Practical Considerations for Building GROMOS-Compatible Small-Molecule Topologies

Justin A. Lemkul, William J. Allen, and David R. Bevan*

Department of Biochemistry, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, United States

Received August 30, 2010

Molecular dynamics simulations are being applied to increasingly complex systems, including those involving small endogenous compounds and drug molecules. In order to obtain meaningful and accurate data from these simulations, high-quality topologies for small molecules must be generated in a manner that is consistent with the derivation of the force field applied to the system. Often, force fields are designed for use with macromolecules such as proteins, making their transferability to other species challenging. Investigators are increasingly attracted to automated topology generation programs, although the quality of the resulting topologies remains unknown. Here we assess the applicability of the popular PRODRG server that generates small-molecule topologies for use with the GROMOS family of force fields. We find that PRODRG does not reproduce topologies for even the most well-characterized species in the force field due to inconsistent charges and charge groups. We assessed the effects of PRODRG-derived charges on several systems: pure liquids, amino acids at a hydrophobic−hydrophilic interface, and an enzyme−cofactor complex. We found that partial atomic charges generated by PRODRG are largely incompatible with GROMOS force fields, and the behavior of these systems deviates substantially from that of simulations using GROMOS parameters. We conclude by proposing several points as "best practices" for parametrization of small molecules under the GROMOS force fields.

INTRODUCTION

Preparation of small-molecule topologies for use in molecular dynamics (MD) simulations is of critical importance in studies of drug−enzyme and drug−receptor interactions, partitioning of drugs into and across membranes, among others. The principal challenge in including small molecules in these simulations is the effort necessary to produce a topology that is consistent with the underlying theory of the force field, which is often based on common macromolecules such as proteins. Thus, less common functional groups present in drug molecules and synthetic compounds are difficult to parametrize. Different force fields frequently used for biomolecular simulation, such as AMBER,[1−4] CHARMM,[5−7] OPLS,[8,9] and GROMOS,[10−14] have different functional forms and underlying assumptions, requiring different parametrization procedures and validation protocols. While parametrization of small molecules for AMBER and CHARMM force fields has been aided by the development of the general AMBER force field (GAFF)[15] and the CHARMM general force field,[16] topology generation for OPLS and GROMOS remains more challenging. Procedures involving quantum mechanical calculations and geometry optimization are thoroughly described for the OPLS-AA parameter set,[8,9] but such detailed information is not publicly available for the GROMOS96 parameter sets. Thus, investigators seeking to use the GROMOS96 force fields must rely on empirical parameter assignment and subsequent validation through the time-consuming process of thermodynamic integration,[14] a challenge that has long been recognized within the simulation field.[17]

The GROMOS96 force fields offer an advantage over the AMBER, CHARMM, and OPLS parameter sets in that they are based on the united-atom approach to parametrization. That is, nonpolar hydrogens are not explicitly represented, decreasing the number of atoms in the system appreciably and speeding up simulations. These force fields are attractive for long-time simulations and large sets of simulations that may be utilized in silico to analyze the dynamics of small molecules and potential drugs.

Automated topology generation is possible through a number of tools, including the very popular PRODRG server.[18,19] Topologies produced by this program have been used widely in studies of protein−peptide,[20] protein−ligand,[21,22] protein−lipid,[23] and small molecule−lipid interactions;[24,25] drug partitioning into lipid membranes;[26] and simulations of small molecules that inhibit amyloid aggregation.[27] Careful inspection of the topologies produced by PRODRG reveals that the charges and charge groups assigned to the functional groups of small molecules are often inconsistent with the same groups present in the GROMOS96 43A1 parameter library. The effects of these inconsistencies are unknown, providing the motivation for the present study. Several methods for generating GROMOS-compatible charges for refining PRODRG topologies have been proposed in the literature,[28−31] but thus far, to the best of our knowledge, no standard has been set for the accuracy and applicability of any of these methods.

Here, we evaluate the quality of small-molecule topologies generated by the automated server PRODRG,[18,19] using small molecules whose functional groups are described by the GROMOS96 43A1 parameter set.[11] We explore a variety of systems, including pure liquids, amino acids at a

* Corresponding author. E-mail: drbevan@vt.edu. Phone: (540) 231-5040. Fax: (540) 231-9070.

