Faraday Discussions
Cite this: *Faraday Discuss.*, 2025, **259**, 614

DISCUSSIONS

View Article Online
View Journal | View Issue

Proteins in asymmetric membranes: general discussion

Francisco N. Barrera, Aileen Cooney, Niclas Paul Decker, Markus Deserno, Milka Doktorova, Thais A. Enoki, Alexander P. Fellows, Giacomo Fiorin, Gonen Golani, Daniel Huster, Manpreet Kaur, Ilya Levental, Kandice R. Levental, Reinhard Lipowsky, Fabio Lolicato, Erwin London, Joseph H. Lorent, Tiemei Lu, Edward Lyman, Jonathan M. Machin, Paramita Manna, Shalini Mishra, Luca Monticelli, Walter Nickel, Georg Pabst, John Seddon, Raya Sorkin and Malavika Varma

DOI: 10.1039/d5fd90021b

Daniel Huster opened the discussion of the paper by Edward Lyman: Can your simulations provide more atomistic details than X-ray or cryo-EM structures, as in those the lipids are often rather blurry? You observed that the cholesterol stayed at the receptor for the entire simulation; what are the molecular interactions between cholesterol and the protein?

Edward Lyman answered: In the simulations we know where every atom is at every instant in time, and so we have a great deal more detail than one finds in a single structure, which as you say may have low resolution for the lipid densities. I expect that the "blurry" density in experimental structures is because they are averaged over an ensemble. If the lipid is in precisely the same configuration in every member of the ensemble (particle in a cryoEM structure, unit cell in a crystal), then you get to see it in fine detail. But this, I think, is the exception rather than the rule. In simulations, "bound" lipids are often quite dynamic. It should be possible to rigorously compare density from an experimental structure to something averaged over a simulation. Phill Stansfeld at Warwick has been working along these lines, and has developed a tool called "LipidDens" for this purpose. Also, I note that in the native membrane structures the resolved lipids are often much higher resolution than in other preparations (*e.g.*, nanodiscs, LCP X-ray structures).

Regarding the side chains that interact with cholesterol, they are shown in Fig. 4D of our paper (**https://doi.org/10.1039/d4fd00210e**).

Daniel Huster asked: Are all the positively charged side chains interacting with phosphatidyl serine (PS)? What is the proximity of the carboxyl or phosphate groups to the positive charges of Arg or Lys? Lysine has different chemistry than

arginine; do you see electrostatics dominating or hydrogen bonding (with Arg)? What is the approximate residence time of a PS molecule on the receptor? Have you considered quantum mechanical MD simulations?

Edward Lyman replied: Good question. We went back and took a look at the trajectories, and we find that both the phosphate the carboxyl group can interact with both Lys and Arg side chains.

As for charge–charge interactions *vs.* hydrogen bonds, standard force fields can't really discriminate between these; nonbonded interactions are all electrostatic and van der Waals. To the extent that one observes "hydrogen bonding" interactions in fixed charge force fields, it is really just a result of the charge distributions on the donor and acceptor, and how they tend to align.

Lys *vs.* Arg is a fascinating comparison, and often underestimated in its import (by physicists, anyway). Force fields do an OK job of telling them apart, because the planar geometry of the Arg side chain is enforced, but there isn't really anything in there that captures the interactions between the π system of Arg and other interaction partners (such as cations), which almost certainly underlies its "specialness". It was Pavel Jungwirth who first pointed this out to me; he has done some really nice work in the context of cell-penetrating peptides while thinking about these things.[1]

1 C. Allolio, A. Magarkar, P. Jurkiewicz, K. Baxová, M. Javanainen, P. E. Mason, R. Šachl, M. Cebecauer, M. Hof, D. Horinek, V. Heinz, R. Rachel, C. M. Ziegler, A. Schröfel and P. Jungwirth, Arginine-rich cell-penetrating peptides induce membrane multilamellarity and subsequently enter via formation of a fusion pore, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, 11923–11928, DOI: **10.1073/pnas.1811520115**.

Erwin London queried: Do you think the cholesterol that did not dissociate during the simulation just has a slow on-rate and a slow off-rate, or really is tightly bound in the thermodynamic sense?

Edward Lyman answered: Well, the on-rate must be fairly fast on the simulation timescale, because it is already "bound" very early in the simulation. While there are just two trajectories in this paper (**https://doi.org/10.1039/d4fd00210e**), we and others have seen similar behavior many times.[1] As for the off-rate, all I can say for sure is that I have here two examples of a residence time that is at least 10 μs. I think this is a critical open question for the field – to develop rigorous methods to obtain the thermodynamics of lipid–protein interactions. There are many examples of potentials of mean force (PMFs) for lipid–protein interactions in the literature, but in our hands most of these methods do not produce reproducible results (we have a big stack of unpublished results using metadynamics, umbrella sampling, *etc.*). Grace Brannigan and Jerome Henin have been working hard on this though, and have published an alchemical free-energy method that looks promising. It is complex, but they have developed a NAMD tutorial to assist people who want to try it out: **https://livecomsjournal.org/index.php/livecoms/article/view/v5i1e2067**.

1 E. Rouviere, C. Arnarez, L. Yang and E. Lyman, Identification of Two New Cholesterol Interaction Sites on the A2A Adenosine Receptor, *Biophys. J.*, 2017, **113**, 2415–2424, DOI: **10.1016/j.bpj.2017.09.027**.

Erwin London asked: What is the present opinion of the field concerning the role of CRAC and CARC motifs in the binding of cholesterol? Are there many examples in which it is involved in cholesterol binding and specific interactions with these residues can be identified?

Edward Lyman answered: I'd say that the opinion of the field is mixed. The issue is that the CARC/CRAC "motif" is extremely promiscuous – scan the sequence of a membrane protein and you will find them. Then look for places where cholesterol interacts with the surface of a membrane protein in a simulation and you will find some (and the definition of 'interacts' is ambiguous, and so one can adjust the definition until the desired result is found).

But do they "predict" cholesterol binding sites? Daniel Isom at Miami reviewed structurally resolved cholesterols and concluded "no".[1] Rosenhouse-Dantsker came to the same conclusion based on a review of ion-channel structures.[2] Irena Levitan has a nice review of these ideas in the discussion of a paper in ref. 3.

1 G. J. Taghon, J. B. Rowe, N. J. Kapolka and D. G. Isom, Predictable cholesterol binding sites in GPCRs lack consensus motifs, *Structure*, 2021, **29**, 499–506, DOI: **10.1016/j.str.2021.01.004**.
2 A. Rosenhouse-Dantsker, Chapter Seven - Insights Into the Molecular Requirements for Cholesterol Binding to Ion Channels, in *Sterol Regulation of Ion Channels*, ed. I. Levitan, Current Topics in Membranes, Academic Press, 2017, ch. 7, vol. 80, pp. 187–208, DOI: **10.1016/bs.ctm.2017.05.003**.
3 N. Barbera, M. A. A. Ayee, B. S. Akpa and I. Levitan, Molecular Dynamics Simulations of Kir2.2 Interactions with an Ensemble of Cholesterol Molecules, *Biophys. J.*, 2018, **115**, 1264–1280, DOI: **10.1016/j.bpj.2018.07.041**.

