The EMBO Journal (2010) 29, 3797–3809 | © 2010 European Molecular Biology Organization | All Rights Reserved 0261-4189/10
www.embojournal.org

THE EMBO JOURNAL

The structural basis for catalysis and substrate specificity of a rhomboid protease

Kutti R Vinothkumar[1,*], Kvido Strisovsky[1], Antonina Andreeva[1], Yonka Christova[1], Steven Verhelst[2] and Matthew Freeman[1,*]

[1]MRC Laboratory of Molecular Biology, Cambridge, UK and [2]Center for Integrated Protein Science Munich, Lehrstuhl Chemie der Biopolymere, Technische Universität München, Freising, Germany

Rhomboids are intramembrane proteases that use a catalytic dyad of serine and histidine for proteolysis. They are conserved in both prokaryotes and eukaryotes and regulate cellular processes as diverse as intercellular signalling, parasitic invasion of host cells, and mitochondrial morphology. Their widespread biological significance and consequent medical potential provides a strong incentive to understand the mechanism of these unusual enzymes for identification of specific inhibitors. In this study, we describe the structure of *Escherichia coli* rhomboid GlpG covalently bound to a mechanism-based isocoumarin inhibitor. We identify the position of the oxyanion hole, and the S_1- and S_2'-binding subsites of GlpG, which are the key determinants of substrate specificity. The inhibitor-bound structure suggests that subtle structural change is sufficient for catalysis, as opposed to large changes proposed from previous structures of unliganded GlpG. Using bound inhibitor as a template, we present a model for substrate binding at the active site and biochemically test its validity. This study provides a foundation for a structural explanation of rhomboid specificity and mechanism, and for inhibitor design.

The EMBO Journal (2010) **29,** 3797–3809. doi:10.1038/emboj.2010.243; Published online 1 October 2010
Subject Categories: proteins; structural biology
Keywords: intramembrane protease; isocoumarin; rhomboid; serine protease; substrate specificity

Introduction

Proteases regulate myriad biological and medically important processes—indeed they are the targets of many important drugs (Turk, 2006). Accordingly, they are some of the most-studied and well-characterised enzymes. Much less well understood, however, are the four recently discovered families of intramembrane proteases (site-2 metalloprotease (Rawson *et al*, 1998), presenilin (De Strooper *et al*, 1999), signal peptide peptidase (Weihofen *et al*, 2002), and rhomboids (Urban *et al*, 2001)), which share the property of cleaving transmembrane regions of substrates, thereby releasing soluble intracellular or luminal/extracellular domains to act as biological effectors (Wolfe and Kopan, 2004; Freeman, 2008). They seem to have evolved as proteases by convergent evolution, bearing no detectable phylogenetic relationship with the classical soluble proteases. Rhomboids were first identified in *Drosophila* as the cardinal activators of the EGF receptor signalling pathway (Wasserman *et al*, 2000; Lee *et al*, 2001; Urban *et al*, 2001). Rhomboids have since been found in all kingdoms of life (Koonin *et al*, 2003; Lemberg and Freeman, 2007b). The list of their biological functions is steadily growing but already includes growth factor signalling, quorum sensing, and protein translocation in bacteria, parasitic invasion by protozoa, and regulation of mitochondrial dynamics (Opitz *et al*, 2002; McQuibban *et al*, 2003; Urban and Freeman, 2003; Howell *et al*, 2005; Cipolat *et al*, 2006; O'Donnell *et al*, 2006; Stevenson *et al*, 2007; Baxt *et al*, 2008). The initial proposal that rhomboids depended on a catalytic serine was indirect, relying on genetic analysis and cell culture-based assays (Lee *et al*, 2001; Urban *et al*, 2001). The ability to overexpress and purify rhomboid homologues from bacterial sources paved the way for *in vitro* assays of proteolysis; this proved that rhomboids by themselves are sufficient to cleave specific substrates without the requirement of additional proteins or subunits (Lemberg *et al*, 2005; Urban and Wolfe, 2005). These *in vitro* studies further supported the serine protease classification, but implied that rhomboids were atypical, using a dyad of serine and histidine rather than the catalytic triad found commonly in classical soluble serine proteases (Hedstrom, 2002; Polgar, 2005; Ekici *et al*, 2008).

Further mechanistic insight was inferred, and the final doubts removed regarding whether the active site was truly within the membrane bilayer, with the reports of the crystal structures of the prokaryotic rhomboids, GlpG from *Escherichia coli* and *Haemophilus influenzae*. They revealed that the core of the structure consisted of a 6TM helical protein (Wang *et al*, 2006; Wu *et al*, 2006; Ben-Shem *et al*, 2007; Lemieux *et al*, 2007). Within this core, a short TM helix-4 and an extended loop-3 are surrounded by other helices to create a large hydrophilic indentation that opens to the periplasm. The catalytic serine found at the tip of TM4 and the general base histidine found in TM6 reside within this indentation, and are brought close together by tight packing of the TM helices. This active-site architecture, in conjunction with biochemical analysis, supported the conclusion that rhomboids are indeed a new family of enzyme.

Despite these advances in structural and biochemical analysis of rhomboids, the lack of any structure of enzyme–substrate or enzyme–inhibitor complexes leaves important questions to be answered. Understanding the catalytic mechanism of this new family of enzymes is of fundamental theoretical interest, but is also necessary for understanding their biological specificity, and being a key step in the identification and design of inhibitors. Indeed, no potent or specific rhomboid inhibitors have been reported, a gap that

*Corresponding author. KR Vinothkumar or M Freeman, MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK. Tel.: +44 122 340 2351; Fax: +44 122 341 2142;
E-mails: vkumar@mrc-lmb.cam.ac.uk or mf1@mrc-lmb.cam.ac.uk

Received: 3 August 2010; accepted: 6 September 2010; published online: 1 October 2010

becomes more striking as evidence increases regarding their medical significance. In fact, the only inhibitor class reported to have any significant, although weak, activity against rhomboids is isocoumarin; potent and irreversible inhibitor of classical serine proteases (Urban *et al*, 2001; Lemberg *et al*, 2005; Urban and Wolfe, 2005).

Although isocoumarin inhibition has contributed to the argument that rhomboids use a variation of the classical serine protease catalytic mechanism, neither their specificity for the presumed catalytic residues of rhomboids nor their mechanism of inhibition of rhomboids has been determined. In this study, we describe the structure of a rhomboid protease covalently bonded to an isocoumarin inhibitor through both the active-site serine and histidine. The position and orientation of the inhibitor locates the oxyanion hole, a key structural characteristic of serine proteases. This proves that rhomboids act by a mechanism that closely resembles their soluble counterparts, thereby setting them in a clear enzymological context. Binding of inhibitor is accompanied by conformational changes in the active site, the loop-5, and TM helices-5 and -6. The acyl enzyme structure suggests how a substrate might bind at the active site and provides a rationale for how rhomboid specificity can be achieved.

Results

Structure of a bacterial rhomboid in complex with an inhibitor

Isocoumarin inhibition of classical serine proteases begins with nucleophilic attack by the catalytic serine, resulting in opening of the isocoumarin ring, and formation of an acyl enzyme (Powers *et al* 2002). This can subsequently react with a nucleophile such as the catalytic histidine, to form a doubly covalent-bonded alkylated acyl enzyme, which is extremely stable (Figure 1A). Dichloroisocoumarin (DCI), a commonly used serine protease inhibitor, has modest activity against rhomboids (Lemberg *et al*, 2005; Urban and Wolfe, 2005). However, its short half-life of 20 min or less in many commonly used buffers (Harper *et al*, 1985) limits its usefulness for crystallography. Indeed, efforts to react DCI with GlpG crystals have been unsuccessful, probably due to its instability in solution and rapid deacylation (Wang and Ha, 2007; KR Vinothkumar *et al*, unpublished data). In a screen of an isocoumarin collection generously provided by Dr Matt Bogyo (Stanford University), we identified a more stable compound, 7-amino-4-chloro-3-methoxy isocoumarin (amino-methoxy-isocoumarin; originally synthesised by Dr James C Powers, Georgia Institute of Technology), with a half-life of $\sim$13 h in phosphate buffer (Harper *et al*, 1985), which inhibits GlpG with a half maximal inhibition of $\sim$6 μM (Figure 1B). Although there was a strong tendency of amino-methoxy-isocoumarin to disrupt GlpG crystals, we found conditions that allowed structural determination of the enzyme–inhibitor complex. The structure described here (PDB code 2XOW) was determined from a single crystal that diffracted to 2.09 Å (Table I).

The change of GlpG crystals from colourless to yellowish orange, and an $\sim$6-Å reduction in the *c*-axis upon soaking with the inhibitor, indicated a probable reaction of the inhibitor, and an associated structural change of the enzyme. Indeed, the initial map after molecular replacement shows a region of continuous positive density connected with active-site residues serine 201 and histidine 254. This represents the inhibitor bound to the enzyme active site, forming an alkylated acyl enzyme (Figure 1A and C). We observed the change in the unit cell and the formation of the doubly covalent bonded inhibitor independently in four different crystals. The inhibitor lies on the active site, $\sim$4.5 Å below the plane of the lipid head groups in the membrane (Figure 2A). The amino group (see diagram of the different parts of the inhibitor molecule in Figure 2B) of the inhibitor points towards the GlpG loop-3, whereas the methoxy group of the inhibitor points towards the gap between TM2 and TM5 and the lipid bilayer. The inhibitor amino group hydrogen bonds to a water molecule, and is in close contact to the main chain carbonyl oxygen of glycine 198. The oxygen atoms of the methyl acetate group form contacts with side chains of H150 and H254 (Figure 2C).

