

# Membrane-embedded protease poses for photoshoot

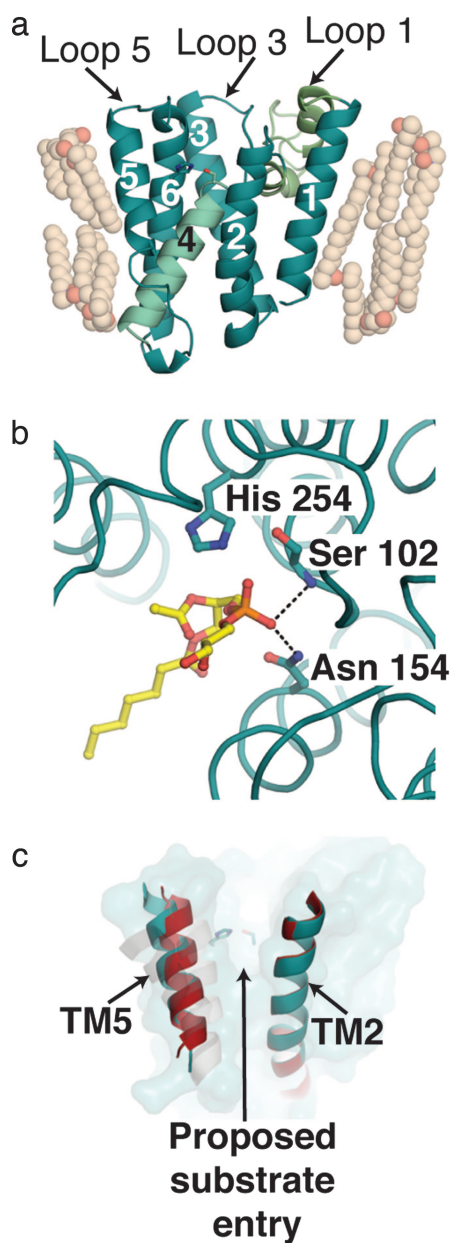
Raquel L. Lieberman and Michael S. Wolfe\*

Center for Neurologic Diseases, Brigham and Women's Hospital, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115

Intramembrane-cleaving proteases (I-CLiPs) are a class of hydrolytic enzymes that cleave hydrophobic substrates within the lipid bilayer (1). Unlike their well understood soluble or membrane-tethered counterparts, I-CLiPs are integral membrane proteins, with catalytic residues thought to lie within the confines of the membrane. Although the identities of these residues have suggested membrane-embedded active sites and mechanistic convergence with other proteases, I-CLiPs bear little or no sequence similarity to their soluble cousins, and their assignment into mechanistic categories has been tentative. Three new articles (2–4), including one by Ben-Shem *et al.* in this issue of PNAS (4), describe the first crystal structures of an I-CLiP, the *Escherichia coli* serine protease GlpG. Together, these reports provide complementary pictures that offer exciting insights into how an I-CLiP handles transmembrane substrates and carries out hydrolysis in a hydrophobic environment.

The known families of I-CLiPs include the site-2 protease (S2P) zinc metalloproteases, presenilin and presenilin-like aspartyl proteases, and the rhomboid family of serine proteases. These enzymes are found throughout the major kingdoms of life, are highly conserved, and play key roles in biology and disease. S2P-type proteases release membrane-tethered transcription factors controlling sterol and fatty acid biosynthesis in animals and sporulation in bacteria (5, 6). Presenilin is the catalytic engine of the four-component  $\gamma$ -secretase complex critical for cell-fate determinations and the pathogenesis of Alzheimer's disease (7). Signal peptide peptidase, the prototype presenilin-like protease, releases remnant signal peptides from the membrane but, unlike presenilin, does not require other protein cofactors (8).

*E. coli* GlpG is a homolog of *Drosophila* rhomboid-1, which was the first identified intramembrane serine protease (9). Rhomboid-1 releases the extracellular domain of the membrane-tethered epidermal growth factor-like protein Spitz as part of a key developmental signaling pathway in flies. The specific physiological roles of most rhomboids, including the seven homologs identified in the human genome and the many of bacterial and parasitic origin, are largely unknown, but re-



**Fig. 1.** Views of GlpG. (a) Molecular architecture of *E. coli* GlpG [Protein Data Bank (PDB) ID code 2IRV]. TMs are numbered sequentially. TM4, which contains the catalytic residue Ser-201, is colored light cyan; and loop 1, which comprises two unusual helices H1 and H2, is colored light green. (b) Phospholipid binding in the proposed active site of GlpG. Dashed lines indicate hydrogen-bonding interactions to the phosphate group that likely parallel the oxyanion hole stabilization during catalysis. (c) GlpG surface (teal) and superimposed TM2 and TM5 helices from two molecules seen in structure solved by Ben-Shem *et al.* (red and teal). A grayscale overlay of TM5 for the other structures of GlpG (PDB ID codes 2NRF and 2IC8) is shown.

search to date points to critical roles in mitochondrial membrane remodeling (10), apoptosis (11), quorum sensing (12), and cell invasion (13). Interestingly, even though GlpG has been shown to cleave Spitz (14), its normal function in *E. coli* remains elusive.