Thais A. Enoki communicated: What are the flip-flop rates of cholesterol in the asymmetric and the symmetric membrane models? Do these rates change in the presence of A_{2A}?

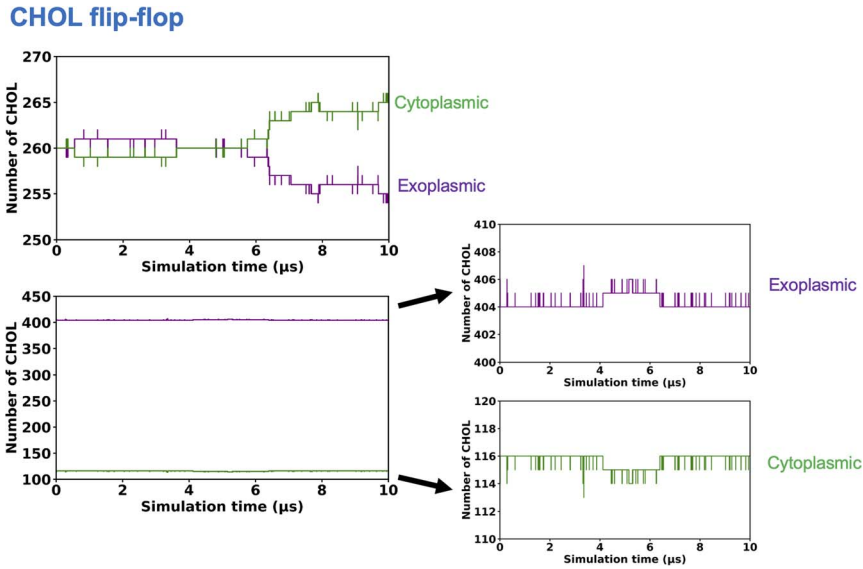

Fig. 1 Cholesterol flip-flop events in the symmetrized membrane model (upper panel) and the asymmetric model (lower panels).

Edward Lyman communicated in reply: At steady state I don't expect that the protein will modify the flip-flop rates very much, although I don't think that we have measured the flip-flop rate in the symmetrized model without protein. In the asymmetric mode, the flip-flop rate is probably too slow to measure with an unbiased simulation (again after the cholesterol distribution has equilibrated). Some data are shown in Fig. 1 in this discussion. Note that in the symmetrized model there is initially a good amount of directional transport, as the membrane equilibrates to the protein, following its insertion.

Gonen Golani remarked: The interaction of cholesterol is with relatively few amino acids. Is it possible to estimate the strength of the interaction with cholesterol based on simulations?

Edward Lyman replied: It is possible, but in our experience it is hard to do so reproducibly and with good statistical precision. One can run "very" long simulations and directly obtain the free energy for the cholesterol/protein interaction by straightforward counting: define the bound state, and accumulate statistics for on/off events. I have only seen one example with an all-atom simulation in work from Janice Roberston and Jose Faraldo Gomez, which will appear in print soon. I don't recall the exact number, but the simulation time was on the order of 50 microseconds, and this is a fairly nonspecific, solvent-like interaction! Grace Brannigan and Jerome Hénin have been developing an alchemical free-energy protocol that looks promising.[1] Many others have published potentials of mean force based on umbrella sampling protocols or metadynamics, but based on our own experience I wouldn't put much faith in any of those results (I have a big stack of unpublished PMFs that weren't reproducible). There are two challenges I think. One is to define the "bound" state. Rarely is cholesterol in a deep cleft where it is maintains fixed interactions between particular atoms of the protein and the sterol "bound" cholesterols are often found to be very dynamic, rotating around the long axis, wiggling their tails, *etc.* If one over-stabilizes a particular subset of states then the entropy of the bound state is underestimated. The second (for umbrella sampling or metadynamics) is to define a low-dimensional collective variable that takes you from the bound state to an unbound one. This is very, very nontrivial, and the difficulty (and importance) of this step tends to be grossly underestimated in the literature in my view. However (and herein lies a major problem) you will always get a decent looking PMF out of whatever protocol you use! Maybe even with nice tight-looking error bars. *Caveat emptor*!

1 R. Salari, T. Joseph, R. Lohia, J. Hénin and G. Brannigan, A Streamlined, General Approach for Computing Ligand Binding Free Energies and Its Application to GPCR-Bound Cholesterol, *J. Chem. Theory Comput.*, 2018, **14**, 6560–6573, DOI: **10.1021/acs.jctc.8b00447**.

Luca Monticelli asked: In the simulations you presented (**https://doi.org/10.1039/d4fd00210e**), is there any asymmetric stress in the membrane? Do you think this will have an effect on protein conformation? How would we assess the extent of such an effect, and whether it is realistic or not?

Edward Lyman answered: Although we have not measured the stress state of our G protein-coupled receptor (GPCR) simulation in this paper (**https://doi.org/**

10.1039/d4fd00210e) (or any other), I think it is a sensible idea. In this case, the initial stress state must be similar to the results from the recent *Cell* paper by Doktorova *et al.*,[1] since that is the membrane model that we used. I think this might be an additional "knob" that one can turn in simulations to try and access different conformational states of the receptor. But is it physical/realistic? Well, that question awaits further work, but so far all the experimental tests (also in the *Cell* paper) suggest that it is. On the other hand, the *Biophys. J.* paper[2] from Varma and Deserno considers a range of different models, and obtains a differential stress with a similar magnitude but the opposite sign! So, clearly quite a lot of work yet to do!

1 M. Doktorova, J. L. Symons, X. Zhang, H.-Y. Wang, J. Schlegel, J. H. Lorent, F. A. Heberle, E. Sezgin, E. Lyman, K. R. Levental and I. Levental, Cell membranes sustain phospholipid imbalance via cholesterol asymmetry, *Cell*, 2025, **188**, 2586–2602, DOI: 10.1016/j.cell.2025.02.034.

2 M. Varma and M. Deserno, Distribution of cholesterol in asymmetric membranes driven by composition and differential stress, *Biophys. J.*, 2022, **121**, 4001–4018, DOI: 10.1016/j.bpj.2022.07.032.

Reinhard Lipowsky said: In your paper, you describe the setup of two lipid bilayers, corresponding to an asymmetric plasma membrane model and its symmetrized counterpart, as shown in Fig. 1 of your paper (https://doi.org/10.1039/d4fd00210e). Each of the two snapshots in Fig. 1 of your paper (https://doi.org/10.1039/d4fd00210e) represents the final configuration of a 10 microsecond all-atom simulation. However, I did not find any information in your paper (https://doi.org/10.1039/d4fd00210e) about different initial conditions for these two setups. Does this imply that the 10 microsecond simulation was performed only once for each setup? You mentioned that the cholesterol molecules do not undergo any flip-flops on the timescale of your simulations. What about lateral diffusion of the cholesterol? Did you observe lateral diffusion of the cholesterol within one leaflet? How large was the corresponding diffusion constant? As far as the free-energy barrier for the flip-flops of cholesterol is concerned, a variety of enhanced sampling methods, such as umbrella sampling or metadynamics, has been developed, by which one can estimate the free-energy barrier for molecular processes. It should be informative to apply one of the enhanced sampling methods to study the flip-flops of cholesterol within your two setups for the lipid bilayers.

