Ten catalytic snapshots of rhomboid intramembrane proteolysis from gate opening to peptide release

Sangwoo Cho, Rosanna P. Baker, Ming Ji and Siniša Urban

Protein cleavage inside the cell membrane triggers various pathophysiological signaling pathways, but the mechanism of catalysis is poorly understood. We solved ten structures of the *Escherichia coli* rhomboid protease in a bicelle membrane undergoing time-resolved steps that encompass the entire proteolytic reaction on a transmembrane substrate and an aldehyde inhibitor. Extensive gate opening accompanied substrate, but not inhibitor, binding, revealing that substrates and inhibitors take different paths to the active site. Catalysis unexpectedly commenced with, and was guided through subsequent catalytic steps by, motions of an extracellular loop, with local contributions from active site residues. We even captured the elusive tetrahedral intermediate that is uncleaved but covalently attached to the catalytic serine, about which the substrate was forced to bend dramatically. This unexpectedly stable intermediate indicates rhomboid catalysis uses an unprecedented reaction coordinate that may involve mechanically stressing the peptide bond, and could be selectively targeted by inhibitors.

Despite their prominent pathophysiological roles, the steps underlying catalysis have never been observed directly for any intramembrane protease; current models only rationalize mechanisms by assuming parallels to their well-characterized soluble protease counterparts, and so it is for rhomboid enzymes. Serine proteases use a catalytic base, which is usually a histidine, to convert the benign serine into a powerful nucleophile by abstracting its proton (Fig. 1)17. The ensuing attack on the carbonyl carbon of the peptide bond results in a short-lived transition state in which the amide nitrogen must be protonated by the enzyme and then stabilized by the enzyme in order for this step to take place. Rapid protonation of the amide nitrogen by the catalytic base facilitates release of the carboxy-terminal peptide product, while the amino-terminal product simultaneously decomposes to an acyl intermediate that is covalently attached to the catalytic serine. The cycle repeats in the second half of the reaction, except this time with a water molecule, which is activated by the enzyme, attacking the carbonyl carbon, thereby releasing the amino-terminal product and restoring a proton to the serine.

While the general chemistry of peptide bond hydrolysis is likely to be similar between soluble serine proteases and rhomboid proteases, meaningful differences should be expected: rhomboid proteases evolved independently by convergent evolution and inside the membrane. Therefore, it is important to delineate the mechanism of rhomboid catalysis directly. Time-resolved crystallography can visualize the discrete actions taken by enzymes during catalysis and thereby directly map out the reaction coordinate: reactions are made to occur in crystals and snapshots are taken in a time-dependent manner19. This approach has revealed some catalytic steps in atomic detail for a handful of soluble serine proteases20–22 and even such complicated enzymes as DNA polymerases23. However, this powerful approach has been difficult to implement with membrane enzymes. Transformative successes include witnessing the biosynthesis and extrusion of nature’s most abundant polymer, cellulose, from the membrane by a bacterial membrane-immersed enzyme24, light-driven steps of the bacteriorhodopsin proton-pumping cycle25.
and water oxidation by massive plant photosystem complexes\textsuperscript{26,27}. Time-resolved crystallography, nevertheless, remains rare with membrane enzymes, and has never been achieved with any membrane protease\textsuperscript{25}. Given the remarkable power of this approach, we sought to examine whether rhomboid proteolysis inside the membrane might be amenable to time-resolved crystallography.

Results
Snapsots of catalysis on a peptide aldehyde inhibitor. Since rhomboid proteases target relatively short regions of their substrates\textsuperscript{20–22} and can take minutes to even hours to catalyze proteolysis of a single peptide bond\textsuperscript{23,24}, we examined whether we might be able to visualize catalytic intermediates using a crystal-soaking approach that we developed previously\textsuperscript{11}. Our starting point was the \textit{Escherichia coli} rhomboid GlpG that we recently crystallized in a bicelle membrane (Fig. 2). This was important for mechanistic studies, because the membrane modulates all known properties of rhomboid proteolysis\textsuperscript{15–17}. In our crystals, GlpG is both active and assumes a partly open conformation that is able to accept peptide inhibitors. Specifically, the L5 loop that overlies that active site, and must lift up to allow substrate access to the active site, was entirely disordered in our starting structure. Having recently characterized the tetrahedral end product that resulted from catalysis by GlpG in crystals soaked with tetrapeptide aldehydes modeled on the classical substrate Gurken\textsuperscript{11}, we sought to test whether we could visualize discrete steps in the inhibition pathway in atomic detail. We exposed crystals to the tetrapeptide aldehyde inhibitor of sequence VRMA (termed P4 to P1 in protease nomenclature) for varying lengths of time, and solved the resulting structures to 2.2–2.4 Å resolution (Table 1). We ultimately succeeded in witnessing five distinct enzymatic steps on this inhibitor, which we termed Snapshot-I to Snapshot-I\textsubscript{5} (Fig. 3 and Supplementary Fig. 1).

The starting apoenzyme structure had a weak 3.29-Å hydrogen bond between the catalytic S201 and H254 residues, and a hydrogen bond between S201 and H150 (which stabilizes the oxyanion with N154) through a bridging water molecule with distances of 3.20 Å and 3.14 Å, respectively (Fig. 3a). The earliest difference we could ever see indicated some density beginning to appear for the inhibitor, but it was very weak and could not be modeled (not shown). Interestingly, the overlying L5 loop had, nevertheless, already started to assume an ordered conformation. In our first snapshot we observed increased density and could model the amino-terminal three residues of the peptide aldehyde, but the P1 alanine next to the catalytic residues remained too disordered to be observed (Fig. 3b). The catalytic residues maintained their positions similar to the apoenzyme structure (including the presence of the bridging water between S201 and H150), but the L5 loop became ordered in all except four residues (L244–M247), and clearly established interactions with the P2, P3 and P4 residues of the inhibitor. Specifically, these included hydrogen bonds between M249 and P3–R at a distance of 3.29 Å, S248 and P2–M at a distance of 2.94 Å and a hydrogen-phobic interaction between P4–V and F146 (Fig. 3c).

