse covariance matrix of the fits. Selection of models between the fully anisotropic, axially symmetric and isotropic models was performed using the statistical *F*-test. Probabilities (*P* %) indicating the possibility that the improvement in fits on increasing model complexity were obtained by chance, were calculated for each of the pairs of models: fully anisotropic/axially symmetric and axially symmetric/isotropic. Values of $P > 5\%$ were not considered to be statistically significant. This analysis yielded an isotropic diffusion tensor and a correlation time of 5.48(±0.04) ns.

Determination of backbone flexibility using the Lipari-Szabo model-free framework.

An analysis of the micro-dynamic motional parameters using the Lipari-Szabo formalism(35, 36) was performed utilizing the program DYNAMICS(37) utilizing the measured relaxation data (at 600 and 800 MHz). Several runs were performed using multiple subsets of the relaxation rates in order to determine the stability of the analysis and the robustness of the selected models and estimated micro-dynamic parameters with respect to completeness of the datasets. Errors (68.3% confidence limits) in the micro-dynamic parameters were obtained from the analytic inverse covariance matrices of the fits. Models used and computational strategies employed were as those described(37).

Measurement of the side-chain ZQ/DQ relaxation rates

The relaxation rates of zero and double-quantum coherences involving methyl ^{13}C nuclei ($^{13}C^{methyl}$) and the ^{13}C nucleus next to it ($^{13}C^{next}$) were measured using the pulse sequences presented by Del Rio *et al(38).* Using the following relaxation delays, τ_M = 1, 4, 7, 14, 16, 22, 28, 38, 42, 44, 50 and 58 ms. The decay curves were fitted to equation (1) below, using procedures discussed in detail previously (see **Figure 4.8(b)**, for a representative example) (38):

$$S_{ZQ/DQ}(\tau_M) = A_{ZQ/DQ}\cos(\pi J_{ZQ/DQ}\tau_M)^m e^{-\Gamma_{ZQ/DQ}\tau_M} \quad (1)$$

where $J_{ZQ/DQ}$ and $\Gamma_{ZQ/DQ}$ are the effective scalar coupling constants(39) and the relaxation rates, respectively, for the zero/double quantum coherences. $A_{ZQ/DQ}$ are constant amplitude factors and m = 0 for coherences involving Ala C^{β}, m = 1 for coherences involving Thr $C^{\gamma 2}$ and m = 2 for all other coherences (NRho contains no Ile residues).

The effects of dipolar couplings with remote protons were estimated using the final cluster of 20 structures described above and chemical shift anisotropy effects were estimated as described(38). No attempt was made to interpret the fast dynamics and those coherences with anomalously large (absolute) differences ($\Delta\Gamma$) between Γ_{DQ} and Γ_{ZQ} were ascertained to be subject to slow conformational exchange.

Determination of the interaction of NRho with C_{16}PN micelles.

NMR spectroscopy.

Stock solutions of ^{15}N-labeled NRho (1 mM) and C_{16}PN (20 mM) were prepared in the NMR buffer. Different amounts of NRho were titrated into 10 mM C_{16}PN to yield eight samples with NRho:C_{16}PN molar ratios of 1:20, 1:25, 1:33.33, 1:40, 1:50, 1:66.66, 1:100, and 1:200. ^{1}H-^{15}N HSQC spectra were recorded at each titration step. Two control ^{1}H-^{15}N HSQC spectra were recorded, one for pure detergent and the other for pure protein. Note that the concentration of C_{16}PN used was 1000-fold higher than its critical micelle concentration (CMC = 0.01 mM) ensuring the formation of C_{16}PN micelles.

The dissociation constant (K_d) for the C_{16}PN/NRho complex was determined by fitting the normalized intensity of the Trp41, $N^{\varepsilon}/H^{\varepsilon}$ resonances corresponding to the bound and free states to equation (2) below:

$$f_B = \frac{K_d + \frac{[D]_0 - C_M}{n} + [P]_0 - \sqrt{\left(K_d + \frac{[D]_0 - C_M}{n} + [P]_0\right)^2 - 4\frac{[D]_0 - C_M}{n}[P]_0}}{2[P]_0} \quad (2)$$

where $f_B = I_b/(I_b + I_f)$ is the fraction of the membrane bound state calculated from the integrated intensities of the peaks corresponding to the bound (I_b) and free (I_f) states, respectively. $[P]_0$ and $[D]_0$ are the total concentration of the protein and the detergent, respectively; C_M is the critical micelle concentration and n is the aggregation number of C_{16}PN. Non-linear least square fits to equation (2) were performed using in-house software that utilized the ODRPACK library(33). The fitted parameters K_d and n (aggregation number) were found to be 8.02(±1.48) μM and 66.76 ± 8.92, respectively.

Far UV CD spectroscopy

Spectra were collected on a Jasco 810 spectro-polarimeter at 25 °C using a cuvette with a 1 cm path-length. The protein concentration was 7.5 μm, in buffer containing 10 mM sodium phosphate dibasic, 2 mM potassium phosphate, 50 mM NaCl at pH 6.5. Spectra were acquired with a resolution of 1 nm from 280 to 200 nm, with a 4 s acquisition time signal-averaged over ten scans. The concentration of detergent was varied to generate four samples with NRho:C_{16}PN molar ratios of 1:20, 1:40, 1:100 and 1:200. A reference spectrum with the same concentration of protein and identical buffer conditions as above but lacking detergent was also measured. Estimates of the helical content were obtained using the following equation(40):

$$f_{\mathrm{H}} = \frac{\theta_{222} - 2220 + 53T}{\theta_{\mathrm{H}} - 2220 + 53T} \tag{3}$$

$$\theta_{\mathrm{H}} = (-44000 + 250T)\left(1 - \frac{3}{N}\right)$$

where f_H is the fractional helical content at temperature T (in °C), θ_{222} is the molar ellipticity measured at 222 nm, θ_H is molar ellipticity of a pure α-helix, and N is the

number of residues in the protein. The results were further confirmed using the CDPRO package published by Sreerama and Woody(41).

Fluorescence spectroscopy

Fluorescence spectra were obtained with a Spex Fluorolog-3 (Horiba Jobin Yvon) spectrofluorometer at 25 °C with the excitation and emission slits at 2 nm. The fluorescence emission spectra (305–450 nm in 1 nm steps and signal-averaged over three scans) were recorded using excitation at 295 nm for each addition of C_{16}PN micelles to the protein sample (9 μM) in NMR buffer. The fluorescence intensity was corrected for protein dilution and the binding parameters were determined by fitting to a simple hyperbolic binding curve using in-house programs that utilized the ODRPACK library(33). The fitted parameters, K_d and n (aggregation number) were found to be 1.39(±0.54) μM and 44.85 ± 3.15, respectively. The fractional change in fluorescence intensity at saturation was found to be 53.01(±1.55)%.

Protein Data Bank accession codes

The coordinates of NRho have been deposited in the RCSB Protein Data Bank with accession code 2GQC.

Results

Solution structure of Nrho

Analysis of the *P. aeruginosa* Rhomboid sequence using the TMHMM membrane protein topology prediction algorithm(42) yielded a six TMD topology (putative TMDs lie between residues 88–111, 144–166, 178–197, 202–224, 231–253 and 263–282), a relatively large N-terminal domain (residues 1–87) and with both N and C termini facing

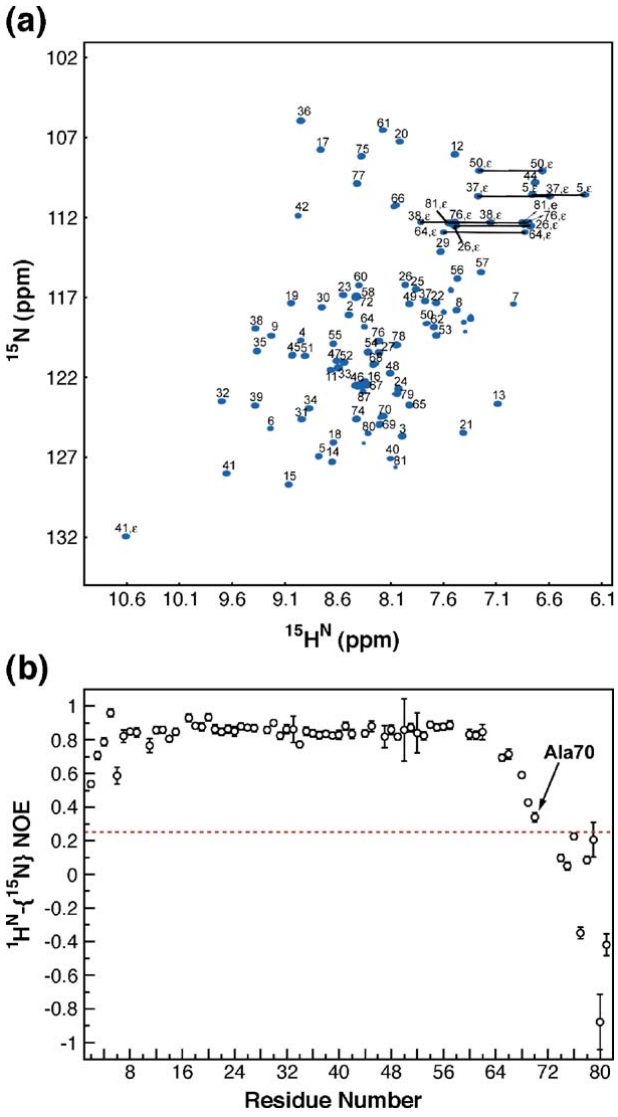

Figure 4.1. NRho forms a folded domain. a). ^{1}H-^{15}N HSQC spectrum of NRho at 700 MHz displaying resonance assignments. The resonances corresponding to residues 82–87 were very weak and not assigned. **b).** $^{1}H^{N}\{^{15}N\}$ NOE of NRho at 800 MHz. The broken red line indicates an NOE value of 0.25. Residues 1–70 were included in the structure calculation.

the cytosol. Thus, a construct (NRho) encoding the first 87 amino acid residues of rhomboid was over expressed in *Escherichia coli* and purified to homogeneity for solution NMR studies. The identity of the purified domain was confirmed by mass spectrometry. NRho exhibited a well-resolved ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectrum (**Figure 4.1(a)**) and had a rotational correlation time of 5.48(±0.04) ns (see **Materials and Methods**), characteristic of a folded, monomeric species in solution. Resonances corresponding to the last six C-terminal amino acid residues (82–87) were extremely weak, most likely due to exchange broadening, and no attempt was made to assign them. An inspection of the $^{1}H^{N}$-{^{15}N} steady-state nuclear Overhauser effect (NOE) values (**Figure 4.1**(b)) revealed that the amino acid residues 71–81 were unstructured with NOE values less than 0.25 at 800 MHz and hence, were excluded from the final structure calculations.

The NRho ensemble (residues 1–70) forms a tight cluster (**Figure 4.2(a)** and **Table 4.1**) and consists of a mixed α/β fold comprising a three-stranded antiparallel β-sheet (β1, 7-9; β2, 30-34; and β3, 37-41) and two approximately orthogonal α-helices (α1, 16-25; and α2, 47-55) (**Figure 4.2(b)**). There are few helix–helix contacts involving the hydrophobic side-chains of Phe18, Leu22 and Leu25 on α1 and Tyr55 on α2. Additional α1–α2 contacts are also observed between Phe18 and Arg52. There are extensive contacts between Leu7 and Phe9 on β1 and Phe18, Val19 and Leu22 on α1, and also between Val6 and Leu7 on β1 and Ala48, Val51 and Arg52 on α2.

NRho is shaped like a triangular wedge and displays a cluster of several acidic residues (Asp14, Glu45, Glu49, Glu53, Glu60, Asp62 and Glu68) localized on one face of the molecule forming a continuous surface (**Figure 4.2(c)**). The role of this negatively charged surface is yet to be elucidated (see below).

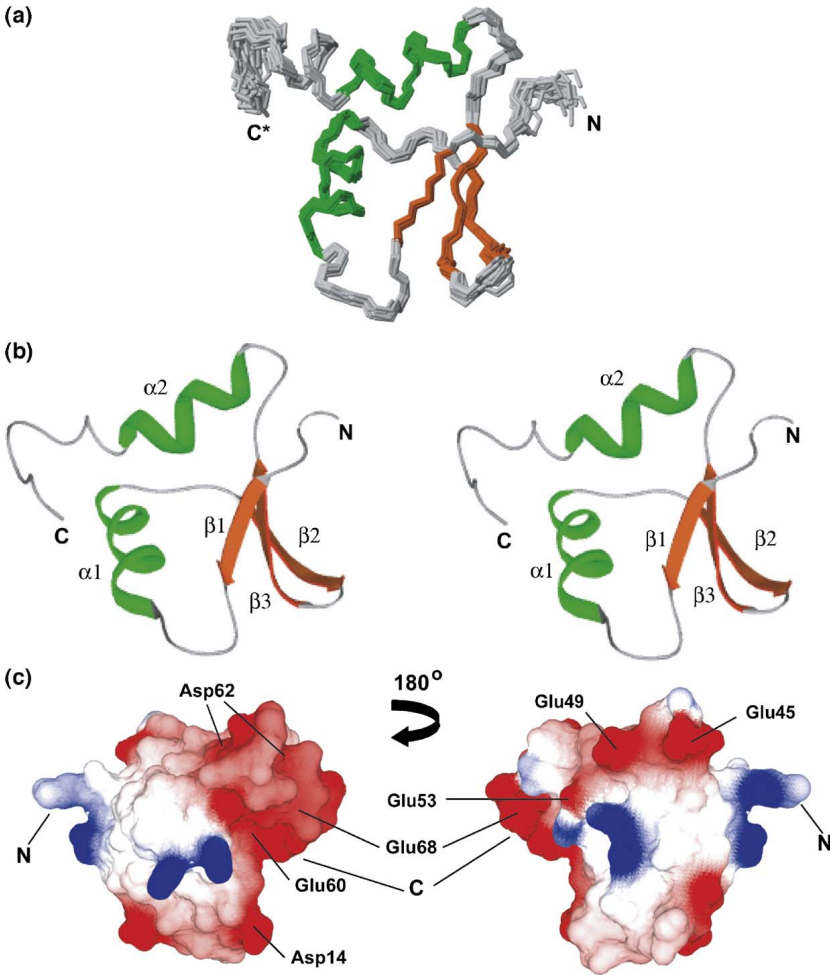

Figure 4.2. Solution structure of NRho. a). Cluster of the final ensemble of 20 NMR structures of NRho. Helices are colored green and the β-strands are colored orange. Only residues 1–66 are shown for clarity. The C* implies that the residues 67–70 have been removed from the C-terminal end. **b).** Stereo-view of a ribbon plot of the lowest energy structure from the NRho ensemble. **c).** NRho provides a highly negatively charged face that may participate in electrostatic interactions with cytosolic domains of Rhomboid targets. Key acidic residues are labeled. Note that the orientation shown is for optimally viewing the charged surface and is different from **a.** and **b.** above.

Table 4.1. Structural statistics for the 20 final NMR structures of Nrho.

NMR constraints	
Distances Constraints	
Total NOE	1830
Unambiguous	1469
Ambiguous	361
Intra-residue[a]	585
Inter-residue[a]	884
Sequential (|i–j| = 1)	344
Medium-range (|i–j| ≤ 4)	305
Long-range (|i–j| > 5)	235
Hydrogen bonds[b]	
Dihedral angle	
Φ	48
Ψ	48
Structure statistics[c]	
Violations	
NOE violations > 0.5 Å	0
Dihedral angle violations > 5°	0
Deviations from idealized geometry	
Bond lengths (Å)	$7.60 \times 10^{-3} \pm 2.38 \times 10^{-4}$
Bond angles (°)	0.96 ± 0.03
Impropers (°)	2.45 ± 0.09
Agreement with experimental restraints	
NOE	$5.46 \times 10^{-2} \pm 2.69 \times 10^{-3}$
Dihedral angles	0.81 ± 0.11
Average RMSD (Å)[d]	
Backbone[e]	0.45±0.09(0.25±0.06)
Heavy atoms[e]	0.89±0.08(0.79±0.13)
Ramachandran plot[f]	
Most favored (%)	69.8 (71.5)(92.9)
Additionally allowed (%)	22.0 (19.4)(7.1)
Generously allowed (%)	7.5 (9.1)(0.0)
Disallowed (%)	0.7 (0.0)(0.0)

a) Splits shown only for unambiguous assignments.
b) Hydrogen bond restraints were H^N–O distance of 1.8–2.3 Å and an N–O distance of 2.8–3.3 Å.
c) Structural characteristics for the final ensemble of 20 water-refined structures.
d) RMSD of the mean structure from individual structures in the ensemble.
e) RMSD for residues 3–66 shown (these residues have $^1H^N$-{^{15}N}NOE at 800 MHz > 0.6). The numbers in the parentheses indicate the RMSD for regions with definite secondary structure.
f) Ramachandran plot data shown for residues 3–66. The first set of numbers in parentheses indicates the statistics for the smaller subset after removal of residues (3, 5, 6, 7, 35, 36 and 42) shown to undergo a significant amount of conformational exchange. The second set indicates the characteristics for those residues in ordered regions of definite secondary structure.