Edward Lyman replied: There are two simulations in this paper (https://doi.org/10.1039/d4fd00210e), each one is 10 μs in duration. Convergence of the lipid environment around the protein was assessed by asking how many first-shell lipids have exchanged out of the first shell over the course of the simulation. This is reported in the electronic supplementary information (ESI) of the paper (https://doi.org/10.1039/d4fd00210e) in Fig. S1. With the exception of the exoplasmic leaflet in the asymmetric model, nearly all of the lipids have exchanged out of the first shell by the end of the simulation. A good additional test (currently underway) is to repeat each simulation several times from a different, random initial configuration of lipids.

Regarding the cholesterol flip-flop and lateral diffusion, I was partly mistaken here. There is some flip-flop in the symmetric model (as seen in Fig. 1 in this

discussion in response to a previous comment), but very little in the asymmetric model. Cholesterol diffuses in the cytoplasmic leaflet with a typical Ld phase value, but the outer leaflet is more like Lo, and so lipid and cholesterol diffusion is slow. We did not measure them in this simulation, but have done so for other complex lipid mixtures in ref. 1.

Regarding enhanced sampling methods for flip-flop: this is something we are currently working quite hard on. Both metadynamics and umbrella sampling (US) are dependent on a good choice of collective variable. Several groups have done such calculations (*e.g.*, Drew Bennet, Wonpil Im).[2,3] We are using our unbiased trajectories with many flip-flop events to learn a one-dimensional collective variable (1D CV) (encoded in a simple neural network (NN), following work by Tiwary[4]). We then do US on this 1D CV. It requires some clever tricks and modifications to the NN, but we are obtaining very sensible and reproducible PMFs for cholesterol flip-flop. We will submit this manuscript in the next few months.

1 K. Pinkwart, F. Schneider, M. Lukoseviciute, T. Sauka-Spengler, E. Lyman, C. Eggeling and E. Sezgin, Nanoscale dynamics of cholesterol in the cell membrane, *J. Biol. Chem.*, 2019, **294**, 12599–12609, DOI: **10.1074/jbc.RA119.009683**.
2 W. F. Drew Bennett, J. L. MacCallum, M. J. Hinner, S. J. Marrink and D. P. Tieleman, Molecular View of Cholesterol Flip-Flop and Chemical Potential in Different Membrane Environments, *J. Am. Chem. Soc.*, 2009, **131**, 12714–12720, DOI: **10.1021/ja903529f**.
3 S. Jo, H. Rui, J. B. Lim, J. B. Klauda and W. Im, Cholesterol Flip-Flop: Insights from Free Energy Simulation Studies, *J. Phys. Chem. B*, 2010, **114**, 13342–13348, DOI: **10.1021/jp108166k**.
4 D. Wang and P. Tiwary, State predictive information bottleneck, *J. Chem. Phys.*, 2021, **154**, 134111, DOI: **10.1063/5.0038198**.

Tiemei Lu asked: Could lipid–protein interactions also occur between lipids and peptides? Does the molecular weight (M_w) of the peptide need to be comparable to that of a full protein for such interactions to happen?

Are lysine (Lys) and arginine (Arg) residues the main contributors to lipid–protein interactions?

How does charge density of the protein influence these interactions?

Edward Lyman replied: Sure, all kinds of peptides interact with lipids. See, for example, the vast literature on antimicrobial peptides.

Milka Doktorova commented: Placing a membrane-spanning protein in a membrane requires removal of lipids from both leaflets. The number of lipids to remove may not be trivial, especially when the protein has an asymmetric shape and is placed in an asymmetric membrane. This, coupled with the increasing evidence that lipid abundance asymmetry exists and is functional in biological membranes, suggests that examining protein–membrane interactions as a function of varying number imbalance may be more informative than looking at a single model for understanding the interaction of the protein with an asymmetric membrane. Do you think that changes in number asymmetry can affect the functional interactions of A_{2A} with the lipids and, if so, what properties do you expect to be most affected?

Edward Lyman answered: Certainly differential stress could couple to conformational changes, just like Michael Brown[1] has shown for curvature stress

for rhodopsin. We are exploring this right now. I don't have any intuition about what properties would be affected, but the basic question is "can differential stress be exploited to drive a GPCR between different conformational states?" Also, I note that we tend to use systems a bit larger than those used by most people for these simulations. This gives us more "bulk membrane" around, helpful when you are trying to look at enrichment of specific lipids. And it mitigates a bit any stress induced during the insertion/deletion part of the system setup. But see also in Fig. 1 in this discussion that we do see about 5 cholesterol molecules translocate from the inner leaflet to the outer one following insertion into the symmetrized model.

1 A. V. Botelho, N. J. Gibson, R. L. Thurmond, Y. Wang, M. F. Brown, Conformational Energetics of Rhodopsin Modulated by Nonlamellar-Forming Lipids, *Biochemistry*, 2002, **41**, 6354–6368, DOI: **10.1021/bi011995g**.

Alexander P. Fellows said: In relation to previous questions and comments regarding the impact of the initial conditions on the simulations and the computational cost associated with running them, could you please comment on such costs in terms of the feasibility of assessing the sensitivity of the results to the starting conditions?

Edward Lyman answered: Here we used a special-purpose computer (Anton2) because it permits very long simulation times (here 10 μs). This mitigates a bit the initial state bias problem. But we don't get enough time on that machine to run a whole bunch of replicas. So, a sensible addition would be to run several additional replicas on a commodity resource. It's not so hard to get trajectories that are 500 ns to 1 μs in duration these days – running these simulations (about 350k–400k atoms) with GROMACS on a single A100 GPU we get around 30 ns per day.

Kandice R. Levental opened the discussion of the paper by Fabio Lolicato: Do you see PIP2 on the exoplasmic leaflet, and, if so, is it evenly distributed or is it puncta? If it's not in the outer leaflet, what would provide the driving force for it to return to the inner leaflet?

Fabio Lolicato and **Walter Nickel** responded: So far, such experiments have not been conducted. However, we expressed and purified a high-affinity PI(4,5)P2 sensor fused to a Halo tag. Along with our single-molecule high-resolution total internal reflection fluorescence (TIRF) microscopy setup, experiments are planned to address this question in living cells.

Ilya Levental said: I was wondering if PIP2 can ever be seen on the outside of the cell. My understanding was that it hasn't been seen yet, probably because there's such a small amount and it likely diffuses away. But if there were some way to increase the throughput of FGF2 secretion (overexpression of it or trafficking components) or enhance PIP2 detection, perhaps that could help?

Fabio Lolicato and **Walter Nickel** responded: So far, we did not attempt to visualize PIP2 at the outer leaflet and there are also no conditions known that

allow for a massive upregulation of FGF2 secretion. Nevertheless, both aspects are something we are looking into and plan to address in the future.

Shalini Mishra queried: Are there any experiments done to test the involvement of redox systems in combination with lipid asymmetry in the *in vitro* condition that can contribute to the translocation of FGF2, as mentioned in the paper (**https://doi.org/10.1039/d4fd00208c**)?

Fabio Lolicato and **Walter Nickel** replied: We are currently working on the identification of redox enzymes involved in oxidative FGF2 dimerisation at the inner plasma membrane leaflet. A hypothesis we are following is a potential involvement of plasma membrane resident NAPH oxidases along with H_2O_2 as the oxidant required for disulfide bridge formation.