Formation of an acyl intermediate is one of the hallmarks of serine proteases, and the first description of serine proteases involved the trapping of an acyl intermediate, even before the sequence or structure was determined (Hartley and Kilby, 1950). With the structure of an alkylated acyl enzyme of GlpG, we now demonstrate unequivocally that rhomboids are serine proteases that use a catalytic mechanism analogous, but not identical, to the classical soluble serine proteases. Moreover, the fact that amino-methoxy-isocoumarin reacts only with S201, confirms that this is indeed the reactive nucleophile in the enzyme. Although this has been suggested by previous biochemical and structural study (for review, see Lemberg and Freeman, 2007a), and was hypothesised when rhomboids were first identified as proteases (Urban *et al*, 2001), they do not share common ancestry with their soluble counterparts, so their mechanism cannot confidently be inferred by evolutionary arguments.

Location of the oxyanion hole

The catalytic efficiency, as defined by increased rate constant or decreased transition state energy, of classical serine proteases is enhanced by the oxyanion hole (Henderson, 1970; Robertus *et al*, 1972; Hedstrom, 2002). In chymotrypsin, the main chain amides of the catalytic serine (S195) and a glycine two residues upstream (G193) stabilise the negatively charged oxyanion, a reaction intermediate that is formed from the carbonyl oxygen of the cleaved peptide bond. Residues contributing to the oxyanion hole in rhomboids have previously been tentatively suggested based on the structures (Ben-Shem *et al*, 2007; Lemieux *et al*, 2007). The benzoyl carbonyl oxygen of amino-methoxy-isocoumarin mimics the substrate peptide carbonyl, allowing us now to identify its position as the oxyanion hole in GlpG, and to identify the probable residues involved (Figure 2C and D). The main chain amide of G199 (equivalent to G193 in chymotrypsin) points away from the active site, so, at least in the case of GlpG, cannot hydrogen bond to the carbonyl oxygen. However, there are four other potential hydrogen-bonding donors in the vicinity of the benzoyl carbonyl oxygen. These include the main chain amides of S201 and L200, the side chain amide of N154 and the nitrogen (Nε) from the imidazole ring of H150. In this structure, the main chain amide of S201 forms a strong hydrogen bond with the benzoyl carbonyl oxygen of the inhibitor, whereas the other residues form hydrogen bonds that can be inferred from their

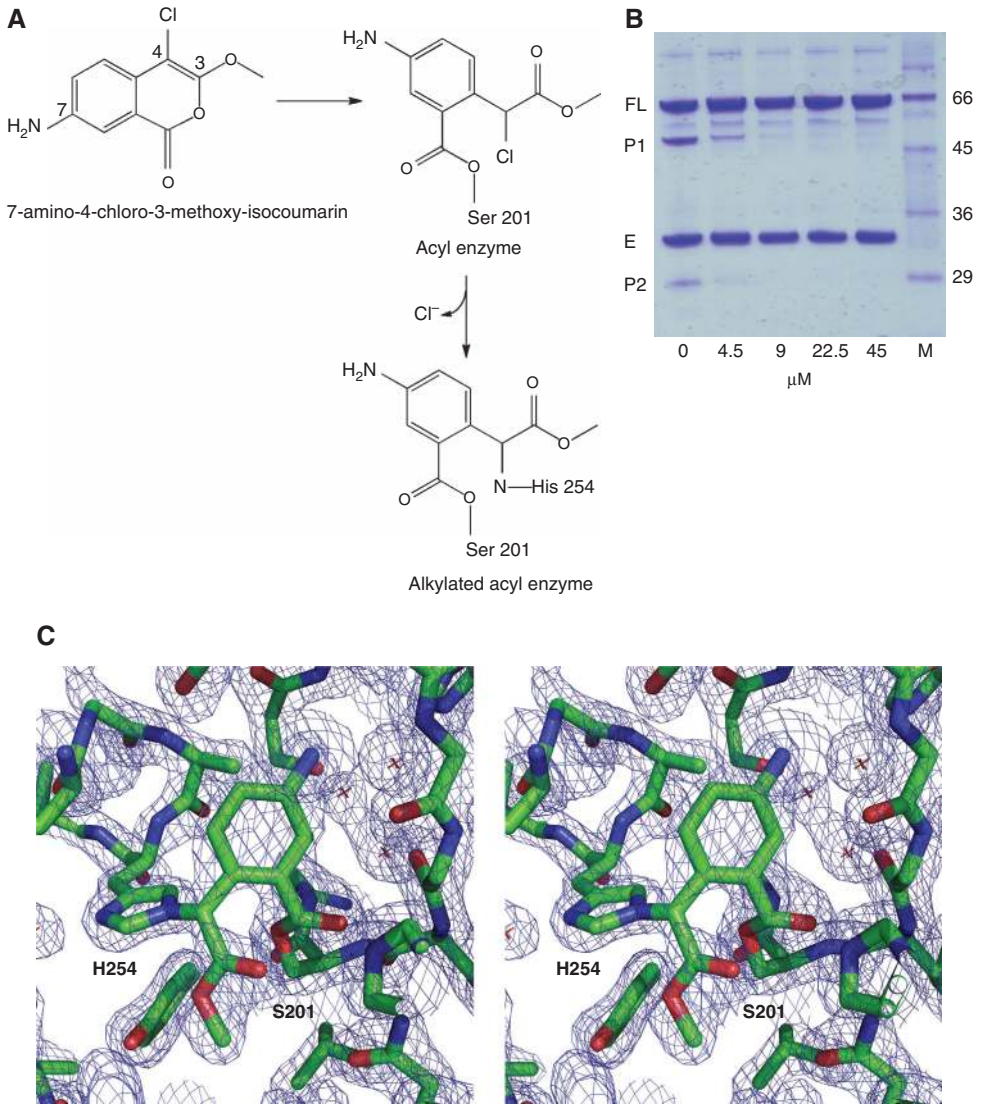

Figure 1 (**A**) Isocoumarins are heterocyclic compounds that inhibit only serine proteases. The presence of chlorine and amino group at positions 4 and 7, respectively, increases the stability of isocoumarins and substitutions at position 3 and 7 can be selectively designed to inhibit specific serine proteases. Nucleophilic attack by the active-site serine on isocoumarin results in ring opening and formation of acyl enzyme. A subsequent reaction with a nucleophile such as histidine can result in a doubly covalent bonded alkylated acyl enzyme. Figure redrawn from Powers *et al*, 2002. (**B**) *In vitro* cleavage and inhibition assay of Gurken fusion protein by GlpG. Bands are labelled as follows: FL for full-length fusion protein; E for enzyme; P1 and P2 for the two products released upon cleavage. Complete cleavage of the substrate has not been observed under these conditions. In this assay, the IC50 of amino-methoxy-isocoumarin is ~6 μM. (**C**) Stereo view of a $2F_o–F_c$ map (blue mesh) contoured at 1.5σ showing the active-site residues and the covalently bonded inhibitor in stick representation (with carbon, oxygen and nitrogen atoms coloured as green, red and blue, respectively). Water molecules are shown as red crosses.

lengths to be weaker (Figure 2D). Residues H150 and N154 are highly conserved in rhomboids. In GlpG, mutation of H150 to an alanine results in complete loss of activity, whereas the N154A mutant shows a reduction in the rate of cleavage (Baker *et al*, 2007). However, in another rhomboid, YqgP from *Bacillus subtilis*, both these residues seem less important for catalysis (Lemberg *et al*, 2005). These observations along with the inhibitor-bound structure suggest that there could be some redundancy in residues that stabilise the oxyanion, with the main chain amide of S201 having the major function.

Structural changes on inhibitor binding

A comparison of the inhibitor-bound acyl enzyme structure with that of the native enzyme without a substrate or inhibitor (PDB code 2XOV, for which an independent data set was collected; see Table I) reveals several significant structural changes (Figure 3A and B). In the native enzyme structure, loop-5, with two bulky methionine side chains (M247 and M249), caps the active site. It was observed that this loop is flexible and easily becomes disordered on soaking the crystals (Wang *et al* 2007). In the acyl enzyme structure, we observe that loop-5 is ordered (Figure 3C) and lifted away from the active site, consistent with the view that it occludes the catalytic centre in the resting state of the enzyme. Accompanied with this movement of loop-5 is a slight displacement of TM5 and TM6 that flank loop-5. The average displacement of both these helices is ~1.25 Å with the greatest deviation observed at the C-terminus of TM5 and N-terminus of TM6, in particular residues 250–252

Table I Data collection and refinement statistics

	GlpG-Native enzyme	Acyl enzyme
Data collection		
Beam line	X06SA	IO2
Space group	R32	R32
Cell dimensions		
a, *b*, *c* (Å)	110.4, 110.4, 127.8	110.6, 110.6, 122.1
γ (deg)	120	120
Resolution (Å)	55.20–1.65 (1.74–1.65)[a]	44.62–2.09 (2.20–2.09)[a]
R_{merge}	0.055 (0.575)	0.054 (0.394)
$I/\sigma I$	12.4 (2.4)	16.3 (2.9)
Completeness (%)	99.8 (100)	97.0 (85.4)
Redundancy	4.5 (4.2)	4.9 (3.5)
Refinement		
Resolution (Å)	34.77–1.65 (1.69–1.65)[a]	31.16–2.09 (2.22–2.09)[a]
No. of reflections	36 038	16 657
R_{work}/R_{free}[b]	0.192/0.218 (0.26/0.275)[a]	0.198/0.242 (0.248/0.276)[a]
No. of atoms		
Total	1733	1606
Protein	1442	1426
Heteroatoms	204	142
Water	87	38
B-factors (Å^2)		
Total	29.7	44.7
Protein	26.7	43.5
Detergent	47.57	58.2
Ligand	—	38.7
Water	38.8	47.0
RMSD		
Bond lengths (Å)	0.006	0.007
Bond angles (deg)	1.0	1.1
Maximum likelihood coordinate error (Å)	0.17	0.24

[a]Values in parentheses are for highest-resolution shell.
[b]R_{free} was calculated using a randomly selected subset of reflections (5%), remaining (95%) reflections was used for calculation of R_{work}.

