Membrane protein simulations with a united-atom lipid and all-atom protein model: lipid–protein interactions, side chain transfer free energies and model proteins

2006 J. Phys.: Condens. Matter 18 S1221

(http://iopscience.iop.org/0953-8984/18/28/S07)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **18** (2006) S1221–S1234 doi:10.1088/0953-8984/18/28/S07

Membrane protein simulations with a united-atom lipid and all-atom protein model: lipid–protein interactions, side chain transfer free energies and model proteins

D Peter Tieleman[1], **Justin L MacCallum, Walter L Ash, Christian Kandt, Zhitao Xu and Luca Monticelli**

Department of Biological Sciences, University of Calgary, 2500 University Drive NW, Calgary AB T2N 1N4, Canada

E-mail: tieleman@ucalgary.ca

Received 23 December 2005
Published 28 June 2006
Online at stacks.iop.org/JPhysCM/18/S1221

Abstract

We have reparameterized the dihedral parameters in a commonly used united-atom lipid force field so that they can be used with the all-atom OPLS force field for proteins implemented in the molecular dynamics simulation software GROMACS. Simulations with this new combination give stable trajectories and sensible behaviour of both lipids and protein. We have calculated the free energy of transfer of amino acid side chains between water and 'lipid-cyclohexane', made of lipid force field methylene groups, as a hydrophobic mimic of the membrane interior, for both the OPLS-AA and a modified OPLS-AA force field which gives better hydration free energies under simulation conditions close to those preferred for the lipid force field. The average error is 4.3 kJ mol^{-1} for water–'lipid-cyclohexane' compared to 3.2 kJ mol^{-1} for OPLS-AA cyclohexane and 2.4 kJ mol^{-1} for the modified OPLS-AA water–'lipid-cyclohexane'. We have also investigated the effect of different methods to combine parameters between the united-atom lipid force field and the united-atom protein force field ffgmx. In a widely used combination, the strength of interactions between hydrocarbon lipid tails and proteins is significantly overestimated, causing a decrease in the area per lipid and an increase in lipid ordering. Using straight combination rules improves the results. Combined, we suggest that using OPLS-AA together with the united-atom lipid force field implemented in GROMACS is a reasonable approach to membrane protein simulations. We also suggest that using partial volume information and free

[1] Author to whom any correspondence should be addressed. http://moose.bio.ucalgary.ca.

energies of transfer may help to improve the parameterization of lipid–protein interactions and point out the need for accurate experimental data to validate and improve force field descriptions of such interactions.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

1.1. MD review

Molecular dynamics (MD) simulations have seen tremendous advances since they were first introduced in the late 1950s (Alder and Wainwright 1957). The first simulation of liquid water was published in 1974 (Stillinger and Rahman 1974) and the first simulation of a small protein was presented in 1977 (McCammon *et al* 1977). Today, MD simulations have been applied to systems as large as a billion atoms (Abraham *et al* 2002), or for periods of time as long as milliseconds (Snow *et al* 2004), and are regularly used to investigate structure–activity relationships in biological macromolecules. Simulations facilitate the interpretation of experimental data and give access to information not directly accessible by experiments. They also form a critical component of atomic-resolution structure determination methods such as x-ray crystallography and NMR.

Currently, simulations are routinely performed on systems of up to 100 000 atoms on a time scale of tens to hundreds of nanoseconds. MD simulations of membrane proteins are of special interest, due to the difficulties particularly in obtaining high-resolution structural experimental information about these proteins (White 2004). Membrane proteins are of high biological and medical relevance as they are key players in crucial processes such as energy conversion, transport, antibiotic resistance, and signal transduction.

At the core of molecular dynamics lie classical equations of motion and statistical mechanics. With knowledge of the forces acting on each particle it is possible to calculate the dynamic behaviour of a system. The palette of possible applications based on this approach is broad, from atoms in a molecule to stars in a galaxy: the principles are the same, only the energy functions used to describe the system are different. For an atomic system, the potential energy component of the Hamiltonian comes from a set of equations that empirically describe bonded and non-bonded interactions between atoms. These energy functions together with their parameters are collectively referred to as the 'force field'. Molecular dynamics force fields usually consist of two major components: one part describing interactions between atoms connected via covalent bonds (such as bond lengths, bond angles, dihedrals) and another treating non-bonded interactions (most often electrostatic interactions between point charges, and a Lennard-Jones function to model van der Waals interactions). Different force fields also use different levels of detail, and can be categorized as 'all-atom', 'united-atom', or 'coarse-grained'. All-atom force fields treat every atom (including hydrogen) explicitly, united-atom force fields combine each aliphatic carbon and associated hydrogens into a single particle, and coarse-grained force fields describe larger molecular units (such as amino acid side-chains and whole water molecules) as single particles. For a more detailed introduction to the method see, for example, (Karplus and Petsko 1990, van Gunsteren and Berendsen 1990, Karplus and McCammon 2002, Ponder and Case 2003).

1.2. Force fields

There are four main force fields in common use for simulating biological macromolecules: AMBER (Weiner *et al* 1984, Cornell *et al* 1995), CHARMM (Brooks *et al* 1983, MacKerell

et al 1998), GROMOS (van Gunsteren 1987) and OPLS (Jorgensen and Tirado-Rives 1988). Each of these has undergone continuous development as parameterization methodology and experimental techniques advance. All four force fields reproduce many protein characteristics satisfyingly well (Ponder and Case 2003, MacKerell 2004), although there remains room for improvement. They share common assumptions, the most serious of which is probably the simplified treatment of electrostatic interactions as point charges centred on atoms. This ignores details of electronic polarizability, limiting the accuracy of properties such as the relative orientation of aromatic residues, interactions with ions, and likely the solvation free energy of amino acid side chains in solvents of different dielectric constants. All force fields started out as 'united-atom' force fields, in which aliphatic carbon atoms and their hydrogen atoms were treated as a single pseudo-atom, but recent versions of AMBER, CHARMM, and OPLS have moved to an all-atom description. Only new versions of GROMOS are united-atom. There are both philosophical and practical reasons behind these developments, which we return to below in the context of lipid force fields.