10.1021/ci100335w
Published on Web 11/30/2010

hydrophobic−hydrophilic interface, and a cofactor bound to an enzyme. By demonstrating the implications of the inherent deficiencies in automated parameter generation and by making recommendations for proper parameter development, we hope to elaborate on the best practices in small molecule topology generation under the GROMOS force fields.

METHODS

To assess the quality of small-molecule topologies in a range of scenarios, several systems were constructed: (1) condensed-phase and gas-phase systems of hexane, ethanol, and *p*-cresol; (2) several amino acids (alanine, asparagine, aspartate, isoleucine, lysine, and serine) at the interface of water and cyclohexane; and (3) the protein UDP-galactopyranose mutase (UGM) with bound flavin adenine dinucleotide (FAD). All systems were prepared and simulations conducted using facilities present in the GROMACS package, version 4.0.7.[32] One set of small-molecule topologies was generated using the PRODRG 2.5 server,[18,19] which is designed to produce topologies consistent with the GROMOS96 43A1 parameter set. The other topologies were generated using known GROMOS96 43A1[11] functional group charges and charge groups, as implemented in the GROMACS distribution. Bonded parameters and atom types for each molecule were kept the same between the different topologies, such that the comparisons involved only charges and charge groups assigned by PRODRG relative to those implemented in the GROMOS96 43A1 parameter set. Unless otherwise noted, all simulations employed a twin-range cutoff scheme for short-range nonbonded interactions, with the real-space contribution to Coulombic terms truncated at 0.8 nm and short-range van der Waals interactions truncated at 1.4 nm. All bond lengths were constrained using the LINCS method,[33] allowing a 2-fs time step. The neighbor list was updated every five simulation steps (10 fs). Long-range electrostatic interactions were calculated with the smooth particle mesh Ewald (PME) method.[34,35] Equilibration simulations utilized the Berendsen weak coupling method[36] to control temperature and/or pressure, while production simulations utilized the Nosé-Hoover thermostat[37,38] and Parrinello-Rahman barostat[39,40] to generate a rigorous NPT ensemble.

Condensed-Phase and Gas-Phase Systems. To measure various thermodynamic and physical properties of pure liquid systems as a function of the molecular topology, systems containing hexane, ethanol, and *p*-cresol were constructed by replicating a single molecule of each compound in three dimensions to create a cubic grid of 512 molecules. For gas-phase systems, molecules were separated by 50 nm and kept in place with weak position restraints, in accordance with GROMOS parametrization methodology, and plain cutoffs were used to calculate electrostatic interactions.[14] Steepest descent minimization was performed on gas-phase and condensed-phase systems alike, followed by 100 ps of isochoric−isothermal (NVT) equilibration at 298 K. Further equilibration under an isothermal−isobaric (NPT) ensemble was performed for 300 ps at the same temperature and 1 bar of pressure. Finally, each system was simulated for another 20 ns production run under the same conditions. Charges and charge groups for these molecules are illustrated in Figure 1.

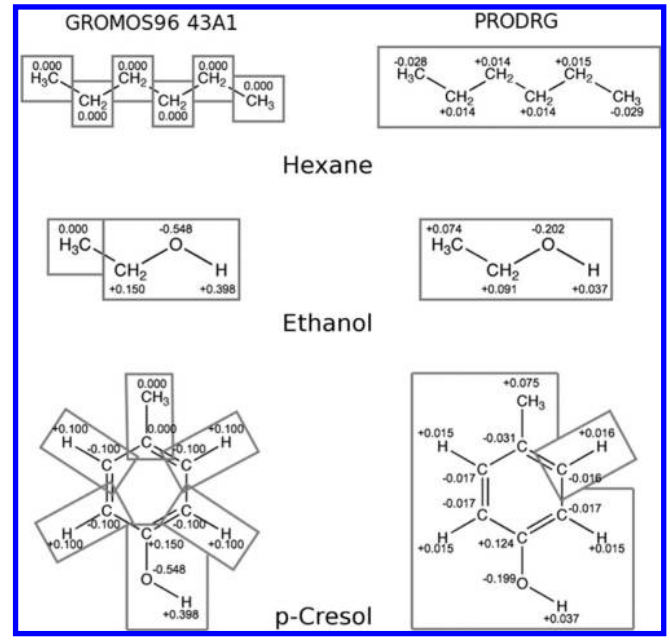

Figure 1. Comparison of GROMOS96 43A1 and PRODRG charges and charge groups for hexane (top), ethanol (middle), and *p*-cresol (bottom). Charges are given in units of *e* and charge groups are indicated by gray boxes.

