Published in final edited form as:

*Curr Opin Struct Biol*. 2007 August ; 17(4): 405–411.

# Structural principles of intramembrane proteases

**Ya Ha**

*Department of Pharmacology, Yale School of Medicine, 333 Cedar Street, New Haven, CT 06520 U.S.A., Telephone: (203)785-7530, Fax: (203)785-7670, Email: ya.ha@yale.edu*

## Summary

Intramembrane proteases are present in most organisms, and are used by cells to send signal across membranes, to activate growth factors, and to accomplish many other tasks that are beyond the capability of their soluble cousins. These enzymes specialize in cleaving peptide bonds that are normally embedded in cell membranes. They contain multiple membrane-spanning segments, and their catalytic residues are often found within these hydrophobic domains. In the past year, a number of important papers have been published that began to address the structural features of these membrane proteins by x-ray crystallography, electron microscopy and biochemical methods, including the first report of an intramembrane protease crystal structure, that of *E. coli* GlpG. Taken together, these studies started to reveal patterns of how intramembrane proteases are constructed, how waters are supplied to the membrane-embedded active site, and how membrane protein substrates interact with them.

## Introduction

Proteolytic enzymes, their structures, and mechanisms are generally thought to be among the best understood in biochemistry. In this review we discuss however a unique, but important, group of proteases whose functions in the cell are just being recognized [1]. Intramembrane proteolysis was not long ago still a controversial concept. Now we know that this reaction can be catalyzed by several families of membrane proteins [2]. The very nature of the reaction, which involves peptide bond hydrolysis within the hydrophobic milieu of cell membranes, suggests fundamental differences in the construction of these enzymes, in their mechanism of action, and in how cells use them for specific functions that are not suitable for soluble proteases.

The biological and medical importance of these proteases is best exemplified by γ-secretase [3]. The action of γ-secretase illustrates a novel type of cross-membrane signaling: for example, by cleaving the transmembrane domain of notch, which releases and activates notch intracellular domain, γ-secretase plays an essential role in notch signaling that controls cell differentiation during development [4,5]. Also, mutations in presenilin, the catalytic component of γ-secretase, and in amyloid precursor protein, a substrate for the secretase, cause Familial Alzheimer's Disease [6]. Most mutations shift the cleavage specificity of γ-secretase, producing a more toxic form of amyloid β-peptide [7].

Figure 1 illustrates the membrane topologies for site-2 protease (S2P) [8], rhomboid protease [9●●] and presenilin [10], representatives of the three mechanistic families, and the location of their catalytic residues relative to the membrane [11–13]. There are quite a few variations

Corresponding author: Ya Ha

from these structures within each family. One interesting example is signal peptide peptidase [16], an aspartic protease like presenilin, whose catalytic motifs adopt a different membrane orientation. This flip results in a specificity switch towards substrates of opposite membrane topology. Similar such examples may also exist in the other two mechanistic families [17, 18].

## The crystal structure of bacterial GlpG, a rhomboid family serine protease

Drosophila rhomboid-1, originally discovered as a key genetic factor in the fly epidermal growth factor receptor signaling pathway, is the founding member of the family [12]. *E. coli* GlpG is a homolog of the fly rhomboid protease. After some effort, experimental conditions to crystallize and derivatize the bacterial membrane protein were finally found, and the resulting crystal structure immediately suggested a solution to the problem of intramembrane proteolysis [9●●]. With this new knowledge, we also start to see clearer patterns in the construction of intramembrane proteases, and in their mechanism of action that are different from soluble proteases.

GlpG is composed primarily of transmembrane helices that form an asymmetric bundle (Figure 2a, b) [9●●]. This overall feature was largely anticipated, but the crystal structure also had many surprises. For instance, the fourth hydrophobic segment of GlpG was once thought to constitute a single membrane-spanning helix. The crystal structure now shows that only part of it adopts a helical conformation (S4), whereas the rest of it forms an extended loop (L3), a region that we will re-visit later in the review. L3 and S4 are positioned roughly at the center of the molecule, surrounded by other transmembrane helices (Figure 2c). This arrangement creates a cavity near L3 that opens to one side of the membrane. One may notice that the structure of GlpG is unlike those of channel proteins, which are also helical bundles, but have a continuous and aqueous path that traverses the membrane. This feature is of great functional importance, and is probably true also for intramembrane proteases of the other families (see below). The active site of GlpG is housed inside the hydrophilic cavity.

Within the active site of GlpG, the composition and spatial arrangement of functional groups are generally similar to those in soluble proteases (Figure 2d) [9●●,19●]. Therefore, the chemical basis of the catalysis is likely preserved. Rhomboid uses a Ser-His catalytic dyad (Ser-201, His-254) for nucleophilic attack on the substrate [20], which is an interesting variation of the more conventional Ser-His-Asp catalytic triad. The oxyanion intermediate of the reaction is likely stabilized by interactions with a backbone amide (Ser-201) and two conserved side chains (His-150, Asn-154) of GlpG [19●,21●,22●], which is much like that of subtilisin, a well-known serine protease.

