

**Computational studies of membrane proteins: from sequence to structure to simulation**

Phillip J. Stansfeld<sup>1\*</sup>

<sup>1</sup>Department of Biochemistry,  
University of Oxford,  
South Parks Road,  
Oxford,  
OX1 3QU,  
UK.

\*to whom correspondence should be addressed:

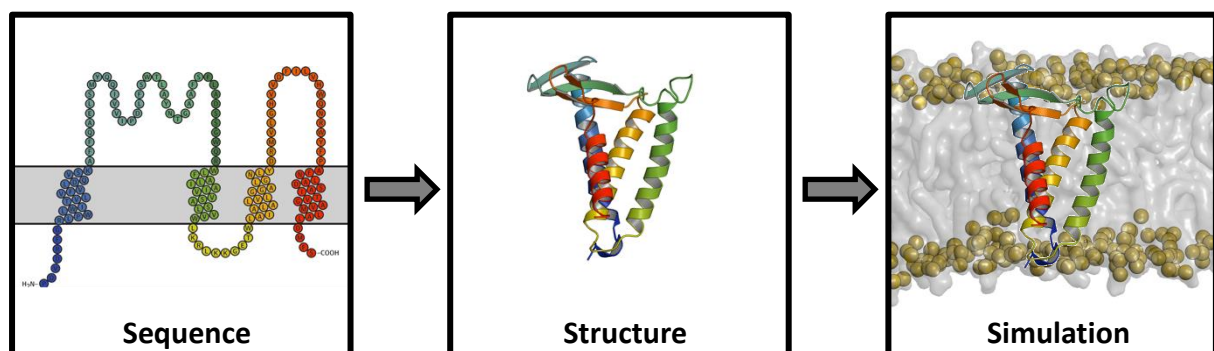
e-mail:           phillip.stansfeld@bioch.ox.ac.uk

phone:           +44-1865-613362

For submission to *Current Opinions in Structural Biology*

**Highlights**

- Databases of membrane protein structures
- Modelling of membrane protein structures using co-evolution information
- Identification and annotation of lipid binding sites on membrane proteins
- Molecular simulations of membrane proteins in complex and biological lipid membranes

**Graphical Abstract**

## Abstract

In this review, I discuss the recent advances in computational approaches to studying membrane protein structures, covering the latest methods for predicting a protein structure from its amino acid sequence, through to methods for assessing the structural dynamics and lipid interactions within molecular simulations of complex biological membranes. These approaches have not only benefited from advances in the computational software and architectures, but have also been assisted by a prodigious rise in the number of both the molecular sequences and experimentally determined membrane protein structures. The former, in part stimulated by metagenomics sequencing techniques, has led to an increased prediction accuracy for the computationally folded protein structures. The latter, assisted by improvements in structural biology approaches, has led to longer, larger and more complex molecular simulations of membrane proteins; many of which have greater relevance to human disease. Here I describe the methods for predicting a membrane protein structure from sequence, discuss the approaches to configure membrane protein simulations and detail the techniques used to identify and characterize specific lipid binding sites to membrane protein structures.

## Introduction

Membrane protein structural biology has firmly come-of-age at the start of the 21<sup>st</sup> century. From the turn of the millennia the number of integral membrane protein structures has risen from marginally over 100 to now roughly 3,000 experimentally determined coordinates (March 2017) [1]. In recent years, this rise has also coincided with an increase in diversity of the proteins solved, permitting distinct folds of proteins to be observed for the first time, as well as structures being captured in multiple conformational states or with distinct ligands bound. Added to this, these proteins are now more frequently mammalian in origin and/or of greater medical relevance.

The rapid and consistent rise in the number of structures has led to the development of databases and resources that catalog and annotate all determined membrane protein structures. These range from a well devised characterisation of the proteins based on function at the Membrane Proteins of Known Structure (MPStruc) database [2], to hydrophobic slab-based methods to accurately delineate the membrane spanning domains of the protein's tertiary fold, at the Orientations of Proteins in Membranes (OPM) database [3] and Protein Data Bank of Transmembrane Proteins (PDBTM) [4]. These independent sites have recently filtered through to the RCSB Protein Data Bank, with the visualisations and annotations now shared for each individual membrane protein PDB entry [5].

## Co-evolutionary modelling of membrane proteins

While the number of membrane protein structures has increased, the rate of their determination is significantly lower than the rise in the characterized number of protein sequences. This has led to the return to a method that was first considered in the late 1980s [6-8]. The foundation of this method is that the interaction of pairs of residues is crucial for the stabilisation of the folded state. If one of the residues involved in the packing arrangement changes, there will be selective pressure for the second residue to adapt and thereby regain optimal packing (Figure 1a). This leads to amino acids appearing to change in tandem to retain a protein's fold and function. By interrogating the sequence databases for homologous proteins with tools such as HHblits [9] and Jackhmmer [10], one can generate diverse alignments of thousands of protein sequences and therefore identify residues that are evolutionarily coupled with one another. Example tools for this include PSICOV [11], DCA<sub>plm</sub> [12], Gremlin [13] and CCMpred [14]. Once residue pairs have been identified they can be turned into molecular constraints

to fold protein structures within FILM3 [15], EVFold [12] and Rosetta [16]. These methods are enhanced by employing transmembrane (TM) helix prediction, e.g. TOPCONS [17] or MEMSAT [18], and secondary structure, e.g. PSIPRED [19], algorithms to add further constraints to the folding methodology. In addition to folding of proteins these contacts are also incredibly successful at predicting residue pairs at the interface of macromolecular complexes [20-22]. In this instance the contacts may be used as constraints in protein-protein docking methods, such as HADDOCK [23], and thereby predict inter-subunit interaction sites.

