

# The MemProtMD database: a resource for membrane-embedded protein structures and their lipid interactions

Thomas D. Newport, Mark S.P. Sansom and Phillip J. Stansfeld\*

Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, UK

Received September 09, 2018; Revised October 05, 2018; Editorial Decision October 08, 2018; Accepted October 16, 2018

## ABSTRACT

**Integral membrane proteins fulfil important roles in many crucial biological processes, including cell signalling, molecular transport and bioenergetic processes. Advancements in experimental techniques are revealing high resolution structures for an increasing number of membrane proteins. Yet, these structures are rarely resolved in complex with membrane lipids. In 2015, the MemProtMD pipeline was developed to allow the automated lipid bilayer assembly around new membrane protein structures, released from the Protein Data Bank (PDB). To make these data available to the scientific community, a web database (<http://memprotmd.bioch.ox.ac.uk>) has been developed. Simulations and the results of subsequent analysis can be viewed using a web browser, including interactive 3D visualizations of the assembled bilayer and 2D visualizations of lipid contact data and membrane protein topology. In addition, ensemble analyses are performed to detail conserved lipid interaction information across proteins, families and for the entire database of 3506 PDB entries. Proteins may be searched using keywords, PDB or Uniprot identifier, or browsed using classification systems, such as Pfam, Gene Ontology annotation, mpstruc or the Transporter Classification Database. All files required to run further molecular simulations of proteins in the database are provided.**

## INTRODUCTION

There are now ~3500 structures of over 1000 unique integral membrane proteins deposited in the Protein Data Bank (PDB) (1,2). Membrane proteins are of considerable biomedical interest, constituting ~25% of published genomes (3) and 50% of current drug targets (4). A near-exponential increase in the number of published membrane protein structures (5) looks set to be sustained through

continuous improvements to detergent solubilization and crystallization protocols, such as Lipidic Cubic Phase (6), HiLiDe (7) and MemGold (8). Meanwhile, the enhanced resolution and variety of structures solved by Cryo-Electron Microscopy (9) opens up a wealth of new possibilities.

With rapid growth in the number of protein structures available, several database systems have been developed in order to organize and annotate these structures in a biologically meaningful way. The PDB is a universal repository for all experimentally derived membrane protein structures, holding a single accession for each deposited structure. Structures deposited in the PDB may be linked to one or more UniProt (10) protein sequence, which in turn can be grouped into larger Pfam (11) families. In addition the Gene Ontology (GO) project (12) defines a standardized way to further annotate proteins with attributes ranging from functional notes to cellular location. Several specialized classification systems also exist to classify membrane proteins; Membrane Proteins of Known Structure (mpstruc; <http://blanco.biomol.uci.edu/mpstruc/>) (13) places membrane protein structures into a hand-curated tree, whilst the Transporter Classification DataBase (TCDB; <http://www.tcdb.org>) (14) provides an automatically generated tree looking specifically at transporters and channels. The GPCRdb provides comprehensive data, diagrams and web tools for G protein-coupled receptors (GPCRs; <http://gpcrdb.org>) (15). The membrane protein database (MPDB; <http://www.mpdb.tcd.ie>) details the conditions and additives used to experimentally solve membrane protein structures (16).

Membrane proteins are embedded in a lipid bilayer, which has significant influence on both the structure and function of the protein. Whilst it is not uncommon to be able to identify a few lipids when determining a membrane protein structure (17), most lipid components of the membrane are highly dynamic and rapidly exchanged (18), so are not readily resolved by current structure determination techniques.

Methods such as Positioning of Proteins in Membrane (PPM) (19), TMDet (20) and Memembed (21) can use the distribution of surface amino acid residues to infer

\*To whom correspondence should be addressed. Tel: +44 1865 613362; Fax: +44 1865 613201; Email: [phillip.stansfeld@bioch.ox.ac.uk](mailto:phillip.stansfeld@bioch.ox.ac.uk)  
Present address: Thomas D. Newport, Oxford Nanopore Technologies, Gosling Building, Edmund Halley Road, Oxford Science Park, OX4 4DQ, UK.

both the tilt of the protein and the thickness of the lipid bilayer, modelling the bilayer as a slab bounded by two planes. PPM is used in the Orientations of Proteins in Membranes (OPM; <https://opm.phar.umich.edu>) database, while TMDET is applied to entries in the Protein Data Bank of Transmembrane Proteins (PDBTM; <http://pdbtm.enzim.hu>). Although these methods can accurately predict the location of transmembrane (TM) domains, they lack explicit lipids and so are unable to capture specific interactions between individual lipid molecules and sites on the surface of the protein. Highly specific protein–lipid interactions can modulate protein function, either directly (22) or by regulating protein–protein interactions such as oligomerization (23,24). These lipid-binding sites yield potential targets for novel allosteric modulatory drugs from within the membrane, e.g. AZ3451 binding to the PAR2 receptor (25). In other instances, interactions between the protein and specific lipid molecules result in chemical modification of the lipid (26), or the translocation of a lipid molecule between leaflets of the bilayer (27). In several cases, the structure of the protein may induce local deformations of the membrane, thinning or thickening the membrane or inducing curvature (28).

Molecular Dynamics (MD) simulations permit detailed models of protein interactions within biological membranes, characterizing the interactions of a protein with explicit lipid molecules. MD simulations are typically performed on one of two scales: atomistic MD simulations consider each atom in the simulated system as a single particle, producing more accurate simulations, whilst Coarse-Grained (CG) MD simulations treat 3–4 heavy atoms as a single particle (29), reducing the complexity of the system and allowing for longer or larger simulations. Multi-scale approaches simulate a system using both CG and atomistic representations, combining the strengths of both approaches (30,31).

