

GRiffin: A Versatile Methodology for Optimization of Protein–Lipid Interfaces for Membrane Protein Simulations

René Staritzbichler,[†] Claudio Anselmi,[‡] Lucy R. Forrest,^{*,†} and José D. Faraldo-Gómez^{*,‡,§}

[†]Computational Structural Biology Group and [‡]Theoretical Molecular Biophysics Group, Max Planck Institute of Biophysics, Frankfurt am Main, Germany

[§]Cluster of Excellence Macromolecular Complexes, Frankfurt am Main, Germany

ABSTRACT: As new atomic structures of membrane proteins are resolved, they reveal increasingly complex transmembrane topologies and highly irregular surfaces with crevices and pores. In many cases, specific interactions formed with the lipid membrane are functionally crucial, as is the overall lipid composition. Compounded with increasing protein size, these characteristics pose a challenge for the construction of simulation models of membrane proteins in lipid environments; clearly, that these models are sufficiently realistic bears upon the reliability of simulation-based studies of these systems. Here, we introduce GRiffin (GRID-based Force Field INput), which uses a versatile framework to automate and improve a widely used membrane-embedding protocol. Initially, GRiffin carves out lipid and water molecules from a volume equivalent to that of the protein, to conserve the system density. In the subsequent optimization phase GRiffin adds an implicit grid-based protein force field to a molecular dynamics simulation of the precarved membrane. In this force field, atoms inside the implicit protein volume experience an outward force that will expel them from that volume, whereas those outside are subject to electrostatic and van der Waals interactions with the implicit protein. At each step of the simulation, these forces are updated by GRiffin and combined with the intermolecular forces of the explicit lipid–water system. This procedure enables the construction of realistic and reproducible starting configurations of the protein–membrane interface within a reasonable time frame and with minimal intervention. GRiffin is a stand-alone tool designed to work alongside any existing molecular dynamics package, such as NAMD or GROMACS.

■ INTRODUCTION

Membrane proteins constitute around a third of all proteins encoded in a typical genome,^{1–3} and yet the microscopic mechanisms of their functions are only just beginning to be described. Major advances in such understanding have been made through the elucidation of the three-dimensional atomic structure of some of these proteins by, e.g., X-ray crystallography. Indeed, the exponential increase in the number of structures being reported⁴ highlights the very significant progress made in this area. However, crystallographic structures capture a single state of what is typically a dynamic conformational equilibrium, intrinsic to the functional mechanism of the protein. A variety of structure-based computational approaches focus on these dynamic properties, in order to complement the experimental data. Prominent among these approaches is molecular dynamics (MD) simulation, for its ability to provide insights at atomic resolution, and its robust and versatile theoretical framework.

A surprising feature of many of the newly discovered membrane protein structures is their complex transmembrane topology, and the highly irregular interfaces they appear to form with the surrounding lipid bilayer. Indeed, for some proteins, such as the voltage-gated and mechanosensitive channels, or osmoregulatory transporters, regulatory mechanisms dependent on lipid composition are likely to be conveyed precisely by this protein–lipid interface.^{5–8} As the structure of the protein–membrane interface is not known experimentally, such complexity poses a serious challenge for the construction of molecular-simulation models. At the same time, whether this interface is realistically

modeled will clearly influence the ability of simulation-based studies to derive plausible mechanistic hypotheses.

A number of methodologies have been proposed for the construction of lipid–protein complexes for simulation.^{9–14} These adopt one of two general strategies, namely either to assemble a lipid bilayer around the protein *de novo*, or to adapt existing and well-optimized membrane models to the protein structure. The second strategy is an appealing and efficient option, because it builds upon much effort and considerable success in the construction of realistic membrane models for MD simulations, for a range of different lipids or mixtures thereof. Atomic coordinates and force fields for such systems are publicly available.

In one of these adaptive strategies, the protein and hydrated lipid bilayer systems are first superimposed in the same coordinate space, and some overlapping lipids and waters are removed, as necessary to preserve the system density.^{12,14} However, the cavity thus carved in the bilayer system does not normally match the shape of the protein, because of the very irregular conformations adopted by lipid hydrocarbon tails. To resolve this problem, one of the authors helped develop a methodology¹² in which the surface of the protein is used in the course of a molecular dynamics simulation to define additional forces that expel these tails from the protein volume. In this way, lipid and solvent atoms become adapted perfectly to the shape of the protein, and there is minimal perturbation of the existing, preoptimized membrane model.

Received: October 7, 2010

Published: March 29, 2011

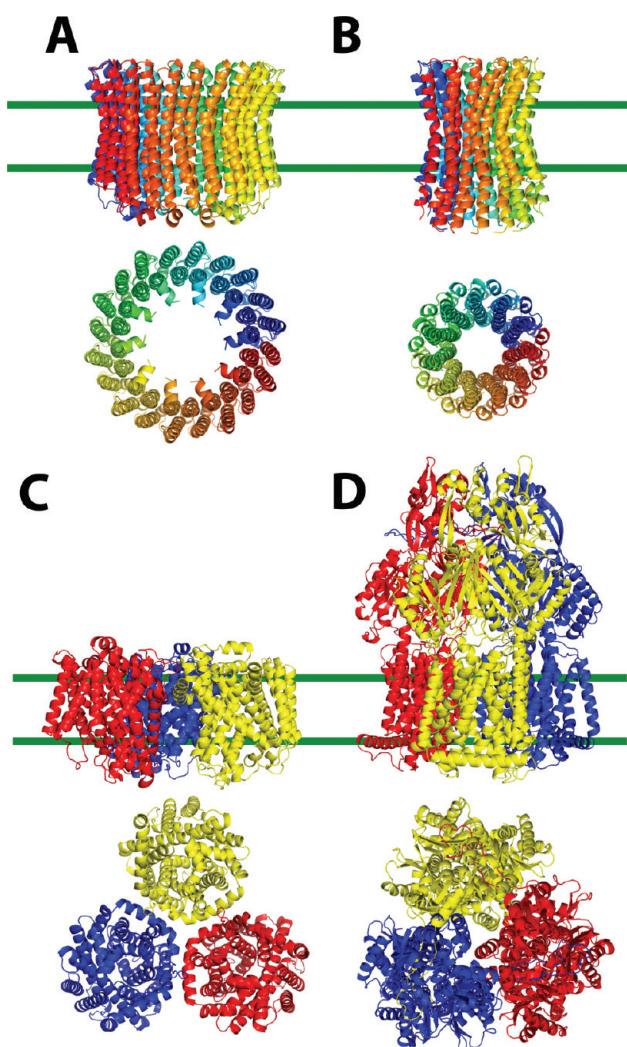


Figure 1. Representative membrane proteins with complex protein–lipid interfaces, used here as test cases for GRIFFIN. Each structure is displayed in cartoon format with different colors for each protein subunit, and viewed either from the plane of the membrane (upper panels) or normal to the membrane (lower panels). The approximate limits of the hydrophobic core of a hypothetical lipid membrane are also indicated (green lines). (A) The K_{10} rotor ring of the V-type Na^+ -ATPase from *Enterococcus hirae*. (B) The c_{11} rotor ring of the F-type Na^+ -ATP synthase from *Ilyobacter tartaricus*. (C) The carnitine/ γ -butyrobetaine antiporter CaiT from *Proteus mirabilis*. (D) The multidrug efflux pump AcrB from *E. coli*.