In 2015, the co-evolutionary methods were employed to provide a 'large scale' prospective determination of a set of over 50 membrane and soluble protein structures [16]. Since the release of this dataset, a number of the predicted structures have been experimentally solved. These X-ray structures illustrate the fidelity of the methodology and highlight the remarkable similarity between the modelled and experimental structures. The X-ray structures include two proteins involved in the lipoprotein biogenesis pathway: the lipoprotein signal peptidase II (LspA) [24] and prolipoprotein diacylglycerol transferase (Lgt) [25]. Other accurate predictions include a dimeric fluoride ion transporter (FluC) [26], a multi-subunit cytochrome bd oxidase [27], and a DMT superfamily transporter RarD, which shares homology with the X-ray structure of YddG [28] (Figure 1b-d).

The same group expanded their methodology earlier this year to include metagenomics sequences [29]. The addition of protein sequences derived from environmental, uncultured microbial samples significantly enriches the number and diversity of the protein sequences and permitted 614 protein families without known three-dimensional structures to be modelled. A third (206) of these prospectively modelled structures are of membrane proteins. As this class of proteins remains notoriously difficult to structurally solve, these methodologies are therefore exceptionally useful for predicting the tertiary and quaternary structures in lieu of their experimental determination. Furthermore, these protocols are considered optimal for membrane proteins as the fold-space within a membrane is relatively limited, i.e. the vast majority of integral membrane proteins are made up of TM helices that cross perpendicular to the membrane, and therefore shield the peptide backbone from the non-aqueous interior. Nevertheless, there are recent examples where the TM helices lie almost parallel to the membrane, e.g. the recently described toppling mechanism for ECF transporters [30] and the position of the  $\alpha$ -subunit for proton transfer in the ATP synthase [31]. The covariance and folding methods are also surprisingly successful at predicting other membrane protein features, e.g. half-helices, interfacial helices and re-entrant loops, while there is a degree of success in predicting additional complexity, e.g.  $3_{10}$ -helices,  $\pi$ -bulges and helical kinks.

### **Co-evolution and conformational change**

One major challenge to the use of covariance analysis is that the modelling methods generally use all the residue contacts as constraints to generate a folded model. In some cases, this means the protein is modelled into a non-native state by using residue constraints that are important for two distinct conformational states. This is best reflected upon by considering the alternating-access conformational landscape of transporters, as discussed in greater detail in other reviews of this series [32-34]. Transporter proteins are one of the most abundant membrane proteins, required to alternate their solvent accessibility from one side of the membrane to the other, thereby facilitating molecular translocation across the bilayer (Figure 2). Secondary transporters accomplish this by the evolutionary design of protein structures that have an almost identical preference for the alternative conformational states. This is achieved by using the same structural repeat for both conformations of the transporter, as described in the accompanying review of the Na<sup>+</sup>-citrate transporter (CitS) [34]. As

a result, if one has solved the inward-facing state of the transporter, the instructions for computationally modelling the outward-facing state is also detailed within the same set of coordinates. Therefore an inverted state model may be generated by mapping the sequences of the individual topological repeats to one another and then switching the residues by conventional homology modelling [35]. The inverted-state model may then be assessed in light of the co-evolution contacts, with state-dependent residue-pairs then identified. In a recent example the conformational changes associated with transport were captured by molecular dynamics simulations of the semiSWEET transporter, a rudimentary dimeric transporter found in *E. coli* [36]. In this study, they identified the motions associated with the transitions from inward-facing, to occluded, to outward-facing state, highlighting that the sugar cargo could be transported without major structural rearrangement.

### **Adding the grease to membrane protein structures**

As highlighted in the accompanying review by Pliotas and Naismith [37], the native lipid bilayer is integral to a membrane protein's stability and function. However, only ca. 100 membrane protein structures contain coordinates of lipid-like molecules, never mind full molecular details for the lipid membrane [38]. The most atomic-level 'membrane' information is for a pair of Aqp0 structures derived from electron crystallography, where the first annular shell of lipids was observed [39,40]. More recently the combined use of Cryo-EM and lipid nanodiscs yielded structures of TRPV1 channels surrounded by a belt of lipids [41]. In this instance, tightly bound lipids and analogous molecules could be modelled into the electron density, but atomic resolution details for the full membrane patch was not possible due to the averaging of the snapshots across the full dataset.