Snapshot-I\textsubscript{2} revealed strong and ordered electron density of the peptide aldehyde that allowed confident modeling of the entire peptide moiety (Fig. 3d). This was the first time point that the P1 alanine residue became visible, making it the last element to bind stably. Despite this residue being essential for proteolysis, its late binding suggested that it does not make a strong interaction that contributes to affinity. This step was also accompanied by the first changes in the catalytic residues. The water molecule that initially bridged S201 and H150 was pushed outwards by the aldehyde oxygen (the future oxyanion) of the inhibitor, and the side chain of H150 assumed a new, flipped-out position by rotating 94.8° with its Nδ1 being displaced 3.93 Å from the future oxyanion. In fact, we could observe both the inward and flipped-out conformation of the H150 side chain in approximately equal proportions at this time point. Notably, S201 and the peptide aldehyde remained too far apart for covalent linkage and without clear connecting electron density, indicating that catalysis had not yet taken place.

In Snapshot-I\textsubscript{3} we witnessed the first evidence of catalysis: clear density connected S201 and the peptide aldehyde, thus mimicking the tetrahedral intermediate (Fig. 3e). No other changes in the substrate were evident along the length of the peptide. Of particular note was the change in position of the catalytic H254, which moved downward by rotating 38° and no longer interacted with S201, which itself rotated by 15°. Such a motion would be expected to accompany proton transfer from S201 during nucleophilic activation\textsuperscript{15}. Surprisingly, we found the oxyanion interacting with only the side chain of N154 at a distance of 3.19 Å, but it became surrounded by three new water molecules, while H150 had assumed the flipped-out position fully.

The final snapshots all revealed a dramatic reorganization of the active site (Fig. 3f): H150 adopted the inward conformation, and all hydrogen bonds with the oxyanion were readjusted to form a tripartite interaction network consisting of the side chains of H150 (at a distance of 3.03 Å) and N154 (at a distance of 3.17 Å), and the backbone of S201 (at a distance of 3.30 Å). The catalytic H254 also shifted to restore its hydrogen bond with S201 (at a distance of 3.15 Å), which remained in covalent complex with the peptide aldehyde. This overall enzyme–inhibitor complex was stable in the bicelle crystal, because we could not see any further conformational changes, even when we extended the incubation time.

Snapshots of catalysis on a transmembrane peptide substrate. Encouraged by the success of our approach with the short tetrapeptide inhibitor, we examined whether we could visualize any binding or catalytic steps on long peptide substrates (lacking any warheads that could help to stabilize them in the active site). We tested several peptide sequences, but ultimately obtained the best results with a 13-mer modeled on the sequence of the classical rhomboid substrate Gurken. The substrate sequence included six residues preceding the cleavage site (RKVRMA) and extended all the way to, and including, the deep transmembrane helix-breaking serine and
proline residues that facilitate unwinding (AIVFSFP). In standard protease nomenclature, residues amino-terminal to the cleavage site are numbered from P1 onwards, while those that are carboxy-terminal to the cleavage site are similarly labeled but with prime designations. Remarkably, we ultimately visualized the entire reaction, including intermediates, and in six snapshots of 2.3–2.6 Å resolution that we termed Snapshot-S1 through S6 (Table 2): gate opening and substrate binding, catalysis forming the tetrahedral intermediate, formation of the acyl intermediate and final product resolution (Fig. 4 and Supplementary Fig. 2).

In the earliest snapshot, soaking GlpG crystallized in a bicelle with the substrate resulted in dramatic changes in GlpG (Fig. 4b)
that are indicative of gate opening (Fig. 2). This was vastly more than we observed at any point with tetrapeptide aldehyde binding and catalysis. Remarkably, the entire TM5 helix and most of the overlying L5 loop became disordered (R227–M247), as did even the side chain residues on TM2, including L148-H150, as well as F153 (the key residue that interdigitates with TM5 in the gate-closed form40,41). Even residues within the active site itself became disorganized, to the point that no hydrogen bond linked the catalytic serine and histidine (moved to 3.55 Å apart).

In dramatic contrast, the first interactions with the substrate were already forming and were quite similar to those we observed with the peptide aldehyde: the very distal part of the L5 loop (S248 and M249) became ordered and started interacting with the incoming peptide substrate at the P4, P3 and P2 positions (Fig. 4b). The rest of the substrate (P6–5, P1–P7′) remained entirely too disordered to model.

In the second snapshot we witnessed evidence of catalysis: the side chain of S201 was in continuous electron density with the P1 carbonyl carbon, and the attack occurred from the uncommon Si face (Fig. 4c). TM5 and the L5 loop assumed an ordered conformation in all except four residues (L244–M247), and the gating side chain W236 on TM5 adopted an ordered and inward-facing conformation identical to its position in the apoenzyme structure. Such an extensive interaction network with substrate is often key precisely at the moment of catalysis. Importantly, we could also for the first time observe substrate electron density for the main chain on the prime side. Also notable was how weak the electron density was for more distal prime side residues, and no electron density whatsoever was evident for any side chain or for the main chain of residues P4′–P7′. The lack of strong electron density and no side chain interactions at residues following the cleavage site suggests that no specific binding pockets exist for substrates at these distal sites.