Interfacial Amino Acid Systems. Amino acids represent some of the most common components of biomolecular force fields and thus should serve as reliable test systems. By using amino acids, we can assess the quality of PRODRG topologies against groups that have been rigorously defined. Six amino acids (alanine, asparagine, aspartate, isoleucine, lysine, and serine) were chosen to represent the 20 common amino acids. The chosen molecules encompass hydrophobic (Ala and Ile), polar uncharged (Asn and Ser), and charged (Asp and Lys) species. Each amino acid was modeled as a dipeptide, with N- and C-terminal groups capped with acetyl and *N*-methyl groups, respectively, to give uncharged termini. Since the goal of the GROMOS96 parameter sets is to accurately describe partitioning of amino acids between polar (water) and hydrophobic (cyclohexane) media,[14] we sought to produce a biphasic system that would model partitioning behavior of the different amino acids.

A 5-nm cubic box of cyclohexane was generated, containing a total of 466 molecules. Standard bond, angle, and dihedral parameters from the GROMOS96 43A1 library were assigned in the cyclohexane topology. All atoms were represented as united-atom CH_2 with zero charge. Following steepest descent minimization, the cyclohexane box was equilibrated first under an isochoric−isothermal (NVT) ensemble for 100 ps (298 K) and then under an isothermal−isobaric (NPT) ensemble for 500 ps (298 K, 1 bar), at which point the box vectors and density had stabilized. The system was simulated for an additional 10 ns to generate the cyclohexane layer used in building the interfacial systems.

Each amino acid was placed at the center of a rectangular box such that both the plane of the amino acid backbone and the Cα-Cβ bond were coincident with the $x-y$ plane of the unit cell (Figure 2). The equilibrated cyclohexane box was placed in the lower half of the simulation cell, while the remainder of the volume was filled with SPC water.[41] A single Cl^- counterion was added to the aqueous phase of the lysine system to compensate for the net charge of the

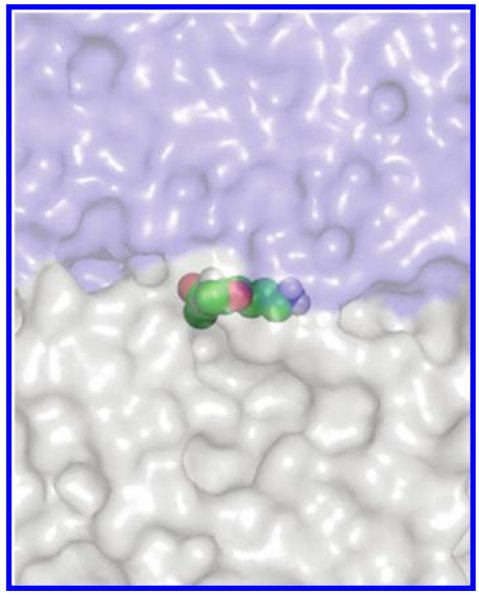

Figure 2. Initial position of lysine in the cyclohexane−water interfacial system. The amino acid backbone is perpendicular to the page and the side chain is extending to the right in the image. Lysine is shown with its atoms colored by element and rendered as van der Waals spheres. The cosolvents are rendered as transparent surfaces, with water in blue and cyclohexane in gray. The image was rendered with PyMOL.[56]

amino acid. In the case of aspartate, a single Na^+ ion was added. No ions were added to systems containing uncharged amino acids. Each system was energy-minimized and equilibrated in the same manner as the cyclohexane box discussed above, with position restraints applied to amino acid heavy atoms during equilibration. Three replicates of each system were generated from different random initial velocities. The data collection period lasted for 20 ns. Charges and charge groups used for the simulations of amino acids are given in Figures 3 and 4.

UDP-Galactopyranose Mutase. To examine the stability of a cofactor bound to a protein, the structure of UDP-galactopyranose mutase (UGM) bound noncovalently with flavin adenine dinucleotide (FAD) was utilized. The starting structure for these simulations was taken from the crystal structure of the *Klebsiella pneumoniae* enzyme−cofactor complex determined by Beis et al. (PDB 2BI7),[42] which contains an oxidized FAD cofactor. Parameters for the protein were assigned from the GROMOS96 43A1 parameter set, and FAD parameters were either taken from this same force field or generated by PRODRG (Figure 5). The protein−cofactor complex was centered in a rhombic dodecahedral unit cell that was subsequently filled with SPC water[41] and sufficient counterions to balance the charge and bring the final ionic concentration to a near-physiological 100 mM. Equilibration procedures similar to the previous systems were implemented. Steepest descent minimization was performed,