Here, we present an improved and more general version of this surface-based approach, based upon a newly developed tool named GRIFFIN (GRId-based Force Field INput). GRIFFIN provides several significant advantages over the earlier implementation. First, forces acting on lipid atoms due to Coulomb and van der Waals interactions with the protein volume are now included along with the repulsive surface-guided forces; this results in chemical specificity at the protein–lipid interface, in addition to shape complementarity. Second, new features are included to selectively guide lipids or other molecules into or out of user-specified regions of the system, and also to prevent trapping of lipid tails in regions of high irregularity. These features thus provide a means to handle the kind of complex protein topologies increasingly uncovered by structural studies.

Third, GRIFFIN is designed to be a stand-alone, fully integrated application, compatible with any MD engine that provides a suitable interface, such as NAMD,¹⁵ GROMACS,¹⁶ or CHARMM.¹⁷ Last, its object-oriented grid-based algorithm, and the resulting computational efficiency, extends the applicability of the method to very large membrane protein complexes.

We demonstrate the effectiveness of this improved methodology with four examples, namely, the secondary transporter CaiT,¹⁸ two membrane rotors from the ATPase family,^{19,20} and the AcrB multidrug efflux pump²¹ (Figure 1). In all cases, irregular features of the protein–lipid interface make other approaches either unsuitable or impractical. These features include a central pore in the ATPase rings, and intersubunit channels and water-filled crevices in the trimeric CaiT and AcrB; AcrB is also one of the largest and most complex membrane proteins structures resolved to date. We will show how GRIFFIN enables the construction of physically realistic starting configurations of the lipid–protein interface for membrane protein simulations, following a reproducible procedure and within an affordable time frame, even for these very challenging cases.

METHODOLOGY

The overall protocol is described in the scheme in Figure 2. Initially, a cavity is carved into a hydrated lipid bilayer system, in which the membrane protein will ultimately be embedded. The protein structure, or rather, its surface, is used to estimate the number of lipid and water molecules to be carved out. In most cases, however, the resulting protein–lipid interface will be unsuitable for energy minimization, let alone simulation. The same surface is therefore employed to define a grid-based force field designed to expel the remaining lipid and/or water atoms from that volume. This force field also includes physical interactions (electrostatic and van der Waals), but only in the region of the grid outside the protein volume. During a subsequent molecular dynamics simulation of the precarved membrane system, the implicit protein force field is added to the standard interaction forces, thus optimizing the protein–lipid and protein–water interfaces. Geometric objects may be used as additional, lipid and/or water specific exclusion volumes, during the carving and/or the simulation.

Initial Carving of the Lipid Membrane. GRIFFIN provides an integrated tool for constructing the surface of the protein and for the initial carving of the membrane. The surface is constructed using a rolling-sphere method,²² with a probe of adjustable radius, typically set to 1.4 Å. Atomic radii are derived from the force field to be used in the simulation, e.g., CHARMM27.²³ The protein surface is used to estimate the volume occupied by the protein within each leaflet of the lipid membrane, as well in the hydration layer. The boundaries of these regions may be user-defined or determined by the program. Based on this volume estimation and the lipid density, the number of lipid molecules to be deleted from each leaflet is determined. These are selected from a list of all lipids that overlap with the protein volume, ranked according to the number of overlapping atoms and their distance to the protein surface. All overlapping water molecules are also deleted. Specific values for these parameters may be also provided by the user, overriding those calculated by the program.

Additional Molecule Type Specific Exclusion Volumes. GRIFFIN offers the possibility of using geometric objects (spheres, rectangular cuboids, and cones) as additional exclusion

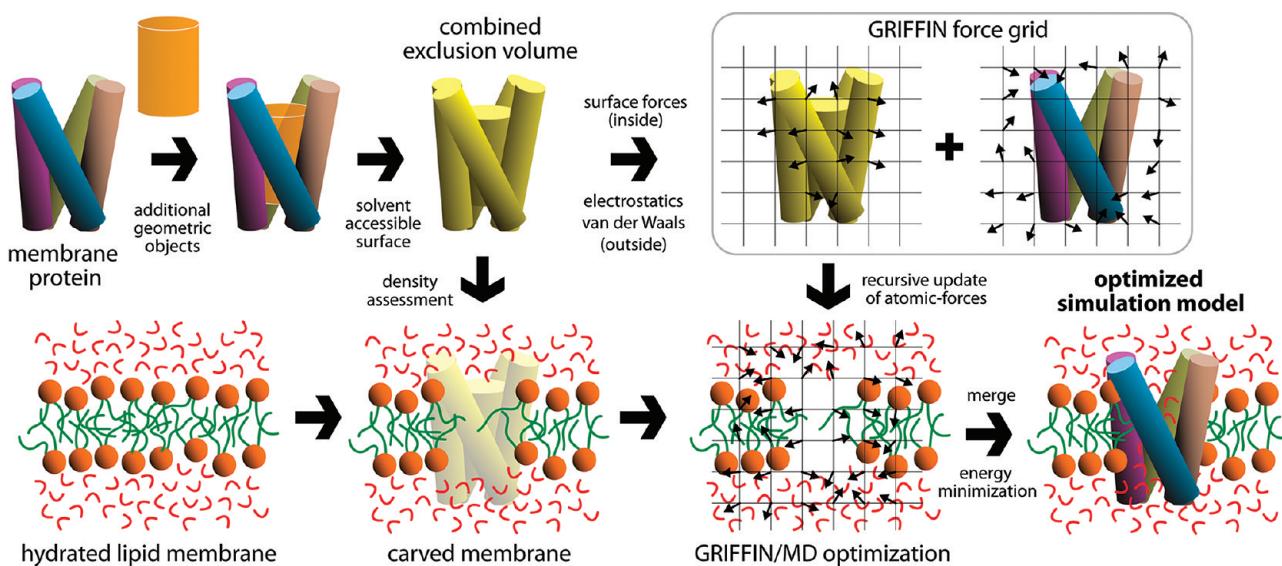


Figure 2. Schematic of the protocol for optimization of a protein–membrane interface using GRIFFIN. In the first step, a hypothetical membrane protein with a central pore (top left) is overlaid with a cylindrical geometric object (orange). The combined volume (yellow) of the protein and object is carved from an explicit system (bottom left) containing water (red) and lipid molecules (hydrocarbon chains are colored green and polar head groups are shown as orange spheres). After carving, some lipid and water atoms typically overlap the protein volume. The combined exclusion volume, and the actual protein structure, are also used to construct a force grid (top right) consisting of surface-directed expelling forces inside the volume and physical interaction forces outside. This force grid is overlaid on the molecular system to compute, at each time step of a molecular dynamics simulation, the actual forces on individual atoms; these are then added to the interaction energies between explicit atoms computed by the molecular dynamics package. In the last step, the GRIFFIN-optimized hydrated membrane is merged with the atomistic model of the membrane protein, and the entire system is energy-minimized to obtain the initial configuration for simulation (bottom right).

regions in the membrane carving, as well as in the optimization stage (see below). These objects are integrated in the volume estimation described above, and may be specific to a user-defined molecule type, e.g., to exclude lipid from a channel pore.