A systematic search of the protein databank (PDB) utilizing the **DALI server‡**(43) did not produce any close apparent structural homologs. Only a remote homolog with a similar overall α/β fold corresponding to the N-terminal domain of EscJ from the enteropathogenic *E. coli* (EPEC) was found (**Figure 4.3**, PDB code 1YJ7; average *Z* value over the final cluster of 20 structures was 2.8 ± 0.1, C^{α} RMSD = 3.4(±0.2) Å; no other structures within the PDB had an average $Z > 2.5$). EscJ is a member of the PrgK/YscJ family and forms an oligomeric ring(44) that is a key component of the needle complex (NC) of type III secretion systems (TTSS) in bacteria(45). EscJ has been shown to be lipidated at its N terminus and localizes to the bacterial cytoplasmic membrane(44), suggesting a role in protein/membrane interaction. NRho does not possess a lipidation signal in its sequence, but its proximity to the bacterial cytoplasmic membrane and overall structural similarity to EscJ raises the possibility that it too may be capable of peripheral association with a membrane.

Interaction of NRho with lysophospholipid-analog micelles

To test whether NRho could specifically interact with membranes, we titrated NRho into a solution containing micelles of the zwitterionic lysophospholipid-analog *n*-hexadecylphosphocholine (C_{16}PN). CD spectra revealed a small change in the secondary structure content in the protein with increasing detergent concentration (see **Figure 4**). The helical content estimated from molar ellipticity values measured at 222 nm (using equation (3), see Materials and Methods) registered a small increase (3.1%, approximately two residues) on going from pure NRho (no detergent) to a NRho:C_{16}PN molar ratio of 1:200. The CD spectra (**Figure 4**) displayed an isodichroic point(46) at 217 nm indicating two discrete states (possibly corresponding to the free and micelle-bound states of NRho). Significant changes were observed in the ^{1}H-^{15}N HSQC spectra of NRho (a representative example is shown in **Figure 4.5(a)**) with

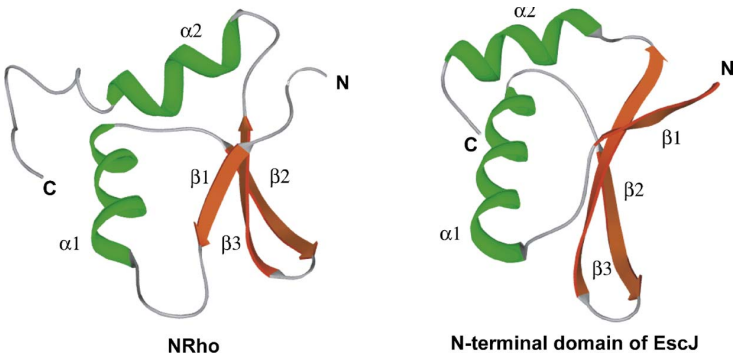

Figure 4.3. A remote structural homolog of NRho. Comparison of the overall fold of NRho (left) and the N-terminal domain of EscJ from the enteropathogenic *E. coli* (right, PDB code: 1YJ7). Both structures reveal an overall α/β fold though the average DALI Z-score on comparison with the final ensemble of 20 structures of NRho was only 2.8 ± 0.1.

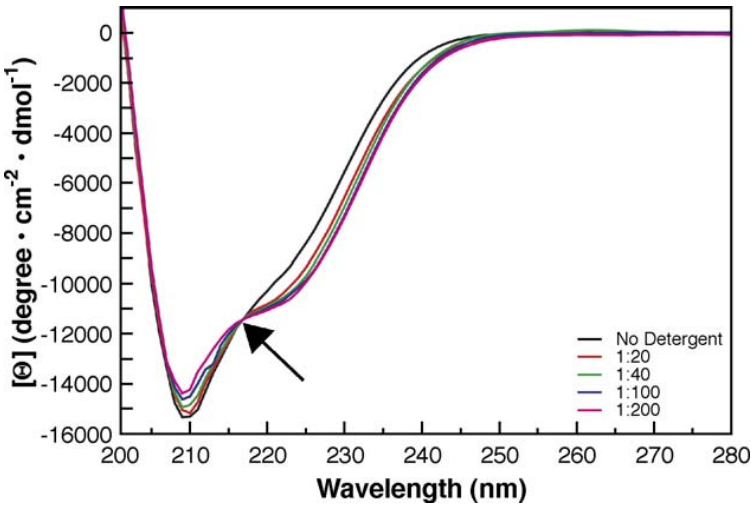

Figure 4.4. NRho binding to C_{16}PN detergent micelles occurs through a two-step process. Far UV circular dichroic (CD) spectra of NRho with varying concentration of C_{16}PN detergent. The C_{16}PN:NRho molar ratios are indicated in the Figure. The isodichroic point is indicated by a black arrow.

increasing detergent concentration. Backbone resonances corresponding to amino acid residues Ser2, Val4, Val6, Leu7, Ser35, Leu40 and Val42 were significantly and preferentially attenuated (all these resonances were undetectable above the noise level at the highest NRho:C_{16}PN molar ratio measured 1:200), as were those corresponding to the side-chains for Gln5 and Gln38. Partial attenuation was also observed in the Ala3 resonance. Notably, no protein aggregation was observed even at the highest protein concentrations. In the course of the titration (with increasing concentration of protein), the peak corresponding to the side-chain $N^{\varepsilon}/H^{\varepsilon}$ resonance of Trp41 indicated a clear transition from the micelle-bound state to the free state, with the intermediate titration steps displaying two sets of peaks (**Figure 5(a)**). The apparent dissociation constant (K_d) for the interaction, determined from an analysis of the peak intensities (using equation (2), see Materials and Methods) of the micelle-bound and free states corresponding to this resonance (**Figure 5(b)**), was found to be 8.02(±1.50) μM, characteristic of a relatively strong interaction.

To confirm the affinity of the NRho/micelle interaction we also used fluorescence spectroscopy. Excitation of NRho at 295 nm resulted in a fluorescence spectrum with maximum intensity at 344 nm. A change of the fluorescence intensity at the emission maximum as a function of C_{16}PN detergent concentration was noted and the resulting hyperbolic saturation curve was used to estimate binding parameters (data not shown). The obtained apparent K_d of 1.39(±0.54) μM was in good agreement with that estimated from the NMR experiments above, confirming that NRho displays high affinity for micelles of the lysophospholipid analog C_{16}PN. Notably, the change in intrinsic fluorescence displayed by NRho can only arise from its single tryptophan residue, thus confirming that the environment of Trp41 changes upon micelle binding.

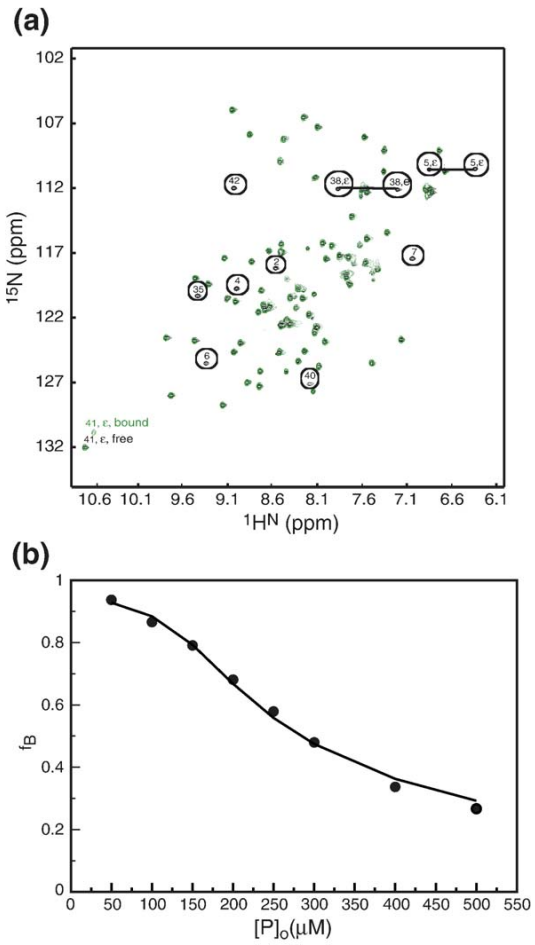

Figure 4.5. Interaction of NRho with detergent micelles. a). Representative ^{1}H-^{15}N HSQC spectrum of NRho in the presence (green) and absence (black) of $C_{16}PN$ micelles showing differential attenuation of resonances. Key resonances are labeled. Two sets of peaks corresponding to the free and micelle-bound states of NRho for the Trp41 side-chain $H^{\varepsilon}/N^{\varepsilon}$ resonance are seen. The molar ratio of NRho:$C_{16}PN$ in the displayed spectrum was 1:40. **b).** Determination of the K_d for the NRho/micelle interaction using equation2. Normalized fraction of the micelle-bound (f_B) state was determined from the intensities of the corresponding peaks for the Trp41, $^{\varepsilon}/_{\varepsilon}$ resonances. Experimental points are indicated by circles and the theoretical curve by the continuous line.

The observed high affinity towards the micelle interface suggests the presence of a specific membrane-binding site in NRho. In fact, the residues that displayed significant spectral changes in the presence of the detergent mapped on to a continuous surface (**Figure (6)**). Furthermore, this surface was completely independent of the negatively charged surface discussed above and lay on the opposite face of NRho. Notably, several of the residues (Ala3, Val4, Leu7, Gln38, Trp41 and Val42) that interact with micelles had buried side-chains, with less than 15% of their total surface area exposed to solvent in the non-micelle bound native state of the domain. Therefore a conformational change would be required for optimal positioning of these residues in the hydrophobic membrane environment upon interaction. An inspection of the structure of NRho reveals that a major conformational change to the NRho fold would not be necessary and that such a change could essentially involve a motion of the N-terminal tail, including the N-terminal end of β1 away from the protein core coupled with minor rearrangement of side-chain conformations for some residues in β2 and β3. This conformational change would be sufficient to expose the key hydrophobic residues (implicated above) to the membrane. It is reasonable to expect that some of these conformational modes would be sampled by the native state of NRho, as demonstrated in other systems(47) and this would be manifested as the effects of slow dynamics in the spin-spin (R_2) relaxation rates as exchange contributions (R_{ex} terms).

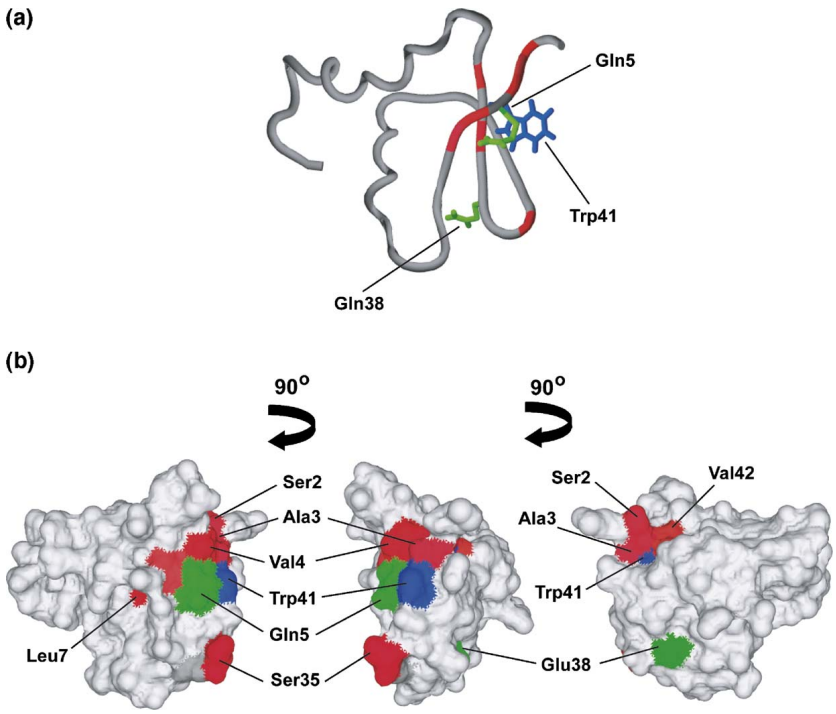

Figure 4.6. Membrane interaction site of NRho. **(a)** Key residues that interact with C_{16}PN micelles lie in the β-sheet region of NRho and in the hairpin connecting β2 and β3. Residues that show attenuation in peak intensity in the presence of C_{16}PN micelles for their backbone resonances are shaded red. Side-chains corresponding to Gln5 and Gln38 that also show peak attenuation in the presence of micelles are shown (green), as is the Trp41 side-chain (blue). **(b)** All residues that interact with C_{16}PN micelles map onto a nearly continuous surface on one face of NRho. Key residues are labeled and the coloring scheme is the same as in (a).

Dynamics of the backbone and methyl-bearing side-chains in NRho

An analysis of a complete set of backbone R_1, R_2 and $^1H^N$-{^{15}N}NOE experiments collected at 600 MHz and 800 MHz using the Lipari-Szabo model-free framework (see Materials and Methods) revealed little disorder in the NRho core (1-70) on the fast (picoseconds–nanoseconds) timescale. The generalized order parameters S^2 (**Figure (7)**) were indicative of a well-ordered structure with some low values at the extreme N and C termini. The helical regions were extremely well ordered ($S^2 > 0.9$) and some disorder (relative to the α-helices) was observed in the loop between β1 and α1, at the C-terminal end of β2 and the N-terminal end of β3. However, there was a significant amount of slow dynamics (microseconds–milliseconds timescale) in the NRho backbone as revealed by the large R_{ex} contributions to the spin-spin relaxation rates (R_2) (**Figure 8(a)**). Gln5 (5.7(±0.2) s^{-1} at 600 MHz), Val6 (29.1(±3.1) s^{-1}), Leu7 (10.2(±0.8) s^{-1}) and Val42 (9.2(±0.4) s^{-1}) revealed extremely large R_{ex} values. Spin-spin relaxation rates corresponding to Val4 and Asp44 contained smaller but significant R_{ex} contributions of 1.9(±0.2) s^{-1} and 1.5(±0.1) s^{-1}, respectively. In addition, residues Ser35 and Gly36 had extremely large R_2 values and data for these residues could not be fitted reliably, suggesting large exchange contributions as well.

To further investigate the nature of slow dynamics in NRho, we measured the difference ($\Delta\Gamma = \Gamma DQ - \Gamma ZQ$) in the relaxation rates ($\Gamma DQ/ZQ$) between zero-quantum (ZQ) and double-quantum (DQ) coherences generated between side-chain nuclei in methyl-bearing residues. The DQ/ZQ coherences were generated between methyl ^{13}C ($^{13}C^{methyl}$) and the ^{13}C nucleus immediately next to it ($^{13}C^{next}$). The coherences involved were C^{β}-$C^{\gamma 2}$ in Thr, C^{α}-C^{β} in Ala, C^{β}-$C^{\gamma 1,2}$ in Val, and C^{γ}-$C^{\delta 1,2}$ in Leu residues(38). In the absence of slow dynamics on the microseconds–milliseconds timescale, the ΔΓ values

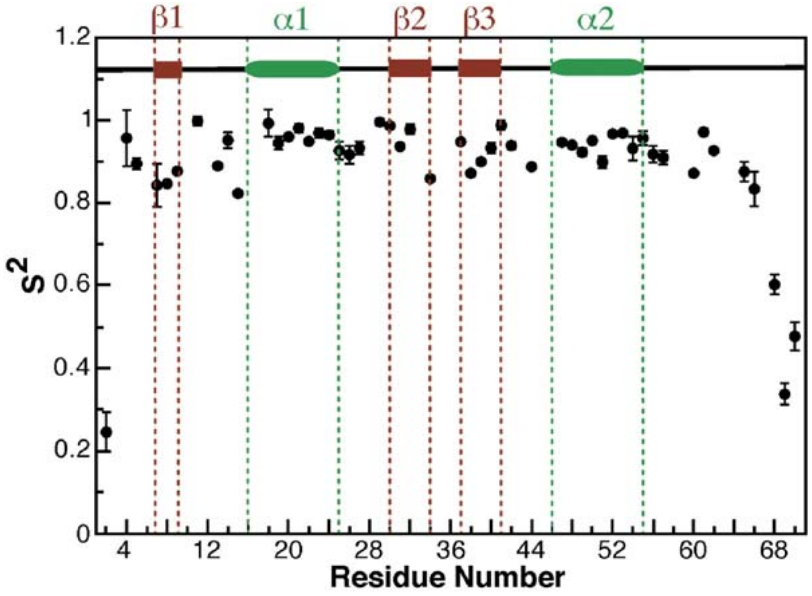

Figure 4.7. NRho is well-ordered on the picoseconds–nanoseconds timescale. S^2 values for NRho obtained by an analysis of backbone R_1, R_2 and $^1H^N\{^{15}N\}$ NOE data at 600 and 800 MHz using the Lipari-Szabo model-free formalism. Regions of secondary structure are indicated and only residues 1–70 are shown.