Several methods exist to reconstitute lipid membranes using MD simulations. The CHARMM-GUI (32) lipid-builder tool can either be used to insert a protein structure into a pre-equilibrated bilayer, or generate a new bilayer by packing lipids around a protein structure, whilst INSANE (33) initially builds a CG membrane on a planar grid around the protein. The MemProtMD pipeline (34) instead uses a CG Self-Assembly (CGSA) approach, whereby lipids are initially placed and oriented randomly around the protein, and a short CG simulation is performed, during which a lipid bilayer forms spontaneously (35). The assembled system can then be simulated further in either CG or atomistic representation (36).

Here, we present the MemProtMD database, an online resource which stores and analyses the results of simulations of a complete set of integral membrane protein structures in explicit dipalmitoylphosphatidylcholine bilayers, set up using the MemProtMD pipeline (34). This database builds on earlier concepts first formulated for the CGDB (37). Protein–lipid interactions, membrane deformations, ensemble analyses and bulk membrane properties from the MD simulations are assessed and stored in a database. These data are combined with annotations from external databases to classify proteins and identify conserved protein–lipid interactions across both proteins and fam-

ilies. The database can be browsed using a freely accessible web application (<http://memprotmd.bioch.ox.ac.uk>), and results of analyses, as well as files and instructions required to perform further simulations can be downloaded.

## MATERIALS AND METHODS

### Implementation

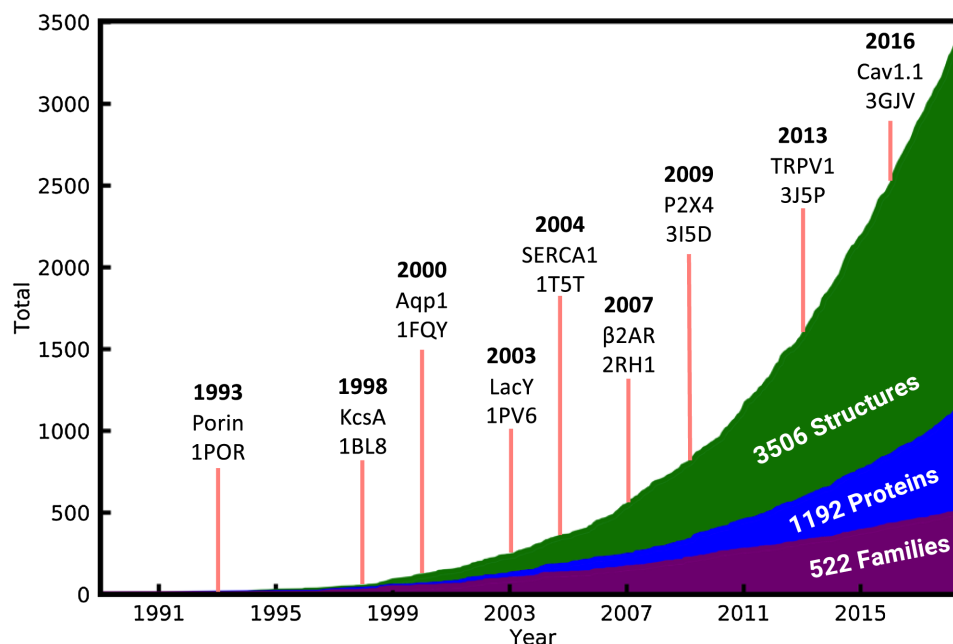
The initial CGSA simulation setup is performed using the MemProtMD pipeline (34). In brief, integral membrane proteins are identified from the PDB based on an Octopus prediction of surface-accessible  $\alpha$ -helical TM domains (38). Potential  $\beta$ -barrel membrane proteins are identified based on the number, length, accessibility and hydrophobicity of their  $\beta$ -strands. Where available, the biological unit in the PDB is prepared for the oligomeric state of the simulated protein. In all instances non-protein atoms are removed from the PDB entry prior to simulation. MD simulations are performed using GROMACS 5.1.4 (39) and the MARTINI 2.2 forcefield (40). Completed simulations are then converted to atomistic representation using CG2AT (30). Analysis of completed simulations is performed using Python, MDAAnalysis (41) and our *in house* mpm-tools, which includes bindings for MUSCLE (42) and a python adaptor for VMD (43), used to render static images. Two dimensional visualizations of data are performed using D3.js. Three dimensional protein visualization uses PV. The database is stored using MongoDB. The web server uses NodeJS and Express to serve a frontend application built on ReactJS and Redux. Server application deployment is performed using Docker Compose.

## RESULTS

### The MemProtMD database and web server contents

At time of writing, the MemProtMD database holds 3506 CGMD simulations of intrinsic membrane proteins inserted into phospholipid bilayers, each based on a single structure deposited in the PDB. Represented within these are TM proteins, which correspond to 1192 unique UniProt proteins and 522 Pfam families. The cumulative number of intrinsic membrane protein structures deposited in the PDB continues to rise at an almost exponential rate, as shown in Figure 1. This figure contains a timeline of a selection of key membrane protein structures (44–52). If the same trend is followed, we expect that in 10 years there will be ~10 500 membrane protein PDB structures, ~3500 unique proteins and ~1200 Pfam entries. The most up-to-date values can be found at <http://memprotmd.bioch.ox.ac.uk/stats>.

