Title

Structural and energetic insights into human rhomboid proteases reveal a unique lateral gating mechanism for orphan family members

Short title

Lateral gating of human rhomboid proteases

Authors

Bryony RJ Clifton[1], Robin A Corey[2,3] and Adam G Grieve[1,3]

Affiliations

[1] School of Biochemistry, University of Bristol, Biomedical Sciences Building, Bristol BS8 1TD

[2] School of Physiology, Pharmacology and Neuroscience, Biomedical Sciences Building, University of Bristol, Bristol, BS8 1TD

[3] Corresponding authors: robin.corey@bristol.ac.uk; adam.grieve@bristol.ac.uk

Abstract

Rhomboid proteases play fundamental roles in human biology and disease by cleaving substrate transmembrane domains. The absence of structural data and the inability to identify substrates for orphan human rhomboids hinder both a comprehensive understanding of their mechanism of action and the development of strategies to modulate their activity. To address this, we characterised the structures, conformational ensembles, and energy landscapes of human rhomboid proteases using AI-based structural models and molecular dynamics simulations. We show that all human rhomboid proteases possess a conformationally dynamic lateral gate within the membrane which transmembrane substrates can enter. Whereas some rhomboids transition between closed and open conformations with minimal energetic cost, orphan rhomboids exhibit unusually narrow lateral gates that require substantial energy to open. These insights reveal an unexpected degree of heterogeneity in substrate engagement mechanisms among human rhomboids, and provide an explanation for the orphan status for evolutionarily recent family members. In doing so, we establish a framework for interrogating membrane protein function.

Teaser

A comprehensive structural, conformational and energetic landscape of human rhomboid proteases

Introduction

Polytopic membrane proteins employ diverse molecular mechanisms to perform their roles, from creating ion pores or protein channels to catalysing complex enzymatic reactions within the membrane bilayer. To understand how these proteins work at the molecular level, their function requires study at the atomic scale. However, many human membrane proteins have proven resistant to structural determination. Advances in single-particle cryo-electron microscopy (cryo-EM) have markedly advanced our understanding of membrane protein function. Nonetheless, membrane proteins with low molecular weights remain technically challenging cryo-EM targets. Furthermore, structural models often capture only isolated snapshots, hindering our ability to fully understand the function of many membrane proteins that frequently rely on transient, metastable intermediate states essential for their activity.

AI-based structure prediction tools such as AlphaFold have revolutionised our ability to obtain insights into 3D biomolecular structures, especially for membrane proteins that are challenging to study using traditional experimental methods (*1*). However, structural predictions for membrane proteins and their membrane-based interactions (in AlphaFold Multimer mode) can be unreliable as they are generated without explicitly accounting for the native membrane environment (*2*). A recent analysis revealed ambiguity and a lack of confidence in many membrane protein structure predictions (*3*). These structural predictions therefore require validation through molecular dynamics (MD) simulations within artificial membrane systems, which allow for capture of realistic conformations and interactions. Despite this, AlphaFold2 (AF2) has shown some promise in being able to predict different conformational states (*4*). A recently developed tool based on the AF2 Evoformer architecture with additional training on MD datasets called Biomolecular Emulator, or BioEmu, has been reported to predict equilibrium conformational ensembles of small- to medium-sized soluble proteins(*5*). In some cases, this may outperform MD at capturing rare conformational states. However, it is currently unclear how well BioEmu can predict conformational ensembles of membrane proteins.

We questioned whether these advances in structural and conformational predictions combined with *in silico* tools for exploring biomolecular mechanics could shed light on the function of dynamic polytopic membrane proteins. As a case study, we analysed the human rhomboids, intramembrane proteases for which no experimental structure has yet been solved (*6, 7*). It is therefore unknown whether features such as their overall structure and conformational dynamics are similar to the structurally solved and well characterised bacterial rhomboid, GlpG (*8, 9*). As dysregulation of rhomboids and intramembrane proteolysis is linked to a variety of human diseases, including Alzheimer's and Parkinson's Disease, it is important that we gain mechanistic insights into intramembrane proteolysis in humans (*10-12*).

The unique structural arrangement of GlpG and gamma-secretase reveal that intramembrane proteases shield their polar active site from the hydrophobic membrane environment and catalyse hydrolysis of transmembrane domains within lipid bilayers (*9, 13-15*). Rhomboid proteases are uniquely characterised by an active site serine, which acts as a nucleophile during proteolysis following abstraction of a proton by the active site histidine residue (*14, 16*). All rhomboids are predicted to share a highly conserved transmembrane fold comprising six transmembrane helices (TMs), termed the rhomboid fold. How substrates access rhomboid active sites remains contentious, but it is generally accepted that rhomboids (and intramembrane proteases more generally) recognise structural elements within transmembrane domains of their substrates, with regions of helical instability enhancing their cleavability (*17-20*). The most prominent model for substrate access is lateral gating, which proposes that transmembrane substrates enter within the plane of the lipid bilayer between transmembrane helices TM2 and TM5 (*21*). In line with this,

preventing the opening of the lateral gate in the bacterial rhomboid GlpG has been shown to prevent proteolysis (*22*). This indicates that conformational regulation of the open or closed states provides a mode for regulating the activity of rhomboid proteases.

The low catalytic efficiency and substrate affinity of the bacterial rhomboid GlpG has been linked to a model in which rhomboids first interrogate their substrates by sampling them within the fold following lateral gate entry. Rhomboids then perform hydrolysis of substrates with helical instability, whereas non-substrates will diffuse out (*23*). Therefore, an open lateral gate is a prerequisite for substrate entry and cleavage. However, this lateral gating mechanism is yet to be confirmed for human rhomboid proteases, as no structural models (with or without substrate) are available.

Mammalian rhomboids can be classified into secretase and mitochondrial proteases. The secretases are rhomboid-like (RHBDL) 1-4, and the mitochondrial rhomboid is presenilin-associated rhomboid-like (PARL) (*7, 24*). As described earlier, the core rhomboid fold has six TMs. RHBDL4 is predicted to have only these six TMs, while RHBDL1-3 have an extra C-terminal domain (designated 6+1), and PARL has an extra N-terminal domain (designated 1+6). PARL is best known for its role in mitochondrial quality control through cleavage of PINK1, a kinase frequently mutated in Parkinsons Disease (*12, 25, 26*). The best characterised mammalian secretase rhomboid, RHBDL2, has several known substrates including single-pass thrombomodulin, ephrin-Bs, adhesion proteins, and more recently the multi-pass membrane protein Orai1 (*20, 27-29*). As a result, it has been implicated in wound healing, cell-cell adhesion, and T cell activation. RHBDL4, like PARL, has also been shown to perform protein quality control. Its known substrates include pre-T cell receptor (pTa), TMED7 and APP, as well as soluble proteins within the lumen of the endoplasmic reticulum (ER), and accordingly it is implicated in ER-associated degradation, inflammatory signalling, and Alzheimer's Disease (*30-34*). The other two mammalian rhomboids, RHBDL1 and RHBDL3, are referred to as the orphans of the family, as no substrates have been identified for them, and so have no ascribed cellular roles. Their conserved active site residues are accessible to activity-based probes, which indicates that they are indeed proteases (*35*). However, as the orphans arose more recently in metazoans, they may have novel roles and biomechanics. As they are enriched in the nervous system, and are linked to ageing and neurological disease, identifying the biological role of orphan rhomboids is crucial for furthering our understanding of these processes (*36, 37*).

Here, we set out to elucidate the mechanisms of human rhomboid proteolysis by employing AI-based tools to predict their structures with and without identified substrates, followed by physics-based MD simulations within membranes to assess the quality of the predicted models and provide conformational dynamics of the rhomboids. Our results reveal that lateral gating appears to be a conserved mechanism for substrate access to human rhomboid proteases. Surprisingly, a comparative analysis reveals that orphan rhomboid proteases present a narrow lateral gate and an atypical energetic requirement for substrate entry, indicating an evolutionarily recent lateral gate adaptation in the rhomboid superfamily. These data provide insights as to why substrates have not yet been identified for RHBDL1 and RHBDL3, and offer paths to discovery of novel mechanisms underlying ageing and neurological disease.

Results

AF2 predictions provide new structural insights into the mammalian rhomboid proteases

With many molecular details surrounding the rhomboid family unclear, including no experimentally-solved structures for any mammalian rhomboid protease, we sought to gain fundamental insights by assessing their predicted structures using AF2 (Fig. 1; for models and quality metrics see https://osf.io/n6348/). Overall, confidence predictions for the AF2 models of human rhomboids were highest for RHBDL2 and PARL (0.84 and 0.85 pTM respectively for rank 1), indicating good predictions (Fig. S1A). RHBDL1 and RHBDL3 models were lower confidence (0.74 and 0.71 respectively), although low confidence regions were generally confined to interhelical loops, or cytoplasmic/IMM domains, which are likely highly flexible, as reflected in pLDDT scores (Fig. S1, B and D). RHBDL4 had the lowest pTM score (0.64), largely due to a highly unstructured C-terminus (Fig. S1E). These cytoplasmic domains (or the IMM domain in PARL) reflect a striking and expected point of heterogeneity between the different rhomboid models. For all systems, the pLDDT scores of the core rhomboid fold were high (Fig. S1, B-G). For all five human rhomboids, the core six TMs comprising the rhomboid fold were highly conserved. Hydrophobic mismatch with the membrane has been found to be central to rhomboid function (*38, 39*); accordingly, all TMs within each rhomboid were observed of different lengths (Fig. S1H). Most models of the rhomboids matched general structural expectations of having seven transmembrane helices, with RHBDL1-3 exhibiting a 6+1 topology, and PARL presenting with a 1+6 topology. However, the AF2 model for RHBDL4 has a 1+6 topology similar to PARL, which contrasts with previous studies indicating it has only six transmembrane domains (*30*). Interestingly, we noticed that the rhomboid fold TMs in RHBDL1-3 appear to have very similar lengths, whereas RHBDL4 and PARL harbour many short TMs which may mirror the membrane thickness of the subcellular compartments to which they target (Fig. S1H). These results indicate an unanticipated similarity between RHBDL4 and PARL (Fig. S2A-D), which is discussed later. Overall, the differences in structures of the human rhomboids mirror their evolutionary relationships, with RHBDL1-3 being most closely related, and PARL and RHBDL4 being divergent from other members of the family (*40-42*).

AF2 models reveal lateral gate differences between characterised and orphan rhomboids

The defining feature of a rhomboid protease is the rhomboid fold, which contains the active serine and histidine dyad that lies behind a lateral gate (*9*). The lateral gate of the bacterial rhomboid GlpG is very well characterised experimentally, and gating has been demonstrated to underpin proteolysis (*43*). To check whether AF2 accurately predicts the structure of the rhomboid fold and therefore provides a route to study lateral gate dynamics in mammalian rhomboids, we compared the lateral gate predictions of human rhomboids to GlpG. Two equivalent residue pairs in the plane of the active site were selected and applied across all rhomboid models for representative lateral gate with quantification. Crystal structures of both an open (PDB 2NRF; lateral gate width ca. 1.55 nm) and closed (PDB 2IC8; lateral gate width ca. 1.09 nm) states of *E.coli* GlpG have been solved (Fig. S3, A and B) (*8, 9*). AF2 predictions of *Ec*GlpG produce a lateral gate width of 1.26 nm (Fig. S3, C-E). This demonstrates that AF2 predictions for lateral gate conformations are in line with solved structures.

AF2 predictions for the human rhomboids reveal a surprising level of heterogeneity in lateral gate width. Top-ranking AF2 models for each rhomboid were structurally aligned to RHBDL2, the best characterised secretase rhomboid in humans. Intriguingly, whilst the RHBDL2 lateral gate is narrower than that of GlpG (Fig. 2, A and F; 1.06 nm and 1.26 nm respectively), the lateral gate was visibly even narrower for the two orphan rhomboids than for RHBDL2, their closest homologue

(Fig. 3B,C; 0.99 nm and 0.96 nm respectively). Indeed, the lateral gates in RHBDL1 and RHBDL3 were narrower than for all rhomboid predictions, even the closed crystal structure of GlpG (ca. 1.09 nm). The lateral gates of both RHBDL4 (1.19 nm) and, to a greater extent, PARL (1.76 nm) are much wider than that of RHBDL2 (1.06 nm) indicating an unexpectedly broad spectrum of lateral gate width across the human proteases (Fig. 2, D and E). The wide lateral gate poses in the RHBDL4 and PARL models may be in part due to a shared beta-sheet structure between TM4 and TM5 (Fig. S2D).

Plotting all five ranks indicate that RHBDL4 predictions were much more heterogeneous (and these also had the lowest pTM scores, Fig. S1A), with 1 of the 5 being as narrow as the orphans (Fig. 2G). This may reflect an ability for AlphaFold2 to predict different biologically relevant conformational states, which has been previously reported, suggesting that the RHBDL4 rhomboid fold is more flexible than the other rhomboids (*4*). As has been reported for GlpG, lateral gate helices and connecting loops show low confidence in nearly all human rhomboid predictions, suggesting they may be dynamic regions across the family (Fig. S1) (*22*).

In summary, the structural predictions of human rhomboids generated with AF2 reveal unique features of the orphan rhomboids in comparison to their better characterised counterparts. Saliently however, none of the models for RHBDL1-4 appear to represent gated or 'open' conformations, when using the lateral gate width of the 'open' model of *Ec*GlpG as a benchmark (1.55 nm) (Fig 2G).

AF2 supports a substrate gating mechanism for mammalian rhomboids

Our recent work revealed that the full-length form of the substrate PINK1 can be modelled within the rhomboid fold of PARL (Fig. 3A) (*44*). In this model, the lateral gate width is ~1.8 nm, the same as seen in our apo models of PARL; both structural models reveal PARL lateral gates that are significantly wider than any of our apo models of the secretase rhomboids. We reasoned that the lateral gate widths measured in the apo models of secretase rhomboids (Fig. 2G) were prohibitively narrow for transmembrane substrate entry, and so must be capable of widening significantly further. To test this hypothesis, we investigated whether AF2 Multimer can model entry of known substrates into the lateral gate of human secretase rhomboids. Accordingly, we ran predictions of RHBDL2 and RHBDL4 with the isolated transmembrane regions of known substrates, either Orai1 TM4 or pTα respectively (Fig. 3, B and C). The resultant structural models of substrate gated RHBDL2 and RHBDL4, where the substrate is positioned adjacent to the rhomboid active site, provide strong support for the model that human rhomboid proteases access their substrates via lateral gating between TM2 and TM5.