The substrate for intramembrane proteases can only move within the plane of the membrane, and therefore has to enter the hydrophilic active site laterally. The entry point needs to be blocked when substrate is absent. This structural requirement is unique to intramembrane proteases. For GlpG, the physical barrier is provided by a mobile loop structure (L5) that caps the active site from the extracellular side: the snug fit of a hydrophobic side chain (Phe-245) that reaches down from the loop completely blocks a gap, between two transmembrane helices (S2, S5), that leads to the catalytic dyad (Figure 2c) [19●]. The rest of the cap sits immediately above the dyad, preventing it from attacking substrate (the orientation of the loop is opposite to that of a substrate, which may explain why it is not cleaved). Shortly after the *E. coli* structure, the apo-structure of *H. influenza* GlpG has also been solved [21●]. Despite differences in the sequence, the structure of the L5 loop, as well as its capping conformation, are identical in the two structures. The peripheral membrane location of the L5 cap is also quite informative: the loop contains both hydrophobic and hydrophilic residues, which enables the cap to operate

smoothly at the lipid:water interface. Intramembrane proteases of other families are expected to have structural elements of similar properties to execute this function.

Crystallographic evidences also indicated that the L5 cap is flexible, and has an intrinsic propensity to flip open [19●]. Whether or not the cap is able to spontaneously "open" in the membrane can not be decided based on crystallographic analysis of the detergent-solubilized protease alone, but biochemical results, which are reviewed below, suggested that the answer is likely yes. With the cap lifted, the top region of substrate transmembrane helix is able to bend into the active site through the narrow gap between S2 and S5, in an unfolded conformation, to become cleaved. It is also clear from the crystal structure that the active site of GlpG, and the path that leads to it, can only accommodate a small portion of the substrate, and that the majority of the substrate helix must remain in the lipid phase [19●]. Since there is yet no structural information for GlpG:substrate complex, these predictions need to be proven by future experiments.

Of the six transmembrane helices in GlpG, S5 is the most loosely packed, and may have a certain degree of flexibility, as indicated by small variations of its position in the three crystal forms of *E. coli* GlpG so-far reported [9●●,22●,23●], and by differences between *E. coli* and *H. influenza* GlpGs [9●●,21●]. In one particular crystal form, S5 in one of the two molecules in the asymmetric unit has tilted dramatically to the side, away from the helical bundle, and exposing the entire hydrophilic interior of the membrane protein [23●]. This observation has led to a rather interesting proposal that substrate entry is gated by the movement of S5. Nevertheless, upon examining crystal packing, one may notice that the helix movement is caused by its contact with a neighboring molecule, a possible artifact that has been overlooked.

One structural element whose role in GlpG function is unclear at the moment is the L1 loop (Figure 2a). This loop is positioned outside the helical bundle, and half-submerged in the membrane rather curiously [9●●]. The amino acid sequence for the submerged part is largely hydrophobic, and is conserved, indicating that this structure is a common feature for the rhomboids. It is worth noting that S2P and presenilin both contain similar long stretches of hydrophobic sequences that do not fully traverse the membrane (Figure 1). These novel structural motifs, and their functions, remain interesting subjects for future studies.

There are two more structural issues whose relevance and implications in the function of intramembrane proteases need to be confirmed, and further explored. One relates to the chirality of the reaction. In the model described above, Ser-201 of GlpG has to attack the scissile bond from the *si*-face, instead of from the *re*-face like most other serine proteases, resulting in a tetrahedral intermediate of opposite handedness [19●]. The second issue relates to the basis of the observed regulatory effect of lipids on protease activity [24]. In one crystal form of *E. coli* GlpG, a phospholipid molecule was found in the protease active site [22●], which suggested one potential mode of inhibition.

## Low-resolution electron microscopic images of γ-secretase

Two reports appeared last year that started to address structural features of this important intramembrane protease using electron microscopy [25●,26]. These studies were still at an early stage, and were based on negatively stained protein solutions, which restricted the resolution. Most unfortunately, the size and shape of the particles observed in the two studies do not match. Therefore, there is some uncertainty about these results at this time.

The most surprising feature of the secretase model from the higher resolution (15 Å) study is a chamber, about 20–40 Å in length and highly irregular in shape, inside the membrane-embedded portion of the protein complex (Figure 3a) [25●]. There are large pores that connect the chamber to both sides of the membrane. As illustrated in Figure 3b, the internal cavity of

GlpG is much smaller, and is open to only one side of the membrane. Because of these obvious differences, one may think that the two proteases must operate quite differently. Nevertheless, since the resolution of the electron microscopic study does not yet permit a trace of the polypeptide chain, the chemical composition of the chamber wall can not be determined, and may not be hydrophilic (most amino acid residues within the membrane-spanning segments of the four protein components are hydrophobic). It is possible that a smaller hydrophilic cavity exists somewhere else, which contains the active site, and is not so obvious at low resolution.

