Coarse-Grained Molecular Dynamics Simulations of the Energetics of Helix Insertion into a Lipid Bilayer[†]

Peter J. Bond,[‡,§] Chze Ling Wee,[‡] and Mark S. P. Sansom*[,‡]

Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K., and Max Planck Institute of Biophysics, Max-von-Laue Strasse 3, 60438 Frankfurt am Main, Germany

Received April 11, 2008; Revised Manuscript Received August 22, 2008

ABSTRACT: Experimental and computational studies have indicated that hydrophobicity plays a key role in driving the insertion of transmembrane α-helices into lipid bilayers. Molecular dynamics simulations allow exploration of the nature of the interactions of transmembrane α-helices with their lipid bilayer environment. In particular, coarse-grained simulations have considerable potential for studying many aspects of membrane proteins, ranging from their self-assembly to the relation between their structure and function. However, there is a need to evaluate the accuracy of coarse-grained estimates of the energetics of transmembrane helix insertion. Here, three levels of complexity of model system have been explored to enable such an evaluation. First, calculated free energies of partitioning of amino acid side chains between water and alkane yielded an excellent correlation with experiment. Second, free energy profiles for transfer of amino acid side chains along the normal to a phosphatidylcholine bilayer were in good agreement with experimental and atomistic simulation studies. Third, estimation of the free energy profile for transfer of an arginine residue, embedded within a hydrophobic α-helix, to the center of a lipid bilayer gave a barrier of ∼15 kT. Hence, there is a substantial barrier to membrane insertion for charged amino acids, but the coarse-grained model still underestimates the corresponding free energy estimate (∼29 kT) from atomistic simulations (Dorairaj, S., and Allen, T. W. (2007) *Proc. Natl. Acad. Sci. U.S.A. 104*, 4943−4948). Coarse-grained simulations were then used to predict the free energy profile for transfer of a simple model transmembrane α-helix (WALP23) across a lipid bilayer. The results indicated that a transmembrane orientation was favored by about −70 kT.

Integral membrane proteins are essential to cell function, and account for ∼25% of open reading frames in most genomes. However, to date only ∼120 distinct, high-resolution structures of membrane proteins have been determined (see http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html for a summary), due to difficulties in overexpression and crystallization. Despite this relative paucity of structural data, and the incompleteness of our knowledge of the interactions between membrane proteins and lipid molecules, progress has been made in experimental (*1*) approaches to understanding the folding and stability of membrane proteins within a bilayer environment. However, in order to exploit this knowledge to predict the structure of novel membrane proteins, advances in computational approaches are required.

The location of a stably folded membrane protein is determined by a complex balance of interactions, including those made with the bilayer core, the interface, and bulk solvent (*2*). While experimental investigation of the partitioning energetics of membrane proteins within bilayers has been hampered by their resistance to reversible folding/unfolding, the free energies of partitioning of their component residues can be estimated by measuring the distribution of amino acid analogues or short peptides (*3*−*5*) between different phases. More recently, Hessa et al. have followed the insertion of a series of designed protein constructs into the ER (endoplasmic reticulum) membrane (*1*). The resultant "biological hydrophobicity" scale suggests that membrane protein insertion is determined, to a first approximation, by the sum of the hydrophobicity of the component amino acids, and also revealed a strong dependence of partitioning on position within the helix for some residues (*1*), consistent with statistical analyses of available high-resolution membrane protein structures (*6*).

It is therefore clear that the overall hydrophobicity/hydrophilicity is a primary driving force in transmembrane (TM[1]) helix insertion. Nevertheless, it is clear that other factors may be important, such as protein flexibility (e.g., helix kinking) and membrane elasticity/deformability (*7*), both of which may lead to helix/bilayer mismatch and/or helix tilting.

[†] This work was supported by the BBSRC, the EPSRC, and the Wellcome Trust. P.J.B. is an EMBO fellow.

* To whom correspondence should be addressed. E-mail: mark.sansom@bioch.ox.ac.uk. Phone: +44-1865-275371. Fax: +44-1865-275273.

[‡] University of Oxford.

[§] Max Planck Institute of Biophysics.

[1] Abbreviations: AT-MD, atomistic molecular dynamics; CG-MD, coarse-grained molecular dynamics; DPPC, dipalmitoyl-phosphatidylcholine; MD, molecular dynamics; PMF, potential of mean force; TI, thermodynamic integration; TM, transmembrane; WHAM, weighted histogram analysis method.

10.1021/bi800642m CCC: $40.75
Published on Web 10/02/2008

Molecular dynamics (MD) simulations provide complementary information to biophysical measurements of partitioning, allowing one to explore the detailed interactions of membrane proteins with their associated lipid bilayer environment (*8*). However, at present, all-atom MD studies are somewhat limited by a practical upper simulation time limit of ~100 ns, which hampers extensive sampling of such complex systems and hence reduces the range of such simulations (*9*−*11*).

In contrast, coarse-grained (CG-MD) simulations (*12*−*18*), in which small groups of atoms are treated as single particles, enable longer timescales to be addressed. Several approaches have been applied to CG models for polymers, lipids, and proteins (*19, 20*). Early folding studies utilized Gō-like models (*21*) based on native protein contacts and spawned a variety of structure-based CG models. Knowledge-based statistical methods also provide a general approach to deriving CG potentials, for prediction of e.g. protein structure or location of proteins within an implicit membrane (*6*). Potentials can also be derived for CG particles based on atomistic functional groups. An empirical functional form for such a model may be parametrized on the basis of all-atom simulations. This may be achieved by iteratively refining the CG potential to achieve a fit with e.g. radial distribution functions from atomistic simulations (*22*). Alternatively, a "force-matching" procedure (*23*), in which forces derived from an atomistic MD simulation are systematically used to obtain effective pairwise CG potentials, has been applied to pure and mixed (*24*) lipid systems, as well as to a mixed atomistic-CG protein/lipid system (*15*). Finally, interparticle CG potentials may be derived by calibration against the physiochemical and thermodynamic properties of atomistic functional groups to which the CG particles correspond. Such an approach has recently been used for development (*12*) and subsequent refinement (*18*) of the lipid CG force field MARTINI.

We recently adapted this model for application to membrane proteins, in which one CG particle is used to represent the peptide bond (*17, 25*), and one or two particles represent the side chain (with zero particles for Gly), yielding 13 different residue types for the 20 amino acids. Similar CG models have been used to study membrane protein self-assembly (*26, 27*) and ion channel gating (*28*). Our model was primarily developed to simulate self-assembly processes associated with membrane proteins, and hence calibrated on the basis of structural data extracted from atomistic MD simulations and biophysical experiments. The model was subsequently shown to be successful in simulating membrane/protein self-assembly (*17, 25*) and insertion (*29, 30*), consistent with simulation and experimental data, and also in agreement with statistical analyses of membrane protein structures (*31*).

Despite the evident successes of the CG approach to membrane proteins, it is important to evaluate its quantitative accuracy (*32*), especially in the context of ongoing attempts to improve the parametrization of amino acid/lipid interactions for atomistic force fields (*10*). Monticelli et al. recently extended the MARTINI force field to proteins, calibrating the oil/water partitioning behavior of CG amino acid side chain particles, and revealing good correlation for free energy profiles across a lipid bilayer with atomistic simulations (*33*). With a similar philosophy, we quantitatively evaluate and refine our protein/lipid CG model, and also focus on the energetics of interaction of charged amino acids and helices with the membrane environment. The CG model is evaluated in three increasingly complex models of insertion of components of a membrane protein in a lipid bilayer. First, we consider individual amino acid side chains. Our original set of CG amino acid parameters (*17, 25*) is used, and iteratively modified where necessary, to calculate amino acid oil/water partition free energies, achieving good agreement with experimental measurements. The chosen CG amino acid parameters are then used to calculate the free energies of transfer of each amino acid across a dipalmitoyl-phosphatidylcholine (DPPC) bilayer. Second, we consider the free energy profile (i.e., potential of mean force; PMF) for translation of a long (101 residue) oligo-leucine α-helix containing a single arginine (LeuN.Arg) residue relative to a DPPC bilayer. A very similar system has been explored via atomistic MD simulations (*34*) as a simple model of the bilayer/lipid interactions formed by the voltage sensor domain of Kv channels (*35*) and thus is of some biological importance. Third, we calculate a free energy profile across a DPPC bilayer for WALP23, a simple model peptide TM α-helix which captures many of the properties of canonical TM helices in more complex membrane proteins (*36*). By exploring this range of systems we are able to show that a CG model is indeed capable of providing semiquantitative insights into the structural principles of membrane proteins.