(Figure 3D). We observe little change in the position of loop-1 (Figure 3A) that points away from the body of the enzyme in the plane of the lipid bilayer. This loop was initially speculated to function in substrate recruitment (Wang *et al*, 2006), but since then has been suggested to have a structural function (Baker *et al*, 2007; Wang and Ha, 2007; Wang *et al*, 2007; Bondar *et al*, 2009). Although our data do not rule out a role for loop-1 in the enzyme function, we see no evidence for its role in catalysis.

We observe two significant changes at the active site of the acyl enzyme. First, covalent binding of the inhibitor to the active-site serine displaces a water molecule that is hydrogen bonded to S201 and H150 in the native enzyme (Figure 3E and F). The benzoyl carbonyl oxygen occupies the position of this water molecule in the acyl enzyme. Second, the covalent bond between the inhibitor and H254 not only results in the displacement of the H254 side chain but also other residues in TM6 (Figure 3D). This movement also affects Y205 in TM4: a rotation of the side chain positions the aromatic ring to be perpendicular to the imidazole ring of H254 in the acyl enzyme structure as opposed to being parallel in the native enzyme (Figure 3E and F). The hydroxyl group of Y205, which hydrogen bonds to a water molecule in native enzyme structure now points towards the main chain carbonyl of W236 in the acyl enzyme structure. To accommodate this movement of Y205, the side chain of W236 in TM5 adopts a different rotamer in the acyl enzyme. Overall, the acyl enzyme structure thus suggests the structural changes that are likely to occur during catalysis, including the opening of the active site by movement of the loop-5, accompanied by small changes in the nearby TM helices.

Substrate-binding subsites in GlpG

Most proteases cleave specific substrates, and such specificity is achieved by precise complementarity between the structures of enzyme and substrate. In particular, specific grooves and cavities in the enzyme structure often represent substrate-binding subsites (Neil *et al*, 1966; Blow, 1974). Surface analysis of the rhomboid acyl enzyme structure reveals, within the hydrophilic indentation, two cavities on either side of the bound inhibitor. First of these cavities is in close proximity to the active site and the amino group of the inhibitor points into it (Figure 4A). This cavity is observed in all described rhomboid structures, and is filled with two or three water molecules (Wang *et al*, 2006; Wu *et al*, 2006; Ben-Shem *et al*, 2007; Lemieux *et al*, 2007). In the native enzyme structure the side chain of methionine 249 from loop-5 plugs this cavity (Figure 4B). The second cavity observed only in the acyl enzyme structure is located on the opposite side of the catalytic dyad and the methyl acetate group of the inhibitor points towards it. This second cavity is largely defined by residues from TM4 and TM5 (Figure 4A and B) and is hydrophobic in character. It is not fully formed in the native enzyme structure, where it resembles a shallow cleft (Figure 4A).

What do these cavities in GlpG represent? Substrate residues flanking the scissile bond are a major determinant in directing site-specific cleavage by proteases (Blow, 1974). By convention, they are defined as P_n–P_1 for the residues N-terminal to the scissile bond and P_1'–P_n' for the residues C-terminal residues to the cleaved P_1–P_1' bond (Schechter and Berger, 1967). Several rhomboid proteases, including *E. coli* GlpG have recently been shown to recognise a substrate sequence motif that defines the cleavage site (Strisovsky *et al*, 2009). Small side chains are preferred at the P_1 position of the substrate, whereas hydrophobic residues are favoured at P_4 and P_2' positions. We assume that corresponding binding sites in the enzyme (S subsites in standard protease nomenclature) are present to accommodate the side chains of these amino acids defined by the recognition motif. As the P_1 residue is the most constrained (Strisovsky *et al*, 2009), the corresponding S_1 site is expected to be most prominent, as is the case for several soluble serine proteases (Blow, 1974; Bode *et al*, 1989; Perona and Craik, 1995). In interpreting the cavities in GlpG, we further rely on two widely accepted views on substrate binding to rhomboids: (1) biological data show that rhomboids release N-terminal substrate domains into extracellular/periplasmic/luminal space (Freeman, 2008), indicating that N-terminus of GlpG substrate points towards the periplasm; and (2) structural and functional evidence suggests that rhomboid substrates interact with the enzyme through the gap between TM2 and TM5 (Baker *et al*, 2007; Ben-Shem *et al*, 2007).

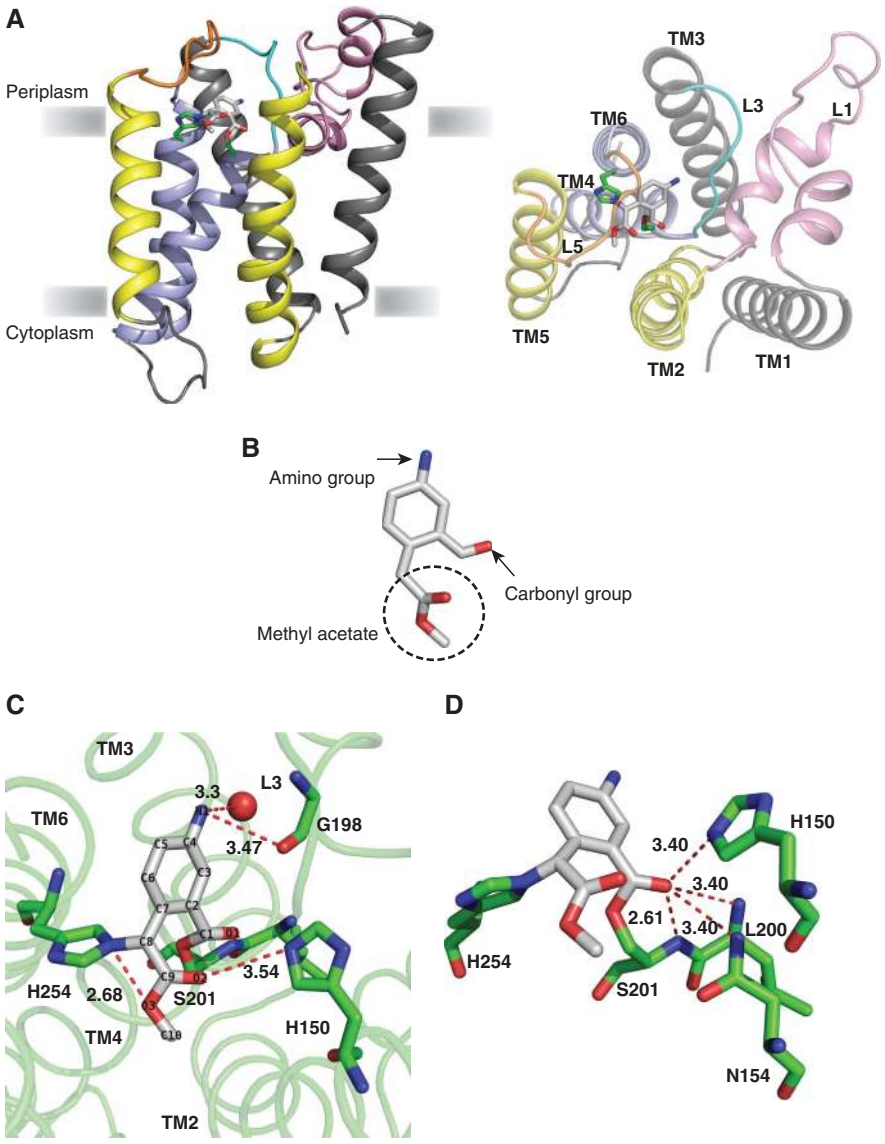

Figure 2 (**A**) Side and top view of the acyl enzyme structure of GlpG. Structural elements of GlpG that perform specific roles are colour-coded individually. TM helices that harbour the active-site serine and histidine (shown in sticks) are coloured light blue. Substrate is thought to interact with the enzyme through the gap between TM helices 2 and 5 (yellow). TM helices 1 and 3, which probably have a supporting structural function, are coloured in grey. Loop-1 (pink) has a unique fold that is partially submerged in the membrane. Its function is not clear yet. Loop-5 (orange) falls over the active site, covering it from the extracellular solution and loop-3 (cyan) could function in substrate binding. The bound inhibitor (in stick representation, with carbon atoms in white) lies in the active site, perpendicular to the membrane plane. (**B**) The structure of the inhibitor molecule after the ring has been opened by nucleophilic attack by serine; distinct groups are described in the text. (**C**) Environment of the bound inhibitor: The amino group of the inhibitor hydrogen bonds to a water molecule (red sphere) and is in close contact with the main chain carbonyl of G198. The oxygen atoms of the methyl acetate group interact with side chains of histidine H150 and H254. The methyl group of the inhibitor has no neighbours within 3.5 Å of the macromolecule. The carbon atoms of the inhibitor are coloured in white, whereas those of the protein are green. (**D**) Oxyanion hole. The benzoyl carbonyl oxygen of the inhibitor is within hydrogen bonding distance of four neighbouring residues. Of these, only the main chain amide of S201 forms a strong hydrogen bond (2.61 Å). The main chain amide of L200, the side chain amide of N154, and the imidazole nitrogen of H150 form weak hydrogen bonds (3.4 Å).