In line with our hypothesis, the lateral gates of both RHBDL2 and RHBDL4 significantly widened in the presence of substrate (1.06 nm to 1.42 nm RHBDL2 apo vs substrate gated; 1.19 nm to 1.45 nm RHBDL4 apo vs substrate gated) (Fig. 3D). This was in stark contrast to the PINK1:PARL model, where the lateral gate remained as wide as in the apo models, suggesting that AF2 captures PARL in a fully open state, both in the presence and absence of substrate. RHBDL4 shows a similar, but more heterogeneous profile to PARL, with the rank 1 AF2 apo model showing only a modest widening following gating, whereas ranks 3 and 4 widen as much as RHBDL2. As both RHBDL4 and PARL perform protein quality control functions, this may reflect their ability to constitutively sample the membrane proteome of their respective organelles. Overall, these data highlight that for human rhomboids to gate substrates, their lateral gates must open to ~1.4 nm – just wide enough to fit an α-helix (diameter ca. 1.2 nm).

MD simulations reveal that the lateral gate of characterised secretase rhomboids can transition to open conformations in the absence of substrate

We next questioned whether the absence of a membrane environment prevents AF2 from modelling open conformations for apo forms of RHBDL1-4. Therefore we sought to use atomistic MD simulations to test the stability and validity of these models within a membrane bilayer environment. For each of the human rhomboids and the bacterial rhomboid GlpG, 5 x 500 ns simulations were performed. All rhomboids were modelled in the same lipid bilayer environment to enable direct comparison. Route mean square deviation (RMSD) analysis was then performed to check stability. As the intracellular (or IMM for PARL) domains of all rhomboids displayed unusually high dynamics throughout simulations, additional RMSD calculations were performed in their absence (Fig. S4 and S7). As the simulations converge upon a stable RMSD, typically below 0.3 nm, the AF2 predictions appear to be successfully capturing energetically favourable conformations for the rhomboids, indicating these are valid predictions.

Next, quantification of lateral gate width was performed to assess the relative dynamics of TM2 and TM5 in each of the rhomboids across the duration of the simulations. With GlpG being the only rhomboid with an experimentally determined structure, simulation data for this rhomboid was assessed first to determine whether it is in line with solved states. The GlpG lateral gate shows small fluctuations, but the lateral gate helices remain within a narrow range throughout all of the runs (Interquartile Range/IQR: 1.27-1.36 nm), suggesting that the state predicted by AF2 is an energy minimum (Fig. S5A). To test whether GlpG can transition to an open state we extended Run 1 to 3 μs (Fig. S5B). This revealed that the lateral gate can substantially widen with an outward movement of TM5 and loop 5, demonstrating that GlpG can open beyond the AF2-predicted state in the absence of substrate. Whilst the widest state sampled is ~1.8 nm (Fig. S5C), a state similar to that of the 'open' 2NRF model (~1.6 nm, RMSD 0.31 nm) appears in this extended run as an energy minimum (Fig. S5D). This demonstrates that MD can sample a broad range of biologically relevant conformational states for rhomboids.

We next turned to the human rhomboid proteases. RHBDL2 exhibited a wide range of lateral gate widths across the five simulations (Interquartile range/IQR 1.01-1.17 nm) (Fig. 4A), but the median width was lower than for GlpG simulations (1.08 vs 1.31 nm respectively) (Fig. 4B, Fig. S5, E and F). Strikingly, a subpopulation for which RHBDL2 is as open as GlpG is seen as a second peak on the histogram (~1.3-1.5 nm), suggesting that the MD simulations sample two distinct conformational states (Fig. 4C). Extending out Run 4, which was the major contributor of this peak, reveals that RHBDL2 can stably exist in this open state (approx. 1.5 nm lateral gate width) (Fig. S5, G and H). RHBDL4 also samples a wide range of conformations, but in contrast to RHBDL2, narrows from its starting width (1.39 nm) in several of the runs (Fig. 4B) (IQR 1.18-1.37 nm). This suggests that RHBDL4 has a more dynamic lateral gate than RHBDL2, and is able to transition between open and closed states with a low energetic cost. This also supports the observation that AF2 models a wider range of states for RHBDL4 compared to predictions of other rhomboids (Fig. 2G). Similar lateral gate dynamics were seen for RHBDL4 using a more ER-like lipid composition (Fig. S6), indicating these differences directly reflect features of the RHBDL4 rhomboid fold. Whilst distribution profiles for RHBDL2 and RHBDL4 lateral gate widths across the simulations are distinct with only a small overlap, a subpopulation for RHBDL2 is seen where it reaches a maximum lateral gate width equal to that of RHBDL4 (Fig. 4, C and D).

Unlike the secretase rhomboids, the lateral gate of PARL remains in an extremely open conformation throughout simulations (IQR 1.65-1.76 nm), sampling almost an entirely distinct ensemble from RHBDL2 or RHBDL4 (Fig. S7), pointing to a distinct mechanism for substrate access.

These data suggest both RHBDL2 and RHBDL4 can sample open and closed states at equilibrium with low energetic penalty, with an observed preference for RHBDL2 to exist in a closed state and RHBDL4 in an open state. Snapshots of the narrowest and widest states are shown in Figure 4E-J. Overall, our data reveal that the lateral gates of RHBDL2 and RHBDL4, both of which have identified substrates, can transiently enter stable conformations compatible with substrate entry.

AF2 models of active rhomboids opened by substrates remain in an open conformation in MD simulations

We next questioned whether the presence of substrate allows secretase rhomboids to cross an energy barrier into a stably open state. To test this, simulations were run using earlier AF Multimer models of rhomboid and substrate, but where substrate atoms were deleted (hereafter termed 'gated') (Fig. 3, B and C; Fig. 5, A and E). In the RHBDL2-gated simulations, the lateral gate of RHBDL2 stably remains much wider than in the apo simulations (Fig. 5, B-D). Unlike the RHBDL2-gated model, the RHBDL4-gated model showed little lateral gate widening versus its rank 1 AF2 model (Fig. 5, E, F). Accordingly, the profiles of the apo- and gated simulations are very similar (Fig. 5G, H). Overall, these simulations indicate that the open state for RHBDL2 is stabilised by substrate entry, whereas RHBDL4 is constitutively open, much like PARL (Fig. 3D).

Overall, this supports the hypothesis that the RHBDL2-gated model is likely an energy minimum and additionally predicts that substrate entry facilitates an energy barrier to be crossed.

BioEmu captures open lateral gate conformations for RHBDL2

During the course of this study, the new Biomolecular Emulator (BioEmu) tool was released (*5*). Following the insight from MD simulations that the relatively well-characterised rhomboid RHBDL2 can sample both open and closed states at equilibrium, despite AF2 only predicting the closed state, this provided a timely opportunity to evaluate the suitability of BioEmu for studying the conformational dynamics associated with membrane protein gating.

Using the same human RHBDL2 input sequence, structural predictions were also generated by BioEmu. Whereas all AF2 predictions had a lateral gate width of 1.01-1.08 nm, of 79 total BioEmu models, 27 had a lateral gate width between 1.00 and 1.10 nm; 16 were narrower; but importantly 36 were wider than this range, approaching and sometimes exceeding 1.4 nm (Fig. 6A). This fits closely with the distribution seen in MD simulations of AF2 modelling (Fig. 4A). When compared to AF2 predictions alone (Fig 6A, dotted line), this demonstrates the ability of both AF2-MD and BioEmu to reveal additional conformational ensembles for membrane proteins. Figures 6B and 6C respectively show the narrow ('closed') extreme model (0.73 nm) and widest non-extreme ('open') models (1.45 nm). These closely resemble the narrowest and widest states from equilibrium MD (Fig. 6, E and F), and also demonstrates that the equilibrium MD may be under-sampling the conformational ensemble.

Overall, these data demonstrate that BioEmu, which is currently largely untested for membrane proteins, can successfully predict a conformational ensemble for a mammalian rhomboid protease similar to that achieved by MD.

MD simulations reveal that orphan rhomboids favour closed lateral gate conformations

As MD simulations of RHBDL2 and RHBDL4 in their apo states revealed that they can sample open lateral gate conformations (Fig. 4), we questioned whether the same was true for the orphan rhomboids. However, unlike their more characterised secretase counterparts, MD simulations seeded from the AF2 models for RHBDL1 and RHBDL3 show that their lateral gates remain relatively narrow, and – in the timescale of our simulations – never approach ~1.4 nm (Fig 7). This suggests that the AF2 predictions also represent energy minima (Fig. 2, B and C; Fig. 7, A and B). We observed in simulations that the RHBDL1 model narrows from its input model (0.90 nm input, IQR 0.83-0.91 nm), and the RHBDL3 model widens (0.88 nm input, IQR 0.89-1.01 nm), producing overlapping profiles with discrete tails (Fig. 7C). The small second peak for RHBDL3 (~1.0-1.2 nm) reveals a state which is similar to the AF2 RHBDL2 input model, and, unlike RHBDL1, overlaps with the IQR of RHBDL2 apo simulations (Fig. 7D). Overall, when comparing across all human rhomboids, RHBDL1 and RHBDL3 both have significantly narrower lateral gates, and more limited conformational dynamics throughout simulations (Fig. 7D). This limited flexibility in lateral gates, particularly for RHBDL1, can be seen in snapshots of the narrowest and widest states (Fig. 7E-J). As rhomboid lateral gate opening is required for substrate entry, these data may explain why no substrate for RHBDL1 or RHBDL3 has been identified.

Having observed different profiles of lateral gate dynamics across MD simulations of human rhomboids, a principal component analysis (PCA) was performed to identify what proportion of the overall dynamics in simulations could be attributed to the lateral gate (Figure S8). RHBDL1-3 were used due to their highly conserved core fold, including having near identical TM lengths. Performing PCA across all trajectories for the simulations based on the apo AF2 states revealed that the first eigenvector accounted for ~41% of the total variance in the system (Fig. S8A). The two extreme states for this eigenvector were visualised and overlaid, revealing that eigenvector 1 represents movement of the lateral gate helices (Fig. S8B). Strikingly, separately projecting each of the trajectories for RHBDL1-3 onto this eigenvector reveals that RHBDL1 and -3 cluster together, with a slight overlap, whereas RHBDL2 is entirely distinct (Fig. S8C).

This confirms that lateral gating is a salient feature of rhomboid dynamics and demonstrates, in an unbiased approach, that the orphans display distinct conformational ensembles associated with lateral gating compared to the characterised rhomboid RHBDL2.

BioEmu predicts lateral gate dynamics for the orphan rhomboids

Having seen that BioEmu can capture conformational states for RHBDL2 consistent with substrate gating, including states not captured by equilibrium MD, we used BioEmu to determine whether the orphan rhomboids can exist in equivalent open conformations, which higher energy barriers may prevent equilibrium MD from sampling.

The majority of the orphan models closely match those generated by AF2, where the lateral gate is narrower than the 1.06 nm average width seen for RHBDL2 AF2 models (63 of 79 RHBDL1 models and 39 of 44 RHBDL3 models) (Fig. 8A). However, several models were as wide as the RHBDL2 AF2 apo models (16 RHBDL1 models and 5 RHBDL3 models), or even the ~1.4 nm width seen in RHBDL2 gated models (1 RHBDL1 model and 2 RHBDL3 models). In these open models, a similar outward movement of TM5 to that seen in gated forms of RHBDL2 and RHBDL4 is observed, suggesting BioEmu can capture a rare conformational state for RHBDL1 and RHBDL3 (Fig. 8B-E). It further reinforces the notion that these orphan rhomboids are capable of substrate gating.

These insights indicate that orphan rhomboids can occupy active conformations. However, why these states are not captured by AF2 predictions or equilibrium MD is unknown. A realistic possibility is that a higher energy barrier exists for lateral gate opening in orphan rhomboids.

Steered MD captures the open conformation of orphan rhomboids

To address whether a higher energy barrier is associated with lateral gate opening for the orphan rhomboids, steered MD was performed to gradually drive the human secretase rhomboid proteases into an extreme open state (Fig. 9, A-C depicts this with RHBDL3 as an example; Fig S9A shows the distance profile from this simulation). This allowed for a broad spectrum of lateral gate widths to be sampled, beyond those achieved by equilibrium MD. The conformational landscape associated with lateral gate dynamics could then be assessed using umbrella sampling combined with potential of mean force (PMF) calculations.

To steer open the lateral gate, three paired residues along TM2 and TM5 were selected upon which a pulling force was applied (termed the 'collective variable' or CV). The position of the CV was selected to ensure uniform pulling across the helices in a manner congruent with the gated or open rhomboid models. It should be noted that, as a result, CV distance measurements vary slightly from lateral gate width measurements in previous analyses, which used 2 residue pairs in the plane of the active site. Where relevant, these two measurements will both be compared for clarity.

Equilibrium MD simulations seeded from the output of steered MD were performed to evaluate whether the human secretase rhomboids can stably exist in these extreme open states (Fig. S9). As expected, all rhomboids narrowed from this extreme open state. However, an intermediate state wider than the AF2 apo prediction is seen for RHBDL1 and RHBDL3 (Fig. S9, B and C). Strikingly, this intermediate state is consistent with the open state predicted by BioEmu (Fig. S9, B and C, dashed line as labelled). This independently validates earlier BioEmu data which shows that the orphan rhomboids can stably occupy open lateral gate conformations consistent with gating (>1.4 nm).

Potential of mean force calculations reveal energy barriers associated with lateral gate opening of secretase rhomboids

Following steered MD to force apart lateral gate helices, the free energy landscape associated with this conformational change was explored by selecting frames at regular intervals throughout this pull simulation. These frames were used to seed umbrella sampling simulations, which apply a biasing potential to the CV, acting as an energetic penalty to deviations from the starting conformation within each window. A relative free energy landscape was then constructed using the Weighted Histogram Analysis Method (WHAM) to reweight the data and correct for the biasing introduced during the umbrella sampling (*45*). Energy landscapes associated with lateral gate dynamics were built for RHBDL1-4. In these landscapes, wells represent local minima, or low energy protein conformations. Umbrella histograms show that all states were sufficiently sampled (Fig. S10). For comparison, we have super-imposed the lateral gate widths from AF2 predictions of apo, and where available, gated rhomboid models onto energy landscapes (Fig 9D-G, triangles). Violin plots are also shown for corresponding AF2-MD equilibrium simulations (data replotted from Figs. 4 and 7), using CV quantification for reference.

