Coarse-Grained MD Simulations of Membrane Protein-Bilayer Self-Assembly

Kathryn A. Scott,[1,2] Peter J. Bond,[1,3] Anthony Ivetac,[1,4] Alan P. Chetwynd,[1] Syma Khalid,[1,5] and Mark S.P. Sansom[1],*

[1]*Structural Bioinformatics and Computational Biochemistry Unit, Department of Biochemistry, University of Oxford,Oxford OX1 3QU, United Kingdom*
[2]*Division of Structural Biology, The Wellcome Trust Centre for Human Genetics, University of Oxford,Roosevelt Drive, Oxford OX3 7BN, United Kingdom.*
[3]*Theoretical Biophysics Group, Max Planck Institute of Biophysics, Max-von-Laue-Strasse 3,60438 Frankfurt am Main, Germany.*
[4]*Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla,CA 92093-0365, USA.*
[5]*School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom. S.Khalid@soton.ac.uk*

Please cite this paper as:
Structure, 2008, 621–630

The publisher's version of this paper is available here:
http://dx.doi.org/10.1016/j.str.2008.01.014

Structure
Article

Coarse-Grained MD Simulations of Membrane Protein-Bilayer Self-Assembly

Kathryn A. Scott,[1,2] Peter J. Bond,[1,3] Anthony Ivetac,[1,4] Alan P. Chetwynd,[1] Syma Khalid,[1,5] and Mark S.P. Sansom[1,*]
[1]Structural Bioinformatics and Computational Biochemistry Unit, Department of Biochemistry, University of Oxford, Oxford OX1 3QU, United Kingdom
[2]Present address: Division of Structural Biology, The Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, United Kingdom.
[3]Present address: Theoretical Biophysics Group, Max Planck Institute of Biophysics, Max-von-Laue-Strasse 3, 60438 Frankfurt am Main, Germany.
[4]Present address: Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0365, USA.
[5]Present address: School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom.
*Correspondence: mark.sansom@bioch.ox.ac.uk
DOI 10.1016/j.str.2008.01.014

SUMMARY

Complete determination of a membrane protein structure requires knowledge of the protein position within the lipid bilayer. As the number of determined structures of membrane proteins increases so does the need for computational methods which predict their position in the lipid bilayer. Here we present a coarse-grained molecular dynamics approach to lipid bilayer self-assembly around membrane proteins. We demonstrate that this method can be used to predict accurately the protein position in the bilayer for membrane proteins with a range of different sizes and architectures.

INTRODUCTION

The importance of membrane proteins is clear from their prevalence in the genome and their roles in cellular function. Despite their physiological and pharmaceutical importance, membrane proteins constitute <1% of known protein structures. For membrane proteins, interaction with lipids is essential for protein function; bilayer properties, such as hydrophobic thickness or lipid composition, can affect membrane protein activity (Hunte, 2005; Lee, 2004, 2005). Although often crystallized as membrane protein-detergent complexes, in most cases only a few tightly bound lipid molecules remain (Fyfe et al., 2001; Marsh and Pali, 2006; Palsdottir and Hunte, 2004). Thus, the crystal structure rarely contains explicit information on where the protein is located in the bilayer.

Experimental methods other than X-ray crystallography can be used to determine protein position in a lipid bilayer (Lee, 2005). Methods include site-directed spin labeling (Fanucci and Cafiso, 2006), cysteine scanning mutagenesis coupled with chemical modification (Guan and Kaback, 2006; Guan et al., 2002), tryptophan scanning mutagenesis with fluorescence spectroscopy (Powl et al., 2003, 2005), and two-dimensional infrared spectroscopy (Mukherjee et al., 2004, 2006). Although such methods can determine the local environment of a protein residue, they are not high-throughput approaches. Given the projected growth in the number of determined membrane protein structures (White, 2004; Lundstrom, 2006), a high-throughput approach is desirable.

Computational methods can be used to predict the position of proteins within the lipid bilayer. A number of methods which treat the bilayer as a hydrophobic "slab" can be used to position membrane peptides and proteins (Basyn et al., 2001, 2003; Ducarme et al., 1998; Im et al., 2003; Ulmschneider et al., 2005, 2006), and have been used, for example, in the Protein Data Bank of Transmembrane Proteins (Tusnady et al., 2004, 2005a, 2005b; http://pdbtm.enzim.hu/) and Orientations of Proteins in Membranes (Lomize et al., 2006a, 2006b; http://opm.phar.umich.edu/) databases. However, such approaches do not account for the complexity of the bilayer head group-water interface (White and Wimley, 1999). Our approach differs from those described above in that the bilayer is formed from discrete lipid molecules. We use a coarse-grained (CG) representation of both protein and lipid molecules (Bond et al., 2007; Marrink et al., 2004; Nielsen et al., 2004; Shelley et al., 2001) and an initial configuration where the protein is surrounded by randomly positioned lipids. During a molecular dynamics (MD) simulation, the bilayer self-assembles around the protein. Importantly, this method allows local structural changes in the region of the bilayer around the protein and can investigate changes in membrane protein positioning with bilayer composition. In this work, we present results for a representative set of 91 proteins embedded in a dipalmitoylphosphatidylcholine (DPPC) bilayer. We also briefly discuss extensions to the method in which different lipids are used. The result of this work is a set of predictions of the positioning of proteins in the membrane bilayer (available at http://sbcb.bioch.ox.ac.uk/cgdb/) and a set of Protein Data Bank (PDB) files for the CG representations that can be converted to atomistic models for later use in multiscale simulations (Shi et al., 2006).

RESULTS

The aim of this study was to use CG-MD self-assembly simulations to position each distinct membrane protein fold in a lipid

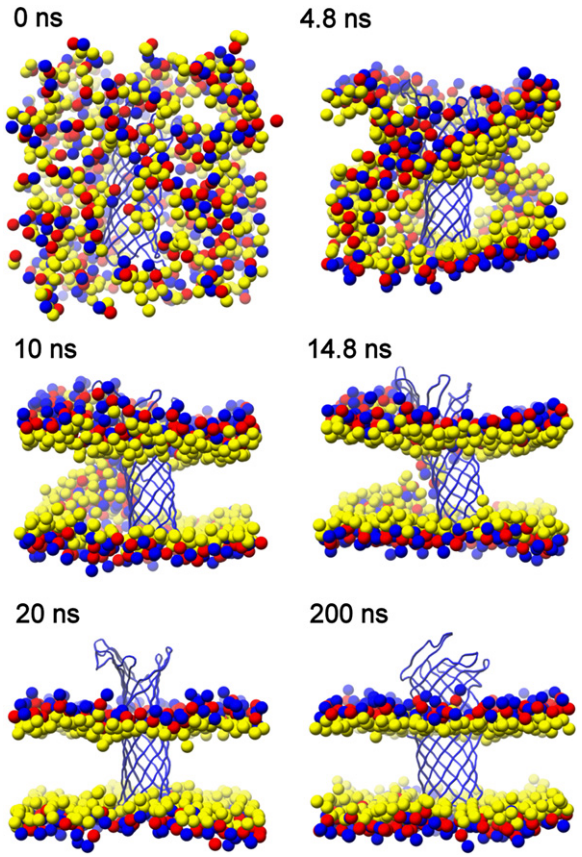

Figure 1. Snapshots of the Self-Assembly Process for DPPC and the Outer Membrane Adhesion Protein OpcA

Particles corresponding to the lipid phosphate group are colored red, the choline group is in blue, and the glycerol backbone is in yellow. The backbone trace of the protein is shown in blue. Water, ion, and lipid chain particles are excluded for clarity. The simulation begins with lipids randomly distributed. For proteins of less than ~200 residues with the majority of the protein in the membrane, bilayer formation is typically complete by ~20 ns. For larger proteins, particularly those with large extramembrane regions, bilayer formation and migration to the transmembrane region is on the order of 100 ns.

bilayer, enabling comparative analysis of the interactions of proteins with a membrane environment. The current data set contains 91 proteins (see the Supplemental Data available with this article online), of which 33 have β-barrel transmembrane (TM) regions and 58 have α-helical TM regions. (Note: we have excluded the voltage-dependent potassium channel KvAP, as the conformation of this protein when in a lipid bilayer remains controversial; Campos et al., 2007; Lee et al., 2005.) Two water-soluble proteins were tested as controls (the spectrin SH3 domain [PDB code: 1SHG], and the retinol-binding protein [PDB code: 1KT7]; ID codes are from the Protein Data Bank throughout) and did not insert into a lipid bilayer.