METHODS

Simulation Details. Simulations were performed using GROMACS (www.gromacs.org) (*37*). CG simulations were performed as described in ref *25*, with CG parameters for lipid molecules (dipalmitoyl-phosphatidylcholine, DPPC), Na^+ and Cl^- ions, and water molecules as in ref *12*. Lennard-Jones interactions were shifted to zero between 0.9 and 1.2 nm, and electrostatics were shifted to zero between 0 to 1.2 nm, with a relative dielectric constant of 20. The non-bonded neighbor list was updated every 10 steps. All simulations were performed at constant temperature, pressure and number of particles. In the amino acid side chain and WALP23 PMF simulations, the temperatures of the protein, lipid, and solvent were each coupled separately using the Berendsen algorithm (*38*) at 323 K with a coupling constant $\tau_T = 1$ ps. The system pressure was semi-isotropically coupled in the bilayer plane (*xy*) and along the bilayer normal (*z*), using the Berendsen algorithm at 1 bar with a coupling constant $\tau_P = 1$ ps and a compressibility of 5×10^{-6} bar^{-1} in each direction. The LeuN.Arg PMF simulations were identically temperature and pressure coupled but at 310 K and with smaller coupling constants $\tau_T = 1$ ps and $\tau_P = 1$ ps. In thermodynamic integration simulations, the temperatures of the amino acid analogues and solvent were coupled separately using the Nosé−Hoover algorithm (*39, 40*) at 298 K with a coupling constant $\tau_T = 5$ ps. The system pressure was isotropically coupled using the Parrinello−Rahman algorithm (*41, 42*) at 1 bar with a coupling constant $\tau_P = 5$ ps and a compressibility of 1×10^{-5} bar^{-1}. The time step for integration was 40 fs. Analyses were performed using GROMACS tools and locally written code. Visualization used VMD (*43*).

CG Protein Parameters. The equilibrium bonded parameters (bond and angle potentials) for backbone and side chain

particles were as in ref *17*. The coordinates for the CG WALP23 peptide model (sequence GWW-$(LA)_8$-LWWA) were generated from an atomistic structure produced by Modeler, using α-helical restraints for all but the terminal residues as described in ref *25*. Similarly, the coordinates for the CG LeuN.Arg (sequence $L_{50}RL_{50}$) and LeuN (sequence L_{101}) helices were generated from atomistic model structures using α-helical restraints for all residues. The CG peptide was composed of a chain of backbone particles with attached side chain particles.

CG Protein Bonded Parametrization. Protein bonds were all harmonically restrained with a force constant of 5000 kJ mol^{-1} nm^{-2}. Equilibrium bond lengths and angles were chosen on the basis of comparison to a selection of high-resolution membrane protein crystal structures (though little variation is obtained when compared to soluble proteins). The equilibrium bond length between backbone particles was set to 0.36 nm independent of secondary structure, compared to a value of 0.35 nm in MARTINI (*33*).

Side chain angles for all residues with two side chain particles were harmonically restrained with a force constant of 250 kJ mol^{-1} rad^{-2} and an equilibrium angle between particles CA−CB−CG of 180° (Arg/Lys), 54° (Trp), or 60° (Phe, Tyr, His). Bond lengths involving side chain particles were chosen by estimating the mean distance between the Cα atom of an amino acid and the center of mass of the heavy side chain atoms (weighted by the number of atoms, assuming a 4:1 atom:CG particle ratio), which the respective CG side chain particle represents. In this way, the CG amino acids were expected to accurately match the size and surface area of the equivalent atomistic structures. Equilibrium bond lengths were as follows: Ala = 0.09 nm (CA−CB); Cys = 0.17 nm (CA−CB); Asp = 0.26 nm (CA−CB); Glu = 0.36 nm (CA−CB); Phe = 0.42 nm (CA−CB, CA−CG, CB−CG); His = 0.38 nm (CA−CB, CA−CG, CB−CG); Ile = 0.26 nm (CA−CB); Lys = 0.31 nm (CA−CB), 0.21 nm (CB−CG); Leu = 0.28 nm (CA−CB); Met = 0.31 nm (CA−CB); Asn = 0.26 nm (CA−CB); Pro = 0.19 nm (CA−CB); Gln = 0.36 nm (CA−CB); Arg = 0.37 nm (CA−CB), 0.25 nm (CB−CG); Ser = 0.16 nm (CA−CB); Thr = 0.20 nm (CA−CB); Val = 0.20 nm (CA−CB); Trp = 0.47 nm (CA−CB), 0.38 nm (CA−CG), 0.28 nm (CB−CG); Tyr = 0.38 nm (CA−CB, CA−CG, CB−CG). Comparing with MARTINI, and excluding ring-containing amino acids (which are treated differently here), equilibrium bond lengths are highly correlated, but ∼0.05 nm shorter here. We found our bonded parameters to most accurately reproduce protein−protein interactions at tight interfaces, as in ref *26*. Irrespective, control thermodynamic integration simulations (below) revealed little variation (i.e., significantly less than the error in individual calculations, <0.5 kT) in partition free energies upon significant variations (i.e., between half to twice) in the equilibrium bond length.

All angles for the protein backbone particles were harmonically restrained with a force constant of 250 or 350 kJ mol^{-1} rad^{-2} for regions of random coil or α-helix, respectively. Equilibrium bond angles were 90° for α-helical segments, 120° for random coil regions, and 130° for β-structures. This may be compared to similar values of 96°, 127°, and 134° in MARTINI (*33*).

CG Protein Secondary Structure Restraints Parametrization. In WALP23, the distance between backbone particles was restrained to mimic secondary structure hydrogen bonds (H-bonds) in the atomistic structure, using a harmonic distance restraint with an equilibrium length of 0.6 nm and a force constant of 1000 kJ mol^{-1} nm^{-2}. In the LeuN.Arg and LeuN helices, backbone particles which were within 3.0 nm of each other were restrained to their initial distances with an identical force constant of 1000 kJ mol^{-1} nm^{-2}. This was necessary to maintain a rigid α-helical structure similar to that employed in published atomistic simulations (*44*) and to avoid the introduction of noise during sampling.

A value of 1000 kJ mol^{-1} nm^{-2} was chosen as the force constant for distance restraints on the basis of iterative comparison of protein backbone root-mean-squared fluctuations (RMSF), a measure of flexibility, for a number of long-time scale atomistic and CG simulations. Using this force constant with the secondary structure H-bonding harmonic distance-restraints approach (*17, 25*), excellent correlation was achieved between RMSF profiles calculated for simulations of simple α-helical peptides or β-barrel proteins with long, flexible connecting loops. Similar results were obtained using the same force constant with an elastic network model (ENM) approach for complex α-helical multipass TM proteins (*17*), and for other protein test cases (*26, 27, 29*). Critically, in each case, the magnitudes of the atomistic and CG RMSF simulation profiles were identical. When a value appreciably below 1000 kJ mol^{-1} nm^{-2} was used for the structural-restraint force constant, significant unfolding and/or distortion of the protein structure was observed, while the root-mean-squared deviation (RMSD) of the backbone structure rose above that typically expected from a normal atomistic membrane protein simulation. Above 1000 kJ mol^{-1} nm^{-2}, no improvement in the correlation of structural fluctuations or conformational deviations between atomistic and CG protein simulations was observed. Much above 5000 kJ mol^{-1} nm^{-2}, numerical instabilities arose in the simulation using a given integration time step.

CG Protein Non-Bonded Parameterization. As described in ref *12*, only four CG particle types (all with the same effective diameter of 0.47 nm) are distinguished, namely "polar" (**P**), "mixed polar/apolar" (**N**), "hydrophobic apolar" (**C**), or "charged" (**Q**) groups, along with further subtypes for the **N** and **Q** particles which allow fine-tuning of Lennard-Jones interactions to reflect hydrogen-bonding capacities ($\mathbf{N_0}$/$\mathbf{Q_0}$ = no H-bonding; $\mathbf{N_d}$ = H-bonding donor; $\mathbf{N_a}$ = H-bonding acceptor; $\mathbf{N_{da}}$ = H-bonding donor and acceptor). The non-bonded interactions between these particles are described by a Lennard-Jones (LJ) potential with five levels of interaction, and the hydrogen-bonding subtypes for **N** and **Q** particles simply modulate this LJ interaction with other particles. **Q** particles also interact via the standard electrostatic Coulombic potential.