## The accessibility of active site by water

At the resolution (2.1 Å) the crystal structure was solved, water molecules are clearly visible within the active site of GlpG [9●●] (Figure 4a). Biochemical studies, using proteases embedded in natural membranes, indicated that water can diffuse directly into the protease.

While the crystal structure of GlpG was being solved, the transmembrane segment that contains the catalytic GLSG motif was independently probed by substituted cysteine accessibility method (Figure 1) [27●]. We now know that this segment is located at the center of the protease, and consists of two structural elements, an extended loop (L3) and a helix (S4). Most cysteines introduced to L3, and to position 201, react readily with alkylating reagents (a water-requiring reaction), thus defining a hydrated and continuous path from aqueous solution to the catalytic dyad (Ser-201, His-254) (Figure 4b). A shorter sequence around His-254 from helix S6 can also be modified. Since some alkylating reagents are fairly large, and unlikely to penetrate into GlpG with its cap closed, their abilities to reach residues buried deep seem to indicate that the flexible L5 cap must be open at certain times, even while the substrate is absent.

The pattern of accessibility in transmembrane domain (TMD) 7 of presenilin, which contains the catalytic GLGD motif, is interestingly similar (Figure 1; Figure 4c) [28]. Most residues, from the cytosolic face (opposite of GlpG) to the catalytic aspartate, when substituted, are susceptible to alkylation, whereas the rest of the domain is not. This result raises the possibility that TMD7 is also composed of two structural elements, a readily accessible and possibly extended region (equivalent of L3 in GlpG) and a buried helix. Some residues from the exposed region may participate in substrate binding since their accessibilities were affected by competitive inhibitors. Residues flanking the other catalytic aspartate on TMD6 are less exposed, but cross-linking experiment indicated that the two aspartates closely oppose each other, and are placed inside a hydrophilic cavity [29]. Drawing lessons from GlpG, one may predict that this local environment is tightly surrounded by transmembrane helices from the side, but is continuous with the aqueous solution.

*E. coli* RseP, a S2P family protease (Figure 1), appears to be somewhat different: cysteines introduced near its catalytic motifs are surprisingly resistant to chemical modifications [30]. This implies that, on top of the basic schemes described above, there may be complex mechanisms that control water access, and substrate binding, to the active site for some intramembrane proteases.

## Conclusions

The past year has witnessed much progress in our structural knowledge about the intramembrane proteases. Now we know that these membrane proteins are constructed from rather conventional, but cleverly arranged, structural elements to accomplish a once-thought impossible task. The chemical apparatus of their catalysis is similar to those used by soluble proteases, but is neatly put inside the framework of an integral membrane protein. Questions remain regarding how transmembrane protein substrates gain access to their active sites. The model where only part of substrate enters the protease is different from ideas that the entire

transmembrane helix becomes engulfed (Figure 3b), and the general applicability of this model requires many future experiments to test.

**Acknowledgements**

The author is grateful to Dr. Akiyama for providing the diagram in Figure 4b, and acknowledges supports from Yale University, the Ellison Medical Foundation, the Neuroscience Education and Research Foundation, and NIH (GM082839).