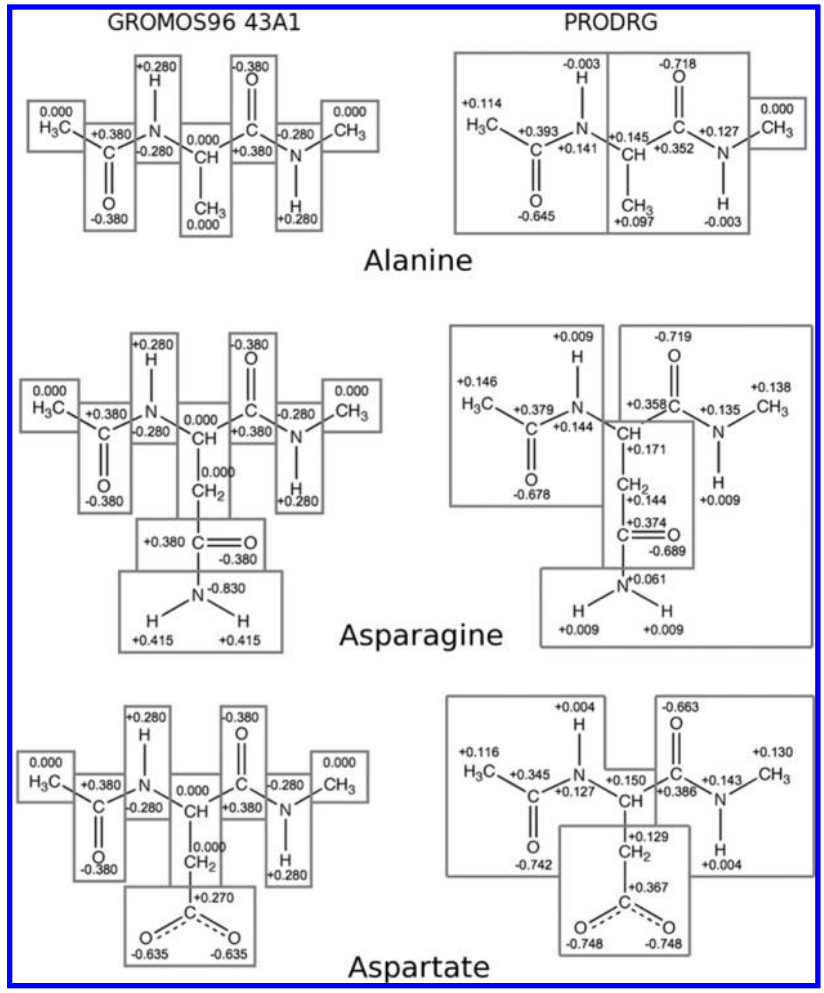

Figure 3. Comparison of GROMOS96 43A1 and PRODRG charges and charge groups for alanine (top), asparagine (middle), and aspartate (bottom). Charges are given in units of *e* and charge groups are indicated by gray boxes.