The energy landscapes show that the RHBDL1 lateral gate is the narrowest of the human rhomboids with a primary energy well seen at ~0.6-0.8 nm (Figure 9D). Applying this CV distance quantification to the original AF2 MD simulations shows that this agrees well with the equilibrium MD data (ca. 0.8 nm; Fig. 7A, C). A series of smaller energy wells are also seen at wider states, e.g. at ~0.9 nm and ~1.1 nm, supporting earlier observations from (Fig. 8A) that these may be rarer conformational states. For RHBDL2, two distinct wells are seen, at CV distances of ~0.95 nm and ~1.25 nm (Fig. 9E). These wells represent the two different states for RHBDL2 sampled in equilibrium MD simulations (ca. 0.9-1.1 nm and 1.3 nm; Fig 4C,D). These two wells are associated with very similar relative free energies and there is a small energy barrier of ca. 3.1 kJ/mol (approximately 1.5 kT) between them. The primary RHBDL3 well is very broad, with a relatively flat region between 0.8-0.95 nm (Fig. 9F). This corresponds well with the equilibrium MD data which shows that RHBDL3 can freely sample a range of states between 0.7-1.0 nm. One well is seen for RHBDL4 between ~1.3-1.4 nm (Fig. 9G), matching AF2 predictions and equilibrium MD data (ca. 1.2-1.4 nm; Fig. 4C,D). However, as the rank 1 RHBDL4 AF2 model was in a more open conformation than those of RHBDL1-3, the narrower states could not be sampled. Presumably, an additional small well would be seen at ca. 1.0 nm were the landscape to be extended downwards.

In summary, we find that models from BioEmu and AF2 indicate that lateral gate distances wider than ~1.4 nm (a distance consistent with substrate gating, Fig. 3) can be sampled by all human rhomboid proteases (Fig. 3, Fig. 6, Fig. 8). RHBDL4 and PARL appear to be able to enter these states with relatively little energetic cost, however the same is not true for RHBDL1-3. Small energy wells around and above 1.4 nm are seen in RHBDL1-3 landscapes, but at greater relative kJ/mol than for RHBDL4, suggestive that these open states are energetically unfavourable without something to stabilise them (discussed later). These findings mirror our equilibrium MD results with apo rhomboids, where lateral gate widths >1.4 nm are frequently observed with RHBDL4, rarely observed with RHBDL2 and never observed with the orphan rhomboids (Fig 9D-G, lower violin plots). Overall, these data reveal that orphan rhomboids have the ability to gate substrates, but compared to all other human rhomboids, higher energy barriers restrict them from occupying open states.

Discussion

The present work demonstrates how insights into functionally relevant protein conformational dynamics can be made using AI-based structure prediction tools combined with physics-based MD. In the example presented here, we address longstanding knowledge gaps surrounding the molecular mechanisms underlying rhomboid intramembrane proteolysis in humans by assessing the validity of the lateral gating model. In doing so, we have revealed distinct mechanisms across human rhomboids and provide insights into the function of the orphan rhomboids RHBDL1 and RHBDL3.

AlphaFold2 models of human rhomboids provide snapshots of an intriguing heterogeneity in lateral gate width (Fig. 2). As this structural feature is hypothesised to be crucial for sampling of substrates by rhomboids, differences in the arrangement of the lateral gate helices could point to differences in the activity of the various family members. MD simulations were performed to test the validity of these models, and these were shown to be stable. Interestingly, rhomboid lateral gates were a dynamic hotspot across simulations and they revealed distinct dynamic profiles for each of the human secretase rhomboids, which provide insights into their functions. For example, RHBDL2 and RHBDL4 have a wider lateral gate in AF2 than the orphans RHBDL1 and RHBDL3, and sample a broader range of conformations at equilibrium (Figs. 2, 4, 7).

Another interesting insight from the AF2 models was the observation that RHBDL4 and PARL contain short TMs, unlike RHBDL1-3 (Fig. S1H). As transmembrane domain length mirrors sub-cellular localisation (*46*), these differences in TM length likely reflect their subcellular localisations, with RHBDL4 being ER-resident, PARL targeting the inner mitochondrial membrane (IMM), and RHBDL1-3 localising to various organelles in the secretory pathway. Surprisingly, RHBDL4 and PARL share unique features not seen in the other rhomboids, perhaps reflecting a shared molecular function. These include a broken TM6 and a beta-sheet in the loop between TM4 and TM5, which sits within the lateral gate (Fig. S2). As RHBDL4 and PARL both feature more open lateral gates across simulations, we hypothesise that this beta-sheet may contribute toward stabilising the lateral gate in an open state. Accordingly, this beta-sheet is a hotspot for dynamics in RHBDL4 MD simulations (Fig. S7, Fig. S11). Combined with the observation that, unlike RHBDL2, RHBDL4 and PARL display little difference in lateral gate distance between apo and gated models, the gating of substrates for the latter two human rhomboids may be less regulated. This would fit with their quality control roles at sites of membrane protein insertion (the ER and inner mitochondrial membrane), where they survey their environment for misfolded or mislocalised proteins(*30, 33, 47*).

The present work presents the case that human rhomboids, like the bacterial rhomboid GlpG (*21, 22, 43*), operate via a lateral gating mechanism. As AF2 predictions and MD simulations of human rhomboids were performed alongside those for GlpG, we could use the GlpG data as a point of reference. The observation that MD data from GlpG AF2 models closely match that of solved structures demonstrates that the AF2 model to MD simulation pipeline used to interrogate the mammalian rhomboids can accurately predict conformational dynamics (Fig. S3, Fig. S5). Across MD simulations of RHBDL1-4, TM5 was consistently responsible for the majority of lateral gate dynamics, with RMSF analysis showing that TM4/5 and TM5/6 loops were hotspots of flexibility (Fig. S11). This is in agreement with previous observations indicating that an outward movement of TM5 drives substrate gating (*21, 22*). Principal component analysis demonstrated that the movement of the lateral gate helices was by far the major contributor to dynamics across simulations of RHBDL1-3, and also independently demonstrates the distinct gating mechanism of orphan rhomboids (Fig. S8).

Data from MD simulations reveal diverse profiles of lateral gate dynamics and, combined with AFM and BioEmu predictions, show that all human rhomboids are capable of occupying states consistent with gating. However, these states are much rarer for the orphan rhomboids RHBDL1 and -3. Steered MD and PMF calculations demonstrate the different energy landscapes associated with lateral gate opening (Fig. 9). For RHBDL2, two energy minima are seen for a narrow and an intermediate state. However, both RHBDL1 and RHBDL3 show only one prominent minimum at a narrow state, supporting observations from equilibrium MD data that these rhomboids favour a closed conformation.

The functional advantage of RHBDL1 and RHBDL3 lateral gates favouring a narrow or more closed conformation is unclear. Perhaps they are unusually specific and only open when they are bound by their exact substrate. Therefore, the reduced sampling of open states in MD simulations directly reflects a reduced capacity to interrogate transmembrane proteins within the crowded membrane environment of a cell. The absence of a second energy minimum at ~1.2 nm (as seen for RHBDL2) in the energy landscape of orphan rhomboids may point to an unprecedented requirement of a lateral gating cofactor for these rhomboid proteases. Such a cofactor may be required for opening of the lateral gate, as our steered MD data indicate that the orphans can stably exist in an open state independently, once an external factor has opened them (Fig. S9). Although all characterised rhomboid proteases to date are capable of cleaving substrates without cofactors, this should not be assumed for RHBDL1 and RHBDL3. RHBDL1-3 show a similar energetic profile around the ~1.3-1.5 nm lateral gate width range corresponding to the open state, yet this conformation is only sampled in equilibrium simulations by RHBDL2. This may indicate that RHBDL1 and RHBDL3 are more tightly regulated, as they may require two separate points of regulation, for instance an interactor to prime the fold, followed by a substrate to fully open the lateral gate. If this speculation were correct, it would be a variation on a previously proposed model where rhomboids are regulated by a requirement to be bound by substrate at both an exosite and the active site for proteolysis to occur (*47, 48*). In this case, the cofactor would bind an exosite, facilitating substrate engagement. The possibility of a membrane-bound regulator is supported by the relationships between the rhomboid pseudoproteases iRhom1/2 and Derlins, with ADAM17 and Hrd1 respectively, which highlight that rhomboid family members can form functional units with other membrane proteins(*49, 50*).

Another possibility is that RHBDL1 and RHBDL3 are regulated via their extended cytoplasmic domains, which both uniquely share a similarly extended N-terminal domain containing predicted and unusual EF hands domains (Fig. 1, Fig. S12) (*51*). The heterogeneity in these cytoplasmic domains across the rhomboid superfamily indicates distinct modes of regulation for localisation, interaction partners and conformational dynamics, which are all likely related to their overall roles. We found that all rhomboids modelled using MD featured unusually dynamic intracellular domains that remained highly flexible throughout the entire simulation runs, despite the core rhomboid fold remaining relatively stable (Figure S4). Stabilisation or regulation of the orphan cytoplasmic domains could propagate conformational changes and lock the proteases into either an open or closed state. This is notably the case for Drosophila rho-4, which requires calcium regulation of cytoplasmic regions for protease activity (*52*).

Finally, it is possible that RHBDL1/3 process a distinct class of protein to that of canonical rhomboid substrates, namely single-pass type I membrane proteins. A substrate which can engage the active site without full opening of the lateral gate would fit this profile, similar to the processing of soluble substrates by RHBDL4 or GlpG (*33, 53*). As the RHBDL family arose in metazoa, and RHBDL1 and RHBDL3 diverged from RHBDL2 later during evolution, these proteases may perform a function which allows metazoa to process other proteins that arose during this time. As the evolution of a nervous system occurred in metazoa, RHBDL1/3 may have evolved to process

neuronal proteins (*54*). If these proteins specifically cleave neuronal proteins, this could also explain why a substrate for them has not yet been identified.

Overall, this work demonstrates that, whilst AF2 predictions can provide important information about biologically relevant conformations of proteins, additional tools such as MD, are required to generate a fuller picture of conformational ensembles. The work also demonstrates the ability of BioEmu, which was developed for medium-sized soluble proteins, to predict conformational ensembles of membrane proteins. The rhomboid proteases have provided a case study for this approach, with these tools providing insights into protease:substrate molecular mechanisms, in addition to furthering our understanding of why orphan rhomboids may currently lack identified substrates. Importantly, this may open the door to the discovery of novel substrates. Although the present work focuses on the rhomboid family, understanding the conformational dynamics of a broad variety of polytopic proteins such as GPCRs, ion channels and transporters, will be essential for discovery of the molecular mechanisms that underlie critical aspects of cell biology and human disease.

Methods

AlphaFold2 modelling

AlphaFold models were generated using AlphaFold2 ColabFold, using a PDB template and relaxation. Coordinate files and confidence scores can be found at https://osf.io/n6348/.

For PARL:PINK1, the models reported previously were used(*44*). Sequences for all models are in Table 1.

Images were generated using ChimeraX(*55*). Structural alignments were performed using ChimeraX using the Matchmaker tool.

AlphaFold2 model	Input sequence
GlpG	All residues of UniProt entry P09391
RHBDL1	All residues of UniProt entry O75783-2
RHBDL2	All residues of UniProt entry Q9NX52
RHBDL3	All residues of UniProt entry P58872
RHBDL4	All residues of UniProt entry Q8TEB9
PARL	Residues 78-379 of UniProt entry Q9H300
RHBDL2:Orai	All residues of UniProt entry Q9NX52; residues 233-257 of Q96D31
RHBDL4:pTa	All residues of UniProt entry Q8TEB9; residues 145-169 of Q6ISU1
PARL:PINK1	Residues 93-368 of UniProt entry Q9H300; all residues Q9BXM7

Table 1 – Input sequences used to generate AF2 models

Lateral gate quantification

To systematically compare lateral gate distances across AF2 models, rhomboids were structurally aligned in ChimeraX and 2 residue pairs which were positioned similarly along transmembrane helices 2 and 5, with one pair above and one pair below the active site. The distance between the alpha-carbon atoms was measured using the measurement wizard and the mean was calculated. Data were then plotted using Prism 10. Residue pairs were as follows:

Rhomboid	Residue pair 1	Residue pair 2
GlpG	G240-M149	W236-N154
RHBDL1	R290-E194	E287-N199
RHBDL2	F230-Q134	D227-N139
RHBDL3	R321-E225	E318-N230
RHBDL4	I183-W79	E179-N84
PARL	D241-F142	I237-Q147

Equivalent residues were also used for quantification of MD simulations, using Gromacs 2024.4 (*56*).

MD simulations

The top-ranking AF2 model was used for MD simulations and generated according with Table 2.

CHARMM-GUI Membrane Builder (*57*) was used to generate the simulation system under a CHARMM36m force field(*58*). Membranes were built from a 3:7 cholesterol:POPC lipid composition using the default parameters. Systems underwent minimisation and equilibration,

using the standard CHARMM-GUI protocol (*59, 60*). Simulations were run using Gromacs 2024.4 (*56*) and performed using the NPT ensemble with V-rescale temperature and Parrinello-Rahman pressure coupling parameters respectively and a timestep of 2 fs. 5 repeats of 500 ns were performed for each simulation and 3 μs extensions were performed.

MD model	**Input model**
GlpG	GlpG AF2 model
RHBDL1	RHBDL1 AF2 model
RHBDL2	RHBDL2 AF2 model
RHBDL3	Residues 35-404 of RHBDL3 AF2 model
RHBDL4	RHBDL4 AF2 model
RHBDL2:Orai	Only the RHBDL2 atoms from the RHBDL2:Orai AF2 model
RHBDL4:pTa	Only the RHBDL4 atoms from the RHBDL2:Orai AF2 model
PARL	Residues 16-292 of PARL AF2 model (94-370 of Q9H300)

Table 2 – input models used to seed MD simulations

Lateral gate helix distance profiles were generated using the earlier described lateral gate quantification. A time step of 2.5 ns was used and a 2nd order smoothing with 4 neighbours was used for the traces. Histograms were generated by appending the distances across 5 simulations into one column and performing a frequency distribution analysis, using Prism. The centre of each bin was plotted along the x axis, and frequency using the 2.5 ns timestep along the y-axis. Violin plots were also generated from the appended tables, and a 'high' smoothing setting was applied.