## References

1. Brown MS, Ye J, Rawson RB, Goldstein JL. Regulated intramembrane proteolysis: a control mechanism conserved from bacteria to humans. Cell 2000;100:391–398. [PubMed: 10693756]
2. Wolfe MS, Kopan R. Intramembrane proteolysis: theme and variations. Science 2004;305:1119–1123. [PubMed: 15326347]
3. Haass C. Take five--BACE and the gamma-secretase quartet conduct Alzheimer's amyloid beta-peptide generation. EMBO J 2004;23:483–488. [PubMed: 14749724]
4. Levitan D, Greenwald I. Facilitation of lin-12-mediated signalling by sel-12, a Caenorhabditis elegans S182 Alzheimer's disease gene. Nature 1995;377:351–354. [PubMed: 7566091]
5. De Strooper B, Annaert W, Cupers P, Saftig P, Craessaerts K, Mumm JS, Schroeter EH, Schrijvers V, Wolfe MS, Ray WJ, et al. A presenilin-1-dependent gamma-secretase-like protease mediates release of Notch intracellular domain. Nature 1999;398:518–522. [PubMed: 10206645]
6. Selkoe DJ. Alzheimer's disease: genes, proteins, and therapy. Physiological Reviews 2001;81:741–766. [PubMed: 11274343]
7. Hardy J. The Alzheimer family of diseases: many etiologies, one pathogenesis? Proc Natl Acad Sci USA 1997;94:2095–2097. [PubMed: 9122152]
8. Zelenski NG, Rawson RB, Brown MS, Goldstein JL. Membrane topology of S2P, a protein required for intramembranous cleavage of sterol regulatory element-binding proteins. J Biol Chem 1999;274:21973–21980. [PubMed: 10419520]
9. Wang Y, Zhang Y, Ha Y. Crystal structure of a rhomboid family intramembrane protease. Nature 2006;444:179–183. [PubMed: 17051161]●● This is the first report of a crystal structure for any intramembrane protease. Besides showing the overall structure of the protease, it demonstrates that the active site is below membrane surface, and contains water. It confirms early hypothesis that rhomboid uses a Ser-His dyad for catalysis
10. Li X, Greenwald I. Additional evidence for an eight-transmembrane-domain topology for Caenorhabditis elegans and human presenilins. Proc Natl Acad Sci USA 1998;95:7109–7114. [PubMed: 9618547]
11. Rawson RB, Zelenski NG, Nijhawan D, Ye J, Sakai J, Hasan MT, Chang TY, Brown MS, Goldstein JL. Complementation cloning of S2P, a gene encoding a putative metalloprotease required for intramembrane cleavage of SREBPs. Mol Cell 1997;1:47–57. [PubMed: 9659902]
12. Urban S, Lee JR, Freeman M. Drosophila rhomboid-1 defines a family of putative intramembrane serine proteases. Cell 2001;107:173–182. [PubMed: 11672525]
13. Wolfe MS, Xia W, Ostaszewski BL, Diehl TS, Kimberly WT, Selkoe DJ. Two transmembrane aspartates in presenilin-1 required for presenilin endoproteolysis and gamma-secretase activity. Nature 1999;398:513–517. [PubMed: 10206644]
14. Yu G, Nishimura M, Arawaka S, Levitan D, Zhang L, Tandon A, Song YQ, Rogaeva E, Chen F, Kawarai T, et al. Nicastrin modulates presenilin-mediated notch/glp-1 signal transduction and betaAPP processing. Nature 2000;407:48–54. [PubMed: 10993067]
15. Takasugi N, Tomita T, Hayashi I, Tsuruoka M, Niimura M, Takahashi Y, Thinakaran G, Iwatsubo T. The role of presenilin cofactors in the gamma-secretase complex. Nature 2003;422:438–441. [PubMed: 12660785]
16. Weihofen A, Binns K, Lemberg MK, Ashman K, Martoglio B. Identification of signal peptide peptidase, a presenilin-type aspartic protease. Science 2002;296:2215–2218. [PubMed: 12077416]

17. Herlan M, Bornhovd C, Hell K, Neupert W, Reichert AS. Alternative topogenesis of Mgm1 and mitochondrial morphology depend on ATP and a functional import motor. J Cell Biol 2004;165:167–173. [PubMed: 15096522]

18. Rudner DZ, Fawcett P, Losick R. A family of membrane-embedded metalloproteases involved in regulated proteolysis of membrane-associated transcription factors. Proc Natl Acad Sci USA 1999;96:14765–14770. [PubMed: 10611287]

19. Wang Y, Ha Y. Open-cap conformation of intramembrane protease GlpG. Proc Natl Acad Sci USA 2007;104:2098–2102. [PubMed: 17277078]● By showing that the cap is flexible and prone to open, this paper makes detailed predictions regarding the lateral entry of substrate, composition of the oxyanion hole, and the chirality of reaction intermediate

20. Lemberg MK, Menendez J, Misik A, Garcia M, Koth CM, Freeman M. Mechanism of intramembrane proteolysis investigated with purified rhomboid proteases. EMBO J 2005;24:464–472. [PubMed: 15616571]

21. Lemieux MJ, Fischer SJ, Cherney MM, Bateman KS, James MN. The crystal structure of the rhomboid peptidase from Haemophilus influenzae provides insight into intramembrane proteolysis. Proc Natl Acad Sci USA 2007;104:750–754. [PubMed: 17210913]● This is only the second new structure of a rhomboid protease, and highlights features that are conserved in the family. It also makes predictions about the reaction mechanism, and suggests that the conserved His-150 participates in oxyanion binding

22. Ben-Shem A, Fass D, Bibi E. Structural basis for intramembrane proteolysis by rhomboid serine proteases. Proc Natl Acad Sci USA 2007;104:462–466. [PubMed: 17190827]● This paper describes the structure of *E. coli* GlpG in a different crystal form. The most intriguing observation is about a phospholipid molecule that occupies one of the two active sites in the crystallographic asymmetric unit

23. Wu Z, Yan N, Feng L, Oberstein A, Yan H, Baker RP, Gu L, Jeffrey PD, Urban S, Shi Y. Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry. Nat Struct Mol Biol 2006;13:1084–1091. [PubMed: 17099694]● This paper describes the structure of *E. coli* GlpG in yet another crystal form. The authors argue that substrate entry is gated by the movement of a transmembrane helix