Calculation of the Implicit Protein Force Field. A second tool within GRIFFIN enables the calculation of the implicit force field for use during the subsequent molecular dynamics optimization stage. This force field is stored on a three-dimensional grid (Figure 3) of user-defined grid-point spacing (with a default value of 0.5 \AA). Within the protein volume (which may be modified to include the optional exclusion regions mentioned above), each grid point stores a force (of magnitude $1 \text{ kcal/mol}/\text{\AA}$) directed toward the nearest point on the protein surface (NPS; Figure 4A). These so-called surface forces will be applied to lipid and water atoms inside the volume during the optimization. Grid points outside of the protein surface map the Coulombic and van der Waals interaction forces between the protein and a probe particle of $q = 1e$, $\epsilon = 1 \text{ kcal/mol}$, and $\sigma = 1 \text{ \AA}$ (Figure 3). User-defined cutoff distances may be specified for each of these interactions (default values are 18 , 12 , and 8 \AA , respectively). The calculation of the force grid is carried out only once during the protocol; however, as it may be time-consuming, this calculation may be distributed over an adjustable number of processors.

Calculation of GRIFFIN Forces during Molecular Dynamics Simulations. GRIFFIN supplies a molecular dynamics simulation with atomic forces derived from the precomputed grid. These are calculated at every simulation step, based on the atomic coordinates of the explicit system (Figure 3F). To map the grid-point forces onto the actual atoms, GRIFFIN uses a trilinear interpolation method. By default, the physical forces are scaled by the actual charge and van der Waals parameters of each atom (a geometric-mean combination rule is used for both

σ and ϵ). Surface forces may also be scaled by a constant; in practice this constant is typically increased in a stepwise manner, as the surface forces ultimately become balanced out by the external pressure (see Results). A user-specified constant factor may also be used to further scale each of the physical interaction forces.

To improve the performance and stability of the algorithm, no surface forces are applied to hydrogen atoms by default. By contrast, all atoms may experience physical forces if outside the protein volume. However, for greater efficiency, physical interactions are computed only for atoms neighboring the protein; a list of such atoms is updated every few steps during the simulation (typically 10), and includes all atoms for which the GRIFFIN force is nonzero at the update step. Lastly, it should be noted that by construction no reaction forces on the protein are considered; it is thus advisable that GRIFFIN is employed alongside stochastic dynamics, with tight temperature and pressure coupling.

A feature specific to lipid molecules is the possibility of redirection of the surface forces. This feature is important when atoms within the same lipid chain are driven in opposite directions (see Results). Redirection is applied when the surface force acting on a given atom is directed away from the overall geometric center of the lipid molecule to which it belongs (Figure 4). An adjustable threshold defines the maximum angle by which the surface force may diverge; this is typically $<110^\circ$.

Simulation Details. All GRIFFIN optimizations described here were carried out with NAMD 2.7¹⁵ with the CHARMM27 protein/lipid force field.²³ The lipid types used were palmitoyl oleyl phosphatidylcholine (POPC; 234 for c_{11} , 542 for K_{10} , and 532 for AcrB) and palmitoyl oleyl phosphatidylethanolamine (POPE; 470 for CaiT). The simulations were carried out at

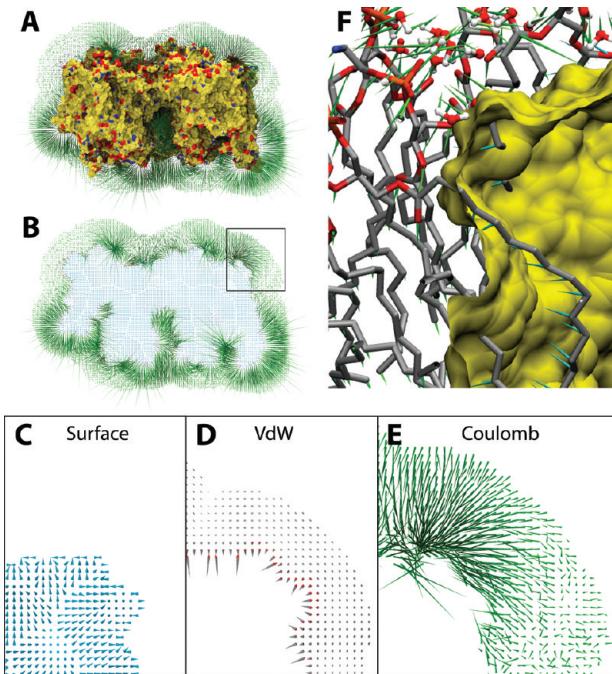


Figure 3. GRIFFIN force field for the CaiT protein. (A–E) The surface of CaiT (C atoms in yellow, O atoms in red, N atoms in blue) is shown overlaid on the corresponding GRIFFIN force field. Lines from a given grid point indicate the direction and magnitude of the forces generated by the protein atoms. These include surface forces (blue) inside the protein volume and electrostatic (green) and van der Waals attractive (gray) and repulsive (red) forces outside. (F) Actual forces derived by GRIFFIN inside (blue) and outside (green) the protein surface (yellow), on explicit lipid and water atoms, after interpolation and scaling of the force grid in (A)–(E).

constant temperature (298 and 310 K for POPC and POPE, respectively), using a Langevin thermostat (collision frequency of 100 ps^{-1}). The pressure was maintained along the membrane normal using a Nosé–Hoover Langevin piston barostat (1 atm), while the surface area of the membrane was kept constant. Electrostatic interactions were calculated using the particle mesh Ewald method (PME) with a real-space cutoff of 12 Å. A cutoff distance of 12 Å was also used for the van der Waals interactions. The MD integration time step was 1 fs; bonds involving hydrogen atoms were constrained with SETTLE. The number of lipid/water atoms in each optimization was 80 172 for c_{11} ring, 187 585 for K_{10} ring, 163 666 for CaiT, and 319 790 for AcrB.

RESULTS

GRIFFIN Forces Expel Lipid and Water Molecules from the Protein Volume. We began by testing the most basic GRIFFIN functionality, namely to gradually empty the volume of a membrane protein during the course of a MD simulation of a precarved, hydrated lipid bilayer. Our test system was the membrane rotor from a bacterial V-type ATPase (Figure 1A). This is a ring-shaped oligomeric assembly of 10 subunits, with a central pore that is plugged by lipid molecules.¹⁹ For this and subsequent tests, we employed NAMD as the MD engine coupled to GRIFFIN.