Biomolecular Emulator

BioEmu predictions were generated by using the same amino acid sequences for human RHBDL1-4 as displayed in Table 1 as inputs in the Google ColabFold notebook. Default parameters were used and the resulting trajectory file was processed using Gromacs to allow individual frames to be extracted as coordinate files for analysis as described.

Principal component analysis

AF2 models of human RHBDL1, RHBDL2 and RHBDL3 were aligned to identify atoms conserved atoms across the core fold, which are shown in Table 3. These residues were taken forward to process trajectories of AF2 MD simulations into one concatenated simulation, with only the conserved C-alpha atoms from these trajectory files used for PCA. Using Gromacs tools, PCA was performed to identify the largest scale movements seen across the 15 x 500 ns simulations. First, a covariance matrix was built and diagonalised before the eigenvectors were projected onto each of the rhomboid simulations, allowing them to be plotted against their eigenvalues. For eigenvector 1, the two extreme states represented by this eigenvector were overlaid.

AF2 model	**Residues used in PCA**
RHBDL1	100-110, 112-299, 302-365
RHBDL2	41-303
RHBDL3	131-141, 143-330, 333-396

Table 3 – Residues used in PCA for RHBDL1-3

Steered MD

Minimised and equilibrated systems of RHBDL1-4 were subjected to steered MD using Gromacs, as above, using Nose-Hoover and Parrinello-Rahman temperature and pressure coupling. A collective variable (CV) was defined as 3 alpha-carbon atoms on each of TM2 and TM5, as in Table 4. A constant velocity of 35 nm/ms was applied in the x and y coordinate axes, and simulations were run for 325 ns (RHBDL2); 400 ns (RHBDL1 and RHBDL3); or 260 ns (RHBDL4). 3 repeats were run and assessed visually in VMD(*61*) to check for artefacts such as bending of helices. RMSD and x-coordinate plots were also assessed to check for smooth changes throughout pulling simulations. The best of the 3 against these metrics was then taken forward for umbrella sampling simulations.

Rhomboid	TM2	TM5
RHBDL1	E194, F198, L202	E287, R290, L294
RHBDL2	Q134, G138, L146	I219, L226, F230
RHBDL3	Q225, L229, and L233.	E318, R321, L325.
RHBDL4	W79, F83, S87	E179, A182, F186

Table 4 – Residues used as collective variable for steered MD

Potential of mean force calculations

Following steered MD to force apart TM2 and TM5, umbrella windows were selected, with ~0.05 nm spacing, using the pmf.py script written by Dr. Owen Vickery (DOI: 10.5281/zenodo.3592318). This produced 29 windows for RHBDL1; 24 for RHBDL2; 30 windows for RHDBL3; and 18 for RHBDL4. Each window was equilibrated for 10 ps, followed by a production run of 50 ns. An additional 50 ns was run to check for convergence and, in all instances, the original 50 ns showed good convergence and was therefore used for subsequent analyses.

The same script was then used to perform potential of mean force (PMF) calculations. The weighted histogram analysis method (WHAM) was then applied to generate an energy landscape (*45*).

References

1. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
2. R. Evans *et al.*, Protein complex prediction with AlphaFold-Multimer. *bioRxiv*, 2021.2010.2004.463034 (2022).
3. M. A. Jambrich, G. E. Tusnady, L. Dobson, How AlphaFold2 shaped the structural coverage of the human transmembrane proteome. *Sci Rep* **13**, 20283 (2023).
4. G. Monteiro da Silva, J. Y. Cui, D. C. Dalgarno, G. P. Lisi, B. M. Rubenstein, High-throughput prediction of protein conformational distributions with subsampled AlphaFold2. *Nat Commun* **15**, 2464 (2024).
5. S. Lewis *et al.*, Scalable emulation of protein equilibrium ensembles with generative deep learning. *Science*, eadv9817 (2025).
6. N. Kühnle, V. Dederer, M. K. Lemberg, Intramembrane proteolysis at a glance: from signalling to protein degradation. *J Cell Sci* **132**, (2019).
7. V. L. Lastun, A. G. Grieve, M. Freeman, Substrates and physiological functions of secretase rhomboid proteases. *Semin Cell Dev Biol* **60**, 10-18 (2016).
8. Z. Wu *et al.*, Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry. *Nat Struct Mol Biol* **13**, 1084-1091 (2006).
9. Y. Wang, Y. Zhang, Y. Ha, Crystal structure of a rhomboid family intramembrane protease. *Nature* **444**, 179-180 (2006).
10. C. Haass *et al.*, Normal cellular processing of the beta-amyloid precursor protein results in the secretion of the amyloid beta peptide and related molecules. *Ann N Y Acad Sci* **695**, 109-116 (1993).
11. M. S. Wolfe, Inhibition and modulation of gamma-secretase for Alzheimer's disease. *Neurotherapeutics* **5**, 391-398 (2008).
12. G. Shi *et al.*, Functional alteration of PARL contributes to mitochondrial dysregulation in Parkinson's disease. *Hum Mol Genet* **20**, 1966-1974 (2011).
13. A. Ben-Shem, D. Fass, E. Bibi, Structural basis for intramembrane proteolysis by rhomboid serine proteases. *Proc Natl Acad Sci U S A* **104**, 462-466 (2007).
14. S. Urban, J. R. Lee, M. Freeman, Drosophila rhomboid-1 defines a family of putative intramembrane serine proteases. *Cell* **107**, 173-182 (2001).
15. X. Guo *et al.*, Molecular mechanism of substrate recognition and cleavage by human γ-secretase. *Science* **384**, 1091-1095 (2024).
16. J. R. Lee, S. Urban, C. F. Garvey, M. Freeman, Regulated intracellular ligand transport and proteolysis control EGF signal activation in Drosophila. *Cell* **107**, 161-171 (2001).
17. S. Urban, M. Freeman, Substrate specificity of rhomboid intramembrane proteases is governed by helix-breaking residues in the substrate transmembrane domain. *Mol Cell* **11**, 1425-1434 (2003).
18. M. K. Lemberg, B. Martoglio, Requirements for signal peptide peptidase-catalyzed intramembrane proteolysis. *Mol Cell* **10**, 735-744 (2002).
19. J. Ye, U. P. Davé, N. V. Grishin, J. L. Goldstein, M. S. Brown, Asparagine-proline sequence within membrane-spanning segment of SREBP triggers intramembrane cleavage by site-2 protease. *Proc Natl Acad Sci U S A* **97**, 5123-5128 (2000).
20. A. G. Grieve *et al.*, Conformational surveillance of Orai1 by a rhomboid intramembrane protease prevents inappropriate CRAC channel activation. *Mol Cell* **81**, 4784-4798.e4787 (2021).
21. R. P. Baker, K. Young, L. Feng, Y. Shi, S. Urban, Enzymatic analysis of a rhomboid intramembrane protease implicates transmembrane helix 5 as the lateral substrate gate. *Proc Natl Acad Sci U S A* **104**, 8257-8262 (2007).

22. C. Bohg *et al.*, The opening dynamics of the lateral gate regulates the activity of rhomboid proteases. *Sci Adv* **9**, eadh3858 (2023).
23. S. W. Dickey, R. P. Baker, S. Cho, S. Urban, Proteolysis inside the membrane is a rate-governed reaction not driven by substrate affinity. *Cell* **155**, 1270-1281 (2013).
24. M. Freeman, The rhomboid-like superfamily: molecular mechanisms and biological roles. *Annu Rev Cell Dev Biol* **30**, 235-254 (2014).
25. A. M. Pickrell, R. J. Youle, The roles of PINK1, parkin, and mitochondrial fidelity in Parkinson's disease. *Neuron* **85**, 257-273 (2015).
26. C. Meissner, H. Lorenz, A. Weihofen, D. J. Selkoe, M. K. Lemberg, The mitochondrial intramembrane protease PARL cleaves human Pink1 to regulate Pink1 trafficking. *J Neurochem* **117**, 856-867 (2011).
27. J. C. Pascall, K. D. Brown, Intramembrane cleavage of ephrinB3 by the human rhomboid family protease, RHBDL2. *Biochem Biophys Res Commun* **317**, 244-252 (2004).
28. O. Lohi, S. Urban, M. Freeman, Diverse substrate recognition mechanisms for rhomboids; thrombomodulin is cleaved by Mammalian rhomboids. *Curr Biol* **14**, 236-241 (2004).
29. N. Johnson *et al.*, Quantitative proteomics screen identifies a substrate repertoire of rhomboid protease RHBDL2 in human cells and implicates it in epithelial homeostasis. *Sci Rep* **7**, 7283 (2017).
30. L. Fleig *et al.*, Ubiquitin-dependent intramembrane rhomboid protease promotes ERAD of membrane proteins. *Mol Cell* **47**, 558-569 (2012).
31. S. Paschkowsky, M. Hamzé, F. Oestereich, L. M. Munter, Alternative Processing of the Amyloid Precursor Protein Family by Rhomboid Protease RHBDL4. *J Biol Chem* **291**, 21903-21912 (2016).
32. Y. C. M. Penalva *et al.*, Loss of the APP regulator RHBDL4 preserves memory in an Alzheimer’s disease mouse model. *Cell Death & Disease* **16**, 280 (2025).
33. J. Bock *et al.*, Rhomboid protease RHBDL4 promotes retrotranslocation of aggregation-prone proteins for degradation. *Cell Rep* **40**, 111175 (2022).
34. J. D. Knopf *et al.*, RHBDL4-triggered downregulation of COPII adaptor protein TMED7 suppresses TLR4-mediated inflammatory signaling. *Nat Commun* **15**, 1528 (2024).
35. E. V. Wolf, A. Zeissler, S. H. Verhelst, Inhibitor Fingerprinting of Rhomboid Proteases by Activity-Based Protein Profiling Reveals Inhibitor Selectivity and Rhomboid Autoprocessing. *ACS Chem Biol* **10**, 2325-2333 (2015).
36. A. Kumar *et al.*, Age-associated changes in gene expression in human brain and isolated neurons. *Neurobiol Aging* **34**, 1199-1209 (2013).
37. A. S. Fröhlich *et al.*, Single-nucleus transcriptomic profiling of human orbitofrontal cortex reveals convergent effects of aging and psychiatric disease. *Nat Neurosci* **27**, 2021-2032 (2024).
38. A. J. B. Kreutzberger, M. Ji, J. Aaron, L. Mihaljević, S. Urban, Rhomboid distorts lipids to break the viscosity-imposed speed limit of membrane diffusion. *Science* **363**, (2019).
39. O. Engberg *et al.*, Rhomboid-catalyzed intramembrane proteolysis requires hydrophobic matching with the surrounding lipid bilayer. *Sci Adv* **8**, eabq8303 (2022).
40. S. Urban, S. W. Dickey, The rhomboid protease family: a decade of progress on function and mechanism. *Genome Biol* **12**, 231 (2011).
41. M. K. Lemberg, M. Freeman, Cutting proteins within lipid bilayers: rhomboid structure and mechanism. *Mol Cell* **28**, 930-940 (2007).

42. Q. Li, N. Zhang, L. Zhang, H. Ma, Differential evolution of members of the rhomboid gene family with conservative and divergent patterns. *New Phytol* **206**, 368-380 (2015).
43. S. Urban, R. P. Baker, In vivo analysis reveals substrate-gating mutants of a rhomboid intramembrane protease display increased activity in living cells. *Biol Chem* **389**, 1107-1115 (2008).
44. J. S. Lorriman, A. G. Grieve, R. A. Corey, I. Collinson, Alternative Import-Channels And Destinations Of Mitochondrial PINK1 Controlled By Trans-Membrane-Domain Structural Plasticity. *bioRxiv*, 2024.2011.2006.622366 (2024).
45. S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, P. A. Kollman, THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry* **13**, 1011-1021 (1992).
46. H. J. Sharpe, T. J. Stevens, S. Munro, A comprehensive comparison of transmembrane domains reveals organelle-specific properties. *Cell* **142**, 158-169 (2010).
47. L. Lysyk *et al.*, Insights into the catalytic properties of the mitochondrial rhomboid protease PARL. *J Biol Chem* **296**, 100383 (2021).
48. E. Arutyunova *et al.*, Allosteric regulation of rhomboid intramembrane proteolysis. *Embo j* **33**, 1869-1881 (2014).
49. F. Lu *et al.*, Cryo-EM reveals that iRhom2 restrains ADAM17 protease activity to control the release of growth factor and inflammatory signals. *Mol Cell* **84**, 2152-2165.e2155 (2024).
50. X. Wu *et al.*, Structural basis of ER-associated protein degradation mediated by the Hrd1 ubiquitin ligase complex. *Science* **368**, (2020).
51. F. Madeira *et al.*, The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Res* **52**, W521-w525 (2024).
52. R. P. Baker, S. Urban, Cytosolic extensions directly regulate a rhomboid protease by modulating substrate gating. *Nature* **523**, 101-105 (2015).
53. S. Maegawa, K. Ito, Y. Akiyama, Proteolytic action of GlpG, a rhomboid protease in the Escherichia coli cytoplasmic membrane. *Biochemistry* **44**, 13543-13552 (2005).
54. G. E. Budd, Early animal evolution and the origins of nervous systems. *Philos Trans R Soc Lond B Biol Sci* **370**, (2015).
55. E. F. Pettersen *et al.*, UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **30**, 70-82 (2021).
56. D. Van Der Spoel *et al.*, GROMACS: fast, flexible, and free. *J Comput Chem* **26**, 1701-1718 (2005).
57. E. L. Wu *et al.*, CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *J Comput Chem* **35**, 1997-2004 (2014).
58. J. Huang *et al.*, CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* **14**, 71-73 (2017).
59. J. Lee *et al.*, CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J Chem Theory Comput* **12**, 405-413 (2016).
60. S. Jo, T. Kim, V. G. Iyer, W. Im, CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* **29**, 1859-1865 (2008).
61. W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics. *J Mol Graph* **14**, 33-38, 27-38 (1996).

Acknowledgments

We thank Mark Dodding (University of Bristol) for proof-reading the manuscript and providing critical feedback. BRJC is funded via the Wellcome Trust Dynamic Molecular Cell Biology PhD programme (218510/Z/19/Z). AGG is funded by a Sir Henry Dale fellowship jointly funded by the Wellcome Trust and Royal Society (221784/Z/20/Z). This project made use of time on ARCHER2 granted via the UK High-End Computing Consortium for Biomolecular Simulation, HECBioSim (http://hecbiosim.ac.uk), supported by EPSRC (EP/R029407/1). The authors acknowledge the use of resources provided by the Isambard 3 Tier-2 HPC Facility. Isambard 3 is hosted by the University of Bristol and operated by the GW4 Alliance (https://gw4.ac.uk) and is funded by UK Research and Innovation; and the Engineering and Physical Sciences Research Council (EP/X039137/1).