However, to our surprise, at this time point the electron density was actually continuous between the catalytic serine and both the P1 and P1′ residues, which indicates that we had captured the tetrahedral intermediate that has proven so elusive to image directly with natural substrates18,21,22. Accordingly, the oxyanion itself was stabilized by all three hydrogen bonds with GlpG (at distances of 3.19 Å with H150, 3.20 Å with N154 and weakly with the backbone of S201 at a distance of 3.41 Å). The residues S201 and H254 also restored their hydrogen bond at a distance of 3.14 Å, which is radically different from prior structures of tetrahedral-mimicking phosphonates that induced a noncatalytic complex where H254 swung away completely and made no bond with S201 (refs. 42,43). The substrate in the active site formed a remarkably abrupt angle precisely at the catalytic serine, adopting a dramatic >90° bend between the carboxy-terminal and amino-terminal halves (Fig. 4d). This structure indicates that mechanically stressing the peptide bond, and/or stabilizing the tetrahedral geometry once formed, is a key strategy for rhomboid enzymes.

The next two snapshots progressively led to acyl intermediate formation. First, in Snapshot-S3, density for the P1′ residue elongated by shifting slightly further from P1, but they remained continuous (Fig. 4e). In Snapshot-S4 we observed evidence of enzyme acylation: while we could still visualize the P1 residue in...
To distinguish between a purely gating role versus catalytic roles, we installed glycines into the L5 loop (Fig. 5a) to enhance its mobility, and assessed the effect on proteolysis. If its role was purely in gating, increasing loop dynamics should facilitate L5 dissociation from the oxyanion (6.65 Å), but through a smaller rotational angle (76.6°). Two new water molecules occupied the space left by the side chains of P2 and P1′ which, relative to the aldehyde structures, established a longer distance from the oxyanion (2.67–2.60 Å). This motion was blocked by steric clash with the substrate to access the active site. However, witnessing L5 loop flipping outwards: this motion was blocked by steric clash with the substrate carbonyl of the acyl intermediate. In our sixth and final snapshot (Fig. 4h), we observed evidence of catalytic resolution: the substrate was no longer covalently attached to S201 and the substrate electron density weakened which was indicative of dissociation, but, nevertheless, the cleaved product remained bound via hydrogen bonds with all three of the oxyanion-stabilizing groups (at distances of 2.88–2.50 Å with H150, 2.96–2.60 Å with N154 and 3.17 Å with the backbone of S201), and maintained prominent interactions with the overlying L5 loop. The catalytic serine and histidine restored the original hydrogen bond (at a distance of 3.09 Å) that readies them for the next round of catalysis.

A critical role for the L5 loop in catalysis. An entirely unexpected finding that emerged from our ability to visualize rhomboid catalysis from start to finish was the continuous role played by the extra-cellular L5 loop. This loop ‘caps’ the active site in the gate-closed state, and, as such, it is assumed that its only role is to lift away for substrate access to the active site. However, witnessing L5 loop actions at nearly every step of catalysis raised the exciting possibility that it plays a direct role in catalysis by progressively positioning substrate through the various reaction steps.

To distinguish between a purely gating role versus catalytic roles for the L5 loop, we installed glycines into the L5 loop (Fig. 5a) to enhance its mobility, and assessed the effect on proteolysis. If its role was purely in gating, increasing loop dynamics should facilitate L5

