

From Coarse Grained to Atomistic: A Serial Multiscale Approach to Membrane Protein Simulations

Phillip J. Stansfeld and Mark S.P. Sansom*

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

 Supporting Information

ABSTRACT: Coarse-grained molecular dynamics provides a means for simulating the assembly and the interactions of membrane protein/lipid complexes at a reduced level of representation, allowing longer and larger simulations. We describe a fragment-based protocol for converting membrane simulation systems, comprising a membrane protein embedded in a phospholipid bilayer, from coarse-grained to atomistic resolution, for further refinement and analysis via atomistic simulations. Overall, this provides a method for generating an accurate and well equilibrated membrane protein/lipid complex. We exemplify the protocol using the acid-sensing/amiloride-sensitive ion channel protein (ASIC) channel protein, a trimeric integral membrane protein. The method is further evaluated using a test set of 10 different membrane proteins of differing size and complexity. Simulations are assessed in terms of protein conformational drift, lipid/protein interactions, and lipid dynamics.

■ INTRODUCTION

Membrane proteins play key roles in cell biology, e.g., in transport and in signaling. As a consequence, membrane proteins account for ~25% of genes¹ and ~50% of the potential drug targets.² There is ongoing progress in the determination of membrane protein structures by X-ray diffraction and other methods,³ which has resulted in ~250 unique structures (see http://blanco.biomol.uci.edu/membrane_proteins_xtal.html for a summary). However, such structures only occasionally (e.g., for AqpO)^{4,5} reveal full details of protein/lipid interactions. At the same time spectroscopic^{6,7} and functional⁸ studies indicate the importance of characterizing the nature of the interactions of membrane proteins with their lipid bilayer environment.

Molecular dynamics (MD) simulations and related methods have an important role in helping us to fully understand the structural dynamics of membrane proteins.^{9,10} However, prior to commencing these simulations the lipid/protein system should first be near optimally configured. The standard computational method for incorporating a protein into a lipid bilayer is to position the protein within the preformed membrane, delete the overlapping lipids, and then equilibrate the resulting complex.¹¹ However, this method requires prior knowledge of the transmembrane region of the protein, with regions of high hydrophobicity, flanked by tyrosine, tryptophan, and basic residues used as indicators.¹² A number of online tools such as the Orientations of Proteins in Membranes (OPM) database¹³ (<http://opm.phar.umich.edu/>) can be used to guide this process, but they simplify the representation of the lipid bilayer to a hydrophobic slab.

A number of studies have shown that coarse-grained molecular dynamics (CG-MD) simulations¹⁴ may be used to characterize the interactions of membrane proteins with bilayer lipids^{15,16} (<http://sbcg.bioch.ox.ac.uk/cgdb>). However, this method simplifies the representation of both protein and lipid and so inevitably, e.g., the energetics of protein/lipid interactions are approximated.^{17,18} Other approaches include use of a mixed

CG-AT system (e.g., refs 19–22). It is therefore desirable to be able to adopt a serial multiscale approach,²³ whereby CG-MD simulations may be used to efficiently explore membrane protein/lipid interactions, yielding system configuration which may then be converted to atomistic resolution and further refined and characterized in detail by atomistic MD (AT-MD) simulations. One challenge in undertaking such an approach is to develop robust and efficient procedures for conversion of complex protein/lipid bilayer systems from CG to AT representations. This can be achieved in a number of ways.^{24,25} Here we describe a fragment-based approach, which we evaluate via application to a test set of 10 different membrane proteins and use as the basis of a comparison of lipid/protein interactions as predicted by CG and AT-MD simulations.

■ METHODS

CG to Atomistic Conversion. The overall problem is to generate an atomistic protein/lipid bilayer system which retains all of the key interactions of the corresponding coarse-grained model. Coarse-grained to atomistic (CG2AT) conversion is an intrinsically under determined problem, and so additional stereochemical information must be added, directly or indirectly. In our approach this additional information is provided by fragment-based libraries. For each protein a CG protein/lipid bilayer system complex, generated by a self-assembly CG-MD simulation,^{16,26} was used as the starting point for the conversion process. Thus, each system had previously been subjected to 500 ns of CG-MD to allow the self-assembly and equilibration of a lipid bilayer around the protein. The conversion starts by renaming all CG lipid and protein particles to their atomistic counterpart. The protocol (written in perl) uses a number of tools from Gromacs v4.5.3,²⁷ with both the standard Martini

Received: October 4, 2010

Published: March 16, 2011

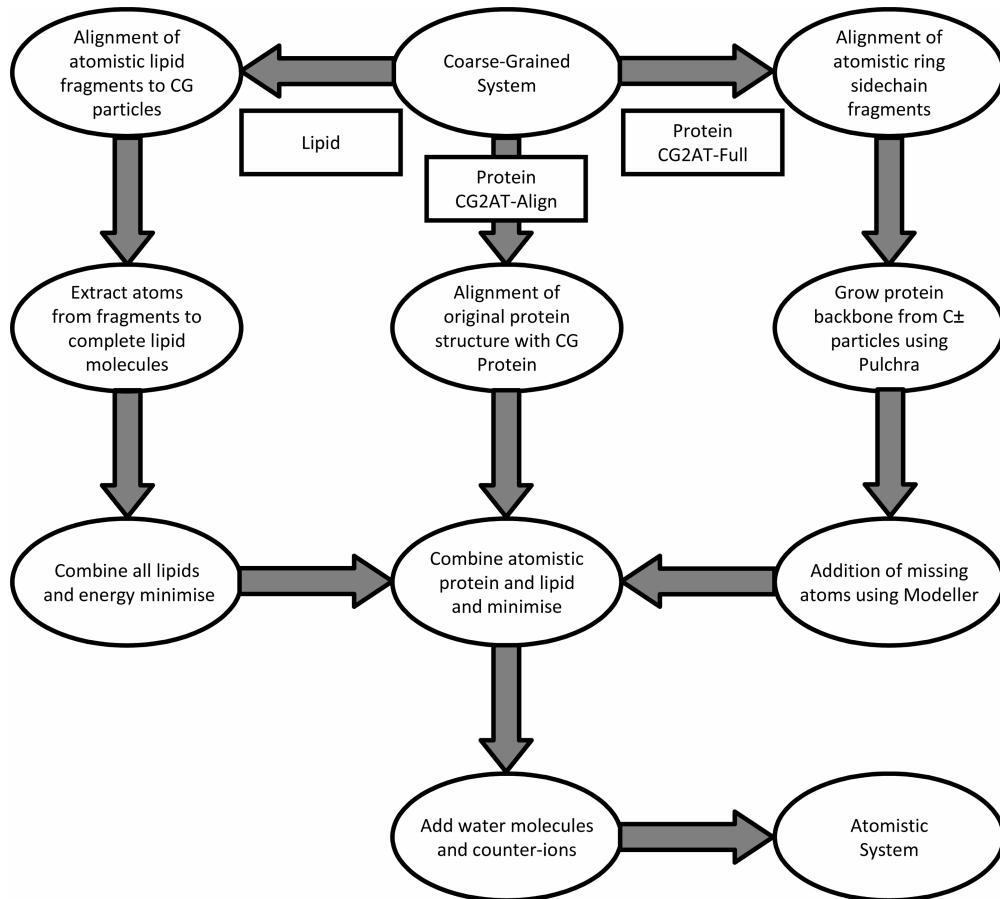


Figure 1. Flowchart describing the CG to AT conversion. Description of the CG2AT conversion methodology. Lipids and proteins are converted independently. The protein may be converted by either realignment of the original atomistic PDB structure (CG2AT-align) or reconstruction of the atomistic coordinates from the CG model using Pulchra and Modeler (CG2AT-full).

v2.1^{28,29} CG model and our local modification of Martini¹⁸ being supported. The protocol currently supports GROMOS, OPLS, and CHARMM36 force fields for the atomistic simulations. A flowchart detailing the conversion process is shown in Figure 1. A typical conversion takes 15 min on a standard Linux workstation for a 25 000 CG-particle system, with relatively small timing differences depending on the options used.