On the basis of these considerations, our data strongly indicates that the cavity in GlpG in which the amino group of the inhibitor points is the S_1 subsite of GlpG. This is consistent with the fact this cavity is observed both in the native and acyl enzyme (Figure 4). In contrast, the second cavity is observed only on inhibitor binding and its relative distance from the catalytic residues leads us to propose that this could represent the S_2' subsite (in which hydrophobic P_2' residue binds). The identity of the S_4 subsite, in which the substrate's characteristic P_4 residue binds is not clear from either of the structures described here.

A model for substrate binding in the active site of GlpG

Combining our knowledge of the recognition motifs in rhomboid substrates, the substrate-binding subsites in enzyme, and the preferred mode of isocoumarin binding in soluble serine proteases (Powers *et al*, 2002; Strisovsky *et al*, 2009), we can now infer how substrate binds at the active site of

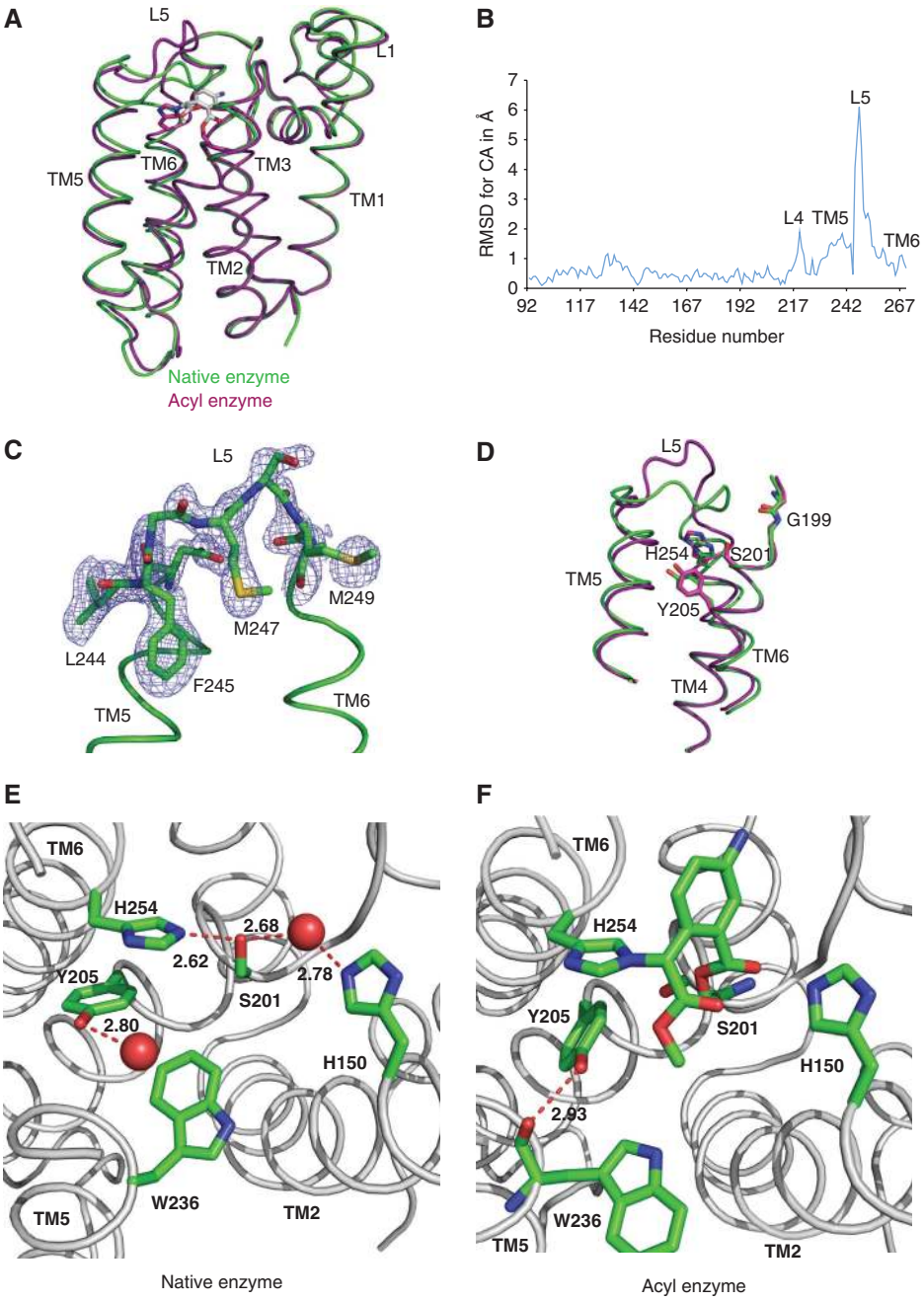

Figure 3 (**A**) A structural overlay of the native enzyme and the acyl enzyme showing the differences observed upon inhibitor binding. (**B**) A graphical representation of the root mean square deviation (RMSD) for the Cα atoms of each residue between the two structures (1.104 Å) determined with the CCP4 program Superpose (1994). Regions where the largest deviations observed are labelled. (**C**) A $2F_o$–F_c map contoured at 1σ showing the density of loop-5; residues 243–250 are shown in sticks. (**D**) An overlay of TM5 and 6, loop-5 and the active-site residues (G199, S201, Y205 and H254; in sticks) displaying the structural change. Native enzyme is coloured in green and the acyl enzyme in purple. (**E**, **F**) Active site of the native and acyl enzyme. A hydrogen-bonding network is observed in the active site of native enzyme with the active-site S201 hydrogen bonded to general base H254 and a water molecule (red sphere). The water molecule is in turn hydrogen bonded to H150, part of oxyanion hole. In the acyl enzyme, this water molecule is displaced due to the binding of inhibitor. The hydroxyl group of Y205 hydrogen bonds to a water molecule in the native enzyme, but due to the movement of H254 upon inhibitor binding, the hydroxyl now points towards the main chain carbonyl of W236.

GlpG. The orientation and the position of the benzoyl carbonyl oxygen of the inhibitor provides a guide to the probable position of the peptidyl carbonyl of the P_1 residue of the substrate—so that it is poised for a nucleophilic attack by the catalytic serine. To be consistent with the second stage of the classical serine protease catalytic mechanism, the amide of the P_1' residue of the substrate (the amino acid immediately C-terminal to the scissile bond) has been constrained in a position to be able to accept a proton from the catalytic histidine.

Proteases and substrates typically form a β-stranded interaction at the active site, frequently engaging local hydrogen bonds (Tyndall *et al*, 2005). With this in mind, we modelled a tetrapeptide of TatA (Thr-Ala-Ala-Phe), a prokaryotic

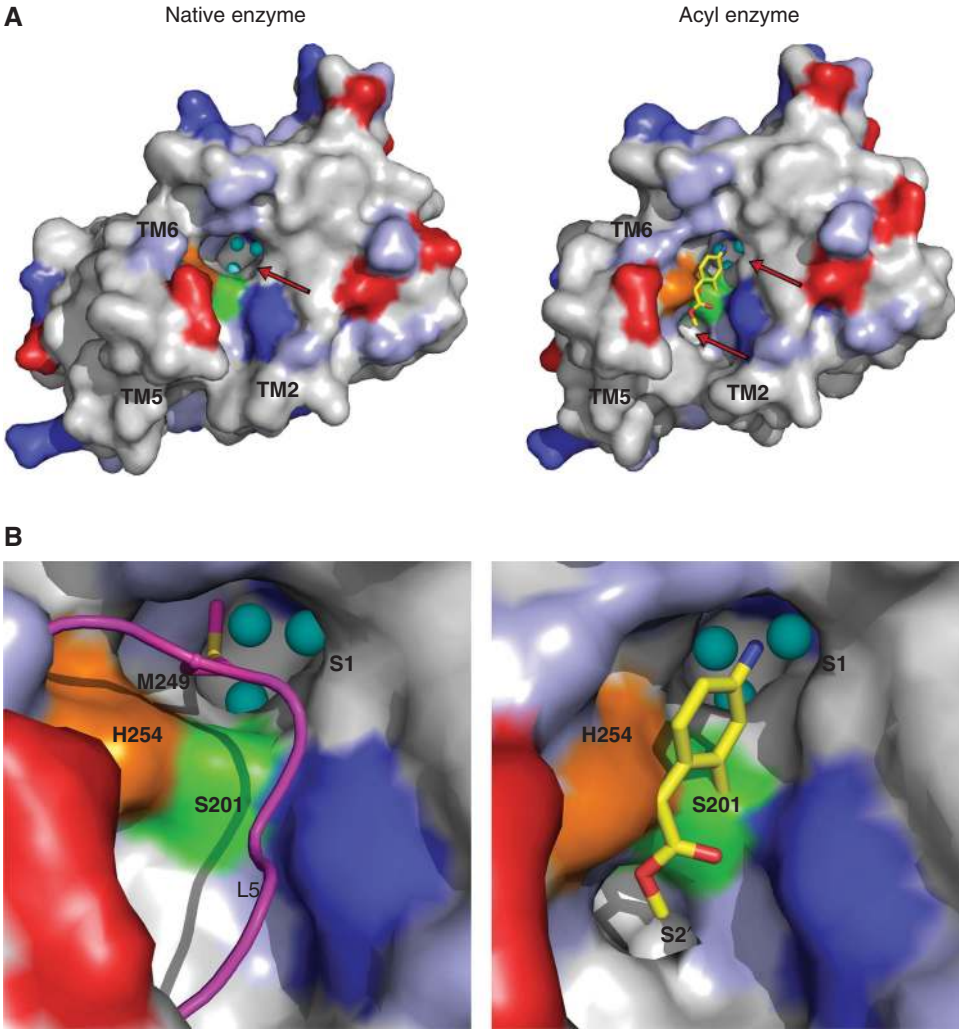

Figure 4 Surface representation of GlpG molecules showing the top view of native and acyl enzyme with loop-5 removed for clarity. (**A**) Positively and negatively charged amino acids are shown in blue and red, respectively, polar residues in light blue, and hydrophobic residues in grey. The active-site serine is coloured green and histidine is coloured orange. A hydrophilic indentation is evident in both structures, and the inhibitor molecule (in yellow stick representation) lies on it. Within this indentation the cavities observed in both structures are marked with red arrows. Water molecules are shown as cyan spheres. (**B**) Close-up view of the cavities in the native and acyl enzyme. The loop-5 in the native enzyme falls over the active site and plugs the S_1 cavity with the side chain of M249 (shown in magenta and stick representation). In the acyl enzyme, the amino group of the inhibitor points towards the S_1 cavity (loop-5 not shown for clarity). In both these structures the S_1 cavity is filled with three water molecules (cyan spheres). A second hydrophobic cavity fully formed only in the acyl enzyme could represent the S_2' cavity.