As part of our workflow, intrinsic membrane proteins are automatically identified and downloaded on a weekly basis as they are released from the PDB in Europe (2), using the MemProtMD pipeline (34). At present, this averages ~8 structures per week. In 10 years' time we expect this to be closer to 20 structures per week. We expect that due to further methodological advances, this is likely to be a conservative estimate. Upon identification, the coordinates are converted to a CG representation for a self-assembly simulation with phospholipids, water and ions (35). This 1  $\mu$ s MD simulation is performed and the final frame converted back to



**Figure 1.** The number of unique PDB, UniProt and Pfam accessions represented in the MemProtMD database over time. A selection of landmark structures are indicated by red lines, denoting their PDB release date and include the first Porin (44),  $K^+$  channel (45), aquaporin (46), solute transporter (47), Sec translocon (48), adrenergic receptor (49), glutamate receptor (50), TRP channel (51) and voltage-gated  $Ca^{2+}$  channel (52). The total number of PDB structures, UniProt proteins and Pfam families are shown.

atomistic representation using CG2AT (30). This process is shown schematically in Figure 2. If the protein successfully inserts into a lipid bilayer during the simulation, a set of analyses are performed to identify contacts between lipid molecules and the protein surface, as well as characterize the surface of each leaflet of the lipid bilayer. These data are then deposited in the MemProtMD database along with CG and atomistic snapshots of the final simulation frame.

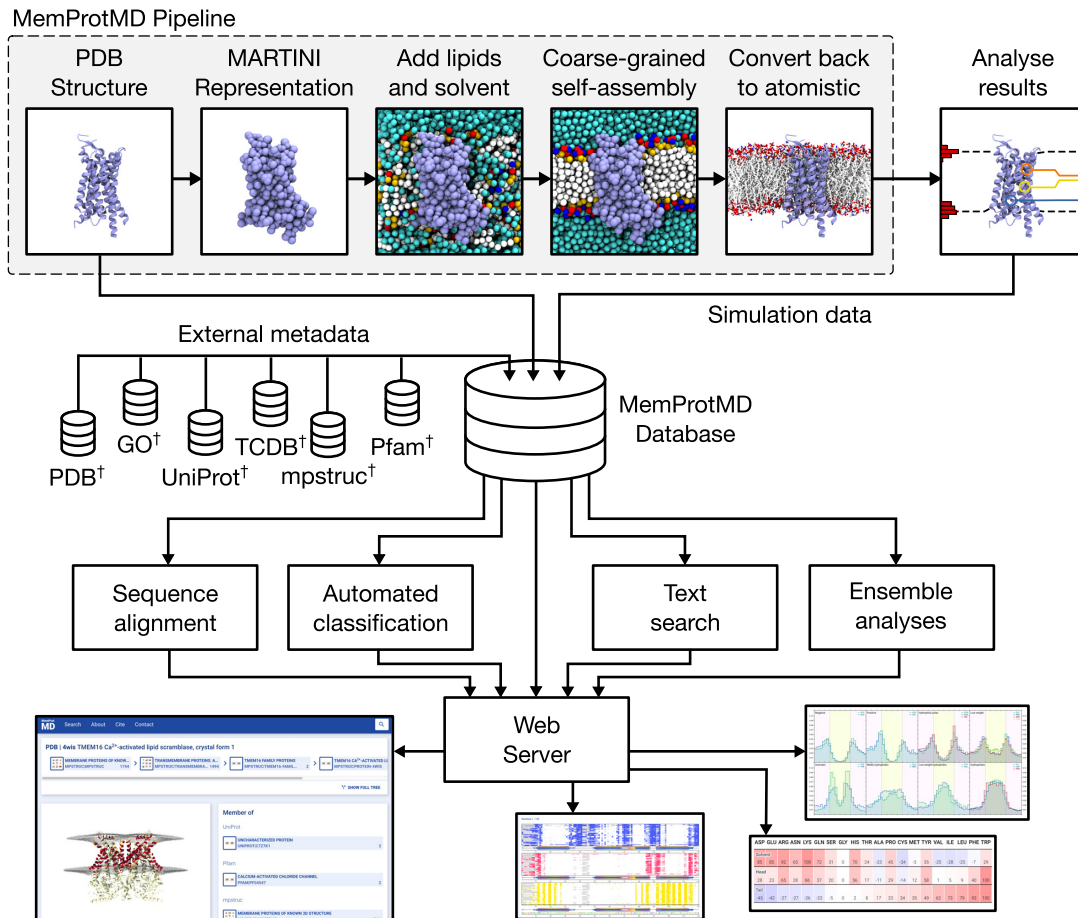
Analysis of specific amino acid residues in structures sharing UniProt or Pfam accessions is then aggregated either according to the aligned amino acid sequence of the protein, or the position of the residue relative to the annular lipids of the simulated bilayer. The results of simulations and subsequent analysis can be visualized and downloaded using the MemProtMD web server, available at <http://memprotmd.bioch.ox.ac.uk>.

### Simulation analysis

Completed simulations pass through an analysis pipeline which performs several assessments to characterize the interactions between lipid bilayer and membrane protein. First, the upper and lower leaflets are identified, as well as any movement of lipid molecules between leaflets. A radial distribution function of lipid head groups around the protein is then calculated and used to determine the shape of annular shells of lipids around the protein. The surfaces formed by each leaflet are then characterized, as well as distortions in these caused by the protein (Figure 3A). Contacts between the protein and different parts of the surrounding lipids and solvents are then calculated, and a set of static images is rendered (Figure 3).

The results of these analyses can be searched for and visualized in a modern web browser using the MemProtMD server (Supplementary Figure S1). An overview for each protein shows metadata derived from external databases, as well as links to download files to run new simulations (Supplementary Figure S2). Pre-rendered images can be browsed (Figure 3A), and a summary of the topology of protein chains can be used to select an individual chain (Figure 3B). This selection is reflected in an online 3D viewer (Figure 3C), which can be configured to show a snapshot of the protein and membrane lipids, coloured by metrics derived from simulation analyses and selectable by the user. When a protein chain is selected, a 2D topology viewer shows a simplified view of each secondary structure element, using a B-spline approximation which preserve bends and kinks (Figure 3D). Unlike other topology viewers, this also captures the thickness changes of the explicit membrane around the protein.