---

Table 2 | Data collection and refinement statistics for substrate snapshots

| Data collection | GlpG-RKVRMAAIVFSFP Snapshot-S1 (PDB 6PJ9) | GlpG-RKVRMAAIVFSFP Snapshot-S2 (PDB 6PJ8) | GlpG-RKVRMAAIVFSFP Snapshot-S3 (PDB 6PJ9) | GlpG-RKVRMAAIVFSFP Snapshot-S4 (PDB 6PJQ) | GlpG-RKVRMAAIVFSFP Snapshot-S5 (PDB 6PJU) | GlpG-RKVRMAAIVFSFP Snapshot-S6 (PDB 6PJU) |
|----------------|---------------------------------------------|---------------------------------------------|---------------------------------------------|---------------------------------------------|---------------------------------------------|---------------------------------------------|
| Space group    | C222(1)                                     | C222(1)                                     | C222(1)                                     | C222(1)                                     | C222(1)                                     | C222(1)                                     |
| Cell dimensions| a, b, c (Å) 71.26, 99.00, 63.21             | a, b, c (Å) 71.24, 99.76, 63.14             | a, b, c (Å) 73.03, 98.56, 62.97             | a, b, c (Å) 71.74, 96.45, 62.57             | a, b, c (Å) 71.28, 99.88, 63.18             | a, b, c (Å) 71.50, 96.65, 62.57             |
| α, β, γ (°)    | 90, 90, 90                                  | 90, 90, 90                                  | 90, 90, 90                                  | 90, 90, 90                                  | 90, 90, 90                                  | 90, 90, 90                                  |
| Resolution (Å) | 57.84–2.50 (2.60–2.50)                      | 58.52–2.60 (2.72–2.60)                      | 58.75–2.40 (2.49–2.40)                      | 57.63–2.50 (2.60–2.50)                      | 58.02–2.30 (2.38–2.30)                      | 57.48–2.50 (2.60–2.50)                      |
| Rsym           | 0.089 (0.687)                               | 0.065 (0.536)                               | 0.061 (0.275)                               | 0.041 (0.451)                               | 0.059 (0.277)                               | 0.042 (0.328)                               |
| i/αl           | 4.5 (1.4)                                   | 4.3 (1.8)                                   | 6.4 (1.8)                                   | 10.3 (2.1)                                  | 5.2 (2.0)                                   | 10.0 (1.2)                                  |
| Completeness (%)| 99.2 (99.2)                                 | 99.5 (99.4)                                 | 95.3 (70.5)                                 | 96.0 (77.0)                                 | 99.7 (99.9)                                 | 98.6 (98.2)                                 |
| Redundancy     | 3.9 (4.0)                                   | 3.9 (3.9)                                   | 5.2 (2.0)                                   | 5.6 (3.7)                                   | 4.1 (4.0)                                   | 8.8 (8.6)                                   |
| Refinement     | Resolution (Å) 57.84–2.50 (2.57–2.50)        | 58.52–2.60 (2.67–2.60)                      | 58.75–2.45 (2.51–2.45)                      | 57.63–2.50 (2.57–2.50)                      | 58.02–2.40 (2.46–2.40)                      | 57.48–2.50 (2.57–2.50)                      |
| No. reflections| 7,588                                       | 6,804                                       | 8,007                                       | 7,304                                       | 8,667                                       | 7,282                                       |
| Rwork / Rfree | 0.269 / 0.256 (0.397 / 0.441)               | 0.256 / 0.269 (0.367 / 0.428)               | 0.250 / 0.261 (0.363 / 0.366)               | 0.247 / 0.287 (0.341 / 0.275)               | 0.236 / 0.248 (0.310 / 0.369)               | 0.249 / 0.285 (0.340 / 0.291)               |
| No. atoms      | 1,442                                       | 1,448                                       | 1,442                                       | 1,419                                       | 1,457                                       | 1,395                                       |
| Protein        | 1,430                                       | 1,438                                       | 1,427                                       | 1,403                                       | 1,433                                       | 1,386                                       |
| Ligand/ion     | 0                                           | 0                                           | 0                                           | 0                                           | 0                                           | 0                                           |
| Water          | 12                                          | 10                                          | 15                                          | 16                                          | 24                                          | 9                                           |
| B factors      | 60.2                                        | 56.3                                        | 47.0                                        | 69.5                                        | 49.7                                        | 70.6                                        |
| Protein        | 60.3                                        | 56.4                                        | 47.4                                        | 69.7                                        | 49.6                                        | 70.8                                        |
| Ligand/ion     | 0                                           | 0                                           | 0                                           | 0                                           | 0                                           | 0                                           |
| Water          | 44.3                                        | 44.4                                        | 40.0                                        | 54.5                                        | 52.4                                        | 49.3                                        |
| R.m.s. deviations | Bond lengths (Å) 0.006       | Bond angles (°) 1.47 | Bond angles (°) 1.38 | Bond angles (°) 1.64 | Bond angles (°) 1.46 | Bond angles (°) 1.40 | Bond angles (°) 1.43 |
| *Values in parentheses are for highest-resolution shell.*
cap opening and thus increase the overall catalytic rate. Conversely, if the L5 loop does indeed repeatedly position substrate for catalysis, then introducing glycines should enhance its initial opening but compromise its subsequent interactions with substrate and lead to an overall decrease in catalytic rate.

Consistent with a key role in catalysis, installing glycines into the L5 loop reduced proteolytic activity in reconstituted proteoliposomes by two orders of magnitude (Fig. 5b). This dramatic reduction was entirely reflected in the catalytic rate constant ($k_{cat}$), without any significant change in initial substrate binding (Michaelis constant $K_m$) (Fig. 5c). Importantly, the glycine mutant exhibited no change in its structural stability, as revealed in a very sensitive and quantitative thermostability assay (Fig. 5d). This is consistent with the folding nucleus being centered near the cytosolic apex of TM2 (ref. 44), and subsequent L5 interactions in the closed state not contributing to the structural stability of GlpG. Finally, by installing a spin probe at the central M247 position of the L5 loop, we were able to use electron paramagnetic resonance spectroscopy (EPR) to confirm that the glycines did indeed noticeably enhance loop dynamics (Fig. 5e). Therefore, as predicted by time-resolved crystallography, the L5 loop plays a critical, yet entirely unanticipated, role directly in progressively ushering substrate through the steps of catalysis.

**Discussion**

In summary, we collected ten time-resolved snapshots of rhomboid proteolysis inside the membrane that encompassed the entire reaction, from substrate binding to hydrolytic enzyme deacylation (Fig. 6).
In addition to broadly mapping out the reaction coordinate from start to finish, these snapshots revealed three counterintuitive features as defining hallmarks of rhomboid catalysis. Our first surprise was the extent of gate opening that facilitates binding of even a relatively short (13 residue) transmembrane substrate peptide: in addition to the L5 loop, the entire TM5 segment,
as well as its interacting residues on top of TM2, became disordered. This structural characterization puts to rest the idea that gating deep inside the membrane cannot occur during substrate proteolysis. On the contrary, it reveals that gating involves considerably greater lateral opening than is currently accepted. Importantly, no such changes occurred with short peptidic inhibitors of identical sequence that can enter the active site directly, revealing that natural substrates and inhibitors/short peptides take different paths to the active site.