rhomboid substrate (Stevenson *et al*, 2007), into the GlpG active site, using docking and energy minimisation, manually modifying side chain rotamers, in particular those of methionines from loop-5. In our preferred model (Figure 5A), the side chain of the P_1 alanine of the TatA peptide points into the proposed S_1 cavity (Figure 5A). The side chain of phenylalanine at the P_2' position fits well into the proposed hydrophobic S_2' cavity. The aromatic ring of this phenylalanine forms contacts with W157, V204, Y205, and W236 residues from TM2, TM4 and TM5, respectively. In contrast, the side chains of the P_2 (Thr) and P_1' (Ala) residues face the solvent, which is consistent with the observed lack of specificity in these positions. With the current position of loop-5 (Figure 5B), the side chains of methionines 247 and 249 and phenylalanine 245 can form contacts with substrate (Figure 5B), suggesting that loop-5 might have a role in holding or stabilizing the substrate at the active site. As it is highly flexible, it is possible that loop-5 adapts to the substrate bound at the active site. Further substrate interactions with GlpG could occur through hydrogen bonding between the substrate and the extended loop-3 upstream of the active site. For instance, the amides of the P_1 and P_2 residues of the substrate are within hydrogen bonding distance of the carbonyl oxygen of G198 and the side chain carbonyl of Q189, respectively (Figure 5C). Importantly, this model of substrate binding is further supported by the fact that a P_2–P_2' fragment of an unrelated rhomboid substrate, *Drosophila* Gurken, with the sequence Met-Ala-His-Ile, could be modelled equally well using the same constraints (Supplementary Figure 1).

Mutations constricting S_1 cavity restrict P_1 specificity of GlpG

The depth and polarity of the proposed S_1 cavity is consistent with the observed requirement for small side chains at the P_1

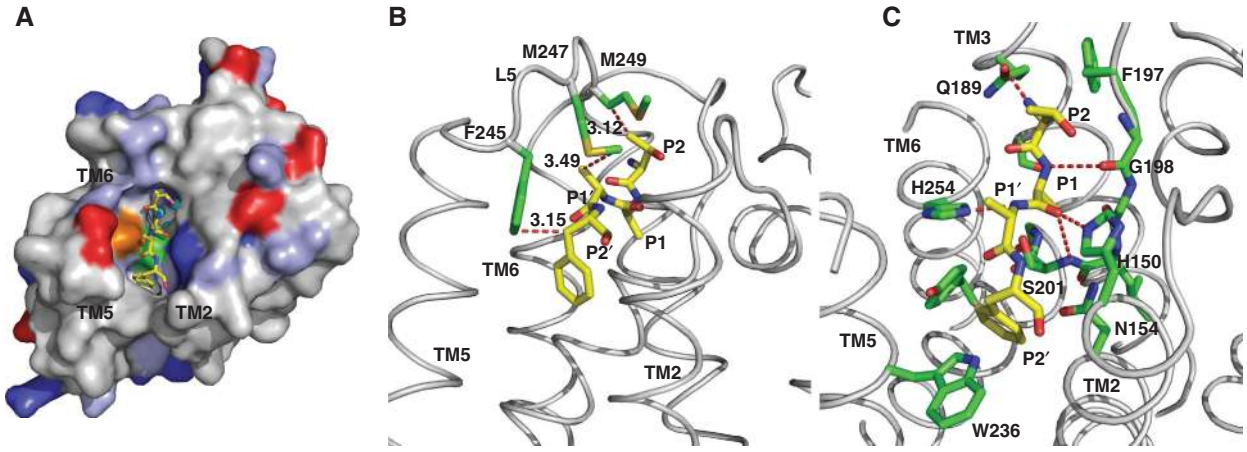

Figure 5 (**A**) A model for substrate binding at the active site of GlpG. Surface representation of GlpG (coloured as in Figure 4) and TatA tetrapeptide shown as stick representation (yellow). (**B**) Possible interaction of P_2 and P_1' side chains of the TatA substrate with loop-5. In our present best model, the side chains of residues from loop-5 (F245, M247 and M249) interact with substrate residues, in particular the side chains at the P_2 and P_1' positions. It is possible that such interactions contribute to stabilisation of the substrate at the active site. (**C**) Possible substrate–enzyme interaction in the GlpG-TatA model. The TatA tetrapeptide consists of a sequence (T-A-A-F) corresponding to the P_2 to P_2' residues. The model shows possible hydrogen bonding interaction between substrate and enzyme (red dashed lines).

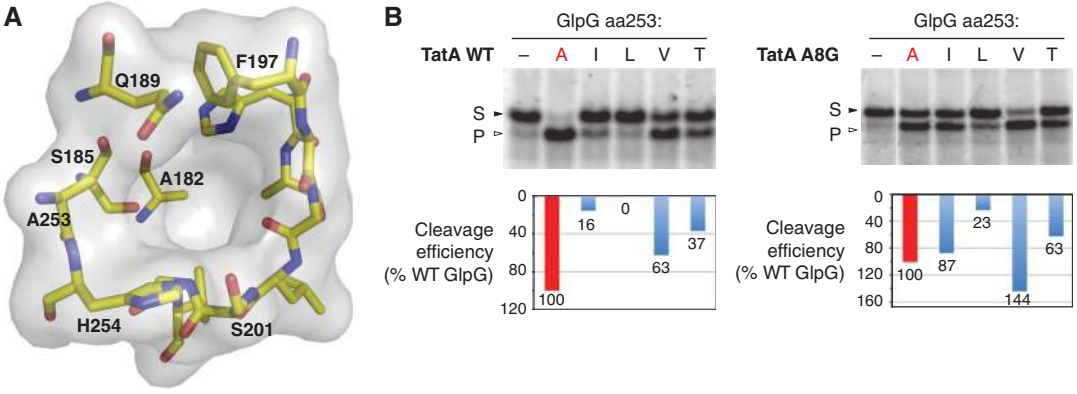

Figure 6 (**A**) S_1 cavity of GlpG: Residues from TM3, TM4, TM6 and loop-1 form the cavity found adjacent to the catalytic dyad. Some of the residues side chain of which points into the cavity are highlighted. (**B**) Mutations in A253 restrict the P_1 specificity of *E. coli* GlpG. *In vitro* translated and radiolabelled TatA and its P1 position mutant A8G (having the smallest possible side chain) were incubated with equimolar amounts of detergent-solubilised and purified wild-type GlpG (253A, red) and its A253I, L,V, and T mutants. Reaction time was 40 and 360 min for wild-type and A8G mutant of TatA, respectively. Reaction products were separated by SDS–PAGE and detected by autoradiography. Substrate conversion values were quantified by densitometry of substrate (S) and product (P) bands. The data shown here are representative of three experiments.

position, such as Cys, Ser, Ala and Gly (Strisovsky *et al*, 2009). The surface of the cavity is formed by a number of amino acids, most notably the side chains of A182, S185, Q189, F197 and A253 (Figure 6A). One prediction of our model would be that occlusion of the S_1 cavity would restrict the P_1 specificity to smaller residues, for example glycine. The residue A253 seems a particularly suitable target for mutation, because its side-chain protrudes directly into the entrance of the cavity. Mutation of A253 into valine, threonine, isoleucine or leucine, designed to sterically interfere with the access to the S_1 cavity, had a progressive inhibitory effect on cleavage of wild-type TatA, which contains alanine in the P_1 position. In contrast, these mutations had much less effect on the cleavage of the TatA A8G mutant that contains glycine, the residue with the smallest possible side-chain, in the P_1 position (Figure 6B). The strongest differential effect is observed with GlpG_A253I that has only 16% of wild-type GlpG activity on wild-type TatA, but 87% of wild-type GlpG activity on TatA_A8G. This is consistent with the partial occlusion of the S_1 cavity by the A253I mutation preventing binding of the wild-type substrate, but still allowing binding of a mutant with a P_1 glycine. These results support our model of substrate binding at the active site and rationalise the requirement for smaller amino acids at the P_1 position of substrates cleaved by GlpG.