24. Urban S, Wolfe MS. Reconstitution of intramembrane proteolysis in vitro reveals that pure rhomboid is sufficient for catalysis and specificity. Proc Natl Acad Sci USA 2005;102:1883–1888. [PubMed: 15684070]

25. Lazarov VK, Fraering PC, Ye W, Wolfe MS, Selkoe DJ, Li H. Electron microscopic structure of purified, active gamma-secretase reveals an aqueous intramembrane chamber and two pores. Proc Natl Acad Sci USA 2006;103:6889–6894. [PubMed: 16636269]● The structural study is of high enough resolution to allow visualization of an internal cavity, pores leading to the cavity, and nicastrin ectodomain

26. Ogura T, Mio K, Hayashi I, Miyashita H, Fukuda R, Kopan R, Kodama T, Hamakubo T, Iwatsubo T, Tomita T, et al. Three-dimensional structure of the gamma-secretase complex. Biochem Biophys Res Commun 2006;343:525–534. [PubMed: 16546128]

27. Maegawa S, Koide K, Ito K, Akiyama Y. The intramembrane active site of GlpG, an *Escherichia coli* rhomboid protease, is accessible to water and hydrolyzes an extramembrane peptide-bond of substrates. Mol Microbiol 2007;64:435–447. [PubMed: 17493126]● Using membrane-embedded GlpG, these authors show that the active site is open to the aqueous solution outside of the membrane. They also argue that the cleavage site on the substrate is exposed to water, a somewhat controversial idea at this time

28. Sato C, Morohashi Y, Tomita T, Iwatsubo T. Structure of the catalytic pore of gamma-secretase probed by the accessibility of substituted cysteines. J Neurosci 2006;26:12081–12088. [PubMed: 17108181]

29. Tolia A, Chavez-Gutierrez L, De Strooper B. Contribution of presenilin transmembrane domains 6 and 7 to a water-containing cavity in the gamma-secretase complex. J Biol Chem 2006;281:27633–27642. [PubMed: 16844686]

30. Koide K, Maegawa S, Ito K, Akiyama Y. Environment of the active site region of RseP, an Escherichia coli regulated intramembrane proteolysis protease, assessed by site-directed cysteine alkylation. J Biol Chem 2007;282:4553–4560. [PubMed: 17179147]

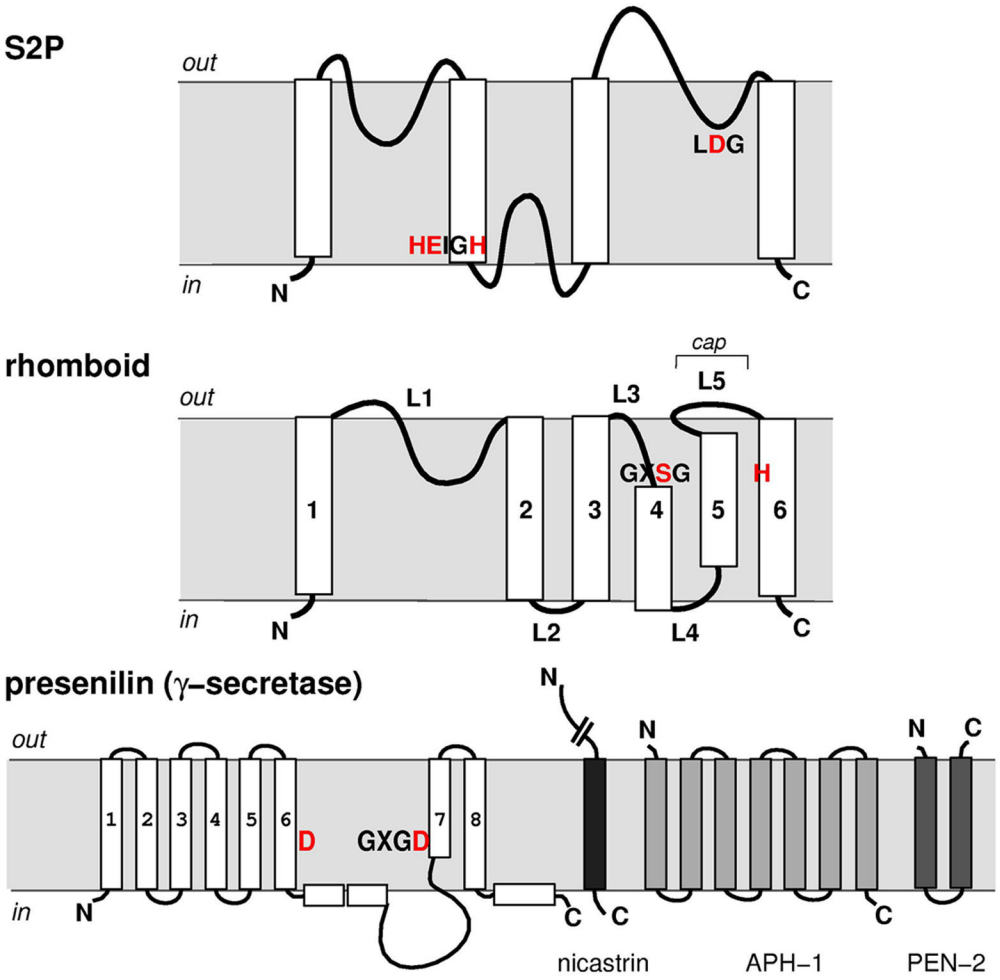


**Figure 1. The membrane topology models of S2P, rhomboid and presenilin**

The catalytic residues are shown in red. Presenilin is the catalytic component of a large membrane protein complex, γ-secretase, that contains three other membrane proteins (nicastrin, APH-1, PEN-2) [14,15].