Author contributions

Conceptualisation: BRJC, RAC, AGG

Investigation: BRJC

Methodology: BRJC, RAC

Supervision: RAC, AGG

Visualisation: BRJC, RAC, AGG

Writing: BRJC, RAC, AGG

Competing interests

Authors declare that they have no competing interests.

Data and materials availability

PDB files for all AF2 models, including confidence scores, can be found at the following online repository: https://osf.io/n6348/.

All other data are available in the main text or the supplementary materials.

Figure Legends

Fig. 1 – AlphaFold2 predictions of mammalian rhomboids. Cartoon representations of rank 1 models for each of the human rhomboid proteases. Active site residues are in red. Core rhomboid folds are similar but soluble domains show high heterogeneity. For RHBDL1-3, the cytoplasmic domain is N-terminal, whereas this is C-terminal for RHBDL4 and PARL due to their 1+6 topology in the models. Additionally, RHBDL1-3 possess two short amphipathic helices between the core fold and the cytoplasmic domain, whereas RHBDL4 and PARL appear to lack these. Whereas the N-terminal domain of RHBDL2 is a single, continuous helix, for RHBDL1 and -3, this domain is extended and folded into a folded domain comprising two EF-hand motifs (see Fig. S12). In contrast, the cytoplasmic domain is predicted to be largely unstructured for RHBDL4. For PARL, this IMM domain is the end of one continuous helix extending down from TM6 and terminating in a small unstructured region.

Fig. 2 – Lateral gate measurements for AlphaFold2 predictions of human rhomboids and *E. coli* GlpG. Equivalent residue pairs in TM2 and TM5 were used for lateral gate quantification of human RHBDL2, all other human rhomboids, and EcGlpG (measurements indicated with dashed line). Images show cartoon representations of the rhomboids, looking through the lateral gate into the active site (yellow for GlpG, red for all other rhomboids). Lateral gate helices (TM2 and TM5) are labelled. **(A)** RHBDL2, **(B-E)** Structural alignments of human rhomboids to RHBDL2. **(F)** Structural alignment of RHBDL2 and GlpG. **(G)** Plot of lateral gate distances for all five AF2 model ranks of each rhomboid. Bar = mean. For comparison, the lateral gate width for the solved structures of GlpG in its closed (2IC8) and open (2NRF) conformation as depicted by dotted lines.

Fig. 3 – AlphaFold2 predicts substrate gating for human rhomboids. (A-C) Overlays of structurally aligned rank 1 AF2 predictions of apo human rhomboids and substrate-bound human rhomboids as labelled. **(D)** Quantification of lateral gate widths for rhomboids either with or without substrate. Dashed lines indicate the threshold for LG width associated with gating of substrate (1.4 nm), which is the width in the rank 1 model of RHBDL2:Orai (post equilibration in a lipid bilayer). Bar = mean value. One-way ANOVA with Šídák's multiple comparisons test. One asterisk represents $P < 0.033$; two asterisks represent $P < 0.002$; three asterisks represent $P < 0.001$.

Fig. 4 – Molecular dynamics simulations of characterised secretase rhomboids RHBDL2 and RHBDL4. Lateral gate distance profiles for **(A)** RHBDL2 and **(B)** RHBDL4 during molecular dynamics simulations. Rank 1 AF2 models were used as inputs following generation of a 7:3 POPC:cholesterol lipid bilayer and energy minimisation. For each rhomboid, five 500 ns simulations were performed. Each simulation is shown as a separate line. Dashed line shows the input lateral gate width (post equilibration in a lipid bilayer). **(C)** Histograms show the lateral gate width distributions across simulations shown in (A) and (B). **(D)** Violin plots comparing RHBDL2 and RHBDL4 lateral gate width profiles**. (E, F)** Snapshots of RHBDL2 narrowest and widest states from (A). **(G)** Overlay of structural alignments of (E) and (F). **(H, I)** Snapshots of RHBDL4 narrowest and widest states from (B). **(J)** Overlay of structural alignments of (H) and (I).

Fig. 5 – Molecular dynamics simulations of gated rhomboids. AFM predictions of gated rhomboids were seeded into simulations with atoms from the transmembrane substrate deleted.

(A) Overlay of the structurally aligned rank1 AF2 model of RHBDL2 apo (light blue) with the AFM model of 'gated' RHBDL2 (navy blue). **(B)** Lateral gate distance profiles from five 500 ns equilibrium simulations seeded with the RHBDL2 gated model shown in (A). Each simulation is shown as a separate line. Dashed line shows the input lateral gate width (post equilibration in a lipid bilayer). **(C)** Histogram showing the frequency distribution of lateral gate widths from RHBDL2 simulations using the apo AF2 input model or the gated AFM model. **(D)** Violin plot comparing lateral gate width distribution in simulations of the RHBDL2 apo or gated models. **(E)** Overlay of the structurally aligned rank1 AF2 model of RHBDL4 apo (light green) with AFM model of RHBDL4 gated (dark green). **(F)** Lateral gate distance profiles from five 500 ns equilibrium simulations seeded with the RHBDL4 gated model shown in (E). Dashed line shows the input lateral gate width. **(G)** Histogram showing the frequency distribution of lateral gate widths of RHBDL4 AF2 apo or AFM gated models. **(H)** Violin plots comparing lateral gate width distribution from simulations using the RHBDL4 apo or gated models.

Fig. 6 – BioEmu predictions of RHBDL2. (A) The full-length amino acid sequence for the human RHBDL2 was used as the input for the Biomolecular Emulator, hosted on ColabFold. The individual frames from the resulting trajectory file were extracted using Gromacs and the average distance between the representative lateral gate residue pairs were plotted for each frame. This is plotted against the average distance between the representative lateral gate residue pairs for equilibrium MD simulations using AF2 input models, with a point plotted every 2.5 ns across five 500 ns simulations. **(B)** Cartoon representation of the narrowest frame from BioEmu (lateral gate width 0.73 nm). **(C)** Cartoon representation of the widest frame from BioEmu (lateral gate width 1.45 nm), besides the extremely wide outlier. **(D)** Overlay of (B) with the narrowest state seen in AF2 equilibrium MD simulations of RHBDL2. **(E)** Overlay of (C) with the widest state seen in AF2 equilibrium MD simulations of RHBDL2.

Fig. 7 – Molecular dynamics simulations of orphan secretase rhomboids show stably narrow lateral gates. Lateral gate width profiles for **(A)** RHBDL1 and **(B)** RHBDL3 during molecular dynamics simulations. Rank 1 AF2 models were used as inputs following generation of a 7:3 POPC:cholesterol lipid bilayer and energy minimisation. For each rhomboid, five 500 ns simulations were performed. Each simulation is shown as a separate line. Dashed lines show the input lateral gate width (post equilibration in a lipid bilayer). **(C)** Histograms show the lateral gate width distributions across simulations shown in (A) and (B). **(D)** Violin plots comparing lateral gate width profiles of all human rhomboid simulations. **(E, F)** Snapshots of narrowest and widest states of RHBDL1 from (A). **(G)** Overlay of structural alignments of (E) and (F). **(H, I)** Snapshots of narrowest and widest states of RHBDL3 from (B). **(J)** Overlay of structural alignments of (H) and (I).

Fig. 8 – BioEmu reveals that orphan rhomboids can exist in open conformations. (A) The full-length amino acid sequences for human RHBDL1-4 were used as inputs for the Biomolecular Emulator, hosted on ColabFold. The individual frames from the resulting trajectory files were extracted using Gromacs and the average distance between the representative lateral gate residue pairs were plotted for each frame. This is plotted against the average distance between the representative lateral gate residue pairs for equilibrium MD simulations using AF2 input models, with a point plotted every 2.5 ns across five 500 ns simulations. **(B)** Cartoon representation of the

widest frame from the BioEmu RHBDL1 prediction ('BioEmu open') (lateral gate width 1.47 nm). **(C)** Overlay of the RHBDL1 BioEmu open model with rank 1 AF2 RHBDL1 model. **(D)** Cartoon representation of the widest frame from the RHBDL3 BioEmu prediction ('BioEmu open') (lateral gate width 1.53 nm). **(E)** Overlay of the RHBDL3 BioEmu open model with the rank 1 AF2 RHBDL3 model.

Fig. 9 – Steered molecular dynamics and energy landscapes associated with gating of secretase rhomboids. Constant velocity steered molecular dynamics were applied to alpha-carbon atoms within 3 residue pairs on TM2 and TM5 ('collective variable' or CV), as indicated by dashed grey lines. Following steered molecular dynamics of RHBDL1-4, frames were selected using even spacing of ~0.05 nm and used to seed umbrella sampling simulations. Potential of mean force (PMF) calculations were performed, allowing an energy landscape associated with lateral gate dynamics to be constructed. **(A)** Cartoon representation of the input model. **(B)** Cartoon representation of the final frame from the pull simulation of RHBDL3 ('post-pull'). **(C)** Overlay of (A) and (B). **(D-G)** Relative energy landscape for RHBDL1, -2, -3 and -4 respectively (top) and corresponding violin plots showing the distribution of lateral gate widths in equilibrium MD simulations of apo models, using equivalent CV quantification as was used for steered MD and PMFs. Dashed lines indicate lateral gate width for either apo or gated AF2 models (post equilibration in a lipid bilayer).

Figure 1

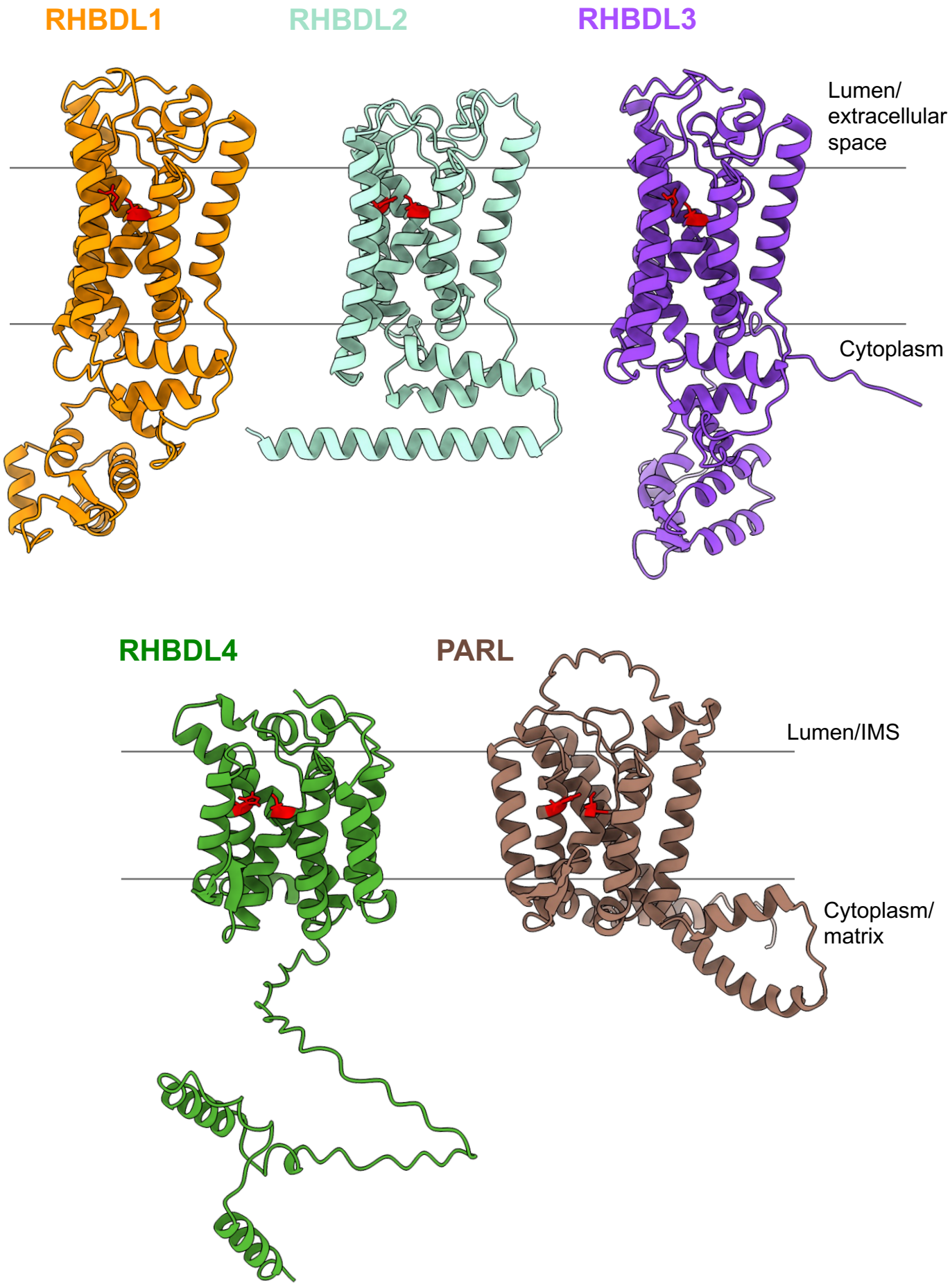
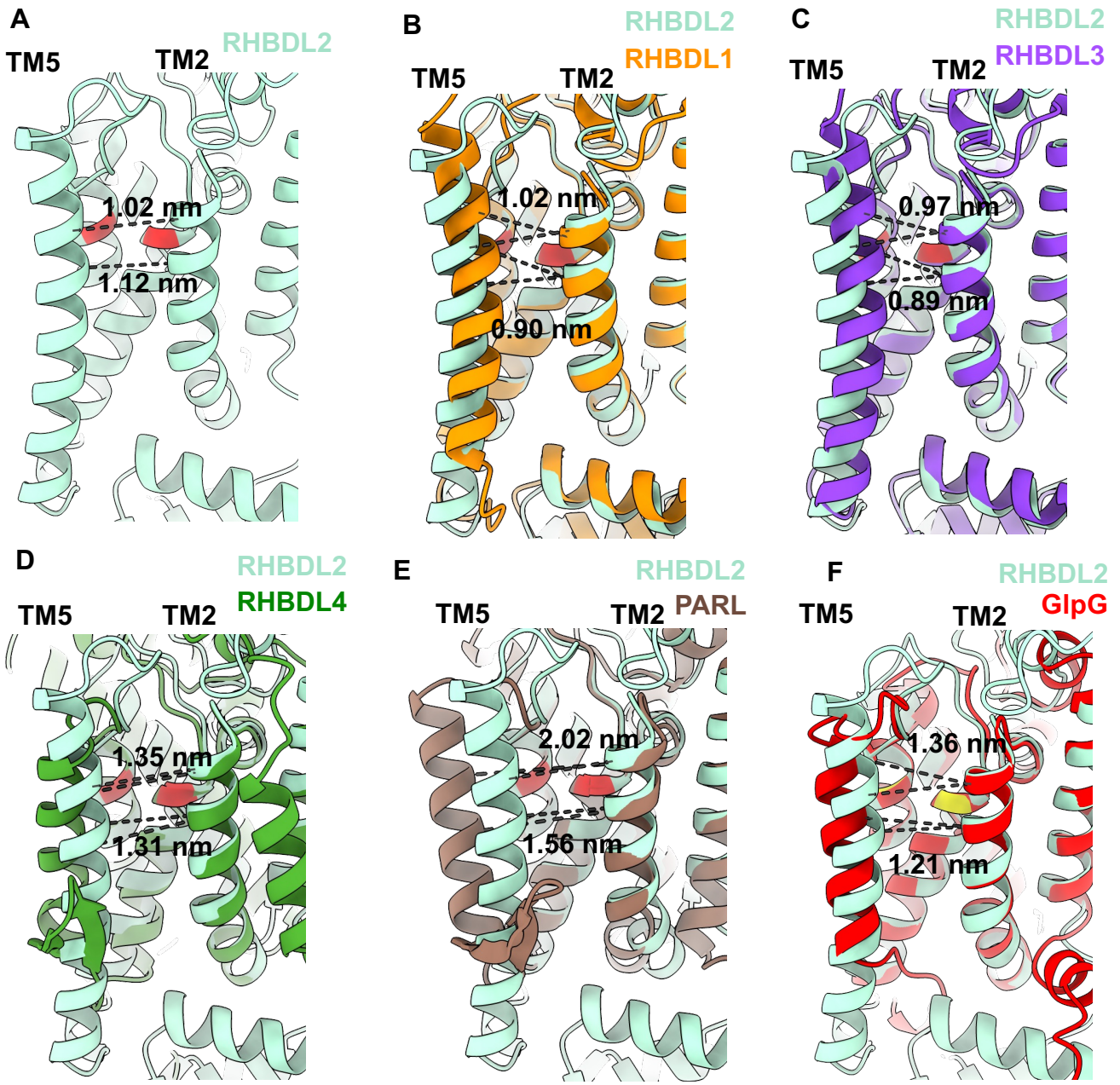