Only two phospholipid force fields are in common use today: one is part of the official CHARMM distribution and one was introduced by Berger and co-workers (Berger *et al* 1997), developed with parameters taken from united-atom versions of OPLS and AMBER. The CHARMM force field is all-atom, whereas the Berger force field is united-atom. Both CHARMM based lipids and Berger lipids reproduce the available experimental information on the structure and dynamics of phospholipid bilayers reasonably well, particular for the experimentally well-studied phosphatidylcholine lipids, and there is no compelling experimental information that indicates either force field is substantially better.

In this paper, we investigate the consequences of combining the Berger force field for lipids with different protein force fields: a united-atom protein force field based on GROMOS87, implemented as ffgmx in GROMACS, the 43A2 and 45A3 versions of the united-atom protein force field GROMOS96 (Schuler *et al* 2001) and the all-atom OPLS-AA force field (Jorgensen *et al* 1996). It is not trivial to combine a protein force field with a lipid force field from a different family of force fields. Although the whole issue of combining parameters can be avoided by using a single unified force field, the only feasible choice at the moment for this is CHARMM. We will make the case that it is desirable to have a united-atom lipid force field, which CHARMM is not. We also identify some issues that may occur when parameters are combined from different force fields. This is relatively common in the literature, and usually not investigated in detail.

There are other issues in setting up membrane protein simulations besides the choice of force field. We have recently reviewed current best practice for such simulations (Kandt *et al* 2006). Other technical discussions and comparisons between different algorithms have also been described previously (Tieleman *et al* 2002, Anezo *et al* 2003).

2. Methods

2.1. Reparameterizing lipid dihedral parameters to deal with OPLS scaling of 1–4 interactions

We were interested in combining all-atom OPLS protein with united-atom Berger lipids, motivated by the belief that the united-atom approximation for the lipids would speed calculations without a significant effect on the accuracy of our calculations. The force fields are mathematically compatible (they use the same functional form for the potential function), but OPLS scales electrostatic and Lennard-Jones interactions between atoms connected by three bonds (1–4 interactions) by a factor of 0.5. With the implementation we use, GROMACS, a 1–4 interaction scaling factor must be universally applied to the force field, precluding a simple

addition of the Berger lipids to OPLS. While a separate set of Lennard-Jones parameters can be used in GROMACS for 1–4 interactions, it is not possible in the current implementation to use a separate set of charges. Thus we decided to replace the 1–4 interactions in the lipids (both the electrostatic and the Lennard-Jones components) with dihedral potentials. The resulting lipids differ only slightly from true Berger lipids, in that the replacement for the missing 1–4 interaction assumes ideal bond lengths and angles. On the other hand, because no 1–4 interactions are present in the modified lipids, the new parameters can be used in combination with any force field sharing the same functional form, independently of the scaling factor used for the 1–4 interaction, including the different versions of GROMOS, OPLS-AA and AMBER force fields for proteins.

Electrostatic and Lennard-Jones energy functions were calculated for each pair of atoms for which a 1–4 interaction is described in the Berger lipid force field, with bonds and angles fixed corresponding to ideal geometries. The interatomic distance (and therefore the Coulomb and Lennard-Jones potential energy) was calculated as a function of the dihedral angle, with constant parameters specifying the two bond angles and three bond lengths, the charges on the terminal atoms and the pair Lennard-Jones parameters, for each unique set of four bonded atoms. We fit to this calculated potential a Ryckaert–Bellemans dihedral function, a six-term series of cosines (Jorgensen and Tirado-Rives 1988), in XMGRACE using nonlinear curve fitting (tolerance: 0.01, iterations: 50). The fits between the Ryckaert–Bellemans dihedral functions and the calculated 1–4 interaction energies were virtually perfect in all cases, with a correlation coefficient higher than 0.9999.

To test that the two approaches give the same lipid properties, three simulations of DOPC bilayers were carried out; in the first case, the original Berger parameters were used, with non-scaled 1–4 interactions, according to the rules for all GROMOS force fields; in the second case, the new parameters were used, with 1–4 interactions replaced by Ryckaert–Bellemans dihedral potentials; the third case was identical to the second, except that C6/C12 Lennard-Jones parameters were converted into sigma end epsilon, for consistency with the OPLS-AA and AMBER force fields. Results for the second and third combination were virtually identical within statistical error margins, and therefore we only report the results for the latter simulation. We will refer to the Berger lipid parameters compatible with OPLS-AA as to Ash lipid parameters. Simulation details are given below.

2.2. Molecular dynamics simulations of WALP23, polyleucine, and OmpF

MD simulations of the WALP23 peptide, the Leu22 peptide and OmpF were performed using two different combinations of force fields: (1) ffgmx for the peptide and the Berger lipid parameters, (2) OPLS-AA with the Ash lipid parameters. The simple point charge (SPC) water model (Berendsen *et al* 1981) was used in the simulations with the ffgmx force field, while the TIP4P model (Jorgensen *et al* 1983) was used in combination with OPLS-AA.

The WALP23 and Leu22 peptides were placed in a DOPC bilayer using the procedure described by Kandt *et al*, starting from a stretched pre-equilibrated DOPC bilayer and translating the lipids back to their original positions (Kandt *et al* 2006). DOPC bilayers consisted of 72 lipids, and the total number of water molecules was 3698. The total simulation time was 60 ns for each force field combination.