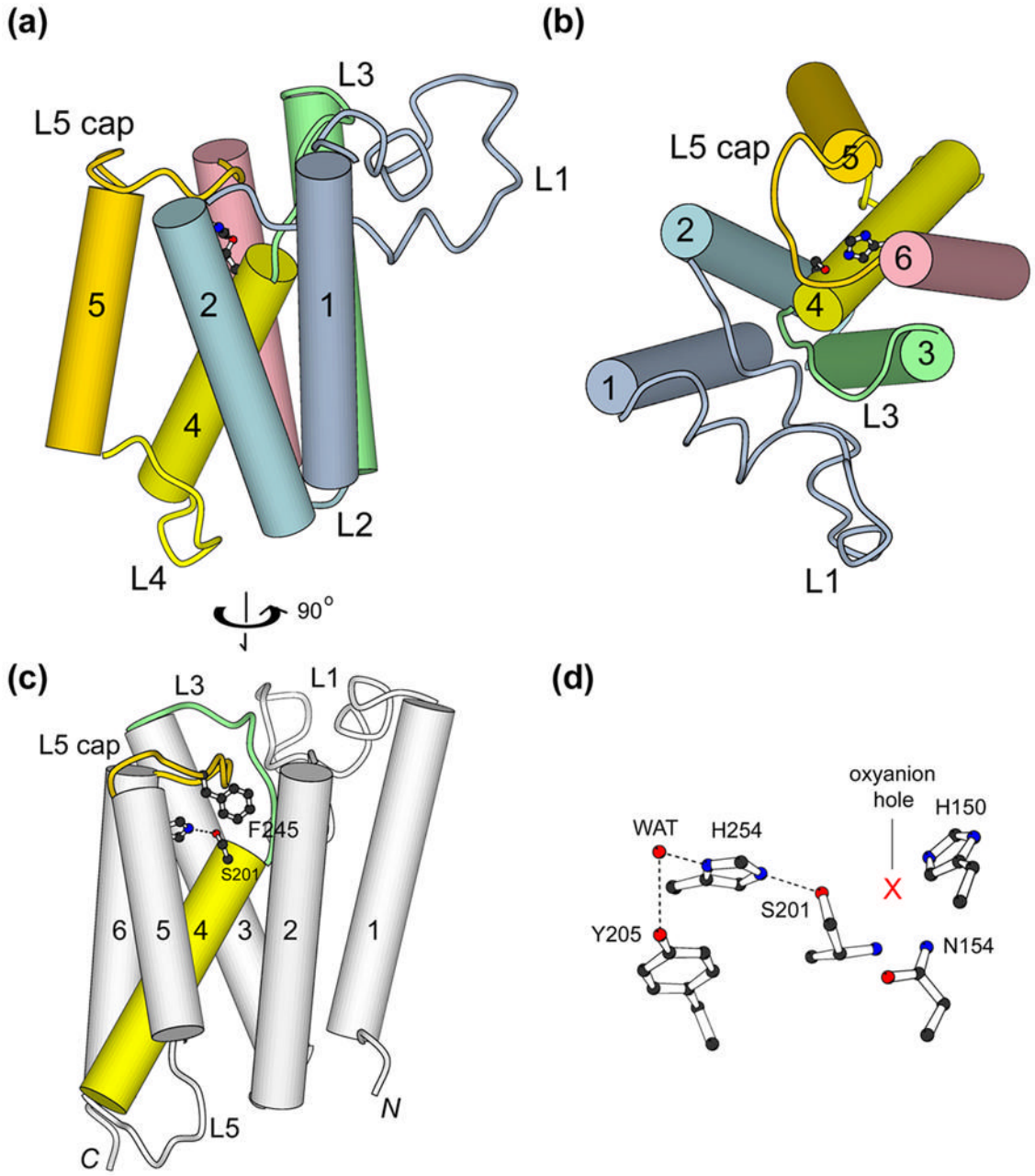


**Figure 2. The crystal structure of *E. coli* GlpG**

(**a**) A cartoon illustration of the molecule (the side view). The six transmembrane helices are shown as cylinders of different colors, and labeled 1–6. The Ser-His dyad is also shown.

(**b**) The top view.