Snapshots from a typical self-assembly simulation (Figure 1; for OpcA; 1K24) show that the bilayer starts to form around the protein within ~10 ns, and by 20 ns the bilayer appears fully formed. Bilayer formation typically occurs within the first ~30 ns of simulation. Subsequently, only small changes in protein position relative to a bilayer occur. The simulations were run for 0.2 or 0.4 μs to allow equilibration of lipid-protein interactions.

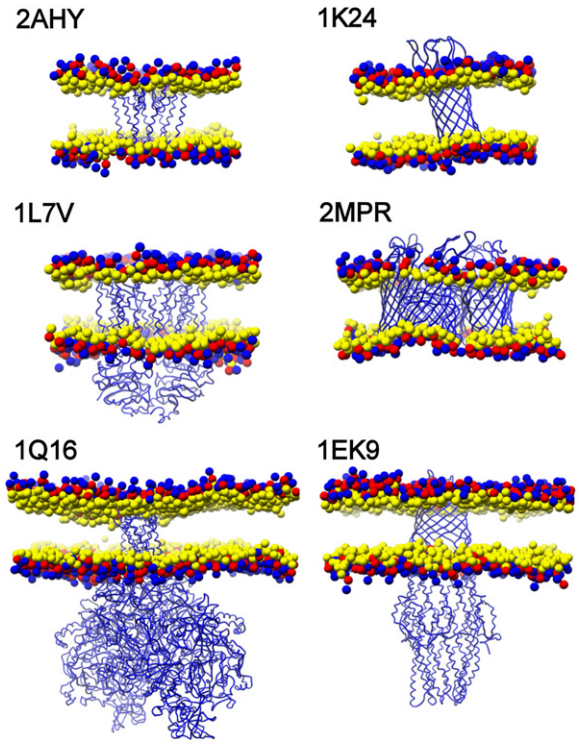

Figure 2. Snapshots of the Final Structure for a Representative Range of Protein Architectures

Particles corresponding to the lipid phosphate group are colored red, the choline group is in blue, and the glycerol backbone is in yellow. The backbone trace of the protein is shown in blue. Water, ion, and lipid chain particles are excluded for clarity. 2AHY is an NaK channel, 1K24 is the outer membrane adhesion protein OpcA, 1L7V is a bacterial ABC transporter, 2MPR is a maltoporin, 1Q16 is respiratory nitrate reductase, and 1EK9 is the TolC efflux pump. The proteins in the data set range from ~200 to ~4000 residues in size, the largest protein being a cytochrome bc1 complex with 3977 residues and a bilayer 200 Å in diameter.

In some cases, in particular for large extramembrane domains, a small number of lipids remained attached to the protein away from the bilayer surface. In curating the database, these lipids were replaced with CG water particles and a further 0.2 μs simulation was carried out to ensure that the stability of the system had not been perturbed. The last 100 ns of simulation were used for analysis; snapshots of the last frame in the simulation are shown for proteins with a range of different architectures and oligomeric states (Figure 2).

Structural Analysis

To assess the extent of structural change over the simulation, the root-mean-square-deviation (rmsd) of the C_α particles was calculated relative to the starting structure. In our CG model, backbone heavy atoms within 7 Å of each other are restrained by an elastic network (Atilgan et al., 2001). Nonprotein elements, such as chromophores, are omitted and represented by the elastic network. Even with these restraints it is possible for proteins to undergo conformational change, with different protein domains moving relative to each other. Histograms of the mean C_α-rmsd for all 91 membrane proteins in the data set (Figure 3) show that only for 5 proteins is the mean C_α-rmsd for residues

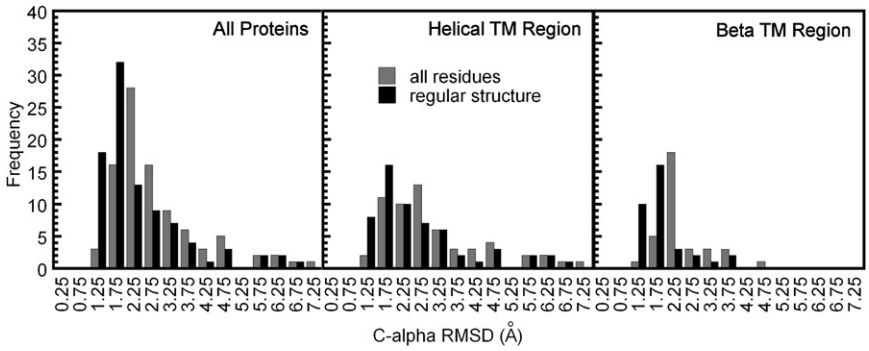

Figure 3. Histograms Showing the Mean C_α-Rmsd over the Last 100 ns of Simulation

C_α-rmsd is calculated relative to the starting structure. Data are shown for the whole data set and are categorized by protein transmembrane architecture; in each case, the C_α-rmsd over all residues and over those residues in regular secondary structure in the starting structure is calculated. The DSSP (Kabsch and Sander, 1983) algorithm implemented in GROMACS was used to determine which residues in the original structure were in regular secondary structure. A bin width of 0.5 Å was used and the median values are shown on the x axis. A higher proportion of proteins with β-barrel than with α-helical TM regions has a C_α-rmsd below 2.5 Å; this is to be expected from the larger number of tertiary structural restraints used in the simulation of the β proteins.

in regular secondary structure greater than 5 Å. One of these proteins is a sarcoplasmic/endoplasmic reticulum Ca^{2+}-ATPase (1SU4), where the high rmsd is due to changes in the relative position of the extramembrane domains. Two of the other proteins which exhibit a high rmsd are light-harvesting complexes (1LGH and 1NKZ). These proteins consist of rings of helices. Their high C_α-rmsd originates from ring distortion due to relative motion of the component helices, possibly demonstrating some limitations of representing the chromophores using the elastic network. The remaining proteins are a cytochrome bc1 complex (1EZV) and a multidrug ABC transporter homolog (Sav1866; 2HYD), both of which have voids between subunits which become smaller during bilayer formation. The majority of proteins undergo relatively small conformational changes during the simulation.

Comparison with Experimental Data

The position of the protein relative to the lipid bilayer has been determined using noncrystallographic experimental approaches for several proteins used in the self-assembly simulations. These include lactose permease LacY (1PV7), rhodopsin (1U19), the mechanosensitive channel of large-conductance MscL (1MSL), the potassium channel KcsA (1K4C), and the outer membrane transporters BtuB (1NQE) and FepA (1FEP). A comparison of the experimental and self-assembly simulation results is shown (Figure 4).

Kaback and coworkers extensively analyzed LacY using site-directed chemical modification, intermolecular crosslinking, and site-directed spin labeling (Ermolova et al., 2003, 2006; Guan et al., 2002; Kwaw et al., 2001; Venkatesan et al., 2000a, 2000b, 2000c; Voss et al., 1996; Zhang et al., 2003; Zhao et al., 1999). Residues accessible to the alkylating agent are located on the interior of the protein near the substrate binding site, in loop regions, and in bands around the top and bottom of the protein perpendicular to the membrane normal. Surface-exposed positions were also identified in crosslinking studies (Ermolova et al., 2003; Guan et al., 2002). Finally, residues exposed to the lipid bilayer were identified by site-directed