The constituent particles for each CG amino acid side chain were initially selected based on the hydrophobicity/hydrophilicity, partial charges, and hydrogen bonding capacity of their component atoms, as detailed in ref *25*. The transfer free energies between polar and apolar solvents were estimated using this initial CG parameter set, by carrying out thermodynamic integration simulations. Following this initial "calibration" stage, the resulting transfer free energies were then used to refine the parameters for some amino acid side chains, guiding the implementation of one or more new particle types. For side chains with new parameters, ther-

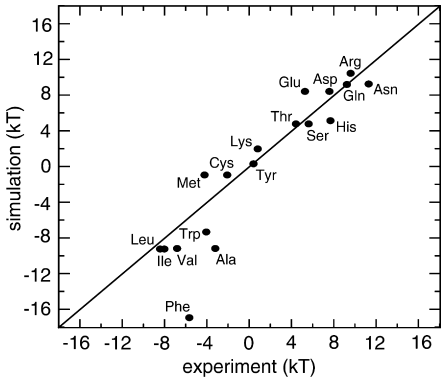

FIGURE 1: Correlation between experimental and CG-MD simulated free energies of transfer for *neutral* amino acid side chain analogues from water (water, CG particle type **P**) to nonpolar solvent (cyclohexane, CG particle type **C**).

modynamic integration calculations were repeatedly performed. In this way, the fit between calculated and experimental free energies (Figure 1) was iteratively improved. The final particle types assigned to each amino acid side chain are as follows: Ala, Ile, Leu, Pro, Val = **C**; Phe = **C** + **C**; Cys, Met = $\mathbf{N_d}$; Ser, Thr = $\mathbf{N_{da}}$; Asn, Gln = **P**; Tyr = **C** + **P**; His = $\mathbf{N_d} + \mathbf{N_{da}}$; Trp = **C** + $\mathbf{N_d}$; Asp, Glu = $\mathbf{Q_a}$ (charge −1); Lys = **C**+$\mathbf{Q_d}$ (charge +1); Arg = $\mathbf{N_d} + \mathbf{Q_d}$ (charge +1). Their free energies of transfer are detailed in Table S1 in the Supporting Information.

It should be noted that the our particle assignment for amino acids is based on the original Marrink CG lipid force field, rather than the MARTINI parameter set, which introduced more particle interaction levels to improve the partitioning free energies of model compounds, but retained the main features of the force field, i.e. a four-to-one mapping of atoms to particles, and use of only four main types of interaction site. A comparison of the transfer free energies presented here with those of MARTINI (*33*) (Table S2 in the Supporting Information) reveals rather close agreement, with an unsigned difference of ∼2 kT, i.e. comparable to the unsigned error with experimental free energies estimated by both groups. This is unsurprising, given that the amino acid parameter sets are similar, i.e. apolar amino acids are represented by **C**-type particles, polar uncharged amino acids by **P**-type particles, small negatively charged side chains by negative **Q**-type particles, and positively charged side chains by positive **Q**-type particles combined with uncharged particles.

The difference in treatment of ring-based side chains should be noted. Here, we chose to treat the ring-based side chains with a combination of *two particles*, bonded to each other and to a single backbone particle, hence forming a simple triangular structure which roughly approximates the planar element of the side chain ring, and provides the necessary membrane-partitioning and side chain orientation behavior for the amphipathic aromatic amino acids, as described in the Results section. On the other hand, Monticelli et al. (*33*) utilize an alternative, smaller ring particle introduced in the newer MARTINI forcefield (*18*), thus allowing more ring-structure detail with the use of *3*−*4 particles* for the ring side chain. Our own testing of such treatment of rings in previous studies (see e.g. ref *17*) suggested that while some improvement in side chain detail is achieved by using three side chain particles, no improvement is observed in the structural detail of protein−protein interactions, or in the effects of e.g. model CG sterol molecules on lipid tail ordering. Attempts to improve this, such as adding bond-constraints to remove out-of-plane distortions, were unremarkable and also required reduction of the maximum integration time step, necessary to avoid numerical instabilities arising from fast fluctuations. Indeed, a maximum time step of 25 fs is reported by Monticelli et al., compared with 40 fs here. Hence, we decided that the improvement in simulation speed sufficiently outweighed the problems related to reduced detail in the ring geometry.

Particle type $\mathbf{N_d}$ was used for backbone particles within the WALP23 helical region, since our PMF calculations (below) show that its free energy for insertion into the center of a DPPC membrane is slightly unfavorable (by ∼1−1.5 kT). This value is consistent with experimental estimates for bilayer partitioning of hydrogen-bonded peptide bonds (*4, 5*). In the LeuN.Arg and LeuN helices, the backbone particle types were assigned based on the participation of the constituent atoms in hydrogen bonding as determined from the atomistic structures (*25*). Thus, the N-terminal 4 residues were assigned backbone particle type $\mathbf{N_d}$ and the C-terminal 4 residues assigned backbone particle type $\mathbf{N_a}$, with all other residues assigned type $\mathbf{N_0}$. It should be pointed out that at both limits of the sampled range along the reaction coordinate, L4 and L98 are at the limits of the non-bonded interaction cutoffs with respect to the bilayer so we expect the LeuN.Arg/LeuN systems to be roughly translational invariant.

Thermodynamic Integration. The partition free energies for CG amino acid analogues were calculated by estimating the solvation free energies (ΔG) in both polar and nonpolar solvents. To calculate each of these, the thermodynamic integration (TI) formula was applied (*45*), where the Hamiltonian is a function of coupling parameter λ:

$$\Delta G = \int_{\lambda=0}^{1} \left\langle \frac{\mathrm{d}H(\lambda)}{\mathrm{d}\lambda} \right\rangle_{\lambda} \mathrm{d}\lambda$$

Thus, the non-bonded interactions between the amino acid analogue and solvent particles were gradually abolished (i.e., the analogue was converted to dummy particles, $\lambda = 0$). While the non-bonded terms were turned off, the masses and bonded terms remained the same. The ensemble average for each amino acid analogue was evaluated at 21 discrete λ-points by performing separate CG-MD simulations with λ increments of 0.05. The integral was subsequently determined numerically.

Errors at each λ-point were estimated using block averaging; these were integrated to obtain an estimate of the total error in ΔG. To avoid singularities, a soft-core potential function was used, with $\alpha = 1.51$ and $\sigma = 0.3$ nm as in ref *45*. The accuracy of the solvation free energy calculations was confirmed by performing control TI simulations in which: greater numbers of λ-points were used; soft-core potential variables were changed; and the TI pathway was reversed (i.e., λ was gradually changed from 0 to 1). In each case, variations in the average ΔG of no more than 0.5 kT were observed.

Thermodynamic integration CG-MD simulations were carried out for each amino acid analogue in both polar (CG particle type **P**) and nonpolar (CG particle type **C**) (*12*) solvent. Only neutral side chain analogues were considered,

so that the $\mathbf{Q_a}$ particles of Asp/Glu and $\mathbf{Q_d}$ particles of Lys/Arg were assigned a charge of 0. All systems contained 1 amino acid analogue and 711 solvent particles, in a cubic box of linear dimension 4.4 nm (nonpolar solvent system) or 4.6 nm (polar solvent system). Prior to simulation, the system was energy minimized using <100 steps of the steepest descent method, to relax any steric conflicts between the amino acid analogue and solvent. Each λ-point simulation was carried out for 400 ps, the first 100 ps of which were discarded as the equilibration period. Using this method, the free energies of partitioning for each side chain analogue were obtained as detailed in Table S1 in the Supporting Information.

It should be noted that a similar approach to parametrization via oil/water partition free energies was taken recently by Monticelli et al. (*33*). However, instead of applying TI, energies were calculated by estimating the equilibrium densities of low concentrations of CG particles dissolved in a water/butane two-phase system. Using this approach, the concentration of the amino acid of interest must approach infinite dilution. This perhaps has the advantage that it seems more physically comprehensible, comparable to experimental measurement of distribution coefficients between different phases. However, for some residues, extremely long timescales may be required to achieve reliable free energy estimates. In some cases they will not achieve convergence (*33*); the density of the most polar amino acids is presumably vanishingly small in the nonpolar phase. For this reason, we chose the TI method to estimate partition free energies. Nevertheless, both approaches are valid, and encouragingly, for common CG particles, very similar results are obtained.

Potential of Mean Forces via Umbrella Sampling. PMFs were calculated by carrying out a series of umbrella sampling simulations, before unbiasing with the weighted histogram analysis method (WHAM) (*46*). The parameters for the PMF calculations of individual amino acid side chain analogues, and the WALP, LeuN.Arg and LeuN helices, were the refined amino acid CG parameters whose partition free energies best matched with experiment, calculated in the previous stage by thermodynamic integration.