(**c**) A view from the back towards the lateral opening (between S2 and S5) of the active site, highlighting L3 (green), S4 (yellow) and the L5 cap (gold). The catalytic dyad (Ser-201, His-254) and Phe-245 from the L5 cap are shown as ball-and-stick models. The dashed line represents hydrogen bond between Ser-201 and His-254.

(**d**) A detailed view of GlpG active site from a similar angle as in (**c**). WAT, water. His-150, Asn-154, Ser-201 and His-254 are conserved in all functional rhomboid proteases.

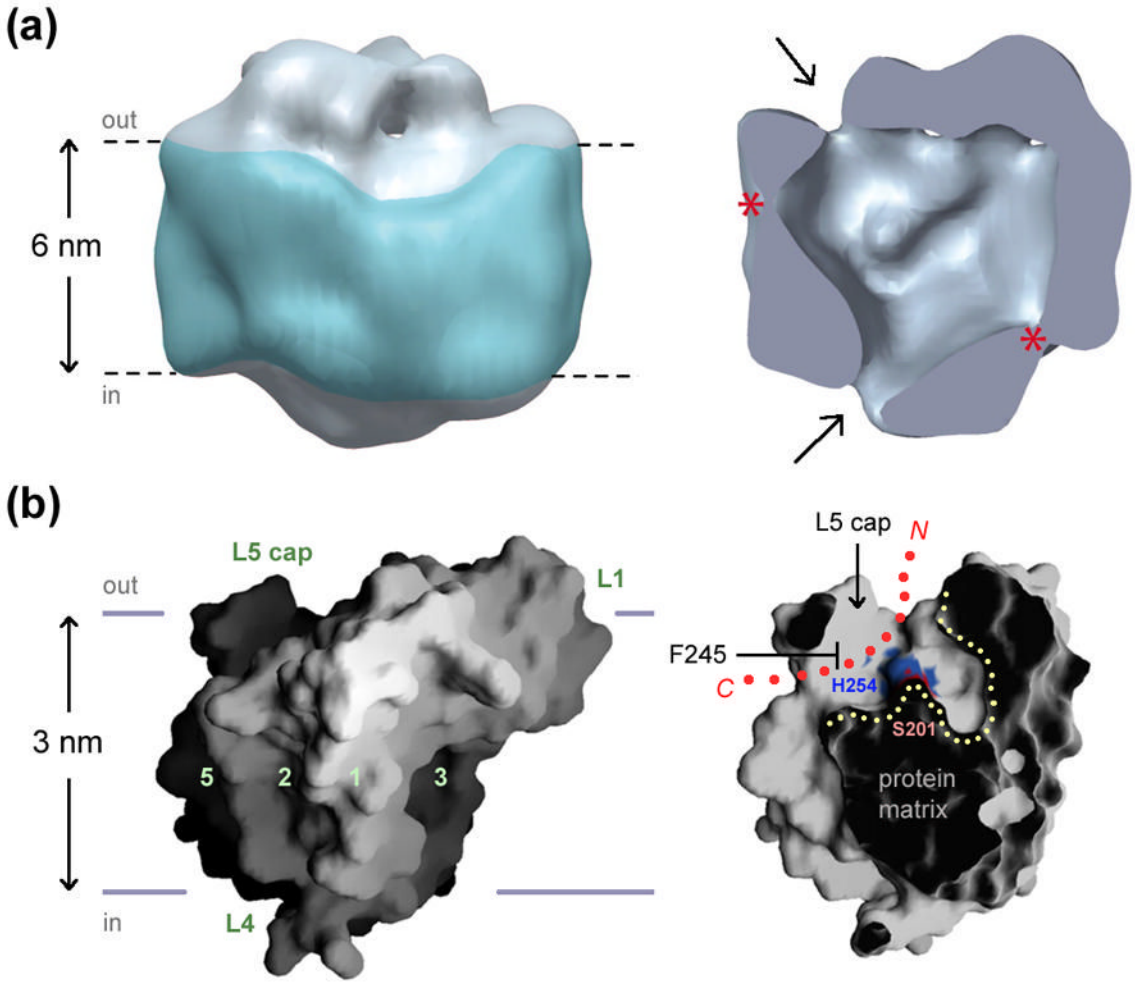


**Figure 3. Internal cavities of γ-secretase and GlpG**

(**a**) The electron microscopic model of γ-secretase (left), and a cut-open view (right). These figures were from reference [25] and copyrighted by The National Academy of Sciences of the USA. The blue surface represents the transmembrane region. The arrows point at openings of the internal chamber.

(**b**) The molecular surface of GlpG (left), and a cut-open view (right). These figures are of a different scale from those in (**a**). The L5 cap is omitted from both pictures to reveal the internal cavity. The hydrophobic core of the membrane is indicated by two horizontal lines. The yellow dotted line marks the lower portion of the internal cavity. The lateral entry is normally blocked by Phe-245, whereas the top opening to the aqueous solution is capped by the rest of L5. The blue surface corresponds to His-254, and the red surface corresponds to Ser-201. The red dotted line marks the potential path of bound substrate (the majority of the transmembrane domain on the C-terminal side remains in the lipid).