We monitor the progress of the procedure using two variables, namely the number of atoms inside the implicit protein volume

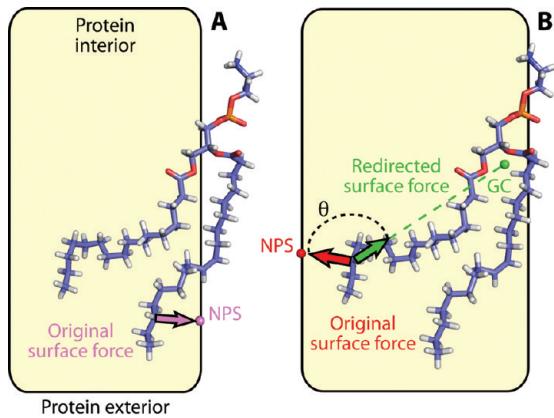


Figure 4. Schematic of force-redirection methodology. (A) Atoms in a lipid alkyl chain (sticks) entering the protein volume (yellow) experience surface forces toward the nearest point on the protein surface (NPS). (B) If the lipid chain spans most of the volume, however, atoms near the left-hand surface experience a surface force directed in the opposite direction from the rest of the chain (red). When the angle θ between the original surface force and the direction of the molecular geometric center (GC) exceeds a given threshold, the force is redirected toward the GC.

(excluding hydrogen atoms for convenience) and the maximum depth of any of those atoms beneath the protein surface. The K_{10} ring system starts out with about 6000 atoms inside the volume, which are up to about 10 Å deep. As shown in Figure 5A, application of the surface forces calculated by GRIFFIN results in a rapid decline in the number of buried atoms, until a plateau is reached, reflecting a balance between the expelling forces and the external pressure from lipid and water. As expected, the rate of decrease and the plateau level are dependent on the strength of the applied surface forces; for example, with a force of 3.0 kcal/mol/Å², the plateau is reached after ~25 ps, having expelled ~3000 atoms. Subsequent stages in which the magnitude of the surface forces is increased (up to 9 kcal/mol/Å² in this case) progressively empty the protein volume further. At the end of the last stage of our test, about 1300 atoms remain inside the implicit volume; however, these are within ~1.5 Å of the protein surface (Figure 5B). As mentioned previously, this surface envelope the actual molecular surface with an offset equal to the probe radius (1.4 Å); hence, at this point in the simulation the volume to be occupied by the protein is in fact empty of lipid or water atoms.

Avoiding Bidirectional Pulling and Trapping of Lipid Tails. The complexity of many membrane protein structures often implies that the curvature of their surface changes drastically in length scales comparable to a lipid molecule. Other features such as pores and gullies create multiple, noncontiguous protein–lipid interfaces. In these cases, a potential shortcoming of the surface-guided repulsion approach is that lipids may be trapped inside the protein volume. This occurs when two sets of atoms in a lipid molecule are pushed in opposite directions, each toward the nearest region of the protein surface.

To overcome this difficulty in GRIFFIN, we implemented a feature we refer to as “force redirection”. This means, roughly speaking, that atomic forces whose direction is opposite to the forces applied to the rest of the lipid molecule are redirected appropriately (see Methodology for a more precise description). To illustrate this feature, we employ a second rotor ring, this time a c_{11} oligomer from a bacterial F-type ATP synthase²⁰ (Figure 1B).

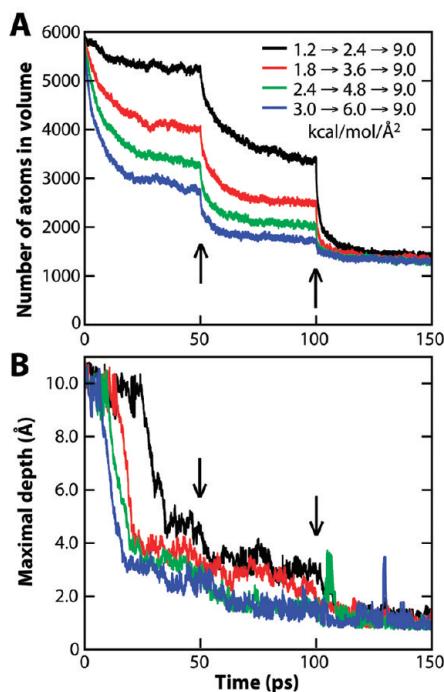


Figure 5. Progressive expulsion of phospholipid and water molecules from the volume of the K_{10} rotor ring due to GRIFFIN surface forces. (A) Number of atoms inside the protein volume, as a function of simulation time. The magnitude of the expelling surface forces calculated with GRIFFIN is increased in stages, at the time points marked with arrows. (B) Maximum depth inside the protein surface of any atom, as a function of simulation time. Hydrogen atoms are not considered in these or subsequent plots.

As shown in Figure 6A, several phospholipids in a membrane precarved for this protein span most of the distance between the inner and outer surfaces of the ring. Application of surface forces without the redirection functionality fails to expel these lipid tails from the protein volume; instead, as Figure 6B shows, the lipid tails are stretched in opposite directions. In this simple case, the situation can be somewhat resolved by increasing the magnitude of the surface forces (Figure 6D), as the lipid tail ultimately commits to one of the two surfaces. However, in general this will not be the case; moreover, large surface forces ($>3\text{ kcal/mol/}\text{\AA}^2$) should be avoided in the first stages of the procedure, as they otherwise result in unrealistic lipid-tail conformations, which subsequent simulations will be unlikely to correct.

By contrast, the redirection of problematic atomic forces rapidly resolves any conflicting lipid configurations, and allows GRIFFIN to empty the protein volume gradually and controllably (Figure 6C–E).

GRIFFIN Generates Lipid–Protein Interfaces with Interaction Specificity. An important improvement of GRIFFIN compared to its predecessor¹² is that it attains specificity in the protein–lipid interface, in addition to shape complementarity. To achieve this, the implicit protein force field overlaid on the explicit membrane system includes both electrostatic and van der Waals interactions; these act on lipid and water atoms outside the protein volume, in contrast to the surface forces, which only apply inside that volume (see Methodology; Figures 2 and 3).

Figure 7 illustrates both the degree of shape complementarity and the electrostatic specificity that may be attained with

GRIFFIN, using again the c_{11} rotor ring as an example. The repulsive surface-directed forces alone cause the lipid molecules to adapt neatly to the shape of the protein volume.¹² However, the electrostatic interactions in particular help to guide lipid groups to appropriate regions of the protein–membrane interface (Figure 7A cf. 7B), and thus reduce its overall electrostatic energy (Figure 7C). As illustrated in Figure 7D, during this optimization specific interactions are formed, for example, between the (explicit) polar head groups and oppositely charged atoms in the (implicit) protein. Arguably, that such interactions are already present in the starting configuration of a simulation is advantageous, as it implies that the subsequent equilibration of the system will be more realistic and efficient.

Geometric Objects for Exclusion or Inclusion of Specific Molecule Types in Desired Locations. Transmembrane pores and access pathways are characteristic features of membrane proteins that function as channels or transporters; many others contain large crevices in the lipid interface (Figure 1). Such cases present an additional challenge to the setup of a membrane protein simulation, since lipid molecules must be excluded from, e.g., hydrated pores, while water should be excluded from, e.g., lipid-filled crevices. To overcome these difficulties, GRIFFIN supports the use of geometric objects that define additional, molecule-type specific exclusion volumes. These objects are added to the protein volume during the initial membrane carving and/or during the subsequent MD/GRIFFIN simulation (see Methodology).