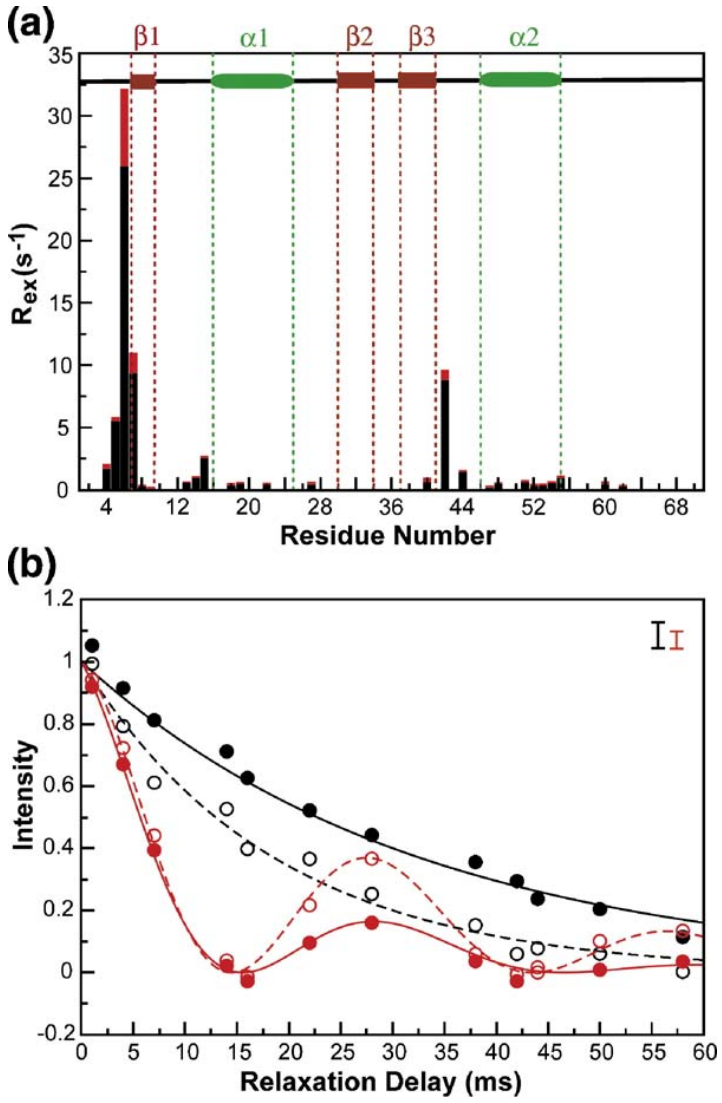

Figure 4.8. NRho shows extensive conformational flexibility for both the backbone and the side-chains on the microseconds–milleseconds timescale. (a) R_{ex} values at 600 MHz for NRho obtained from an analysis of backbone R_1, R_2 and $^1H^N\{^{15}N\}$ NOE data at 600 and 800 MHz using the Lipari-Szabo model-free formalism. Errors in the R_{ex} values are indicated by red risers. R_{ex} values scale as the square of the static magnetic field. (b) Fits of the Ala3 (black) and Leu7 (red) C^{methyl}-C^{next} zero-quantum coherences (ZQ, experimental points, open circles; theoretical curve, dotted line) and double-quantum coherences (DQ, experimental points, filled circles; theoretical curve, continuous line) to equation (1). Errors in the experimental data points (Ala3, black; Leu7, red) are indicated by the error bars at the top right-hand corner of the Figure.

lie within a narrow range depending on the range of mobility of the C^{next}–H^{next} (where H^{next} denotes the proton attached to the C^{next} nucleus) bond-vector. This range of ΔΓ values is dominated by local effects, and may be estimated from simple analytical expressions in the case of a completely ordered or a completely disordered C^{next}–H^{next} bond when the rotational correlation time is known (38). Further, the variance in these rates with residue-type and with the protein fold can be estimated when the structure of the protein is known. Thus, for NRho, these values were expected to lie between –2.7 s^{-1} (completely ordered) and –10.9 s^{-1} (fully disordered). However, in the presence of slow conformational exchange leading to the modulation of isotropic chemical shifts (which also lead to the increase of spin-spin relaxation rates through large R_{ex} contributions) the can be of either sign and have very large magnitudes. Thus, analysis of values obtained from fits of the data to equation (1) (see Materials and Methods below, **Figure 8(b)**) revealed that for three sets of coherences, namely the Leu7 C^{γ}-$C^{\delta 1,2}$ coherences (−16.3(±3.5) s^{-1} and +25.8(±4.0) s^{-1}) and the Ala3 C^{α}−C^{β} coherence (−23.9(±4.5) s^{-1}), these values occurred far beyond the expected range and that the side-chains for these residues were subject to a significant amount of conformational exchange on the microseconds–milliseconds timescale.

Thus, from the analysis of backbone and side-chain relaxation properties of NRho, the following residues: Ala3, Gln5, Val6, Leu7, Ser35, Gly36 and Val42, were ascertained to be dynamic on the slow (microseconds–milliseconds) timescale. These residues mapped on to a surface (**Figure 9**) including and contiguous to that implicated in membrane interaction (see **Figure 9(c)**). The present studies thus confirm the presence of a significant amount of structural plasticity on the microseconds–milliseconds timescale at the membrane interaction site.

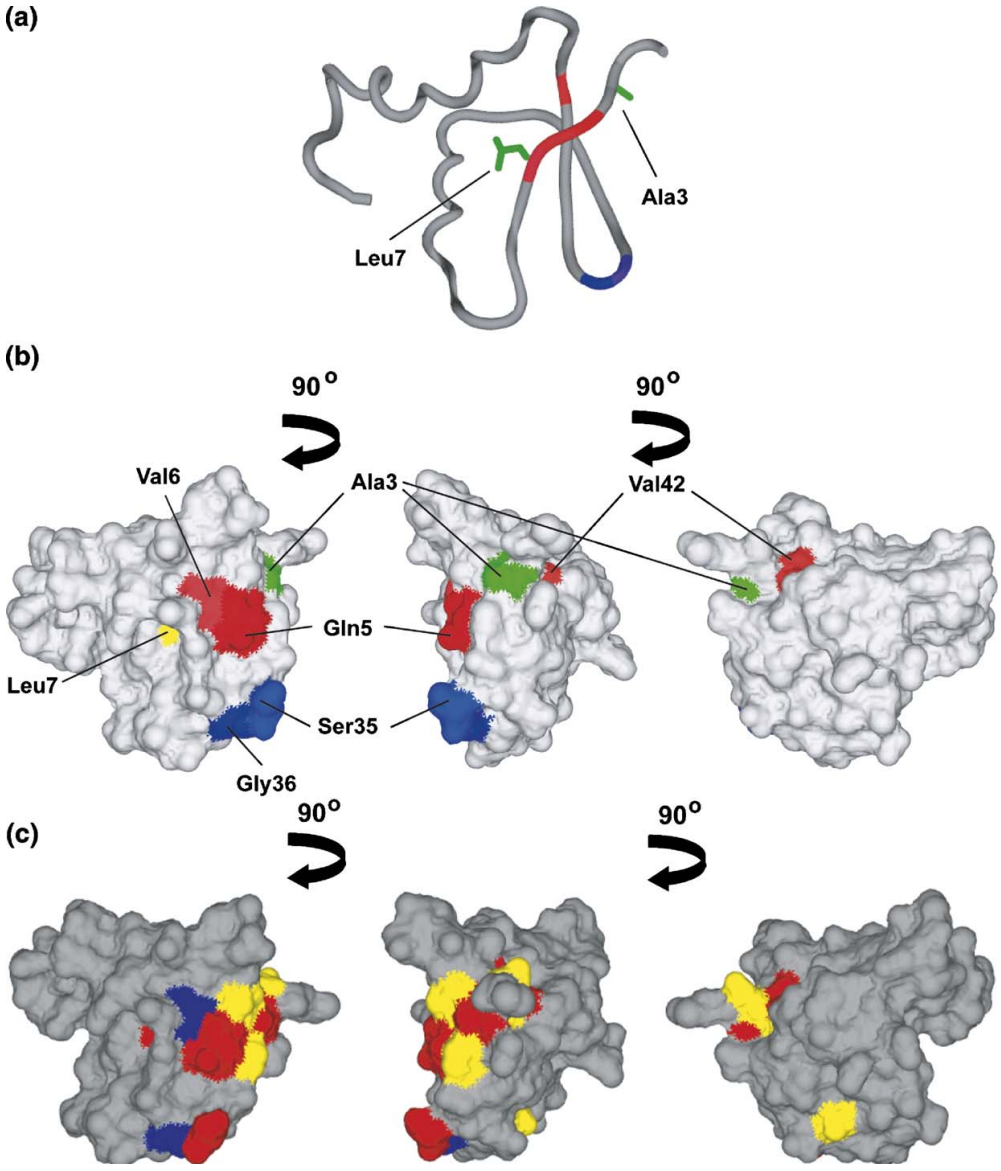

Figure 4.9. Residues that display dynamics on the microseconds-milleseconds timescale map on to a continuous surface. (**a**) Residues that display large R_{ex} values (>5 s^{-1}), as obtained from an analysis of backbone R_1, R_2 and $^1H^N\{^{15}N\}$ NOE data at 600 and 800 MHz using the Lipari-Szabo model-free formalism, are shaded red. Backbone relaxation data corresponding to Ser35 and Gly36 could not be analyzed accurately due to large R_2 values (see the text); these residues are shaded blue. Ala3 and Leu7 side-chains that were shown to display slow dynamics as determined from multiple-quantum experiments (see the text) involving methyl groups, are displayed and shaded green. (**b**) The residues that display slow dynamics map onto a continuous surface

contiguous to that implicated in membrane-interaction. The shading scheme is the same as in (a) except in the case Leu7 that shows slow dynamics both in the backbone and side-chain regions and is shaded yellow. (**c**) Residues that interact with C_{16}PN liposomes and display slow microseconds-milliseconds timescale motion (common with Figure 6) are shown in red. Residues that interact with C_{16}PN liposomes but are not significantly dynamic on the slow timescale (seen only in Figure 6, see the text) are shown in yellow. Additional residues that are dynamic on the slow timescale are shown in blue.

Activity of the membrane domain of GlpG and PA3086 rhomboids.

The possible regulatory role on the activity and the cleavage site selectivity was study using membrane domain of GlpG and PA3086: ΔNT-GlpG (NRho) and ΔNT-PA3086 respectively, to determine activity with the substrates we designed. In the **Figure 10(a)** and **10(b)** we can see the activity of the membrane domains with all the six substrates. We can observe in these gels that the activity is not being modified for the lacking of the cytoplasmic domain and both truncated enzymes show a similar pattern than the obtained with the full-length enzymes showed before (chapter 2).

Discussion

As described previously, there is not a significant alignment between rhomboids and they show a high structural conservancy between them and with certain residues that are involved in the active site. More specifically the conservancy of the cytoplasmic is restricted to the rhomboid proteins in other bacteria that belong to Pseudomonas genre and few other bacteria (Fig 11 and 12).

As noted above, NRho adopts a novel α/β fold and displays a strong and specific association with micelles of the zwitterionic lysophospholipid-analog C_{16}PN. C_{16}PN contains the same polar head group as phosphatidylcholine and its micelle is a good mimic of the interface between the head group of phospholipids and bulk-water characteristic of cellular membranes. Therefore, we predict that this domain will bind with a comparable affinity to the cytoplasmic membrane of bacteria. The membrane interaction site in the domain is well defined and shows slow dynamics on the microseconds–milliseconds timescale. The detection of slow dynamics at the membrane interaction site implies the presence of a small population of pre-existing conformers(47)

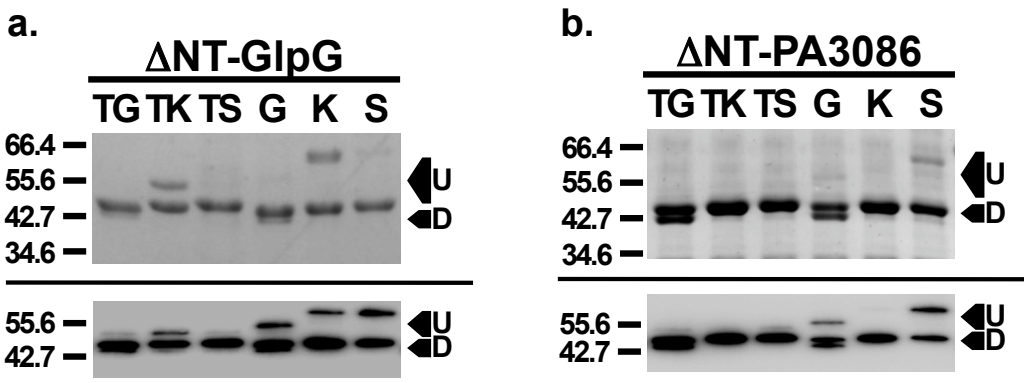

Figure 4.10. Activity of transmembrane regions of GlpG (ΔNT-GlpG) and PA3086 (ΔNT-PA3086). a) 12% SDS-PAGE analysis of coomassie stained digestion products of substrates Gurken-TMD, Keren-TMD, Spitz-TM, Gurken-TMD/CT, Keren-TMD/CT and Spitz-TMD/CT (TG, TK, TS, G, K and S, respectively) were incubated with purified ΔNT-GlpG at 37^{o}C for 8 hr in 0.1% DDM. Below the gel is shown the resulting western blot using anti-MBP as primary antibody. **b).** As in the a) but for the membrane domain of *P. aeruginosa* rhomboid ΔNT-PA3086. Protein bands corresponding to the undigested (**U**) and digested (**D**) substrates are indicated with an arrow on the right and the molecular weight marker positions are shown on the left.

in the native-state ensemble. This population is likely to be better suited to interact with the membrane (high-affinity conformers: possibly with some of the side-chains of the key hydrophobic residues exposed to solvent) than the native structure (low-affinity conformers: the most-populated conformational states in the native-state ensemble that have the hydrophobic side-chains buried as is expected for a soluble protein) in solution. Membrane binding thus causes a reorganization of the populations of the native-state ensemble with possible relaxation towards a bound-state ensemble. This two-step process (selective binding of conformers and relaxation to bound-state) for binding has been proposed for protein–protein interactions by Grunberg *et a*l(48). These structural characteristics are coincidental with the conservancy degree observed as observed in Fig 12 where it is show that the conservancy is higher in the regions where the secondary structures of NRh of *Pseudomonas aeruginosa* are present.

It has been shown that in some cases Ca^{2+} ions can act as an electrostatic trigger in the association of proteins with zwiterionic membranes by neutralizing charged residues and decreasing the desolvation penalty(39). However, NRho displays no observable Ca^{2+} affinity (data not shown) and we find no evidence for this mode of interaction. We therefore propose the following mode of the association of NRho with the bacterial cytoplasmic membrane: NRho interacts with the membrane utilizing an aromatic residue (Trp41) while undergoing a change in its average conformation (transition form the native-state ensemble to the bound-state ensemble) during the process. The presence of the membrane is sufficient to trigger this conformational change.

What could be the role of NRho in the context of the intramembrane protease activity of rhomboids? As mentioned above, Lee *et al*.(11), using in-cell assays as well as *in vivo* experiments have demonstrated that a variant of Rhomboid lacking the N-

terminal domain cleaves substrate with a significantly reduced activity. Further, phylogenetic analysis points towards the NRho having a regulatory role in Rhomboid activity(13), but the precise nature of this regulatory role is not clear. Notably however, Urban *et al.* have demonstrated that substrate specificity in rhomboids is governed by the presence of helix-breaking residues (a Gly-Ala motif) on the luminal side of the substrate TMD(49). Studies *in vivo* revealed that displacement of this motif by two residues up towards the membrane/water interface, or two, five or seven residues further into the hydrophobic core of the membrane abolished substrate cleavage (nevertheless if this motif is displaced one or three residues towards the membrane center the substrate could still be cleaved)(49). Thus, the positioning of the substrate TMD relative to an axis normal to the membrane plane appears to play a significant role in substrate cleavage. It is thus also reasonable to expect that the position of the Rhomboid catalytic site relative to the substrate TMD, i.e. relative to the membrane plane, would be similarly important. A possible mechanism by which NRho could modulate the position of the substrate TMD relative to the Rhomboid active site and *vice versa* can be proposed. NRho could modulate the active site height by affecting the tilt or position of the TMDs in the membrane region of the protein relative to the membrane plane. The observed ability of NRho to strongly interact with membranes could serve as an anchor point from which the necessary TMD tilt and therefore active site height could be derived.

A secondary role for NRho could also be envisaged; it could be involved in substrate binding and function to increase the local concentration of substrate. NRho contains a conspicuously large negatively charged surface (**Figure 2(c)**). Inspection of the sequences of the two primary Rhomboid targets, namely Spitz and Keren reveals that both possess cytoplasmic C-terminal sequences approximately 70 residues in length with a basic amino acid content of approximately 20–25%. In addition, Gurken,

which differs from both Spitz and Keren in its TMD characteristics(10, 50), its role in the fly life-cycle (it appears only during oogenesis) and possibly also in the details of its regulation, but nevertheless has been shown to be a Rhomboid target(50), also contains a cytoplasmic sequence, which is 23 residues long of which nine residues are basic. Thus, the negatively charged surface in NRho may serve to interact with the substrate through electrostatic interactions with their C-terminal soluble components. Note that no effector molecules are required for the proteolytic activity of rhomboids, only the presence of the substrate is sufficient for proteolysis(3). Thus NRho could provide substrate specificity and increase local substrate concentration. Another consequence of the interaction could be the correct positioning of the substrate transmembrane domain relative to the Rhomboid active site. A role for the interactions between the cytoplasmic domains of Rhomboid homologs and their targets would not be unprecedented. In case of the human rhomboid ortholog RHBDL2, the interaction between its N-terminal cytosolic domain and the C-terminal cytosolic domain of its target, thrombomodulin (an anti-coagulant cell surface protein), has been demonstrated to modulate cleavage(18), though the molecular details of this interaction are currently unknown.