From an X-ray perspective, lipidic cubic phase (LCP) crystallization remains an important method for capturing membrane protein structures. This method has been used to solve proteins ranging from the single-pass TM helices of glycophorin [42] through to large multi-subunit G-protein receptor complexes [43]. This method frequently retains atomic resolution information for lipid binding sites, which are identified by the typical LCP lipid, monoolein. Molecular simulations have been used to address the interactions of G-protein coupled receptors within LCP [44], whilst also characterizing the protein and lipid driving forces behind the LCP crystallization process [45].

The recent LCP-based determination of two enzyme structures of diacylglycerol kinase (DgkA) and the previously discussed LspA, provided a means to capture monoolein binding sites to these structures. In both instances the observed monoolein coordinates were used to guide the computational modelling of substrate binding (Figure 3). Monoolein itself is a substrate for DgkA, but its native lipid in *E. coli* is diacylglycerol (DAG) and therefore the coordinates of two adjacent molecules of monoolein were amalgamated to create the bound configuration of DAG [46]. Similarly, the binding of diacylcysteine to LspA was modelled based on the monoolein crystallographic coordinates [24]. This not only permits the molecular modelling of the bound lipids but also creates coordinate sets for commencing molecular simulations to capture the molecular motions of the protein-lipid complexes. These membrane protein-lipid complexes are best simulated within a lipid membrane.

### **Molecular simulations of proteins in membranes**

There are now a range of methods for computationally inserting a membrane protein structure into a bilayer. These techniques vary from implanting the protein directly into preformed membranes, as performed using e.g. CHARMM-GUI [47] and INSANE [48], to permit self-assembly of the lipid membrane around the protein [49]. We recently developed an automated pipeline for setting-up membrane protein simulation upon their deposition and release from the PDB (Figure 4). This permits the membrane insertion of all solved protein structures and their subsequent upload to our online

resource, MemProtMD (<http://memprotmd.bioch.ox.ac.uk>) [1]. In this instance, the initial simulations are performed, for speed, using coarse-grained (CG) molecular dynamics (MD) simulations (in which groups of ca. 4 atoms are represented by a single particle or 'bead'), with the end snapshots converted back to atomistic resolution to permit more accurate determination of molecular interactions [50,51].

The primary aim of this methodology is to highlight the importance of the lipid molecules that embrace the membrane protein structure. This is especially important for computationally determining specific lipid binding sites within the annular shell and assessing protein-induced membrane deformation. While the initial MemProtMD simulations are performed with a model DPPC membrane, this can be accurately replaced with a more physiological composition of the native lipids that would ordinarily encapsulate the protein [52]. Examples of these simulations have permitted the identification of PIP<sub>2</sub> lipid binding sites on Kir channels [53], reproduced the electron crystallography binding sites of lipids in the annular shell around aquaporins [54-56], divulged lipid interactions with single-pass receptor tyrosine kinases (RTKs) [57], structurally rationalized the role for cardiolipin in the dimerization of the leucine transporter (LeuT) [58], and identified cardiolipin binding sites to respiratory chain complexes [59-61] and their role in supercomplex formation [62]. For a comprehensive review of lipid-binding simulations see [63].

### **Calculating binding energies for lipid binding and increasing the molecular complexity**

Once a binding site has been identified for a lipid to a membrane protein, either experimentally or computationally, molecular simulations may then be used to assay the lipid affinity for the site. By performing potential of mean force (PMF) calculations one can calculate a free energy of binding and assess the influence of mutations [64]. Sufficient sampling is critical for these calculations and therefore the majority of lipid-protein PMFs have so far been performed using CGMD simulations. As part of the respiratory chain study referenced above, PMFs were performed for cardiolipin to its identified binding sites on cytochrome c oxidase to reveal that 'sites I and II' had the greatest affinity. Subsequent simulations revealed that binding 'site II' was also selective for cardiolipin over other lipid species tested. Equivalent studies have been performed to assess the affinity of cardiolipin to the three binding sites on the adenine nucleotide translocase (ANT) [65], and to evaluate the free energy associated with PIP<sub>2</sub> binding to the Kir2.2 channel [64].

The identification of binding sites for crucial lipid species highlights the importance of including full molecular complexity of lipid membranes around the simulated membrane protein. As a result, CGMD has been used to simulate larger molecular systems, with multiple copies of integral membrane proteins, which attempt to accurately model the dynamics of biological membranes. Key examples are multi-component lipid mixtures that simulate plasma membranes from mammals [66-68], thylakoid membranes in cyanobacteria and plants with [69] and without [70] photosystem II and bacterial proteins within models of the outer membrane [71-73]. Similar studies have built and simulated atomic models of lamellar [74] and vesicular chromatophores [75] from purple bacteria. Vesicular models have also been used as starting points for molecular simulations of entire virion membranes from both influenza A [76] and dengue [77,78].