For the amino acid analogues, starting configurations for each umbrella window were obtained by simple superposition of each amino acid analogue into the center of a preassembled bilayer system, which was produced in a previous simulation (*25*). For the WALP peptide and the LeuN.Arg and LeuN helices, a few lipids were removed prior to superposition to reduce steric overlap between lipid and protein. For umbrella sampling, the center of mass of the amino acid/WALP peptide/LeuN.Arg and LeuN helices was harmonically restrained to subsequent positions, each separated by $\Delta z = 0.1$ nm, along the membrane normal, with a restraining force constant of 1000 kJ mol^{-1} nm^{-2}. For each amino acid analogue, 85 windows were used with limits of +4.2 and −4.2 nm with respect to the bilayer center, ensuring good sampling into the bulk solvent region. For the WALP peptide, 121 windows were used between +6.0 and −6.0 nm with respect to the bilayer center; the increased number of windows ensured that the minimum distance between peptide and bilayer was beyond non-bonded interaction cutoffs. For the LeuN.Arg and LeuN helices, 80 windows were used between +4.0 and −4.0 nm with respect to the bilayer center. At the limits, the Arg residue was located in bulk solvent beyond the non-bonded interaction cutoffs. The choice of umbrella sampling parameters allows the molecule of interest to diffuse into neighboring windows, as confirmed by analysis of sampling histograms. In the LeuN.Arg/LeuN PMF simulations, each helix had a tendency to "collapse" into the bilayer to minimize exposure of the hydrophobic segments of the helix to the polar water particles. To prevent this, the backbone particles of the terminal residues were restrained in x and y to ensure the helix remained transmembrane relative to the bilayer.

For the amino acid analogues, the final system had dimensions $6.4 \times 6.4 \times 10.9$ nm^3, and consisted of 1 amino acid analogue molecule, 128 DPPC lipids, and 2181 CG water particles (for charged amino acid analogues, 2 CG "solvated ion" particles of charge ±0.5 e replaced 2 water particles, to maintain overall system electroneutrality). For the WALP peptide, the final system had dimensions $9.0 \times 9.0 \times 12.1$ nm^3, and consisted of 1 peptide, 252 DPPC lipids, and 5139 CG water molecules. For the LeuN.Arg and LeuN helices, the final system had dimensions $12.2 \times 10.5 \times 27.9$ nm^3, and consisted of 1 helix, ~410 DPPC lipids, and ~25000 CG water molecules (similarly, 2 CG "solvated ion" particles of charge −0.5 e were added for LeuN.Arg). Prior to simulation, each window system was energy minimized, using <500 steps of the steepest decent method, to relax any steric conflicts between protein, lipid and solvent. For the amino acid analogues, each window simulation was run for 25 ns, the initial 5 ns being discarded as equilibration time. For the WALP peptide, each window simulation was run for 50 ns, the initial 20 ns being discarded as equilibration time. The longer equilibration time for the peptide primarily reflects the greater configurational phase space available to the helix as it rotates around its restrained center of mass. For the LeuN.Arg and LeuN helices, each window simulation was run for 20 ns with the initial 5 ns discarded as equilibration time. Unlike the rotational freedom available to WALP23 about its center of mass, the LeuN.Arg helix was less complex with the crucial degree of freedom being the conformational space explored by the Arg side chain, thus allowing for a shorter simulation duration. A single PMF comprised 1.7, 3.6 or 1.2 μs of simulation time, for an amino acid analogue, WALP peptide, or LeuN.Arg/LeuN helix, respectively. This is equivalent to ~30−90 million configurations of the respective sampled molecule (at a simulation time step of 40 fs).

Histograms were unbiased using WHAM, with 200−400 bins and a tolerance of 10^{-5} kT for window offsets. PMFs were converged with respect to these variables, as established by calculating control PMFs with a variety of bins and tolerance values. PMFs were converged with respect to number of simulation windows and equilibration time, as established by block analysis, whereby PMF curves calculated over the first and second halves of equilibrated simulations overlapped. Error bars for each PMF window represent the difference between these two halves. It should be noted that while the one-dimensional PMF is normally ill-defined in bulk regions where the motion of a test molecule is not bounded orthogonal to a reaction coordinate, the use of a coarse-grained representation and long simulation times here enabled sufficient sampling to provide a converged (and symmetrical) PMF in the lipid bilayer and in bulk solvent.

RESULTS

The aim of the current study is to evaluate the quantitative accuracy of the CG-MD method with respect to prediction of insertion of α-helices into lipid bilayers. This is of especial interest in the context of the ongoing debate concerning the free energy of insertion of α-helices containing charged residues such as the Arg-containing S4 helix of voltage-sensitive ion channels (*29, 32, 35, 44, 47*).

To this end we perform a series of free energy calculations on systems of increasing complexity. First, we compare simulated (CG-MD) and experimental partitioning free energies for neutral amino acid side chain analogues between a polar and a hydrophobic solvent. Having established the accuracy of the CG side chains, we next apply the umbrella sampling method to calculate the PMF for each CG amino acid analogue, using the newly parametrized CG amino acids, to obtain their free energy profile along the bilayer normal. This permits comparison with equivalent distributions from statistical analyses of membrane protein structures. Having thus evaluated our model vs experimental observations we explore the energetics of two model systems. The first is an oligo-Leu α-helix containing a single Arg residue (LeuN.Arg) which has been explored as a simple model of a TM helix containing a basic side chain in detailed atomistic simulations (*34*). The second is an idealized model of a canonical TM α-helix, WALP23, which consists of a hydrophobic core (of alternating Ala and Leu residues) and which is capped at either end with Trp residues. The WALP peptides have been explored in some detail as model TM helices both by experiment (*36, 48*) and by atomistic simulation (*49*).

Free Energy of Partitioning for Side Chain Analogues. Partition free energies for neutral side chain analogues were calculated by estimating their solvation energies in water (CG particle type **P**) and in a nonpolar alkane solvent (CG particle type **C**) via thermodynamic integration (Table S1 in the Supporting Information). Comparison with experimental transfer free energies for transfer from water to cyclohexane (*3, 6*) yielded a correlation coefficient of 0.93 for all amino acids (Figure 1), and an average unsigned error of ~2.3 kT. This is similar to errors from all-atom (AT) forcefields such as OPLS, and almost twice as small as the error obtained for GROMOS96 (*10*). It also comparable to the ~2 kT difference estimated from equilibrium densities of low concentrations of CG particles dissolved in a water/butane two-phase system, obtained with the MARTINI forcefield (*33*).

In previous all-atom studies, transfer from water to cyclohexane for most side chains was more favorable than expected compared to experimental values (*10*). We observed this only for some of the hydrophobic residues, suggesting that the interaction between nonpolar particles may be too strong. This is primarily evident for the side chain of Phe, where simultaneous representation of its size, and limited hydrophobicity estimated by partitioning experiments, proved difficult. This is because of the absence of the unsaturated aromatic ring in the CG model. Experimentally, aromaticity of molecules leads to their aqueous solubility being higher than expected (*3*), which may be in part due to association effects as a result of ring stacking. However, in a bilayer environment such effects are absent, and in this context the CG representation of Phe is expected to be reasonable (see next section). Moreover, as has been noted (*10*), excess hydrophobicity may not unduly affect the location of membrane proteins in predictions.

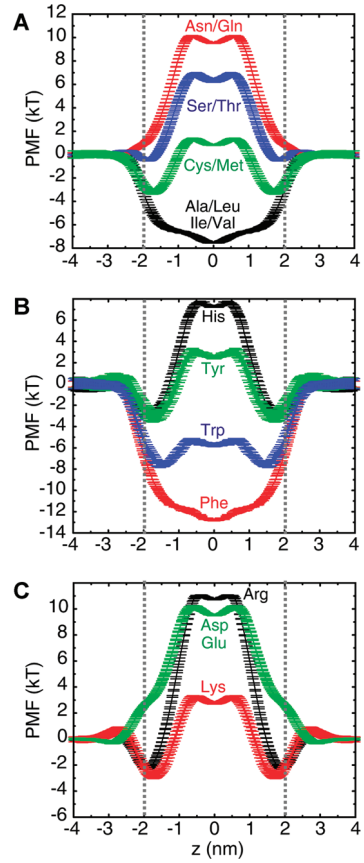

FIGURE 2: Potentials of mean force (i.e., ΔG profiles) for transfer of amino acid side chain analogues across the normal (z) of a DPPC bilayer and into bulk solvent. PMFs are shown for small aliphatic and polar amino acids (A), aromatic amino acids (B), and ionizable amino acids (C). The curve for each amino acid is indicated by inset labels. In each PMF, error bars represent the difference between PMFs calculated for the first and second halves of each simulation. Gray dotted lines indicate centers of phosphate particle peaks (Figure S1 in the Supporting Information). In calculating these PMFs, windows were separated by $\Delta z = 0.1$ nm, with a single PMF comprising 1.7 μs of simulation time.