For OmpF, we used a previously published structure of the protein inserted into a DMPC bilayer (Robertson and Tieleman 2002). Snapshots of some of the simulation systems are shown in figures 1 and 2. OmpF was simulated in a DMPC bilayer containing 397 lipids, with 14353 water molecules, 40 chloride, 40 potassium, and 21 sodium ions. Simulation parameters were as described below, except for the use of a real-space interaction cutoff of 1.0 nm and

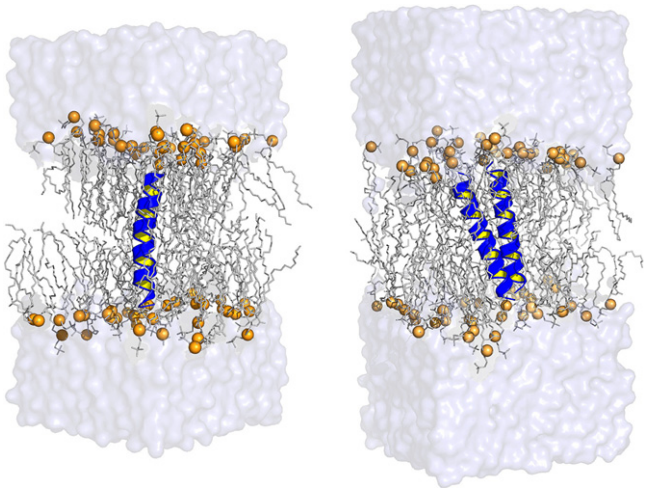

Figure 1. Simulation snapshots of two peptide simulation systems: WALP23 and the WALP23 dimer.

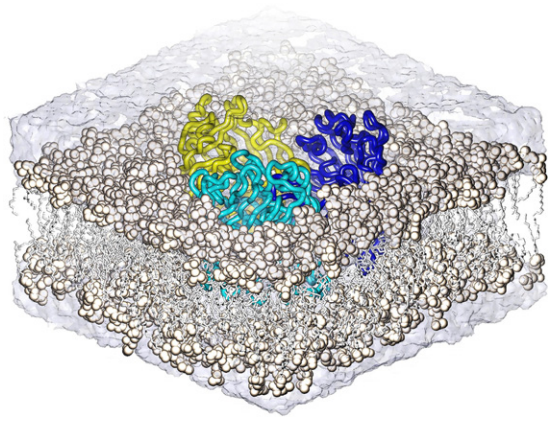

Figure 2. Simulation snapshot of the OmpF simulation.

truncated Lennard-Jones interactions at 1.0 nm. The total simulation time was 10 ns for each of the two force field combinations, with analysis performed on the last 5 ns.

2.3. Volume calculations of WALP23

Simulations were carried out to calculate the partial volume of the WALP23 peptide in water and in octane. Two sets of simulations were performed, in the presence and in the absence of the peptide. Simulations in the absence of the peptide were used to calculate the partial volume of the solvent (water or octane) with the different force fields; the partial volume of the peptide was calculated as the difference between the total volume of the simulation box and the partial volume of the solvent times the number of solvent molecules. Three water models were used: SPC (together with ffgmx for the peptide), TIP3P and TIP4P (together with OPLS-AA for the peptide); and five different combinations of force fields for peptide–octane simulations: (1) ffgmx CH2/CH3 atom types for the octane with ffgmx for the peptide, (2) Berger's LP2/LP3

atom types for the octane with ffgmx for the peptide, with special Lennard-Jones interactions distributed with ffgmx between octane and the protein, (3) Berger's LP2/LP3 atom types for the octane with ffgmx for the peptide, with combination rules for the lipid–protein interactions, (4) Berger's LP2/LP3 atom types for the octane with OPLS-AA for the peptide, (5) GROMOS 45a3 parameters for both the octane and the peptide. All the simulations including the WALP23 peptide were 40 ns long.

2.4. Transfer free energy calculations

We calculated the hydration and water–cyclohexane transfer free energies of analogues of 18 or 20 amino acid residues (excluding glycine and valine) using the same procedure as employed previously (MacCallum and Tieleman 2003). In the present work, free energies calculated using three force field combinations were compared with each other and with experimental values; the force field combinations included: (1) the OPLS-AA force field for the side chain analogues combined with the standard OPLS-AA parameters for aliphatic hydrocarbons; (2) the OPLS-AA force field for the side chain analogues combined with a model of cyclohexane based on the Berger lipid parameters; (3) a version of OPLS-AA with modified charges (Xu *et al* 2006) combined with a model of cyclohexane based on the Berger lipid parameters. Free energies of solvation were calculated using the thermodynamic integration method. The interactions between the side chain analogue and the solvent were turned off with the use of a coupling parameter λ. For each side chain analogue, a series of 21 simulations was performed with the λ coupling parameter fixed at values between 0.0 and 1.0, in increments of 0.05. Each simulation was 350 ps long, and the first 50 ps were discarded as equilibration.

2.5. Simulation details

All simulations were carried out using the GROMACS package version 3.2 (Lindahl *et al* 2001). All bond lengths were constrained to their equilibrium values using the SETTLE algorithm (Miyamoto and Kollman 1992) for water and the LINCS algorithm for the peptides, lipids, and protein (Hess *et al* 1997). The integration time step was 2 fs and the neighbour list for the calculation of non-bonded interactions was updated every 10 time steps. Periodic boundary conditions were used and the calculation of electrostatic forces utilized the particle mesh Ewald (PME) method (Darden *et al* 1993, Essmann *et al* 1995), with a cutoff of 0.9 nm for real-space interactions; the reciprocal-space interactions were evaluated on a 0.12 nm grid with fourth-order B-spline interpolation. A twin-range cutoff of 0.9–1.4 nm was used for the calculation of Lennard-Jones interactions. In each simulation the peptide and solvent were coupled separately to a temperature bath at 300 K, using the Berendsen algorithm with $\tau_T = 0.1$ ps, and the pressure was kept at 1 bar using weak pressure coupling with $\tau_P = 4.0$ ps (Berendsen *et al* 1984).