### **Conclusions**

This review scratches the surface of the computational methods that may be performed for studying the structural biology of membrane proteins. In the first section of the review I have focused on means to acquire accurate three-dimensional structures for a given membrane protein sequence. This methodology is expected to improve further as both the number of metagenomics sequences and solved protein-folds increase. At present this methodology has its greatest use for proteins found in prokaryotes, due to the extensive variety of these sequences. Of interest in the future will be the

application of this methodology to capture novel eukaryotic (e.g. human) membrane protein structures, given current metagenomics sequencing efforts for capturing the diversity of fungal and other simple eukaryotic sequences. The ability to visualize a protein structure prior to its experimental elucidation also provides a structure-based approach to design optimal constructs for structural determination, to provide search candidates for molecular replacement or to utilize the contact pairs as restraints in the interpretation of electron density.

The subsequent sections in this review highlight the use of experimentally-derived structures in molecular simulations to predict lipid binding sites and their affinity, and to assess dynamics within larger and more complex lipid membranes. As more membrane proteins structures emerge it will become increasingly important to assess their membrane interactions and full molecular dynamics. This will be best performed by considering the native and physiological environment of the protein, by taking into account the full molecular complexity of the lipid membrane in light of lipidomics data taken from e.g. the LIPID MAPS gateway [79].

As confidence increases in the folded structures and computational processing speeds continue to rise, there is the potential for accurate molecular simulations of a membrane protein structures in complex lipid membranes, given only a protein single-letter sequence. These approaches would facilitate membrane protein dynamics, including identification and characterisation of lipid binding sites, prior to the actual determination of an experimental structure; essentially computationally modelling a membrane protein from sequence to structure to simulation.

#### Conflict of interest

Nothing declared.

#### References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

#### Acknowledgements

I would like thank Mark Sansom for critical reading of the manuscript, the Biotechnology and Biological Sciences Research Council (BBSRC grants BB/I019855/1, BB/L002531/1, BB/P01948X/1) for funding, and the Irish Center for High End Computing (ICHEC) for the provision of computational facilities and support that contributed towards publications referenced in this review.

#### References

1. •• Stansfeld PJ, Goose JE, Caffrey M, Carpenter EP, Parker JL, Newstead S, Sansom MS: **MemProtMD: Automated Insertion of Membrane Protein Structures into Explicit Lipid Membranes.** *Structure* 2015, **23**:1350-1361.  
A pipeline and database for inserting membrane protein structures into lipid bilayers. All results may be accessed online at MemProtMD (<http://memprotmd.bioch.ox.ac.uk>)
2. White SH: **Biophysical dissection of membrane proteins.** *Nature* 2009, **459**:344-346.
3. Lomize MA, Lomize AL, Pogozheva ID, Mosberg HI: **OPM: Orientations of proteins in membranes database.** *Bioinformatics* 2006, **22**:623-625.
4. Tusnady GE, Dosztanyi Z, Simon I: **PDB\_TM: selection and membrane localization of transmembrane proteins in the protein data bank.** *Nucl. Acids Res.* . 2005, **33**:D275-D278.
5. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE: **The Protein Data Bank.** *Nucl. Acids Res.* 2000, **28**:235-242.

6. Altschuh D, Vernet T, Berti P, Moras D, Nagai K: **Coordinated amino acid changes in homologous protein families.** *Protein Eng* 1988, **2**:193-199.
7. Shindyalov IN, Kolchanov NA, Sander C: **Can three-dimensional contacts in protein structures be predicted by analysis of correlated mutations?** *Protein Eng* 1994, **7**:349-358.
8. Gobel U, Sander C, Schneider R, Valencia A: **Correlated mutations and residue contacts in proteins.** *Proteins* 1994, **18**:309-317.
9. Remmert M, Biegert A, Hauser A, Soding J: **HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment.** *Nat Methods* 2011, **9**:173-175.
10. Johnson LS, Eddy SR, Portugaly E: **Hidden Markov model speed heuristic and iterative HMM search procedure.** *BMC Bioinformatics* 2010, **11**:431.
11. Jones DT, Buchan DW, Cozzetto D, Pontil M: **PSICOV: precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments.** *Bioinformatics* 2012, **28**:184-190.
12. Hopf TA, Colwell LJ, Sheridan R, Rost B, Sander C, Marks DS: **Three-dimensional structures of membrane proteins from genomic sequencing.** *Cell* 2012, **149**:1607-1621.
13. Kamisetty H, Ovchinnikov S, Baker D: **Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era.** *Proc Natl Acad Sci U S A* 2013, **110**:15674-15679.
14. Seemayer S, Gruber M, Soding J: **CCMpred--fast and precise prediction of protein residue-residue contacts from correlated mutations.** *Bioinformatics* 2014, **30**:3128-3130.
15. Nugent T, Jones DT: **Accurate de novo structure prediction of large transmembrane protein domains using fragment-assembly and correlated mutation analysis.** *Proc Natl Acad Sci U S A* 2012, **109**:E1540-1547.
16. •• Ovchinnikov S, Kinch L, Park H, Liao Y, Pei J, Kim DE, Kamisetty H, Grishin NV, Baker D: **Large-scale determination of previously unsolved protein structures using evolutionary information.** *Elife* 2015, **4**.  
The co-evolution-based modelling of over 50 membrane protein structures, using GREMLIN co-evolution analysis and Rosetta *ab initio* modelling.
17. Bernsel A, Viklund H, Hennerdal A, Elofsson A: **TOPCONS: consensus prediction of membrane protein topology.** *Nucleic Acids Res* 2009, **37**:W465-468.
18. Nugent T, Jones DT: **Transmembrane protein topology prediction using support vector machines.** *BMC Bioinformatics* 2009, **10**:159.
19. Jones DT: **Protein secondary structure prediction based on position-specific scoring matrices.** *J Mol Biol* 1999, **292**:195-202.
20. Hopf TA, Scharfe CP, Rodrigues JP, Green AG, Kohlbacher O, Sander C, Bonvin AM, Marks DS: **Sequence co-evolution gives 3D contacts and structures of protein complexes.** *Elife* 2014, **3**.
21. Ovchinnikov S, Kamisetty H, Baker D: **Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information.** *Elife* 2014, **3**:e02030.
22. • Alcock F, Stansfeld PJ, Basit H, Habersetzer J, Baker MA, Palmer T, Wallace MI, Berks BC: **Assembling the Tat protein translocase.** *Elife* 2016, **5**.  
The use of co-evolutionary data to model the interactions between TatA, TatB and TatC.
23. Dominguez C, Boelens R, Bonvin AM: **HADDOCK: a protein-protein docking approach based on biochemical or biophysical information.** *J Am Chem Soc* 2003, **125**:1731-1737.
24. •• Vogeley L, El Arnaout T, Bailey J, Stansfeld PJ, Boland C, Caffrey M: **Structural basis of lipoprotein signal peptidase II action and inhibition by the antibiotic globomycin.** *Science* 2016, **351**:876-880.  
The crystal structure and molecular simulations of the lipoprotein signal peptidase II (LspA).
25. • Mao G, Zhao Y, Kang X, Li Z, Zhang Y, Wang X, Sun F, Sankaran K, Zhang XC: **Crystal structure of E. coli lipoprotein diacylglycerol transferase.** *Nat Commun* 2016, **7**:10198.  
The crystal structure of the lipoprotein diacylglycerol transferase (Lgt).