Erwin London asked: Would it be possible to crosslink FGF2 with a reagent that would resist cleavage of a crosslink under reducing conditions to allow it to be co-isolated with bound PIP2?

Fabio Lolicato and **Walter Nickel** answered: We did crosslink FGF2 and identified dimers in a C95-dependent manner.[1] However, those experiments aimed at the analysis of the FGF2 dimerization interface and were not used for co-isolation experiments with PIP2. The interaction of PIP2 with FGF2 in cells and *in vitro* has been demonstrated in a number of ways, including super-resolution microscopy in cells and *in vitro* liposome binding studies, as well as functional studies of various kinds, demonstrating that a loss of interaction between FGF2 and PIP2 results in a block of FGF2 secretion from cells.

1 F. Lolicato, J. P. Steringer, R. Saleppico, D. Beyer, J. Fernandez-Sobaberas, S. Unger, S. Klein, P. Riegerová, S. Wegehingel, H.-M. Müller, X. J. Schmitt, S. Kaptan, C. Freund, M. Hof, R. Šachl, P. Chlanda, I. Vattulainen and W. Nickel, Disulfide bridge-dependent dimerization triggers FGF2 membrane translocation into the extracellular space, *eLife*, 2024, **12**, RP88579, DOI: **10.7554/eLife.88579.3**.

Reinhard Lipowsky remarked: In your talk and in Fig. 1 of your paper (**https://doi.org/10.1039/d4fd00208c**), you describe the formation of tetramers of the FGF2 protein and the binding of four to five PI(4,5)P_2 molecules, which then form a localized cluster under the FGF2 tetramer. In the caption of Fig. 1 of your paper (**https://doi.org/10.1039/d4fd00208c**), you also refer to cholesterol-enriched nanodomains with an asymmetric transbilayer distribution of PI(4,5)P_2 molecules. In the cartoon of Fig. 1 of your paper (**https://doi.org/10.1039/d4fd00208c**), such a nanodomain is depicted by the blueish bilayer patch, which has a large lateral size compared to the size of the FGF2 tetramer. What are your estimates for the lateral size of the FGF2 tetramer and of the nanodomain? What is the experimental evidence for the relatively large nanodomain as depicted in Fig. 1 of your paper (**https://doi.org/10.1039/d4fd00208c**)? Do you think such a nanodomain will typically accommodate a single FGF2 tetramer or several such tetramers?

Fabio Lolicato and **Walter Nickel** replied: At this point, the size of the proposed nanodomains housing the FGF2 translocation machinery is unknown. It is also

unknown whether only a single or multiple FGF2 oligomers along with the Na,K-ATPase, Tec kinase and GPC1 are present in one such nanodomain. Fig. 1 in our paper (**https://doi.org/10.1039/d4fd00208c**) is therefore just a schematic illustration of data pointing to a liquid-ordered nanodomain as the organisational principle of the FGF2 membrane translocation machinery. In studies from other groups (Hell, Jahn; MPI Göttingen, Germany),[1] PI(4,5)P2-containing nanodomains have been analysed by super-resolution light microscopy and reported to have diameters in the range of 80 nm. Maybe this is a size range that could also apply for the PI(4,5)P_2-containing nanodomains housing the FGF2 membrane translocation machinery.

1 G. van den Bogaart, K. Meyenberg, H. J. Risselada, H. Amin, K. I. Willig, B. E. Hubrich, M. Dier, S. W. Hell, H. Grubmüller, U. Diederichsen and R. Jahn, Membrane protein sequestering by ionic protein–lipid interactions, *Nature*, 2011, **479**, 552–555, DOI: **10.1038/nature10545**.

Raya Sorkin commented: You describe three steps: dimerization, oligomerization, and pore formation. Which of these steps is affected in the synthetic *vs.* the cell system? Which one is the rate-limiting step?

Fabio Lolicato and **Walter Nickel** responded: All three of these steps are critical in both the *in vitro* and the cell-based experiments. Mutants of FGF2 (*e.g.* C95A) cannot dimerise, oligomerise or form pores. They are also not secreted from cells. The rate-limiting step is likely to be PI(4,5)P_2-dependent oxidative FGF2 dimerisation at the inner plasma membrane leaflet.

Paramita Manna addressed all: My question concerns the peptide–lipid membrane interaction and specifically the conformations of a peptide upon association with the lipid membrane. It is well established that membrane permeability is influenced by the biophysical properties of the membranes, such as rigidity, with fluid-phase membranes generally exhibiting increased permeability in the presence of membrane-active or cell-penetrating peptides. In these cases, the peptide remains embedded in the lipid bilayer. However, in cases where permeability remains low in the presence of a peptide, such as gel-phase lipids, how can one demonstrate that the peptide is embedded within the membrane bilayer rather than extending outward into the aqueous phase to engage with a distal surface? What parameters should I look into? Is it possible to establish any correlation between the peptide's spatial orientation on or within the membrane and the membrane's fluidity?

Georg Pabst replied: The positioning of membrane-active peptides relative to the lipid bilayer is influenced by the physicochemical properties of both the peptides and the lipid bilayer, as well as the concentrations of each. For amphipathic peptides that adopt linear alpha-helical secondary structures in the presence of a smooth membrane in the lamellar gel phase, it is likely that they will be adsorbed onto the membrane surface. However, this behavior can change dramatically in the presence of membrane defects.

The absence of a pretransition in phosphatidylcholine (PC), phosphatidylglycerol (PG), or sphingomyelin (SM) membranes may suggest a scenario where the

peptides are predominantly surface-bound, although this interpretation is not entirely definitive.

To investigate these dynamics, a robust experimental approach would involve a combination of molecular dynamics (MD) simulations, small-angle X-ray scattering (SAXS) or small-angle neutron scattering (SANS), along with careful hydrogen/deuterium (H/D) contrast variation. This can be achieved by using chain-deuterated lipids, headgroup-deuterated lipids, and adjusting the H_2O/D_2O ratio to accurately determine the peptide's location within the membrane.

In a subsequent analysis, it would be valuable to explore the correlation between peptide positioning and membrane fluidity. However, I anticipate that this correlation may not be straightforward. It is also crucial to consider the lipid concentration required for each specific technique, rather than solely focusing on the lipid-to-peptide ratio. For a more detailed discussion of these nuanced issues, please refer to ref. 1.

1 E. F. Semeraro, P. Pajtinka, L. Marx, I. Kabelka, R. Leber, K. Lohner, R. Vácha and G. Pabst, Magainin 2 and PGLa in bacterial membrane mimics IV: Membrane curvature and partitioning, *Biophys. J.*, 2022, **121**, 4689–4701, DOI: **10.1016/j.bpj.2022.10.018**.

Edward Lyman opened the discussion of the paper by Jonathan M. Machin: My question is about hysteresis – can you obtain reversible folding curves (*e.g.*, as Karen Fleming's group has done), so that the free energy of folding can be extracted from the measurement?