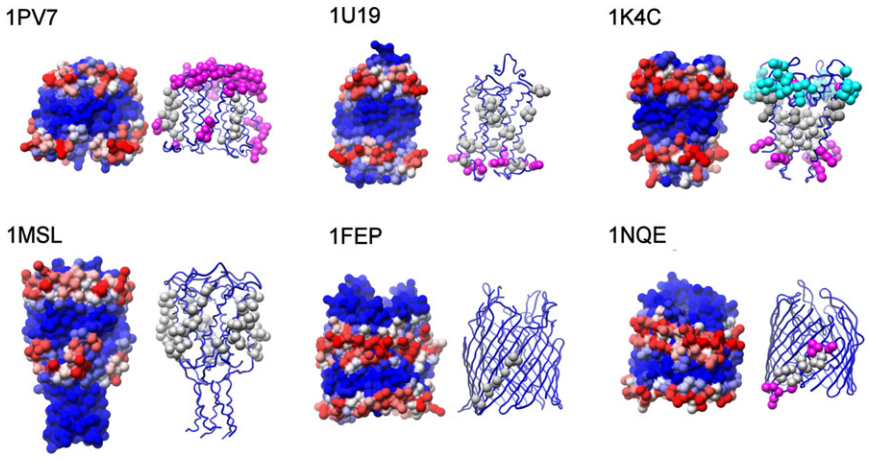

Figure 4. Comparison with Experimental Data for Six Proteins

In all panels, the representation on the left shows a space-filled CG protein colored according to the fraction of the time contacts are made between the protein and the lipid head group/glycerol backbone. Residues colored blue interact with these particles for 0%, those colored white for 50%, and those colored red for 100% of the simulation. The representation on the right shows the backbone trace of the protein in blue and residues for which there are experimental data. 1PV7: LacY. Residues that can be Cys crosslinked or chemically modified by N-ethylmaleimide (NEM) are shown in magenta, and residues that are resistant to NEM or shown to be buried in site-directed spin labeling studies are shown in gray. 1U19: Rhodopsin. Residues that can be labeled by a hydrophilic probe are shown in magenta, and those that can be modified by a hydrophobic probe are in gray. 1K4C: KcsA. Residues shown to be exposed to the aqueous environment are colored magenta, and those that are accessible to a spin label attached to a lipid molecule are shown in cyan. Residues that are not accessible to the aqueous environment are shown in gray. 1MSL: MscL. Residues shown by Trp fluorescence to be located in the hydrophobic region of the membrane are colored gray. 1FEP: Ferric enterobactin receptor FepA. Residues exposed to the hydrophobic region of the bilayer are shown in gray. 1NQE: BtuB membrane transporter. Residues exposed to the hydrophobic region of the bilayer are shown in gray, and those in the interfacial region are shown in magenta.

spin labeling (Voss et al., 1996; Zhao et al., 1999). Our data are, in general, consistent with these results (Figure 4). Residues accessible to N-ethylmaleimide (NEM) or that form intermolecular crosslinks are exposed to solvent or in the region of the protein that interacts with the lipid head group and glycerol backbone particles in the simulations. Those residues that are shown by spin labeling to be in contact with lipid lie in the region in contact with lipid tails in the simulation.

The membrane/water accessibility of residues in ovine rhodopsin has been determined in native membrane by chemical modification with hydrophobic and hydrophilic probes (Barclay and Findlay, 1984; Davison and Findlay, 1986). All lysine residues that reacted with the hydrophilic probe are either solvent exposed or in the head group region of the membrane in the simulation (Figure 4). Most residues modified by the hydrophobic probe interact with the lipid tails throughout the simulations; exceptions are either in contact with the glycerol backbone particles for the majority of the simulation, or (in two cases, Lys231 and Cys316) predominantly interact with the lipid head groups.

MscL from *Mycobacterium tuberculosis* has been studied using fluorescence spectroscopy (Powl et al., 2003, 2005). These studies suggest that Leu69 and a residue between Val91 and Tyr94 are located in the interfacial region of the bilayer, and that Phe80 is located near the center of the bilayer. Our simulations locate Phe80 near the center of the bilayer in all five MscL subunits. In the upper leaflet, Leu69 interacts with lipid tail particles throughout the simulation and with glycerol backbone particles for ~50% of the simulation. The region of the protein interacting with head group and glycerol backbone particles is broader in the lower leaflet than in the upper leaflet. Val91 and Trp94 both interact with lipid tail, glycerol backbone, and PC head groups at different times in the simulation. Both make contact with the glycerol backbone during ~80% of the simulation but Tyr94 interacts with PC head groups for ~70% and Val91 for ~40% of the simulation.

The interaction of KcsA with lipid has been probed by site-directed spin labeling (Gross et al., 1999; Gross and Hubbell, 2002; Perozo et al., 1998). It was shown that Arg27 is exposed to solvent at the N terminus of TM1. In the C-terminal region of TM1, residues from 45 onward are accessible to quenching reagents and Arg52 is solvent exposed. In TM2, residues 86–90 and 117 onward are accessible to quenching reagents, whereas Val97, Phe103, Val106, Ala109, Leu110, and Trp113 are in the hydrophobic region of the membrane. Our simulations show that the first residue in TM1 to interact with lipid head groups is Ser22 and that Arg27 is ''snorkeling'' and thus within 6 Å of the solvent, phosphatidyl choline (PC) head groups, and glycerol backbone. Further along TM1, Leu49 interacts with the glycerol backbone in the upper leaflet. Arg52 is the last residue of TM1 to make prolonged contact with the bilayer interacting with both the PC head groups and solvent. In TM2, Thr85 is the first residue to make contact with PC head groups. Trp87 interacts with the glycerol backbone. At the C terminus, Trp113 interacts with the glycerol backbone and Arg117 with the glycerol backbone, PC head groups, and solvent.

Site-directed spin labeling of two β strands of BtuB (Fanucci et al., 2002) indicate that residues 154, 156, 166, and 168 are in the center of the bilayer, whereas 150, 158, 162, 164, and 172 lie in the interface. Our simulations show residues Gln150, Gln158, Thr164, and Tyr172 interact principally with the glycerol backbone, whereas Asp162 is partially solvent exposed and interacts with the PC head groups. Residues Val154, Thr156, Val166, and Leu168 contact lipid tail particles for 99%–100% of the last 100 ns of simulation. In FepA, similar studies of a single β strand (residues 245–253) (Klug et al., 1997) suggested all examined residues on the outside of the strand had interfacial locations. Our simulations show similar results. Tyr253 interacts with glycerol backbone and lipid tails throughout the simulation. Gln245 contacts lipid tails for 90% of the simulation but also contacts the glycerol backbone for ~20%. Residues 246–252 exclusively contact the lipid tail.

Several surveys have been conducted of lipid and/or detergent molecules bound to membrane proteins in crystal structures (Lee, 2003; Marsh and Pali, 2004, 2006; Palsdottir and Hunte, 2004). There are a number of protein structures with multiple bound lipid or detergent molecules organized in a bilayer-like fashion, including the high-resolution (1.9 Å; 2B6O) cryoelectron microscopy structure of Aqp0 (Gonen et al., 2005) containing eight dimyristoyl phosphatidylcholine molecules per monomer. The X-ray structure (1.96 Å resolution; 2NS1) of the AmtB/GlnK complex (Gruswitz et al., 2007) also contains eight detergent (β-octyl glucoside) molecules per AmtB monomer. We compared these lipid/detergent contacts with the CG-MD simulation predictions. In each case, we identified residues in the experimental structures forming close (<4 Å) contacts to potential H-bonding atoms of the lipid/detergent head groups. For AmtB, these residues were in contact with PC head groups for 79% of the simulation with a mean number of contacts of 2.7 per residue and 63% and 2.0 for Aqp0. Averaging across all residues in these two proteins, one obtains values of 14% and 0.38, respectively. Thus, in both examples, residues interacting with lipid/detergent head groups in the crystal structures formed significant such interactions in the CG-MD simulations.

In summary, detailed comparison of our simulation results for eight proteins for which there are extensive experimental data demonstrated good agreement in terms of lipid-protein interactions.

Protein-Lipid Contacts

The distribution of amino acid residues making contact with the lipid bilayer as a function of the distance normal to the bilayer was analyzed. This is of interest for sequence-based predictions of the structure of membrane proteins (Bowie, 2005), and also for comparison with studies of translocon-mediated insertion of TM α helices (Hessa et al., 2005). For this analysis, the inner leaflet was defined to have a negative distance from the bilayer center of mass. The inside of the membrane was defined as the cytoplasmic side of the membrane in Gram-positive bacteria, the cytoplasmic side of the inner membrane in Gram-negative bacteria, the periplasmic side of the outer membrane in Gram-negative bacteria, the stromal side of the thylakoid membrane, and the matrix side of the inner mitochondrial membrane. As expected, characteristic distributions were observed for each residue type (Figure 5). The results obtained when all simulations are analyzed together are described here.