To illustrate this functionality, we consider as an example the trimeric carnitine/ γ -butyrobetaine antiporter CaiT from *Proteus mirabilis*,¹⁸ each protomer of which contains an aqueous vestibule leading toward the center of the membrane (Figures 1C and 8). Without using any additional exclusion volumes during the initial membrane carving, several lipid molecules are found inside these access pathways, amounting to ~ 160 atoms, not counting hydrogens (Figure 8A). Throughout a subsequent MD/GRIFFIN simulation—also without the objects—the majority of these atoms remain stubbornly in place (Figure 8B). If during the MD simulation, however, a spherical exclusion object is overlaid on each protomer so that it encompasses the access pathway (spheres, Figure 8), these regions are progressively emptied of lipid molecules (Figure 8, orange line). This result shows that the geometric objects are indeed integrated correctly into the protein volume, and that the expelling forces are calculated accordingly. Nevertheless, it should be noted that to use these additional exclusion volumes during the MD stage, but not the carving stage, implies that the overall area per lipid will be too small, particularly in the vicinity of the protein interface. Thus, it is recommended that the same objects be used also during the initial carving of the membrane. In our example, this excludes all but ~ 15 of the lipid atoms originally within those volumes (Figure 8C). At the end of the subsequent MD/GRIFFIN simulation, the access pathways in CaiT are entirely clear of lipid molecules (Figure 8D).

Performance and Portability. GRIFFIN is designed to be a stand-alone tool; i.e., it is intended to be compatible with any MD package, provided a minimal input/output interface. This interface, already included in, e.g., NAMD and GROMACS, allows for an external, accessory program (in this case GRIFFIN) to be executed at every time step of the simulation, to which the current coordinates of the molecular system are made available. If the program returns a set of atomic forces (and related quantities, e.g., energy, virial, etc), these will be added to the main algorithm and reflected in the simulation.

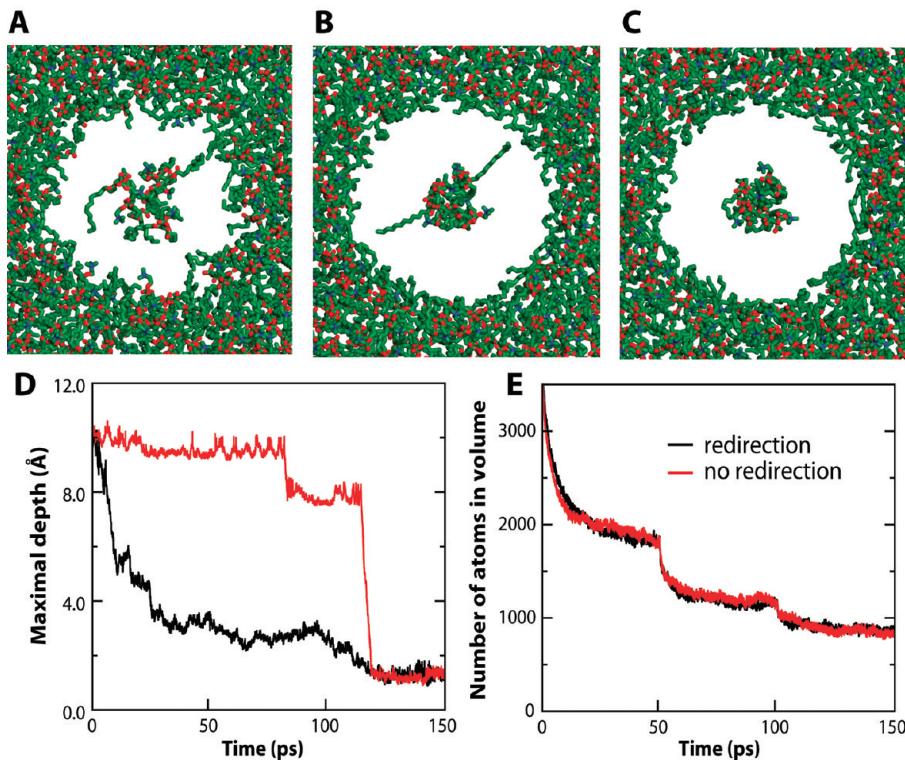


Figure 6. Effect of force redirection on the expulsion of lipid tails from the volume of the c_{11} rotor ring. (A–C) Several configurations of the lipids (green sticks) in the lipid bilayer, viewed along the normal to the membrane. Hydrogen and water atoms are omitted for clarity. Snapshots are taken (A) after carving, at $t = 0$ ps, and after the first stage of the GRIFFIN simulations, at $t = 100$ ps, either (B) without or (C) with force redirection. (D) The maximal depth inside the protein surface of any atom, as a function of simulation time. (E) Number of atoms inside the protein volume, as a function of simulation time. The magnitude of the forces was increased from 3.0 to 6.0 kcal/mol/Å² (at $t = 50$ ps) and 9.0 kcal/mol/Å² (at $t = 100$ ps). The force redirection angle was set to 110° (see Methodology).

As described previously, GRIFFIN precomputes a three-dimensional force grid before the start of the optimization phase, based on the membrane protein structure and its interactions with a probe particle positioned throughout the grid. During the GRIFFIN/MD simulation, the grid-point forces are interpolated and scaled to yield actual atomic forces. As mentioned, the initial force-grid calculation can be distributed among an adjustable number of computers, by portioning the grid accordingly. The computation of GRIFFIN atomic forces during run time can also be parallelized, using an MPI-compatible version of the program. As shown in Figure 9, the scalability of the atomic-force calculations is very good, especially for systems of about $\sim 100\,000$ atoms, such as the c_{11} ring. This scalability is, however, limited; the limit is imposed by the time required to exchange coordinates and forces with the MD software, which is a fixed overhead that depends on the computing infrastructure. Nevertheless, the total execution time, including the input/output of information, is still significantly reduced as the number of processor increases, at least within the range typically required. This allows GRIFFIN optimizations to be feasible within an affordable time frame even for very large systems such as AcrB, which comprises $\sim 300\,000$ atoms (Figure 10).

■ DISCUSSION

As new technologies lead to ever-growing computational power, and as algorithms improve in efficiency and accuracy, molecular dynamics simulations will be increasingly capable of

providing unique insights into the molecular mechanisms of membrane proteins. The strength of this theoretical tool—founded on statistical thermodynamics—lies in the atomic resolution at which the energetics and dynamics of the molecular system is simulated. These qualities, however, also pose a challenge in practice, as in most cases the structure of the environment of the membrane protein of interest is not known in atomic (or even molecular) detail. If this environment is modeled unrealistically to begin with, the results of the subsequent calculations will be questionable, as simulations are not designed to overcome large systematic errors. Therefore, it is generally advisable that a simulation research project comprises an initial stage focused not on the membrane protein but on the optimization of its lipid and solvent environment.