A sequence view of the selected chain (Figure 3E) shows metrics mapped to the amino acid sequence of the protein. Two coloured series at the top identify amino acid residues which contact leaflets flipping between two bilayers and amino acids which contact water and ions. Contacts with each CG bead of membrane lipids are then shown for each amino acid residue, starting with the choline bead of the upper leaflet, progressing through phosphate beads, glycerol beads and the four beads representing acyl tails of the lipid. Below this is a plot showing contacts with three larger groups of beads; lipid head-group beads, lipid acyl tail beads and solvent beads, represented as stacked bars. The displacement of each protein residue from the bilayer centre is calculated, as well as the displacement of the closest point on the surface of each membrane leaflet. Using this



**Figure 2.** Overview of the MemProtMD pipeline and database. Membrane protein structures deposited in the PDB are first converted to MARTINI representation. Lipids and solvent are then added with random orientations and a 1  $\mu$ s self-assembly simulation is performed. The simulation is converted back to atomistic representation using CG2AT and several analyses are performed. Analysis results, as well as metadata derived from the original PDB structure and metadata downloaded from a range of databases are deposited into the MemProtMD database, which is then used to perform multiple sequence alignments, automatically classify membrane proteins, perform text searches and perform ensemble analyses.

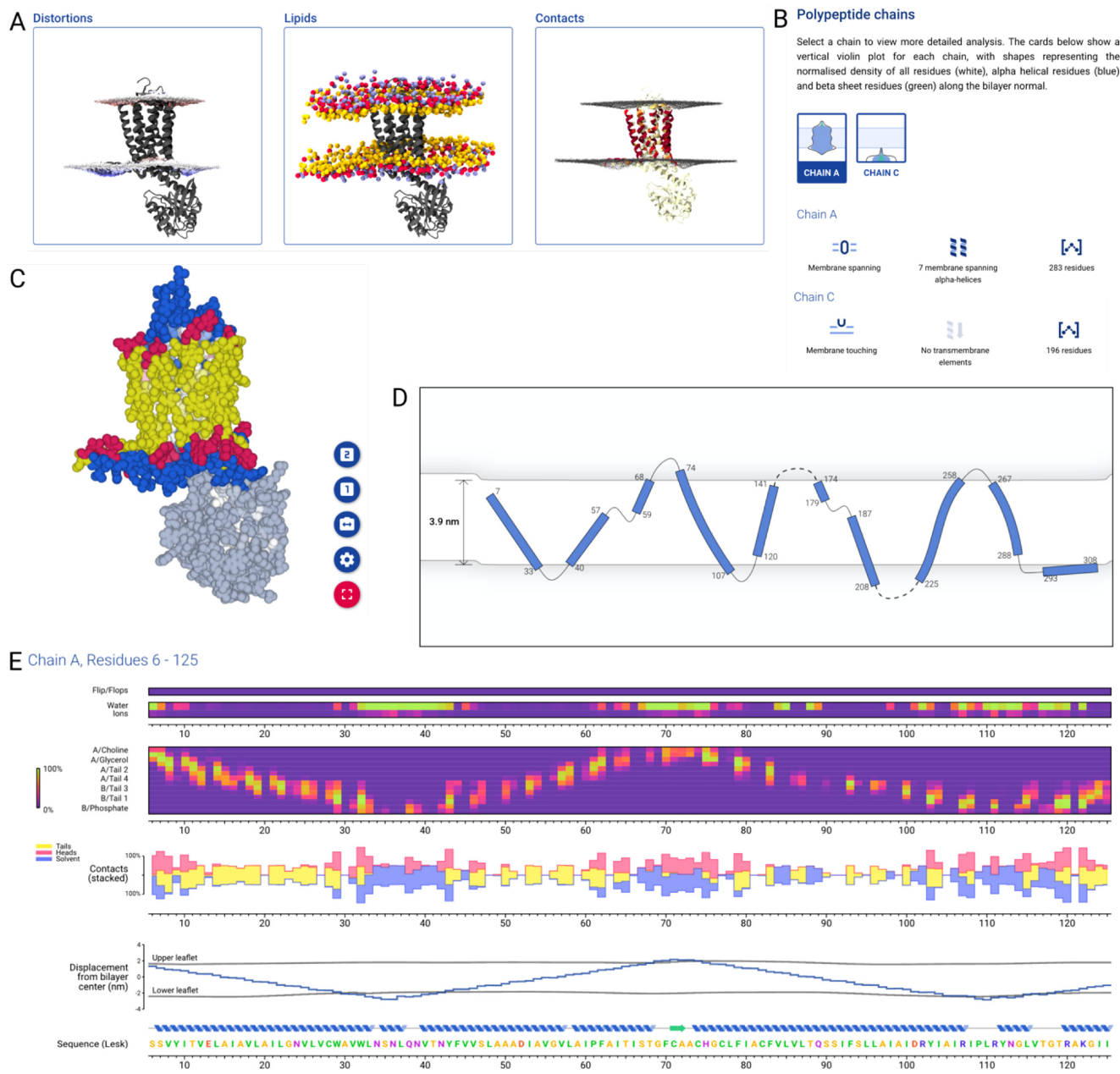
calculation, it is possible to display bilayer thickness and distortions along the amino acid sequence, identifying sets of residues which may be involved in distorting the membrane.