3. Results

3.1. Ash and Berger parameters

Since the Ash parameters with OPLS-AA rules and Berger parameters with ffgmx rules are nearly the same mathematically, they should give the same results for the properties of lipids within statistical accuracy, in the absence of any peptides. We looked both at conformational properties of the individual lipid molecules and at simple properties of the lipid bilayer.

Because 1–4 interactions depend on the distance between two atoms, while Ryckaert–Bellemans dihedrals depend on the dihedral angle, the assumption of ideal geometries (and in

Table 1. Area per lipid in simulations of pure DOPC and DOPC in the presence of hydrophobic transmembrane helical peptides, and the total area of the OmpF trimer in a DMPC bilayer. Error estimates are based on block averaging on the last 40 ns (DOPC, peptides) or 10 ns (OmpF).

System	ffgmx + Berger Area (nm^2), (std. error)	OPLS-AA + Ash Area (nm^2), (std. error)
DOPC	0.645 (0.005)	0.621 (0.009)
Leu22 (1)	0.578 (0.002)	0.631 (0.003)
Leu22 (2)	0.577 (0.008)	0.635 (0.011)
WALP23 monomer (1)	0.568 (0.003)	0.635 (0.007)
WALP23 monomer (2)	0.560 (0.008)	0.645 (0.004)
WALP23 dimer (1)	0.595 (0.003)	0.641 (0.004)
WALP23 dimer (2)	0.590 (0.004)	0.636 (0.008)
OmpF trimer (DMPC)	160.2 (0.5)	168 (2)

particular bond angles) implies that the two functions are not exactly equivalent. In particular, the local geometry is more likely to be different from the ideal one when particles separated by three bonds carry high partial charges, because the Ryckaert–Bellemans function does not take into account the distortion from the ideal angles due to the electrostatic interaction. The differences in the local conformation between the Berger lipids (with 1–4 interactions) and the Ash lipids (with no 1–4 interactions) can be quantified by comparing the distribution of dihedral angles (assuming that the conformational equilibrium is established on a timescale shorter than the simulation time). This comparison shows that conformational differences are found only when both atoms in the pair carry high partial charges, e.g., in the lipid head group, while the lipid tails have indistinguishable conformational properties.

The simplest property to look at in a lipid bilayer is the area per lipid, as this is correlated with all other structural parameters (Nagle and Tristram-Nagle 2000). As shown in table 1, the results are in reasonable agreement within the statistical error margins, with the Ash lipid parameters yielding slightly lower values. We attribute the difference in area between Berger lipids and Ash lipids to the coupling between angles and dihedrals, which is different in the presence and in the absence of 1–4 interactions. We also compared transverse bilayer density profiles and deuterium order parameters for the acyl chains; both were in good agreement (not shown). When, as a control, the original Berger parameters were used with OPLS-AA rules, the area rapidly shrinks by about 10%, with a simultaneous increase in the lipid chain ordering.

3.2. Peptide and protein simulations give a large difference in behaviour between the two force fields

We calculated the area per lipid in the presence of different hydrophobic helical transmembrane peptides. In this case, we did not take into account the area occupied by the peptide, and the numbers reported in table 1 represent the total area divided by the number of lipids on each leaflet. We did not attempt to isolate the area of the peptide and the area of the lipid, as this is difficult to define rigorously. For the direct comparison between equivalent systems this is not an issue, however, and our 'normalized' areas are a practical measure. All of the peptides formed stable transmembrane helices during all simulations as verified by a number of standard analyses, including secondary structure analyses (details not shown).

In the presence of peptides, the area per lipid appears to shrink consistently in all the simulations. When using the ffgmx force field and the Berger lipid parameters, the shrinkage is approximately by 10–12% in the case of a single transmembrane peptide and 7% in the case of two WALP23 peptides, despite the fact the peptides are present in addition to the same

Table 2. Partial volume of WALP23 peptide in water and octane, with different force fields.

	No peptide					With peptide			
Force field combination	T	n sol	V (nm^3)	V/mol (Å^3)	Density	n sol	V sol	V_{tot} (nm^3)	$V_{peptide}$ (nm^3), std. err.
SPC–ffgmx	298	2054	62.08	30.23	989.7	8228	248.70	256.51	7.81 (0.02)
TIP3P–OPLS-AA	298	2054	61.30	29.84	1002.5	8236	245.80	254.94	9.15 (0.02)
TIP4P–OPLS-AA	298	2125	63.21	29.74	1004.9	8105	241.08	248.51	7.44 (0.03)
Octane CH2–ffgmx	298	200	49.77	248.83	762.3	817	203.29	207.44	4.14 (0.03)
Octane LP2–ffgmx	298	200	58.89	294.44	644.3	817	240.55	243.57	3.01 (0.02)
Octane LP2mod–ffgmx	298	200	58.88	294.42	644.3	817	240.54	244.10	3.56 (0.02)
Octane LP2–OPLS-AA	298	200	58.89	294.44	644.3	812	239.08	242.80	3.71 (0.02)
Octane—experiment	293				702.0				

number of lipids as in the pure bilayer. When OPLS-AA was used for the protein and the Ash parameters for the lipids, the shrinkage was approximately 2–4% in the case of a single peptide and less in the case of two peptides. In both OmpF simulations (figure 2), the protein remained stable (total protein RMSD relative to the starting structure 0.25 nm (ffgmx) and 0.19 (OPLS-AA)), but as with the smaller peptides, our modified lipids combined with OPLS-AA gave a larger total bilayer area (table 1).

3.3. Partial volumes

In order to understand the reason for the large decrease in the area per lipid upon addition of peptides, we calculated the partial volumes for the WALP23 peptide in water and octane using different force fields (table 2). During all the simulations the WALP23 peptide maintains its α-helical structure, except for the octane–CH2 simulation, in which unfolding of the peptide occurs within a few nanoseconds. A moderate decrease of the partial volume may be expected when the peptide is embedded in octane, compared to water, due to the high hydrophobicity of the peptide; octane interacts with the peptide more strongly compared to water.