Protein Conversion. Two alternative approaches were used for CG2AT conversion of the protein. The first (CG2AT-full) uses Modeller and Pulchra to construct an atomistic protein structure from the CG particles.^{30,31} The second (CG2AT-align) structurally aligns the original protein PDB structure on the CG model of the protein. The decision as to which to use depends on whether or not one wishes to include any protein conformational changes that may have occurred during the CG-MD simulation. Thus, CG2AT-full would be used if one wished to carry over a (limited) protein conformational change occurring in the protein during the CG-MD step, e.g., due to the interaction of a nonmembrane domain with lipid headgroups. In contrast, one might better use CG2AT-align if CG-MD simulations were being used for a relatively rigid membrane protein simply to establish an optimal lipid bilayer environment as a starting point for extensive AT-MD simulations. Ultimately, which approach to use depends on the nature of the specific questions being addressed in a given simulation study.

In CG2AT-full the backbone of the protein is grown from the CG $\text{C}\alpha$ particles using the Pulchra algorithm.³¹ To guide the reconstruction of the aromatic side chains, fragments of ring structures are first aligned to the CG particles. The missing side chain atoms are then added using the complete_pdb function in Modeller,³⁰ this method preserves the original coordinates from the CG particles (Supporting Information, Figure S1). Subsequent conjugate gradients energy minimization is then applied using Modeller to reduce the internal steric clashes of the model. The protein is then energy minimized further through 500 steps of steepest descents using Gromacs. The stereochemical quality of the generated models was evaluated using Procheck.³²

In CG2AT-align the original atomistic structure used as the starting point for the CG simulations is structurally aligned with the protein CG particles and then energy minimized using 500 steps of steepest descents. There is also an option to align based only on the transmembrane region of the protein (selected using a consensus of sequence-based TM helix predictions).³³ This latter option allows for an improved alignment of the transmembrane region of the protein, especially for proteins with large, mobile nonmembrane domains.

Lipid Conversion. Lipids were converted by alignment of atomistic lipid fragments (Figure 2) to the CG particles of each lipid molecule. This is repeated for all lipids in the system, which is then subjected to 5000 steps of steepest descents energy minimization. For PG and PS containing lipids, a further limited

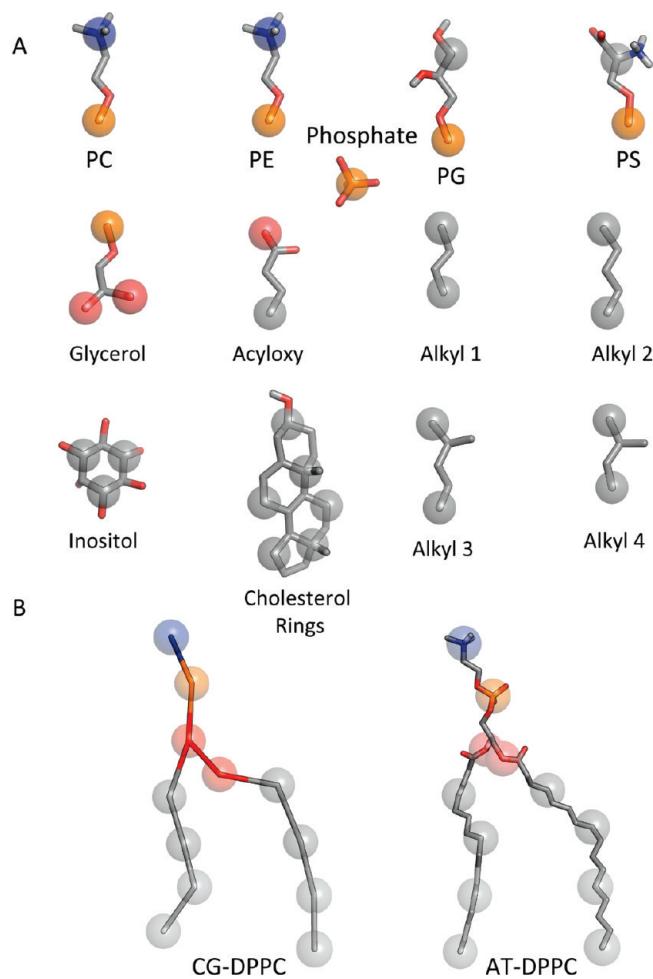


Figure 2. (A) Library of lipid fragments. Atomistic fragments used for the conversion of the lipid molecules. The atoms that are aligned with the CG particles are shown in spheres, with the remainder of the atoms shown as sticks. (B) An example of the lipid conversion is shown for DPPC.

memory Broyden–Fletcher–Goldfarb–Shanno (l-bfgs) energy minimization was required to reduce steric clashes between the headgroups of neighboring lipid molecules. Lipids currently available include POPC, POPE, POPG, POPS, DPPC, DHPC, DMPC, DMPG, DOPC, DSPE, BNG (β -nonyl glucoside), cholesterol, PIP₂, and PIP₃.

Final Steps. Following conversion of the protein and lipids the two energy minimized components were then combined and energy minimized further. In the case of protein converted using CG2AT-align, lipids within 1 Å of the protein were deleted to remove any unavoidable protein/lipid clashes. The system is then solvated, with any waters sitting within the hydrophobic core of the bilayer removed. Counterions are added to neutralize the system. The solvated system was finally energy minimized in preparation for the MD simulations.

Atomistic MD Simulations. Atomistic MD simulations were performed using Gromacs v4.5.3 with the GROMOS96 43a2 force field.³⁴ Simulations were performed using semi-isotropic pressure coupling with the Parrinello–Rahman barostat,³⁵ while the temperature of the lipid, protein, and solvent (water and counterions) was separately coupled to an external bath held at 323 K, using the Berendsen thermostat.³⁶ The water model used

was SPC.³⁷ The LINCS algorithm was used to constrain bond lengths.³⁸ Long-range electrostatic interactions beyond 10 Å were modeled using the particle mesh Ewald (PME) method.³⁹ A cutoff of 10 Å was used for van der Waals' interactions. Each converted system was first subjected to 1 ns of protein-restrained simulation, during which all heavy atoms of the protein were harmonically restrained with a force constant of 1000 kJ/mol/nm³. These restraints were then removed for 50 ns of production simulation with coordinates saved every 10 ps for analysis.