Jonathan M. Machin answered: Although highly desirable, reversible folding of outer membrane proteins (OMPs) (*i.e.*, without hysteresis) is very hard to achieve. As noted, there is a single report in the literature where this was done, with assays conducted at a pH of 3.8 for the protein OmpLA.[1] We have previously attempted to repeat this for other OMPs unsuccessfully, suggesting that these conditions are likely protein specific and will need optimising for each OMP, if indeed such conditions exist for a given OMP. OmpA has advantages for the study here (**https://doi.org/10.1039/d4fd00180j**) due to its very high stability once folded, making the formation of stable and long-lived proteoliposomes possible. However, given that (WT-)OmpA minimally unfolds, even after incubation in 8 M urea at 50 °C overnight in multiple lipid systems,[2] it seems possible that for OmpA such hysteresis may be unavoidable.

1 C. P. Moon, S. Kwon and K. G. Fleming, Overcoming Hysteresis to Attain Reversible Equilibrium Folding for Outer Membrane Phospholipase A in Phospholipid Bilayers, *J. Mol. Biol.*, 2011, **413**(2), 484–494, DOI: **10.1016/j.jmb.2011.08.041**.
2 J. M. Machin, A. C. Kalli, N. A. Ranson and S. E. Radford, Protein–lipid charge interactions control the folding of outer membrane proteins into asymmetric membranes, *Nat. Chem.*, 2023, **15**(12), 1754–1764, DOI: **10.1038/s41557-023-01319-6**.

Georg Pabst commented: OmpA folding into lipid vesicles was performed using DLPC – the shortest lipid that forms a bilayer – and DMPC at its melting transition, where significant density fluctuations occur. In both cases, the bilayers are less densely packed than in more physiologically relevant lipid systems. Have you tested whether OmpA can also fold into more tightly packed, physiologically representative membranes?

Jonathan M. Machin responded: OmpA will intrinsically fold into a limited range of membranes. It will successfully (but slowly) fold to ~80% completion into DMPC membranes in the fluid phase (30 °C), but poorly into longer saturated acyl-chain lengths (*e.g.* DPPC was ~30% folded at completion).[1,2] We chose DLPC for this work because the timescales are more experimentally accessible (15 hours), while the hydrophobic width of the bilayer is similar to the thin native outer membrane. Unsaturated longer lipid acyl chains will support folding (for example, POPC and DOPC) to varying extents.[1,3] Note that *in vitro* folding into liposomes catalysed by reconstituted BAM (the native foldase) will support folding into a much wider variety of membranes than OmpA (or other OMPs) will intrinsically fold into.[2]

1 J. M. Machin, A. C. Kalli, N. A. Ranson and S. E. Radford, Protein–lipid charge interactions control the folding of outer membrane proteins into asymmetric membranes, *Nat. Chem.*, 2023, **15**(12), 1754–1764, DOI: **10.1038/s41557-023-01319-6**.
2 S. Hussain and H. D. Bernstein, The Bam complex catalyzes efficient insertion of bacterial outer membrane proteins into membrane vesicles of variable lipid composition, *J. Biol. Chem.*, 2018, **293**(8), 2959–2973, DOI: **10.1074/jbc.RA117.000349**.
3 J. H. Kleinschmidt and L. K. Tamm, Folding Intermediates of a β-Barrel Membrane Protein. Kinetic Evidence for a Multi-Step Membrane Insertion Mechanism, *Biochemistry*, 1996, **35**(40), 12993–13000, DOI: **10.1021/bi961478b**.

Erwin London remarked: In terms of the positive-inside rule, would you predict the tendency of excess anionic lipid on the inner leaflet to induce positive inside orientation to be mostly cancelled by the excess of cationic protein residues near the surface of the membrane? If so, is there a good way to test this?

Jonathan M. Machin replied: Electrostatics at the membrane surface are determined by protein and lipid features, as well as the broader solvent/context in which the membrane resides – and all these features will combine to govern protein orientation *via* the positive-inside rule. In addition, protein crowding will significantly affect the electrostatic context. How significant an effect the titration of anionic lipids with positive proteins would have on orientation determination likely depends on whether the orientation determination is driven by a global net electrostatic environment (in which case, one would expect an effect), or local effects where positive charges on the protein bind and are anchored by one (or a few) negatively charged lipids (in which case, given the excess of lipids, providing there is adequate diffusion in the membrane there would likely be small effects). Note also work from the Hegde group which suggests that, when folding, many α-helical proteins partition into a partially-occluded patch of membrane, which would help shield any positive-inside electrostatic interactions from the larger membrane context, as well as providing lipids for interactions to form with.[1,2]

While the work presented here is on OMPs (which have a positive-outside rule and so the findings here may or may not be transferable to α-helical transmembrane proteins), it suggests that the charge balance is important not just for stability but also for the rate and efficiency of the folding process – both of which may be important for determining orientation. To test these hypotheses on α-helical proteins *in vitro* requires good refoldable model systems, inherently much more difficult for α-helical proteins that tend to be much less soluble and much less readily refolded into bilayers than their OMP counterparts. That said, it may be possible to test for loss of orientation control as increasing amounts of protein

are incorporated into membranes, and tested over a range of membranes with differing anionic lipid concentrations.

1 P. J. Chitwood and R. S. Hegde, An intramembrane chaperone complex facilitates membrane protein biogenesis, *Nature*, 2020, **584**(7822), 630–634, DOI: 10.1038/s41586-020-2624-y.
2 L. Smalinskaitė, M. K. Kim, A. J. O. Lewis, R. J. Keenan and R. S. Hegde, Mechanism of an intramembrane chaperone for multipass membrane proteins, *Nature*, 2022, **611**(7934), 161–166, DOI: 10.1038/s41586-022-05336-2.

Francisco N. Barrera asked: After the insertion of the porin with the negatively charged loops, what is the transmembrane potential change?

Membrane potential probes, like di-8-ANEPPS and similar molecules, could maybe provide this kind of information.

Jonathan M. Machin responded: We are confident in the directionality of the dipole, but not its precise value, which we did not measure. A comparative metric of per-lipid charge-difference across the membrane can be used to compare between liposomes ($\pm\sim$0.04 charges per lipid, for the positive and negative OmpA variant used), but clearly this is imprecise. We agree that to further this work a direct quantitation of the membrane potential would be useful, and would facilitate studying the effects of titrating the magnitude of the protein-induced membrane potential.

Ilya Levental communicated: My question is about the positive-inside rule. The general rule that positively charged residues are enriched on the cytosolic side of membrane proteins seems to hold for mammalian cells. This seems to make sense, since the cytosolic leaflet of the PM (plasma membrane) is negatively charged. But actually, it's quite strange, as the topology of these proteins is originally established in the ER (endoplasmic reticulum) during their biogenesis, which does not seem to be charge-asymmetric. So how did this "rule" evolve?

Jonathan M. Machin communicated in reply: This is an interesting question. A positive-inside rule for transmembrane helix insertion appears to be broadly conserved across SEC-inserted proteins, although it is somewhat relaxed for proteins inserted by the YidC family.[1,2] From an evolutionary perspective, the complex membrane architectures found in eukaryotes (like the ER) were not present when the rule first developed. It is of course hard to interpret what such ancient membranes may have looked like, including their lipid composition, membrane asymmetry and the state of the protein-inserting complexes SEC/YidC families, but given broad conservation of a negatively charged cytosolic leaflet of plasma membranes. It thus seems plausible that this charge asymmetric membrane could have been the original context for such a rule to evolve, which has then been retained. I also wonder whether a strictly charge-asymmetric membrane is actually required for the rule to develop? It is possible that simply the interaction of positive charges on the protein with negatively charged lipids may be sufficient to anchor one end of the helix and prevent it traversing the membrane, regardless of the overall membrane asymmetry. This would be particularly true in light of work from the Hegde lab

suggesting many transmembrane proteins partition through the membrane at least partially exposed to the bilayer,[3,4] in which case the unfavorability of positive charges crossing the membrane and their relatively high pK_a (making deprotonation challenging) would also contribute to holding positive charges on the inside of the membrane.