The highest reduction in the partial volume is observed when LP2/LP3 atom types are used in combination with the ffgmx force field for the peptide, and the Lennard-Jones parameters for the interactions between the peptide and the octane are calculated based on CH2/CH3 atom types instead of LP2/LP3 atom types. This is the most common set of parameters used in simulations of peptides and lipids with the ffgmx force field. The reason for the large effect is that interactions between CH2/CH3 atom types are too highly attractive, as demonstrated by the high density of pure octane obtained with these parameters.

The lowest reduction is observed when CH2/CH3 atom types from the ffgmx force field are used to describe octane and the ffgmx force field is used for the peptide. In this case, the peptide unfolds rapidly; this exposes the amide groups (at least partially) to the solvent, which decreases the hydrophobicity of the peptide and causes an increase in its partial volume.

If we exclude the octane–CH2 simulation set, the lowest reduction in the partial volume of the peptide is found in the simulation using the Berger lipid parameters and the OPLS-AA force field for the peptide; in this case, the secondary structure is stable for the whole duration of the simulation, and the higher partial volume is due to slightly weaker interactions between the peptide and the octane. The Lennard-Jones parameters for these interactions are calculated based on LP2/LP3 atom types and OPLS-AA atom types for the peptide, using combination rules.

Table 3. Cyclohexane to water transfer free energies of the OPLS-AA force field.

		OPLS-AA[b]		OPLS-AA w/Berger lipids		OPLS-AA + modified charges w/Berger lipids	
Residue	Experiment[a] ΔG (kJ mol^{-1})	ΔG (kJ mol^{-1})	Error	ΔG (kJ mol^{-1})	Error	ΔG (kJ mol^{-1})	Error
Asn	−27.7	−21.4 ± 1.0	6.3	−21.0 ± 0.9	6.7	−25.6 ± 0.6	2.1
Arg	−24.2	−23.6 ± 2.1	0.6	−19.8 ± 1.8	4.4		
Gln	−22.9	−13.0 ± 1.9	9.9	−19.4 ± 1.6	3.5	−20.2 ± 0.8	2.7
His	−18.7	−7.5 ± 1.3	11.2	−8.2 ± 1.2	10.5	−31.1 ± 1.2	−12.4
Asp	−18.6	−17.3 ± 1.0	1.3	−17.3 ± 0.9	1.3	−13.1 ± 0.6	5.5
Ser	−14.2	−14.4 ± 0.9	−0.2	−14.1 ± 0.8	0.1		
Glu	−13.0	−13.7 ± 1.1	−0.7	−13.2 ± 0.8	−0.2	−8.9 ± 0.7	4.1
Thr	−11.1	−12.1 ± 1.1	−1.0	−9.8 ± 0.9	1.3		
Lys	−1.6	5.2 ± 1.5	6.8	10.0 ± 1.3	11.6		
Tyr	−0.8	4.8 ± 1.7	5.7	5.8 ± 1.4	6.6	−0.6 ± 1.6	0.2
Cys	5.2	8.1 ± 1.0	3.0	7.6 ± 0.9	2.4		
Ala	7.7	8.4 ± 0.7	0.7	10.4 ± 0.6	2.7		
Trp	9.5	15.5 ± 1.8	6.0	14.5 ± 1.7	5	4.3 ± 1.2	−5.2
Met	9.7	9.1 ± 2.7	−0.6	10.3 ± 2.1	0.6		
Phe	14.1	18.8 ± 1.8	4.7	19.8 ± 1.6	5.7		
Val	16.7	17.8 ± 1.4	1.1	20.6 ± 0.9	3.9		
Ile	20.2	20.7 ± 1.4	0.5	26.4 ± 1.1	6.2		
Leu	20.5	22.8 ± 1.4	2.4	26.2 ± 1.2	5.7		
Average error			3.2		4.3		2.4[c]

[a] Taken from Radzicka and Wolfenden (1988).
[b] Taken from MacCallum and Tieleman (2003).
[c] Average error includes error from OPLS-AA residues that were not modified.

We also ran simulations using the ffgmx force field for the peptide, LP2/LP3 atom types for the octane and Lennard-Jones parameters for peptide–octane interactions based on LP2/LP3 atom types, instead of CH2/CH3. The partial volume obtained in this case is close to the one obtained with OPLS-AA and the Berger lipid parameters.

3.4. Side chain free energies of transfer between water and cyclohexane

The results of the water–cyclohexane transfer free energies are given in table 3. All force fields tested systematically overestimate the water–cyclohexane transfer free energy. The results using the Berger lipid parameters to model the cyclohexane are slightly worse than those obtained using the OPLS-AA force field to model the cyclohexane. The average error per residue increases from 3.2 to 4.3 kJ mol^{-1}. In particular the residues which have large aliphatic groups (Ile, Leu, Val, Arg, Lys) appear to be more hydrophobic than when using OPLS-AA cyclohexane.

4. Discussion

4.1. Combining the OPLS-AA protein force field with united-atom lipids

We have replaced 1–4 interactions (described in the Berger parameter set) with Ryckaert–Bellemans dihedral potentials in order to obtain the same lipid properties as in the well-tested Berger lipids. This allows the combination of the united-atom lipid force field of Berger *et al* with the OPLS-AA protein force field (as well as any force field with the same functional form). Both the conformational properties of the DOPC head group and the bilayer properties

display differences between the Berger parameters and the Ash parameters in the absence of peptides. The observation that differences in area per lipid, density profile and order parameters are minor suggests that the properties of a lipid bilayer are not very sensitive to the details of the conformation of the head groups.

In the test simulation of WALP23, Leu22, and OmpF, both force field combinations behave reasonably. OPLS-AA gives a somewhat smaller RMSD for OmpF, staying closer to the experimentally determined crystal structure. Overall, however, we have no basis based on these simulations to decide which protein force field is better, and our goal was merely to make it possible to use OPLS-AA in lipid simulations.