Pure Lipid Bilayer Simulations. In addition to studying the conversion of membrane protein complexes we also assessed this methodology with pure lipid bilayers, consisting of either POPC, POPE, DOPC, DPPC, or DMPC. Each bilayer was self-assembled by a 500 ns CG-MD simulation, before conversion to an atomistic systems, and used to start a 10 ns AT-MD simulation. The final snapshot of these simulations, along with the parameters used, can be found in Lipidbook (<http://lipidbook.bioch.ox.ac.uk/>).⁴⁰

RESULTS

Example: An Ion Channel Protein. The method is best illustrated by following in detail a specific example. For this we have selected the ASIC protein. This is an acid-sensing ion channel the structure of which has been determined at 1.9 Å resolution.⁴¹ It is trimeric, with a transmembrane (TM) domain containing six helices and an extensive extracellular domain. It is therefore a good example of a moderately complex membrane protein. It has been the subject of some simulation studies.⁴² The conversion process is illustrated in Figure 3. It can be seen that the CG-MD self-assembly process ‘correctly’ inserts the protein in a bilayer, i.e., with presumed TM domain in a bilayer spanning orientation and with the extracellular domain making few contacts with the lipids.

ASIC is of interest in that the trimer is asymmetric. This is reflected in the asymmetric, tilted orientation of the trimer relative to the bilayer. This is seen in the CG-MD simulation and increases during the 50 ns AT-MD simulation (Figure 4A and B). The orientation during the AT-MD simulation may be compared with the orientation predicted in the OPM database (<http://opm.phar.umich.edu/>), which treats the bilayer via an implicit bilayer method.¹³ Both the simulations and the implicit bilayer method suggest that the TM domain of ASIC is tilted (Figure 4). However the degree to which the protein changes over the course of the atomistic simulations, as the ectodomain interacts with the interfacial region of the bilayer. Thus for the three subunits, considering the TM segments only, the simulation-based tilt angles (relative to the bilayer normal, averaged over the last 10 ns of the two AT-MD simulations) are: subunit A, ~19°; subunit B, ~60°; and subunit C, ~48°. For OPM the corresponding tilt angles are: 2°, 25°, and 25°. In either case, the tilt is greater than that which would be assumed by simple ‘manual’ positioning of the ASIC protein in a lipid bilayer, guided by inspection transmembrane regions of the X-ray structure (Figure 4D). This difference in tilt angle is likely to be important in functional predictions of, e.g., the electrostatic potential surface around the protein where it is embedded in the bilayer (see e.g. ref 42).

We can compare the lipid headgroup contacts predicted by CG-MD and those maintained in the AT-MD simulation of ASIC. It can be seen there is an excellent agreement between the lipid contacts suggested by the CG simulations and those

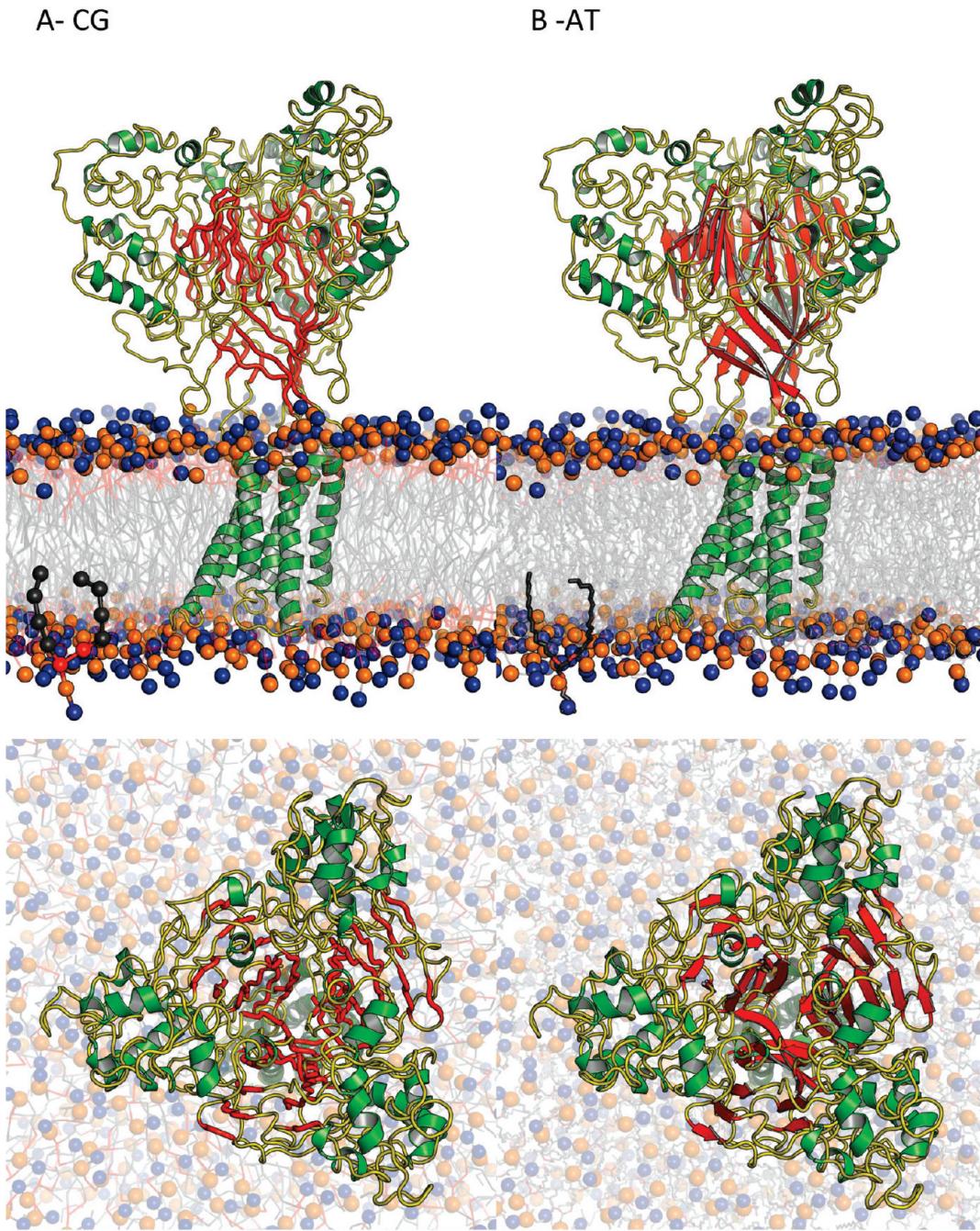


Figure 3. Visual comparison between CG and AT representations. A comparison between the (A, C) CG and (B, D) atomistic system of ASIC shown perpendicular to (top) and down (bottom) the bilayer normal.

preserved after CG-to-AT conversion and AT-MD simulation (Figure 5), regardless of whether the conversion used the ‘full’ or ‘aligned’ procedure for the protein (see above and Figure 1). This is reflected in correlation coefficients of between 0.79 to 0.81 for the lipid contacts seen in the three simulations.

Benchmark Systems. We employed a benchmark set of 10 membrane proteins against which to evaluate the CG2AT procedure and the behavior and lipid interactions of the proteins in the subsequent short AT-MD simulations (Figure 6). These 10 systems were selected to span a range of membrane proteins, simple and complex, with different overall architectures and differing patterns of interaction with lipids. Thus, there are

two relatively simple integral membrane proteins—for which most of the protein mass is α -helical and is located in the bilayer—namely LeuT and an aquaporin. There are three proteins with extensive extracellular (ELIC, ASIC) or equivalent (Cyt Ox) domains and three proteins with extensive intracellular domains (KcsA, SERCA and β_2 AdR/lysozyme). β_2 AdR/lysozyme is of especial interest as it is an artificial chimeric construct in which lysozyme has been inserted into an intracellular loop of a GPCR protein. The two other major classes of membranes proteins—outer membrane β -barrels and monotopic membrane proteins—are represented by OmpC and OSC, respectively.