These two possible roles of NRho, discussed above, are not mutually exclusive and could occur simultaneously (as one would expect for proper functioning), since the two surfaces involved in membrane interactions and putative electrostatic interactions are well differentiated. Further experiments are needed to elucidate the exact role of NRho in substrate specificity. The lessons learned could have relevant implications for other intramembrane proteases, such as γ-secretase, where substrate recognition and peptide bond selectivity have serious implications in the pathogenesis of Alzheimer disease(51).

Acknowledgements

This work has been supported by the following grants: MCB-0347100 and MCB-0546087 from the National Science Foundation to R.G. and I.U.-B., respectively; 5G12 RR03060 towards support for the NMR facilities at the City College of New York and P41 GM-66354 to the New York Structural Biology Center from the National Institutes of Health. The authors thank Dr Matthew Freeman (Laboratory of Molecular Biology, MRC, Cambridge, UK) for a generous gift of the *P. aeruginosa* cDNA, Celia Torres (City College) for her initial work on the expression of NRho, Dr Andrea Piserchio (New York Structural Biology Center) for extensive help with fluorescence experiments on NRho and Dr David Cowburn (New York Structural Biology Center) for useful discussions and a critical reading of the manuscript.

Bibliography

1. Brown MS, Ye J, Rawson RB, & Goldstein JL (2000) *Cell* **100,** 391-398.

2. Urban S & Freeman M (2002) *Curr Opin Genet Dev* **12,** 512-518.

3. Wolfe MS & Kopan R (2004) *Science* **305,** 1119-1123.

4. Wolfe MS, De Los Angeles J, Miller DD, Xia W, & Selkoe DJ (1999) *Biochemistry* **38,** 11223-11230.

5. De Strooper B, Annaert W, Cupers P, Saftig P, Craessaerts K, Mumm JS, Schroeter EH, Schrijvers V, Wolfe MS, Ray WJ*, et al*. (1999) *Nature* **398,** 518-522.

6. Ni CY, Murphy MP, Golde TE, & Carpenter G (2001) *Science* **294,** 2179-2181.

7. Sisodia SS & St George-Hyslop PH (2002) *Nat Rev Neurosci* **3,** 281-290.

8. Rawson RB, Zelenski NG, Nijhawan D, Ye J, Sakai J, Hasan MT, Chang TY, Brown MS, & Goldstein JL (1997) *Mol Cell* **1,** 47-57.

9. Sturtevant MA, Roark M, & Bier E (1993) *Genes Dev* **7,** 961-973.

10. Freeman DM (2004) *Nature Reviews* **5,** 188-197.

11. Lee JR, Urban S, Garvey CF, & Freeman M (2001) *Cell* **107,** 161-171.

12. Urban S, Lee JR, & Freeman M (2001) *Cell* **107,** 173-182.

13. Koonin EV, Makarova KS, Rogozin IB, Davidovic L, Letellier MC, & Pellegrini L (2003) *Genome Biol* **4,** R19.

14. Rather PN, Ding X, Baca-DeLancey RR, & Siddiqui S (1999) *J Bacteriol* **181,** 7185-7191.

15. Gallio M, Sturgill G, Rather P, & Kylsten P (2002) *Proc Natl Acad Sci U S A* **99,** 12208-12213.

16. Urban S, Schlieper D, & Freeman M (2002) *Curr Biol* **12,** 1507-1512.

17. McQuibban GA, Saurya S, & Freeman M (2003) *Nature* **423,** 537-541.

18. Lohi O, Urban S, & Freeman M (2004) *Curr Biol* **14,** 236-241.

19. Kanaoka MM, Urban S, Freeman M, & Okada K (2005) *FEBS Lett* **579,** 5723-5728.

20. Lemberg MK, Menendez J, Misik A, Garcia M, Koth CM, & Freeman M (2005) *EMBO J* **24,** 464-472.

21. Wang Y, Zhang Y, & Ha Y (2006) *Nature* **444,** 179-180.

22. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, & Bax A (1995) *J Biomol NMR* **6,** 277-293.

23. Johnson BA (2004) *Methods Mol Biol* **278,** 313-352.

24. Sattler M, Schleucherb, J., Griesinger, C (1999) *Progress in Nuclear Magnetic Resonance Spectroscopy* **34,** 93–158.

25. Grzesiek FCaS (1999) *J. Am. Chem. Soc.* **121,** 1601-1602.

26. G.W. Vuister AB (1993) *J. Am. Chem. Soc.* **115,** 7772-7777.

27. Cornilescu G, Delaglio F, & Bax A (1999) *J Biomol NMR* **13,** 289-302.

28. Habeck M, Rieping W, Linge JP, & Nilges M (2004) *Methods Mol Biol* **278,** 379-402.

29. Linge JP, Williams MA, Spronk CA, Bonvin AM, & Nilges M (2003) *Proteins* **50,** 496-506.

30. Koradi R, Billeter M, & Wuthrich K (1996) *J Mol Graph* **14,** 51-55, 29-32.

31. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, & Thornton JM (1996) *J Biomol NMR* **8,** 477-486.

32. Palmer AG, 3rd (2001) *Annu Rev Biophys Biomol Struct* **30,** 129-155.

33. P.T. Boggs JRD, R.H. Byrd and R.B. Schnabel (1989) *ACM Trans. Math. Software* **15,** 348-364.

34. Ghose R, Fushman D, & Cowburn D (2001) *J Magn Reson* **149,** 204-217.

35. Szabo GLaA (1982) *J. Am. Chem. Soc.* **104,** 4546–4559.

36. G.M. Clore AS, A. Bax, L.E. Kay, P.C. Driscoll and A.M. Gronenborn (1990) *J. Am. Chem. Soc.* **112,** 4936–4989.

37. Fushman D, Cahill S, & Cowburn D (1997) *J Mol Biol* **266,** 173-194.

38. Del Rio A, Anand A, & Ghose R (2006) *J Magn Reson* **180,** 1-17.

39. Cho W & Stahelin RV (2005) *Annu Rev Biophys Biomol Struct* **34,** 119-151.

40. Luo P & Baldwin RL (1997) *Biochemistry* **36,** 8413-8421.

41. Sreerama N & Woody RW (2004) *Methods Enzymol* **383,** 318-351.

42. Sonnhammer EL, von Heijne G, & Krogh A (1998) *Proc Int Conf Intell Syst Mol Biol* **6,** 175-182.

43. Holm L & Sander C (1993) *J Mol Biol* **233,** 123-138.

44. Yip CK, Kimbrough TG, Felise HB, Vuckovic M, Thomas NA, Pfuetzner RA, Frey EA, Finlay BB, Miller SI, & Strynadka NC (2005) *Nature* **435,** 702-707.

45. Mota LJ, Sorg I, & Cornelis GR (2005) *FEMS Microbiol Lett* **252,** 1-10.

46. Dutta K, Alexandrov A, Huang H, & Pascal SM (2001) *Protein Sci* **10,** 2531-2540.

47. Volkman BF, Lipson D, Wemmer DE, & Kern D (2001) *Science* **291,** 2429-2433.

48. Grunberg R, Leckner J, & Nilges M (2004) *Structure* **12,** 2125-2136.

49. Urban S & Freeman M (2003) *Mol Cell* **11,** 1425-1434.

50. Urban S, Lee JR, & Freeman M (2002) *EMBO J* **21,** 4277-4286.

51. Hardy J & Selkoe DJ (2002) *Science* **297,** 353-356.

Chapter 5

Introduction

Regulated intramembrane proteolysis is an ancient mechanism to control cell metabolism, differentiation and development in organisms ranging from bacteria to humans (1-3). In intramembrane proteolysis single-pass membrane proteins can be cleaved within their transmembrane domain (TMD) to yield soluble fragments that can act as molecular effectors. Examples include the release of transcriptional activators in the Notch (4) and ErbB-4 (5) cascades; the processing of the β-amyloid precursor protein (APP) to release the neuropathogenic β-amyloid peptides (Aβs) (6-9); and the liberation of cellular growth factors (10, 11). The proteases responsible for this intramembrane cleavage (known as intramembrane-cleaving proteases; i-CLiPs) constitute a prominent class of integral membrane proteins. In analogy to water-soluble proteases, i-CLiPs can be divided into aspartic proteases, metalloproteases and serine proteases (12, 13). In this thesis we have focused on rhomboids, which are serine intramembrane proteases that play a pivotal role in the development of *Drosophila*, in the integrity of mitochondria, and in quorum sensing in bacteria. Specifically, our research has addressed the following questions. (1) Do rhomboids display substrate specificity? (2) How is this specificity governed?

In this chapter we will summarize the results and insight obtained in this thesis, with emphasis in discussing our data in the context of other intramembrane proteases. In the last section of the chapter, we will also frame remaining questions that need to be addressed to better understand the mechanism of rhomboid mediated intramembrane proteolysis.

Proteolytic profiles of intramembrane proteases

The substrate specificity and peptide bond selectivity in intramembrane proteases is not well understood. For example, one of the most intriguing characteristics of presenilin, the aspartyl intramembrane protease, in the γ-secretase complex is its ability to process more than 50 unique type I transmembrane proteins (14). These include the single-pass integral membrane proteins APP, APP-like proteins, E-Cadherin, CD44, lipoprotein receptor related protein, Notch, interferon response element and activated transcription factor 6 (for reviews see (15, 16)). These substrates appear not to share any consensus sequence around their transmembrane cleavage sites and the main prerequisite for cleavage appears to be prior proteolytic removal of an ectodomain (17). In addition, presenilin can process its substrates at multiple sites, a property that can have serious implications in the pathogenesis of Alzheimer's disease. Indeed, a major feature of both sporadic and familial forms of Alzheimer's disease is the accumulation of senile plaques and neurofibrillary tangles in specific brain areas (18). Senile plaques are composed of dystrophic neurites surrounding extracellular aggregates of Aβs. Sequential degradation of APPs produces these peptides (19): first, β-secretase sheds the ectodomain of APP to yield a C-terminal fragment (β-CTF) containing the Aβ domain, and subsequently, γ-secretase catalyzes the intramembrane cleavage of β-CTF to liberate Aβs (9). Indeed, γ-secretase can cleave the TMD of β-CTF at multiple sites (referred to as γ-, ε- and ζ-site). The ε cleavage closest to the cytoplasmic border of the TMD releases the intracellular domain of β-CTF into the cytosol (20). The remaining membrane-anchored fragment undergoes an intermediate scission ~3 residues N-terminal to the ε-site at the so-called ζ-site (21). Thereafter, Aβ is released by final cuts at the γ-site. The γ-cleavage is variable and can occur after Aβ amino acid residues 38, 40 or 42. The balance between these γ-cleavages can have an

important role on the pathogenicity of Aβs. As the longer peptide Aβ42 - relative to the Aβ40 species - is prone to form aggregates, an increase in the Aβ42/40 ratio appears to be central to the pathogenesis of early-onset familial Alzheimer's disease (22). Indeed, in normal individuals, APP processing leads to low Aβ42/40 ratios. In contrast, individuals with mutations in presenilins or in APP can have elevated Aβ42/40 ratios(23-31).

Our experiments with designed substrate chimeras in which the transmembrane and cytoplasmic regions of the physiological substrates Gurken, Keren and Spitz were preserved while their EGFR ligand ectodomain was replaced by maltose binding protein, demonstrated that rhomboids cleave their substrates at a single site. The four rhomboids tested rhomboid-1, *H. sapiens* RHBDL-2, *P. aeruginosa* PA3086 and *E. coli* GlpG cleaved after Ala138 for the Spitz chimeras, after Ala122 for the Keren chimeras, and after Ala245 for the Gurken chimeras, suggesting a conservation of proteolytic profiles among prokaryotic and eukaryotic rhomboids. Thus the proteolytic profile displayed by rhomboids with the existence of a single cleavage site is in marked contrast to γ-cleavage, which as stated above displays a clear cleavage site promiscuity. In analogy to rhomboids, SPP-catalyzed intramembrane proteolysis seems to occur predominantly at one position after a helix-breaking amino acid residue (32).

Substrate specificity in intramembrane proteases

As we mentioned above, the substrate specificity in intramembrane proteases is not well understood. The aspartyl intramembrane protease presenilin can process more than 50 unique type I transmembrane proteins (14), without these substrates sharing any apparent consensus sequence around their transmembrane cleavage sites. Having said that, there are examples of single point mutations abolishing γ-secretase cleavage,

suggesting that the enzyme requires a certain sequence motif/characteristics for cleavage. Indeed, for example single point mutations of the ErbB4 receptor tyrosine kinase (V673I) and the Notch-1 receptor (V1744G) are known to abrogate proteolysis (33). In both these cases, the mutated valine residue is located immediately adjacent to the site of cleavage. However, mutation of a similarly positioned conserved valine in other γ-secretase substrates, including APP (34), CD43 (35), and DCC (36) does not affect proteolysis. Thus, the relevance of primary structure in substrate specificity of intramembrane proteases is unclear.

In our substrate specificity studies described in chapter 3 we found that rhomboid orthologs from different species display P1, P1' and P2' specificity against substrates derived from Spitz, Keren and Gurken. Recently, Strisovsky et al (37) found the positions that are important in the selection of the cleavage site by rhomboids to be P4, P1 and P2'. Strisovsky et al. arrived to this conclusion by analyzing the effect on AarA activity of different mutants of the substrate TatA. In this case P4 was occupied by isoleucine, P1 by alanine and P2' by phenylalanine. We didn't find any change in activity and/or a shift in the cleavage site when we mutated this P4 position to alanine. Regarding the position P2' they did not observe an increasing in AarA activity with any of their mutations. Arguably, the differences observed might be due to the different systems employed by the two laboratories. We note however, that Strisovsky et al found AarA to cleave Spitz and Gurken at the same position as GlpG, PA3086, Rho-1 and RHBDL2.

More specifically, we found a preference for an amino acid with a small side chain at P1 position. Sequence alignment of known rhomboid substrates reveals the amino acid alanine to be highly conserved at this P1 site and small side chain residues to be conserved at the P1' site (**Fig. 5.1**). These substrates include: (1) rho-1 substrates Spitz, Keren and Gurken; (2) AarA substrate TatA; (3) pfROM4 substrate EBA-175

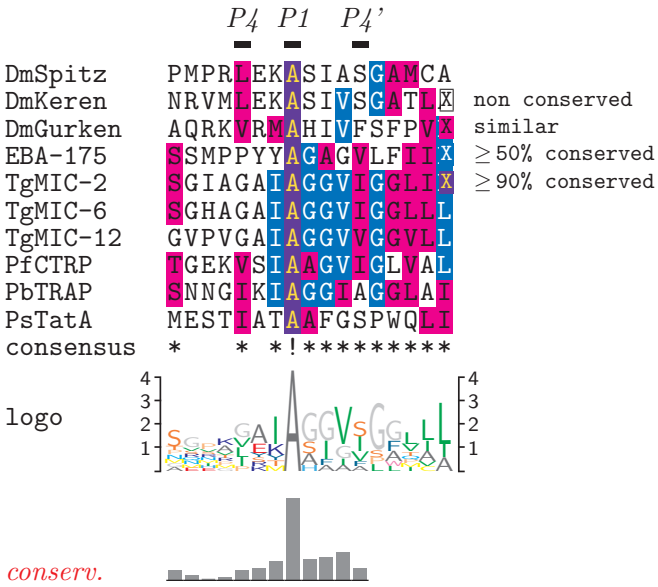

Figure 5.1. Alignment of the sequences of different rhomboids cleavage region of different rhomboid substrates. Region of the cleavage site of different rhomboid substrates is shown. Position P4, P1 and P4" are indicated in the top of the alignment. (Alignments were produced using TEXshade (Eric Beitz (2000), TEXshade: shading and labeling multiple sequence alignments using LATEX2ε. Bioinformatics: 16, 135–139)

substrate and (4) microneme protein protease-1 substrates Mic-2, Mic-6 and Mic-12. Thus, the question of whether rhomboids from all species recognize a small residue at the P1 site arises. On the one hand we note that the rhomboids tested in this thesis retain the substrate cleavage site when alanine at the P1 site is mutated to a conservative residue like glycine or serine, whereas mutation to bulky amino acid residues such as isoleucine and phenylalanine alter the proteolytic profile. On the other hand, it is known that rhomboid orthologs such as *S. cerevisae* Rbd1 cleave citochrome c peroxidase (Ccp-1) and Mgm1 after a relatively bulky threonine **(38)**. Remarkably, SPP intramembrane proteases also show specificity for small side chain residues (alanine and serine) in the P1 position (**Fig. 5.2a**). SPP recognizes the signal peptide present in preproteins that are going to be secreted such as like bovine prolactin (32, 39) and the MHC class I preproteins (39, 40). SPP also plays a role in the infection of pestiviruses by processing the C terminus of the core protein of classic swine fever virus and hepatitis C virus (41). Finally a SPP-like class of protease, SPPL2b, processes tumor necrosis factor alpha (42). We note that the sequence alignment of the transmembrane region of these SPP substrates shows (**Fig. 5.2b**) the overall conservation to be very poor. However it is clear that side from a preference of small side chain residues in the P1 position, SPP displays specificity for a bulky residue at the P1' site. The conservation at key substrate sites observed with rhomboid and SPP substrates is not unique. Indeed The substrates of another important intramembrane protease, the site-2 metalloprotease, also show a high level of conservation in the region around the cleavage site (**Fig. 5.3**). In the position P4 arginine appears to be conserved for the human substrates SREBP-1, SREBP-2, ATF6, Luman, BDNF, CREB-H and CREB4 (43). In addition, leucine in position P1 is also conserved l (44). In contrast to SPP, we found that small side chain at P1' position and a bulky side chain at P2' position to be important for rhomboid specificity.