### Ensemble analyses

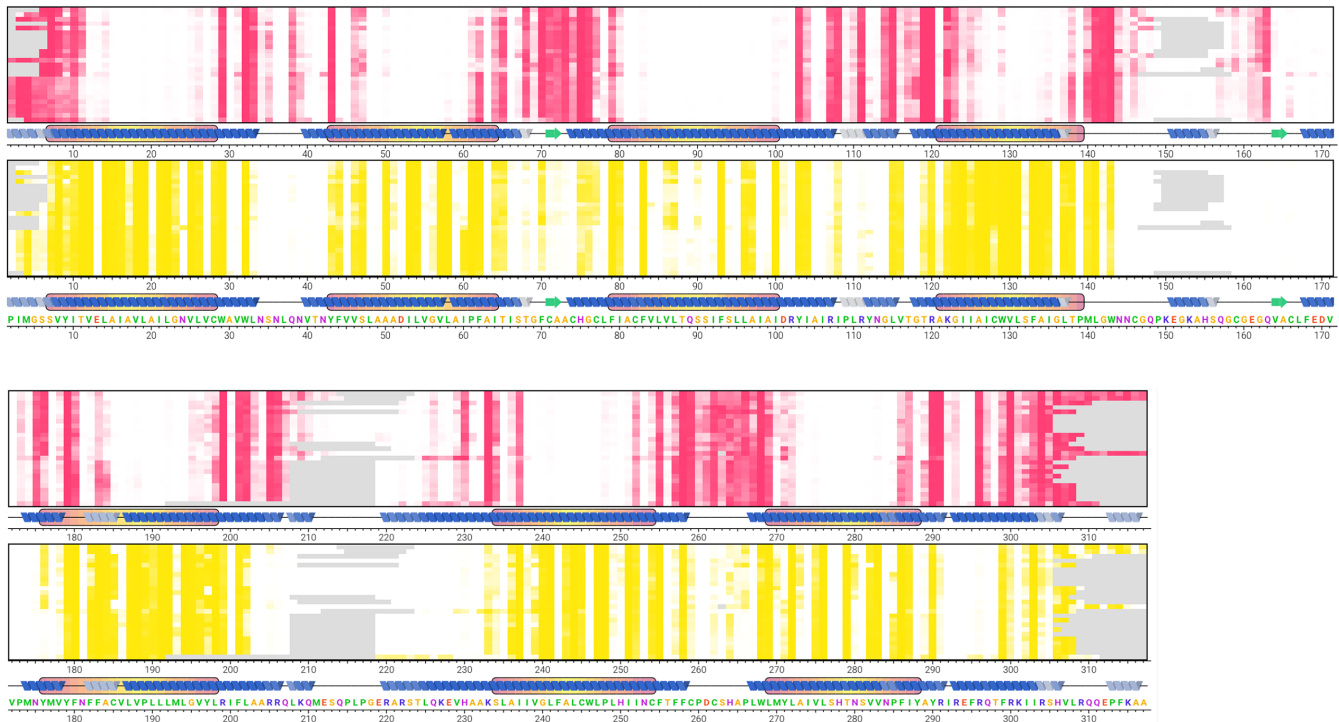
The 3506 membrane protein simulations, shown in Supplementary Figure 3, enable single-entry, collective and large-scale evaluation of general membrane protein properties. For a given UniProt identifier with multiple PDB entries, per-residue analyses of the aggregated simulations may be viewed along the protein sequence, shown here for the multiple structures of the human A<sub>2A</sub> adenosine receptor (Figure 4). This reveals a regular and highly conserved pattern of membrane head group and tail contacts, in addition to solvent exposed residues. As part of this analysis consensus, secondary structure elements and TM domains of the protein are shown. This methodology is also extended to each Pfam family by aligning all related UniProt sequences using MUSCLE (42). By incorporating the lipid contact information, an equivalent sequence alignment may then be produced for each of the 522 Pfam families in MemProtMD, as

we performed previously, in a non-automated manner, for the Aquaporin family (53). This reveals a highly conserved pattern of lipid contacts across all GPCRs within the 7TM receptor Pfam (PF00001), which currently includes 222 protein structures (Supplementary Figure S4). Other exemplar Pfam entries include that of the POT family (PF00854), the Sodium:neurotransmitter symporter family (PF00209) and an updated version of the Aquaporin family, now with 51 entries, using the Major intrinsic protein Pfam (PF00230).

Using amalgamated simulation data from every single protein in the MemProtMD database, further light can be shed on the distribution of amino acids within lipid bilayers. For each residue type, the unbiased probability of a contact between that residue and either lipid head-groups, lipid acyl tails or solvent is calculated (Supplementary Figure S5A). The probabilities are used to construct a normalized scale, with glycine assigned a value of 0 and the maximum probability being 100. Due to our explicit lipid methodology, we can unpack these data further, capturing the specific portion of the lipid that interacts preferentially with each amino acid residue, as we previously performed for aquaporins (53). As we observed for the aquaporin dataset, the two basic residues, arginine and lysine, preferentially in-



**Figure 3.** Online analysis and visualization of a simulation of the A<sub>2A</sub> adenosine receptor. (A) Pre-rendered images available to view and download showing (from left to right) protein with membrane surface shaded to show thinning (red) and thickening (blue), protein with CG lipid head-group beads (yellow: glycerol, red: phosphate, blue: choline) and protein surface shaded according to contacts with lipid head-groups. (B) Chain-by-chain overview of the protein, showing violin plot of protein density along membrane normal, classification of chain membrane interactions ('Membrane spanning'), topology summary ('7 membrane spanning  $\alpha$  helices') and residue count. (C) Online 3D protein view showing protein in sphere representation coloured by membrane contact type (yellow: acyl tails, red: lipid head-groups, blue: solvent). The selected chain (Chain A) is shown in full colour, whilst other parts of the protein are desaturated. (D) Topology viewer showing tilt and curvature of TM helices. Non-continuous loops are shown as dashed lines. The membrane thickness and degree of protein-induced deformation is shown. (E) Sequence viewer showing several metrics along the amino acid sequence: (from top to bottom) contacts with lipids flipping from one leaflet to another; contacts with water and ions; contacts with lipids, broken down by chemical group; stacked plot of contacts with acyl tails, head-groups and solvent; position of the protein and local membrane thickness; protein secondary structure and amino acid sequence.