26. • Stockbridge RB, Kolmakova-Partensky L, Shane T, Koide A, Koide S, Miller C, Newstead S: **Crystal structures of a double-barrelled fluoride ion channel**. *Nature* 2015, **525**:548-551.  
The dimeric X-ray structure of a bacterial fluoride channel.
27. • Safarian S, Rajendran C, Muller H, Preu J, Langer JD, Ovchinnikov S, Hirose T, Kusumoto T, Sakamoto J, Michel H: **Structure of a bd oxidase indicates similar mechanisms for membrane-integrated oxygen reductases**. *Science* 2016, **352**:583-586.  
The crystal structure of the multimeric cytochrome bd oxidase, which used co-evolutionary data to optimise the model building from the electron density map.
28. • Tsuchiya H, Doki S, Takemoto M, Ikuta T, Higuchi T, Fukui K, Usuda Y, Tabuchi E, Nagatoishi S, Tsumoto K, et al.: **Structural basis for amino acid export by DMT superfamily transporter YddG**. *Nature* 2016, **534**:417-420.  
The crystal structure of the RarD homologue, YddG.
29. •• Ovchinnikov S, Park H, Varghese N, Huang PS, Pavlopoulos GA, Kim DE, Kamisetty H, Kyrpides NC, Baker D: **Protein structure determination using metagenome sequence data**. *Science* 2017, **355**:294-298.  
The use of metagenomics data to enrich the sequences used in the covariance analysis and the subsequent modelling of over 600 structures using the residue pairs as constraints.
30. Swier LJ, Guskov A, Slotboom DJ: **Structural insight in the toppling mechanism of an energy-coupling factor transporter**. *Nat Commun* 2016, **7**:11072.
31. Allegretti M, Klusch N, Mills DJ, Vonck J, Kuhlbrandt W, Davies KM: **Horizontal membrane-intrinsic alpha-helices in the stator a-subunit of an F-type ATP synthase**. *Nature* 2015, **521**:237-240.
32. Kazmier K, Claxton DP, McHaourab HS: **Alternating access mechanisms of LeuT-fold transporters: trailblazing towards the promised energy landscapes**. *Curr Opin Struct Biol* 2016, **45**:100-108.
33. Newstead S: **Recent advances in understanding proton coupled peptide transport via the POT family**. *Curr Opin Struct Biol* 2016, **45**:17-24.
34. Lolkema JS, Slotboom DJ: **Structure and elevator mechanism of the Na<sup>+</sup>-citrate transporter CitS**. *Curr Opin Struct Biol* 2016, **45**:1-9.
35. Forrest LR, Zhang YW, Jacobs MT, Gesmonde J, Xie L, Honig BH, Rudnick G: **Mechanism for alternating access in neurotransmitter transporters**. *Proc. Natl. Acad. Sci. USA* 2008, **105**:10338-10343.
36. Latorraca NR, Fastman NM, Venkatakrisnan AJ, Frommer WB, Dror RO, Feng L: **Mechanism of Substrate Translocation in an Alternating Access Transporter**. *Cell* 2017, **169**:96-107 e112.
37. Pliotas C, Naismith JH: **Spectator no more, the role of the membrane in regulating ion channel function**. *Curr Opin Struct Biol* 2016, **45**:59-66.
38. Yeagle PL: **Non-covalent binding of membrane lipids to membrane proteins**. *Biochim Biophys Acta* 2014, **1838**:1548-1559.
39. Gonen T, Cheng Y, Sliz P, Hiroaki Y, Fujiyoshi Y, Harrison SC, Walz T: **Lipid-protein interactions in double-layered two-dimensional AQP0 crystals**. *Nature* 2005, **438**:633-638.
40. Hite RK, Li Z, Walz T: **Principles of membrane protein interactions with annular lipids deduced from aquaporin-0 2D crystals**. *EMBO J* 2010, **29**:1652-1658.
41. • Gao Y, Cao E, Julius D, Cheng Y: **TRPV1 structures in nanodiscs reveal mechanisms of ligand and lipid action**. *Nature* 2016, **534**:347-351.  
The use of nanodiscs and cryo-EM to capture the structure of a membrane protein surrounded by a belt of lipid molecules.
42. Trenker R, Call ME, Call MJ: **Crystal Structure of the Glycophorin A Transmembrane Dimer in Lipidic Cubic Phase**. *J Am Chem Soc* 2015, **137**:15676-15679.
43. Rosenbaum DM, Zhang C, Lyons JA, Holl R, Aragao D, Arlow DH, Rasmussen SG, Choi HJ, Devree BT, Sunahara RK, et al.: **Structure and function of an irreversible agonist-beta(2) adrenoceptor complex**. *Nature* 2011, **469**:236-240.
44. Khelashvili G, Albornoz PB, Johnner N, Mondal S, Caffrey M, Weinstein H: **Why GPCRs behave differently in cubic and lamellar lipidic mesophases**. *J Am Chem Soc* 2012, **134**:15858-15868.