Discussion

We have determined the structure of an acyl enzyme of a rhomboid protease, thereby finally proving that, despite being phylogenetically unrelated and structurally very different to the classical serine proteases, rhomboids have convergently evolved a serine protease mechanism that resembles that of their better-understood counterparts, although with some significant differences. A hallmark of serine proteases is the oxyanion hole, typically defined by the main chain amides of the catalytic serine and a neighbouring glycine, but variations have been found where the side chain

of asparagine or the hydroxyl of serine can also take part in stabilizing the oxyanion (Robertus *et al*, 1972; Bryan *et al*, 1986; Carlos *et al*, 2000). On the basis of the native GlpG enzyme structure, it has been suggested that a structural rearrangement at the active site might result in the amide of G199 pointing into the active site (Wang *et al*, 2006), but the acyl enzyme structure provides no evidence for this. The fact that the carbonyl oxygen of the inhibitor makes a good hydrogen bond to the main chain amide of the catalytic S201 supports the idea that the inhibitor mimics a substrate peptide bond. In turn, this makes a strong case that other nearby residues, like the main chain amide of L200 or the side chains of H150 or N154, have a role in the stabilisation of oxyanion hole. In addition, the importance of N154 and H150 in retaining GlpG activity (Baker *et al*, 2007) leads us to propose that either of these residues could substitute for chymotrypsin G193 in the oxyanion hole.

Some of the structural changes observed in the acyl enzyme, such as the lifting of the loop-5 or the movement of TM5, were anticipated, but others, such as the movement of Y205 or TM6, were not. Interestingly, the aromatic ring of Y205 has been proposed to stabilise the position of the imidazole ring of the catalytic histidine (Wang *et al*, 2006), although the significance of its rotation in the acyl enzyme is presently unclear. It is possible that such a rotation, accompanied by movement of TM helices, is required for the formation of the S_2' site. The second covalent bond of the inhibitor to the catalytic histidine does not occur during normal catalysis, and could potentially lead to changes in the protein, such as those observed in complexes of isocoumarin with elastase and trypsin (Supplementary Figure 2; Meyer *et al*, 1985; Chow *et al*, 1990; Powers *et al*, 1990; Vijayalakshmi *et al*, 1991). The displacement of TM6 in the acyl enzyme structure is probably an example of such a change. It will therefore be important to obtain a structure with an inhibitor that forms only a single covalent bond to the nucleophilic serine to verify the movement of TM6.

Conformational changes that might occur during rhomboid catalysis have been controversial (Ha, 2009; Urban, 2009). Although the overall architecture of the multiple reported *E. coli* GlpG structures (Wang *et al*, 2006; Wu *et al*, 2006; Ben-Shem *et al*, 2007) is similar, variations in TM5 and loop-5 are observed, sometimes even within two molecules in the same crystal (Figure 7A and B; Supplementary Table 1). The differences in TM5 are particularly pronounced, varying from an ~1.5-Å displacement in one molecule to a tilt of 35° in another (Figure 7A and B). Yet another variation is found in the structure of GlpG from *H. influenzae* (Lemieux *et al*, 2007), in which TM5 adopts a different conformation but loop-5 still points to the active site (Figure 7C). Such variations combined with biochemical studies have led to the proposal that large movement of TM5 occurs during catalysis. The molecule with the most extreme tilt of the C-terminus of TM5 (molecule A of 2NRF) has been called the open conformation of enzyme, whereas other structures with less apparent movement of TM5 have been proposed to be intermediate states (for review, see Urban, 2009). It is, however, notable that none of these structures of empty enzymes resemble the conformation of the TM5/loop-5, which we observe for the inhibitor-bound enzyme (Figure 7). This leads us to suspect that at least the more extreme variations in TM5 position in these published structures are an effect of the detergent environment and crystal contacts. Consistent with this, the structural change and the movement of TM5 in the acyl enzyme structure are very subtle, confined to displacement by a few angstroms. Of course, it is possible that soaking inhibitor into preformed crystals may restrict the structural change that can occur in the enzyme, and in the absence of a structure of a full enzyme–substrate complex, it is impossible to know for sure whether extreme displacement of TM5 is biologically relevant.

On the basis of the inhibitor-bound structure described here, we propose a relatively simple mechanism for substrate binding and catalysis by rhomboids. In the resting, or closed state of the enzyme, described previously (Wang *et al*, 2006) and confirmed by us, loop-5 physically caps the active site of GlpG, and the side chains of methionine (247 and 249) points into the active site. The orientation of the inhibitor, in particular the methyl group pointing towards the putative bilayer and small displacement of TM5, add to the growing view that a rhomboid substrate, a type I or III TM protein, interacts with the enzyme at the interface between TM2/TM5

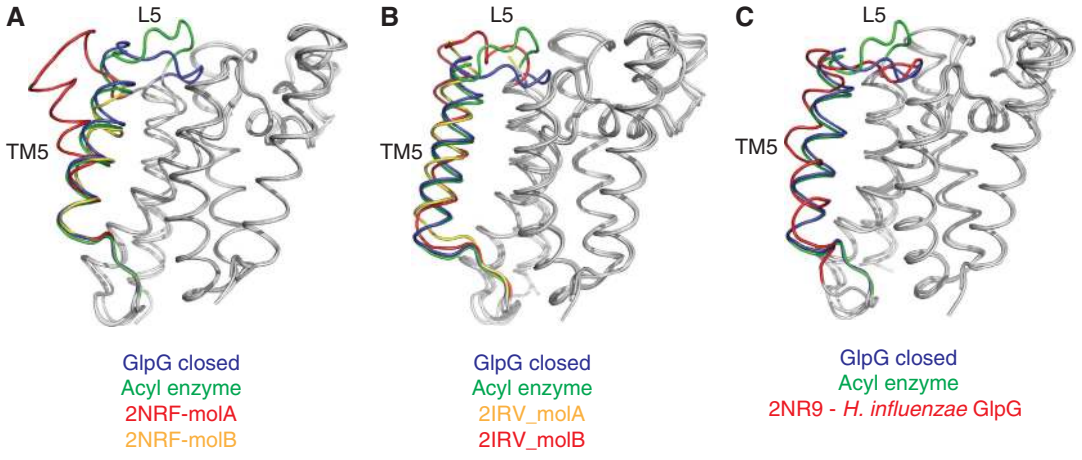

Figure 7 Panels (**A**) and (**B**) show *E. coli* GlpG crystallised in space groups P31 and P21, respectively. In both these structures there are two molecules in the asymmetric unit (red and yellow) that differ significantly. Panel (**C**) shows GlpG from *H. influenzae* crystallised in space group C2. The 'closed' state of GlpG (PDB coordinates 3B45 and 2XOV) is shown in blue in all panels. It has been proposed that the molecule of GlpG in which TM5 is very tilted (2NRF-molA, shown in red in panel (**A**) is called the open conformation. However, the present acyl enzyme structure (pdb coordinates 2XOW), shown in green in all panels, reveals that even in the presence of mechanism-based inhibitor such a large change is not seen.

(Baker *et al*, 2007; Ben-Shem *et al*, 2007; for review, see Lemberg and Freeman, 2007a). This initial interaction between the substrate and the enzyme is likely to involve the displacement of loop-5, followed by interaction of the recognition motif of the substrate with the active site of GlpG, defined by the hydrophilic indentation above the catalytic dyad (Figures 4 and 5). This interaction requires flexibility of the polypeptide chain in the region around the cleavage site. If the residues of the scissile bond (P_1–P_1') are within the transmembrane (TM) domain, then their alpha helical conformation need to be destabilised, presumably promoted by the presence of TM-helix-destabilising residues (Urban and Freeman, 2003; Maegawa *et al*, 2007; Strisovsky *et al*, 2009). However, if the cleavage site is outside the membrane, TM-helix-destabilising residues become less important. In this case, the juxtamembrane segment linking the recognition motif and the TM domain probably provides the necessary flexibility (Strisovsky *et al*, 2009).

Although the sequence of events described above provides a simple view of substrate binding and catalysis in rhomboids, there are several key facts that remain to be addressed. Our inhibitor-bound structure of GlpG shows that the movement of TM5 is probably necessary but it does not tell us when it occurs. Does the initial substrate binding trigger a conformational change or is it caused by binding at the active site? This will become clear when a structure with a substrate TM domain or perhaps an isocoumarin inhibitor with a longer hydrophobic group at position 3 (Figure 1) is determined. At present, we also cannot determine the exact path of the N-terminal region of the substrate, which contains the characteristic substrate P_4 residue. A specific cavity is not evident in this region but based on the current peptide model and distance constraints, we suggest that it might involve residues from C-terminal part of TM3, the N-terminal residues of loop-3, and probably loop-1. A role for residues from loop-1, either in interaction with the substrate or as a supporting structural feature to loop-3, would be consistent with loss of activity observed in many mutants at L1–L3 interface (Baker *et al*, 2007).

Rhomboids that have been studied are quite specific: they cleave only a small subset of TM domains (Urban *et al*, 2002; Urban and Freeman, 2003; Lohi *et al*, 2004; Baker *et al*, 2006; Strisovsky *et al*, 2009). On the basis of the structure and biochemical data, we can now identify the S_1 subsite of the enzyme, and the probable location for the S_2' subsite in GlpG. The substrate recognition motif first identified in the TatA rhomboid substrate is also used by diverse rhomboids (Strisovsky *et al*, 2009). However, it is conceivable that other members of the rhomboid family may have a different or less stringent recognition motif. For example, cleavage sites in the mitochondrial rhomboid substrates do not seem to contain a TatA-like recognition motif (Herlan *et al*, 2003; Tatsuta *et al*, 2007), and widely different sequences are apparently accepted at their cleavage site (Schafer *et al*, 2010). It is likely that different rhomboid specificities will be determined by the size, shape, and position of their S_1, S_2', and S_4 (or any other) subsites. This study therefore provides a foundation for a structural understanding of biological specificity of rhomboid proteases.