**Figure 4.** Online sequence view for 25 simulations of the UniProt AA2AR\_HUMAN entry (Uniprot id: P29274) contacts with lipid head-groups (red) and lipid acyl tails (yellow) are shown arranged along the protein amino acid sequence, with each row representing a single simulated structure. A consensus secondary structure is shown at the bottom of each row for each PDB file, shaded darker where secondary structure is more highly conserved within the structures. TM domains are shown by a box around the secondary structure, shaded according to mean depth within the membrane, from red (shallow) to yellow (deep), where a certain residue was not present in a structure the cell is shaded grey.

teract with the phosphate group, tyrosine and tryptophan form the majority of interactions with the glycerol region of the lipid, while phenylalanine, leucine, isoleucine, tryptophan and valine most commonly form contacts with the terminal lipid tails. (Supplementary Figure S5B). These data dynamically updates on a weekly basis, with the latest data found at <http://memprotmd.bioch.ox.ac.uk/stats>.

Also displayed on this page is the mean distribution of surface amino acid residues for all simulations. This is visualized along the bilayer normal (Supplementary Figure S6). Pore inner residues are discounted from the analysis, and each protein residue is assigned a coordinate for displacement along the membrane normal, normalized to account for variations in membrane thicknesses. Residue types with similar distributions are clustered together. Histograms for each amino acid are normalized to sum to 1. The data may be dynamically viewed based on the absolute residue distribution or normalized to the relative number of the amino acid species. As observed for a smaller dataset (54), hydrophobic residues isoleucine, leucine, valine and phenylalanine are found in the membrane core, the aromatic residues tryptophan and tyrosine are found at the membrane-solvent interface, near the more solvent exposed arginine and lysine.

Global analysis also illustrates a clear difference in the membrane thickness (phosphate centroid to phosphate centroid) proximal to the protein, between bilayers containing  $\alpha$ -helical proteins, typically  $\sim 37$  Å in thickness, and membranes of  $\beta$ -barrel proteins, typically  $\sim 33$  Å in thickness

(Supplementary Figure S7). This is likely due to the reduced hydrophobic thickness of bacterial outer membranes, in which most of the  $\beta$ -barrel membrane proteins reside.

### Integration with other databases

Protein structures are initially identified by their PDB deposition. In MemProtMD, the links to the PDB are maintained, with PDB metadata used to populate the database. Structures are grouped according to their constituent proteins, as defined by UniProt accessions, and according to their family, defined by Pfam accessions. A grouping of proteins is also created from each GO annotation. Each node defined by the TCDB and mpstruc trees is also used to group proteins, using the classifications from the external databases. A page showing all simulations in the defined group, as well as statistics for that group (where the group has more than 50 structures) can be viewed online. We have also configured our own 'mpm' annotation scheme, with links to groupings of GPCRs, Ion channels, Outer membrane proteins, Aquaporins and ATP Binding Cassette Transporters, which are shown on the home page of the website.

### Search and classification tools

The MemProtMD database automatically classifies proteins based on a variety of criteria. These may include keywords, references to accessions in external databases or metrics derived from simulations of the protein. These clas-

sifications are applied automatically and are then refined manually as needed. An example of the automated classification is shown in Supplementary Figure S2 for the A<sub>2A</sub> GPCR protein. The MemProtMD database can also be searched using the MemProtMD web server using a text-based search against keywords, titles, accessions and descriptions of entries in external databases. Results are returned, ordered by relevance and may include individual simulations or collections comprising of several simulations sharing a particular reference to an external database.

## CONCLUSION AND OUTLOOK

Here we have described the MemProtMD database, which houses automated MD simulations and analyses for membrane protein structures embedded in explicit lipid membranes. In addition to displaying data for individual PDB entries, ensemble analyses have also been performed to characterize lipid interactions with multiple Uniprot entries and Pfam families. The database is designed for sustainability with automated weekly updates based on the latest release of membrane protein structures. Going forward we will expand the database to incorporate a diverse array a membrane lipid species (33), selected based on the nature of the membrane, e.g. bacterial or eukaryotic, endosomal or plasma membrane. This will enable the identification of specific lipid binding sites within the annular shell.

## DATA AVAILABILITY

The MemProtMD database can be accessed through the web server at <http://memprotmd.bioch.ox.ac.uk>. Regular updates of new entries are provided from the Twitter account @memprotmd.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We would like to thank members of both groups, other labs in the Department of Biochemistry, Oxford and many other structural biologists, too many to name, for extremely helpful discussions.

## FUNDING

Wellcome [208361/Z/17/Z]; BBSRC [BB/P01948X/1, BB/I019855/1]; EPSRC Doctoral Training Centre studentship (to TDN); ARCHER UK National Supercomputing Service, HECBioSim; UK High End Computing Consortium for Biomolecular Simulation ([hecbiosim.ac.uk](http://hecbiosim.ac.uk)), EPSRC [EP/L000253/1]. Funding for open access charge: Wellcome Trust.

*Conflict of interest statement.* None declared.