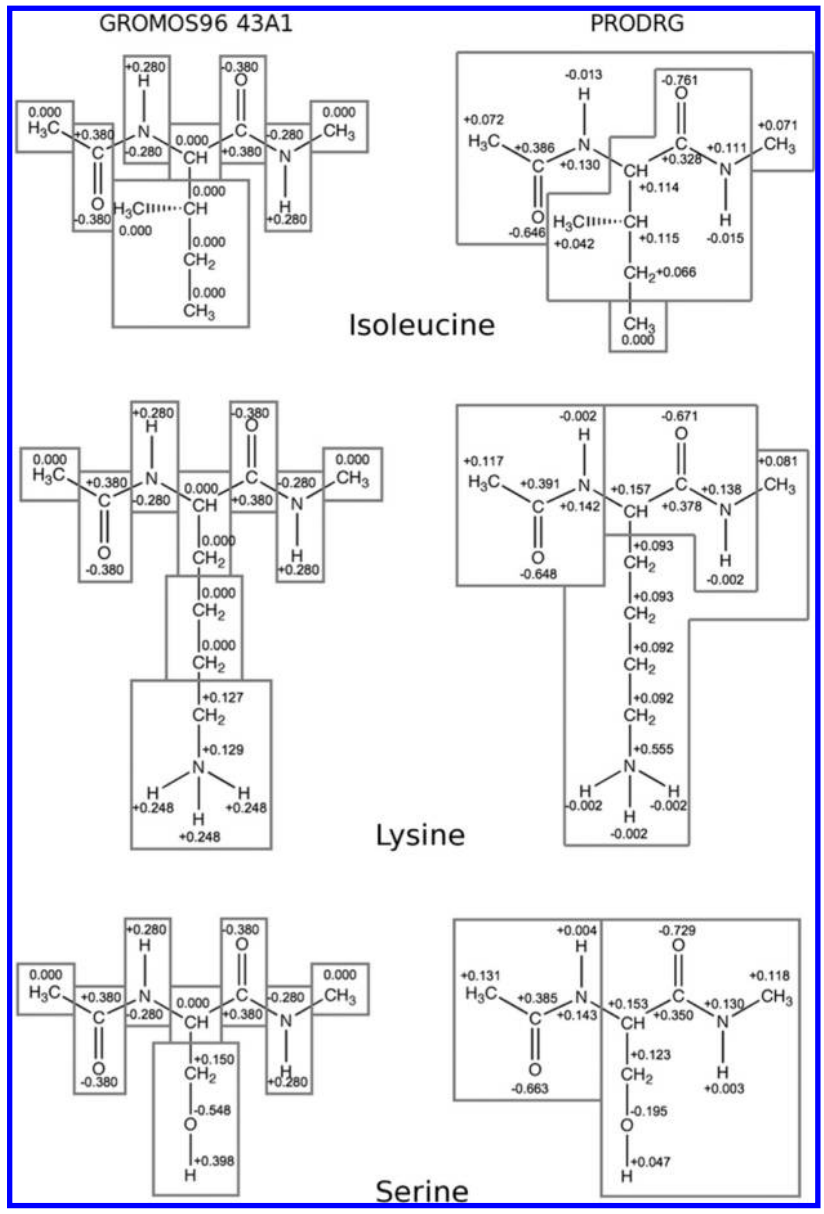

Figure 4. Comparison of GROMOS96 43A1 and PRODRG charges and charge groups for isoleucine (top), lysine (middle), and serine (bottom). Charges are given in units of *e* and charge groups are indicated by gray boxes.

followed by 100 ps of isochoric−isothermal (NVT) equilibration. Next, 500 ps of equilibration under an isothermal−isobaric (NPT) ensemble was performed, followed by 40-ns production simulations. Position restraints were applied to the backbone of the protein during equilibration and released for the production run. Different random initial velocities were used to generate independent trajectories.

RESULTS AND DISCUSSION

The goals of our approach include the assessment of several physical and thermodynamic parameters in the liquid- and gas-phase systems, behavior of amino acids in the presence of hydrophobic and polar media, and the interactions of an enzyme-bound molecule with its partner protein. These scenarios represent some of the most common uses of small-molecule topologies in the literature and provide us with versatile systems in which to test the quality of these topologies. We provide an analysis of the various intermolecular interactions that give rise to the observed behavior, focusing primarily on electrostatic and van der Waals interactions, hydrogen bonding, and relative orientation of the different molecules. While these quantities are largely interdependent, a comprehensive analysis of these terms is appropriate as they relate directly to quantities that are often evaluated in MD simulations.

For all comparisons between the results of using PRODRG and GROMOS96 43A1 topologies, the behavior observed in the GROMOS96 43A1 simulations is considered the benchmark for all comparisons. Parameters for novel species must be derived in a manner consistent with the existing, validated force field, which is taken as the standard for accuracy, and thus, we evaluate the results of using PRODRG parameters in this context.

Condensed-Phase and Gas-Phase Systems. The parametrization of the GROMOS96 force field parameter sets relies on accurate reproduction of condensed-phase behavior and thermodynamic data. In the most recent edition of the force field (53A6),[14] the criteria used to assess the accuracy of