45. Johnner N, Mondal S, Morra G, Caffrey M, Weinstein H, Khelashvili G: **Protein and lipid interactions driving molecular mechanisms of in meso crystallization.** *J Am Chem Soc* 2014, **136**:3271-3284.
46. •• Li D, Stansfeld PJ, Sansom MS, Keogh A, Vogeley L, Howe N, Lyons JA, Aragao D, Fromme P, Fromme R, et al.: **Ternary structure reveals mechanism of a membrane diacylglycerol kinase.** *Nat Commun* 2015, **6**:10140.  
The crystal structures of diacylglycerol kinase with monoolein lipids and an ATP-analogue (AMP-PCP) bound, with simulations performed with diacylglycerol modelled on the monoolein coordinates.
47. Jo S, Lim JB, Klauda JB, Im W: **CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes.** *Biophys. J.* 2009, **97**:50-58.
48. Wassenaar TA, Ingolfsson HI, Bockmann RA, Tieleman DP, Marrink SJ: **Computational Lipidomics with insane: A Versatile Tool for Generating Custom Membranes for Molecular Simulations.** *J Chem Theory Comput* 2015, **11**:2144-2155.
49. Bond PJ, Sansom MSP: **Insertion and assembly of membrane proteins via simulation.** *J. Am. Chem. Soc.* 2006, **128**:2697-2704.
50. Stansfeld PJ, Sansom MSP: **From coarse-grained to atomistic: a serial multi-scale approach to membrane protein simulations.** *J. Chem. Theor. Comp.* 2011, **7**:1157–1166.
51. Wassenaar TA, Pluhackova K, Bockmann RA, Marrink SJ, Tieleman DP: **Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models.** *J Chem Theory Comput* 2014, **10**:676-690.
52. Koldso H, Shorthouse D, Helie J, Sansom MS: **Lipid clustering correlates with membrane curvature as revealed by molecular simulations of complex lipid bilayers.** *PLoS Comput. Biol.* 2014, **10**:e1003911.
53. Stansfeld PJ, Hopkinson RJ, Ashcroft FM, Sansom MSP: **The PIP<sub>2</sub> binding site in Kir channels: definition by multi-scale biomolecular simulations.** *Biochem.* 2009, **48**:10926–10933.
54. Aponte-Santamaria C, Briones R, Schenk AD, Walz T, de Groot BL: **Molecular driving forces defining lipid positions around aquaporin-0.** *Proc Natl Acad Sci U S A* 2012, **109**:9887-9892.
55. Stansfeld PJ, Jefferys EE, Sansom MS: **Multiscale simulations reveal conserved patterns of lipid interactions with aquaporins.** *Structure* 2013, **21**:810-819.
56. Briones R, Aponte-Santamaria C, de Groot BL: **Localization and Ordering of Lipids Around Aquaporin-0: Protein and Lipid Mobility Effects.** *Front Physiol* 2017, **8**:124.
57. Hedger G, Sansom MS, Koldso H: **The juxtamembrane regions of human receptor tyrosine kinases exhibit conserved interaction sites with anionic lipids.** *Sci Rep* 2015, **5**:9198.
58. • Gupta K, Donlan JA, Hopper JT, Uzdavinyas P, Landreh M, Struwe WB, Drew D, Baldwin AJ, Stansfeld PJ, Robinson CV: **The role of interfacial lipids in stabilizing membrane protein oligomers.** *Nature* 2017, **541**:421-424.  
The combination of mass spectrometry and molecular dynamics to detect lipid binding sites at the interface of membrane protein oligomers.
59. Arnarez C, Marrink SJ, Periole X: **Identification of cardiolipin binding sites on cytochrome c oxidase at the entrance of proton channels.** *Sci Rep* 2013, **3**:1263.
60. Arnarez C, Mazat JP, Elezgaray J, Marrink SJ, Periole X: **Evidence for cardiolipin binding sites on the membrane-exposed surface of the cytochrome bc<sub>1</sub>.** *J Am Chem Soc* 2013, **135**:3112-3120.
61. Poyry S, Cramariuc O, Postila PA, Kaszuba K, Sarewicz M, Osyczka A, Vattulainen I, Rog T: **Atomistic simulations indicate cardiolipin to have an integral role in the structure of the cytochrome bc<sub>1</sub> complex.** *Biochim Biophys Acta* 2013, **1827**:769-778.
62. Arnarez C, Marrink SJ, Periole X: **Molecular mechanism of cardiolipin-mediated assembly of respiratory chain supercomplexes.** *Chemical Science* 2016, **7**:4435-4443.
63. • Hedger G, Sansom MS: **Lipid interaction sites on channels, transporters and receptors: Recent insights from molecular dynamics simulations.** *Biochim Biophys Acta* 2016, **1858**:2390-2400.