## REFERENCES

- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
- Velankar, S., Alhroub, Y., Alili, A., Best, C., Boutselakis, H.C., Caboche, S., Conroy, M.J., Dana, J.M., van Ginkel, G., Golovin, A. *et al.* (2011) PDB: Protein Data Bank in Europe. *Nucleic Acids Res.*, **39**, D402–D410.
- Bakheet, T.M. and Doig, A.J. (2009) Properties and identification of human protein drug targets. *Bioinformatics*, **25**, 451–457.
- Fagerberg, L., Jonasson, K., von Heijne, G., Uhlen, M. and Berglund, L. (2010) Prediction of the human membrane proteome. *Proteomics*, **10**, 1141–1149.
- White, S.H. (2004) The progress of membrane protein structure determination. *Prot. Sci.*, **13**, 1948–1949.
- Caffrey, M. (2015) A comprehensive review of the lipid cubic phase or in meso method for crystallizing membrane and soluble proteins and complexes. *Acta Crystallogr. F Struct. Biol. Commun.*, **71**, 3–18.
- Gourdon, P., Andersen, J.L., Hein, K.L., Bublitz, M., Pedersen, B.P., Liu, X.-Y., Yatime, L., Nyblom, M., Nielsen, T.T., Olesen, C. *et al.* (2011) HiLiDe—Systematic approach to membrane protein crystallization in lipid and detergent. *Cryst. Growth Des.*, **11**, 2098–2106.
- Newstead, S., Ferrandon, S. and Iwata, S. (2008) Rationalizing alpha-helical membrane protein crystallization. *Prot. Sci.*, **17**, 466–472.
- Liao, M., Cao, E., Julius, D. and Cheng, Y. (2014) Single particle electron cryo-microscopy of a mammalian ion channel. *Curr. Opin. Struct. Biol.*, **27**, 1–7.
- Bairoch, A., Apweiler, R., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M. *et al.* (2005) The universal protein resource (UniProt). *Nucleic Acids Res.*, **33**, D154–D159.
- Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J. *et al.* (2014) Pfam: the protein families database. *Nucleic Acids Res.*, **42**, D222–D230.
- Huntley, R.P., Sawford, T., Mutowo-Muullenet, P., Shypitsyna, A., Bonilla, C., Martin, M.J. and O'Donovan, C. (2015) The GOA database: gene Ontology annotation updates for 2015. *Nucleic Acids Res.*, **43**, D1057–D1063.
- White, S.H. (2009) Biophysical dissection of membrane proteins. *Nature*, **459**, 344–346.
- Saier, M.H. Jr., Tran, C.V. and Barabote, R.D. (2006) TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Res.*, **34**, D181–D186.
- Pandy-Szekeres, G., Munk, C., Tsonkov, T.M., Mordalski, S., Harpsoe, K., Hauser, A.S., Bojarski, A.J. and Gloriam, D.E. (2018) GPCRdb in 2018: adding GPCR structure models and ligands. *Nucleic Acids Res.*, **46**, D440–D446.
- Raman, P., Cherezov, V. and Caffrey, M. (2006) The Membrane Protein Data Bank. *Cell. Mol. Life Sci.*, **63**, 36–51.
- Yeagle, P.L. (2014) Non-covalent binding of membrane lipids to membrane proteins. *Biochim. Biophys. Acta*, **1838**, 1548–1559.
- Lee, A.G. (2004) How lipids affect the activities of integral membrane proteins. *Biochim. Biophys. Acta*, **1666**, 62–87.
- Lomize, M.A., Pogozheva, I.D., Joo, H., Mosberg, H.I. and Lomize, A.L. (2012) OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.*, **40**, D370–D376.
- Tusnady, G.E., Dosztanyi, Z. and Simon, I. (2005) TMDet: web server for detecting transmembrane regions of proteins by using their 3D coordinates. *Bioinformatics*, **21**, 1276–1277.
- Nugent, T. and Jones, D.T. (2013) Membrane protein orientation and refinement using a knowledge-based statistical potential. *BMC Bioinformatics*, **14**, 276.
- Poveda, J.A., Giudici, A.M., Renart, M.L., Molina, M.L., Montoya, E., Fernandez-Carvajal, A., Fernandez-Ballester, G., Encinar, J.A. and Gonzalez-Ros, J.M. (2014) Lipid modulation of ion channels through specific binding sites. *Biochim. Biophys. Acta*, **1838**, 1560–1567.
- Arkhipov, A., Shan, Y., Das, R., Endres, N.F., Eastwood, M.P., Wemmer, D.E., Kuriyan, J. and Shaw, D.E. (2013) Architecture and membrane interactions of the EGF receptor. *Cell*, **152**, 557–569.
- Gupta, K., Donlan, J.A., Hopper, J.T., Uzdavinys, P., Landreh, M., Struwe, W.B., Drew, D., Baldwin, A.J., Stansfeld, P.J. and Robinson, C.V. (2017) The role of interfacial lipids in stabilizing membrane protein oligomers. *Nature*, **541**, 421–424.