The hydrophobic residues Ile, Leu, Val, and Ala are the most abundant types, accounting for 39% of all TM residues in contact with lipid. The distribution of hydrophobic residues relative

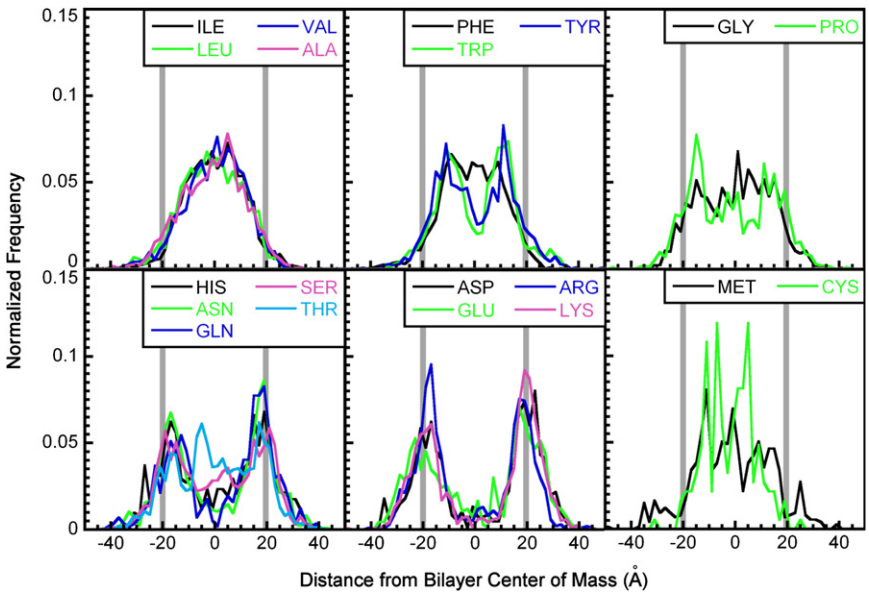

Figure 5. Residue Distributions Perpendicular to the Bilayer Center of Mass

Only protein residues in contact with the lipid bilayer for >33% of the last 100 ns of simulation are considered (similar results are seen for contact times of 10% and 5%). The position of each residue was calculated as the mean distance perpendicular to the plane defining the bilayer center of mass over the last 100 ns of simulation. The data for each residue type were normalized to the number of each residue type making contact with lipid. Residues are classified according to type: hydrophobic, aromatic, other, charged, polar, and sulfur-containing. Gray lines are shown at ±20 Å from the bilayer center of mass to facilitate comparison between residue types. The distribution for Cys and Met is noisy owing to the very small number of residues. His is not charged in the CG representation.

to the bilayer normal shows a single peak centered on the bilayer center of mass, demonstrating preferential interaction with lipid tail particles. The distributions for the aromatic residues Trp and Tyr clearly show two peaks at ±10–12 Å from the bilayer center of mass where they interact with the DPPC glycerol backbone particles. The interfacial location of Trp was not surprising (Killian and von Heijne, 2000), but the interfacial preference of Tyr was clearer than in previous studies (Ulmschneider et al., 2005). In contrast, a greater proportion of Phe residues interact with the lipid tails. With the exception of Thr, the polar and charged residues show two major peaks in their distributions at around ±18–20 Å from the bilayer center of mass where they interact with the PC head groups. The distributions for Arg, Lys, Glu, and Asp are asymmetric, with a higher frequency of Arg in the inner leaflet and Lys, Glu, and Asp in the outer leaflet. Significantly, Gly and Pro residues (accounting for 6% and 4% of residues, respectively) occurred across the whole TM region. Met and Cys are too poorly sampled for conclusions to be drawn about their preferential locations. For all amino acids, the width of the distribution of residues relative to the bilayer center of mass is influenced by the dynamic nature of the simulations and the combination of proteins with different TM hydrophobic thickness. Our results are in good agreement with the observations by von Heijne and coworkers on the positioning of amino acids which favor partitioning into the membrane in the translocon (Hessa et al., 2005).

Properties of the Membrane Bilayer

Our approach enables us to explore local perturbations of the bilayer properties in the vicinity of the protein. One metric of bilayer distortion around a protein is the separation of the phosphate head groups in the upper and lower leaflets as a function of distance from the protein (Bond and Sansom, 2007; Sands and Sansom, 2007). The phosphate-phosphate separation was measured by dividing the membrane into 5 × 5 Å bins in the plane of the membrane and calculating the mean distance of the phosphate group in the upper and lower membrane leaflets from the bilayer center of mass over the course of the simulation. The distance between phosphate particles in the upper and lower leaflets was then calculated for each bin. In our simulations, unless a large extracellular region interacts with the bilayer, the influence of a protein typically extends to a distance of ~30 Å; beyond 30 Å, the mean phosphate-phosphate separation is 42 ± 1 Å. Closer to the protein, the bilayer behavior varies depending on protein architecture. This is illustrated for two proteins, the β-barrel outer membrane enzyme PagP (1THQ) and the α-helical LacY (Figure 6). The mean phosphate-phosphate separation of lipids making contact with the protein can deviate from that of the bulk bilayer by several Å (as also seen in recent CG-MD simulations of rhodopsin; Periole et al., 2007), whereas the standard deviation is also larger closer to the protein, reflecting both lower sampling and protein asymmetry.

To investigate the effect of differing protein TM architectures on bilayer thickness, data from multiple proteins were combined. The resultant frequency distributions (Figure 7) show that with increasing distance from the protein, the bilayer thickness approaches 40–42 Å for both classes of TM proteins. There is a large variation in the bilayer thickness less than 10 Å from the protein; however, some general properties can be identified by looking at the behavior of the lipids farther away from the protein. For both α-helical and β-barrel proteins, the distributions of phosphate-phosphate separation at 10–15 Å are broad, with the distribution for α-helical proteins centered on ~41 Å and that for β-barrel proteins centered on ~39 Å. A higher proportion of β-barrel proteins have a bilayer thickness in the range 35–40 Å at 10–15 Å from the protein compared to α-helical proteins.

The distributions of phosphate-phosphate separation at a distance of 10–15 Å from the protein are broad. This breadth might result from proteins being surrounded by a bilayer of uniform thickness, with the thickness varying between proteins. Alternatively, each protein might be surrounded by a bilayer of nonuniform thickness. Examples of both are seen in our CG-MD data set, although local distortion of the bilayer leading to nonuniform bilayer thickness is most prevalent. The magnitude of the local

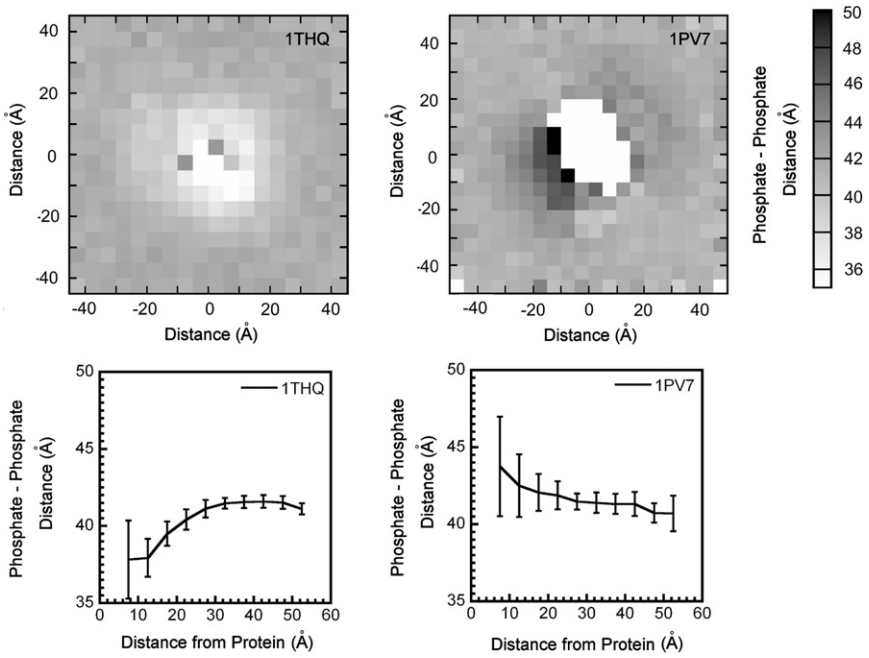

Figure 6. Bilayer Distortion around Transmembrane Proteins

The upper panel shows the mean phosphate-phosphate distance over the last 100 ns of simulation. The bilayer is divided into 5 × 5 Å bins, with the color intensity indicating the mean phosphate-phosphate distance over the last 100 ns of simulation. For this analysis, the bilayer is centered on the protein center of mass. The lower panel shows the mean phosphate-phosphate distance as a function of the distance from the protein. In PagP (left), the bilayer thickness at distances of 10–15 and 15–20 Å from the protein are clearly lower than that in the bulk lipid. In LacY (right), the mean bilayer thickness is the same (within error) at all distances from the protein.

distortion is protein dependent. For MscL, at 10–15 Å from the protein surface, there is a mean phosphate-phosphate distance of 40 Å with a standard deviation of 5.0 Å, whereas α-hemolysin has a mean distance of 37 Å with a standard deviation of only 1.6 Å.