1 K. Xie and R. E. Dalbey, Inserting proteins into the bacterial cytoplasmic membrane using the Sec and YidC translocases, *Nat. Rev. Microbiol.*, 2008, **6**, 234–244, DOI: **10.1038/nrmicro3595**.
2 A. N. Gray, J. M. Henderson-Frost, D. Boyd, S. Shirafi, H. Niki and M. B. Goldberg, Unbalanced Charge Distribution as a Determinant for Dependence of a Subset of *Escherichia coli* Membrane Proteins on the Membrane Insertase YidC, *mBio*, 2011, **2**, e00238-11, DOI: **10.1128/mBio.00238-11**.
3 P. J. Chitwood and R. S. Hegde, An intramembrane chaperone complex facilitates membrane protein biogenesis, *Nature*, 2020, **584**(7822), 630–634, DOI: **10.1038/s41586-020-2624-y**.
4 L. Smalinskaitė, M. K. Kim, A. J. O. Lewis, R. J. Keenan and R. S. Hegde, Mechanism of an intramembrane chaperone for multipass membrane proteins, *Nature*, 2022, **611**(7934), 161–166, DOI: **10.1038/s41586-022-05336-2**.

Manpreet Kaur asked: How does the presence of other lipids modulate the protein folding? Given that *E. coli* has large amounts of phosphoethanolamine (PE), I wonder would the protein folding be better in tensed or more fluidic membrane?

Jonathan M. Machin responded: Folding is highly sensitive to the physiochemical properties of the lipids used, including their headgroups. While small unilamellar vesicles (SUVs) containing large amounts of PE lipid cannot be generated due to PE's negative curvature, titrating in small amounts of PE symmetrically slow down the folding.[1] We have shown that PE asymmetrically distributed such that it is only on the outer leaflet (*i.e.*, the equivalent of the inner leaflet in the native membrane) will accelerate folding relative to the symmetric case, and also that changing ratios of DMPC and DMPG lipid symmetrically or asymmetrically has a large impact on folding kinetics and stability.[2] OMPs tend not to intrinsically fold into gel-phase membranes, but do in the fluid state. In contrast, when OMPs are *in vitro*-folded catalysed by their native foldase, the BAM complex, they will fold into both gel-state and fluid membranes. Indeed, the native membrane is highly rigid due to protein and lipopolysaccharide non-covalent crosslinking, and increasing the membrane tension with high concentrations of extracellular sucrose accelerates folding.[3] Thus it appears that membrane tension may play an important role in facilitating folding *in vivo*, although this remains to be explored.

1 D. Gessmann, Y. H. Chung, E. J. Danoff, A. M. Plummer, C. W. Sandlin, N. R. Zaccai and K. G. Fleming, Outer membrane β-barrel protein folding is physically controlled by periplasmic lipid head groups and BamA, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**(16), 5878–5883, DOI: **10.1073/pnas.1322473111**.
2 J. M. Machin, A. C. Kalli, N. A. Ranson and S. E. Radford, Protein–lipid charge interactions control the folding of outer membrane proteins into asymmetric membranes, *Nat. Chem.*, 2023, **15**(12), 1754–1764, DOI: **10.1038/s41557-023-01319-6**.
3 M. T. Doyle, J. R. Jimah, T. Dowdy, S. I. Ohlemacher, M. Larion, J. E. Hinshaw and H. D. Bernstein, Cryo-EM structures reveal multiple stages of bacterial outer membrane protein folding, *Cell*, 2022, **185**(7), 1143–1156, DOI: **10.1016/j.cell.2022.02.016**.

John Seddon remarked: Have you tried reducing the phospholipid chain length even further, to say C10, to see if this speeds up the folding of OmpA?

Jonathan M. Machin replied: The intrinsic folding (*i.e.*, uncatalysed folding into membranes) of OMPs is known to be modulated by acyl-chain length. With saturated lipids, OMPs will fold more slowly as the acyl chains are made longer (at the same temperature, with all membranes in the fluid phase), due to the increased barrier to partitioning the hydrophilic loops of the protein across the membrane, required to adopt a stable transmembrane fold. DL lipids were chosen for this work because folding into DL lipids is faster than into DM lipids, making experiments more accessible, but it retains a reasonable approximation of the hydrophobic thickness of OmpA's native membrane. Previous work has characterised the chain-length dependence on intrinsic folding, and shorter lipids like C11 and C10 do indeed speed up the folding.[1,2]

1 B. Schiffrin, A. N. Calabrese, A. J. Higgins, J. R. Humes, A. E. Ashcroft, A. C. Kalli, D. J. Brockwell and S. E. Radford, Effects of Periplasmic Chaperones and Membrane Thickness on BamA-Catalyzed Outer-Membrane Protein Folding, *J. Mol. Biol.*, 2017, **429**(23), 3776–3792, DOI: **10.1016/j.jmb.2017.09.008**.
2 N. K. Burgess, T. P. Dao, A. M. Stanley and K. G. Fleming, β-Barrel Proteins That Reside in the *Escherichia coli* Outer Membrane *in Vivo* Demonstrate Varied Folding Behavior *in Vitro*, *J. Biol. Chem.*, 2008, **283**(39), 26748–26758, DOI: **10.1074/jbc.M802754200**.

Tiemei Lu remarked: You provided a great example showing that OmpA can induce lipid membrane phase separation. What other types of proteins can also induce phase separation? What are the key factors involved? Additionally, what changes or advancements occur in the lipid membrane after phase separation?

Jonathan M. Machin replied: In the outer membrane OmpA is not thought to induce lipid membrane phase separation alone. Rather it sits within a membrane that is at least partially phase separated into lipid- and protein-rich patches. OmpA has been shown to be important for maintenance of the OM organisation[1] so it likely does support this phase separation, but the precise mechanisms of its formation and maintenance remain unclear. Many factors in membranes (and their environments) can contribute to lipid and/or protein phase separation within the membrane. Key factors include, but are not limited to, preferential protein–protein or lipid–lipid interactions, membrane-external organising features (*e.g.* cytoskeleton, signalling molecules). Within bacteria, one major example is the chemosensing protein arrays (*e.g.* CheA) that form in the cytoplasmic membrane, driven by protein–protein interactions and lipid-exclusion.[2] In addition, liquid–liquid phase separation in the cell can drive membrane phase separation, and *vice versa*.[3]

Within the context of the outer membrane, one of the consequences of the phase separation is that the individual components of the membrane are rarely free to diffuse – indeed, diffusion rate is on the same order of magnitude as cell growth.[4] Other changes in membranes due to phase separation may include changes to the lipid order and dynamics in the local environment, and changes to the lipid–lipid/protein–lipid/protein–protein contacts due to altered accessible contacts (which may in turn alter the biological function).