Potential of Mean Force for Transferring Side Chain Analogues through the Membrane. Each amino acid analogue was transferred along the normal (z) of a 128-lipid DPPC bilayer, to yield a PMF via a series of 20 ns umbrella sampling simulations (Figure 2). Encouragingly, a correlation coefficient of 0.99 was observed between transfer energies from bulk water to the bilayer center and the water−hydrophobic solvent partition energies calculated above, confirming the agreement of the two methods of approaching free energies. As well as showing good agreement with experimental water/cyclohexane transfer free energies, there was also good correlation (~0.83) with the recently derived "biological" hydrophobicity scale (*6*).

Beyond correlation with experimental hydrophobicity scales, the PMF curves broadly agree with data obtained from statistical analyses of the distribution of residues in membrane protein structures (*6, 31*) and with positional variations in "biological" hydrophobicity (*6*), as well as all-atom PMFs (*10*). For example, among the smaller amino acids, those with aliphatic side chains (Figure 2A) partition favorably

within the entire bilayer, with a deep minimum at the center of the lipid tails. Cys/Met partition equally between solvent and the bilayer center but favor the interface slightly, reflecting the small hydrogen-bonding capacity of their sulfhydryl groups. This can be contrasted with the polar side chains of Ser/Thr and Asn/Gln for which transfer to the center of the bilayer is highly unfavorable. Ser/Thr also has a small minimum at the bilayer interface, favoring the glycerol/phosphate region over the choline groups. Comparable agreement was observed in recent PMFs for amino acids transferred through a dioleoylphosphatidylcholine (DOPC) bilayer using the MARTINI forcefield (*33*), although in that case, a small barrier to insertion was also observed for the hydrophobic aliphatic residues in the proximity of the lipid headgroup region.

The balance of transfer free energies for the ring-containing side chains (Figure 2B) is particularly important given the well-known interfacial roles of Trp/Tyr in "anchoring" membrane proteins. Phe is purely hydrophobic and partitions extremely favorably throughout the membrane, with a similar PMF to the small aliphatic analogues, but with a deeper minimum by ~5 kT. A large free energy minimum was also observed for Phe in MARTINI (*33*), although the well was less deep by ~4 kT. His, Trp and Tyr show more complex partitioning behavior. Although His partitions at the bilayer interface, it is rather polar and a large barrier is observed at the center of the membrane. Both Tyr and Trp have pronounced local energy wells at the membrane interface, as expected for these "anchoring" amino acids. This is reflected in the angle adopted by the aromatic plane with respect to the bilayer normal, the amphipathic distribution of CG particles for their side chains orienting with respect to the hydrophobic-polar boundary (Figure S2 in the Supporting Information). Nevertheless, there is a barrier to Tyr entering the middle of the bilayer, reflecting its hydroxyl group, whereas the more hydrophobic Trp still favors partitioning into the bilayer center compared to bulk solvent. The pattern of partitioning for Trp vs Phe is supported by all-atom equilibrium simulations (*50*). Again, comparable PMF curves were observed for MARTINI (*33*), with local and global energy minima at the membrane interface and membrane center respectively, for Tyr and Trp, but only a single free energy minimum at the interface and a large barrier of ~7 kT in the core of the membrane for His.

The charged residues experience large barriers to insertion into the hydrophobic core of the bilayer (Figure 2C), as expected from analysis of membrane protein structures (*6*) as well as the "biological" hydrophobicity scale, particularly with respect to their positional dependency (*6*). Arg has the largest barrier to membrane insertion, followed by Asp/Glu, and Lys, as observed in all-atom simulations (*10*). Also in agreement, large free energy minima are observed for Arg and Lys in the vicinity of the interfacial glycerol/phosphate groups, reflecting their charge and "snorkeling" ability (*9, 34*). In contrast, no local free energy minima are observed for the small and negatively charged Asp/Glu, also consistent with atomistic (AT-MD) simulations (*11*). The "snorkeling" of the basic residues is reflected in the angle adopted by the side chains with respect to the bilayer normal (Figure S2 in the Supporting Information). They evidently reorient around the interfacial region, with their charged terminal particles pointing toward the polar glycerol backbone and negatively charged phosphate groups. In absolute terms, while the transfer free energies from bulk solvent to the bilayer center for the neutral forms of Arg Lys, Glu, and Asp are comparable to all-atom estimates, the transfer energies for the charged forms are ~50% of their atomistic counterparts (*10*). Similarly, in MARTINI the shapes of the atomistic PMF curves were reproduced for the charged residues, but the absolute barrier to membrane insertion was significantly underestimated (*33*). This discrepancy is addressed in more detail below.

Translating a LeuN.Arg Helix through a Bilayer. To explore the issue of possible insertion of TM helices containing a basic residue into a lipid bilayer (discussed above) we have explored the PMF for translation of LeuN.Arg along the bilayer normal. This was designed to reproduce as closely as possible the AT-MD simulations of Allen and colleagues (*34*), permitting comparison between the CG and AT PMFs. As a control simulation, we also calculated the PMF for translation of a purely hydrophobic (LeuN) α-helix along the bilayer normal. The simulation system employed is shown in Figure 3A. For the LeuN.Arg simulations, this consists of a 101-residue α-helix, with an Arg residue located at the center of the α-helix flanked by 50 leucines on either side, and a bilayer of ~410 DPPC lipid molecules. Thus the helix is 15.4 nm long. Thus the range of translation (from $z = -4$ to $+4$ nm) is such that the Arg is moved right across the bilayer. (Note that at the limits, the Arg residue is beyond the non-bonded interaction cutoffs with respect to the bilayer.) Furthermore, at all stages the helix termini sit well clear of the membrane in the aqueous compartment on either side.

The resulting PMFs for LeuN vs LeuN.Arg are shown in Figure 3B. It can be seen that, as anticipated, the PMF for LeuN is essentially flat with an average (over z, ±standard deviation) of -0.2 ± 0.5 kT. In contrast PMF for LeuN.Arg shows a pronounced (14 to 15 kT) barrier for $z = -0.7$ to $+0.7$ nm, corresponding to the Arg side chain in the center of the bilayer. The shape of the profile is comparable to that observed in the AT simulations, although somewhat broader at the top. The barrier height in the CG-MD simulations (~15 kT) is about half of the estimated from AT-MD (~29 kT) (*34*). This correlates with the single amino acid side chain PMF results (see above). Interestingly, the Arg side chain was seen to snorkel toward the water as it was translated toward the bilayer core (Figure S3 in the Supporting Information), which was also observed in previous AT-MD simulations (*34*). This allowed the charged Arg side chain to interact favorably with the lipid phosphates and interfacial waters. The approximate rotational symmetry of Figure S3 in the Supporting Information about $z = 0$ Å provides further confidence that our umbrella sampling simulations were of sufficient duration.

That the barrier height experienced in the center of the lipid bilayer for Arg is significantly underestimated in comparison with atomistic simulations is perhaps not unexpected. Atomistic simulations of a basic side chain located in a membrane, either as an isolated side chain (*10*) or as part of a hydrophobic α-helix (*9, 34*), reveal large water defects which accompany the charged side chain within the low dielectric milieu. While bilayer deformation and some degree of water penetration are observed in our CG model, they are not as pronounced as those observed in correspond-

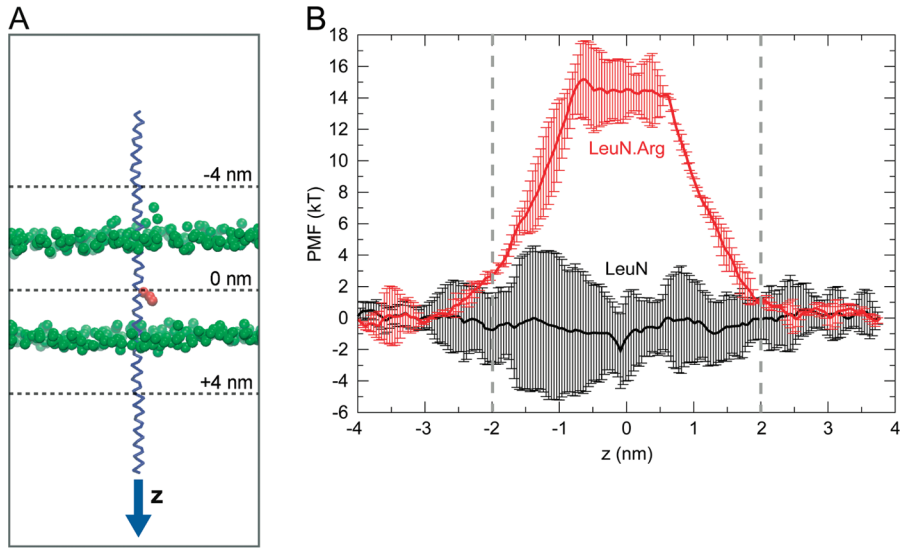

FIGURE 3: A: System setup for the LeuN.Arg simulation. The LeuN.Arg helix backbone is shown in blue, with the central Arg residue shown in red space-filling format. The phosphate particles of the DPPC bilayer are shown in green space-filling format. All other particles are omitted for clarity. B: PMF translation of LeuN and of LeuN.Arg helices along the bilayer normal (z). Gray dotted lines indicate centers of phosphate particle peaks (Figure S1 in the Supporting Information). The error bars indicate differences between PMFs calculated for first and second halves of simulation.)