RHBDL1
RHBDL2
RHBDL3
Lumen/
extracellular
space
Cytoplasm
RHBDL4
PARL
Lumen/IMS
Cytoplasm/
matrix

Figure 2


A
TM5
TM2
RHBDL2
1.02 nm
1.12 nm
B
RHBDL2
RHBDL1
TM5
TM2
1.02 nm
0.90 nm
C
RHBDL2
RHBDL3
TM5
TM2
0.97 nm
0.89 nm
D
RHBDL2
RHBDL4
TM5
TM2
1.35 nm
1.31 nm
E
RHBDL2
PARL
TM5
TM2
2.02 nm
1.56 nm
F
RHBDL2
GlpG
TM5
TM2
1.36 nm
1.21 nm

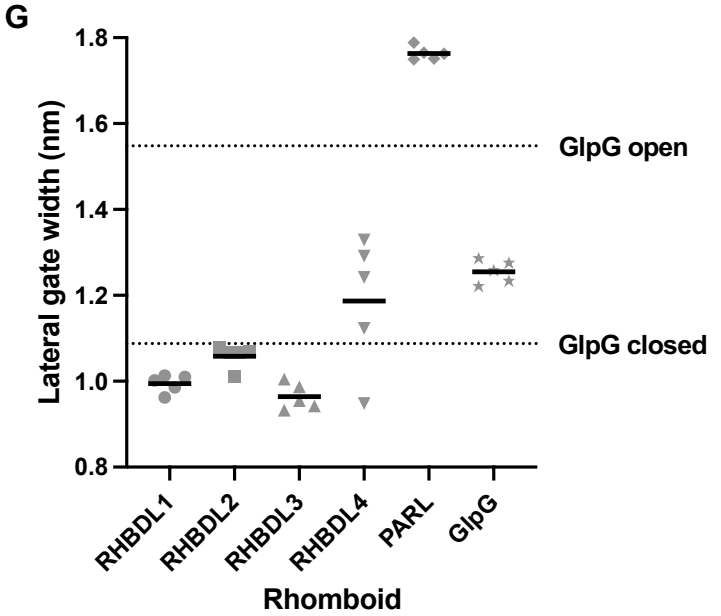
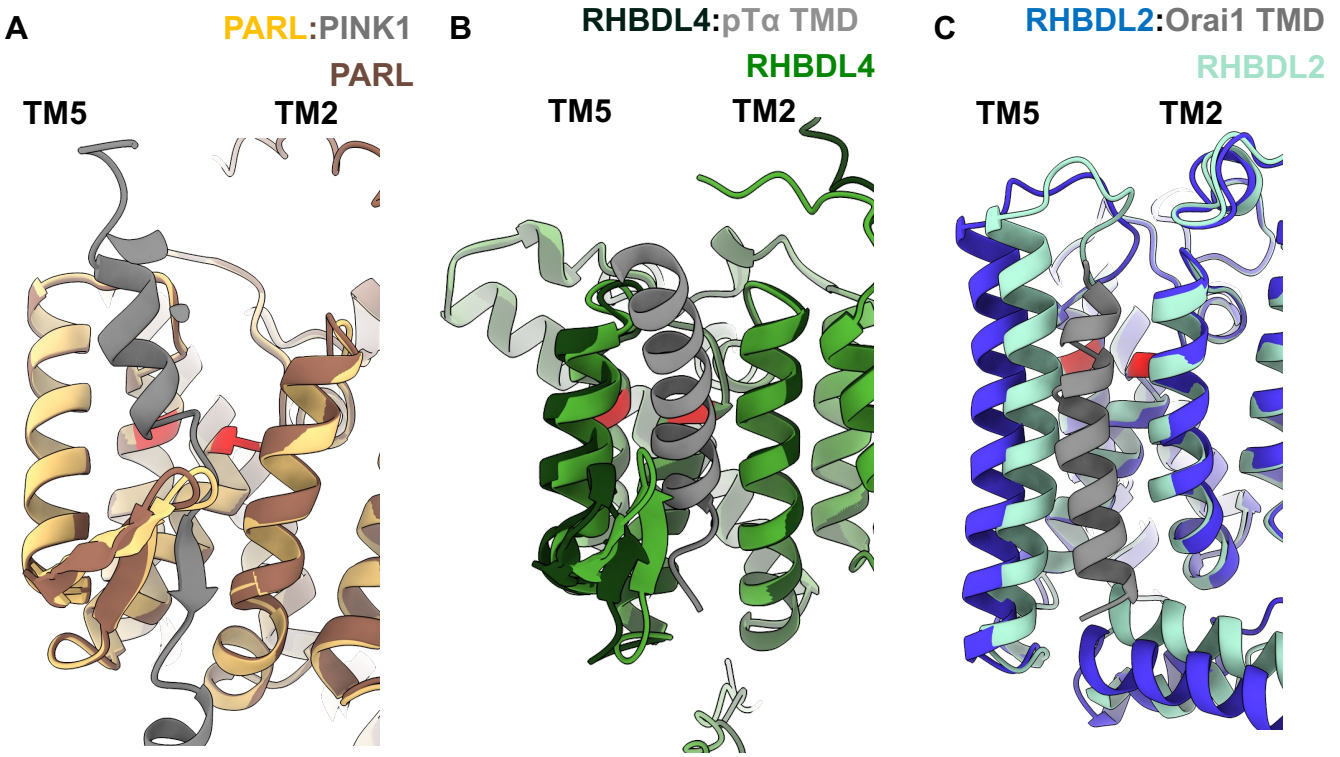

G
Lateral gate width (nm)
1.8
1.6
1.4
1.2
1.0
0.8
GlpG open
GlpG closed
RHBDL1
RHBDL2
RHBDL3
RHBDL4
PARL
GlpG
Rhomboid

Figure 3


A
PARL:PINK1
PARL
TM5
TM2
B
RHBDL4:pTα TMD
RHBDL4
TM5
TM2
C
RHBDL2:Orai1 TMD
RHBDL2
TM5
TM2

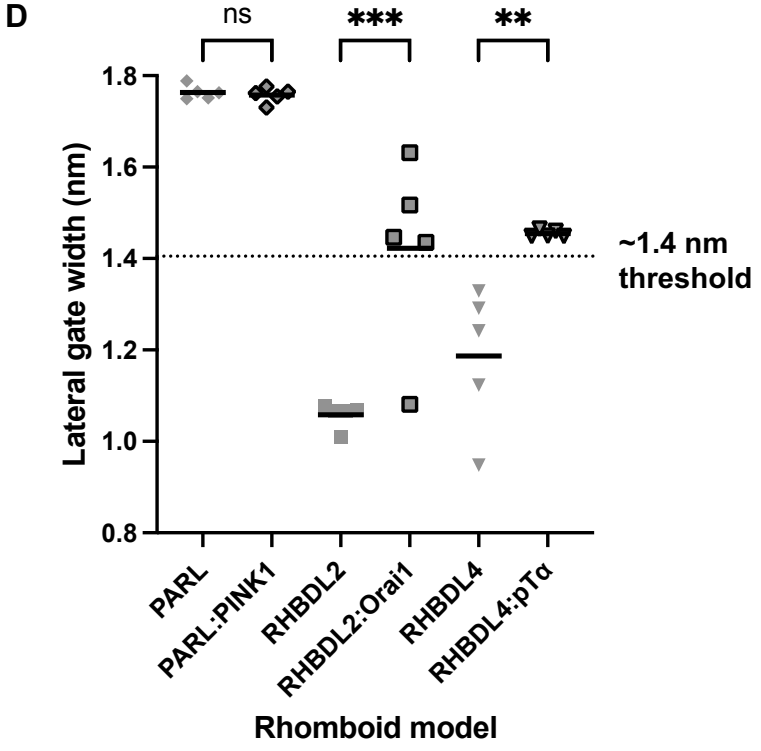

D
ns
***
**
1.8
1.6
1.4
1.2
1.0
0.8
Lateral gate width (nm)
~1.4 nm
threshold
PARL
PARL:PINK1
RHBDL2
RHBDL2:Orai1
RHBDL4
RHBDL4:pTα
Rhomboid model

Figure 4

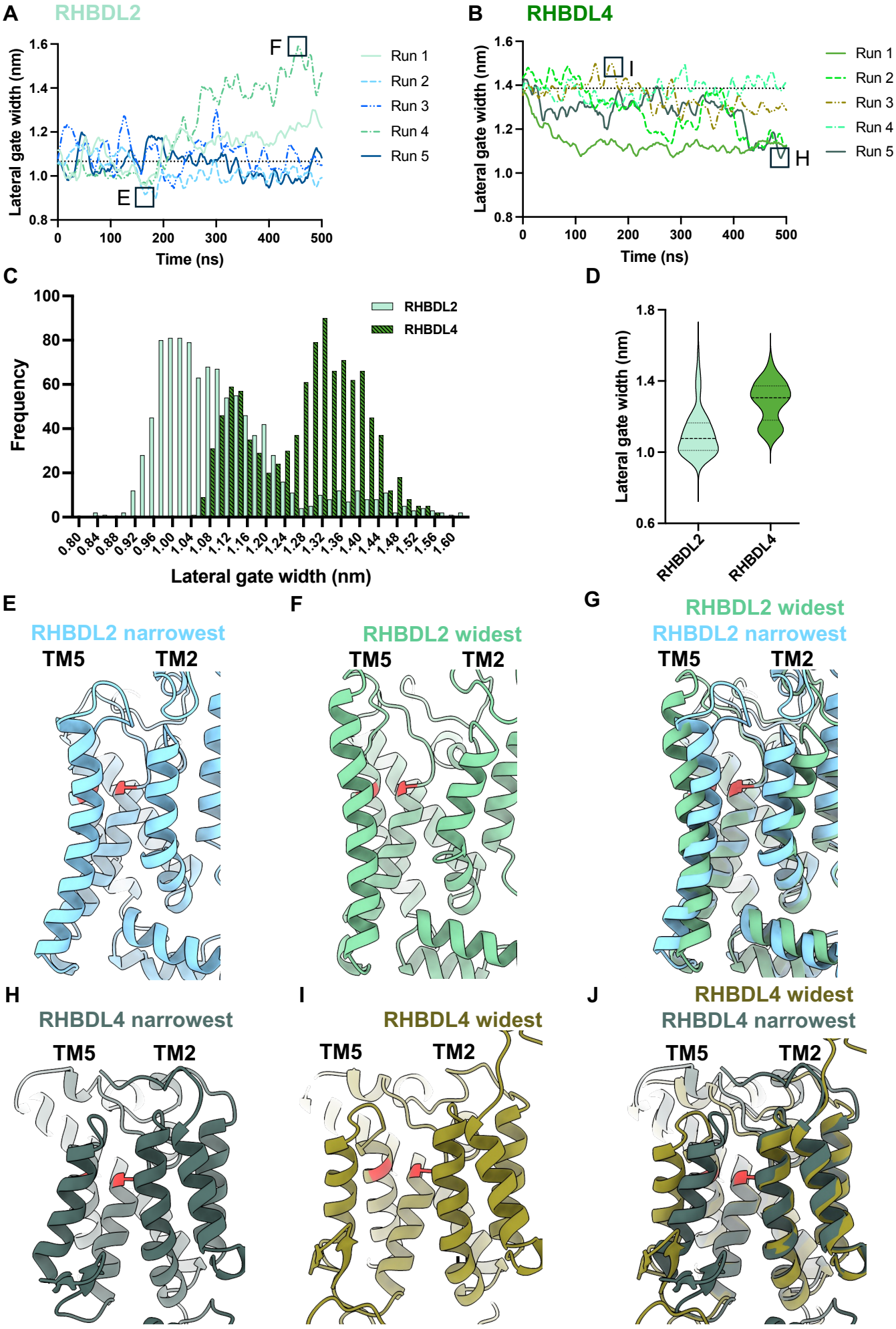

A
RHBDL2
Lateral gate width (nm)
Time (ns)
Run 1
Run 2
Run 3
Run 4
Run 5
E
F
B
RHBDL4
I
H
C
Frequency
Lateral gate width (nm)
RHBDL2
RHBDL4
D
Lateral gate width (nm)
RHBDL2
RHBDL4
E
RHBDL2 narrowest
TM5
TM2
F
RHBDL2 widest
TM5
TM2
G
RHBDL2 widest
RHBDL2 narrowest
TM5
TM2
H
RHBDL4 narrowest
TM5
TM2
I
RHBDL4 widest
TM5
TM2
J
RHBDL4 widest
RHBDL4 narrowest
TM5
TM2