Current methodologies for preparing membrane protein simulations fall in two classes: those in which a lipid membrane is constructed around the protein of interest *de novo*, and those in which an existing lipid membrane model is somehow adapted. In the former class, one method uses a library of lipid conformations (derived from simulations of hydrated lipid bilayers) and progressively assembles individual molecules at designated locations around the protein; this is followed by a series of rigid-body rotations, translations, and energy minimizations.^{10,24–27} A second methodology in this class involves coarse-grained self-assembly simulations of a protein–lipid–water system, to which atomic detail is ultimately added.⁹

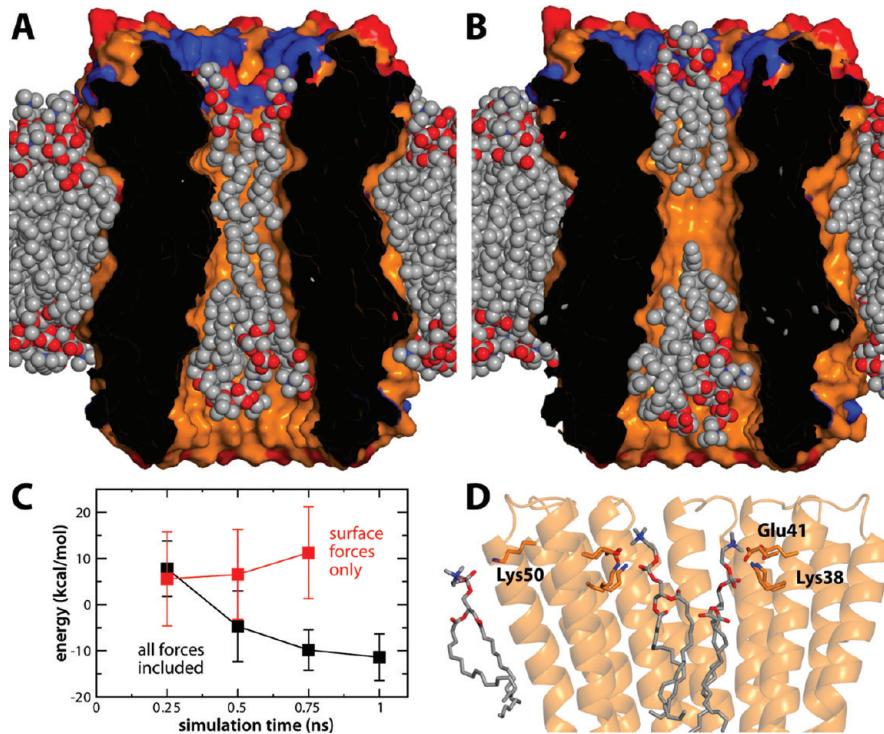


Figure 7. Shape and electrostatic complementarity between a hydrated lipid bilayer and the surface of the c_{11} rotor ring. (A) GRIFFIN-optimized interfaces between lipid (spheres) and protein (orange surface; basic and acidic residues in blue and red, respectively), after an extended GRIFFIN setup of the c_{11} ring. The system is viewed along the plane of the membrane at a cross section-through the center of the protein. Water molecules are omitted for clarity. (B) Same view as in (A), for an analogous GRIFFIN simulation in which only volume-exclusion forces were included. Note how the absence of electrostatic and van der Waals interactions with the protein causes the lipids in the inner pore to separate, driven by the gain in hydration by the surrounding waters. (C) Coulombic energy associated with interactions between charged side chains and lipid head groups, during the simulations in (A) and (B), shown as a function of simulation time. Note that no electrostatic complementarity develops when only volume-excluding forces are used. (D) Specific interactions between lipid head groups and charged side chains on the protein surface, in both the interior and the exterior of the ring, formed during the simulation in (A). The interaction distances, e.g., between amino and phosphate groups, approximately correspond to a hydrogen bond ($<3.5\text{ \AA}$).

In the class of adaptive methodologies, the general aim is to embed the membrane protein structure within an existing hydrated lipid membrane. Two recent approaches employ coordinate scaling schemes to accomplish this. In the first of those, the membrane is expanded by scaling up the intermolecular lipid distances; the protein is then inserted into the space thus created and the membrane is recompressed gradually.¹¹ In the second, it is the protein structure that is compressed (onto a one-dimensional object perpendicular to the membrane plane), inserted into the lipid bilayer, and progressively expanded.¹³ A third methodology in this class, used widely, has been to employ the surface of the protein as a template with which to carve and shape a cavity within the hydrated membrane system, into which the protein itself may be then inserted.¹²

All these methodologies are in principle valid, but they have different strengths and weaknesses. An advantage of the adaptive approaches relative to the de novo methods is that they build upon the work carried out to optimize atomic-resolution molecular simulation models of lipid systems, be they single components or mixtures, and bilayers, bicelles, or nanodiscs. By contrast, de novo approaches require reconstruction of all of the interactions in the system, that is, not only those between the protein and its environment but also lipid–lipid and lipid–water interactions. This may be an important shortcoming given the extremely long time scales characteristic of lipid motions.

For example, lipid rotation about its long axis has a time scale of $\sim 100\text{ ns}$.²⁸ Additionally, the library-based approach may be too time-consuming for large membrane proteins, unless coarse-grained simulations of self-assembly are used, although these can be unpredictable and are somewhat arbitrary in restoring the atomic detail. Within the adaptive class, on the other hand, the scaling methods are more limited than de novo approaches in the case of complex membrane topologies, e.g., those with lipid-filled pores or gullies. Lastly, the surface-based approach is also suitable for any protein topology, while sharing the advantages of the other adaptive methods. Nevertheless, in the original version of this protocol the protein was represented by its shape alone, and thus only nonspecific interactions between protein and its environment could be optimized.¹² Another practical disadvantage of the original surface-based methodology is that it required software from several different sources (some unsupported) and also lacked automation.

The methodology presented here, referred to as GRIFFIN, builds on the surface-based adaptive approach, removing its original shortcomings and adding new features to expand its versatility and transferability. First, the preparation of the optimization stage (surface calculation, precarving of the hydrated membrane, etc.) has been automated and integrated into a single program. Also, it is now possible to refine the exclusion volume defined by the protein surface by adding geometric objects,

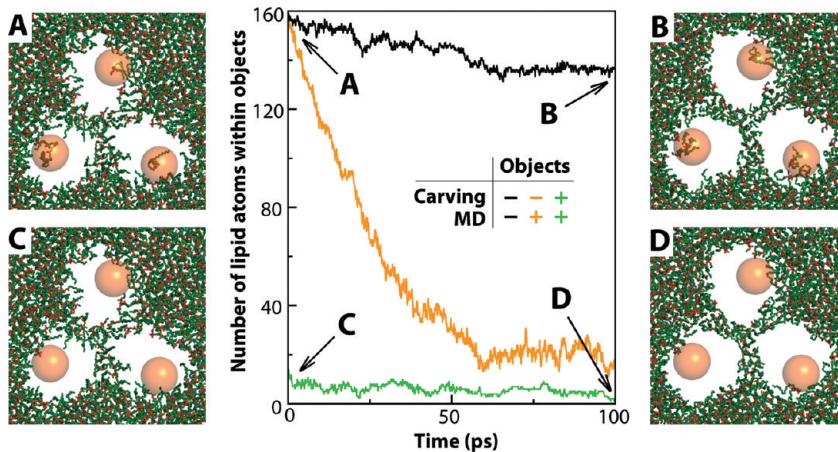


Figure 8. Use of geometric lipid-exclusion volumes to clear the aqueous substrate-access pathways in CaiT during GRIFFIN carving and MD simulations. The number of lipid atoms within three 22-Å-diameter spherical objects (spheres), positioned to encompass the access pathway in each of the three protomers, is plotted as a function of simulation time. Three cases are considered depending on whether the spherical exclusion volumes are used during the carving and the MD simulation stages. (A–D) The configuration of the lipid bilayer is depicted for (A) the precarved bilayer without objects; (B) at the end of the subsequent GRIFFIN/MD simulation, also without objects; (C) for the bilayer carved with objects; and (D) after subsequent GRIFFIN/MD simulations also using objects.