1 G. Benn, C. Borrelli, D. Prakaash, A. N. T. Johnson, V. A. Fideli, T. Starr, D. Fitzmaurice, A. N. Combs, M. Wühr, E. R. Rojas, S. Khalid, B. W. Hoogenboom and T. J. Silhavy, OmpA controls order in the outer membrane and shares the mechanical load, *Proc. Natl. Acad. Sci. U. S. A.*, 2024, **121**(50), e2416426121, DOI: **10.1073/pnas.2416426121**.
2 G. L. Hazelbauer, J. J. Falke and J. S. Parkinson, Bacterial chemoreceptors: high-performance signaling in networked arrays, *Trends Biochem. Sci.*, 2007, **33**(1), 9–19, DOI: **10.1016/j.tibs.2007.09.014**.
3 J. A. Ditlev, Membrane-associated phase separation: organization and function emerge from a two-dimensional milieu, *J. Mol. Cell Biol.*, 2021, **13**(4), 319–324, DOI: **10.1093/jmcb/mjab010**.
4 J. E. Horne, D. J. Brockwell and S. E. Radford, Role of the lipid bilayer in outer membrane protein folding in Gram-negative bacteria, *J. Biol. Chem.*, 2020, **295**(30), 10340–10367, DOI: **10.1074/jbc.REV120.011473**.

Tiemei Lu asked: Why didn't you choose GUVs (giant unilamellar vesicles, μm-scale instead of nanometer) as the membrane model system? Is the size of the GUVs a problem?

Jonathan M. Machin answered: OmpA-WT will intrinsically fold into GUVs, although folding into large unilamellar vesicles (LUVs) is much better characterised in the literature. However, it is easier to make large quantities of size-controlled LUVs. Having minimal size-variation is important for ensuring that the dipoles are similar between different liposomes in the same ensemble (*i.e.*, it means there are fewer sources of variation in the downstream folding kinetic experiments).

Aileen Cooney asked: Does changing the charge also affect the conformation of protein in a way that would impact the results?

Jonathan M. Machin answered: The gross structure of the protein does not change – all the OmpA variants still form transmembrane β-barrels with substantial β-sheet content. However, the organisation of the non-membrane (natively extracellular) loops will alter in conformation due to the altered charge density and electrostatic interactions. For the experiments described in our paper (**https://doi.org/10.1039/d4fd00180j**), where the OmpA protein was used simply as a scaffold to generate the charge dipoles, these conformational changes should not impact the results – the conferred charge dipole will be very similar, regardless of the loop orientation.

Georg Pabst asked: Regarding Fig. 3g and h in your paper (**https://doi.org/10.1039/d4fd00180j**), can you rule out an influence of variations in OmpA copy number between the different samples on the observed melting behavior? What was the extent of sample-to-sample variation?

Jonathan M. Machin responded: Controlling for OmpA copy number is an important consideration, not just for the changes in observed melting temperature, but also to ensure that the later folding experiments are controlled with a similar amount of protein in the pre-formed proteoliposomes across all samples. The experimental procedure was set up to ensure that the final, purified proteoliposomes with the different OmpA-variants were approximately matched in protein concentration. There are differences in the folding efficiencies of the different OmpA variants, as shown in Fig. 3c of our paper (**https://doi.org/10.1039/**

d4fd00180j), ranging from ~50% (OmpA-Pos) to ~90% (OmpA-WT) folded at equilibrium. However, by changing the starting lipid : protein ratio (LPR) of each sample to match the folding efficiency (*i.e.*, 160 : 1 for OmpA-Pos, 290 : 1 for OmpA-WT), it is possible to create proteoliposomes with similar amounts of folded OmpA variant, but differing amounts of unfolded OmpA still in solution (demonstrated in Fig. 3d of our paper [https://doi.org/10.1039/d4fd00180j] – the intensity of the folded bands approximately match, but the unfolded bands are different between variants). This is effective because the final folding efficiency of OmpA is minimally dependent on the LPR (at least down to 160 : 1). Next, trypsin digest and liposome pelleting by ultracentrifugation respectively degrade and remove OmpA's C-terminal domain, as well as the unfolded OmpA remaining in each sample. This leaves proteoliposomes with only the correctly folded OmpA, at approximately the same LPR (320 : 1) for each of the different membrane-embedded OmpA variants (Fig. 3e in our paper [https://doi.org/10.1039/d4fd00180j]). For example, for OmpA-Pos at an initial LPR of 160 : 1, 50% of the OmpA-Pos folded into the membrane and 50% remained unfolded, and so the LPR of only the folded protein is 2× less than the LPR of the total proteins (320 : 1). By applying this over all the OmpA-variants, adjusted for their folding efficiency, results in proteoliposome samples of approximately the same LPR for the downstream experiments. The sample-to-sample variation is indicated by the error in Fig. 3c of our paper (https://doi.org/10.1039/d4fd00180j) – typically 5–10%. All downstream folding experiments were repeated over at least three independent proteoliposome preparations.

Edward Lyman opened the discussion of the paper by Joseph H. Lorent: We've been thinking a lot lately about assembling databases of membrane proteins (for different organisms, for different membranes) from predicted protein structures, like the AlphaFold (AF) structural database. Algorithms for detecting transmembrane regions are quite good these days; structures of transmembrane domains (TMDs) are often well predicted, and curating out weird AF stuff has been done by several groups, for example, TmAfDd from Tusnady's group: https://tmalphafold.ttk.hu/search/protein/tmem. But there is a problem – how do you get the orientation correct? Plenty of algorithms exist for this too, going all the way back to early work from von Heijne.[1] Many of these algorithms have built the positive-inside rule into their prediction framework, so if you assemble a database of TMDs from some interesting corner of the tree of life and you find that they follow the positive inside rule, do they? Or are you just observing the bias built into the algorithm? To solve this chicken-and-egg problem it would be great to have a few examples of membranes that are known to follow some other rule, like the positive-outside rule in some bacterial membranes. But many of the algorithms are also trained on bacterial membrane proteins. We really need some funky membranes that the algorithms have never seen before to test whether this bias exists.

1 G. Heijne, The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology, *EMBO J.*, 1986, **5**, 3021–3027, DOI: 10.1002/j.1460-2075.1986.tb04601.x.

Joseph H. Lorent responded: For some of the transmembrane proteins that were analyzed, the annotations of type I–IV transmembrane proteins (TMPs) exist

based on experimental evidence, but many orientations of transmembrane proteins are based on prediction algorithms. It would be useful in this regard to systematically determine the topology of plasma membrane proteins in different species experimentally and verify if the algorithms catch the real orientations in the majority of TMPs.

Giacomo Fiorin remarked: In your analysis of transmembrane domain (TMD) properties, and specifically charge among them, you have treated histidine as an aromatic residue. Based on the results of this analysis, and on available experimental data regarding pH in different cellular compartments, do you expect that accounting for pH-dependent charges could improve the ability to predict the localization of peptides in different membranes?

Joseph H. Lorent replied: Since the pK_a of histidine is around 6.5 we would definitely expect a difference in charge for species that live in a low or very high pH environment, such as certain extremophiles, in different tissues such as the gastric cavity with a pH from 1–3 or even in certain sub-cellular localizations, such as acid lysosomes with a pH of 4–6. It would be interesting to integrate the environmental pH in the charge analysis.