Another unanticipated feature was the return of the L5 loop after substrate had already entered the active site. The current view is that the L5 loop only has to move away for substrates to access the active site; all subsequent catalytic steps are thought to be caused by local active site residue motions. While this opening is an incontrovertible first step, L5 returning was actually the key step that ‘seemed’ substrate binding in the active site itself through stabilizing interactions. This insight supports a controversial view of intramembrane proteolysis centered on substrate dynamics: rhomboid enzymes initially identify their substrates by interrogating transmembrane helix dynamics where unstable helices that spontaneously unwind will be cut, while those that maintain their secondary structure will not. Our time-resolved crystallography now reveals that L5 loop ‘clamping down’ restrains substrates into a catalytically competent conformation for proteolysis, thereby converting the ‘interrogation’ complex to a ‘scission’ complex, as predicted by inhibition kinetics.

Perhaps even more revealing were L5 motions that progressively formed and loosened distal interactions with substrate throughout the reaction. These unexpected features suggest that it is primarily the actions of the remote, extramembranous L5 loop that progressively repositions substrate as a means to ‘usher’ it through the entire reaction coordinate. This is strongly supported by the dramatic ~100-fold decrease specifically in the catalytic rate $k_c$ displayed by the glycan variant of GlpG, which renders the L5 loop more mobile and thus compromised in its ability to control the position of the substrate during catalysis. Local active site residues also changed position during catalysis, some of which prior mutagenesis experiments support as key for catalysis, while others less so. For example, we and others observed various interactions of W236 with both substrates and inhibitors that imply a role in catalysis. Yet mutagenesis of W236 results in a 5–10-fold increase in proteolysis. Thus, in contrast to the L5 loop, the gating role of W236 dominates its role in the catalytic cycle, with subsequent substrate interactions having minor, if any, meaningful functions during catalysis. More enigmatic is H150, which is both key in stabilizing the oxyanion, but also undergoes a dramatic outward flipping motion. While this may facilitate water entry during the second stage of catalysis, the precise and/or full function(s) of this flipping motion remain unclear, because H150 mutants cannot be studied using an enzymatic approach as they render GlpG catalytically inert. Similar limitations hinder dissecting the multifaceted roles of H254 and N154 (refs. 40,41,48).

The third and most surprising finding in our study was visualizing the elusive tetrahedral intermediate in two separate snapshots as a stable entity. This is very unexpected, because in other serine proteases the tetrahedral intermediate is a short-lived transition state. Visualizing it during rhomboid catalysis suggests that rhomboid is the only serine protease known to have evolved the tetrahedral intermediate to be a stable stalling point, and thus not the transition state. Achieving this unprecedented stability may result from serine attack from the rare Si face, tripartite oxyanion stabilization that only rhomboid is known to employ and/or the effect of accommodating a dramatic >90° bend of the substrate at the point of catalysis. These features have not, to our knowledge, been observed in any other protease. Since serine protease mechanisms historically provided the foundation on which our modern understanding of enzyme theory is based, characterizing the unusual reaction coordinate that our structures mapped out could add a new dimension to our understanding of general enzyme mechanisms. Although measuring energetics is challenging with membrane enzymes, these studies could expose new vulnerabilities for designing novel inhibitors to target rhomboid enzymes selectively.

Finally, while this work provides a structural glimpse of rhomboid proteolysis inside the membrane, prototypic members of other intramembrane protease families have also been crystallized, are able to act on short peptides and have comparably slow catalytic rates. Implementing our time-resolved soaking strategy should, therefore, also enable delineating their gating and catalytic cycles in atomic detail.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41594-019-0296-9.

Received: 26 March 2019; Accepted: 9 August 2019; Published online: 30 September 2019