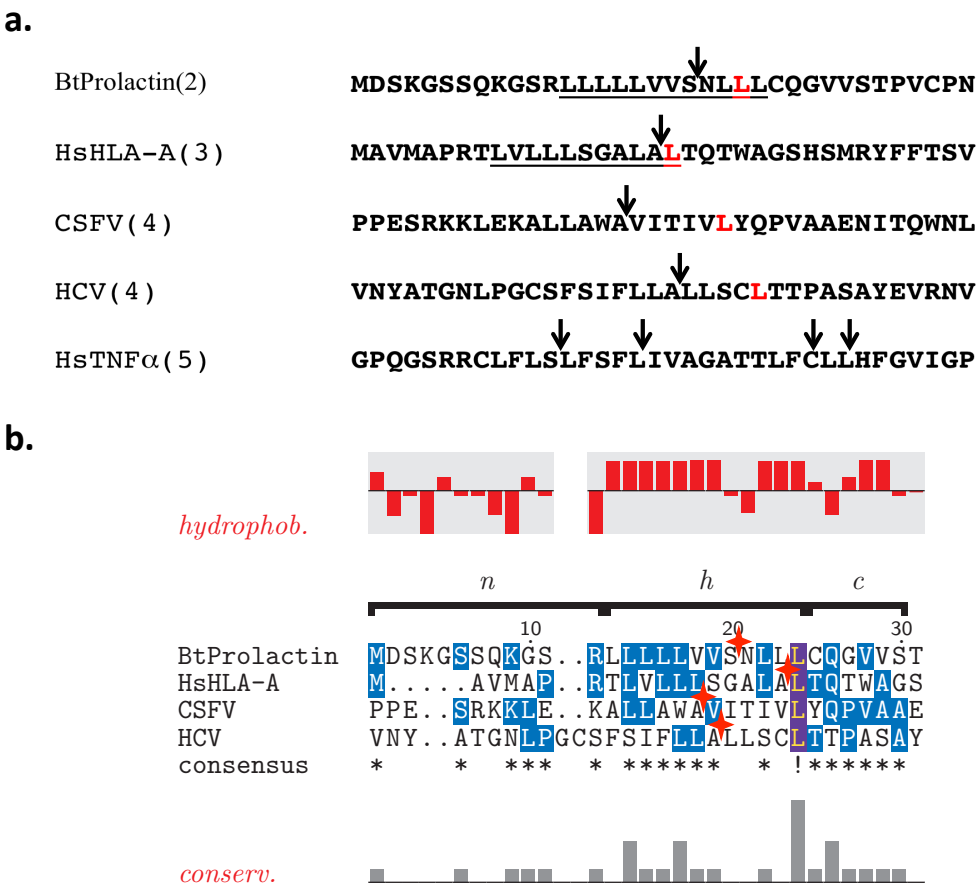

Figure 5.2. Cleavage site and alignment of signal peptide peptidase. a). Sequence of the signal peptides of bovine Prolactin, human HLA, classic swine fever virus, hepatitis c virus and human tumor necrosis factor are shown. Cleavage site are indicated with black arrows. **b).** Alignment of the signal peptides showing the three regions that are important for the signal peptide peptidase. (Alignments were produced using TEXshade (Eric Beitz (2000), TEXshade: shading and labeling multiple sequence alignments using LATEX2ε. Bioinformatics: 16, 135–139)

In the case of the P1' site, the sequence alignment suggests that residues with small side chain such as serine, glycine and alanine - with the exception of histidine in the position P1' of Gurken – are preferred (**Fig. 5.1**). A possibility to explain the presence of histidine in Gurken is the helix-breaking capacity of this residue (see next section). In our mutagenesis studies we found that the activity of rhomboid is strongly inhibited when a helix stabilizing residue, like isoleucine, is introduced at the P1' site of Spitz (Chapter 3). We found that the position P2' has a preference for isoleucine, however this isoleucine at the P2' site is not well conserved among other rhomboid substrates (**Fig. 5.1**). Thus the role of this amino acid residue is controversial. In our hands it appears this P2' site is key in the selection of the primary, over the secondary, cleavage site in Spitz. Finally, the P2 position appeared to accommodate a variety of amino acid side chains without a significant perturbation of the cleavage profile.

In summary, we have found that rhomboids are capable of displaying substrate specificity by recognizing conserved positions along the substrate. The arising question then is, what is the mechanism by which these positions affect specificity.

Substrate helicity and rhomboid activity

The intramembrane cleavage catalyzed by rhomboids poses a problem: how does the enzyme gain access to the scissile peptide bond that is otherwise protected, due to its predicted α-helical conformation, by main-chain H-bonds. One conceptual solution to this problem posits that the transmembrane domain of the substrate has inherent conformational instability and thus it populates conformations that deviate from canonical α-helicity that can distort main chain H-bonding (45).

The role of the substrate helicity in the cleavage mechanism by intramembrane proteases is still under discussion. The substrate in SPPs has three well-defined regions: (1) the n region that is hydrophilic and contains positively charged residues, (2) the h region that is hydrophobic, and (3) the c region containing generally small and polar residues (46, 47) (**Fig. 5.2b**). The presence of helix breaking residues in the region h is believed to be important for SPP activity (48, 49). In contrast, for the site-2 peptidases helicity does not appear to be key (49-51). For rhomboids we found the role of helicity to be unclear. The inhibitory effect observed for the Spitz mutant S139I suggest that a helix-breaking residue is necessary in the position P1'. When the effect of the different mutants on the activity of rhomboids we observed this activity in inhibited by the presence of residues with a hydrophobic and larger size side chain like isoleucine and phenylalanine (chapter 3).

Residue S139 was mutated into F, G, T and I. A brief insight in this position leads to the conclusion that a small side chain residue is needed in this position. In the case of wild type spitz, a serine residue in present in P1' and if this position is mutated to phenylalanine an inhibition occurs. When the helicity and hydrophobicity is analyzed, we found that there is a tendency to reduce the activity of rhomboids GlpG and PA3086 when the helicity and hydrophobicity is increased (**Fig 3.10**). Strisovsky (37) found that in the position P1' the introduction of helix-stabilizing residues inhibits the activity by AarA and, when residues that have small side chains or, residues that are helix destabilizing are introduced in this position, the activity of rhomboid AarA is not affected. A brief insight in this position leads to the conclusion that a small side chain and helix-destabilizing residue is needed.

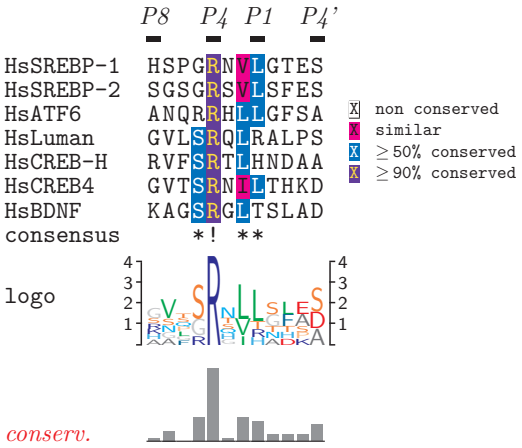

Figure 5.3. Alignment of the sequences of different S2P cleavage region. Regions of the cleavage site of different S2P substrates are shown. Position P8, P4, P1 and P4" are indicated in the top of the alignment. (Alignments were produced using TEXshade (Eric Beitz (2000), TEXshade: shading and labeling multiple sequence alignments using LATEX2ε. Bioinformatics: 16, 135–139).

As a summary we propose a substrate motif for rhomboids taking into consideration all data presented here. This substrate motif has alanine in the position P1, a small size amino acid in the position P1' and isoleucine or phenylalanine in the position P2'. We think also that the position P2 should contain a helix-breaking residue like lysine (**Fig. 5.4**).

Future work

Insight into rhomboid kinetics.

Three Michaelis–Menten parameters describe a typical enzyme reaction: kcat, Km and kcat/Km. The kcat value is the turnover number and it measures the amount of product formed per enzyme molecule when all of the enzyme has bound substrate. Since some selection may occur during substrate binding, kcat is an inadequate measure of substrate specificity. Km is also a complex kinetic constant and an apparent dissociation constant of all enzyme-bound substrate complexes. Km is also an inadequate measure of specificity, since it does not take into account the rate of substrate turnover. In contrast, kcat/Km provides an accurate measure of specificity. This parameter is an apparent second-order rate constant for the reaction of the free enzyme and the free substrate. Since different substrates will compete for the free enzyme, a comparison of the values of kcat/Km would describe specificity. Not surprisingly, kcat/Km is also known as the specificity constant. Obtaining the kcat/Km for rhomboid against selected substrate mutants will provide much needed detailed insight into rhomboid specificity. For such experiments a rapid and sensitive activity assay is a requisite. Substrates that incorporate Fluorescence Resonance Energy Transfer (FRET) based technologies have been used successfully to study protease biochemistry (52-55). We propose to design small peptide substrates containing the minimal substrate

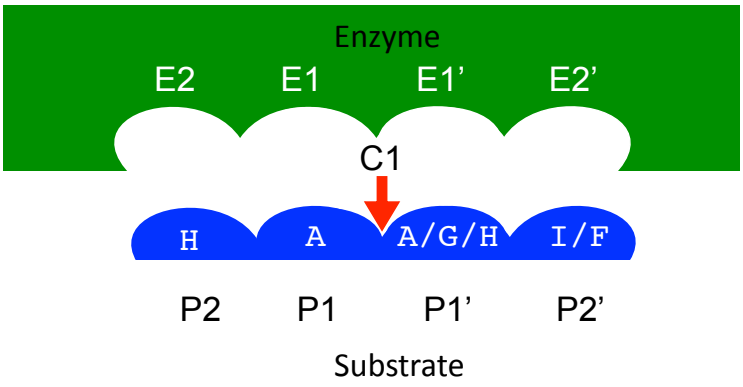

Figure 5.4. Proposed motif for rhomboid. The proposed substrate motif for rhomboid intramembrane protease. In green is represented the protease and in blue the substrate. E2, E1, E1' and E2' are the binding pockets in the enzyme that recognize the positions P2, P1, P1' and P2' in the substrate. C1 and the red arrow indicate the position of the peptide bond is cleaved by the enzyme after the recognition of the substrate motif. The proposed motif are represented in white letters.

motif, which are internally quenched by FRET. Upon rhomboid activity the fluorophores quenching will be alleviated and thus a large increase in the fluorescence signal will immediately appear.

Explore for protocols and models to obtain higher expression level of eukaryotic rhomboid.

One of the more difficult experimental objectives is the enzymatic study of rhomboid is the expression of eukaryotic rhomboid in enough quantity for the determination of activity and also the determination of the cleavage site of these rhomboids. In our project we tried to express eukaryotic rhomboid: Rho-1 from *D. melanogaster* and RHBDL-2 from *H. sapiens*, however as we discussed earlier the expression level was so poor that the enzyme was only detectable by western blot after Ni-bead purification. Nevertheless these enzyme preparations were active enough to detect using the chimaeras substrates and determine the cleavage site. However this work was very difficult and time consuming. One alternative to approach the solution of this task is to try with different in vivo tests like Lei and Li, 2009 (56) who demonstrate that human rhomboid RHBDL-2 has to be processed by proteolysis in order to gain activity in transfected cells and even a RHBDL-2ase was proposed as a regulator protease. These results can explain why our full length RHBDL-2 preparation had a very low activity and expression level although. The same reason can be used as an argument that account for the low activity of our Rho-1 preparation. It has to be emphasized that this experimental problem is a major obstacle for a proper in vitro biochemical studies that can confirm the biochemical characterization using purified components of the rhomboid activity. An experimental opportunity to prevent this problem is to use other models for the over expression of eukaryotic rhomboids. In this

aspect the heterologous expression of membrane proteins using yeast cell as *Picchia pastoris* is an important option in the expression of eukaryotic proteins (57, 58).

PARL and Parkinson disease

PARL is a rhomboid protein that is localized in the mitochondrial inner membrane of mouse (59, 60) and its substrate is another membrane protein: OPA-1 (Optic atrophy-1). The digestion of OPA-1 is involved in the regulation of apoptosis (61). *D. melanogaster* and *S. cerevizae* orthologs of PARL (Rhomboid-7 and Mgm1) are also localized in the mitochondrial inner membrane and they play similar roles in the physiology of mitochondria. In Parkinson's disease PARL protein plays an important role because it is in charge of the proteolysis of Pink1 (PTEN- induced putative kinase) (62, 63). The mutations of Pink-1 orthologs in D. melanogaster show symptoms that are similar to the mutations in Rhomboid-7 proteins that are also mitochondrial protein (64, 65). Finally, the role of PARL is proposed to regulate the traffic of Pink-1 protein to the inter membrane space. This traffic is part f the regulation of the mitophagy, process which leads to the degradation of the mithocondria before the apoptosis takes place. When this process is impaired it can be the etiology of Parkinson's disease (66, 67). The study of this pathway is then very important for future development of therapies of this terrible disease. The availability of the sequence of the substrates of PARL: Pink1 and OPA1, empowers the possibility to develop artificial substrates in order to characterize enigmatically this important enzyme. Furthermore, it is possible to study a great variety of human PARL orthologs in different organism. Efforts should be making in order to over-express and purified PARL proteins in order to obtain the structure of this complex molecule.

Structural studies.

From the structure of GlpG, a gating mechanism that permits the entrance of water molecules and also interacts with the substrate was proposed. The main component of this gating mechanism is the transmembrane domain V (68, 69). The blending of the transmembrane domain permits three conformations of this gating system: closed, semi-opened and opened (68). This is a mechanism that is common to presenilin, S2P and signal peptide peptidase (70-72). The introduction of helix-breaking residues in the TMD V increases the activity of GlpG dramatically. This is in accordance with the theory that the blend of the TMD V is necessary for the accommodation of the TMD of the substrate in the active site of rhomboid. This characteristic indicates that the recognition of the TMD of the substrates occurs during a possible interaction between the TMD of the substrate and the TMD V of GlpG. This mechanism is confirmed in in vivo studies (73). There are multiple mutagenesis studies in GlpG that are reviewed by Urban, S. (74). These studies can be extended in two different levels. The first is the confirmation of the role of the TMD V in other rhomboids that have a known activity with the substrates we study: PA3086 from *P. aeruginosa*, AarR from *P. stuartii*, hiGlpG from *H. influenza*. The other experimental approach is the analysis of complex GlpG-Substrate or PA3086-Susbtrate. These experiments can be performed using the substrates (wt and mutants with reduced activity) developed in our lab (with or without the enzymatic elimination of the N-terminal domain) and the mutants GlpG-N154A that lost activity with gurken. Native gels can be used to study the formation of these complexes with the advantage that these techniques are set in our lab.

Bibliography

1. Brown MS, Ye J, Rawson RB, & Goldstein JL (2000) *Cell* **100,** 391-398.

2. Heldin CH & Ericsson J (2001) *Science* **294,** 2111-2113.

3. Urban S & Freeman M (2002) *Curr Opin Genet Dev* **12,** 512-518.

4. De Strooper B, Annaert W, Cupers P, Saftig P, Craessaerts K, Mumm JS, Schroeter EH, Schrijvers V, Wolfe MS, Ray WJ*, et al.* (1999) *Nature* **398,** 518-522.

5. Ni CY, Murphy MP, Golde TE, & Carpenter G (2001) *Science* **294,** 2179-2181.

6. Sherrington R, Rogaev EI, Liang Y, Rogaeva EA, Levesque G, Ikeda M, Chi H, Lin C, Li G, Holman K*, et al.* (1995) *Nature* **375,** 754-760.

7. Levy-Lahad E, Wijsman EM, Nemens E, Anderson L, Goddard KA, Weber JL, Bird TD, & Schellenberg GD (1995) *Science* **269,** 970-973.

8. Rogaev EI, Sherrington R, Rogaeva EA, Levesque G, Ikeda M, Liang Y, Chi H, Lin C, Holman K, Tsuda T*, et al.* (1995) *Nature* **376,** 775-778.

9. Sisodia SS & St George-Hyslop PH (2002) *Nat Rev Neurosci* **3,** 281-290.

10. Lee JR, Urban S, Garvey CF, & Freeman M (2001) *Cell* **107,** 161-171.

11. Urban S, Lee JR, & Freeman M (2001) *Cell* **107,** 173-182.

12. Wolfe MS & Selkoe DJ (2002) *Science* **296,** 2156-2157.

13. Weihofen A & Martoglio B (2003) *Trends Cell Biol* **13,** 71-78.

14. Beel AJ & Sanders CR (2008) *Cell Mol Life Sci* **65,** 1311-1334.

15. Marambaud P & Robakis NK (2005) *Genes Brain Behav* **4,** 134-146.

16. Wolfe MS & Kopan R (2004) *Science* **305,** 1119-1123.

17. Shah S, Lee SF, Tabuchi K, Hao YH, Yu C, LaPlant Q, Ball H, Dann CE, 3rd, Sudhof T, & Yu G (2005) *Cell* **122,** 435-447.