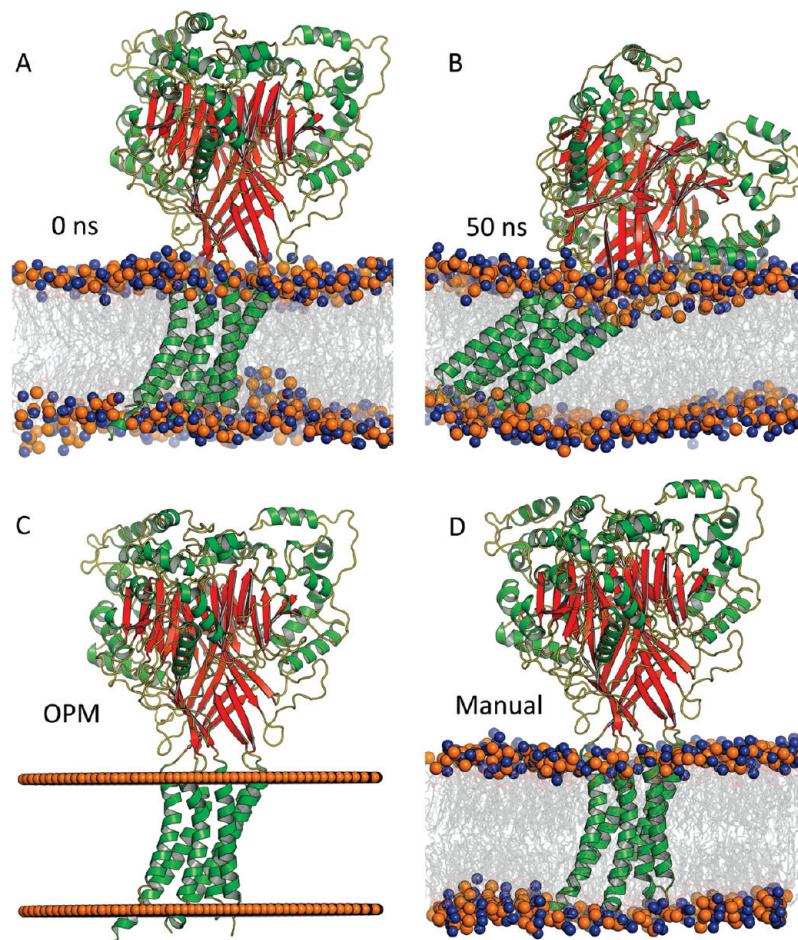


Figure 4. ASIC AT-MD simulations. Snapshots of ASIC in a lipid bilayer at the start (A) and end (B) of the AT-MD simulation from the CG2AT-align conversion. (C) ASIC oriented relative to the bilayer plane as predicted by OPM¹³ (<http://opm.phar.umich.edu/>). (D) ASIC oriented ‘manually’ in a bilayer based on helix locations in the X-ray structure and the sequence-based prediction of TM helices.

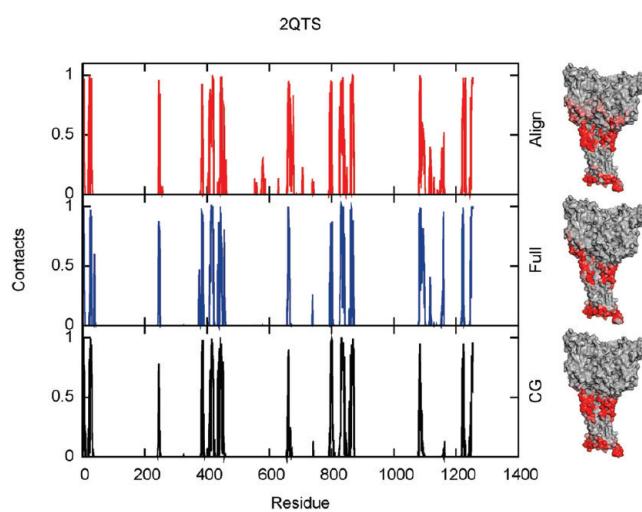


Figure 5. ASIC lipid/protein interactions. (A) CG and AT structures of ASIC color coded according to the frequency of lipid headgroup contacts averaged over the respective simulations; blue = no contacts made. (B) Lipid headgroup contact frequencies (as a fraction of simulation time) as a function of residue number for the CG-MD and the two AT-MD simulations of ASIC.

Behavior of the Proteins. One may compare the protein structures in terms of $C\alpha$ root-mean-square deviations (rmsds) before and after the CG2AT conversion. Unsurprisingly, the $C\alpha$ rmsd is small when the protein structure converted using Pulchra and Modeler, with an average value of 1.3 Å. For the CG2AT-align protocol the rmsd fit between the CG protein, and the overlaid X-ray structure is lower, with an average of 2.5 Å across all of the test set of proteins. This reflects a degree of (local) conformational change from the initial structures in the CG simulations, largely determined by the elastic network restraints used to model the protein tertiary structures in the CG method.

One may also monitor the degree of conformational drift over the course of the 50 ns AT-MD simulations. All simulations have a relatively low $C\alpha$ rmsd drift, especially for the transmembrane domains (Figure 7). Nevertheless, the overall degree of conformational drift is somewhat lower (by ca. 0.5 Å on average) for the simulations starting from the X-ray protein coordinates (i.e., from CG2AT-align) than those starting with remodeled protein coordinates (i.e., from CG2AT-full). An interesting outlier in terms of $C\alpha$ rmsd is the β_2 AdR/lysosome chimeric protein. For this the rmsd is much lower if one considers only the β_2 -adrenoceptor domain within the chimeric construct.