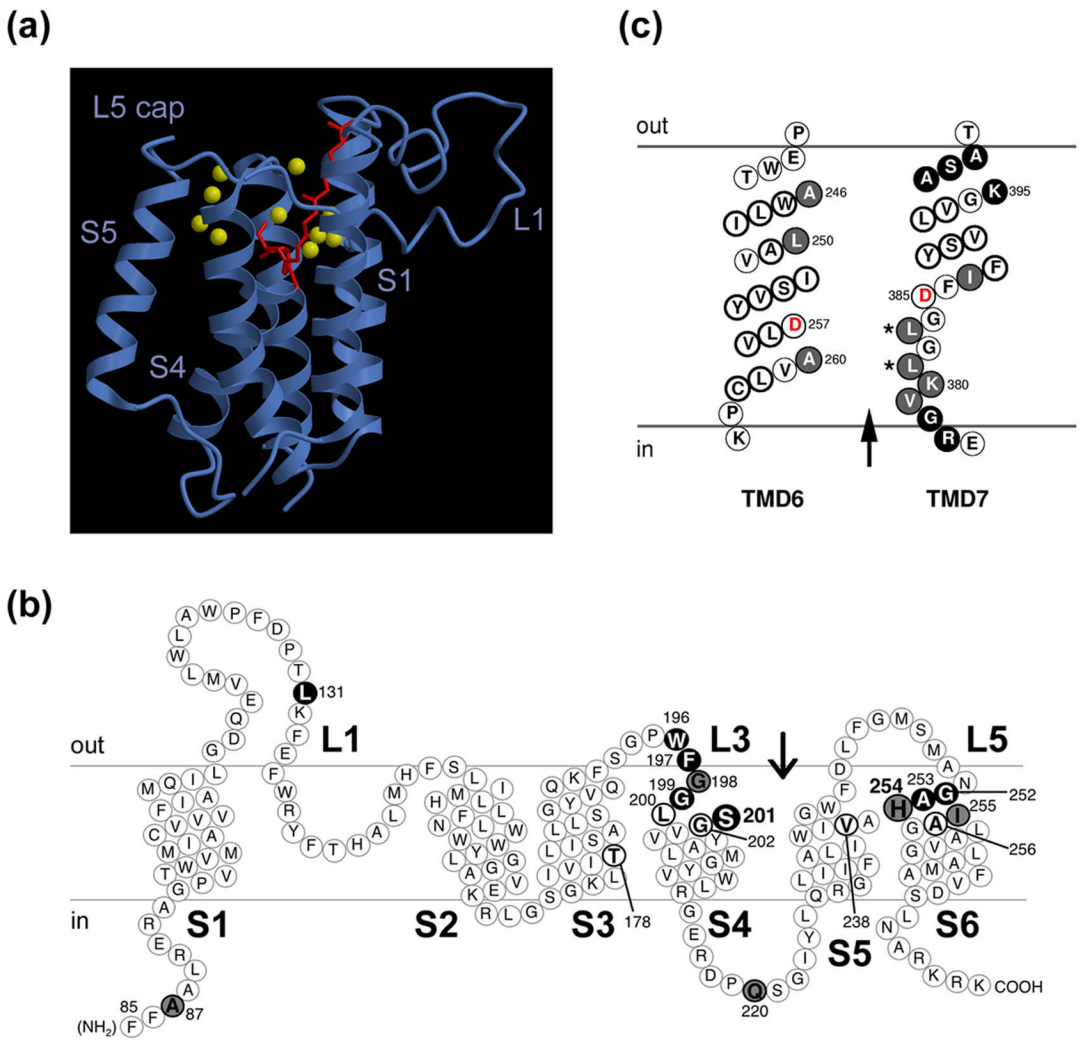


**Figure 4. The opening of active site to aqueous solution**

(**a**) Water molecules (yellow) found inside GlpG. The peptide backbone from Trp-196 to Gly-202 (L3), and the sidechain of Ser-201, are shown in red.

(**b**) A schematic representation of GlpG (kindly provided by Dr. Akiyama). Residues highlighted in black or grey are accessible to water even while GlpG is embedded in cell membranes [27]. The arrow marks the opening to periplasm, which is confirmed by the x-ray structure.

(**c**) A schematic representation of presenilin TMD6 and TMD7 [28]. The catalytic aspartates are shown as red letters. Residues highlighted in black or grey are accessible to water, whereas those represented by the thick circles are not. The stars mark residues protected by a competitive inhibitor. The arrow marks the opening to cytosolic solution. Results from a similar study differ slightly, but they agree on the observation that residues on TMD7 from cytosol to Asp-385 are very accessible to water [29].