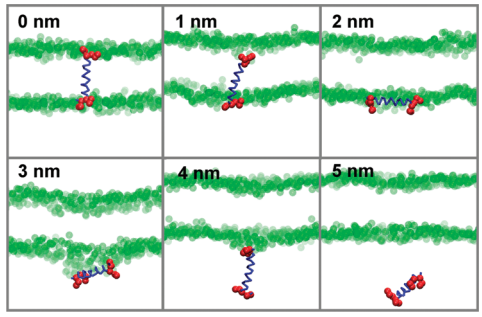

FIGURE 4: Representative simulation snapshots of the WALP23 peptide restrained at different points along z during umbrella sampling simulations. The peptide backbone is shown in blue, with flanking Trp residues of the WALP helix shown in red space-filling format, and the glycerol backbone particles of the bilayer shown in green space-filling format.

ing atomistic simulations (*44, 51*). The lack of a dipole on the simplified CG water particle leads to a reduction in desolvation energy, and also prevents formation of extended "water wires" within the membrane.

Potential of Mean Force for Transferring WALP23 through the Membrane. We calculated the partitioning free energy profile of the α-helical WALP23 peptide within a 252-DPPC bilayer as a model of a canonical TM α-helix. The WALP family of engineered peptides consist of Trp-flanked oligo-(Leu-Ala) sequences, simulations of which suggest that an interfacially bound intermediate may be important during folding/insertion (*52*). The PMF (derived from a total of 3.6 μs of simulation) was calculated via a series of 50 ns umbrella sampling simulations, during which the peptide center of mass (*com*) was restrained at positions separated by $\Delta z = 0.1$ nm along the membrane normal, z (Figure 4). A number of trial simulations, on this and more complex systems (*53*), suggested that 50 ns was long enough for other degrees of freedom to equilibrate.

As expected for a TM helix, a deep energy minimum is observed in the center of the membrane, at $z = 0$ nm (Figure 5A). The peptide remains in a TM orientation within $z \approx \pm 1$ nm on either side of the bilayer center (Figure 5B), with the terminal Trp residues anchoring the peptide to either leaflet (Figure 4). A small degree of positive bilayer mismatch (~3.9 nm, compared to ~4.1 nm in a pure bilayer) is observed in this region, as estimated by the mean distance between phosphate particles of opposing membrane leaflets within 0.5 nm of the peptide (Figure 5C). However, consistent with experimental data (*48*), only limited helix tilting occurs in response to this, with an optimal tilt angle of ~10−15° (Figure 5B), which is comparable to previous CG simulations using our model (*25*) and MARTINI (*33*).

After $z \approx \pm 1$ nm, the boundary of the TM energy well is reached, with a total free energy change of ~10 kT (Figure 5A), as a pair of peptide-anchoring Trp residues are fully removed from one leaflet (Figure 4). The helix adopts a locally favorable interfacial in-plane orientation (Figure 5B). Interestingly, atomic force microscopy (AFM) pulling studies of WALP23 peptides also revealed single barriers to removal, at ~0.75 nm, corresponding to mechanical removal of the helix out of the hydrophobic membrane core and into the interface (*54*).

Following the observed switch to an interfacial location, changes in free energy (of ~40 kT) up to $z \approx \pm 3$ nm (Figure 5A) are largely manifested by an increasingly negative bilayer mismatch (Figure 5C) as the peptide locally drags lipid headgroups from the membrane surface (Figure 4). After this z range, the bilayer distortion is rapidly abolished as the peptide reorients (Figure 5B) so that only the terminal region of the helix, with its flanking Trp residues, makes contact with lipid (Figure 4). It should be noted that the peptide might subsequently be expected to unfold, exposing backbone hydrogen bonds to solvent, as suggested by combined AFM/MD studies (*54*), but during CG-MD, helical restraints are applied. After $z \approx \pm 4$ nm, the PMF plateaus as the helix is completely extracted (Figure 5A).

DISCUSSION

The studies reported in this paper suggest that a relatively simple CG model is capable of capturing the energetics of

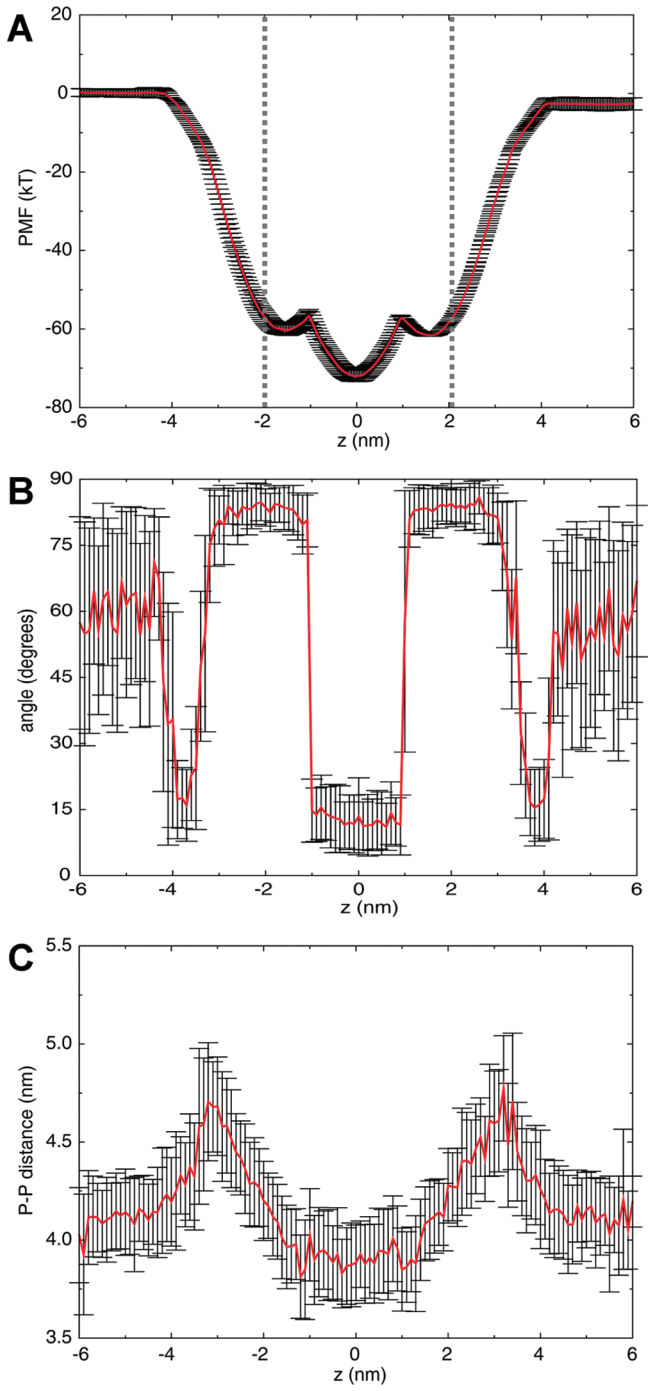

FIGURE 5: A: Potential of mean force (i.e., ΔG profile) for transfer of a WALP23 α-helix across the normal (z) of a DPPC bilayer and into bulk solvent. (Error bars: difference between PMFs calculated for first and second halves of simulation.) B: Average angle (±standard deviation) between bilayer normal and vector formed by a line of best-fit for the backbone particles of the α-helical region of WALP23, calculated as a function of z. C: Local deformation of the bilayer indicated by the average distance (±standard deviation) between upper and lower phosphate (**P**) particles (d_{PP}) within 0.5 nm of the WALP23 peptide center of mass, calculated as a function of z.

interaction of amino acid side chains, and hence of TM helices, with a lipid bilayer. Examination of the energetics of transfer from an aqueous solvent to a hydrophobic solvent or the hydrophobic bilayer core suggests that the CG model stands up to quantitative scrutiny, and that the overall accuracy of this approach (i.e., averaging over all amino acid side chain types) is comparable with that of widely used atomistic forcefields such as OPLS and GROMOS. Along with this study, Monticelli et al. (*33*) have recently used a similar approach to the calibration of CG amino acid parameters. Thus, it would seem that the CG methodology is of sufficient accuracy to be employed in e.g. studies aimed at prediction of TM protein structure. The properties of various model peptides such as WALP within a bilayer environment using the CG approach show good qualitative agreement with experimental measurements and atomistic simulations, for both our model (*25*) and MARTINI (*33*). Here, we have provided further support for the CG approach by successfully calculating the complete membrane PMF for WALP as a canonical test system.