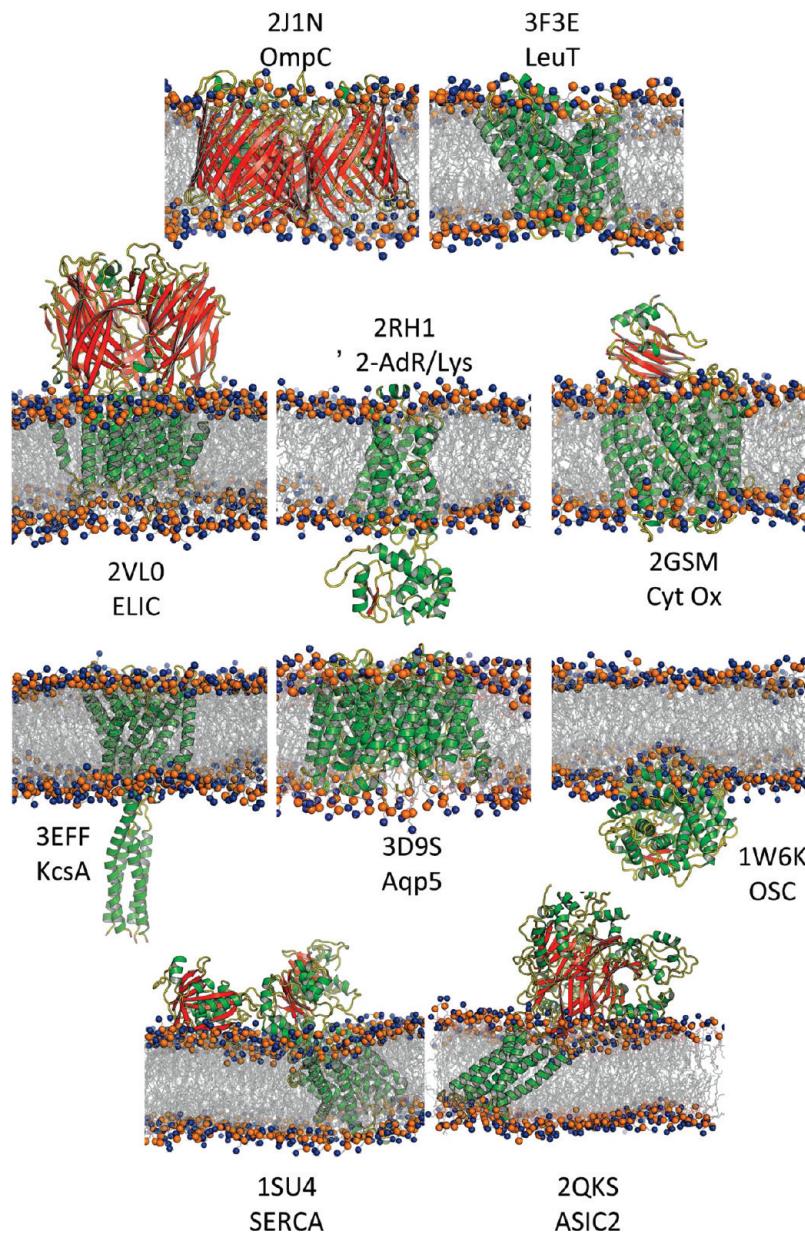


Figure 6. Benchmark systems. The 10 systems studied, from the AT-MD simulations. In each case the protein is shown by a ribbon, colored by secondary structure; green, α -helices; red, β -strands; and yellow, loops. The phosphorus (orange) and nitrogen (blue) atoms of the DPPC lipids are shown as spheres, to indicate the position of the bilayer.

Behavior of the Lipids. The average rmsd between CG and atomistic lipids was 1 Å, when aligning the atomic coordinates with their counterpart CG particle. One of the main benefits of the CGMD approach for simulating bilayer arrangement around a protein is it can allow for any local or global membrane deformation by the protein. The converted atomistic system (from CG2AT-full) retains the bilayer distortions of the CG simulation, allowing tight packing around the membrane protein. In contrast, in the CG2AT-align method on average 25 lipid molecules were deleted to remove any protein/lipid clashes. Although the deletion of the lipids removes close contacts between lipid and protein, these contacts are regained to a certain extent in the initial 1 ns equilibration step of the subsequent AT-MD simulation during which the protein is restrained. Such a lipid deletion step can be avoided by

reducing the protein flexibility during the CG-MD bilayer self-assembly step by, e.g., increasing the CG protein elastic network model¹⁴³ cutoff to ≥ 10 Å. This allows for better retention of the key protein/lipid interactions predicted by the CG-MD simulations while maintaining the X-ray structure of the protein.

We calculated the lipid tail order parameter (S_{CD} ; Figure 8) profiles from the AT-MD simulations both for pure lipid bilayers generated by CG2AT and for the two sets (align and full) of lipid/protein CG2AT systems. Those for the lipid only simulations closely resemble those from standard atomistic simulations of lipid bilayers thus suggesting that the fragment-based conversion methods from CG lipids preserves a stable fluid-phase lipid bilayer. The order parameter curves for proteins are very similar whichever CG2AT method is used for the protein and in

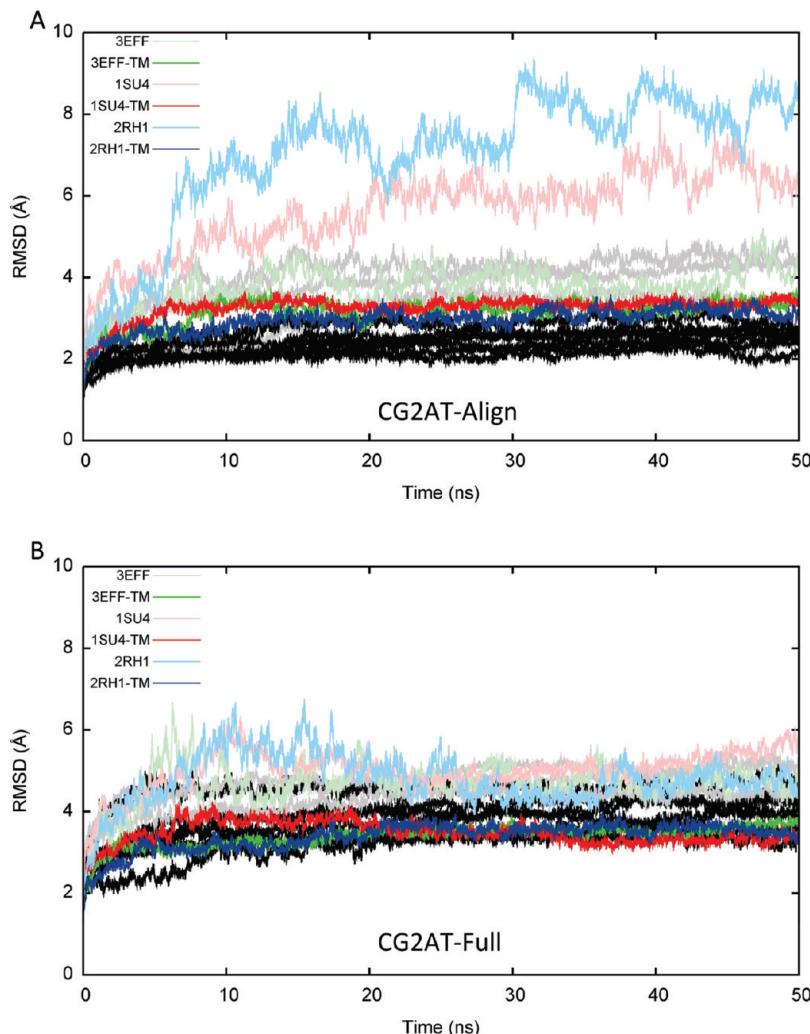


Figure 7. Protein conformational drift for the AT-MD simulations. The $\text{C}\alpha$ rmsd as a function of time the AT-MD simulations of the 10 systems are shown, using either the (A) CG2AT-align or (B) CG2AT-full method for protein conversion. The rmsds for the complete proteins are shown in gray, and those for the transmembrane domains only are in black. The β_2 -adrenoceptor structure (2RH1) shows the greatest drift from the starting structure (light blue), when the entire engineered complex is considered. Nevertheless, if one considers only the GPCR component, the structure is very stable (red). The SERCA structure (1SU4; entire structure, pink; TM only, red) and full-length KcsA (3EFF; entire structure, light green; TM only, green) also shown to be relatively dynamic structures in the non-TM regions.

general suggest a degree of (local) increase in lipid tail order by the proteins, as might be anticipated.

Lipid/Protein Interactions. The lipid headgroups contacting the proteins were analyzed as a function of protein residue number in a similar fashion to that for ASIC (see above). It is evident that there are high-correlation coefficients between the contacts observed in the three simulations (Table 1). Therefore we may conclude that the CG2AT conversion procedure preserves the key lipid/protein contacts generated by self-assembly CG-MD and that these contacts remain, at least over the short (50 ns) duration of the AT-MD simulations.