Georg Pabst asked: Could you comment on potential pitfalls of focusing solely on transmembrane domain (TMD) properties to determine membrane architecture?

Joseph H. Lorent answered: The biggest problem is that single-pass transmembrane domain properties not only reflect membrane properties but also have certain functions such as large membrane anchors. Certain TMD multimerization motifs are also functional but can in some cases be related to membrane properties or lipid composition, such as cholesterol binding motifs, *etc.* At the current state of knowledge it is also not clear to what extent asymmetric TMD properties of certain underexplored species correspond really to asymmetric membrane properties. This is obviously true for species where we don't know anything about membrane asymmetry. Further experimental data, specifically from bacteria and intracellular organelles, would be very useful in this regard.

Markus Deserno commented: You set out to solve the very challenging problem of extracting complicated patterns or rules from typically high dimensional data sets. In the olden days we would probably scratch our heads and make inspired guesses, which we would then check against the data, but we would be missing so much of the richness and subtlety of the problem – patterns that are hard to just guess.

Today we instead have all these amazing tools from statistical data analysis and machine learning at our disposal, and so we can leave it to the computer, in some unsupervised learning exercise, to find complex patterns. But, of course, with great power there must also come great responsibility, and we need to be mindful of the pitfalls these techniques have. In particular, if the possible patterns are intricate, and the data sets are not actually all that big, there is a risk that the rules we learn are quite specific to the available data instead of being generic statements about nature.

I therefore think it is important to not merely have a few dedicated benchmark cases at hand but to run one's analysis in a way that would ensure we can spot possible overfitting. For instance, if you have 200 data sets, you could more or less randomly split them into two, use the algorithm to learn a pattern on the first dataset, but then use the learned rules to make predictions about the second "fresh" data set. And if it turns out that the learned rules make predictions about the second data set that do not really make sense, we know we learned the wrong rules. Given that it seems you are about to embark on a bigger data journey, I was curious about what plans you have to guard against such common "overfitting pitfalls".

Joseph H. Lorent replied: In this study, we tried to avoid overweighting of certain features by excluding transmembrane proteins of very similar structures. We compared the "shortened" dataset which contains around 80% of the TMDs of the original dataset and there were no differences observed in terms of average structural information. Nevertheless, the structural information in our case could still be biased by certain factors, such as the algorithm that is used by the UniProt database to determine whether a type I–IV single-pass transmembrane domain is present. This information is relevant to indicate the direction of the TMD in the membrane. However, in some cases, this information is obtained by experiment. We therefore did an analysis of the asymmetric properties of experimentally annotated orientations – when enough proteins were available. The absolute averages were sometimes different but the overall observed trends regarding differences between major clades and subcellular organelles remained. To be precise, in this work (**https://doi.org/10.1039/d4fd00199k**) we did not use supervised machine learning to predict certain patterns but only determined average properties per position of the centrally aligned transmembrane domains including their flanking residues. This approach avoids machine-learning-generated biaises to a large part (see above). However, it would be very interesting to follow the proposed approach as an alternative and "learn" average properties per species per position.

Milka Doktorova said: Linking TMD structure to membrane asymmetry and properties can open exciting opportunities for the analysis of membranes that are hard to purify and examine by other means. A key factor for this to work is having a database of TMD structures and knowing which membrane they reside in within the cell. How easy is it in general to find this data or obtain it if no prior information exists (*e.g.* when working with a new cell line or under specific conditions)? I suppose that at least for TMDs localized in the plasma membrane that should be feasible.

Joseph H. Lorent replied: There exist several databases that allow the subcellular localization of single-pass transmembrane proteins to be found. For bacteria that do not possess intracellular membrane-bound organelles, this information is obviously not relevant. For eukaryotic cells, the UniProt databank that was used in this study (**https://doi.org/10.1039/d4fd00199k**) annotates where the protein resides, which permits us to compare single-pass transmembrane proteins of the cell membrane with the Golgi and the endoplasmic reticulum. In certain cases, it is useful in this regard to exclude transmembrane proteins that

are present in multiple organelles to better focus on the differences. Since the information that we gathered is species-dependent, we ignore differences that might arise between certain tissues or organs. To obtain this information, it will be useful to browse the scientific literature for proteomics analysis of different organelles or databases such as the Membranome database (**https://membranome.org**).[1]

1 A. L. Lomize, K. A. Schnitzer, S. C. Todd, S. Cherepanov, C. Outeiral, C. M. Deane and I. D. Pogozheva, Membranome 3.0: Database of single-pass membrane proteins with AlphaFold models, *Protein Sci.*, 2022, **31**(5), e4318, DOI: **10.1002/pro.4318**.

Niclas Paul Decker asked: Have you examined homologous proteins conserved across the major domains of life to determine how their structural adaptations differ when embedded in, for example, bacterial *versus* eukaryotic membranes?

Joseph H. Lorent replied: Here, we just compared average structural differences but excluded very similar proteins from the analysis to avoid a possible bias in average properties. This means that such an analysis was not possible within this dataset – but it would be a very good idea to do this analysis to gather information on evolutionary adaptations.

Ilya Levental added: Yeah, this is definitely an interesting question.

I do think this is a potentially cool direction, though I really don't know how many membrane proteins are meaningfully conserved from bacteria to eukaryotes. My gut feeling is not very many. But I could be wrong, in which case it would be quite interesting to examine commonalities in their adaptations.

Malavika Varma commented: Your machine learning analysis involves different taxa. Are time correlations between these different organisms on an evolutionary timescale considered while grouping the organisms and running algorithms across these different groups? Is that something that can be, or needs to be, disentangled?

Edward Lyman responded: This refers to the other *Faraday Discussions* paper presented at this meeting with my name on it (**https://doi.org/10.1039/d4fd00210e**)! There is no machine learning or evolution in this one.

Joseph H. Lorent replied: Evolutionary timescales were not considered in this work (**https://doi.org/10.1039/d4fd00199k**) but could be included in a follow-up work. Another factor that would be interesting to include are environmental factors.

Edward Lyman addressed all: I think this discussion was about whether we can infer the membrane environment (*e.g.*, localization within a eukaryotic cell) from membrane protein structure. We have been trying to train several different deep learning models to do this task, and we can't seem to do better than an F_1 score of about 65% (which is similar to existing tools like WoLF PSORT). At this point I am not sure whether this is a too-little-data problem (for structure-based methods we

are limited to a thousand or so structures), or a data-curation problem (if we use sequences or AF2-predicted structures we have to rely on GO annotations to get membrane environment labels). Or, maybe TMDs are just too promiscuous regarding the membranes that will accommodate, and this prediction task is hitting some fundamental upper bound. If anyone has ideas, please send me an email!

Aileen Cooney asked: Have you also looked into or have any idea of the impact of local cooperativity between proteins, namely whether nearby protein–lipid interactions have knock-on effects that influence the explicit protein being studied?

Edward Lyman answered: In the GPCR field there is a longstanding interest in receptor dimerization as an additional level of modulation, but we have not attempted to use simulations to study this. It's very hard (maybe impossible) to do well with all-atom models.

Conflicts of interest

There are no conflicts to declare.