Other Lipid Bilayers

The CG-MD self-assembly method can be used to explore the interactions of membrane proteins with different bilayer environments. To illustrate this, for eight proteins (1E12, 1EK9, 1FEP, 1OGV, 1PV7, 1Z98, 1THQ, and 2J7A), in addition to simulations using DPPC at 323K and 303K, simulations were carried out using palmitoyl-oleoyl phosphatidylcholine (POPC) at 323K and 303K; a 4:1 mixture of palmitoyl-oleoyl phosphatidylethanolamine (POPE) and palmitoyl-oleoyl phosphatidylglycerol (POPG) (Marrink et al., 2004); palmitoyl-lauroyl phosphatidylcholine (PLPC); and dilauroyl phosphatidylcholine (DLPC). The CG representation of oleic acid has five particles, palmitic acid has four particles, and lauric acid has three particles. This allowed us to investigate differences in protein-lipid interactions with fatty acid chain length and also head group type. For each protein, the fraction of time residues interact with lipid-lipid head groups was compared for pairs of simulations over all residues. For six of the proteins, 1E12, 1FEP, 1EK9, 1PV7, 1Z98, and 1THQ, similar regions of the protein interact with lipid head groups irrespective of chain length and head group type (pairwise correlation coefficients for comparison of the fraction of time residues interact with lipid head groups range from $r^2 \approx 0.8$ to 0.9). The data for 1FEP are shown as an example in Figures S1 and S2. As might be expected, the greatest differences are seen between the POPE:POPG bilayer and the DLPC bilayer, where the difference in fatty acid chain length is greatest and the charge of the head groups differs. For the other two proteins, 1OGV and 2J7A, much greater differences are seen between simulations with different lipids, and also between duplicate simulations using a different random seed (pairwise correlation coefficients for comparison of the fraction of time residues interact with lipid head groups range from $r^2 \approx 0.6$ to 0.7). The data for 1OGV are shown as an example in Figures S3 and S4. For 1OGV, a photosynthetic reaction center, some of the differences between simulations arise owing to the presence of lipid molecules in regions of the protein where

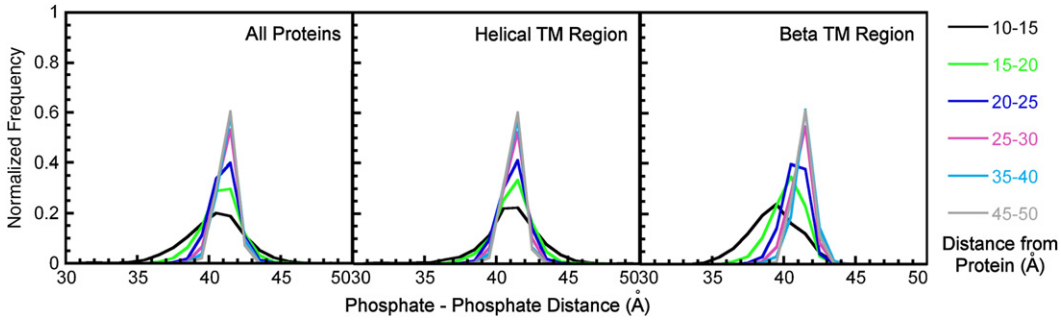

Figure 7. Bilayer Distortion around Proteins with Different Transmembrane Architectures

The mean phosphate-phosphate distance was calculated as a function of the distance from the protein. The calculation was performed for three categories: the whole data set, α-helical TM proteins, and β-sheet TM proteins. The distance from the protein is split into 5 Å bins and the phosphate-phosphate distance into 1 Å bins. As the distance from the protein increases, the distribution moves toward 41–42 Å.

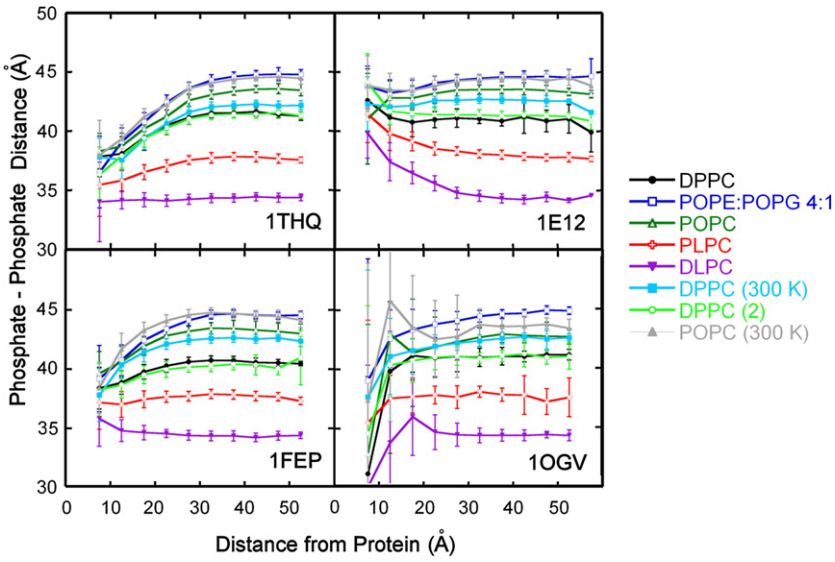

Figure 8. Other Lipid Bilayers

The mean phosphate-phosphate distance is shown as a function of the distance from the protein surface for four different proteins, with verticle bars representing the standard deviation of the mean [PagP (1THQ), FepA (1FEP), halorhodopsin (1E12), and a photosynthetic reaction center (1OGV)]. For PagP, FepA, and halorhodopsin, the extent of local bilayer distortion depends upon the degree of hydrophobic mistmatch between the protein and the lipid. For 1OGV, there is a large variation in the position of lipids close to the protein, illustrated by the magnitude of the standard deviation of the mean.

cofactors would be bound, whereas others reflect a difference in positioning in the bilayer. 2J7A, part of the NrfHA complex, has only two helices inserted into the membrane. In this case, some of the differences between simulations appear to arise from changes in the relative orientation of the TM helices in response to the formation of the lipid bilayer, whereas others arise from changes in the position of domains that are not fully inserted into the membrane.

In our simulations, for proteins without extramembrane domains that interact with the lipid surface, similar regions of the protein typically interact with the lipid irrespective of tail length and head group charge. This is further illustrated by the comparison of local bilayer distortion in different simulations (Figure 8). At distances greater than 30 Å from the protein surface, the bilayer thickness is not influenced by the presence of these proteins. Closer to the protein, the extent of the distortion is dependent upon the magnitude of the hydrophobic mismatch between the protein and lipid bilayer. Thus, for 1THQ (PagP), whose hydrophobic thickness matches DLPC, distortion is increased with longer tail length. For 1E12 (halorhodopsin), whose hydrophobic thickness matches lipid POPC, distortion is decreased with longer tail length. In contrast, for 1OGV, there is a large variation in lipid positions close to the protein. It should be noted that in our simulations, the protein is restrained using an elastic network model. Although this does not prevent structural fluctuation, particularly in α-helical TM proteins where the majority of restraints are within rather than between helices, in our simulations the response of the protein to hydrophobic mismatch is typically a change in tilt angle rather than large structural rearrangement. This is seen, for example, in PagP, where the tilt angle increases with decreasing hydrophobic thickness.