- A timely and comprehensive review of the application of molecular dynamics simulations to study lipid interactions with membrane proteins
64. Domanski J, Hedger G, Best RB, Stansfeld PJ, Sansom MS: **Convergence and Sampling in Determining Free Energy Landscapes for Membrane Protein Association.** *J Phys Chem B* 2016, 10.1021/acs.jpcc.6b08445.
  65. Hedger G, Rouse SL, Domanski J, Chavent M, Koldso H, Sansom MS: **Lipid-Loving ANTs: Molecular Simulations of Cardiolipin Interactions and the Organization of the Adenine Nucleotide Translocase in Model Mitochondrial Membranes.** *Biochemistry* 2016, **55**:6238-6249.
  66. • Ingolfsson HI, Melo MN, van Eerden FJ, Arnarez C, Lopez CA, Wassenaar TA, Periole X, de Vries AH, Tieleman DP, Marrink SJ: **Lipid organization of the plasma membrane.** *J Am Chem Soc* 2014, **136**:14554-14559.  
Molecular simulations of a biologically realistic plasma membrane with 63 different lipid species.
  67. • Koldso H, Sansom MS: **Organization and Dynamics of Receptor Proteins in a Plasma Membrane.** *J Am Chem Soc* 2015, **137**:14694-14704.  
Large-scale molecular dynamics of multiple membrane proteins within a model plasma membrane.
  68. Koldso H, Reddy T, Fowler PW, Duncan AL, Sansom MS: **Membrane Compartmentalization Reducing the Mobility of Lipids and Proteins within a Model Plasma Membrane.** *J Phys Chem B* 2016, **120**:8873-8881.
  69. van Eerden FJ, de Jong DH, de Vries AH, Wassenaar TA, Marrink SJ: **Characterization of thylakoid lipid membranes from cyanobacteria and higher plants by molecular dynamics simulations.** *Biochim Biophys Acta* 2015, **1848**:1319-1330.
  70. van Eerden FJ, van den Berg T, Frederix PWJM, de Jong DH, Periole X, Marrink SJ: **Molecular Dynamics of Photosystem II Embedded in the Thylakoid Membrane.** *The Journal of Physical Chemistry B* 2016, 10.1021/acs.jpcc.6b06865.
  71. Balusek C, Gumbart JC: **Role of the Native Outer-Membrane Environment on the Transporter BtuB.** *Biophys J* 2016, **111**:1409-1417.
  72. Gu Y, Li H, Dong H, Zeng Y, Zhang Z, Paterson NG, Stansfeld PJ, Wang Z, Zhang Y, Wang W, et al.: **Structural basis of outer membrane protein insertion by the BAM complex.** *Nature* 2016, **531**:64-69.
  73. Samsudin F, Ortiz-Suarez ML, Piggot TJ, Bond PJ, Khalid S: **OmpA: A Flexible Clamp for Bacterial Cell Wall Attachment.** *Structure* 2016, **24**:2227-2235.
  74. Chandler DE, Strumpfer J, Sener M, Scheuring S, Schulten K: **Light harvesting by lamellar chromatophores in Rhodospirillum rubrum.** *Biophys J* 2014, **106**:2503-2510.
  75. Sener M, Strumpfer J, Singharoy A, Hunter CN, Schulten K: **Overall energy conversion efficiency of a photosynthetic vesicle.** *Elife* 2016, **5**.
  76. Reddy T, Shorthouse D, Parton DL, Jefferys E, Fowler PW, Chavent M, Baaden M, Sansom MS: **Nothing to sneeze at: a dynamic and integrative computational model of an influenza A virion.** *Structure* 2015, **23**:584-597.
  77. Marzinek JK, Holdbrook DA, Huber RG, Verma C, Bond PJ: **Pushing the Envelope: Dengue Viral Membrane Coaxed into Shape by Molecular Simulations.** *Structure* 2016, **24**:1410-1420.
  78. Reddy T, Sansom MS: **The Role of the Membrane in the Structure and Biophysical Robustness of the Dengue Virion Envelope.** *Structure* 2016, **24**:375-382.
  79. Fahy E, Subramaniam S, Murphy RC, Nishijima M, Raetz CR, Shimizu T, Spener F, van Meer G, Wakelam MJ, Dennis EA: **Update of the LIPID MAPS comprehensive classification system for lipids.** *J Lipid Res* 2009, **50 Suppl**:S9-14.