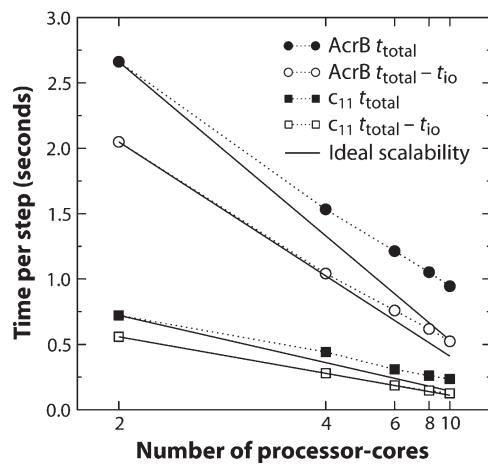


Figure 9. Performance and scalability of GRIFFIN, as a function of the number of processor-cores used. Calculation times per step are plotted for two membrane protein systems, namely the c_{11} ring and AcrB; both are embedded in a POPC membrane, comprising $\sim 80\,000$ and $\sim 320\,000$ atoms, respectively. Actual timings are shown with (t_{total}) and without ($t_{total} - t_{io}$) the overhead imposed by the input/output (i.e., coordinate/forces) exchange with the main MD program (dashed lines, symbols). These are compared with timings assuming ideal scalability relative to a two-core calculation (solid lines).

designed to be used as molecule type specific exclusion regions, e.g., for water-accessible pathways (as in CaiT; Figure 8) or lipid-filled cavities (as in AcrB or the c rings). In the actual optimization phase, which continues to involve a series of MD simulations, GRIFFIN now considers electrostatic and van der Waals interactions between protein and the hydrated membrane, alongside the expelling surface forces (Figures 2 and 3). These result in a much improved complementarity between the protein and its lipid and water environment, especially when the optimization is carried out gradually (Figure 7). This implies that subsequent simulations of the system will be more realistic and efficient. A lattice-based approach replaces the

original particle-based algorithm to meet the additional computational requirements due to the calculation of physical interactions during run time. To this end, a three-dimensional force grid including physical and exclusion forces is precomputed once, based on the protein structure (Figure 2). This implicit force field is overlaid on the explicit hydrated lipid membrane during the MD-based optimization; GRIFFIN thus derives actual atomic forces from the force grid at every step of the MD simulations and makes them available to the underlying MD engine (Figure 3F). Moreover, the code for the calculation of both the initial force grid and the time-dependent atomic forces is now written in C++ and parallelized; the scalability of the GRIFFIN force calculations is very good within the range of processor counts required in typical applications (Figure 9).

A practical caveat of the original surface-based method¹² was its limited compatibility, as it was integrated with a version of the MD software GROMACS that eventually became outdated. Therefore, the methodology has not been available to users of the recent versions of GROMACS or other MD packages. (Incidentally, lack of transferability among MD programs is a common limitation of the existing membrane-embedding methods mentioned above.) GRIFFIN, by contrast, is designed to be a stand-alone tool; it will work with any MD software that provides a minimal input/output (I/O) interface. This interface must be able to write out the coordinates of the explicit membrane system at every step; to execute a user-defined system command, e.g., to run GRIFFIN; and finally to read in and append a set of atomic forces (and related physical quantities) to the underlying MD algorithm. This procedure is clearly not optimal in terms of computational performance; however, as we have shown, the actual I/O overhead is not critical in practice even for systems of $\geq 100\,000$ atoms (Figure 9). Nevertheless, ongoing improvements are focused on the I/O protocol (and also on, e.g., memory management) since transferability is, in our view, paramount. At the present time, GRIFFIN can be readily employed alongside both NAMD 2.7¹⁵ and GROMACS v4.¹⁶ GRIFFIN and related materials will be made publicly available

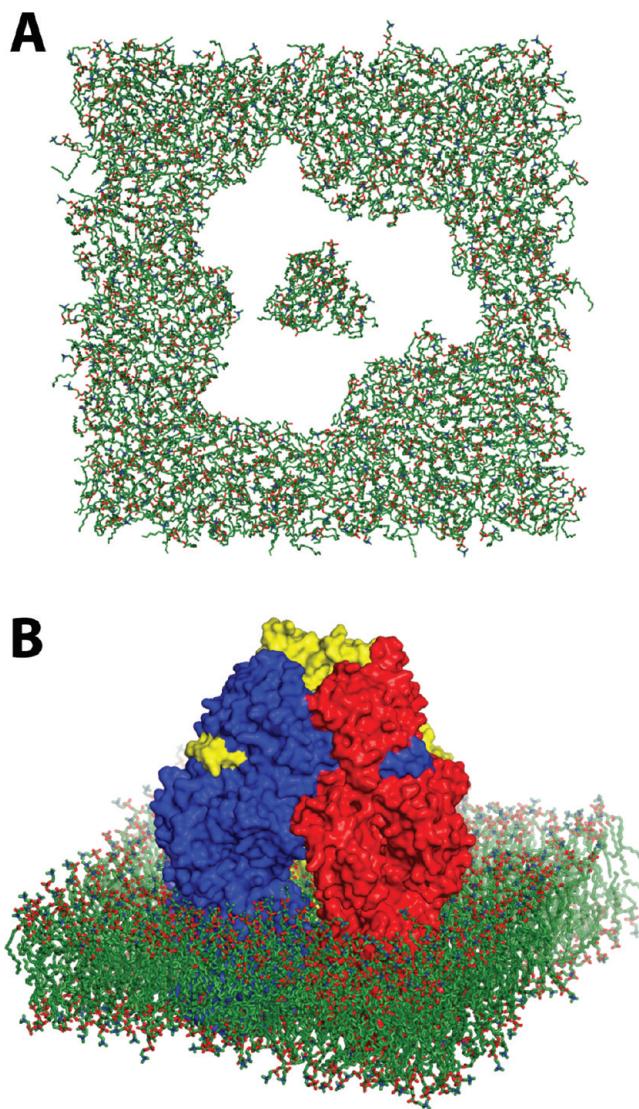


Figure 10. Molecular simulation model of AcrB in a lipid membrane. (A) The lipid membrane, following the GRIFFIN/MD optimization phase. (B) The complete system, after embedding and energy minimization of the actual AcrB structure into the GRIFFIN-optimized exclusion volume. Water is omitted for clarity.

for academic users at www.faradolab.org and www.forrestlab.org.