Careful biochemical and cell biological studies have contributed to our understanding of hydrolysis by intramembrane serine proteases since their discovery in 2001. Such studies have suggested a six- or seven-transmembrane architecture, a Ser-His catalytic dyad (14), and a critical Trp-Arg (WR) motif (9). Substrate cleavage is predicted to occur several residues into the transmembrane helix on the periplasmic/extracellular side and requires helix-destabilizing residues at or near the cleavage site (15). The crystal structure confirms and helps clarify several of these predicted features in molecular detail and also enables new structure-based hypotheses for catalysis by this family of enzymes.

As predicted, the crystal structure of GlpG reveals a core molecular architecture composed of six transmembrane helices (TM) plus two helices, H1 and H2, that are part of the unusual loop 1 connecting TM1 and TM2 (Fig. 1a). Although membrane proteins are typically thought to form a bundle of uniform-length helices that cross the membrane and are parallel to each other, the TM helices in GlpG are of varying lengths and angles with respect to the membrane normal. The particular spatial arrangement creates a hydrophilic cavity within the protein and emphasizes the importance of the overall protein scaffold in forming the unique chemical environment amenable to hydrolysis. A major contribution of the GlpG structures is in establishing that residues implicated in catalysis by site-directed mutagenesis are indeed located within the boundaries of the membrane,  $\approx 10$  Å below the surface within a water-filled cavity on the periplasmic side of the membrane. These residues, Ser-201 and His-254, which form the proposed catalytic dyad, are hydrogen-bonded with a distance of  $\approx 3.1$  Å and are located at

Author contributions: R.L.L. and M.S.W. wrote the paper.

The authors declare no conflict of interest.

See companion article on page 462.

\*To whom correspondence should be addressed. E-mail: mwolfe@rics.bwh.harvard.edu.

© 2007 by The National Academy of Sciences of the USA

the start of TM4 and in the middle of TM6, respectively.

The active-site cavity is shaped by the angled and central TM4 (Fig. 1*a*), in which the helical structure begins only at Ser-201, well within the membrane. This substructure is caused by the invariant and helix-destabilizing glycine residues that flank Ser-201. In addition, surrounding TM helices contribute a number of polar residues and enable water molecules to be sequestered in this cavity. TM4 appears to act as a stopper to prevent the water molecules from penetrating beyond Ser-201. Below Ser-201, residues on TM4 interact with other hydrophobic residues from surrounding TM helices. Lipid entry to the active site is further prevented by interactions of surrounding loop residues. Although not conserved among all species, side chains from residues on loop 3 form hydrophobic interactions with similar residues in the connecting region between H1 and H2 of loop 1 and act as a further barrier to lipid entry into the hydrophilic cavity from that side. On the other side is loop 5, which may mediate substrate entry (see below) and is ordered in only one of the two independent views of GlpG in the Ben-Shem *et al.* (4) structure.

The ensemble of GlpG structures adds important catalytic insights. In support of site-directed mutagenesis studies (14), Asn-154, initially proposed to be the third residue of a catalytic triad (9), similar to the aspartate associated with many soluble serine proteases, is distant from the side chains of the catalytic residues. Instead, Asn-154 may provide the oxyanion hole that stabilizes the tetrahedral intermediate expected after nucleophilic attack of the peptide carbonyl by Ser-201. Seen only in the structure by Ben-Shem *et al.* (4), the phosphate head group of a lipid molecule found near Asn-154, His-254, and Ser-201 mimics the tetrahedral intermediate expected in the reaction and provides further support that this is the catalytic center of rhomboid-like proteases (Fig. 1*b*).

In addition, the presence of the ordered lipid molecule near the hydrophilic catalytic center and extending into the hydrophobic region of TM2 and TM5 suggests that this surface may be used for substrate presentation. TM5 is one of the least sequence-conserved regions among rhomboids and may play a

## The transmembrane helices in GlpG are of varying lengths and angles.

role in tuning substrate specificity. Even among the available structures of GlpG, the greatest structural variation is in TM5 and the adjacent loop 5 (Fig. 1*c*). It is not known whether these differences are influenced by crystallization, but in the structure by Ben-Shem *et al.* (4), loop 5 is only ordered when the lipid is also present. How changes observed in TM5 and loop 5 might be triggered *in vivo* is unclear, but the variability of loop 5 and TM5 within the asymmetric unit demonstrates that such movement is possible.