Even with only a relatively few cases having been well studied, it is clear that rhomboids control a wide range of biologically important events. Moreover, there are increasing indications of their potential medical relevance in areas as diverse as protozoan parasite infection, bacterial infection, mitochondrial diseases, diabetes, and cancer. Although there is not yet a fully validated therapeutic opportunity, the case for identifying specific and potent inhibitors of rhomboids—both as research tools and as leads for possible future drug development—is strong. Understanding the enzyme mechanism and the architecture of the active site covalently bound to an isocoumarin provides an important foundation for the design and optimisation of such inhibitors.

Materials and methods

Protein expression and purification

Full-length GlpG in pET25b vector with a C-terminal His tag was expressed in *E. coli* BL21 C43(DE3) cells (Miroux and Walker, 1996) in 2YT medium. Cells were induced with 0.5 mM isopropyl thiogalactoside (IPTG) at an OD value of 0.6. At this point, the temperature was reduced to 24°C and cells were grown overnight. Cells were collected and broken by two passages in an Emulsiflex (Avestin Systems). Unbroken cells were removed by low-speed centrifugation. The membrane fraction was collected by centrifugation at 100 000 *g* at 4°C for 90 min, and resuspended in 25 mM Tris (pH 8). Membranes at a final protein concentration of 5 mg/ml were solubilised in 1.25 % β-D-decyl maltoside (DM, Anatrace) for 45 min at room temperature (RT). The soluble fraction was collected by centrifugation at 100 000 *g* for 30 min. NaCl and imidazole were added to the soluble fraction at a final concentration of 0.3 M and 10 mM, respectively, which was then applied to a Ni–NTA column (Qiagen) pre-equilibrated with Tris–HCl (pH 8.0), 300 mM NaCl, and 10 mM imidazole containing 0.2% DM. Two washing steps with 10 mM and 30 mM imidazole removed nonspecifically bound proteins. GlpG was eluted with 0.2 M imidazole. The N-terminal domain of GlpG was removed by digestion with chymotrypsin for 36 h at a ratio of 1:50 (w/w; Wang *et al*, 2006). After cleavage, buffer was exchanged to reduce the ionic strength and truncated protein was passed through a Mono-Q (GE Healthcare) ion exchange column pre-equilibrated with 25 mM Tris (pH 8.0) and 0.2% DM. The majority of the truncated GlpG did not bind to column. It eluted in the flow-through peak and allowed much of the N-terminal cleaved product to be separated. GlpG without the N-terminal domain was concentrated with a 10-kDa MWCO concentrator (Vivaspin) and loaded onto a Superdex-200 (16/60) column pre-equilibrated with 25 mM Tris (pH 8.0), 0.15 M NaCl and 0.5% β-D-nonyl glucoside (NG, Anatrace). Peak fractions at ~13 ml were collected and concentrated to 7–8 mg/ml before crystallisation.

Rhomboid activity assay and inhibition

The TM domain of Gurken, a rhomboid substrate from *Drosophila*, was expressed as a fusion protein with an N-terminal maltose-binding protein domain and a C-terminal thioredoxin domain and a His tag (construct pKS34 from Strisovsky *et al*, 2009) in *E. coli* C43(DE3). Expression, membrane preparation, solubilisation, and purification with nickel affinity column were carried out as described above for GlpG. In a typical rhomboid assay, full-length GlpG (1.3 μM) and substrate (1.45 μM) in a 10 μl reaction in 25 mM HEPES buffer (pH 7.5), 0.15 M NaCl, and 0.2% DM were incubated at 37°C for 2 h. Cleavage products were separated on a 10% bis-Tris gel (Invitrogen) developed with MES buffer and bands stained with Coomassie brilliant blue. For inhibition assays, freshly diluted inhibitor at a given concentration was incubated with enzyme (1.3 μM) for 30 min at RT before the addition of substrate. Gels were scanned and bands quantified with ImageQ software (Amersham Pharmacia).

Cloning, purification, and radioactive assay to analyse GlpG S_1 cavity mutants

Point mutations were introduced by the QuikChange method (Stratagene) and all mutant constructs were verified by sequencing the whole open reading frames. Full-length wild-type GlpG and its A253 mutant variants were expressed in *E. coli* C43(DE3), and purified from the membrane fraction in the presence of n-dodecyl-β-

D-maltoside (DDM) as described previously (Lemberg *et al*, 2005, Stevenson *et al*, 2007). Radiolabelled TatA, extended at the N-terminus by a $(GS)_4$ tag, and its A8G mutant were generated by *in vitro* transcription and translation as described previously (Strisovsky *et al*, 2009). For activity comparison of GlpG A253 mutants, the purified enzymes (0.5 μg) were incubated at equimolar concentrations with 2 μl of radiolabelled substrate in a total of 20 μl of reaction buffer (50 mM HEPES–NaOH (pH 7.5), 0.4 M NaCl, 5 mM EDTA, 10% (v/v) glycerol, and 0.05% (w/v) DDM) at 37°C. After the indicated time, the reaction products were precipitated by 10% trichloroacetic acid, separated on 10% Bis-Tris–MES SDS–PAGE, and autoradiographed. Substrate conversion was quantified by densitometry as described previously (Strisovsky *et al*, 2009).

Crystallisation

Our attempts to use constructs corresponding to the truncated version of GlpG often resulted in poorly diffracting crystals, but GlpG truncated with chymotrypsin invariably gave well-diffracting crystals. Crystals of truncated GlpG WT were obtained by mixing a solution of 2.5–3 M ammonium chloride with protein at ratio of 1:1 in hanging drops at 25°C. Numerous inhibitors that have been shown to react with soluble serine proteases were tried extensively under many different conditions. We observed that 7-amino-4-chloro-3-methoxy isocoumarin (available as JLK-6 from Tocris) showed most promise. Inhibitor was dissolved in 100% DMSO at a concentration of 45 mM and stored at −20°C. For soaking, JLK-6 was first diluted to 4.5 mM in buffer resembling the mother liquor (25 mM Bis-Tris (pH 7.0) and 2.5 M ammonium chloride) with 20% DMSO. Then, single crystals were incubated in a diluted solution with a final inhibitor and DMSO concentration of 1 mM and 7.5%, respectively, for 20 h at 25°C. All crystals were cryo-protected by adding 25% glycerol to mother liquor and flash-frozen in liquid nitrogen.

Crystallography

Numerous data sets of GlpG were collected at the I02/I03/I04 beam lines at the Diamond Light Source (Harwell) and at X06SA beam line at the Swiss Light Source (Villgen) to verify whether the enzyme had reacted with inhibitor. Diffraction data were indexed and integrated with imosflm and reduced with SCALA (Powell, 1999; Evans, 2006; Leslie, 2006). For the structure with inhibitor bound, the coordinates of GlpG (PDB -3B45) with the active-site residues serine, histidine, and loop-5 corresponding to residues 245–249 were omitted as an input model for Phaser (McCoy *et al*, 2007). A single solution with an RFZ = 10.1, TFZ = 34.9, and LLG = 1372 was obtained. After rebuilding the active-site serine and histidine, rigid body refinement with simulated annealing was carried out with PHENIX (Adams *et al*, 2010), followed by manual model building in COOT (Emsley and Cowtan, 2004) and refinement of individual sites and B factors. Model and library files of ring-opened isocoumarin were obtained from the PRODRG (Schuttelkopf and van Aalten, 2004) server and link files were generated with Jligand. The final model of the inhibitor of GlpG consists of residues 92–270, 16 partially ordered detergent molecules, and 38 water molecules.

For the structure determination of native enzyme, we used coordinates of GlpG (PDB -3B45) with loop-5 corresponding to residues 245–249 omitted as an input model for Phaser, followed by rigid body refinement and individual sites and B factors by PHENIX.

Peptide modelling

The coordinates of the acyl enzyme were used for modelling a substrate at the active site. A tetrapeptide corresponding to the P_2 to P_2' site of TatA (Thr-Ala-Ala-Phe) or Gurken (Met-Ala-His-Ile) was manually docked using the constraints that the carbonyl oxygen of scissile bond points towards the oxyanion hole and the amide of P_1' is within hydrogen bonding distance to His254. The resulting coordinates were energy minimised with reduced weight from the X-ray data in PHENIX. All the figures were generated with Pymol (Delano, 2002).

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (http://www.embojournal.org).

Acknowledgements

We thank Richard Henderson for his constant support and advice on many aspects of this project; Andrew Leslie and Jade Li for advice on crystallography; Ya Ha for sharing his experience on crystallisation of GlpG; Werner Kuhlbrandt for his support to VK at the initial stage of the project at the Max-Planck Institute of Biophysics in Frankfurt; and all colleagues at the LMB for critical reading of the paper. We are particularly grateful to Matt Bogyo (Stanford University) for generously providing access to the isocoumarin library that he compiled from compounds synthesised in several laboratories, and to Olivier Pierrat for his help with sourcing JLK-6. We thank the beam line staff at the Diamond Light source and the Swiss light source for their excellent support. VK acknowledges the European Union for a Marie Curie Intraeuropean Fellowship and the Medical Research Council for a career developmental fellowship. KS acknowledges support from a Marie Curie Intraeuropean Fellowship, an EMBO Long-Term Fellowship, and a Medical Research Council Career Development Fellowship. The coordinates of the native and acyl enzyme have been deposited at the PDB under the code 2xov and 2xow.