Figure 5

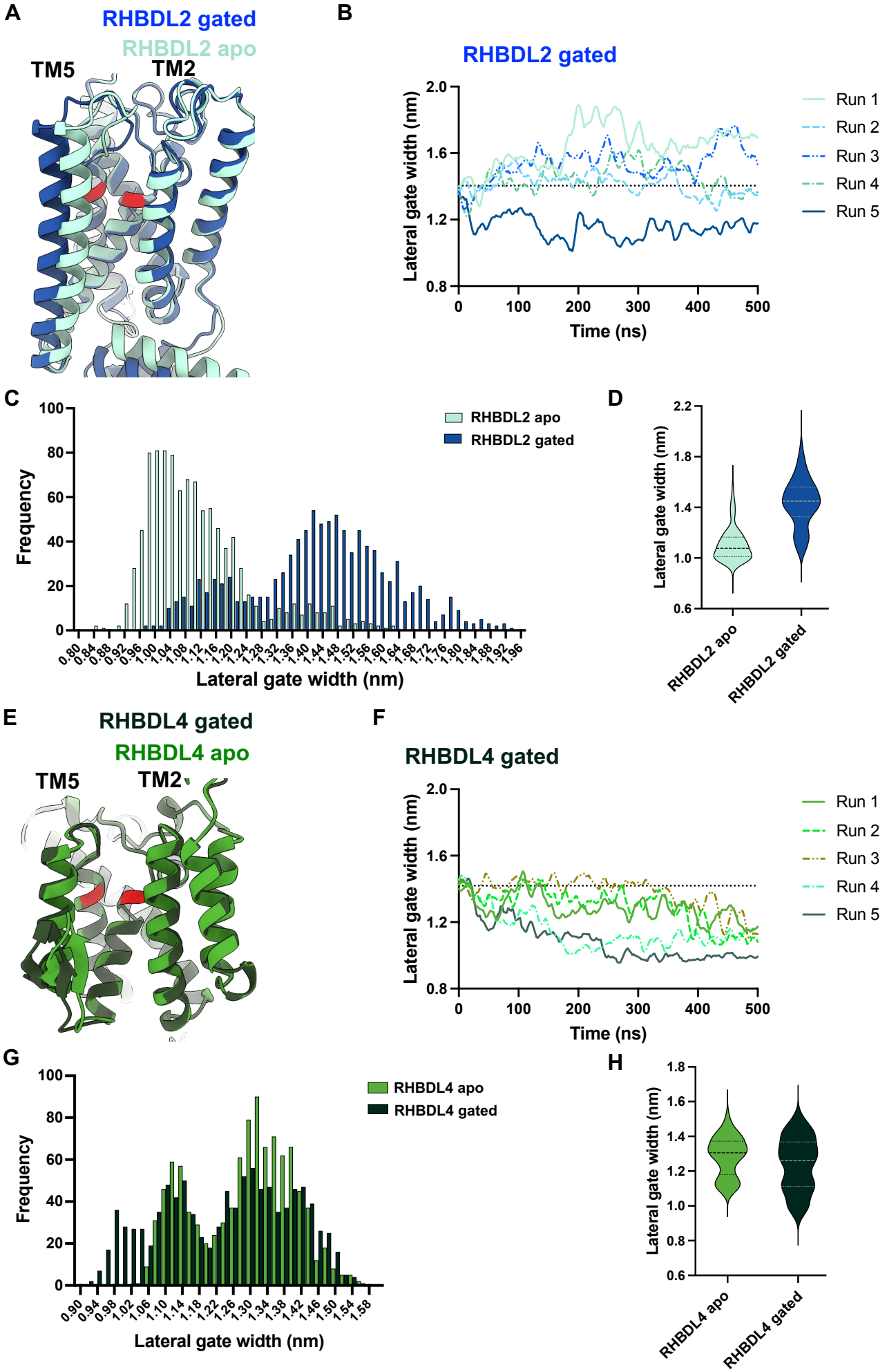

A
RHBDL2 gated
RHBDL2 apo
TM5
TM2
B
RHBDL2 gated
Lateral gate width (nm)
Time (ns)
Run 1
Run 2
Run 3
Run 4
Run 5
C
Frequency
Lateral gate width (nm)
RHBDL2 apo
RHBDL2 gated
D
Lateral gate width (nm)
RHBDL2 apo
RHBDL2 gated
E
RHBDL4 gated
RHBDL4 apo
TM5
TM2
F
RHBDL4 gated
Lateral gate width (nm)
Time (ns)
Run 1
Run 2
Run 3
Run 4
Run 5
G
Frequency
Lateral gate width (nm)
RHBDL4 apo
RHBDL4 gated
H
Lateral gate width (nm)
RHBDL4 apo
RHBDL4 gated

Figure 6

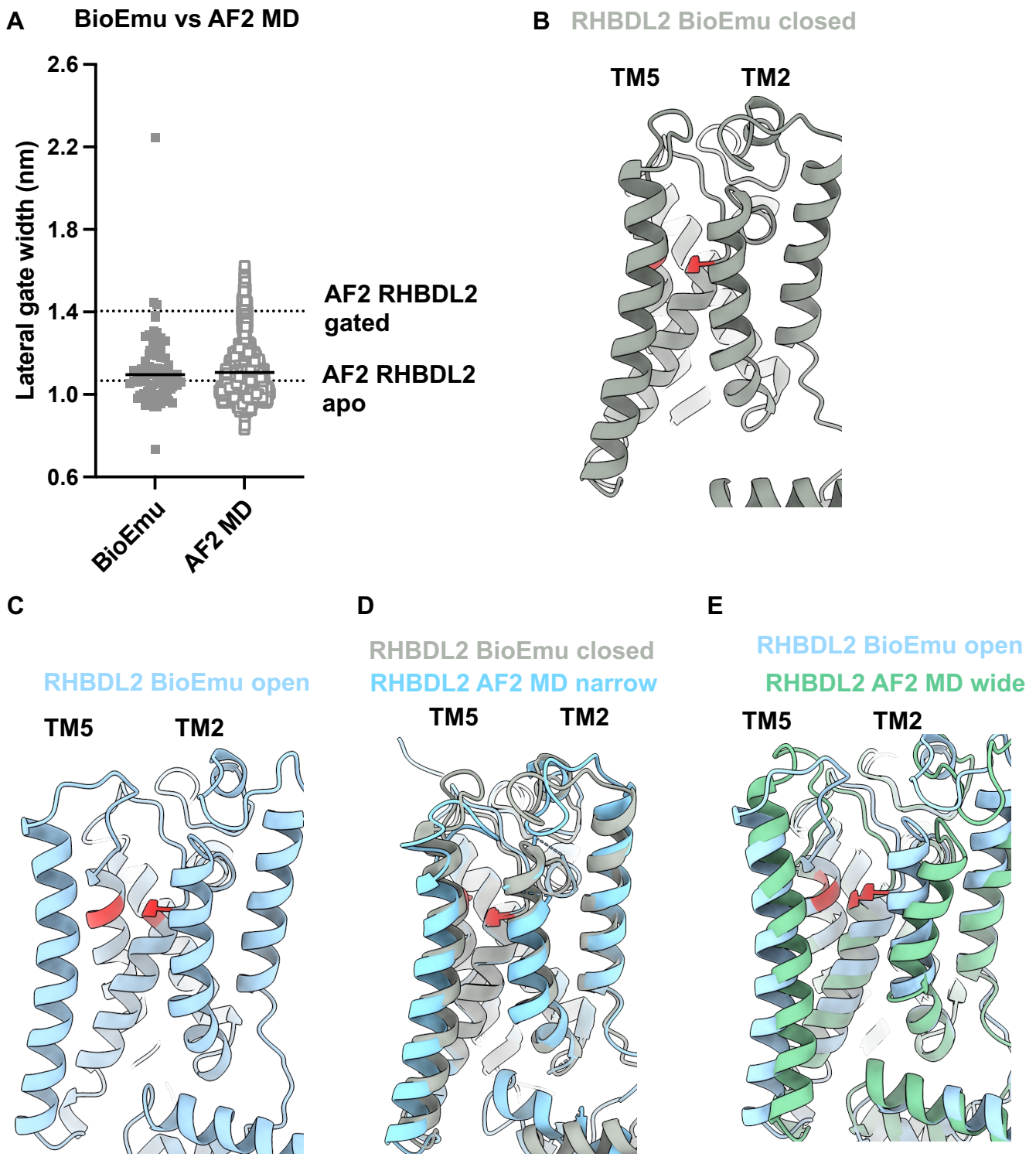
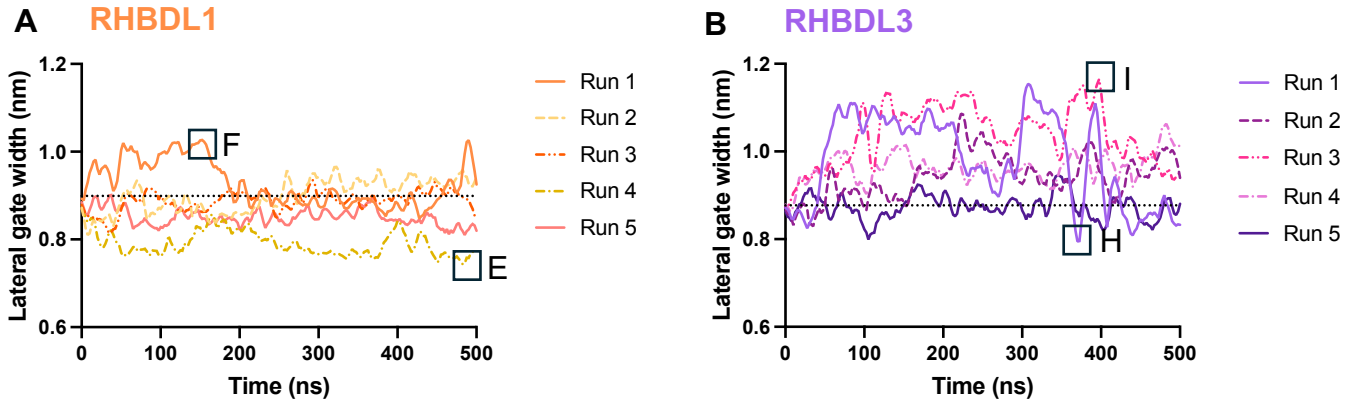
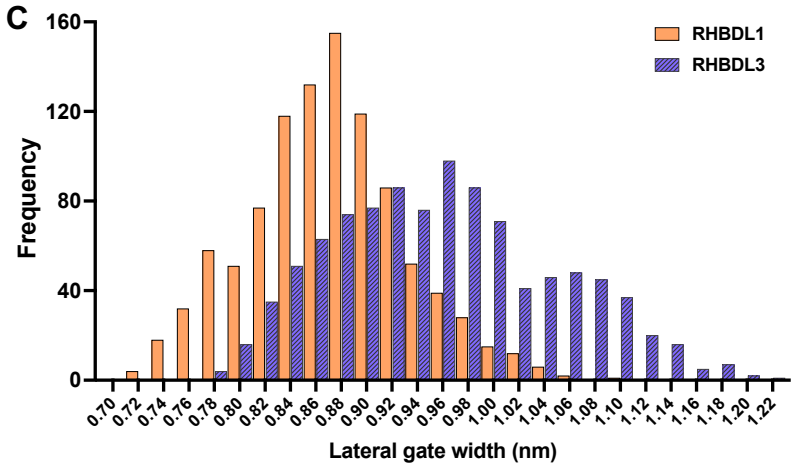
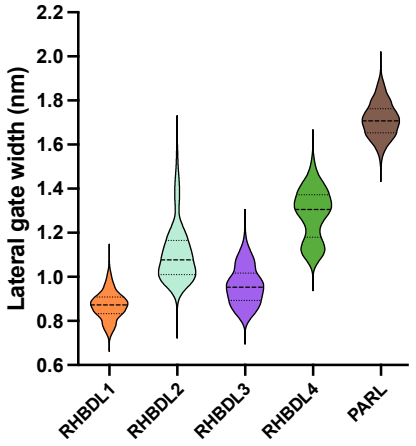

A
BioEmu vs AF2 MD
Lateral gate width (nm)
2.6
2.2
1.8
1.4
1.0
0.6
BioEmu
AF2 MD
AF2 RHBDL2 gated
AF2 RHBDL2 apo
B
RHBDL2 BioEmu closed
TM5
TM2
C
RHBDL2 BioEmu open
TM5
TM2
D
RHBDL2 BioEmu closed
RHBDL2 AF2 MD narrow
TM5
TM2
E
RHBDL2 BioEmu open
RHBDL2 AF2 MD wide
TM5
TM2

Figure 7


A
RHBDL1
Lateral gate width (nm)
Time (ns)
Run 1
Run 2
Run 3
Run 4
Run 5
F
E
B
RHBDL3
Lateral gate width (nm)
Time (ns)
I
H