References

1. Brown, M. S., Ye, J., Rawson, R. B. & Goldstein, J. L. Regulated intramembrane proteolysis: a control mechanism conserved from bacteria to humans. Cell 100, 391–398 (2000).
2. Urban, S. SnapShot: cartography of intramembrane proteolysis. Cell 167, 1898–1908.e1 (2016).
3. Verhelst, S. H. L. Intramembrane proteases as drug targets. FEBS J. 284, 1489–1502 (2017).
4. Dusterhoft, S., Kunzel, U. & Freeman, M. Rhomboid proteases in human disease: mechanisms and future prospects. Biochim. Biophys. Acta 1864, 2200–2209 (2017).
5. De Strooper, B. & Chavez Gutierrez, L. Learning by failing: ideas and concepts to tackle gamma-secretases in Alzheimer’s disease and beyond. Annu Rev. Pharmacol. Toxicol. 55, 419–437 (2015).
6. Urban, S. Making the cut: central roles of intramembrane proteolysis in pathogenic microorganisms. Nat. Rev. Microbiol. 7, 411–423 (2009).
7. Rawson, R. B. et al. Complementation cloning of 52P, a gene encoding a putative metalloprotease required for intramembrane cleavage of SRERPs. Mol. Cell 1, 47–57 (1997).
8. De Strooper, B. et al. A presenilin-1-dependent gamma-secretase-like protease mediates release of Notch intracellular domain. Nature 398, 518–522 (1999).
9. Manolardis, I. et al. Mechanism of farnesylated CAAX protein processing by the intramembrane protease Rce1. Nature 504, 301–305 (2013).
10. Urban, S., Lee, J. R. & Freeman, M. Drosophila rhomboid-1 defines a family of putative intramembrane serine proteases. Cell 107, 173–182 (2001).
11. Kinch, L. N. & Grishin, N. V. Bioinformatics perspective on rhomboid intramembrane protease evolution and function. Biochim. Biophys. Acta 1828, 2937–2943 (2013).
12. Stevenson, L. G. et al. 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 (2007).
13. Baker, R. P., Wijetilaka, R. & Urban, S. Two plasmidium rhomboid proteases preferentially cleave different adhesins implicated in all invasive stages of malaria. PLoS Pathog. 2, e113 (2006).
14. O’Donnell, R. A. et al. Intramembrane proteolysis mediates shedding of a key adhesin during erythrocyte invasion by the malaria parasite. J. Cell Biol. 174, 1023–1033 (2006).
15. Riestra, A. M. et al. A Trichomonas vaginalis rhomboid protease and its substrate modulate parasite attachment and cytolsis of host cells. PLoS Pathog. 11, e1005294 (2015).
16. Lastun, V. L., Grieve, A. G. & Freeman, M. Substrates and physiological functions of secretase rhomboid proteases. Semin. Cell Dev. Biol. 60, 10–18 (2016).
17. Spinauzzi, M. & De Strooper, B. PARL: the mitochondrial rhomboid protease. Semin. Cell Dev. Biol. 60, 19–28 (2016).
18. Hedstrom, L. Serine protease mechanism and specificity. Chem. Rev. 102, 4501–4524 (2002).
19. Samara, N. L., Gao, Y., Wu, J. & Yang, W. Detection of reaction intermediates in Mg$^{2+}$-dependent DNA synthesis and RNA degradation by time-resolved X-ray crystallography. Methods Enzymol. 592, 283–327 (2017).
20. Radisky, E. S., Lee, J. M., Lu, C. J. & Koshland, D. E. Jr Insights into the serine protease mechanism from atomic resolution structures of trypsin reaction intermediates. Proc. Natl Acad. Sci. USA 103, 6835–6840 (2006).
21. Liu, B., Schofield, C. J. & Wilmouth, R. C. Structural analyses on intermediates in serine protease catalysis. *J. Biol. Chem.* **281**, 24024–24035 (2006).
22. Wilmouth, R. C. et al. X-ray snapshots of serine protease catalysis reveal a tetrahedral intermediate. *Nat. Struct. Biol.* **8**, 689–694 (2001).
23. Raper, A. T., Reed, A. J. & Suo, Z. Kinetic mechanism of DNA polymerases: contributions of conformational dynamics and a third divalent metal ion. *Chem. Rev.* **118**, 6000–6025 (2018).
24. Morgan, J. L. et al. Observing cellulose biosynthesis and membrane translocation in *crystallo*. *Nature* **531**, 329–334 (2016).
25. Nango, E. et al. A three-dimensional movie of structural changes in bacteriorhodopsin. *Science* **354**, 1552–1557 (2016).
26. Kupitz, C. et al. Serial time-resolved crystallography of photosystem II using a femtosecond X-ray laser. *Nature* **513**, 261–265 (2014).
27. Segui, M. et al. Light-induced structural changes and the site of O-O bond formation in PSI elucidated by XFEL. *Nature* **543**, 131–135 (2017).
28. Sun, L., Li, X. & Shi, Y. Structural biology of intramembrane proteases: mechanistic insights from rhomboid and S2P to gamma-secretase. *Curr. Opin. Struct. Biol.* **37**, 97–107 (2016).
29. Strizovsky, K., Sharpe, H. J. & Freeman, M. Sequence-specific intramembrane proteolysis: Identification of a recognition motif in rhomboid substrates. *Mol. Cell* **36**, 1048–1059 (2009).
30. Urban, S. & Freeman, M. Substrate specificity of rhomboid intramembrane proteases is governed by helix-breaking residues in the substrate transmembrane domain. *Mol. Cell* **11**, 1425–1434 (2003).
31. Cho, S., Dickey, S. W. & Urban, S. Crystal structures and inhibition kinetics reveal a two-stage catalytic mechanism with drug design implications for rhomboid proteolysis. *Mol. Cell* **61**, 329–340 (2016).
32. Zoll, S. et al. Substrate binding and specificity of rhomboid intramembrane protease revealed by substrate-peptide complex structures. *EMBO J.* **33**, 2408–2421 (2014).
33. Dickey, S. W., Baker, R. P., Cho, S. & Urban, S. Proteolysis inside the membrane is a rate-limited reaction not driven by substrate affinity. *Cell* **155**, 1270–1281 (2013).
34. Ticha, A. et al. Sensitive versatile fluorogenic transmembrane peptide substrates for rhomboid intramembrane proteases. *J. Biol. Chem.* **292**, 2703–2713 (2017).
35. Moin, S. M. & Urban, S. Membrane immersion allows rhomboid proteases to achieve specificity by reading transmembrane segment dynamics. *eLife* **1**, e00173 (2012).
36. Urban, S. & Moin, S. M. A subset of membrane-anchoring agents and gamma-secretase modulators provoke nonsubstrate cleavage by rhomboid proteases. *Cell Rep.* **8**, 1241–1247 (2014).
37. Urban, S. & Wolfe, M. S. Reconstitution of intramembrane proteolysis in vitro reveals that pure rhomboid is sufficient for catalysis and specificity. *Proc. Natl Acad. Sci. USA* **102**, 1883–1888 (2005).
38. Zhou, Y. & Zhang, Y. Serine protease acylation proceeds with a subtle re-orientation of the histidine ring at the tetrahedral intermediate. *Chem. Commun. (Camb.)* **47**, 1577–1579 (2011).
39. Schechter, I. Mapping of the active site of proteases in the 1960s and rational design of inhibitors/drugs in the 1990s. *Curr. Protein Pept. Sci.* **6**, 501–512 (2005).
40. Baker, R. P. & Urban, S. Architectural and thermodynamic principles underlying intramembrane protease function. *Nat. Chem. Biol.* **8**, 759–768 (2012).
41. Baker, R. P., Young, K., Feng, L., Shi, Y. & Urban, S. Enzymatic analysis of a rhomboid intramembrane protease implicates transmembrane helix 5 as the lateral substrate gate. *Proc. Natl Acad. Sci. USA* **104**, 8257–8262 (2007).
42. Xue, Y. & Ha, Y. Catalytic mechanism of rhomboid protease GlpG probed by 3,4-dichloroisocoumarin and diisopropyl fluorophosphonate. *J. Biol. Chem.* **287**, 3099–3107 (2012).
43. Xue, Y. et al. Conformational change in rhomboid protease GlpG induced by inhibitor binding to its S’ subsites. *Biochemistry* **51**, 3723–3731 (2012).
44. Paslawski, W. et al. Cooperative folding of a polytopic alpha-helical membrane protein involves a compact N-terminal nucleus and nonnative loops. *Proc. Natl Acad. Sci. USA* **112**, 7978–7983 (2015).
45. Xue, Y. & Ha, Y. Large lateral movement of transmembrane helix S5 is not required for substrate access to the active site of rhomboid intramembrane protease. *J. Biol. Chem.* **288**, 16645–16654 (2013).
46. Baker, R. P. & Urban, S. Cytosolic extensions directly regulate a rhomboid protease by modulating substrate gating. *Nature* **523**, 101–105 (2015).
47. Wang, Y. & Ha, Y. Open-cap conformation of intramembrane protease GlpG. *Proc. Natl Acad. Sci. USA* **104**, 2098–2102 (2007).
48. Urban, S. & Baker, R. P. In vivo analysis reveals substrate-gating mutants of a rhomboid intramembrane protease display increased activity in living cells. *Biol. Chem.* **389**, 1107–1115 (2008).
49. Drag, M. & Salvesen, G. S. Emerging principles in protease-based drug discovery. *Nat. Rev. Drug Discov.* **9**, 690–701 (2010).
50. Kamp, F. et al. Intramembrane proteolysis of β-amyloid precursor protein by γ-secretase is an unusually slow process. *Biophys. J.* **108**, 1229–1237 (2015).
51. Szaruga, M. et al. Alzheimer’s-causing mutations shift αβ length by destabilizing gamma-secretase-abetan interactions. *Cell* **170**, 443–456 e14 (2017).
52. Bolduc, D. M., Montagna, D. R., Gu, Y. & Wolfe, M. S. Nicasrin functions to sterically hinder γ-secretase-substrate interactions driven by substrate transmembrane domain. *Proc. Natl Acad. Sci. USA* **113**, E509–E518 (2016).