4.2. Lipid–protein interactions

Although the protein/peptide properties do not differ much for our test cases of single helices, models of helical dimers and OmpF porin, the lipid properties show dramatic differences between the different combinations of force fields. The most commonly used combination of the GROMACS ffgmx force field with a set of combination parameters introduced seven years ago (Tieleman *et al* 1999a, 1999b) gives a dramatic change in lipid properties compared to other combinations. The shrinkage appears to be due mainly to the fact that interactions between the hydrophobic peptides and DOPC lipids are stronger than interactions of lipids with themselves. This effect is also evident from the low density of alkanes simulated with the Berger lipid parameters, and simulations with OmpF.

Because the lipid properties for pure lipids show some differences in the two parameter sets with and without 1–4 interactions, this could be attributed both to different protein–lipid interactions and to differences in the lipids themselves. In order to understand which contribution is more important, we carried out partial volume calculations with different combinations of protein and lipid force fields. In the simulations of WALP23 in octane, the volume of the peptide is a measure for the interactions between lipid parameters and protein parameters. When the octane is modelled by the same carbon parameters as in the peptide, the volume of WALP23 is the highest because the peptide unfolds, exposing polar amide groups to the solvent. More interesting are the simulations in which the octane is modelled by carbon parameters taken from the Berger lipid force field, while three different force fields are used for the peptide: (1) ffgmx, (2) a modified version of ffgmx with combination rules and (3) OPLS-AA; because the interactions between octane molecules are identical in all three cases, the difference in the partial volume of the peptide can only be attributed to differences in the peptide–octane interactions. When ffgmx is used for the peptide, with the most commonly used combination parameters for lipid–protein interactions, the volume is unreasonably low. When straightforward mixing rules are used, both with ffgmx and with OPLS-AA protein force field, the volume is intermediate. This mixing rule is likely to be a better approach than the parameters commonly used in simulations of membrane proteins with GROMACS. In principle, further improvement should be possible by balancing lipid–protein interaction parameters, provided the difference between the protein force field and the lipid force field is not too large. OPLS-AA appears to behave reasonably well in this respect when combining with the Ash lipid parameters. This is not surprising, considering that the Berger (and therefore the Ash) lipid parameters ultimately contain a large number of united-atom OPLS parameters.

4.3. Free energies of transfer

We previously calculated the free energy of transfer of OPLS-AA side chain analogues between water and cyclohexane (MacCallum and Tieleman 2003), using the standard OPLS-AA parameters for aliphatic hydrocarbons and following the work of Villa and Mark, who

calculated these energies for a GROMOS96 force field (Villa and Mark 2002). Shirts and Pande did the same calculation for a number of other force fields (Shirts *et al* 2003). In all cases, there are substantial errors in the solvation free energies, both in water and cyclohexane. We recently refitted the partial charges on the OPLS-AA residues that gave the largest error for the free energy of solvation in water, which improved the hydration free energies substantially under the conditions used (Xu *et al* 2006). These conditions are not exactly the same as those used in parameterizing OPLS-AA, but are useful for membrane simulations. Here we calculated the solvation free energy in cyclohexane using the Berger lipid parameters for cyclohexane, combined both with the standard OPLS-AA force field for proteins and with the reparameterized OPLS-AA parameter set (Xu *et al* 2006). Although the hydration energy has been improved, the effect on transfer free energies is mixed because some compensation of errors no longer occurs. Overall, the difference between cyclohexane modelled by OPLS-AA parameters and cyclohexane modelled by the Berger united-atom lipid parameters is modest compared to the absolute magnitude of the errors.

Practically, the fact that the transfer free energies are systematically too high for all combinations studied may or may not present a problem depending on the particular system being studied. Most simulations of membrane proteins should be relatively insensitive to the force field being too hydrophobic on average. Transmembrane helices are highly hydrophobic already and an increase in hydrophobicity is unlikely to cause any qualitative change in membrane protein behaviour. However, there are many systems that will be very sensitive to errors in transfer free energies. For example, the equilibrium between surface bound and transmembrane forms of a peptide could be very sensitive to changes in the force field. Although the transfer free energies are slightly worse, the use of OPLS-AA proteins with the Berger lipid parameters is reasonable for general membrane protein simulations. The errors are unlikely to make much qualitative difference and there is a very large increase in simulation speed considering all-atom versus united-atom lipids. For systems that are more sensitive to the transfer free energies it would be preferable to use a combination with a lower average error and no extreme errors.

A major advantage of a united-atom hydrocarbon force field is the greatly reduced computational cost compared to an all-atom hydrocarbon force field. To illustrate this, we carried out a simulation of united-atom cyclohexane (554 molecules) and one of all-atom cyclohexane (554) molecules, with the same simulation parameters and about the same density. The united-atom simulation took 4.3 h ns^{-1} per nanosecond, the all-atom simulation 66.9 h ns^{-1}, more than 15 times as long. Both were done on the same Xeon 3.06 GHz processor, using GROMACS 3.1.4, a fairly recent GROMACS version. There are several reasons for this difference in speed. The most obvious reason is that all-atom cyclohexane (C_6H_{12}) has three times as many atoms compared to united-atom cyclohexane (six CH_2 groups), but this accounts for a factor of 3 at most (actually less, due to the way neighbour searching and some parts of the code are optimized to work on small groups of atoms instead of individual atoms). The second reason is that the number of pair interactions in a given volume is now much higher. In the united-atom case, there is only one interaction between each pair of two CH_2 groups, whereas there are nine pairs between two all-atom CH_2 groups. This is the most important additional cost. Finally, in the case of united-atom CH_2 each group is neutral, and no electrostatic interactions need to be calculated, whereas in the case of all-atom CH_2 each atom has a partial charge that requires accounting for. In a lipid bilayer, only a fraction of the system is actually hydrocarbon. Depending on the amount of water and the size of the proteins in the system, reasonable estimates range from 10% to 50% lipid atoms in the total number of atoms, so the overall increase in computational cost for all-atom versus united-atom for the same number of lipids will be significantly lower than the extreme 15 times for cyclohexane.