We also analyzed the frequency distributions along the bilayer normal (z) of amino acid residue types that make contacts with the lipid molecules in the AT-MD simulations for at least 30% of the time (Supporting Information, Figure S2). These distributions show the same general patterns as seen in CG-MD simulations of a wide range of membrane proteins.¹⁶ Thus amphipathic aromatic residues (e.g., Trp and Tyr) are located in the lipid/water interfacial regions, as are Arg and Lys side chains (but

at higher somewhat $|z|$ values allowing interactions with lipid phosphate groups), whereas hydrophobic side chains are preferentially localized in the bilayer core. This further demonstrates that the CG2AT conversion has preserved the positions and the interactions of lipid-exposed amino acids on the surface of the TM domains of membrane proteins.

■ DISCUSSION

We have described a method for accurate and automatic conversion of a CG description of a complex membrane/protein system to atomistic detail, enabling subsequent AT-MD simulation. This method has been evaluated against a test set of 10 membrane proteins and shown to provide stable AT-MD simulation systems which in turn provide information on, e.g., protein/lipid interactions. Thus, we have provided a protocol for serial multiscale (as defined by Voth and colleagues)²³ simulation of membrane proteins. This combines efficient sampling of

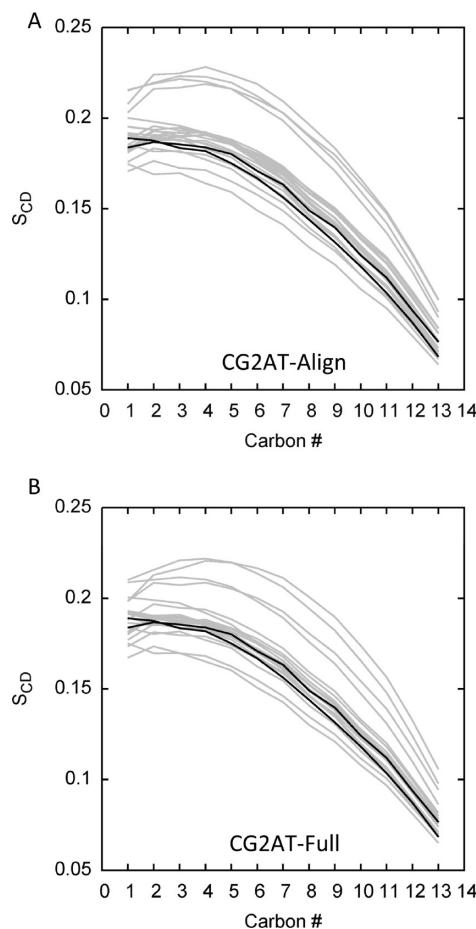


Figure 8. Lipid order parameters. The lipid order parameters (S_{CD}) from a 50 ns simulation of pure DPPC lipid bilayers (black) are compared to the order parameters derived from the systems containing protein (gray) that have been converted either through (A) structural alignment or (B) the full conversion method.

lipid/protein interactions in CG-MD simulations with refinement of these interactions on a shorter time scale by AT-MD.

Currently there are relatively few protocols available for preparing complex, mixed lipid systems around a membrane protein. OPM^{13,44} and related prediction methods indicate the regions of the protein that will interact with lipid, and implicit bilayer models^{45–47} can provide efficient sampling in a bilayer-like environment and allow use enhanced sampling methods, such as replica exchange.⁴⁸ However, particle-based approximations (either CG-MD as in the current study, or DPD as in, e.g., refs 49–52), are needed to provide models of more specific lipid/protein interactions, including local distortions of bilayer thickness and/or selective interactions with lipid headgroups. By combining such approaches with CG2AT, it is possible to initiate detailed atomistic simulations to refine the resultant models of such interactions.

We have described two variants on the CG2AT method, differing in whether or not changes in protein conformation from CG-MD are retained. The CG2AT-align method should be used where preserving the exact starting structure of the protein is important for the subsequent atomistic molecular simulations, e.g., if known ligand binding sites need to be accurately retained. Additionally, in order to retain the local protein/lipid

Table 1. Lipid/Protein Contacts: Correlations Between Simulations^a

protein	PDB id	correlation coefficient		
		CG vs AT-full	AT-align	AT-align
SERCA	1SU4	0.73	0.71	0.78
OSC	1W6K	0.66	0.53	0.67
Cyt Ox	2GSM	0.79	0.78	0.81
OmpC	2J1N	0.80	0.85	0.86
ASIC	2QTS	0.81	0.77	0.78
B ₂ AdR/Lys	2RH1	0.81	0.84	0.82
ELIC	2VL0	0.77	0.80	0.76
AqpS	3D9S	0.80	0.79	0.76
KcsA	3EFF	0.70	0.54	0.59
LeuT	3F3E	0.77	0.78	0.81

^a These proteins include calcium ATPase (SERCA) (1SU4),⁶¹ oxidosqualene cyclase (OSC) (PDB id: 1W6K),⁶² cytochrome c oxidase (2GSM),⁶³ outer membrane protease C (OmpC) (2J1N),⁶⁴ acid-sensing ion channel 2 (ASIC) (2QTS),⁴¹ β -2 adrenoreceptor with lysozyme (β 2-Adr/lys) (2RH1),⁶⁵ ELIC pentameric ion channel (2VL0),⁶⁶ aquaporin P5 (AqpS) (3D9S),⁶⁷ KcsA potassium channel (3EFF),⁶⁸ and leucine transporter (LeuT) (3F3E).⁶⁹

contacts, the CG-MD simulations may be performed with tighter elastic network model⁴³ restraints imposed on the protein structure to prevent any major conformational changes and therefore avoid the need for lipid deletion upon conversion. In addition, the GROMACS g_membed protocol⁵³ has been included in more recent versions of the conversion process.

The major strength of our approach is that it enables detailed studies of lipid/protein interactions by providing an optimal orientation of a complex membrane protein in a bilayer for subsequent AT-MD simulations. This is illustrated here for the ion channel ASIC, enabling comparison of the multiscale simulation approach with other methods (e.g., ‘manual’ insertion or use of the OPM model) for prediction of how the protein might be arranged relative to a bilayer and indicating significant differences in the final result. In the current paper we have used a relatively simple lipid bilayer (DPPC), but recent studies have shown that a similar approach may be applied to more complex protein/lipid systems, e.g., models of Kir channels and their possible interactions with PIP₂.⁵⁴

On balance we think that our method for CG2AT conversion performs well for membrane protein systems, allowing a balance between utility and accuracy to be achieved in the context of a given biological problem. Thus the current method, based on fragment assembly, provides a possible alternative to other methods that have used simulated annealing⁵⁵ or force matching.⁵⁶ It also appears to be a less time-intensive mechanism for conversion that that used previously in our laboratory.⁵⁷ However, the current approach does not balance the energetics of both levels of granularity. An alternative approach would be to use mixed resolution methods which in principle could achieve detailed balance.^{58,59} However, at present such methods are likely to be difficult to applying to complex multicomponent membrane systems.

It should also be noted that we do not attempt to convert the water particles from the CG simulation, as these are somewhat

simplified in MARTINI and related CG models. However, it should be possible to include such a conversion if appropriate, e.g., if polarizable CG water models⁶⁰ are used.