Nevertheless, further development of the CG-MD approach necessitates maintaining a critical approach to its evaluation. In this context, it is clear that there are some clear physical deficiencies in the current approach. This is perhaps most evident in the case of the Arg side chain of the LeuN.Arg test helix, the PMF for which reveals that the barrier height experienced in the center of the lipid bilayer is ∼50% of that estimated from atomistic simulations (*34*), and that only limited water defects accompanied entry of charged side chains within the membrane. In this context, a recent implementation of an earlier version of the CG model (in which a less polar combination of particle types (**C** + **Qd**) was used for the Arg side chain (*25*)) resulted in an estimate of only ∼8 kT for the barrier height, compared to a value of ∼27 kT from atomistic simulations (*51*). This estimate of the CG barrier is lower than our CG current barrier estimate of ∼15 kT, which is based on (**Nd** + **Qd**) for Arg. Thus, as Vorobyov et al. (*51*) suggest, careful parametrization of CG models is indeed required if meaningful quantitative comparisons with AT models are to be made.

Similarly, the free energy barrier was also found to be too low for CG charged amino acids by Monticelli et al. (*33*) within the core of a DOPC bilayer. This result is unsurprising, given the neglect of long-range electrostatics, and the use of a simplified water model containing no dipole. Use of the original Marrink CG forcefield has been shown to successfully model dendrimer pore formation in lipid bilayers (*55*) with the inclusion of long-range electrostatics. On the other hand, our own measurements suggest that use of a longer electrostatics cutoff range, or of Ewald summation, leads to only limited improvements in the bilayer PMF for charged particles (data not shown). Some improvement in modeling the interaction strength between polar/charged residues within the low dielectric membrane core has been demonstrated via a combination of decreased screening and increased Lennard-Jones repulsion with nonpolar particles (*18, 33*). The fundamental problem of implicit screening remains, however. Without substantially altering the treatment of the CG water model, significant underestimation of desolvation energy will likely remain. Use of a more sophisticated CG water model (e.g., the soft sticky dipole model used in some recent CG studies (*56*)) may help in better approximating these phenomena.

Another limitation of the approach is the lack of resolution in the amino acid particles. For example, while the treatment of ring-based side chains here enables a reasonable reproduction of their partitioning behavior, and for Trp and Tyr, membrane position-specific reorientation predicted for their "anchoring" role in membrane proteins, significant deficiencies remain. Use

of a larger number of special, small particles to represent aromatic rings in MARTINI (*18, 33*) allows an improved description of the geometric detail of such ring structures. However, a more complicated description of the bonded interactions to preserve rigidity and prevent out-of-plane distortions results in a smaller integration time step. Our own preliminary testing of such treatment of rings (as in e.g. ref *17*) suggests that while some improvement in side chain detail is achieved by using three side chain particles, only limited improvement is observed in the structural detail of protein−protein interactions, or in the expected effects of e.g. low densities of model CG sterol molecules on lipid tail ordering. It will also prove difficult to incorporate such complex phenomena as ring stacking or cation−π interactions within a CG model. More generally, the lack of atomistic detail leads to limitations in the treatment of protein−protein and protein−lipid interactions. Nevertheless, a surprising sensitivity is evident within the CG model. Preliminary studies suggest that the approximation of side chains by just one or two particles can still provide considerable subtlety in spontaneous side chain dependent conformational changes at protein−protein interfaces, with the successful prediction of the structures of oligomeric α-helical bundles (*25*−*27*), and of the effects of point mutations upon them (*26*). Similarly, small changes in the sequence of TM peptides (*29, 33*) have been shown to exhibit consistent, reproducible membrane partitioning behavior with the CG representation, in agreement with biophysical measurements demonstrating the "biological hydrophobicity scale" (*47*).

Finally, in our CG protein model, secondary structure was maintained by restraining the distance between backbone particles to mimic backbone hydrogen bonds in the equivalent atomistic structure. Thus, the secondary structure is effectively fixed. In contrast to distance restraints, MARTINI imposes proper dihedral potentials to quartets of CG backbone particles with the same secondary structure, with defined equilibrium dihedral angles for α- or β-secondary structure. Although technically different, this results in the same effect; that is, a fixed protein structure. Indeed, as pointed out by Monticelli et al. (*33*), conformational changes of protein secondary structure are not adequately modeled with either approach, and transitions between different kinds of secondary structure are impossible. On the other hand, realistic movement of secondary structure elements with respect to each other is possible, as revealed by recent application of our CG model to the effects of mutations on formation of simple model TM dimers (*26*), and the tetramerization of viral channel structures (*27*), and of MARTINI to e.g. gating motions of voltage-gated channels (*28*).

In conclusion, a CG model which treats the peptide backbone as a single particle is unlikely to achieve sufficient resolution to enable an "unrestrained" (be it by harmonic restraints or dihedral potentials) approach to protein structural changes. Similarly, the conformational protein landscape results from a balance of fine-tuned interactions between protein, lipid, and solvent.; parameter tuning of CG particles will only allow limited improvements in the description of conformational change or (un)folding. However, it is clear that the CG approach will prove of suitable accuracy to generate e.g. initial models of insertion of single α-helices, α-helix hairpins (*17*), or helix dimers (*26*) which may be refined by subsequent extended atomistic simulation (*57*). Looking beyond such predictions, the CG approach is able to capture larger-scale processes, such as protein/lipid self-assembly, protein oligomerization/aggregation, and even substantial relative motions of secondary structural regions in multidomain membrane proteins. The likely accuracy of the model explains the excellent agreement between CG and AT simulations of the KvAP voltage sensor domain (*29, 58*), and the agreement of CG simulation with recent experimental data on S4/lipid interactions (*59*) despite some reservations expressed on the basis of a preliminary comparison of atomistic and CG simulations (*34*). Elastic network models have already been combined with our CG parameters to simulate limited conformational changes (*17, 29*); a refined multibasin network approach might yield further insight. More detailed studies will likely benefit from a multigrained approach, for example, by simulating the protein in atomistic detail while treating the solvent and/or lipid at a coarse-grained level (*15*). Equally, a promising direction involves alternating between atomistic and CG representations of an entire simulation system (*27*), either sequentially, or in an analogous way to the replica exchange method, enabling the simultaneous study of related phenomena spanning different time- and length-scales.

ACKNOWLEDGMENT

We thank our colleagues, especially Philip Biggin and Oliver Beckstein, for valuable discussions concerning this work.

SUPPORTING INFORMATION AVAILABLE

A comparison of the free energies of transfer for amino acid side chain analogues from water to a nonpolar solvent for the current study and for the MARTINI forcefield with experimental values is provided, in addition to supplementary analysis results mentioned in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., ersson, H., Nilsson, I., White, S. H., and von Heijne, G. (2005) Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature 433*, 377–381.
2. Popot, J. L., and Engelman, D. M. (2000) Helical membrane protein folding, stability, and evolution. *Annu. Rev. Biochem. 69*, 881–922.
3. Radzicka, A., and Wolfenden, R. (1988) Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. *Biochemistry 27*, 1664–1670.
4. Wimley, W. C., Creamer, T. P., and White, S. H. (1996) Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. *Biochemistry 35*, 5109–5124.
5. Wimley, C. W., and White, S. H. (1996) Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Biol. 3*, 842–848.
6. 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: Struct., Funct., Bioinf. 59*, 252–265.
7. Andersen, O. S., and Koeppe, R. E. (2007) Bilayer thickness and membrane protein function: an energetic perspective. *Annu. Rev. Biophys. Biomol. Struct. 36*, 107–130.
8. Lindahl, E., and Sansom, M. S. P. (2008) Membrane proteins: molecular dynamics simulations. *Curr. Opin. Struct. Biol. 18*, 425–431.
9. Johansson, A. C. V., and Lindahl, E. (2006) Amino-acid solvation structure in transmembrane helices from molecular dynamics simulations. *Biophys. J. 91*, 4450–4463.