18. Price DL, Sisodia SS, & Borchelt DR (1998) *Science* **282,** 1079-1083.

19. Selkoe DJ (2001) *Physiol Rev* **81,** 741-766.

20. Weidemann A, Eggert S, Reinhard FB, Vogel M, Paliga K, Baier G, Masters CL, Beyreuther K, & Evin G (2002) *Biochemistry* **41,** 2825-2835.

21. Zhao G, Cui MZ, Mao G, Dong Y, Tan J, Sun L, & Xu X (2005) *J Biol Chem* **280,** 37689-37697.

22. Price DL & Sisodia SS (1998) *Annu Rev Neurosci* **21,** 479-505.

23. Siman R, Reaume AG, Savage MJ, Trusko S, Lin YG, Scott RW, & Flood DG (2000) *J Neurosci* **20,** 8717-8726.

24. Moehlmann T, Winkler E, Xia X, Edbauer D, Murrell J, Capell A, Kaether C, Zheng H, Ghetti B, Haass C*, et al.* (2002) *Proc Natl Acad Sci U S A* **99,** 8025-8030.

25. Schroeter EH, Ilagan MX, Brunkan AL, Hecimovic S, Li YM, Xu M, Lewis HD, Saxena MT, De Strooper B, Coonrod A*, et al.* (2003) *Proc Natl Acad Sci U S A* **100,** 13075-13080.

26. Qi Y, Morishima-Kawashima M, Sato T, Mitsumori R, & Ihara Y (2003) *Biochemistry* **42,** 1042-1052.

27. Walker ES, Martinez M, Brunkan AL, & Goate A (2005) *J Neurochem* **92,** 294-301.

28. Kumar-Singh S, Theuns J, Van Broeck B, Pirici D, Vennekens K, Corsmit E, Cruts M, Dermaut B, Wang R, & Van Broeckhoven C (2006) *Hum Mutat* **27,** 686-695.

29. Bentahir M, Nyabi O, Verhamme J, Tolia A, Horre K, Wiltfang J, Esselmann H, & De Strooper B (2006) *J Neurochem* **96,** 732-742.

30. Shen J & Kelleher RJ, 3rd (2007) *Proc Natl Acad Sci U S A* **104,** 403-409.

31. Kaneko H, Kakita A, Kasuga K, Nozaki H, Ishikawa A, Miyashita A, Kuwano R, Ito G, Iwatsubo T, Takahashi H*, et al.* (2007) *J Neurosci* **27,** 13092-13097.

32. Sato T, Nyborg AC, Iwata N, Diehl TS, Saido TC, Golde TE, & Wolfe MS (2006) *Biochemistry* **45,** 8649-8656.

33. Vidal GA, Naresh A, Marrero L, & Jones FE (2005) *J Biol Chem* **280,** 19777-19783.

34. Lichtenthaler SF, Wang R, Grimm H, Uljon SN, Masters CL, & Beyreuther K (1999) *Proc Natl Acad Sci U S A* **96,** 3053-3058.

35. Andersson CX, Fernandez-Rodriguez J, Laos S, Baeckstrom D, Haass C, & Hansson GC (2005) *Biochem J* **387,** 377-384.

36. Taniguchi Y, Kim SH, & Sisodia SS (2003) *J Biol Chem* **278,** 30425-30428.

37. Strisovsky K, Sharpe HJ, & Freeman M (2009) *Mol Cell* **36,** 1048-1059.

38. Herlan M, Vogel F, Bornhovd C, Neupert W, & Reichert AS (2003) *J Biol Chem* **278,** 27781-27788.

39. Lemberg MK, Bland FA, Weihofen A, Braud VM, & Martoglio B (2001) *J Immunol* **167,** 6441-6446.

40. Dev KK, Chatterjee S, Osinde M, Stauffer D, Morgan H, Kobialko M, Dengler U, Rueeger H, Martoglio B, & Rovelli G (2006) *Eur J Pharmacol* **540,** 10-17.

41. Heimann M, Roman-Sosa G, Martoglio B, Thiel HJ, & Rumenapf T (2006) *J Virol* **80,** 1915-1921.

42. Fluhrer R, Grammer G, Israel L, Condron MM, Haffner C, Friedmann E, Bohland C, Imhof A, Martoglio B, Teplow DB*, et al.* (2006) *Nat Cell Biol* **8,** 894-896.

43. Seidah NG, Khatib AM, & Prat A (2006) *Biol Chem* **387,** 871-877.

44. Yang T, Espenshade PJ, Wright ME, Yabe D, Gong Y, Aebersold R, Goldstein JL, & Brown MS (2002) *Cell* **110,** 489-500.

45. Ubarretxena-Belandia I & Engelman DM (2001) *Curr Opin Struct Biol* **11,** 370-376.

46. Haeuptle MT, Flint N, Gough NM, & Dobberstein B (1989) *J Cell Biol* **108,** 1227-1236.

47. von Heijne G (1985) *J Mol Biol* **184,** 99-105.

48. Feng L, Yan H, Wu Z, Yan N, Wang Z, Jeffrey PD, & Shi Y (2007) *Science* **318,** 1608-1612.

49. Zelenski NG, Rawson RB, Brown MS, & Goldstein JL (1999) *J Biol Chem* **274,** 21973-21980.

50. Kanehara K, Akiyama Y, & Ito K (2001) *Gene* **281,** 71-79.

51. Brown MS & Goldstein JL (1999) *Proc Natl Acad Sci U S A* **96,** 11041-11048.

52. Kolygo K, Ranjan N, Kress W, Striebel F, Hollenstein K, Neelsen K, Steiner M, Summer H, & Weber-Ban E (2009) *J Struct Biol* **168,** 267-277.

53. Nyborg AC, Herl L, Berezovska O, Thomas AV, Ladd TB, Jansen K, Hyman BT, & Golde TE (2006) *Mol Neurodegener* **1,** 16.

54. Schaal R, Kupfahl C, Buchheidt D, Neumaier M, & Findeisen P (2007) *J Microbiol Methods* **71,** 93-100.

55. Swanson R, Raghavendra MP, Zhang W, Froelich C, Gettins PG, & Olson ST (2007) *J Biol Chem* **282,** 2305-2313.

56. Lei X & Li YM (2009) *J Mol Biol* **394,** 815-825.

57. Karlsson M, Fotiadis D, Sjovall S, Johansson I, Hedfalk K, Engel A, & Kjellbom P (2003) *FEBS Lett* **537,** 68-72.

58. Kukulski W, Schenk AD, Johanson U, Braun T, de Groot BL, Fotiadis D, Kjellbom P, & Engel A (2005) *J Mol Biol* **350,** 611-616.

59. Cipolat S, Rudka T, Hartmann D, Costa V, Serneels L, Craessaerts K, Metzger K, Frezza C, Annaert W, D'Adamio L*, et al.* (2006) *Cell* **126,** 163-175.

60. Frezza C, Cipolat S, Martins de Brito O, Micaroni M, Beznoussenko GV, Rudka T, Bartoli D, Polishuck RS, Danial NN, De Strooper B*, et al.* (2006) *Cell* **126,** 177-189.

61. Pellegrini L & Scorrano L (2007) *Cell Death Differ* **14,** 1275-1284.

62. Meissner C, Lorenz H, Weihofen A, Selkoe DJ, & Lemberg MK *J Neurochem* **117,** 856-867.

63. Shi G, Lee JR, Grimes DA, Racacho L, Ye D, Yang H, Ross OA, Farrer M, McQuibban GA, & Bulman DE *Hum Mol Genet* **20,** 1966-1974.

64. Whitworth AJ, Lee JR, Ho VM, Flick R, Chowdhury R, & McQuibban GA (2008) *Dis Model Mech* **1,** 168-174.

65. McQuibban GA, Lee JR, Zheng L, Juusola M, & Freeman M (2006) *Curr Biol* **16,** 982-989.

66. Deas E, Plun-Favreau H, Gandhi S, Desmond H, Kjaer S, Loh SH, Renton AE, Harvey RJ, Whitworth AJ, Martins LM*, et al. Hum Mol Genet* **20,** 867-879.

67. Jin SM, Lazarou M, Wang C, Kane LA, Narendra DP, & Youle RJ *J Cell Biol* **191,** 933-942.

68. Wu Z, Yan N, Feng L, Oberstein A, Yan H, Baker RP, Gu L, Jeffrey PD, Urban S, & Shi Y (2006) *Nat Struct Mol Biol* **13,** 1084-1091.

69. Wang Y, Zhang Y, & Ha Y (2006) *Nature* **444,** 179-180.

70. Hayrapetyan V, Rybalchenko V, Rybalchenko N, & Koulen P (2008) *Cell Calcium* **44,** 507-518.

71. Rybalchenko V, Hwang SY, Rybalchenko N, & Koulen P (2008) *Int J Biochem Cell Biol* **40,** 84-97.

72. Fraering PC (2007) *Curr Genomics* **8,** 531-549.

73. Urban S & Baker RP (2008) *Biol Chem* **389,** 1107-1115.

74. Urban S (2010) *Biochem J* **425,** 501-512.

Bibliography.

Chapter 1

1. Sik A, Passer BJ, Koonin EV, & Pellegrini L (2004) *J Biol Chem* **279,** 15323-15329.

2. Martys-Zage JL, Kim SH, Berechid B, Bingham SJ, Chu S, Sklar J, Nye J, & Sisodia SS (2000) *J Mol Neurosci* **15,** 189-204.

3. Brown MS, Ye J, Rawson RB, & Goldstein JL (2000) *Cell* **100,** 391-398.

4. Kanaoka MM, Urban S, Freeman M, & Okada K (2005) *FEBS Lett* **579,** 5723-5728.

5. Li T, Wen H, Brayton C, Das P, Smithson LA, Fauq A, Fan X, Crain BJ, Price DL, Golde TE*, et al.* (2007) *J Biol Chem* **282,** 32264-32273.

6. Kim YS, Kim SG, Park JE, Park HY, Lim MH, Chua NH, & Park CM (2006) *Plant Cell* **18,** 3132-3144.

7. Cipolat S, Rudka T, Hartmann D, Costa V, Serneels L, Craessaerts K, Metzger K, Frezza C, Annaert W, D'Adamio L*, et al.* (2006) *Cell* **126,** 163-175.

8. Gallio M, Sturgill G, Rather P, & Kylsten P (2002) *Proc Natl Acad Sci U S A* **99,** 12208-12213.

9. Urban S (2006) *Genes Dev* **20,** 3054-3068.

10. Koide K, Maegawa S, Ito K, & Akiyama Y (2007) *J Biol Chem* **282,** 4553-4560.

11. Stevenson LG, Strisovsky K, Clemmer KM, Bhatt S, Freeman M, & Rather PN (2007) *Proc Natl Acad Sci U S A* **104,** 1003-1008.

12. Urban S, Lee JR, & Freeman M (2002) *EMBO J* **21,** 4277-4286.

13. De Strooper B, Annaert W, Cupers P, Saftig P, Craessaerts K, Mumm JS, Schroeter EH, Schrijvers V, Wolfe MS, Ray WJ*, et al.* (1999) *Nature* **398,** 518-522.

14. Ni CY, Murphy MP, Golde TE, & Carpenter G (2001) *Science* **294,** 2179-2181.

15. Urban S, Lee JR, & Freeman M (2001) *Cell* **107,** 173-182.

16. Denisov IG, Hung SC, Weiss KE, McLean MA, Shiro Y, Park SY, Champion PM, & Sligar SG (2001) *J Inorg Biochem* **87,** 215-226.

17. Rawson RB, Zelenski NG, Nijhawan D, Ye J, Sakai J, Hasan MT, Chang TY, Brown MS, & Goldstein JL (1997) *Mol Cell* **1,** 47-57.

18. Bihel F, Das C, Bowman MJ, & Wolfe MS (2004) *J Med Chem* **47,** 3931-3933.

19. Wolfe MS (2009) *J Biol Chem* **284,** 13969-13973.

20. Lehninger AL, Nelson DL, & Cox MM (2005) *Lehninger principles of biochemistry* (W.H. Freeman, New York).

21. Berg JM, Tymoczko JL, & Stryer L (2002) *Biochemistry* (W.H. Freeman, New York).

22. Mitra K, Ubarretxena-Belandia I, Taguchi T, Warren G, & Engelman DM (2004) *Proc Natl Acad Sci U S A* **101,** 4083-4088.

23. Nagle JF & Tristram-Nagle S (2000) *Biochim Biophys Acta* **1469,** 159-195.

24. Simons K & Vaz WL (2004) *Annu Rev Biophys Biomol Struct* **33,** 269-295.

25. Rajendran L & Simons K (2005) *J Cell Sci* **118,** 1099-1102.

26. Pike LJ (2005) *Biochim Biophys Acta* **1746,** 260-273.

27. Carruthers A & Melchior DL (1988) *Annu Rev Physiol* **50,** 257-271.

28. Rothman JE & Engelman DM (1972) *Nat New Biol* **237,** 42-44.

29. Szabo G (1974) *Nature* **252,** 47-49.

30. Hinz HJ & Sturtevant JM (1972) *J Biol Chem* **247,** 3697-3700.

31. Dowhan W, and Bogdanov, M. (2002) in *Biochemistry of Lipids, Lipoproteins and Membranes*, ed. Vance DEVaJE (Elsevier Science B. V.), p. 35.

32. Lomize AL, Pogozheva ID, Lomize MA, & Mosberg HI (2006) *Protein Sci* **15,** 1318-1333.

33. Lomize AL, Pogozheva ID, Lomize MA, & Mosberg HI (2007) *BMC Struct Biol* **7,** 44.

34. Lomize MA, Lomize AL, Pogozheva ID, & Mosberg HI (2006) *Bioinformatics* **22,** 623-625.

35. Piknova B, Perochon E, & Tocanne JF (1993) *Eur J Biochem* **218,** 385-396.

36. Jordan P, Fromme P, Witt HT, Klukas O, Saenger W, & Krauss N (2001) *Nature* **411,** 909-917.

37. Sui H, Han BG, Lee JK, Walian P, & Jap BK (2001) *Nature* **414,** 872-878.

38. Sung MT, Lai YT, Huang CY, Chou LY, Shih HW, Cheng WC, Wong CH, & Ma C (2009) *Proc Natl Acad Sci U S A* **106,** 8824-8829.

39. Openshaw AE, Race PR, Monzo HJ, Vazquez-Boland JA, & Banfield MJ (2005) *J Biol Chem* **280,** 35011-35017.

40. Colussi T, Parsonage D, Boles W, Matsuoka T, Mallett TC, Karplus PA, & Claiborne A (2008) *Biochemistry* **47,** 965-977.

41. Hwang PM, Zhou N, Shan X, Arrowsmith CH, & Vogel HJ (1998) *Biochemistry* **37,** 4288-4298.

42. Hunter HN, Fulton DB, Ganz T, & Vogel HJ (2002) *J Biol Chem* **277,** 37597-37603.

43. Sasaki K, Dockerill S, Adamiak DA, Tickle IJ, & Blundell T (1975) *Nature* **257,** 751-757.

44. Florkin M (1957) *Rev Med Liege* **12,** 139-144.

45. Northrop JH (1929) *Science* **69,** 580.

46. Herriott RM & Northrop JH (1936) *Science* **83,** 469-470.

47. Rawlings ND, Barrett AJ, & Bateman A (2010) *Nucleic Acids Res* **38,** D227-233.

48. Manning G, Whyte DB, Martinez R, Hunter T, & Sudarsanam S (2002) *Science* **298,** 1912-1934.

49. Ren Q & Paulsen IT (2007) *J Mol Microbiol Biotechnol* **12,** 165-179.

50. Pitzer F, Dantes A, Fuchs T, Baumeister W, & Amsterdam A (1996) *FEBS Lett* **394,** 47-50.

51. Delic J, Morange M, & Magdelenat H (1993) *Mol Cell Biol* **13,** 4875-4883.

52. Havens CG, Ho A, Yoshioka N, & Dowdy SF (2006) *Mol Cell Biol* **26,** 4701-4711.

53. Sakai J, Duncan EA, Rawson RB, Hua X, Brown MS, & Goldstein JL (1996) *Cell* **85,** 1037-1046.

54. Lyko F, Martoglio B, Jungnickel B, Rapoport TA, & Dobberstein B (1995) *J Biol Chem* **270,** 19873-19878.

55. Arbige MV & Pitcher WH (1989) *Trends in Biotechnology* **7,** 330-335.

56. Gupta R, Beg QK, & Lorenz P (2002) *Appl Microbiol Biotechnol* **59,** 15-32.

57. Maurer KH (2004) *Curr Opin Biotechnol* **15,** 330-334.

58. Gupta R & Ramnani P (2006) *Appl Microbiol Biotechnol* **70,** 21-33.

59. Singh A, Ghosh VK, & Ghosh P (1994) *Letters in Applied Microbiology* **18,** 177-180.

60. James P, Quadroni M, Carafoli E, & Gonnet G (1993) *Biochem Biophys Res Commun* **195,** 58-64.

61. Yates JR, 3rd, Speicher S, Griffin PR, & Hunkapiller T (1993) *Anal Biochem* **214,** 397-408.

62. Chatterjee S, Schoepe J, Lohmer S, & Schomburg D (2005) *Protein Expr Purif* **39,** 137-143.

63. Zhu F, Han B, Kumar P, Liu X, Ma X, Wei X, Huang L, Guo Y, Han L, Zheng C*, et al.* (2010) *Nucleic Acids Res* **38,** D787-791.