25. Cheng, R.K.Y., Fiez-Vandal, C., Schlenker, O., Edman, K., Aggeler, B., Brown, D.G., Brown, G.A., Cooke, R.M., Dumelin, C.E., Dore, A.S. *et al.* (2017) Structural insight into allosteric modulation of protease-activated receptor 2. *Nature*, **545**, 112–115.
26. Fagone, P. and Jackowski, S. (2009) Membrane phospholipid synthesis and endoplasmic reticulum function. *J. Lipid Res.*, **50**, S311–S316.
27. Brunner, J.D., Lim, N.K., Schenck, S., Duerst, A. and Dutzler, R. (2014) X-ray structure of a calcium-activated TMEM16 lipid scramblase. *Nature*, **516**, 207–212.
28. Pliotas, C., Dahl, A.C., Rasmussen, T., Mahendran, K.R., Smith, T.K., Marius, P., Gault, J., Banda, T., Rasmussen, A., Miller, S. *et al.* (2015) The role of lipids in mechanosensation. *Nat. Struct. Mol. Biol.*, **22**, 991–998.
29. Monticelli, L., Kandasamy, S.K., Periole, X., Larson, R.G., Tieleman, D.P. and Marrink, S.J. (2008) The MARTINI coarse grained force field: extension to proteins. *J. Chem. Theor. Comp.*, **4**, 819–834.
30. Stansfeld, P. and Sansom, M. (2011) From coarse grained to atomistic: a serial Multi scale approach to membrane protein simulations. *J. Chem. Theory Comput.*, **7**, 1157–1166.
31. Ayton, G.S. and Voth, G.A. (2009) Systematic Multi scale simulation of membranes protein systems. *Curr. Opin. Struct. Biol.*, **19**, 138–144.
32. Lee, J., Cheng, X., Swails, J.M., Yeom, M.S., Eastman, P.K., Lemkul, J.A., Wei, S., Buckner, J., Jeong, J.C., Qi, Y. *et al.* (2016) CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *J. Chem. Theory Comput.*, **12**, 405–413.
33. Wassenaar, T.A., Ingolfsson, H.I., Bockmann, R.A., Tieleman, D.P. and Marrink, S.J. (2015) Computational lipidomics with insane: a versatile tool for generating custom membranes for molecular simulations. *J. Chem. Theory Comput.*, **11**, 2144–2155.
34. Stansfeld, P.J., Goose, J.E., Caffrey, M., Carpenter, E.P., Parker, J.L., Newstead, S. and Sansom, M.S. (2015) MemProtMD: automated insertion of membrane protein structures into explicit lipid membranes. *Structure*, **23**, 1350–1361.
35. Bond, P.J. and Sansom, M.S.P. (2006) Insertion and assembly of membrane proteins via simulation. *J. Am. Chem. Soc.*, **128**, 2697–2704.
36. Stansfeld, P.J. and Sansom, M.S.P. (2011) From coarse-grained to atomistic: a serial multi-scale approach to membrane protein simulations. *J. Chem. Theor. Comp.*, **7**, 1157–1166.
37. Chetwynd, A.P., Scott, K.A., Mokrab, Y. and Sansom, M.S.P. (2008) CGDB: a database of membrane protein/lipid interactions by coarse-grained molecular dynamics simulations. *Mol. Membr. Biol.*, **25**, 662–669.
38. Viklund, H. and Elofsson, A. (2008) OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. *Bioinformatics*, **24**, 1662–1668.
39. Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B. and Lindahl, E. (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **1–2**, 19–25.
40. Monticelli, L., Kandasamy, S.K., Periole, X., Larson, R.G., Tieleman, D.P. and Marrink, S.J. (2008) The MARTINI Coarse-Grained force field: extension to proteins. *J. Chem. Theory Comput.*, **4**, 819–834.
41. Michaud-Agrawal, N., Denning, E.J., Woolf, T.B. and Beckstein, O. (2011) MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.*, **32**, 2319–2327.
42. Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.
43. Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD - Visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38.
44. Weiss, M.S. and Schulz, G.E. (1992) Structure of porin refined at 1.8 Å resolution. *J. Mol. Biol.*, **227**, 493–509.
45. Doyle, D.A., Cabral, J.M., Pfuetzner, R.A., Kuo, A., Gulbis, J.M., Cohen, S.L., Cahit, B.T. and MacKinnon, R. (1998) The structure of the potassium channel: molecular basis of K<sup>+</sup> conduction and selectivity. *Science*, **280**, 69–77.
46. Murata, K., Mitsuoaka, K., Hirai, T., Walz, T., Agre, P., Heymann, J.B., Engel, A. and Fujiyoshi, Y. (2000) Structural determinants of water permeation through aquaporin-1. *Nature*, **407**, 599–605.
47. Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H.R. and Iwata, S. (2003) Structure and mechanism of the lactose permease of *Escherichia coli*. *Science*, **301**, 610–615.
48. van den Berg, B., Clemons, W.M., Collinson, I., Modis, Y., Hartmann, E., Harrison, S.C. and Rapoport, T.A. (2004) X-ray structure of a protein-conducting channel. *Nature*, **427**, 36–44.
49. Cherezov, V., Rosenbaum, D.M., Hanson, M.A., Rasmussen, S.G., Thian, F.S., Kobilka, T.S., Choi, H.J., Kuhn, P., Weis, W.I., Kobilka, B.K. *et al.* (2007) High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science*, **318**, 1258–1265.
50. Sobolevsky, A.I., Rosconi, M.P. and Gouaux, E. (2009) X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor. *Nature*, **462**, U745–U766.
51. Liao, M., Cao, E., Julius, D. and Cheng, Y. (2013) Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature*, **504**, 107–112.
52. Wu, J., Yan, Z., Li, Z., Qian, X., Lu, S., Dong, M., Zhou, Q. and Yan, N. (2016) Structure of the voltage-gated calcium channel Ca(v)1.1 at 3.6 Å resolution. *Nature*, **537**, 191–196.
53. Stansfeld, P.J., Jefferys, E.E. and Sansom, M.S. (2013) Multi scale simulations reveal conserved patterns of lipid interactions with aquaporins. *Structure*, **21**, 810–819.
54. Scott, K.A., Bond, P.J., Ivetic, A., Chetwynd, A.P., Khalid, S. and Sansom, M.S.P. (2008) Coarse-grained MD simulations of membrane protein-bilayer self-assembly. *Structure*, **16**, 621–630.