C
Frequency
Lateral gate width (nm)
RHBDL1
RHBDL3


D
Lateral gate width (nm)
RHBDL1
RHBDL2
RHBDL3
RHBDL4
PARL

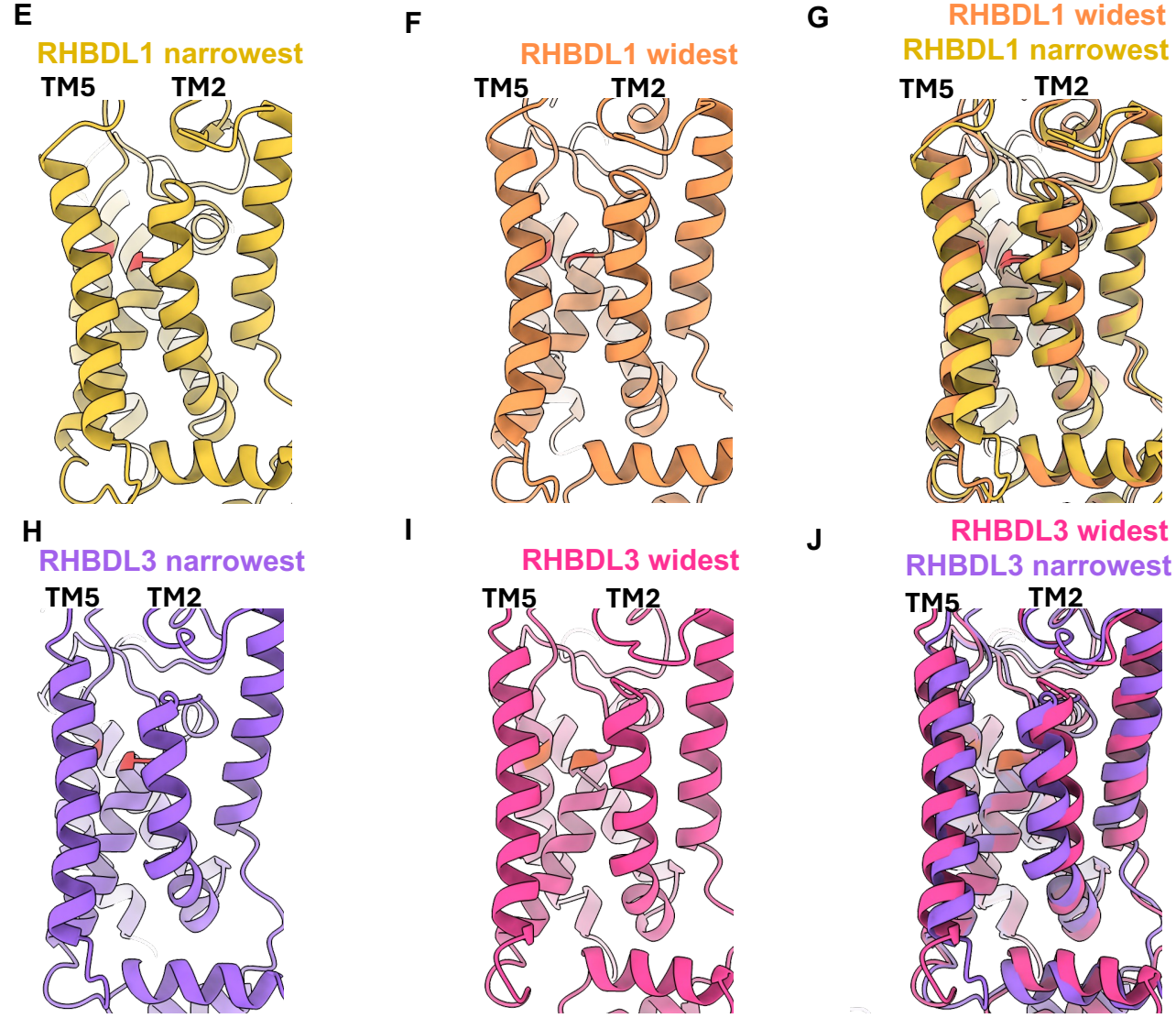
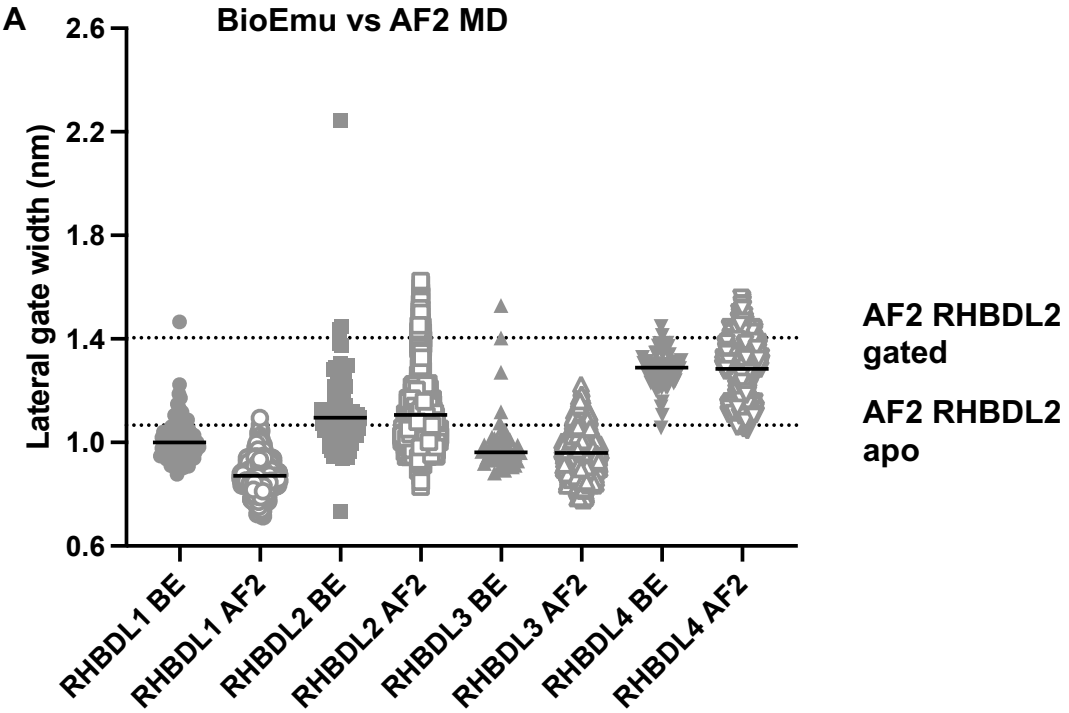

E
RHBDL1 narrowest
TM5
TM2
F
RHBDL1 widest
TM5
TM2
G
RHBDL1 widest
RHBDL1 narrowest
TM5
TM2
H
RHBDL3 narrowest
TM5
TM2
I
RHBDL3 widest
TM5
TM2
J
RHBDL3 widest
RHBDL3 narrowest
TM5
TM2

Figure 8


A
BioEmu vs AF2 MD
Lateral gate width (nm)
2.6
2.2
1.8
1.4
1.0
0.6
AF2 RHBDL2 gated
AF2 RHBDL2 apo
RHBDL1 BE
RHBDL1 AF2
RHBDL2 BE
RHBDL2 AF2
RHBDL3 BE
RHBDL3 AF2
RHBDL4 BE
RHBDL4 AF2

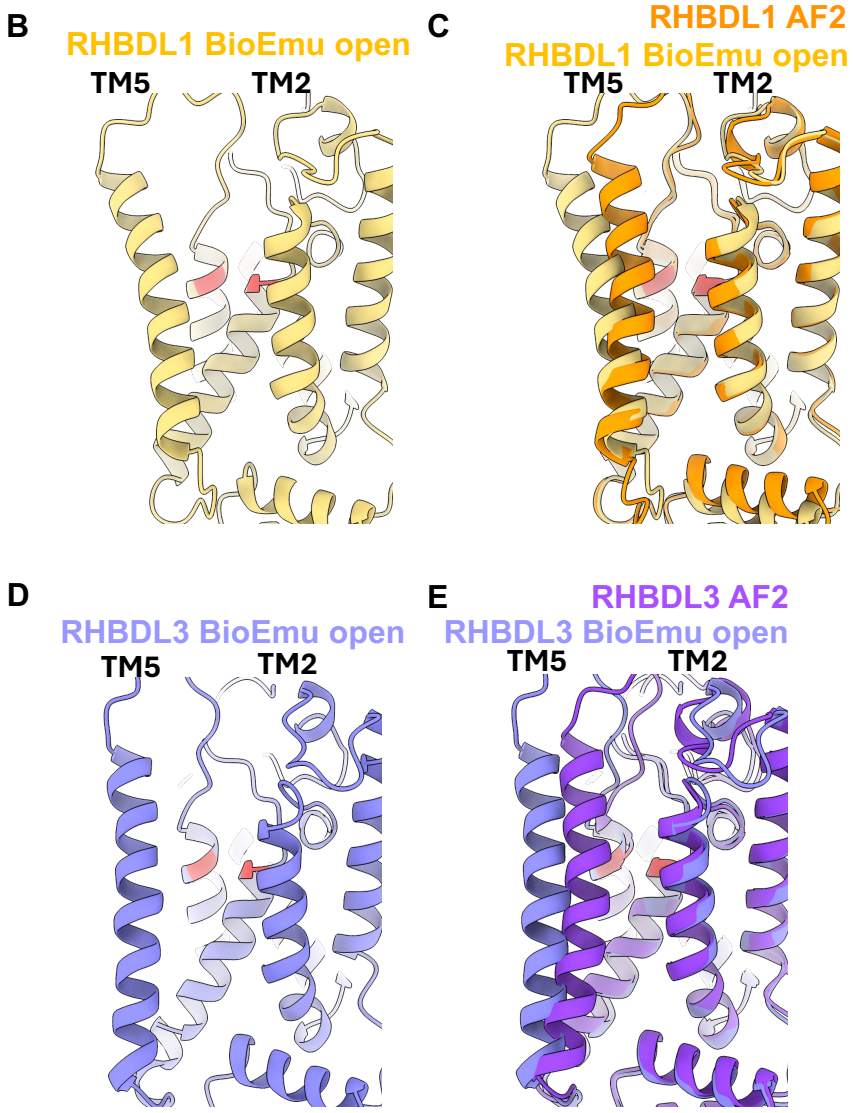
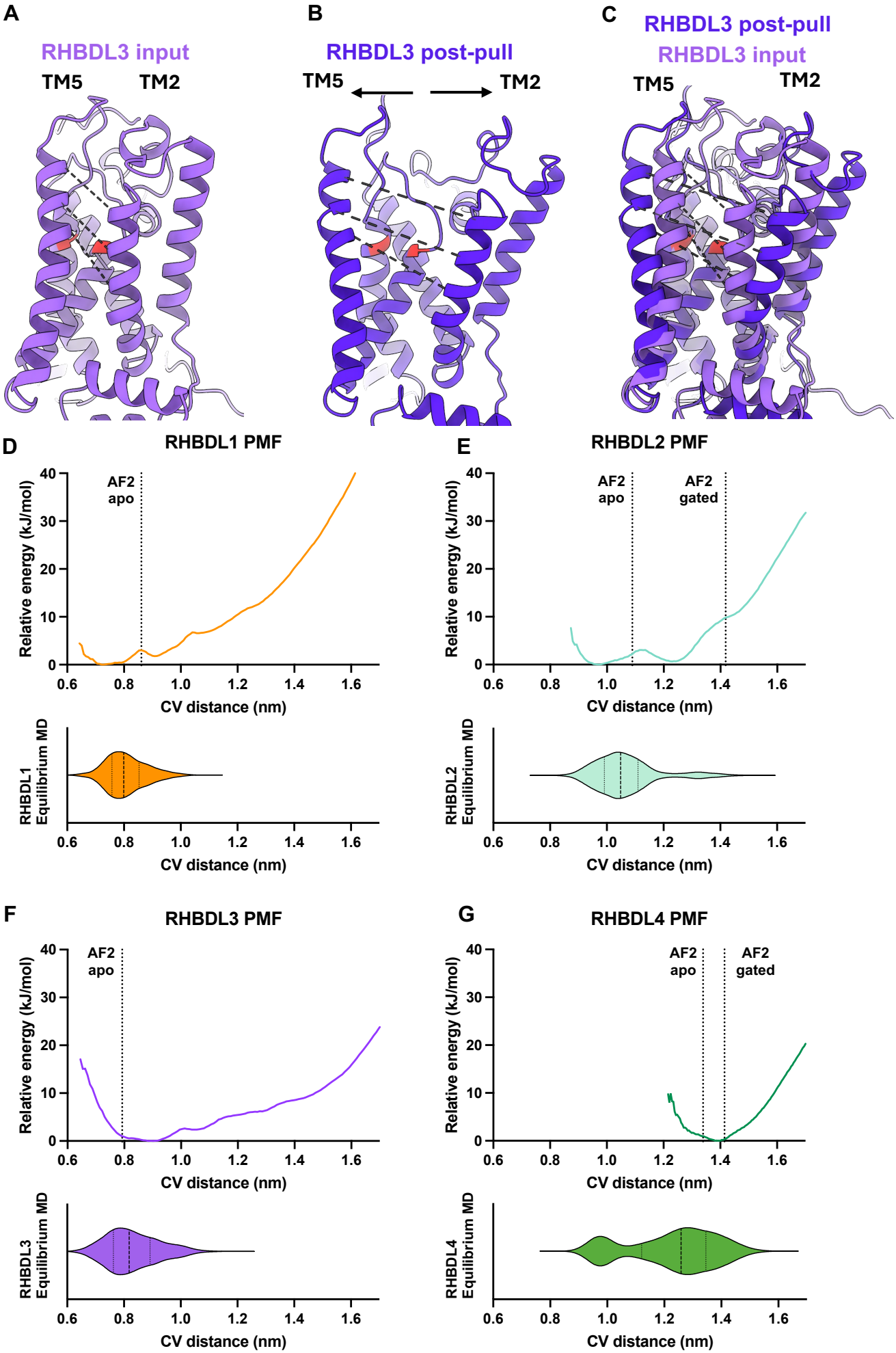

B
RHBDL1 BioEmu open
TM5
TM2
C
RHBDL1 AF2
RHBDL1 BioEmu open
TM5
TM2
D
RHBDL3 BioEmu open
TM5
TM2
E
RHBDL3 AF2
RHBDL3 BioEmu open
TM5
TM2

Figure 9


A
RHBDL3 input
TM5
TM2
B
RHBDL3 post-pull
TM5
TM2
C
RHBDL3 post-pull
RHBDL3 input
TM5
TM2
D
RHBDL1 PMF
AF2
apo
Relative energy (kJ/mol)
CV distance (nm)
RHBDL1
Equilibrium MD
E
RHBDL2 PMF
AF2
apo
AF2
gated
Relative energy (kJ/mol)
CV distance (nm)
RHBDL2
Equilibrium MD
F
RHBDL3 PMF
AF2
apo
Relative energy (kJ/mol)
CV distance (nm)
RHBDL3
Equilibrium MD
G
RHBDL4 PMF
AF2
apo
AF2
gated
Relative energy (kJ/mol)
CV distance (nm)
RHBDL4
Equilibrium MD