Acknowledgements

We are grateful to current and past members of the Urban laboratory for discussions and support, as well as the expert, friendly and helpful staff at the Cornell High Energy Synchrotron Source (CHESS) for access to beam time. This work was supported by a National Institutes of Health grant (no. R01AI066025 to S.U.). All X-ray diffraction data were collected using instruments at CHESS (beamline supported by National Science Foundation grant no. DMR-0963684 and National Institutes of Health grant no. GM-103485).

Author contributions

S.U. and S.C. conceived the research. S.U. made all DNA constructs. S.C. performed protein purification, enzyme kinetics and thermostability analyses. M.J. performed protein purification, X-ray crystallography and structure determination. R.P.B. conducted EPR spectroscopy. S.U. wrote the paper and all authors approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41594-019-0296-9.

Correspondence and requests for materials should be addressed to S.U.

Peer review information Ines Chen was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019
Methods

Crystallization, data collection and refinement. Crystals of E. coli rhomboid GlpG (the Y205F variant with the amino-terminal 87 residues removed) were obtained in a bicine composed of dimyristoyl phosphatidylcholine:3-[N,N-cholamidopropyl]dimethylammonio]-2-hydroxy-1-propanesulfonate (2.8:1) exactly as described previously31. Reservoir buffer contained 0.1 M NaOAc pH 5.5, 3 M NaCl and 5% ethylene glycol. For time-resolved analysis with the peptide aldehyde inhibitor (Ac-VRMA-CHO), apoenzyme crystals were incubated in 2.5 mM peptide aldehyde dissolved in the same reservoir buffer at room temperature for a series of time points, and flash-frozen in a nitrogen stream for data collection. Specifically, the time points used to generate the inhibitor snapshots in this study were: 30 min (Snapshot-I1), 3 h (Snapshot-I2), 5 h (Snapshot-I3) and 7 h (Snapshot-I4). For the time-resolved analysis with the peptide substrate (Ac-RKVRMAAVFSFP-NH2), the peptide was dissolved at 5 mM in crystallization buffer (a 1:4 mixture of bicine and reservoir buffer) overnight, and undissolved peptide was removed by ultracentrifugation. The apoenzyme crystals were soaked in the peptide solution at room temperature for a series of time points and then flash-frozen in a nitrogen stream for data collection. Specifically, the time points used to generate the substrate snapshots in this study were: 5 min (Snapshot-S1), 2 h (Snapshot-S2), 5 h (Snapshot-S3), 6 h (Snapshot-S4), 7 h (Snapshot-S5) and 3 d (Snapshot-S6). All X-ray diffraction datasets were collected at the F1 station of the Cornell High Energy Synchrotron Source, and were processed with iMosflm v.7.2.2 in the CCP4 program suite53. Structures were solved by molecular replacement using Molrep and further refined by refmac5 and PHENIX with iterative manual model building using Coot54. Final Ramachandran statistics were: 92.2% favored, 0% outliers (Snapshot-I1); 93.5% favored, 0% outliers (Snapshot-I2); 92.0% favored, 0% outliers (Snapshot-I3); 93.1% favored, 0.52% outliers (Snapshot-I4); 100% favored, 0% outliers (Snapshot-S1); 98.3% favored, 0.6% outliers (Snapshot-S2); 97.5% favored, 0.5% outliers (Snapshot-S3); 97.7% favored, 0% outliers (Snapshot-S4); 98.8% favored, 0% outliers (Snapshot-S5); 97.6% favored, 0% outliers (Snapshot-S6).