CONCLUSIONS

We have introduced a versatile and efficient methodology to prepare molecular dynamics simulations of membrane proteins, specifically aimed at optimizing the interface between the protein structure and its lipid environment. This methodology, based on a new grid-based simulation tool named GRIFFIN, is uniquely suited for membrane proteins of intricate topologies and irregular interfaces, such as those increasingly found among channels and transporters. Another advantage of GRIFFIN is that it builds upon already optimized simulation models of lipid–solvent systems, be they simple homogeneous bilayers, multicomponent membranes, or even nanodiscs. Lastly, GRIFFIN is a stand-alone tool that is designed to work alongside any existing molecular dynamics package, such as NAMD or GROMACS.

AUTHOR INFORMATION

Corresponding Author

*Phone: +49 69 6303 1600 (L.R.F.), +49 69 6303 1500 (J.D.F.-G.).
Fax: +49 69 6303 1502 (L.R.F.), +49 69 6303 1502 (J.D.F.-G.).
E-mail: lucy.forrest@biophys.mpg.de (L.R.F.), jose.faraldo@biophys.mpg.de (J.D.F.-G.).

ACKNOWLEDGMENT

We thank Gerrit Groenhof (MPI of Biophysical Chemistry), for his assistance with the GROMACS interface for GRIFFIN, and Wenchang Zhou (University of Konstanz), for his involvement in the AcrB simulations. This work was supported in part by the DFG Collaborative Research Center 807 “Transport and Communication across Biological Membranes” (R.S. and L.R.F.), the DFG Cluster of Excellence “Macromolecular Complexes” (J.D.F.-G.), and the Behrens-Weise-Stiftung (C.A.). Computational resources were in part provided by the Jülich Supercomputing Center.

REFERENCES

- (1) Wallin, E.; von Heijne, G. Genome-wide analysis of integral membrane proteins from eubacterial, archean, and eukaryotic organisms. *Protein Sci.* **1998**, *7*, 1029–1038.
- (2) Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **2001**, *305*, 567–580.
- (3) Gerstein, M. A structural census of genomes: comparing bacterial, eukaryotic, and archaeal genomes in terms of protein structure. *J. Mol. Biol.* **1997**, *274*, 562–576.
- (4) White, S. H. Biophysical dissection of membrane proteins. *Nature* **2009**, *459*, 344–346.
- (5) Ozcan, N.; Ejsing, C. S.; Shevchenko, A.; Lipski, A.; Morbach, S.; Kramer, R. Osmolality, temperature, and membrane lipid composition modulate the activity of betaine transporter BetP in *Corynebacterium glutamicum*. *J. Bacteriol.* **2007**, *189*, 7485–7496.
- (6) Swartz, K. J. Sensing voltage across lipid membranes. *Nature* **2008**, *456*, 891–897.
- (7) Vasquez, V. A structural mechanism for MscS gating in lipid bilayers. *Science* **2008**, *321*, 1210–1213.
- (8) Watt, I. N.; Montgomery, M. G.; Runswick, M. J.; Leslie, A. G. W.; Walker, J. E. Bioenergetic cost of making an adenosine triphosphate molecule in animal mitochondria. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 16823–16827.
- (9) Scott, K. A.; Bond, P. J.; Ivetac, A.; Chetwynd, A. P.; Khalid, S.; Sansom, M. S. P. Coarse-grained MD simulations of membrane protein-bilayer self-assembly. *Structure* **2008**, *16*, 621–630.
- (10) Woolf, T. B.; Roux, B. Molecular dynamics simulation of the gramicidin channel in a phospholipid bilayer. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11631–11635.
- (11) Kandt, C.; Ash, W. L.; Peter Tieleman, D. Setting up and running molecular dynamics simulations of membrane proteins. *Methods* **2007**, *41*, 475–488.
- (12) Faraldo-Gómez, J. D.; Smith, G. R.; Sansom, M. S. P. Setting up and optimization of membrane protein simulations. *Eur. Biophys. J.* **2002**, *31*, 217–227.
- (13) 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.
- (14) Shen, L.; Bassolino, D.; Stouch, T. Transmembrane helix structure, dynamics, and interactions: multi-nanosecond molecular dynamics simulations. *Biophys. J.* **1997**, *73*, 3–20.

- (15) Phillips, J.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.; Kale, L.; Schulten, K. Scalable molecular dynamics in NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- (16) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (17) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
- (18) Schulze, S.; Koester, S.; Geldmacher, U.; Terwisscha van Scheltinga, A. C.; Kuehlbrandt, W. Structural basis of cooperative substrate binding and Na^+ -independent transport in the carnitine/butyrobetaine antiporter CaiT. *Nature* **2010**, *467*, 233–237.
- (19) Murata, T.; Yamato, I.; Kakinuma, Y.; Leslie, A. G. W.; Walker, J. E. Structure of the Rotor of the V-Type Na^+ -ATPase from *Enterococcus hirae*. *Science* **2005**, *308*, 654–659.
- (20) Meier, T.; Krah, A.; Bond, P. J.; Pogoryelov, D.; Diederichs, K.; Faraldo-Gómez, J. D. Complete ion-coordination structure in the rotor ring of Na^+ -dependent F-ATP synthases. *J. Mol. Biol.* **2009**, *391*, 498–507.
- (21) Seeger, M. A.; Schiefner, A.; Eicher, T.; Verrey, F.; Diederichs, K.; Pos, K. M. Structural asymmetry of AcrB trimer suggests a peristaltic pump mechanism. *Science* **2006**, *313*, 1295–1298.
- (22) Connolly, M. L. Analytical molecular surface calculations. *J. Appl. Crystallogr.* **1983**, *16*, 548–558.
- (23) MacKerrell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCartney, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (24) Woolf, T.; Roux, B. Structure, energetics, and dynamics of lipid-protein interactions: A molecular dynamics study of the gramicidin A channel in a DMPC bilayer. *Proteins: Struct., Funct., Genet.* **1996**, *24*, 92–114.
- (25) Roux, B.; Woolf, T., Molecular dynamics of Pf1 coat protein in a phospholipid bilayer. In *Biological Membranes*; Merz, K. M. J., Roux, B., Eds.; Birkhauser: Boston, 1996; pp 555–587.
- (26) Petrache, H.; Grossfield, A.; MacKenzie, K. R.; Engelman, D.; Woolf, T. B. Modulation of glycoporin A transmembrane helix interactions by lipid bilayers: molecular dynamics calculations. *J. Mol. Biol.* **2000**, *302*, 727–746.
- (27) Jo, S.; Kim, T.; Im, W. Automated builder and database of protein/membrane complexes for molecular dynamics simulations. *PLoS One* **2007**, *2*, e880.
- (28) Klauda, J. B.; Roberts, M. F.; Redfield, A. G.; Brooks, B. R.; Pastor, R. W. Rotation of lipids in membranes: molecular dynamics simulation, ^{31}P spin-lattice relaxation, and rigid-body dynamics. *Biophys. J.* **2008**, *94*, 3074–3083.