DISCUSSION

The full determination of membrane protein structure requires both the three-dimensional coordinates of the protein atoms and the location of the lipid bilayer. This study aims to address the second of these questions. A number of methods have previously been used to position proteins relative to a bilayer treated as a hydrophobic slab (Basyn et al., 2001, 2003; Ducarme et al., 1998; Im et al., 2003; Lomize et al., 2006a, 2006b; Tusnady et al., 2005a, 2005b; Ulmschneider et al., 2005, 2006). Our approach uses CG-MD to predict the membrane protein position relative to a more detailed bilayer model. The CG approach is intermediate between high-resolution atomic simulations and a very low resolution hydrophobic slab model. Our results provide a set of predictions for membrane protein position in a lipid bilayer, and also model lipid-protein interactions. The CG data can be used to generate atomic coordinates for further simulation as part of a multiscale approach (Shi et al., 2006).

Two important tests for our model are how well our results compare with experimental data for bilayer position and whether the CG bilayer can adapt to match the hydrophobic thickness of the protein. Experimental data exist for several proteins in our database. We have presented a comparison of our CG results with data for six α-helical and two β-barrel TM proteins showing a high qualitative similarity between the experimental and simulation results. Observed discrepancies might arise from differing lipid composition between the experimental and CG studies, or the fact that the protein secondary structure is restrained in simulation-limiting protein dynamics. These factors will be addressed in future extensions to the CG approach.

One of the important ways in which the CG method differs from the hydrophobic slab approximation is that the CG lipid bilayer can distort to match the hydrophobic thickness of the protein. The efficiency of hydrophobic matching has been demonstrated experimentally (Williamson et al., 2002). In this study, we have used the same bilayer composition for 91 proteins, the equilibrium hydrophobic thickness of which often differs from the hydrophobic thickness of the protein. Our results show that the CG bilayer responds to the properties of the protein. The CG model shows specificity of lipid-protein interaction; the extent of bilayer distortion typically depends upon the degree of hydrophobic mismatch, maintaining lipid-protein contacts. Close to the protein, we see large variation in bilayer thickness, as expected for proteins sourced from membranes with differing lipid composition. We see some generalized differences in bilayer distortion between proteins with different TM architectures; β-barrel TM regions show a smaller bilayer thickness than α-helical TM regions.

The lipid bilayer can be modeled as a central hydrophobic core and an interfacial region of polar lipid head groups surrounded by solvent. The free energy of transfer of different amino acids into different regions of the membrane varies enormously. As a result, different amino acids are preferentially located in different regions of the membrane (Nyholm et al., 2007). The amino acid distribution in the lipid bilayer, especially in α-helical proteins, has been the focus of a number of studies (Arkin and Brunger, 1998; Hurwitz et al., 2006; Pellegrini-Calace et al., 2003; Senes et al., 2007; Ulmschneider et al., 2005; Yarov-Yarovoy and Baker, 2006). Distributions for α-helical TM proteins have been used to develop empirical potential functions for membrane protein insertion (Senes et al., 2007; Ulmschneider et al., 2005). The methods differ in the structures used to determine residue frequency and in the equation used in the potential function. Despite these differences, the amino acid distributions reported are similar. The distribution of amino acid residues in β-barrel TM proteins has recently been reported for both the internal and external surfaces of the β strands of outer membrane proteins (Valavanis et al., 2006).

The distribution of residues normal to the plane of the membrane is similar in our CG simulations to those reported in the previously published statistical analyses (Senes et al., 2007; Ulmschneider et al., 2005). In all cases, hydrophobic residues are seen to have a strong preference for the membrane core where they interact with the lipid tails. Aromatic residues, especially Trp and Tyr, are believed important in anchoring the protein in the membrane through interaction with the glycerol backbone and lipid head groups (Ridder et al., 2000; Yau et al., 1998). For Trp and Tyr, we observe maxima in the distributions at the glycerol backbone for both protein architectures. With the exception of Ser and His, the polar residues behave similarly in our data set and in the statistical analyses. In previous analyses, Ser has shown no preference for either the polar or hydrophobic region of the membrane. In our analysis, the distribution for Ser shows two maxima at the lipid head groups. In the analysis by Ulmschneider et al., His had a distribution similar to Trp and Tyr, with maxima at the interfacial region of the membrane. Our analysis shows that for His, the maxima occur at a similar position to those of the charged residues. Charged residues are also important in protein membrane positioning, with a preference for positively charged residues occurring on the inside of the membrane (Nyholm et al., 2007; von Heijne, 1992). In our analysis, we see approximately symmetrical distributions for the charged residues; this is not inconsistent with previous results, as we have only considered residues making contact with the lipid bilayer.

The work presented here demonstrates that it is possible to accurately predict the position of a membrane protein in the lipid bilayer using CG methods based on the protein structure. The position of the protein in the bilayer appears independent of membrane lipid composition. In the future, this method will be extended to use more complex, and hence more biologically realistic, membrane compositions.

EXPERIMENTAL PROCEDURES

The CG representation for protein molecules was as described by Sansom and coworkers (Bond et al., 2007; Bond and Sansom, 2006), whereas that for lipid, detergent, water, and ion particles was as used by Marrink et al. (2004). The CG protein model was generated from the corresponding atomistic structure. A full list of PDB files used is given in the Supplemental Data. The CG representation of protein and lipid molecules used in this study has been tested previously (Bond et al., 2007; Bond and Sansom, 2006, 2007; Marrink et al., 2004). The protein backbone is represented by a single particle for each residue with between zero and three side chain particles, depending on residue type. As in previous studies (Bond et al., 2007), the side chains of Phe, Trp, and Tyr are represented by three particles and are planar. Protein structure was maintained using an elastic network model. Harmonic restraints were applied between all backbone particles within 7 Å of one another. The equilibrium bond length was equal to the separation of the particles in the starting structure and the force constant was 10 kJ mol^{-1} $Å^{-2}$. Cofactors were excluded in the CG representation. A number of the atomistic structures were missing density in loop regions; these regions were omitted in the CG representation.

All simulations were performed using GROMACS (http://www.gromacs.org/; Lindahl et al., 2001; van der Spoel et al., 2005). Lennard-Jones interactions were shifted to zero between 9 and 12 Å and electrostatic interactions were shifted to zero between 0 and 12 Å, with a relative dielectric constant of 20 used for explicit screening. The nonbonded neighbor list was updated every ten steps. Simulations were performed at constant temperature, pressure, and number of particles. The temperature of the protein, lipid/detergent, and solvent was each coupled separately using the Berendsen algorithm (Berendsen et al., 1984). The temperature was 323K and the coupling constant τ_T = 40 ps. The system pressure was semi-isotropically coupled using the Berendsen algorithm at 1 bar with a coupling constant τ_P = 40 ps and a compressibility of 1×10^{-5} bar. The time step for integration was 40 fs.

Following generation of the CG representation of the protein, the CG model was subjected to 100 steps of steepest descent energy minimization to relax any steric clashes within the protein model. The protein was aligned along the z axis of the simulation box and surrounded by randomly placed CG lipid/detergent molecules. The size of the box and the number of lipids used in each simulation is given in the Supplemental Data. The system was solvated with CG water particles and CG Na^+ or Cl^- counterions were added where necessary such that the overall charge was zero. The system was subjected to between 200 and 400 steps of steepest decent energy minimization prior to the production simulation. To facilitate analysis of protein-lipid interaction, the database was curated. Following bilayer formation, lipids outside the membrane-bilayer region were replaced with water particles and the system was subjected to ten steps of steepest descent minimization. Velocities were reassigned and a 200 ns simulation was performed to ensure that the system had not been perturbed. Analysis was performed using GROMACS tools and in-house software. Molecular graphics images were produced using the UCSF Chimera package (Pettersen et al., 2004).

SUPPLEMENTAL DATA

Supplemental Data include four figures, one table, Supplemental Experimental Procedures, and Supplemental References and can be found with this article online at http://www.structure.org/cgi/content/full/16/4/621/DC1/.