64. Quesada V, Ordonez GR, Sanchez LM, Puente XS, & Lopez-Otin C (2009) *Nucleic Acids Res* **37,** D239-243.

65. Alcalay RN, Caccappolo E, Mejia-Santana H, Tang MX, Rosado L, Ross BM, Verbitsky M, Kisselev S, Louis ED, Comella C*, et al. Arch Neurol* **67,** 1116-1122.

66. Gribouval O, Gonzales M, Neuhaus T, Aziza J, Bieth E, Laurent N, Bouton JM, Feuillet F, Makni S, Ben Amar H*, et al.* (2005) *Nat Genet* **37,** 964-968.

67. Zhu S, Hsu AP, Vacek MM, Zheng L, Schaffer AA, Dale JK, Davis J, Fischer RE, Straus SE, Boruchov D*, et al.* (2006) *Hum Genet* **119,** 284-294.

68. De Strooper B (2003) *Neuron* **38,** 9-12.

69. De Strooper B & Annaert W (2001) *J Cell Biol* **152,** F17-20.

70. Petek E, Windpassinger C, Vincent JB, Cheung J, Boright AP, Scherer SW, Kroisel PM, & Wagner K (2001) *Am J Hum Genet* **68,** 848-858.

71. Kopan R & Ilagan MX (2004) *Nat Rev Mol Cell Biol* **5,** 499-504.

72. Weihofen A, Lemberg MK, Ploegh HL, Bogyo M, & Martoglio B (2000) *J Biol Chem* **275,** 30951-30956.

73. Haass C & Steiner H (2002) *Trends Cell Biol* **12,** 556-562.

74. Steiner H, Kostka M, Romig H, Basset G, Pesold B, Hardy J, Capell A, Meyn L, Grim ML, Baumeister R*, et al.* (2000) *Nat Cell Biol* **2,** 848-851.

75. Sato C, Takagi S, Tomita T, & Iwatsubo T (2008) *J Neurosci* **28,** 6264-6271.

76. Wang J, Brunkan AL, Hecimovic S, Walker E, & Goate A (2004) *Neurobiol Dis* **15,** 654-666.

77. Wang J, Beher D, Nyborg AC, Shearman MS, Golde TE, & Goate A (2006) *J Neurochem* **96,** 218-227.

78. Iben LG, Olson RE, Balanda LA, Jayachandra S, Robertson BJ, Hay V, Corradi J, Prasad CV, Zaczek R, Albright CF*, et al.* (2007) *J Biol Chem* **282,** 36829-36836.

79. Friedmann E, Lemberg MK, Weihofen A, Dev KK, Dengler U, Rovelli G, & Martoglio B (2004) *J Biol Chem* **279,** 50790-50798.

80. Martoglio B & Golde TE (2003) *Hum Mol Genet* **12 Spec No 2,** R201-206.

81. Xia W & Wolfe MS (2003) *J Cell Sci* **116,** 2839-2844.

82. Okamoto K, Moriishi K, Miyamura T, & Matsuura Y (2004) *J Virol* **78,** 6370-6380.

83. Weihofen A, Binns K, Lemberg MK, Ashman K, & Martoglio B (2002) *Science* **296,** 2215-2218.

84. Weihofen A & Martoglio B (2003) *Trends Cell Biol* **13,** 71-78.

85. Lemberg MK & Martoglio B (2002) *Mol Cell* **10,** 735-744.

86. Hua X, Sakai J, Brown MS, & Goldstein JL (1996) *J Biol Chem* **271,** 10379-10384.

87. Sakai J, Rawson RB, Espenshade PJ, Cheng D, Seegmiller AC, Goldstein JL, & Brown MS (1998) *Mol Cell* **2,** 505-514.

88. Duncan EA, Dave UP, Sakai J, Goldstein JL, & Brown MS (1998) *J Biol Chem* **273,** 17801-17809.

89. Baonza A, Casci T, & Freeman M (2001) *Curr Biol* **11,** 396-404.

90. Wasserman JD & Freeman M (1998) *Cell* **95,** 355-364.

91. Freeman M (1994) *Mech Dev* **48,** 25-33.

92. Wasserman JD, Urban S, & Freeman M (2000) *Genes Dev* **14,** 1651-1663.

93. Urban S, Brown G, & Freeman M (2004) *Development* **131,** 1835-1845.

94. Fuss B & Hoch M (2002) *Curr Biol* **12,** 171-179.

95. Tsruya R, Wojtalla A, Carmon S, Yogev S, Reich A, Bibi E, Merdes G, Schejter E, & Shilo BZ (2007) *EMBO J* **26,** 1211-1220.

96. Urban S, Schlieper D, & Freeman M (2002) *Curr Biol* **12,** 1507-1512.

97. Klambt C (2002) *Curr Biol* **12,** R21-23.

98. Yogev S, Schejter ED, & Shilo BZ (2008) *EMBO J* **27,** 1219-1230.

99. Koonin EV, Makarova KS, Rogozin IB, Davidovic L, Letellier MC, & Pellegrini L (2003) *Genome Biol* **4,** R19.

100. Lemieux MJ, Fischer SJ, Cherney MM, Bateman KS, & James MN (2007) *Proc Natl Acad Sci U S A* **104,** 750-754.

101. Brenner S (1988) *Nature* **334,** 528-530.

102. Wang Y, Zhang Y, & Ha Y (2006) *Nature* **444,** 179-180.

103. Clemmer KM, Sturgill GM, Veenstra A, & Rather PN (2006) *J Bacteriol* **188,** 3415-3419.

104. Wang Y & Ha Y (2007) *Proc Natl Acad Sci U S A* **104,** 2098-2102.

105. Nakagawa T, Guichard A, Castro CP, Xiao Y, Rizen M, Zhang HZ, Hu D, Bang A, Helms J, Bier E*, et al.* (2005) *Dev Dyn* **233,** 1315-1331.

106. Pascall JC & Brown KD (2004) *Biochem Biophys Res Commun* **317,** 244-252.

107. Urban S & Freeman M (2003) *Mol Cell* **11,** 1425-1434.

108. Li YM, Lai MT, Xu M, Huang Q, DiMuzio-Mower J, Sardana MK, Shi XP, Yin KC, Shafer JA, & Gardell SJ (2000) *Proc Natl Acad Sci U S A* **97,** 6138-6143.

109. Wu Z, Yan N, Feng L, Oberstein A, Yan H, Baker RP, Gu L, Jeffrey PD, Urban S, & Shi Y (2006) *Nat Struct Mol Biol* **13,** 1084-1091.

110. Baker RP, Young K, Feng L, Shi Y, & Urban S (2007) *Proc Natl Acad Sci U S A* **104,** 8257-8262.

111. Maegawa S, Ito K, & Akiyama Y (2005) *Biochemistry* **44,** 13543-13552.

112. Maegawa S, Koide K, Ito K, & Akiyama Y (2007) *Mol Microbiol* **64,** 435-447.

113. Strisovsky K, Sharpe HJ, & Freeman M (2009) *Mol Cell* 36, 1048-1059.

114. Lee JR, Urban S, Garvey CF, & Freeman M (2001) Cell 107, 161-171.

Chapter 2

1. Wolfe MS (2009) *J Biol Chem* **284,** 13969-13973.

2. Pascall JC & Brown KD (1998) *FEBS Lett* **429,** 337-340.

3. Gottlieb E (2006) *Cell* **126,** 27-29.

4. O'Donnell RA & Blackman MJ (2005) *Curr Opin Microbiol* **8,** 422-427.

5. Kanaoka MM, Urban S, Freeman M, & Okada K (2005) *FEBS Lett* **579,** 5723-5728.

6. Dowse TJ, Pascall JC, Brown KD, & Soldati D (2005) *Int J Parasitol* **35,** 747-756.

7. Dowse TJ & Soldati D (2005) *Trends Parasitol* **21,** 254-258.

8. Howell SA, Hackett F, Jongco AM, Withers-Martinez C, Kim K, Carruthers VB, & Blackman MJ (2005) *Mol Microbiol* **57,** 1342-1356.

9. Koonin EV, Makarova KS, Rogozin IB, Davidovic L, Letellier MC, & Pellegrini L (2003) *Genome Biol* **4,** R19.

10. Urban S, Schlieper D, & Freeman M (2002) *Curr Biol* **12,** 1507-1512.

11. Urban S, Lee JR, & Freeman M (2002) *EMBO J* **21,** 4277-4286.

12. Lee JR, Urban S, Garvey CF, & Freeman M (2001) *Cell* **107,** 161-171.

13. Pascall JC, Luck JE, & Brown KD (2002) *Biochem J* **363,** 347-352.

14. Tsruya R, Schlesinger A, Reich A, Gabay L, Sapir A, & Shilo BZ (2002) *Genes Dev* **16,** 222-234.

15. Freeman DM (2004) *Nature Reviews* **5,** 188-197.

16. Urban S, Lee JR, & Freeman M (2001) *Cell* **107,** 173-182.

17. McQuibban GA, Saurya S, & Freeman M (2003) *Nature* **423,** 537-541.

18. McQuibban GA, Lee JR, Zheng L, Juusola M, & Freeman M (2006) *Curr Biol* **16,** 982-989.

19. Lohi O, Urban S, & Freeman M (2004) *Curr Biol* **14,** 236-241.

20. Rather PN, Ding X, Baca-DeLancey RR, & Siddiqui S (1999) *J Bacteriol* **181,** 7185-7191.

21. Gallio M, Sturgill G, Rather P, & Kylsten P (2002) *Proc Natl Acad Sci U S A* **99,** 12208-12213.

22. Stevenson LG, Strisovsky K, Clemmer KM, Bhatt S, Freeman M, & Rather PN (2007) *Proc Natl Acad Sci U S A* **104,** 1003-1008.

23. Wang Y, Zhang Y, & Ha Y (2006) *Nature* **444,** 179-180.

24. Wu Z, Yan N, Feng L, Oberstein A, Yan H, Baker RP, Gu L, Jeffrey PD, Urban S, & Shi Y (2006) *Nat Struct Mol Biol* **13,** 1084-1091.

25. Ben-Shem A, Fass D, & Bibi E (2007) *Proc Natl Acad Sci U S A* **104,** 462-466.

26. Lemieux MJ, Fischer SJ, Cherney MM, Bateman KS, & James MN (2007) *Proc Natl Acad Sci U S A* **104,** 750-754.

27. Baker RP, Young K, Feng L, Shi Y, & Urban S (2007) *Proc Natl Acad Sci U S A* **104,** 8257-8262.

28. Maegawa S, Koide K, Ito K, & Akiyama Y (2007) *Mol Microbiol* **64,** 435-447.

29. Alexandrov A, Dutta K, & Pascal SM (2001) *Biotechniques* **30,** 1194-1198.

30. Cadene M & Chait BT (2000) *Anal Chem* **72,** 5655-5658.

31. Fenyo D, Wang Q, DeGrasse JA, Padovan JC, Cadene M, & Chait BT (2007) *J Vis Exp***,** 192.

32. Wasserman JD, Urban S, & Freeman M (2000) *Genes Dev* **14,** 1651-1663.

33. Wasserman JD & Freeman M (1998) *Cell* **95,** 355-364.

34. Freeman M (1994) *Mech Dev* **48,** 25-33.

35. Freeman M, Kimmel BE, & Rubin GM (1992) *Development* **116,** 335-346.

36. Urban S & Wolfe MS (2005) *Proc Natl Acad Sci U S A* **102,** 1883-1888.

37. Del Rio A, Dutta K, Chavez J, Ubarretxena-Belandia I, & Ghose R (2007) *J Mol Biol* **365,** 109-122.

38. Pascall JC & Brown KD (2004) *Biochem Biophys Res Commun* **317,** 244-252.

39. Lemberg MK, Menendez J, Misik A, Garcia M, Koth CM, & Freeman M (2005) *Embo J* **24,** 464-472.

40. le Coutre J, Whitelegge JP, Gross A, Turk E, Wright EM, Kaback HR, & Faull KF (2000) *Biochemistry* **39,** 4237-4242.

41. Strisovsky K, Sharpe HJ, & Freeman M (2009) *Mol Cell* **36,** 1048-1059.

42. Beel AJ & Sanders CR (2008) *Cell Mol Life Sci* **65,** 1311-1334.

43. Lei X, Ahn K, Zhu L, Ubarretxena-Belandia I, & Li Y (2008) *Biochemistry*.

44. Snijder HJ, Ubarretxena-Belandia I, Blaauw M, Kalk KH, Verheij HM, Egmond MR, Dekker N, & Dijkstra BW (1999) *Nature* **401,** 717-721.

45. Wang Y & Ha Y (2007) *Proc Natl Acad Sci U S A* **104,** 2098-2102.

46. Meissner M, Reiss M, Viebig N, Carruthers VB, Toursel C, Tomavo S, Ajioka JW, & Soldati D (2002) *J Cell Sci* **115,** 563-574.

47. Li YM, Lai MT, Xu M, Huang Q, DiMuzio-Mower J, Sardana MK, Shi XP, Yin KC, Shafer JA, & Gardell SJ (2000) *Proc Natl Acad Sci U S A* **97,** 6138-6143.

48. Urban S & Freeman M (2003) *Mol Cell* **11,** 1425-1434.

49. O'Donnell RA, Hackett F, Howell SA, Treeck M, Struck N, Krnajski Z, Withers-Martinez C, Gilberger TW, & Blackman MJ (2006) *J Cell Biol* **174,** 1023-1033.

50. Maegawa S, Ito K, & Akiyama Y (2005) *Biochemistry* **44,** 13543-13552.

51. Opitz C, Di Cristina M, Reiss M, Ruppert T, Crisanti A, & Soldati D (2002) *EMBO J* **21,** 1577-1585.

52. Tatsuta T, Augustin S, Nolden M, Friedrichs B, & Langer T (2007) *EMBO J* **26,** 325-335.

53. Urban S (2003) *Molecular Cell* **11,** 1425-1434.

54. Mitra K, Ubarretxena-Belandia I, Taguchi T, Warren G, & Engelman DM (2004) *Proc Natl Acad Sci U S A* **101,** 4083-4088.

55. Wang Y, Maegawa S, Akiyama Y, & Ha Y (2007) *J Mol Biol* **374,** 1104-1113.

Chapter 3

1. Urban S, Lee JR, & Freeman M (2001) *Cell* **107,** 173-182.

2. Urban S & Freeman M (2003) *Mol Cell* **11,** 1425-1434.

3. Akiyama Y & Maegawa S (2007) *Mol Microbiol* **64,** 1028-1037.

4. Maegawa S, Ito K, & Akiyama Y (2005) *Biochemistry* **44,** 13543-13552.

5. Urban S & Wolfe MS (2005) *Proc Natl Acad Sci U S A* **102,** 1883-1888.

6. Baker RP, Young K, Feng L, Shi Y, & Urban S (2007) *Proc Natl Acad Sci U S A* **104,** 8257-8262.

7. Ha Y (2008) *Semin Cell Dev Biol*.

8. Erez E & Bibi E (2009) *Biochemistry* **48,** 12314-12322.

9. Strisovsky K, Sharpe HJ, & Freeman M (2009) *Mol Cell* **36,** 1048-1059.

10. Alexandrov A, Dutta K, & Pascal SM (2001) *Biotechniques* **30,** 1194-1198.

11. Landolt-Marticorena C, Williams KA, Deber CM, & Reitchmeier RAF (1993) *J Mol Biol* **229,** 602-608.

12. Arkin IT & Brunger AT (1998) *Biochim Biophys Acta* **1429,** 113-128.

13. Hedstrom L (2002) *Chem Rev* **102,** 4501-4524.

14. Tulumello DV & Deber CM (2009) *Biochemistry* **48,** 12096-12103.

15. Tang YC & Deber CM (2002) *Biopolymers* **65,** 254-262.

16. Liu LP & Deber CM (1998) *J Biol Chem* **273,** 23645-23648.

17. Deber CM & Li SC (1995) *Biopolymers* **37,** 295-318.

18. Fluman N, Cohen-Karni D, Weiss T, & Bibi E (2009) *J Biol Chem* **284,** 32296-32304.

19. Erez E, Fass D, & Bibi E (2009) *Nature* **459,** 371-378.

20. Lemberg MK, Menendez J, Misik A, Garcia M, Koth CM, & Freeman M (2005) *EMBO J* **24,** 464-472.

21. Maegawa S, Koide K, Ito K, & Akiyama Y (2007) *Mol Microbiol* **64,** 435-447.

22. Weihofen A, Binns K, Lemberg MK, Ashman K, & Martoglio B (2002) *Science* **296,** 2215-2218.

23. Heimann M, Roman-Sosa G, Martoglio B, Thiel HJ, & Rumenapf T (2006) *J Virol* **80,** 1915-1921.

24. Bohm C, Seibel NM, Henkel B, Steiner H, Haass C, & Hampe W (2006) *J Biol Chem* **281,** 14547-14553.

25. Lemberg MK & Martoglio B (2002) *Mol Cell* **10,** 735-744.

Chapter 4.