If a united-atom would be demonstrably less accurate than an all-atom force field, a case-by-case argument would have to be made whether the extra accuracy of an all-atom force field is worth the additional computational effort, and whether this effort might not be better spent on additional control simulations or better sampling (longer simulation). However, it is currently challenging to show that an all-atom force field is better for lipids.

First, the best lipid force fields at the moment all have issues that can be improved, likely without running into the limitations imposed by a united-atom model for quite a while.

Second, we believe it is likely that there will be more fundamental constraints that cannot easily be solved by either an all-atom or a united-atom force field. As an example, a recent parameterization of the GROMOS force field reported that it was impossible to obtain accurate solvation free energies for both water and cyclohexane at the same time, presumably due to the lack of polarizability in most current force fields (Oostenbrink *et al* 2004). An all-atom force field has more degrees of freedom, and it will be interesting to see if it is possible to parameterize an all-atom force field to give good agreement in this case or whether this is a fundamental limitation of the potential function used.

Third, we lack accurate experimental data that would help identify shortcomings in lipid models. There are accurate data on lipids from NMR and from diffraction experiments, but in both cases these are averages that hide many details. In the case of lipid–protein interactions the need for accurate experimental information is even more pressing. There are only a handful of peptides for which there is sufficiently detailed experimental information. We are pursuing several of them, including WALP23 (Petrache *et al* 2002, Sparr *et al* 2005), Ac-WLXLL peptides (Aliste *et al* 2003, Aliste and Tieleman 2005) from the hydrophobicity scale of Wimley and White (1996) and alamethicin. Gramicidin A is another interesting peptide for which there is extensive experimental information (Allen *et al* 2003, Andersen *et al* 2005). Detailed simulation studies of such systems combined with further development of parameters and perhaps potential functions are necessary to improve current force fields and methods.

In a sense, the use of united-atom lipids is akin in spirit to coarse-graining details that are not likely to be relevant for the properties of interest, a sound principle of physical modelling. Recently developed coarse-grained models proved to be useful in predicting properties of lipid bilayers (Marrink *et al* 2004), lipid phase transformations (Marrink *et al* 2005) and self-assembly of lipid monolayers (Lopez *et al* 2002). Why would we not want to apply the same principle to protein force fields? This questions falls outside the scope of this paper, but except for GROMOS, most major force fields have switched to all-atom representations. The extra computational cost of having an all-atom description of a protein in solution versus a united-atom description is negligible compared to the difference found for lipids, removing one substantial incentive to pursue a united-atom protein force field. Having a way of mixing united-atom lipids with all-atom proteins will allow studies to benefit from the substantial ongoing all-atom parameterization efforts on proteins and other biomolecules.

5. Conclusions

We explored some of the consequences of mixing different protein force fields with a united-atom lipid force field. The commonly used combination of ffgmx with Berger *et al* lipid parameters appears to significantly overestimate the strength of lipid–protein interactions. Straight combination rules are more realistic but still overestimate the strength of lipid–protein interactions. OPLS-AA parameters perform better. We calculated free energies of transfer for OPLS-AA between water and cyclohexane using the same hydrocarbon parameters as in the lipid force field, and find a similar degree of accuracy as for the united-atom ffgmx protein force field. Revised OPLS-AA parameters by Xu *et al* improve the hydration free energy of

side chains in water under the simulation conditions used, but only have a marginal average effect on free energies of transfer between water and hydrocarbons. This work highlights the complexities of obtaining accurate MD force field parameters for both aqueous and hydrophobic environments and the nontrivial nature of combining parameters from different force fields.

Acknowledgments

This work was supported by NSERC and CIHR. JLM and WLA hold NSERC Canada Graduate Studentships. In addition, JLM is supported by Alberta Ingenuity, and WLA by the Alberta Heritage Foundation for Medical Resarch (AHFMR). CK, ZX, and LM hold AHFMR postdoctoral fellowships. DPT is an AHFMR Senior Scholar, CIHR New Investigator, and Sloan Foundation Fellow. We thank WestGrid for use of computational resources. We thank John Nagle for discussions that led to the idea of looking at peptide volumes.

References

Abraham F F *et al* 2002 Simulating materials failure by using up to one billion atoms and the world's fastest computer: brittle fracture *Proc. Natl Acad. Sci. USA* **99** 5777–82

Alder B J and Wainwright T E 1957 *J. Chem. Phys.* **27** 1208

Aliste M P *et al* 2003 Molecular dynamics simulations of pentapeptides at interfaces: salt bridge and cation–pi interactions. *Biochemistry* **42** 8976–87

Aliste M P and Tieleman D P 2005 Computer simulation of partitioning of ten pentapeptides Ace-WLXLL at the cyclohexane/water and phospholipid/water interfaces. *BMC Biochem.* **6** 30

Allen T W *et al* 2003 Structure of gramicidin a in a lipid bilayer environment determined using molecular dynamics simulations and solid-state NMR data. *J. Am. Chem. Soc.* **125** 9868–77

Andersen O S *et al* 2005 Gramicidin channels *IEEE Trans. Nanobiosci.* **4** 10–20

Anezo C *et al* 2003 Methodological issues in lipid bilayer simulations *J. Phys. Chem.* B **107** 9424–33

Berendsen H J C *et al* 1981 Interaction models for water in relation to protein hydration *Intermolecular Forces* ed B Pullman (Dordrecht: Reidel) pp 331–42

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** 3684–90

Berger O *et al* 1997 Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature *Biophys. J.* **72** 2002–13

Brooks B R *et al* 1983 CHARMM—a program for macromolecular energy, minimization, and dynamics calculations *J. Comput. Chem.* **4** 187–217