ACKNOWLEDGMENTS

This work was funded by the Wellcome Trust, the Biotechnology and Biological Sciences Research Council, and the Membrane Protein Structure Initiative. Our thanks to Janet Thornton, Marialuisa Pellegrini-Calace, and Mahon Maguire for valuable discussions.

Received: August 19, 2007
Revised: January 2, 2008
Accepted: January 7, 2008
Published: April 8, 2008

REFERENCES

Arkin, I.T., and Brunger, A.T. (1998). Statistical analysis of predicted transmembrane α-helices. Biochim. Biophys. Acta *1429*, 113–128.

Atilgan, A.R., Durell, S.R., Jernigan, R.L., Demirel, M.C., Keskin, O., and Bahar, I. (2001). Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys. J. *80*, 505–515.

Barclay, P.L., and Findlay, J.B. (1984). Labelling of the cytoplasmic domains of ovine rhodopsin with hydrophilic chemical probes. Biochem. J. *220*, 75–84.

Basyn, F., Charloteaux, B., Thomas, A., and Brasseur, R. (2001). Prediction of membrane protein orientation in lipid bilayers: a theoretical approach. J. Mol. Graph. Model. *20*, 235–244.

Basyn, F., Spies, B., Bouffioux, O., Thomas, A., and Brasseur, R. (2003). Insertion of X-ray structures of proteins in membranes. J. Mol. Graph. Model. *22*, 11–21.

Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., and Haak, J.R. (1984). Molecular dynamics with coupling to an external bath. J. Chem. Phys. *81*, 3684–3690.

Bond, P.J., and Sansom, M.S.P. (2006). Insertion and assembly of membrane proteins via simulation. J. Am. Chem. Soc. *128*, 2697–2704.

Bond, P.J., and Sansom, M.S.P. (2007). Bilayer deformation by the Kv channel voltage sensor domain revealed by self-assembly simulations. Proc. Natl. Acad. Sci. USA *104*, 2631–2636.

Bond, P.J., Holyoake, J., Ivetac, A., Khalid, S., and Sansom, M.S.P. (2007). Coarse-grained molecular dynamics simulations of membrane proteins and peptides. J. Struct. Biol. *157*, 593–605.

Bowie, J.U. (2005). Solving the membrane protein folding problem. Nature *438*, 581–589.

Campos, F.V., Chanda, B., Roux, B., and Bezanilla, F. (2007). Two atomic constraints unambiguously position the S4 segment relative to S1 and S2 segments in the closed state of Shaker K channel. Proc. Natl. Acad. Sci. USA *104*, 7904–7909.

Davison, M.D., and Findlay, J.B. (1986). Identification of the sites in opsin modified by photoactivated azido[^{125}I]iodobenzene. Biochem. J. *236*, 389–395.

Ducarme, P., Rahman, N., and Brasseur, R. (1998). IMPALA: a simple restraint field to simulate the biological membrane in structure studies. Proteins *30*, 357–371.

Ermolova, N., Guan, L., and Kaback, H.R. (2003). Intermolecular thiol cross-linking via loops in the lactose permease of *Escherichia coli*. Proc. Natl. Acad. Sci. USA *100*, 10187–10192.

Ermolova, N., Madhvani, R.V., and Kaback, H.R. (2006). Site-directed alkylation of cysteine replacements in the lactose permease of *Escherichia coli*: helices I, III, VI, and XI. Biochemistry *45*, 4182–4189.

Fanucci, G.E., and Cafiso, D.S. (2006). Recent advances and applications of site-directed spin labeling. Curr. Opin. Struct. Biol. *16*, 644–653.

Fanucci, G.E., Cadieux, N., Piedmont, C.A., Kadner, R.J., and Cafiso, D.S. (2002). Structure and dynamics of the β-barrel of the membrane transporter BtuB by site-directed spin labeling. Biochemistry *41*, 11543–11551.

Fyfe, P.K., McAuley, K.E., Roszak, A.W., Isaacs, N.W., Codgell, R.J., and Jones, M.R. (2001). Probing the interface between membrane proteins and membrane lipids by X-ray crystallography. Trends Biochem. Sci. *26*, 106–112.

Gonen, T., Cheng, Y., Sliz, P., Hiroaki, Y., Fujiyoshi, Y., Harrison, S.C., and Walz, T. (2005). Lipid-protein interactions in double-layered two-dimensional AQP0 crystals. Nature *438*, 633–638.

Gross, A., and Hubbell, W.L. (2002). Identification of protein sidechains near the membrane-aqueous interface: a site-directed spin labelling study of KcsA. Biochemistry *41*, 1123–1128.

Gross, A., Columbus, L., Hideg, K., Altenbach, C., and Hubbell, W.L. (1999). Structure of the KcsA potassium channel from *Streptomyces lividans*: a site-directed spin labeling study of the second transmembrane segment. Biochemistry *38*, 10324–10335.

Gruswitz, F., O'Connell, J., and Stroud, R.M. (2007). Inhibitory complex of the transmembrane ammonia channel, AmtB, and the cytosolic regulatory protein, GlnK, at 1.96 Å. Proc. Natl. Acad. Sci. USA *104*, 42–47.

Guan, L., and Kaback, H.R. (2006). Lessons from lactose permease. Annu. Rev. Biophys. Biomol. Struct. *35*, 67–91.

Guan, L., Murphy, F.D., and Kaback, H.R. (2002). Surface-exposed positions in the transmembrane helices of the lactose permease of *Escherichia coli* determined by intermolecular thiol cross-linking. Proc. Natl. Acad. Sci. USA *99*, 3475–3480.

Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., Nilsson, I., White, S.H., and von Heijne, G. (2005). Recognition of transmembrane helices by the endoplasmic reticulum translocon. Nature *433*, 377–381.

Hunte, C. (2005). Specific protein-lipid interactions in membrane proteins. Biochem. Soc. Trans. *33*, 938–942.

Hurwitz, N., Pellegrini-Calace, M., and Jones, D.T. (2006). Towards genome-scale structure prediction for transmembrane proteins. Philos. Trans. R. Soc. Lond. B Biol. Sci. *361*, 465–475.

Im, W., Feig, M., and Brooks, C.L. (2003). An implicit membrane generalized Born theory for the study of structure, stability, and interactions of membrane proteins. Biophys. J. *85*, 2900–2918.

Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern-recognition of hydrogen-bonded and geometrical features. Biopolymers *22*, 2577–2637.

Killian, J.A., and von Heijne, G. (2000). How proteins adapt to a membrane-water interface. Trends Biochem. Sci. *25*, 429–434.

Klug, C.S., Su, W.Y., and Feix, J.B. (1997). Mapping of the residues involved in a proposed β-strand located in the ferric enterobactin receptor FepA using site-directed spin-labeling. Biochemistry *36*, 13027–13033.

Kwaw, I., Zen, K.C., Hu, Y.L., and Kaback, H.R. (2001). Site-directed sulfhydryl labeling of the lactose permease of *Escherichia coli*: helices IV and V that contain the major determinants for substrate binding. Biochemistry *40*, 10491–10499.

Lee, A.G. (2003). Lipid-protein interactions in biological membranes: a structural perspective. Biochim. Biophys. Acta *1612*, 1–40.

Lee, A.G. (2004). How lipids affect the activities of integral membrane proteins. Biochim. Biophys. Acta *1666*, 62–87.

Lee, A.G. (2005). How lipids and proteins interact in a membrane: a molecular approach. Mol. Biosyst. *1*, 203–212.

Lee, S.Y., Lee, A., Chen, J., and MacKinnon, R. (2005). Structure of the KvAP voltage-dependent K^+ channel and its dependence on the lipid membrane. Proc. Natl. Acad. Sci. USA *102*, 15441–15446.

Lindahl, E., Hess, B., and van der Spoel, D. (2001). GROMACS 3.0: a package for molecular simulation and trajectory analysis. J. Mol. Model. *7*, 306–317.

Lomize, A.L., Pogozheva, I.D., Lomize, M.A., and Mosberg, H.I. (2006a). Positioning of proteins in membranes: a computational approach. Protein Sci. *15*, 1318–1333.