As the number of membrane protein structures determined by X-ray crystallography and other high-resolution methods expands,³ there is an increasing need for multiscale simulations of membrane proteins. Building upon our earlier database of CG simulations (<http://sbcb.bioch.ox.ac.uk/cgdb/>),^{16,26} it should now be possible to provide multiscale simulations of all membrane proteins as their structures are determined. This in turn will enable structural bioinformatics studies, such as data mining of membrane protein/lipid interactions.

■ ASSOCIATED CONTENT

S Supporting Information. The CG particle to atomistic mapping for the protein sidechains for (A) our in house CG protein parameters and (B) Martini. Distribution along the bilayer normal of amino acids making contact with the bilayer over the course of the two sets of twenty 10 ns atomistic simulations of the membrane protein complexes. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: mark.sansom@bioch.ox.ac.uk.

■ ACKNOWLEDGMENT

We thank members of MSPS laboratory for helpful discussions. Research in MSPS laboratory is supported by the Wellcome Trust and BBSRC.

■ REFERENCES

- Wallin, E.; von Heijne, G. Genome-wide analysis of integral membrane proteins from eubacterial, archean, and eukaryotic organisms. *Protein Sci.* **1998**, *7*, 1029–1038.
- Terstappen, G. C.; Reggiani, A. In silico research in drug discovery. *Trends Pharmacol. Sci.* **2001**, *22*, 23–26.
- White, S. H. Biophysical dissection of membrane proteins. *Nature* **2009**, *459*, 344–6.
- Gonen, T.; Cheng, Y.; Sliz, P.; Hiroaki, Y.; Fujiyoshi, Y.; Harrison, S. C.; Walz, T. Lipid–protein interactions in double-layered two-dimensional AQP0 crystals. *Nature* **2005**, *438*, 633–638.
- Hite, R. K.; Li, Z. L.; Walz, T. Principles of membrane protein interactions with annular lipids deduced from aquaporin-0 2D crystals. *EMBO J.* **29**, 1652–1658.
- Powl, A. M.; Wright, J. N.; East, J. M.; Lee, A. G. Identification of the hydrophobic thickness of a membrane protein using fluorescence spectroscopy: Studies with the mechanosensitive channel MscL. *Biochemistry* **2005**, *44*, 5713–5721.
- Anbazhagan, V.; Qu, J.; Kleinschmidt, J. H.; Marsh, D. Incorporation of outer membrane protein OmpG in lipid membranes: Protein-lipid interactions and beta-barrel orientation. *Biochemistry* **2008**, *47*, 6189–6198.
- Gold, V. A. M.; Robson, A.; Bao, H.; Romantsov, T.; Duong, F.; Collinson, I. The action of cardiolipin on the bacterial translocon. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 10044–10049.
- Lindahl, E.; Sansom, M. S. P. Membrane proteins: molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **2008**, *18*, 425–431.
- Niemela, P. S.; Miettinen, M. S.; Monticelli, L.; Hammaren, H.; Bjelkmar, P.; Murtola, T.; E., L.; Vattulainen, I. Membrane proteins diffuse as dynamic complexes with lipids. *J. Am. Chem. Soc.* **2010**, *132*, 7574–7575.
- Faraldo-Gómez, J. D.; Smith, G. R.; Sansom, M. S. P. Setup and optimization of membrane protein simulations. *Eur. Biophys. J.* **2002**, *31*, 217–227.
- Killian, J. A.; von Heijne, G. How proteins adapt to a membrane-water interface. *Trends Biochem. Sci.* **2000**, *25*, 429–434.
- Lomize, M. A.; Lomize, A. L.; Pogozheva, I. D.; Mosberg, H. I. OPM: Orientations of proteins in membranes database. *Bioinformatics* **2006**, *22*, 623–625.
- Voth, G. A. *Coarse-Graining of Condensed Phase and Biomolecular Systems*; CRC Press: Boca Raton, FL, 2008.
- Periole, X.; Huber, T.; Marrink, S. J.; Sakmar, T. P. G protein-coupled receptors self-assemble in dynamics simulations of model bilayers. *J. Am. Chem. Soc.* **2007**, *129*, 10126–10132.
- Scott, K. A.; Bond, P. J.; Ivetac, A.; Chetwynd, A. P.; Khalid, S.; Sansom, M. S. P.; Coarse-grained, M. D. simulations of membrane protein-bilayer self-assembly. *Structure* **2008**, *16*, 621–630.
- Vorobyov, I.; Li, L.; Allen, T. W. Assessing atomistic and coarse-grained force fields for protein-lipid interactions: the formidable challenge of an ionizable side chain in a membrane. *J. Phys. Chem. B* **2008**, *112*, 9588–9602.
- Bond, P. J.; Wee, C. L.; Sansom, M. S. P. Coarse-grained molecular dynamics simulations of the energetics of helix insertion into a lipid bilayer. *Biochemistry* **2008**, *47*, 11321–11331.
- Shi, Q.; Izvekov, S.; Voth, G. A. Mixed atomistic and coarse-grained molecular dynamics: simulation of a membrane bound ion channel. *J. Phys. Chem. B* **2006**, *110*, 15045–15048.
- Lu, L. Y.; Izvekov, S.; Das, A.; Andersen, H. C.; Voth, G. A. Efficient, regularized, and scalable algorithms for multiscale coarse-graining. *J. Chem. Theory Comp.* **2010**, *6*, 954–965.
- Orsi, M.; Sanderson, W. E.; Essex, J. W. Permeability of small molecules through a lipid bilayer: a multiscale simulation study. *J. Phys. Chem. B* **2009**, *113*, 12019–12029.
- Orsi, M.; Essex, J. W. Permeability of drugs and hormones through a lipid bilayer: insights from dual-resolution molecular dynamics. *Soft Matter* **2010**, *6*, 3797–3808.
- Ayton, G. A.; Noid, W. G.; Voth, G. A. Multiscale modeling of biomolecular systems: in serial and in parallel. *Curr. Opin. Struct. Biol.* **2007**, *17*, 192–198.
- Thogersen, L.; Schiott, B.; Vosegaard, T.; Nielsen, N. C.; Tajkhorshid, E. Peptide aggregation and pore formation in a lipid bilayer - a combined coarse grained and all atom molecular dynamics study. *Biophys. J.* **2008**, *95*, 4337–4347.
- Rzepiela, A. J.; Schafer, L. V.; Goga, N.; Risselada, H. J.; De Vries, A. H.; Marrink, S. J. Reconstruction of atomistic details from coarse-grained structures. *J. Comput. Chem.* **2010**, *31*, 1333–1343.
- Chetwynd, A. P.; Scott, K. A.; Mokrab, Y.; Sansom, M. S. P. CGDB: a database of membrane protein/lipid interactions by coarse-grained molecular dynamics simulations. *Mol. Membr. Biol.* **2008**, *25*, 662–669.
- van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. GROMACS: fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
- Marrink, S. J.; Risselada, J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI forcefield: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.
- Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J. The MARTINI coarse grained force field: extension to proteins. *J. Chem. Theor. Comp.* **2008**, *4*, 819–834.
- Sali, A.; Blundell, T. L. Comparative protein modeling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779–815.
- Rotkiewicz, P.; Skolnick, J. Fast procedure for reconstruction of full-atom protein models from reduced representations. *J. Comput. Chem.* **2008**, *29*, 1460–5.
- Laskowski, R. A.; Macarthur, M. W.; Moss, D. S.; Thornton, J. M. Procheck - a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283–291.