Cornell W D *et al* 1995 A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules *J. Am. Chem. Soc.* **117** 5179–97

Darden T *et al* 1993 Particle mesh Ewald—an n.log(n) method for Ewald sums in large systems *J. Chem. Phys.* **98** 10089–92

Essmann U *et al* 1995 A smooth particle mesh Ewald method *J. Chem. Phys.* **103** 8577–93

Hess B *et al* 1997 LINCS: A linear constraint solver for molecular simulations *J. Comput. Chem.* **18** 1463–72

Jorgensen W L *et al* 1983 Comparison of simple potential functions for simulating liquid water *J. Chem. Phys.* **79** 926–35

Jorgensen W L and Tirado-Rives J 1988 The OPLS potential functions for proteins—energy minimizations for crystals of cyclic-peptides and crambin *J. Am. Chem. Soc.* **110** 1657–66

Jorgensen W L *et al* 1996 Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids *J. Am. Chem. Soc.* **118** 11225–36

Kandt C *et al* 2006 Setting up and running membrane protein simulations *Methods* at press

Karplus M and McCammon J A 2002 Molecular dynamics simulations of biomolecules *Nat. Struct. Biol.* **9** 646–52

Karplus M and Petsko G A 1990 Molecular dynamics simulations in biology *Nature* **347** 631–9

Lindahl E *et al* 2001 GROMACS 3.0: a package for molecular simulation and trajectory analysis *J. Mol. Modeling* **7** 306–17

Lopez C F *et al* 2002 Self-assembly of a phospholipid Langmuir monolayer using coarse-grained molecular dynamics simulations *J. Phys.: Condens. Matter* **14** 9431–44

MacCallum J L and Tieleman D P 2003 Calculation of the water–cyclohexane transfer free energies of neutral amino acid side-chain analogs using the OPLS all-atom force field *J. Comput. Chem.* **24** 1930–5

MacKerell A D 2004 Empirical force fields for biological macromolecules: overview and issues *J. Comput. Chem.* **25** 1584–604

MacKerell A D *et al* 1998 All-atom empirical potential for molecular modeling and dynamics studies of proteins *J. Phys. Chem.* B **102** 3586–616

Marrink S J *et al* 2004 Coarse grained model for semiquantitative lipid simulations *J. Phys. Chem.* B **108** 750–60

Marrink S J *et al* 2005 Simulation of gel phase formation and melting in lipid bilayers using a coarse grained model *Chem. Phys. Lipids* **135** 223–44

McCammon J A *et al* 1977 Dynamics of folded proteins *Nature* **267** 585–90

Miyamoto S and Kollman P A 1992 Settle—an analytical version of the shake and rattle algorithm for rigid water models *J. Comput. Chem.* **13** 952–62

Nagle J F and Tristram-Nagle S 2000 Structure of lipid bilayers *Biochim. Biophys. Acta* **1469** 159–95

Oostenbrink C *et al* 2004 A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6 *J. Comput. Chem.* **25** 1656–76

Petrache H I *et al* 2002 Hydrophobic matching mechanism investigated by molecular dynamics simulations *Langmuir* **18** 1340–51

Ponder J W and Case D A 2003 Force fields for protein simulations *Protein Simul.* **66** 27

Radzicka A and Wolfenden R 1988 Comparing the polarities of the amino-acids-side-chain distribution coefficients between the vapour-phase, cyclohexane, 1-octanol, and neutral aqueous-solution *Biochemistry* **27** 1664–70

Robertson K M and Tieleman D P 2002 Orientation and interactions of dipolar molecules during transport through OmpF porin *FEBS Lett.* **528** 53–7

Schuler L D *et al* 2001 An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase *J. Comput. Chem.* **22** 1205–18

Shirts M R *et al* 2003 Extremely precise free energy calculations of amino acid side chain analogs: comparison of common molecular mechanics force fields for proteins *J. Chem. Phys.* **119** 5740–61

Snow C D *et al* 2004 Trp zipper folding kinetics by molecular dynamics and temperature-jump spectroscopy *Proc. Natl Acad. Sci. USA* **101** 4077–82

Sparr E *et al* 2005 Self-association of transmembrane alpha-helices in model membranes: importance of helix orientation and role of hydrophobic mismatch *J. Biol. Chem.* **280** 39324–31

Stillinger F H and Rahman A 1974 Improved simulation of liquid water by molecular dynamics *J. Chem. Phys.* **60** 1545–57

Tieleman D P *et al* 1999a An alamethicin channel in a lipid bilayer: molecular dynamics simulations *Biophys. J.* **76** 1757–69

Tieleman D P *et al* 1999b Alamethicin helices in a bilayer and in solution: molecular dynamics simulations *Biophys. J.* **76** 40–9

Tieleman D P *et al* 2002 Analysis and evaluation of channel models: simulations of alamethicin *Biophys. J.* **83** 2393–407

van der Spoel D *et al* 1991–2005 *GROMACS User Manual*

van Gunsteren W F 1987 *GROMOS. Groningen Molecular Simulation Program Package* (Groningen: University of Groningen)

van Gunsteren W F and Berendsen H J C 1990 Computer simulation of molecular-dynamics—methodology, applications, and perspectives in chemistry *Angew. Chem. Int. Edn Engl.* **29** 992–1023

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** 548–53

Weiner S J *et al* 1984 A new force-field for molecular mechanical simulation of nucleic-acids and proteins *J. Am. Chem. Soc.* **106** 765–84

White S H 2004 The progress of membrane protein structure determination *Protein Sci.* **13** 1948–9

Wimley W C and White S H 1996 Experimentally determined hydrophobicity scale for proteins at membrane interfaces *Nat. Struct. Biol.* **3** 842–8

Xu Z *et al* 2006 Improving hydration free energies for OPLS-AA amino acid side chains: new atomic charges and an off-plane charge model for aromatic residues *J. Comput. Chem* at press