Author contributions: VK conceived and designed experiments, performed all structural analysis, analysed results and wrote the paper; KS conceived and designed experiments, performed the biochemical analysis of structural mutants, analysed results and wrote the paper; AA performed *in silico* analysis of the GlpG active site, predicting residues that contribute to the S_1 cavity; YC and SV identified amino-methoxy-isocoumarin as an efficient rhomboid inhibitor; MF conceived and designed experiments, analysed results and wrote the paper.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66** (Part 2)**:** 213–221

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

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

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

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

Blow D (1974) Structure and Substrate Specificity in the serine enzymes. *Isr J Chem* **12:** 483–494

Bode W, Meyer Jr E, Powers JC (1989) Human leukocyte and porcine pancreatic elastase: X-ray crystal structures, mechanism, substrate specificity, and mechanism-based inhibitors. *Biochemistry* **28:** 1951–1963

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

Bryan P, Pantoliano MW, Quill SG, Hsiao HY, Poulos T (1986) Site-directed mutagenesis and the role of the oxyanion hole in subtilisin. *Proc Natl Acad Sci USA* **83:** 3743–3745

Carlos JL, Klenotic PA, Paetzel M, Strynadka NC, Dalbey RE (2000) Mutational evidence of transition state stabilization by serine

88 in *Escherichia coli* type I signal peptidase. *Biochemistry* **39:** 7276–7283

Chow MM, Meyer Jr EF, Bode W, Kam CM, Radhakrishnan R, Vijayalakshmi J, Powers JC (1990) The 2.2 Åresolution X-ray crystal structure of the complex of trypsin inhibited by 4-chloro-3-ethoxy-7-guanidinoisocoumarin: a proposed model of the thrombin–inhibitor complex. *J Am Chem Soc* **112:** 7783–7789

Cipolat S, Rudka T, Hartmann D, Costa V, Serneels L, Craessaerts K, Metzger K, Frezza C, Annaert W, D'Adamio L, Derks C, Dejaegere T, Pellegrini L, D'Hooge R, Scorrano L, De Strooper B (2006) Mitochondrial rhomboid PARL regulates cytochrome c release during apoptosis via OPA1-dependent cristae remodeling. *Cell* **126:** 163–175

De Strooper B, Annaert W, Cupers P, Saftig P, Craessaerts K, Mumm JS, Schroeter EH, Schrijvers V, Wolfe MS, Ray WJ, Goate A, Kopan R (1999) A presenilin-1-dependent gamma-secretase-like protease mediates release of Notch intracellular domain. *Nature* **398:** 518–522

Delano WL (2002) *The PyMOL Molecular Graphics System*, (http://www.pymol.org)

Ekici OD, Paetzel M, Dalbey RE (2008) Unconventional serine proteases: variations on the catalytic Ser/His/Asp triad configuration. *Protein Sci* **17:** 2023–2037

Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60** (Part 12, Part 1)**:** 2126–2132

Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62**(Part 1)**:** 72–82

Freeman M (2008) Rhomboid proteases and their biological functions. *Annu Rev Genet* **42:** 191–210

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

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

Hartley BS, Kilby BA (1950) Inhibition of chymotrypsin by diethyl p-nitrophenyl phosphate. *Nature* **166:** 784–785

Hedstrom L (2002) An overview of serine proteases. *Curr Protoc Protein Sci Chapter* **21:** Unit 21 10

Henderson R (1970) Structure of crystalline alpha-chymotrypsin. IV. The structure of indoleacryloyl-alpha-chyotrypsin and its relevance to the hydrolytic mechanism of the enzyme. *J Mol Biol* **54:** 341–354

Herlan M, Vogel F, Bornhovd C, Neupert W, Reichert AS (2003) Processing of Mgm1 by the rhomboid-type protease Pcp1 is required for maintenance of mitochondrial morphology and of mitochondrial DNA. *J Biol Chem* **278:** 27781–27788

Howell SA, Hackett F, Jongco AM, Withers-Martinez C, Kim K, Carruthers VB, Blackman MJ (2005) Distinct mechanisms govern proteolytic shedding of a key invasion protein in apicomplexan pathogens. *Mol Microbiol* **57:** 1342–1356

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

Lee JR, Urban S, Garvey CF, Freeman M (2001) Regulated intracellular ligand transport and proteolysis control EGF signal activation in *Drosophila*. *Cell* **107:** 161–171

Lemberg MK, Freeman M (2007a) Cutting proteins within lipid bilayers: rhomboid structure and mechanism. *Mol Cell* **28:** 930–940

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

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

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

Leslie AG (2006) The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr* **62**(Part 1)**:** 48–57

Lohi O, Urban S, Freeman M (2004) Diverse substrate recognition mechanisms for rhomboids; thrombomodulin is cleaved by mammalian rhomboids. *Curr Biol* **14:** 236–241

Maegawa S, Koide K, Ito K, Akiyama Y (2007) The intramembrane active site of GlpG, an *E. coli* rhomboid protease, is accessible to water and hydrolyses an extramembrane peptide bond of substrates. *Mol Microbiol* **64:** 435–447

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**(Part 4)**:** 658–674

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

Meyer Jr EF, Presta LG, Radhakrishnan R (1985) Steresospecific reaction of 3-methoxy-4-chloro-7-aminoisocoumarin with crystalline porcine pancreatic elastase. *J Am Chem Soc* **107:** 4091–4093

Miroux B, Walker JE (1996) Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol* **260:** 289–298

Neil GL, Niemann C, Hein GE (1966) Structural specificity of alpha-chymotrypsin: polypeptide substrates. *Nature* **210:** 903–907

O'Donnell RA, Hackett F, Howell SA, Treeck M, Struck N, Krnajski Z, Withers-Martinez C, Gilberger TW, Blackman MJ (2006) Intramembrane proteolysis mediates shedding of a key adhesin during erythrocyte invasion by the malaria parasite. *J Cell Biol* **174:** 1023–1033

Opitz C, Di Cristina M, Reiss M, Ruppert T, Crisanti A, Soldati D (2002) Intramembrane cleavage of microneme proteins at the surface of the apicomplexan parasite *Toxoplasma gondii*. *EMBO J* **21:** 1577–1585

Perona JJ, Craik CS (1995) Structural basis of substrate specificity in the serine proteases. *Protein Sci* **4:** 337–360

Polgar L (2005) The catalytic triad of serine peptidases. *Cell Mol Life Sci* **62:** 2161–2172

Powell HR (1999) The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta Crystallogr D Biol Crystallogr* **55** (Part 10)**:** 1690–1695

Powers JC, Asgian JL, Ekici OD, James KE (2002) Irreversible inhibitors of serine, cysteine, and threonine proteases. *Chem Rev* **102:** 4639–4750

Powers JC, Oleksyszyn J, Narasimhan SL, Kam CM (1990) Reaction of porcine pancreatic elastase with 7-substituted 3-alkoxy-4-chloroisocoumarins: design of potent inhibitors using the crystal structure of the complex formed with 4-chloro-3-ethoxy-7-guanidinoisocoumarin. *Biochemistry* **29:** 3108–3118

Rawson RB, Cheng D, Brown MS, Goldstein JL (1998) Isolation of cholesterol-requiring mutant Chinese hamster ovary cells with defects in cleavage of sterol regulatory element-binding proteins at site 1. *J Biol Chem* **273:** 28261–28269

Robertus JD, Kraut J, Alden RA, Birktoft JJ (1972) Subtilisin; a stereochemical mechanism involving transition-state stabilization. *Biochemistry* **11:** 4293–4303

Schafer A, Zick M, Kief J, Steger M, Heide H, Duvezin-Caubet S, Neupert W, Reichert AS (2010) Intramembrane proteolysis of Mgm1 by the mitochondrial rhomboid protease is highly promiscuous regarding the sequence of the cleaved hydrophobic segment. *J Mol Biol* **401:** 182–193

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

Schuttelkopf AW, van Aalten DM (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr* **60**(Part 8)**:** 1355–1363

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

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

Tatsuta T, Augustin S, Nolden M, Friedrichs B, Langer T (2007) m-AAA protease-driven membrane dislocation allows intramembrane cleavage by rhomboid in mitochondria. *EMBO J* **26:** 325–335

The CCP4 suite: programs for protein crystallography (1994) *Acta Crystallogr D Biol Crystallogr* **50**(Part 5): 760–763

Turk B (2006) Targeting proteases: successes, failures and future prospects. *Nat Rev Drug Discov* **5:** 785–799

Tyndall JD, Nall T, Fairlie DP (2005) Proteases universally recognize beta strands in their active sites. *Chem Rev* **105:** 973–999

Urban S (2009) Taking the plunge: integrating structural, enzymatic and computational insights into a unified model for membrane-immersed rhomboid proteolysis. *Biochem J* **425:** 501–512

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

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

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

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

Vijayalakshmi J, Meyer Jr EF, Kam CM, Powers JC (1991) Structural study of porcine pancreatic elastase complexed with 7-amino-3-(2-bromoethoxy)-4-chloroisocoumarin as a nonreactivatable doubly covalent enzyme-inhibitor complex. *Biochemistry* **30:** 2175–2183

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

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

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

Wasserman JD, Urban S, Freeman M (2000) A family of rhomboid-like genes: *Drosophila* rhomboid-1 and roughoid/rhomboid-3 cooperate to activate EGF receptor signaling. *Genes Dev* **14:** 1651–1663

Weihofen A, Binns K, Lemberg MK, Ashman K, Martoglio B (2002) Identification of signal peptide peptidase, a presenilin-type aspartic protease. *Science* **296:** 2215–2218

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

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