Lomize, M.A., Lomize, A.L., Pogozheva, I.D., and Mosberg, H.I. (2006b). OPM: orientations of proteins in membranes database. Bioinformatics *22*, 623–625.

Lundstrom, K. (2006). Structural genomics for membrane proteins. Cell. Mol. Life Sci. *63*, 2597–2607.

Marrink, S.J., de Vries, A.H., and Mark, A.E. (2004). Coarse grained model for semiquantitative lipid simulations. J. Phys. Chem. B *108*, 750–760.

Marsh, D., and Pali, T. (2004). The protein-lipid interface: perspectives from magnetic resonance and crystal structures. Biochim. Biophys. Acta *1666*, 118–141.

Marsh, D., and Pali, T. (2006). Lipid conformation in crystalline bilayers and in crystals of transmembrane proteins. Chem. Phys. Lipids *141*, 48–65.

Mukherjee, P., Krummel, A.T., Fulmer, E.C., Kass, I., Arkin, I.T., and Zanni, M.T. (2004). Site-specific vibrational dynamics of the CD3ζ membrane peptide using heterodyned two-dimensional infrared photon echo spectroscopy. J. Chem. Phys. *120*, 10215–10224.

Mukherjee, P., Kass, I., Arkin, I.T., and Zanni, M.T. (2006). Picosecond dynamics of a membrane protein revealed by 2D IR. Proc. Natl. Acad. Sci. USA *103*, 3528–3533.

Nielsen, S.O., Lopez, C.F., Srinivas, G., and Klein, M.L. (2004). Coarse grain models and the computer simulation of soft materials. J. Phys. Condens. Matter *16*, R481–R512.

Nyholm, T.K.M., Ozdirekcan, S., and Killian, J.A. (2007). How protein transmembrane segments sense the lipid environment. Biochemistry *46*, 1457–1465.

Palsdottir, H., and Hunte, C. (2004). Lipids in membrane protein structures. Biochim. Biophys. Acta *1666*, 2–18.

Pellegrini-Calace, M., Carotti, A., and Jones, D.T. (2003). Folding in lipid membranes (FILM): a novel method for the prediction of small membrane protein 3D structures. Proteins *50*, 537–545.

Periole, X., Huber, T., Marrink, S.J., and Sakmar, T.P. (2007). G protein-coupled receptors self-assemble in dynamics simulations of model bilayers. J. Am. Chem. Soc. *129*, 10126–10132.

Perozo, E., Cortes, D.M., and Cuello, L.G. (1998). Three-dimensional architecture and gating mechanism of a K^+ channel studied by EPR spectroscopy. Nat. Struct. Biol. *5*, 459–469.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Powl, A.M., East, J.M., and Lee, A.G. (2003). Lipid-protein interactions studied by introduction of a tryptophan residue: the mechanosensitive channel MscL. Biochemistry *42*, 14306–14317.

Powl, A.M., Wright, J.N., East, J.M., and Lee, A.G. (2005). Identification of the hydrophobic thickness of a membrane protein using fluorescence spectroscopy: studies with the mechanosensitive channel MscL. Biochemistry *44*, 5713–5721.

Ridder, A., Morein, S., Stam, J.G., Kuhn, A., de Kruijff, B., and Killian, J.A. (2000). Analysis of the role of interfacial tryptophan residues in controlling the topology of membrane proteins. Biochemistry *39*, 6521–6528.

Sands, Z.A., and Sansom, M.S.P. (2007). How does a voltage-sensor interact with a lipid bilayer? Simulations of a potassium channel domain. Structure *15*, 235–244.

Senes, A., Chadi, D.C., Law, P.B., Walters, R.F.S., Nanda, V., and DeGrado, W.F. (2007). E(z), a depth-dependent potential for assessing the energies of insertion of amino acid side-chains into membranes: derivation and applications to determining the orientation of transmembrane and interfacial helices. J. Mol. Biol. *366*, 436–448.

Shelley, J.C., Shelley, M.Y., Reeder, R.C., Bandyopadhyay, S., and Klein, M.L. (2001). A coarse grain model for phospholipid simulations. J. Phys. Chem. B *105*, 4464–4470.

Shi, Q., Izvekov, S., and Voth, G.A. (2006). Mixed atomistic and coarse-grained molecular dynamics: simulation of a membrane bound ion channel. J. Phys. Chem. B *110*, 15045–15048.

Tusnady, G.E., Dosztanyi, Z., and Simon, I. (2004). Transmembrane proteins in the Protein Data Bank: identification and classification. Bioinformatics *20*, 2964–2972.

Tusnady, G.E., Dosztanyi, Z., and Simon, I. (2005a). PDB_TM: selection and membrane localization of transmembrane proteins in the Protein Data Bank. Nucleic Acids Res. *33*, D275–D278.

Tusnady, G.E., Dosztanyi, Z., and Simon, I. (2005b). TMDET: web server for detecting transmembrane regions of proteins by using their 3D coordinates. Bioinformatics *21*, 1276–1277.

Ulmschneider, M.B., Sansom, M.S.P., and Di Nola, A. (2005). Properties of integral membrane protein structures: derivation of an implicit membrane potential. Proteins *59*, 252–265.

Ulmschneider, M.B., Sansom, M.S.P., and Di Nola, A. (2006). Evaluating tilt angles of membrane-associated helices: comparison of computational and NMR techniques. Biophys. J. *90*, 1650–1660.

Valavanis, I.K., Bagos, P.G., and Emiris, I.Z. (2006). β-barrel transmembrane proteins: geometric modelling, detection of transmembrane region, and structural properties. Comput. Biol. Chem. *30*, 416–424.

van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., and Berendsen, H.J. (2005). GROMACS: fast, flexible, and free. J. Comput. Chem. *26*, 1701–1718.

Venkatesan, P., Hu, Y.L., and Kaback, H.R. (2000a). Site-directed sulfhydryl labeling of the lactose permease of *Escherichia coli*: helix X. Biochemistry *39*, 10656–10661.

Venkatesan, P., Kwaw, I., Hu, Y.L., and Kaback, H.R. (2000b). Site-directed sulfhydryl labeling of the lactose permease of *Escherichia coli*: helix VII. Biochemistry *39*, 10641–10648.

Venkatesan, P., Liu, Z.L., Hu, Y.L., and Kaback, H.R. (2000c). Site-directed sulfhydryl labeling of the lactose permease of *Escherichia coli*: N-ethylmaleimide-sensitive face of helix II. Biochemistry *39*, 10649–10655.

von Heijne, G. (1992). Membrane protein structure prediction. Hydrophobicity analysis and the positive inside rule. J. Mol. Biol. *225*, 487–494.

Voss, J., He, M.M., Hubbell, W.L., and Kaback, H.R. (1996). Site-directed spin labeling demonstrates that transmembrane domain XII in the lactose permease of *Escherichia coli* is an α-helix. Biochemistry *35*, 12915–12918.

White, S.H. (2004). The progress of membrane protein structure determination. Protein Sci. *13*, 1948–1949.

White, S.H., and Wimley, W.C. (1999). Membrane protein folding and stability: physical principles. Annu. Rev. Biophys. Biomol. Struct. *28*, 319–365.

Williamson, I.M., Alvis, S.J., East, J.M., and Lee, A. (2002). Interactions of phospholipids with the potassium channel KcsA. Biophys. J. *83*, 2026–2038.

Yarov-Yarovoy, V., and Baker, J.S.D. (2006). Multipass membrane protein structure prediction using Rosetta. Proteins *62*, 1010–1025.

Yau, W.M., Wimley, W.C., Gawrisch, K., and White, S.H. (1998). The preference of tryptophan for membrane interfaces. Biochemistry *37*, 14713–14718.

Zhang, W., Hu, Y.L., and Kaback, H.R. (2003). Site-directed sulfhydryl labeling of helix IX in the lactose permease of *Escherichia coli*. Biochemistry *42*, 4904–4908.

Zhao, M., Zen, K.C., Hernandez-Borrell, J., Altenbach, C., Hubbell, W.L., and Kaback, H.R. (1999). Nitroxide scanning electron paramagnetic resonance of helices IV and V and the intervening loop in the lactose permease of *Escherichia coli*. Biochemistry *38*, 15970–15977.