Enzymatic analysis of intramembrane proteolysis in proteoliposomes. Full-length GlpG proteins were analyzed for protease activity under steady-state conditions, thermostability and protein dynamics, as established previously. Briefly, site-directed mutagenesis was used to engineer a GlpG variant with L5 glycine (L5-Gly), and verified by sequencing the entire GlpG open reading frame. Briefly, site-directed mutagenesis was used to engineer a GlpG variant with L5 glycine (L5-Gly), and verified by sequencing the entire GlpG open reading frame. Intramembrane proteolytic activity was measured in a real-time fluorogenic assay with a 34-residue fluorescein isothiocyanate (FITC)-TatA transmembrane peptide substrate co-reconstituted with GlpG into liposomes formed from E. coli polar lipids30. Briefly, increasing amounts of FITC-TatA substrate, ranging from 37.6 pmol to 1.28 nmol, were co-reconstituted at pH 4 in 1 mg/ml−1 liposomes with 5 pmol of wild-type GlpG or 100 pmol of the L5-Gly variant. Substrate cleavage was monitored for 30 min at 37 °C using a BioTek Synergy H4 plate reader to detect FITC fluorescence. Slopes from linear reaction time courses plotted against substrate concentration in molar percent (relative to liposome phospholipids) were fit using the Michaelis–Menten equation to derive $k_\text{cat}$ and $V_{\text{max}}$ for each enzyme. Average $K_m$ and $k_{\text{cat}}$ values were calculated from three separate biological replicate experiments.

Spectroscopic analyses. Thermostability analysis of wild type and L5-Gly GlpG proteins was performed in a StarGazer instrument (Harbinger Biotechnology) as described55. Proteins were diluted to a concentration of ~2.5 µM in elution buffer, placed in wells of an optical-bottom black 384-well plate (Nunc, 242764) and exposed to increasing temperature from 25 °C to 85 °C at a heating rate of 1 °C min−1. Differential static light scattering was detected from a 620 nm light-emitting diode light source and imaged by charge-coupled device photography every 30 s. Scattering intensity data were extracted from each image and plotted against temperature to generate curves that were fit by a Boltzmann sigmoidal equation (GraphPad Prism). The transition midpoints derived from curve fits were averaged from nine replicates for each protein. Nitroxide spin labeling and continuous-wave EPR spectroscopy was performed on GlpG proteins (lacking their amino-terminal 87 residues and harboring a C104A mutation) reconstituted into liposomes formed from E. coli polar lipids, exactly as has been described56.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Coordinates of all structures have been deposited into the Protein Data Bank under accession codes 6PJ4 (Snapshot-I1), 6PJS (Snapshot-I2), 6PIT (Snapshot-I3), 6P18 (Snapshot-I4), 6P99 (Snapshot-S1), 6P1A (Snapshot-S2), 6P1P (Snapshot-S3), 6PQJ (Snapshot-S4), 6PR (Snapshot-S5) and 6PJU (Snapshot-S6). Source data for Fig. 3b–d are available with the paper online. Any other data are available from the authors upon reasonable request.

References

53. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D 50, 760–765 (1994).

54. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. Acta Crystallogr. D 60, 2126–2132 (2004).

55. Baker, R. P. & Urban, S. An inducible reconstitution system for the real-time kinetic analysis of protease activity and inhibition inside the membrane. Methods Enzymol. 584, 229–253 (2017).

56. Kreutzberger, A. B., Ji, M., Aaron, J., Mihaljevic, L. & Urban, S. Rhomboid distorts lipids to break the viscosity-imposed speed limit of membrane diffusion. Science 363, eaa0076 (2019).
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a | Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
  Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
  Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection | No new or custom code was used in this work. The standard programs used in structure determination and refinement are described in detail in the Methods section.

Data analysis | No new or custom code was used in this work. The standard programs used in structure determination and refinement and analysis of biochemical data are described in detail in the Methods section.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data and materials availability: coordinates of all structures have been deposited into the Protein Data Bank under the accession codes 6P14 (Snapshot-11), 6P15 (Snapshot-12), 6P17 (Snapshot-13), 6P18 (Snapshot-14), 6P19 (Snapshot-51), 6P3A (Snapshot-52), 6P3P (Snapshot-53), 6PJC (Snapshot-54), 6PJU (Snapshot-55), and 6PJU (Snapshot-56). Source data for Figure 5b-d are available with the paper online. Any other data are available upon reasonable request from the authors.
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size | Disclosed for all for all biochemical experiments (with data in graphs and sample size in legend). |
|-------------|--------------------------------------------------------------------------------------------------|
| Data exclusions | Rfree was calculated over reflections in a test set (~5%) that was not included in atomic refinement. |
| Replication | Disclosed for all biochemical experiments (data in graphs, sample size in legend). |
| Randomization | Not applicable |
| Blinding | Not applicable |

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

| Materials & experimental systems | Methods |
|---------------------------------|---------|
| Involved in the study           | Involved in the study |
| Antibodies                      | ChIP-seq |
| Eukaryotic cell lines           | Flow cytometry |
| Palaeontology                   | MRI-based neuroimaging |
| Animals and other organisms     |         |
| Human research participants     |         |
| Clinical data                   |         |