An alternative hypothesis is that the substrate is presented to the hydrophilic cavity after a substantial rearrangement of H1, H2, and the rest of loop 1 to reveal a large opening between TM1 and TM3 (2). The unusual short helices that protrude partway into the membrane contain the WR motif, are amphipathic, and thus may be metastable. However, this loop appears unaltered in all GlpG structures to date. Moreover, mutation of the WR motif, expected in this model to destabilize the unusual structure and therefore increase substrate access and activity, instead reduces or abolishes protease activity (9, 14). In addition, the catalytic residues do not appear poised for catalysis from this direction. Another possible purpose of this unusual

region is in protein–protein interactions, which might transmit cell signals to control catalysis (4).

The first snapshots of GlpG have provided new clues about intramembrane proteolysis, and future efforts are likely to address details of the biochemical mechanism of this family of enzymes. To distinguish among various hypotheses for substrate entry, crystallographic trapping of a complex between GlpG and a substrate- or product-mimicking inhibitor, or a complex of a catalytically inactive GlpG and a substrate, would be invaluable. A cocrystal structure would provide evidence for any initial docking site for the substrate before presentation at the active site as well as residues involved in unwinding the helical substrate to help the nucleophile gain access to the cleavage site. Homology modeling and structures of other rhomboids are likely to expand our understanding of the functional landscape of this family of enzymes. Detailed knowledge of intramembrane proteolysis by rhomboids could eventually lead to the design of molecules to modulate pathological activities.

The sophisticated way in which rhomboid sequesters its active site, which requires hydrophilic residues and water, is likely to have parallels for other intramembrane proteases. For example, studies of presenilin and the  $\gamma$ -secretase complex point to a hydrophilic cavity in this intramembrane aspartyl protease (16–18). On the other hand, results using chemical probes of  $\gamma$ -secretase suggest that substrate access involves the interface between two distinct parts of presenilin (19), a mechanism that may be distinct from that proposed for rhomboid. Molecular structures of other I-CLiPs, such as presenilin/ $\gamma$ -secretase, signal peptide peptidase, and S2P, may reveal common features and general principles of protein hydrolysis in the membrane. Such future snapshots will eventually join those of rhomboid to form a family album of these fascinating enzymes.

1. Wolfe MS, Kopan R (2004) *Science* 305:1119–1123.
2. Wang Y, Zhang Y, Ha Y (2006) *Nature* 444:179–180.
3. 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.
4. Ben-Shem A, Fass D, Bibi E (2007) *Proc Natl Acad Sci USA* 104:462–466.
5. Rawson RB (2003) *Nat Rev Mol Cell Biol* 4:631–640.
6. Rudner DZ, Fawcett P, Losick R (1999) *Proc Natl Acad Sci USA* 96:14765–14770.
7. Wolfe MS (2006) *Biochemistry* 45:7931–7939.
8. Martoglio B, Golde TE (2003) *Hum Mol Genet* 12:R201–R206.
9. Urban S, Lee JR, Freeman M (2001) *Cell* 107:173–182.
10. McQuibban GA, Saurya S, Freeman M (2003) *Nature* 423:537–541.
11. 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.
12. Gallio M, Sturgill G, Rather P, Kylsten P (2002) *Proc Natl Acad Sci USA* 99:12208–12213.
13. Brossier F, Jewett TJ, Sibley LD, Urban S (2005) *Proc Natl Acad Sci USA* 102:4146–4151.
14. Lemberg MK, Menendez J, Misik A, Garcia M, Koth CM, Freeman M (2005) *EMBO J* 24:464–472.
15. Urban S, Freeman M (2003) *Mol Cell* 11:1425–1434.
16. Tolia A, Chavez-Gutierrez L, De Strooper B (2006) *J Biol Chem* 281:27633–27642.
17. Sato C, Morohashi Y, Tomita T, Iwatsubo T (2006) *J Neurosci* 26:12081–12088.
18. Lazarov VK, Fraering PC, Ye W, Wolfe MS, Selkoe DJ, Li H (2006) *Proc Natl Acad Sci USA* 103:6889–6894.
19. Kornilova AY, Bihel F, Das C, Wolfe MS (2005) *Proc Natl Acad Sci USA* 102:3230–3235.