1. Brown MS, Ye J, Rawson RB, & Goldstein JL (2000) *Cell* **100,** 391-398.

2. Urban S & Freeman M (2002) *Curr Opin Genet Dev* **12,** 512-518.

3. Wolfe MS & Kopan R (2004) *Science* **305,** 1119-1123.

4. Wolfe MS, De Los Angeles J, Miller DD, Xia W, & Selkoe DJ (1999) *Biochemistry* **38,** 11223-11230.

5. De Strooper B, Annaert W, Cupers P, Saftig P, Craessaerts K, Mumm JS, Schroeter EH, Schrijvers V, Wolfe MS, Ray WJ*, et al.* (1999) *Nature* **398,** 518-522.

6. Ni CY, Murphy MP, Golde TE, & Carpenter G (2001) *Science* **294,** 2179-2181.

7. Sisodia SS & St George-Hyslop PH (2002) *Nat Rev Neurosci* **3,** 281-290.

8. Rawson RB, Zelenski NG, Nijhawan D, Ye J, Sakai J, Hasan MT, Chang TY, Brown MS, & Goldstein JL (1997) *Mol Cell* **1,** 47-57.

9. Sturtevant MA, Roark M, & Bier E (1993) *Genes Dev* **7,** 961-973.

10. Freeman DM (2004) *Nature Reviews* **5,** 188-197.

11. Lee JR, Urban S, Garvey CF, & Freeman M (2001) *Cell* **107,** 161-171.

12. Urban S, Lee JR, & Freeman M (2001) *Cell* **107,** 173-182.

13. Koonin EV, Makarova KS, Rogozin IB, Davidovic L, Letellier MC, & Pellegrini L (2003) *Genome Biol* **4,** R19.

14. Rather PN, Ding X, Baca-DeLancey RR, & Siddiqui S (1999) *J Bacteriol* **181,** 7185-7191.

15. Gallio M, Sturgill G, Rather P, & Kylsten P (2002) *Proc Natl Acad Sci U S A* **99,** 12208-12213.

16. Urban S, Schlieper D, & Freeman M (2002) *Curr Biol* **12,** 1507-1512.

17. McQuibban GA, Saurya S, & Freeman M (2003) *Nature* **423,** 537-541.

18. Lohi O, Urban S, & Freeman M (2004) *Curr Biol* **14,** 236-241.

19. Kanaoka MM, Urban S, Freeman M, & Okada K (2005) *FEBS Lett* **579,** 5723-5728.

20. Lemberg MK, Menendez J, Misik A, Garcia M, Koth CM, & Freeman M (2005) *EMBO J* **24,** 464-472.

21. Wang Y, Zhang Y, & Ha Y (2006) *Nature* **444,** 179-180.

22. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, & Bax A (1995) *J Biomol NMR* **6,** 277-293.

23. Johnson BA (2004) *Methods Mol Biol* **278,** 313-352.

24. Sattler M, Schleucherb, J., Griesinger, C (1999) *Progress in Nuclear Magnetic Resonance Spectroscopy* **34,** 93–158.

25. Grzesiek FCaS (1999) *J. Am. Chem. Soc.* **121,** 1601-1602.

26. G.W. Vuister AB (1993) *J. Am. Chem. Soc.* **115,** 7772-7777.

27. Cornilescu G, Delaglio F, & Bax A (1999) *J Biomol NMR* **13,** 289-302.

28. Habeck M, Rieping W, Linge JP, & Nilges M (2004) *Methods Mol Biol* **278,** 379-402.

29. Linge JP, Williams MA, Spronk CA, Bonvin AM, & Nilges M (2003) *Proteins* **50,** 496-506.

30. Koradi R, Billeter M, & Wuthrich K (1996) *J Mol Graph* **14,** 51-55, 29-32.

31. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, & Thornton JM (1996) *J Biomol NMR* **8,** 477-486.

32. Palmer AG, 3rd (2001) *Annu Rev Biophys Biomol Struct* **30,** 129-155.

33. P.T. Boggs JRD, R.H. Byrd and R.B. Schnabel (1989) *ACM Trans. Math. Software* **15,** 348-364.

34. Ghose R, Fushman D, & Cowburn D (2001) *J Magn Reson* **149,** 204-217.

35. Szabo GLaA (1982) *J. Am. Chem. Soc.* **104,** 4546–4559.

36. G.M. Clore AS, A. Bax, L.E. Kay, P.C. Driscoll and A.M. Gronenborn (1990) *J. Am. Chem. Soc.* **112,** 4936–4989.

37. Fushman D, Cahill S, & Cowburn D (1997) *J Mol Biol* **266,** 173-194.

38. Del Rio A, Anand A, & Ghose R (2006) *J Magn Reson* **180,** 1-17.

39. Cho W & Stahelin RV (2005) *Annu Rev Biophys Biomol Struct* **34,** 119-151.

40. Luo P & Baldwin RL (1997) *Biochemistry* **36,** 8413-8421.

41. Sreerama N & Woody RW (2004) *Methods Enzymol* **383,** 318-351.

42. Sonnhammer EL, von Heijne G, & Krogh A (1998) *Proc Int Conf Intell Syst Mol Biol* **6,** 175-182.

43. Holm L & Sander C (1993) *J Mol Biol* **233,** 123-138.

44. Yip CK, Kimbrough TG, Felise HB, Vuckovic M, Thomas NA, Pfuetzner RA, Frey EA, Finlay BB, Miller SI, & Strynadka NC (2005) *Nature* **435,** 702-707.

45. Mota LJ, Sorg I, & Cornelis GR (2005) *FEMS Microbiol Lett* **252,** 1-10.

46. Dutta K, Alexandrov A, Huang H, & Pascal SM (2001) *Protein Sci* **10,** 2531-2540.

47. Volkman BF, Lipson D, Wemmer DE, & Kern D (2001) *Science* **291,** 2429-2433.

48. Grunberg R, Leckner J, & Nilges M (2004) *Structure* **12,** 2125-2136.

49. Urban S & Freeman M (2003) *Mol Cell* **11,** 1425-1434.

50. Urban S, Lee JR, & Freeman M (2002) *EMBO J* **21,** 4277-4286.

51. Hardy J & Selkoe DJ (2002) *Science* **297,** 353-356.

Chapter 5.

1. Brown MS, Ye J, Rawson RB, & Goldstein JL (2000) *Cell* **100,** 391-398.

2. Heldin CH & Ericsson J (2001) *Science* **294,** 2111-2113.

3. Urban S & Freeman M (2002) *Curr Opin Genet Dev* **12,** 512-518.

4. De Strooper B, Annaert W, Cupers P, Saftig P, Craessaerts K, Mumm JS, Schroeter EH, Schrijvers V, Wolfe MS, Ray WJ*, et al.* (1999) *Nature* **398,** 518-522.

5. Ni CY, Murphy MP, Golde TE, & Carpenter G (2001) *Science* **294,** 2179-2181.

6. Sherrington R, Rogaev EI, Liang Y, Rogaeva EA, Levesque G, Ikeda M, Chi H, Lin C, Li G, Holman K*, et al.* (1995) *Nature* **375,** 754-760.

7. Levy-Lahad E, Wijsman EM, Nemens E, Anderson L, Goddard KA, Weber JL, Bird TD, & Schellenberg GD (1995) *Science* **269,** 970-973.

8. Rogaev EI, Sherrington R, Rogaeva EA, Levesque G, Ikeda M, Liang Y, Chi H, Lin C, Holman K, Tsuda T*, et al.* (1995) *Nature* **376,** 775-778.

9. Sisodia SS & St George-Hyslop PH (2002) *Nat Rev Neurosci* **3,** 281-290.

10. Lee JR, Urban S, Garvey CF, & Freeman M (2001) *Cell* **107,** 161-171.

11. Urban S, Lee JR, & Freeman M (2001) *Cell* **107,** 173-182.

12. Wolfe MS & Selkoe DJ (2002) *Science* **296,** 2156-2157.

13. Weihofen A & Martoglio B (2003) *Trends Cell Biol* **13,** 71-78.

14. Beel AJ & Sanders CR (2008) *Cell Mol Life Sci* **65,** 1311-1334.

15. Marambaud P & Robakis NK (2005) *Genes Brain Behav* **4,** 134-146.

16. Wolfe MS & Kopan R (2004) *Science* **305,** 1119-1123.

17. Shah S, Lee SF, Tabuchi K, Hao YH, Yu C, LaPlant Q, Ball H, Dann CE, 3rd, Sudhof T, & Yu G (2005) *Cell* **122,** 435-447.

18. Price DL, Sisodia SS, & Borchelt DR (1998) *Science* **282,** 1079-1083.

19. Selkoe DJ (2001) *Physiol Rev* **81,** 741-766.

20. Weidemann A, Eggert S, Reinhard FB, Vogel M, Paliga K, Baier G, Masters CL, Beyreuther K, & Evin G (2002) *Biochemistry* **41,** 2825-2835.

21. Zhao G, Cui MZ, Mao G, Dong Y, Tan J, Sun L, & Xu X (2005) *J Biol Chem* **280,** 37689-37697.

22. Price DL & Sisodia SS (1998) *Annu Rev Neurosci* **21,** 479-505.

23. Siman R, Reaume AG, Savage MJ, Trusko S, Lin YG, Scott RW, & Flood DG (2000) *J Neurosci* **20,** 8717-8726.

24. Moehlmann T, Winkler E, Xia X, Edbauer D, Murrell J, Capell A, Kaether C, Zheng H, Ghetti B, Haass C*, et al.* (2002) *Proc Natl Acad Sci U S A* **99,** 8025-8030.

25. Schroeter EH, Ilagan MX, Brunkan AL, Hecimovic S, Li YM, Xu M, Lewis HD, Saxena MT, De Strooper B, Coonrod A*, et al.* (2003) *Proc Natl Acad Sci U S A* **100,** 13075-13080.

26. Qi Y, Morishima-Kawashima M, Sato T, Mitsumori R, & Ihara Y (2003) *Biochemistry* **42,** 1042-1052.

27. Walker ES, Martinez M, Brunkan AL, & Goate A (2005) *J Neurochem* **92,** 294-301.

28. Kumar-Singh S, Theuns J, Van Broeck B, Pirici D, Vennekens K, Corsmit E, Cruts M, Dermaut B, Wang R, & Van Broeckhoven C (2006) *Hum Mutat* **27,** 686-695.

29. Bentahir M, Nyabi O, Verhamme J, Tolia A, Horre K, Wiltfang J, Esselmann H, & De Strooper B (2006) *J Neurochem* **96,** 732-742.

30. Shen J & Kelleher RJ, 3rd (2007) *Proc Natl Acad Sci U S A* **104,** 403-409.

31. Kaneko H, Kakita A, Kasuga K, Nozaki H, Ishikawa A, Miyashita A, Kuwano R, Ito G, Iwatsubo T, Takahashi H*, et al.* (2007) *J Neurosci* **27,** 13092-13097.

32. Sato T, Nyborg AC, Iwata N, Diehl TS, Saido TC, Golde TE, & Wolfe MS (2006) *Biochemistry* **45,** 8649-8656.

33. Vidal GA, Naresh A, Marrero L, & Jones FE (2005) *J Biol Chem* **280,** 19777-19783.

34. Lichtenthaler SF, Wang R, Grimm H, Uljon SN, Masters CL, & Beyreuther K (1999) *Proc Natl Acad Sci U S A* **96,** 3053-3058.

35. Andersson CX, Fernandez-Rodriguez J, Laos S, Baeckstrom D, Haass C, & Hansson GC (2005) *Biochem J* **387,** 377-384.

36. Taniguchi Y, Kim SH, & Sisodia SS (2003) *J Biol Chem* **278,** 30425-30428.

37. Strisovsky K, Sharpe HJ, & Freeman M (2009) *Mol Cell* **36,** 1048-1059.

38. Herlan M, Vogel F, Bornhovd C, Neupert W, & Reichert AS (2003) *J Biol Chem* **278,** 27781-27788.

39. Lemberg MK, Bland FA, Weihofen A, Braud VM, & Martoglio B (2001) *J Immunol* **167,** 6441-6446.

40. Dev KK, Chatterjee S, Osinde M, Stauffer D, Morgan H, Kobialko M, Dengler U, Rueeger H, Martoglio B, & Rovelli G (2006) *Eur J Pharmacol* **540,** 10-17.

41. Heimann M, Roman-Sosa G, Martoglio B, Thiel HJ, & Rumenapf T (2006) *J Virol* **80,** 1915-1921.

42. Fluhrer R, Grammer G, Israel L, Condron MM, Haffner C, Friedmann E, Bohland C, Imhof A, Martoglio B, Teplow DB*, et al.* (2006) *Nat Cell Biol* **8,** 894-896.

43. Seidah NG, Khatib AM, & Prat A (2006) *Biol Chem* **387,** 871-877.

44. Yang T, Espenshade PJ, Wright ME, Yabe D, Gong Y, Aebersold R, Goldstein JL, & Brown MS (2002) *Cell* **110,** 489-500.

45. Ubarretxena-Belandia I & Engelman DM (2001) *Curr Opin Struct Biol* **11,** 370-376.

46. Haeuptle MT, Flint N, Gough NM, & Dobberstein B (1989) *J Cell Biol* **108,** 1227-1236.

47. von Heijne G (1985) *J Mol Biol* **184,** 99-105.

48. Feng L, Yan H, Wu Z, Yan N, Wang Z, Jeffrey PD, & Shi Y (2007) *Science* **318,** 1608-1612.

49. Zelenski NG, Rawson RB, Brown MS, & Goldstein JL (1999) *J Biol Chem* **274,** 21973-21980.

50. Kanehara K, Akiyama Y, & Ito K (2001) *Gene* **281,** 71-79.

51. Brown MS & Goldstein JL (1999) *Proc Natl Acad Sci U S A* **96,** 11041-11048.

52. Kolygo K, Ranjan N, Kress W, Striebel F, Hollenstein K, Neelsen K, Steiner M, Summer H, & Weber-Ban E (2009) *J Struct Biol* **168,** 267-277.

53. Nyborg AC, Herl L, Berezovska O, Thomas AV, Ladd TB, Jansen K, Hyman BT, & Golde TE (2006) *Mol Neurodegener* **1,** 16.

54. Schaal R, Kupfahl C, Buchheidt D, Neumaier M, & Findeisen P (2007) *J Microbiol Methods* **71,** 93-100.

55. Swanson R, Raghavendra MP, Zhang W, Froelich C, Gettins PG, & Olson ST (2007) *J Biol Chem* **282,** 2305-2313.

56. Lei X & Li YM (2009) *J Mol Biol* **394,** 815-825.

57. Karlsson M, Fotiadis D, Sjovall S, Johansson I, Hedfalk K, Engel A, & Kjellbom P (2003) *FEBS Lett* **537,** 68-72.

58. Kukulski W, Schenk AD, Johanson U, Braun T, de Groot BL, Fotiadis D, Kjellbom P, & Engel A (2005) *J Mol Biol* **350,** 611-616.

59. Cipolat S, Rudka T, Hartmann D, Costa V, Serneels L, Craessaerts K, Metzger K, Frezza C, Annaert W, D'Adamio L*, et al.* (2006) *Cell* **126,** 163-175.

60. Frezza C, Cipolat S, Martins de Brito O, Micaroni M, Beznoussenko GV, Rudka T, Bartoli D, Polishuck RS, Danial NN, De Strooper B*, et al.* (2006) *Cell* **126,** 177-189.

61. Pellegrini L & Scorrano L (2007) *Cell Death Differ* **14,** 1275-1284.

62. Meissner C, Lorenz H, Weihofen A, Selkoe DJ, & Lemberg MK *J Neurochem* **117,** 856-867.

63. Shi G, Lee JR, Grimes DA, Racacho L, Ye D, Yang H, Ross OA, Farrer M, McQuibban GA, & Bulman DE *Hum Mol Genet* **20,** 1966-1974.

64. Whitworth AJ, Lee JR, Ho VM, Flick R, Chowdhury R, & McQuibban GA (2008) *Dis Model Mech* **1,** 168-174.

65. McQuibban GA, Lee JR, Zheng L, Juusola M, & Freeman M (2006) *Curr Biol* **16,** 982-989.

66. Deas E, Plun-Favreau H, Gandhi S, Desmond H, Kjaer S, Loh SH, Renton AE, Harvey RJ, Whitworth AJ, Martins LM*, et al. Hum Mol Genet* **20,** 867-879.

67. Jin SM, Lazarou M, Wang C, Kane LA, Narendra DP, & Youle RJ *J Cell Biol* **191,** 933-942.

68. Wu Z, Yan N, Feng L, Oberstein A, Yan H, Baker RP, Gu L, Jeffrey PD, Urban S, & Shi Y (2006) *Nat Struct Mol Biol* **13,** 1084-1091.

69. Wang Y, Zhang Y, & Ha Y (2006) *Nature* **444,** 179-180.

70. Hayrapetyan V, Rybalchenko V, Rybalchenko N, & Koulen P (2008) *Cell Calcium* **44,** 507-518.

71. Rybalchenko V, Hwang SY, Rybalchenko N, & Koulen P (2008) *Int J Biochem Cell Biol* **40,** 84-97.

72. Fraering PC (2007) *Curr Genomics* **8,** 531-549.

73. Urban S & Baker RP (2008) *Biol Chem* **389,** 1107-1115.

74. Urban S (2010) *Biochem J* **425,** 501-512.