10. MacCallum, J. L., Bennett, W. F. D., and Tieleman, D. P. (2007) Partitioning of amino acid side chains into lipid bilayers: results from computer simulations and comparison to experiment. *J. Gen. Physiol. 129*, 371–377.
11. MacCallum, J. L., Bennett, W. F. D., and Tieleman, D. P. (2008) Distribution of amino acids in a lipid bilayer from computer simulations. *Biophys. J. 94*, 3393–3404.
12. 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.
13. 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.
14. Izvekov, S., and Voth, G. A. (2005) A multiscale coarse-graining method for biomolecular systems. *J. Phys. Chem. B 109*, 2469–2473.
15. 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.
16. Shih, A. Y., Arkhipov, A., Freddolino, P. L., and Schulten, K. (2006) Coarse grained protein-lipid model with application to lipoprotein particles. *J. Phys. Chem. B 110*, 3674–3684.
17. 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.
18. Marrink, S. J., Risselada, J., Yefimov, S., Tieleman, D. P., and de Vries, A. H. (2007) The MARTINI forcefield: coarse grained model for biomolecular simulations. *J. Phys. Chem. B. 111*, 7812–7824.
19. Tozzini, V. (2005) Coarse-grained models for proteins. *Curr. Opin. Struct. Biol. 15*, 144–150.
20. Ayton, G. A., Noid, W. G., and Voth, G. A. (2007) Multiscale modeling of biomolecular systems: in serial and in parallel. *Curr. Opin. Struct. Biol. 17*, 192–198.
21. Go, N., and Abe, H. (1981) Noninteracting local-structure model of folding and unfolding transition in globular proteins I. Formulation. *Biopolymers 20*, 991–1011.
22. Reith, D., Putz, M., and Muller-Plathe, F. (2003) Deriving effective mesoscale potentials from atomistic simulations. *J. Comput. Chem. 24*, 1624–1636.
23. Zhou, J., Thorpe, I., Izvekov, S., and Voth, G. A. (2007) Coarse-grained peptide modeling using a systematic force-matching approach. *Biophys. J. 92*, 4289–4303.
24. Izvekov, S., and Voth, G. A. (2006) Multiscale coarse-graining of mixed phospholipid/cholesterol bilayers. *J. Chem. Theory Comput. 2*, 637–648.
25. Bond, P. J., and Sansom, M. S. P. (2006) Insertion and assembly of membrane proteins via simulation. *J. Am. Chem. Soc. 128*, 2697–2704.
26. Psachoulia, E., Bond, P. J., Fowler, P. W., and Sansom, M. S. P. (2008) Helix-helix interactions in membrane proteins: coarse-grained simulations of glycophorin helix dimerization. *Biochemistry 47*, 10503–10512.
27. Carpenter, T., Bond, P. J., Khalid, S., and Sansom, M. S. P. (2008) Self-assembly of a simple membrane protein: coarse-grained molecular dynamics simulations of the influenza M2 channel. *Biophys. J.* (in press), doi: 10.1529/biophysj.108.131078.
28. Treptow, W., Marrink, S.-J., and Tarek, M. (2008) Gating motions in voltage-gated potassium channels revealed by coarse-grained molecular dynamics simulations. *J. Phys. Chem. B 112*, 3277–3282.
29. 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. U.S.A. 104*, 2631–2636.
30. Wee, C. L., Bemporad, D., Sands, Z. A., Gavaghan, D., and Sansom, M. S. P. (2007) SGTx1, a Kv channel gating-modifier toxin, binds to the interfacial region of lipid bilayers. *Biophys. J. 92*, L07−L09.
31. Scott, K. A., Bond, P. J., Ivetac, A., Chetwynd, A. P., Khalid, S., and Sansom, M. S. P. (2008) Coarse-grained MD simulations of membrane protein-bilayer self-assembly. *Structure 16*, 621–630.
32. Allen, T. W. (2007) Modeling charged protein side chains in lipid membranes. *J. Gen. Physiol. 130*, 237–240.
33. Monticelli, L., Kandasamy, S. K., Periole, X., Larson, R. G., Tieleman, D. P., and Marrink, S. J. (2008) The MARTINI coarse grained force field: extension to proteins. *J. Chem. Theory Comput. 4*, 819–834.
34. Dorairaj, S., and Allen, T. W. (2007) On the thermodynamic stability of a charged arginine side chain in a transmembrane helix. *Proc. Natl. Acad. Sci. U.S.A. 104*, 4943–4948.
35. Freites, J. A., Tobias, D. J., von Heijne, G., and White, S. H. (2005) Interface connections of a transmembrane voltage sensor. *Proc. Natl. Acad. Sci. U.S.A. 102*, 15059–15064.
36. Killian, J. A. (2003) Synthetic peptides as models for intrinsic membrane proteins. *FEBS Lett. 555*, 134–138.
37. 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.
38. 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.
39. Nose, S. (1984) A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys. 52*, 255–268.
40. Hoover, W. G. (1985) Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A31*, 1695–1697.
41. Parrinello, M., and Rahman, A. (1981) Polymorphic transitions in single-crystals - a new molecular-dynamics method. *J. Appl. Phys. 52*, 7182–7190.
42. Nose, S., and Klein, M. L. (1983) Constant pressure molecular-dynamics for molecular-systems. *Mol. Phys. 50*, 1055–1076.
43. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD - Visual Molecular Dynamics. *J. Mol. Graphics 14*, 33–38.
44. Dorairaj, S., and Allen, T. W. (2006) Free energies of arginine-lipid interactions: potential of mean force of a transmembrane helix through a membrane. *Biophys. J. 90*, 1022-Pos.
45. Villa, A., and Mark, A. E. (2002) Calculation of the free energy of solvation for neutral analogs of amino acid side chains. *J. Comput. Chem. 23*, 548–553.
46. Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A., and Rosenberg, J. M. (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. 1. The method. *J. Comput. Chem. 13*, 1011–1021.
47. Hessa, T., White, S. H., and von Heijne, G. (2005) Membrane insertion of a potassium-channel voltage sensor. *Science 307*, 1427.
48. van der Wel, P. C. A., Strandberg, E., Killian, J. A., and Koeppe, R. E. (2002) Geometry and intrinsic tilt of a tryptophan-anchored transmembrane a-helix determined by ^{2}H NMR. *Biophys. J. 83*, 1479–1488.
49. Petrache, H. I., Zuckerman, D. M., Sachs, J. N., Killian, J. A., Koeppe, R. E., and Woolf, T. B. (2002) Hydrophobic matching mechanism investigated by molecular dynamics simulations. *Langmuir 18*, 1340–1351.
50. Norman, K. E., and Nymeyer, H. (2006) Indole localization in lipid membranes revealed by molecular simulation. *Biophys. J. 91*, 2046–2054.
51. Vorobyov, I., Li, L., and Allen, T. W. (2008) Assessing atomistic and coarse-grained force fields for protein-lipid interactions: the formidable challenge of an ionizable side chain in a membrane. *J. Phys. Chem. B 112*, 9588–9602.
52. Nymeyer, H., Woolf, T. B., and Garcia, A. E. (2005) Folding is not required for bilayer insertion: replica exchange simulations for an α-helical peptide with an explicit lipid bilayer. *Proteins: Struct., Funct., Bioinf. 59*, 783–790.
53. Wee, C. L., Gavaghan, D., and Sansom, M. S. P. (2008) Lipid bilayer deformation and the free energy of interaction of a K_v channel gating-modifier toxin. Biophys. J. (in press), doi: 10.1529/biophysj.108.130971.
54. Ganchev, D. N., Rijkers, D. T. S., Snel, M. M. E., Killian, J. A., and de Kruijff, B. (2004) Strength of integration of transmembrane α-helical peptides in lipid bilayers as determined by atomic force spectroscopy. *Biochemistry 43*, 14987–14993.
55. Lee, H., and Larson, R. G. (2008) Coarse-grained molecular dynamics studies of the concentration and size dependence of fifth- and seventh-generation PAMAM dendrimers on pore formation in DMPC bilayer. *J. Phys. Chem. B* (in press).
56. Orsi, M., Haubertin, D. Y., Sanderson, W. E., and Essex, J. W. (2008) A quantitative coarse-grain model for lipid bilayers. *J. Phys. Chem. B 112*, 802–815.
57. Cuthbertson, J. M., Bond, P. J., and Sansom, M. S. P. (2006) Transmembrane helix-helix interactions: comparative simulations of the glycophorin A dimer. *Biochemistry 45*, 14298–14310.
58. 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.
59. Long, S. B., Tao, X., Campbell, E. B., and MacKinnon, R. (2007) Atomic structure of a voltage-dependent K^+ channel in a lipid membrane-like environment. *Nature 450*, 376–382.

BI800642M