## Figures

**Figure 1. From sequence to structure.** (a) Extensive multiple sequences alignments, containing thousands of amino acid sequences, are used to identify residues that co-vary, i.e. they change at the same time as one another in different sequences, as shown here for leucine and valine. The covariance can be rationalized structurally by considering that these pairs of residues are in contact within three-dimensional space and therefore the correlated mutation is required to provide optimal protein packing. These pairs of residue contacts may be used as either distance constraints to fold a protein structure or to guide molecular docking of proteins into a macromolecular complex. Molecular models of membrane protein structures, based on co-evolution derived residue pair restraints, compared with X-ray structures of the same protein, or equivalent: (b) Prolipoprotein diacylglycerol transferase (Lgt; PDB id: 5AZB), (c) Lipoprotein signal peptidase (LspA; PDB id: 5DIR), (d) Divalent metal transporter (DMT) superfamily member RarD, compared with the homologous X-ray structure YddG (PDB id: 5I20), (e) The fluoride channel (FluC; PDB id: 5A40).

**Figure 2. Covariance and conformational change.** The covariance data is mapped onto the (a, c, e) outward-open (PDB id: 5GXB) and (b, e, f) inward-open (PDB id: 1PV6) structures of LacY, shown from the (a, b) periplasm, (c, d) membrane and (f, g) cytoplasm. Predicted residue pairs that are less than 12 Å C $\alpha$  distance apart are connected by a green dashed line. The residue pairs greater than 12 Å are coloured red. While many of the contacts are less than 12 Å in both conformational states, some contacts are state-specific, and therefore are greater than 12 Å apart in the alternate conformation.

**Figure 3. Enzymes and lipid substrates.** Exemplar X-ray structures of two integral membrane enzymes with resolved electron density for monoolein lipids. (a) Diacylglycerol kinase (DgkA) with two bound monoolein molecules (purple sticks) and the non-hydrolysable ATP analogue, AMP-PCP (grey sticks)(PDB id: 4UXX). (b) Although DgkA can also phosphorylate monoolein lipids, its native lipid substrate is diacylglycerol. Molecular modelling was used to convert the two monoolein molecules to a single diacylglycerol (purple sticks) and also modify AMP-PCP to ATP (grey sticks). (c) The lipoprotein signal peptidase II (LspA) binds a protein substrate with a lipid-modified cysteine group. Two monoolein molecules (purple sticks) were observed close to the active site of this enzyme in the X-ray structure, adjacent to the bound globomycin antibiotic (grey sticks)(PDB id: 5DIR). (d) Molecular modelling was used to convert the monoolein molecules to a lipid-modified diacylglycerolcysteine (purple sticks), while the globomycin - a peptide mimetic - was used to partially orient the binding of the protein component of the lipoprotein substrate (grey cartoon). The lipoprotein residues that overlap with globomycin are shown in grey sticks.

**Figure 4. The MemProtMD pipeline.** (a) The crystal structure of the  $\beta_2$  adrenergic receptor ( $\beta_2$ -AR)-G-protein complex, with the lysozyme fusion domain and bound nanobody (grey surface)(PDB id: 3SN6). (b) The non-physiological domains are removed leaving only the  $\beta_2$ AR-G-protein complex. (c) The structure is converted to a coarse-grained (CG) representation and submerged in a random assembly of lipids (yellow sticks), lipid phosphates are shown in orange spheres, with solvent shown in blue dots. (d) CG molecular dynamics (CGMD) simulations are performed to assemble the lipid bilayer around the integral membrane protein. (e) The molecular system is converted back to atomistic detail using the CG coordinates. Ions are illustrated as dots, but water molecules are not shown for clarity. (f) Lipid contact analysis is displayed on the protein structure, based on the CGMD simulations, on a white (lipid contacts in no frames) to red (lipid contacts in all frames) scale. The deformations of the bilayer around the protein are shown for the upper and lower leaflet, with thickening of the bilayer coloured red, and thinning coloured blue.