- (33) Cuthbertson, J. M.; Doyle, D. A.; Sansom, M. S. P. Transmembrane helix prediction: a comparative evaluation and analysis. *Prot. Eng. Des. Sel.* **2005**, *18*, 295–308.
- (34) Scott, W. R. P.; Hunenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Kruger, P.; van Gunsteren, W. F. The GROMOS biomolecular simulation program package. *J. Phys. Chem. A* **1999**, *103*, 3596–3607.
- (35) Parrinello, M.; Rahman, A. Polymorphic transitions in single-crystals - a new molecular-dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (36) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (37) Hermans, J.; Berendsen, H. J. C.; van Gunsteren, W. F.; Postma, J. P. M. A consistent empirical potential for water-protein interactions. *Biopolymers* **1984**, *23*, 1513–1518.
- (38) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (39) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald - an N. log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (40) Domański, J.; Stansfeld, P. J.; Sansom, M. S. P.; Beckstein, O. Lipidbook - a public repository for force field parameters used in membrane simulations. *J. Membr. Biol.* **2010**, *236*, 255–258.
- (41) Jasti, J.; Furukawa, H.; Gonzales, E. B.; Gouaux, E. Structure of acid-sensing ion channel 1 at 1.9 Å resolution and low pH. *Nature* **2007**, *449*, 316–23.
- (42) Shaikh, S. A.; Tajkhorshid, E. Potential cation and H⁺ binding sites in acid sensing ion channel-1. *Biophys. J.* **2008**, *95*, 5153–5164.
- (43) Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* **2001**, *80*, 505–515.
- (44) Lomize, A. L.; Pogozheva, I. D.; Lomize, M. A.; Mosberg, H. I. Positioning of proteins in membranes: A computational approach. *Protein Sci.* **2006**, *15*, 1318–1333.
- (45) Ulmschneider, M. B.; Sansom, M. S. P.; Di Nola, A. Properties of integral membrane protein structures: derivation of an implicit membrane potential. *Proteins: Struct., Funct., Bioinf.* **2005**, *59*, 252–265.
- (46) Ulmschneider, M. B.; Ulmschneider, J. P.; Sansom, M. S. P.; Di Nola, A. A generalized Born implicit membrane representation based on experimental insertion free energies. *Biophys. J.* **2007**, *92*, 2338–2349.
- (47) Tanizaki, S.; Feig, M. Molecular dynamics simulations of large integral membrane proteins with an implicit membrane model. *J. Phys. Chem. B* **2006**, *110*, 548–556.
- (48) Sayadi, M.; Tanizaki, S.; Feig, M. Effect of membrane thickness on conformational sampling of phospholamban from computer simulations. *Biophys. J.* **2010**, *98*, 805–814.
- (49) de Meyer, F. J. M.; Venturoli, M.; Smit, B. Molecular simulations of lipid-mediated protein-protein interactions. *Biophys. J.* **2008**, *95*, 1851–1865.
- (50) de Meyer, F. J. M.; Rodgers, J. M.; Willem, T. F.; Smit, B. Molecular simulation of the effect of cholesterol on lipid-mediated protein-protein interactions. *Biophys. J.* **2010**, *99*, 3629–3638.
- (51) Schmidt, U.; Guigas, G.; Weiss, M. Cluster formation of transmembrane proteins due to hydrophobic matching. *Phys. Rev. Lett.* **2008**, *101*, 128104.
- (52) Schmidt, U.; Weiss, M. Hydrophobic mismatch-induced clustering as a primer for protein sorting in the secretory pathway. *Biophys. Chem.* **2010**, *151*, 34–38.
- (53) Wolf, M. G.; Hoefling, M.; Aponte-Santamaría, C.; Grubmüller, H.; Groenhof, G. g_membed: Efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation. *J. Comput. Chem.* **2010**, *31*, 2169–2174.
- (54) Stansfeld, P. J.; Hopkinson, R. J.; Ashcroft, F. M.; Sansom, M. S. P. The PIP₂ binding site in Kir channels: definition by multi-scale biomolecular simulations. *Biochemistry* **2009**, *48*, 10926–10933.
- (55) Rzepiela, A. J.; Schäfer, L. V.; Goga, N.; Risselada, H. J.; de Vries, A. H.; Marrink, S. J. Reconstruction of atomistic details from coarse grained structures. *J. Comput. Chem.* **2010**, *31*, 1333–1343.
- (56) Izvekov, S.; Voth, G. A. A multiscale coarse-graining method for biomolecular systems. *J. Phys. Chem. B* **2005**, *109*, 2469–2473.
- (57) Carpenter, T.; Bond, P. J.; Khalid, S.; Sansom, M. S. P. Self-assembly of a simple membrane protein: coarse-grained molecular dynamics simulations of the influenza M2 channel. *Biophys. J.* **2008**, *95*, 3790–3801.
- (58) Liu, P.; Voth, G. A. Smart resolution exchange: an efficient algorithm for exploring complex energy landscapes. *J. Chem. Phys.* **2007**, *126* (045106), 1–6.
- (59) Liu, P.; Shi, Q.; Lyman, E.; Voth, G. A. Reconstructing atomistic detail from coarse-grained models with resolution exchange. *J. Chem. Phys.* **2008**, *129* (114103), 1–8.
- (60) Yesylevskyy, S. O.; Schäfer, L. V.; Sengupta, D.; Marrink, S. J. Polarizable water model for the coarse-grained MARTINI force field. *PLoS Comp. Biol.* **2010**, *6*, e1000810.
- (61) Toyoshima, C.; Nakasako, M.; Nomura, H.; Ogawa, H. Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature* **2000**, *405*, 647–655.
- (62) Thoma, R.; Schulz-Gasch, T.; D'Arcy, B.; Benz, J.; Aeby, J.; Dehmlow, H.; Hennig, M.; Stihle, M.; Ruf, A. Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase. *Nature* **2004**, *432*, 118–122.
- (63) Qin, L.; Hiser, C.; Mulichak, A.; Garavito, R. M.; Ferguson-Miller, S. Identification of conserved lipid/detergent-binding sites in a high-resolution structure of the membrane protein cytochrome c oxidase. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16117–22.
- (64) Basle, A.; Rummel, G.; Storici, P.; Rosenbusch, J. P.; Schirmer, T. Crystal structure of osmoporin OmpC from *E. coli* at 2.0 Å. *J. Mol. Biol.* **2006**, *362*, 933–942.
- (65) 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.; Stevens, R. C. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **2007**, *318*, 1258–65.
- (66) Hilf, R. J. C.; Dutzler, R. X-ray structure of a prokaryotic pentameric ligand-gated ion channel. *Nature* **2008**, *452*, 375–379.
- (67) Horsefield, R.; Norden, K.; Fellert, M.; Backmark, A.; Tornroth-Horsefield, S.; Terwisscha van Scheltinga, A. C.; Kvassman, J.; Kjellbom, P.; Johanson, U.; Neutze, R. High-resolution x-ray structure of human aquaporin 5. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 13327–32.
- (68) Uysal, S.; Vasquez, V.; Tereshko, V.; Esaki, K.; Fellouse, F. A.; Sidhu, S. S.; Koide, S.; Perozo, E.; Kossiakoff, A. Crystal structure of full-length KcsA in its closed conformation. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 6644–9.
- (69) Singh, S. K.; Piscitelli, C. L.; Yamashita, A.; Gouaux, E. A competitive inhibitor traps LeuT in an open-to-out conformation. *Science* **2008**, *322*, 1655–61.