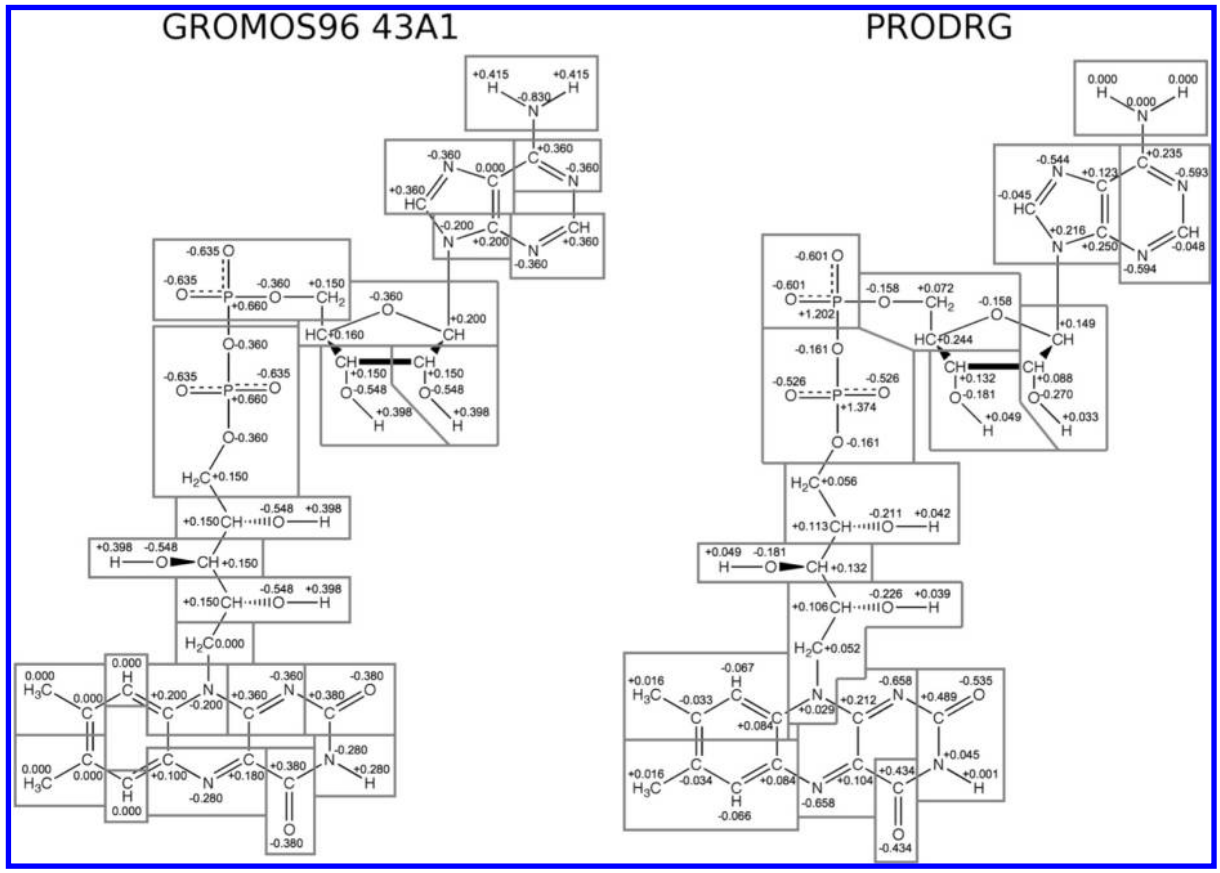

Figure 5. Comparison of GROMOS96 43A1 and PRODRG charges and charge groups for FAD. Charges are given in units of *e* and charge groups are indicated by gray boxes.

Table 1. Density and ΔH_{vap} for Small Molecules[54,55]

	density (g mL^{-1})			ΔH_{vap} (kJ mol^{-1})		
system	43A1	PRODRG	expt	43A1	PRODRG	expt
ethanol	0.766 ± 0.008	0.972 ± 0.007	0.789[a]	20.3 ± 0.2	9.4 ± 0.4	42.3[d]
hexane	0.681 ± 0.005	0.679 ± 0.005	0.661[b]	17.2 ± 0.3	17.2 ± 0.3	31.6[e]
p-cresol	1.097 ± 0.004	1.045 ± 0.005	1.019[c]	48.7 ± 0.5	35 ± 2	73.1[f]

[a] Reference 54, pp 3−206. [b] Reference 54, pp 3−308. [c] Reference 54, pp 3−134. [d] Reference 54, pp 6−111. [e] Reference 54, pp 6−117. [f] Reference 55.

the derived parameters included density and the heat of vaporization, ΔH_{vap}. These quantities can be calculated from MD simulations of hexane, ethanol, and *p*-cresol.

Table 1 summarizes the results of density and ΔH_{vap} calculations. The equation used to calculate ΔH_{vap} was

$$\Delta H_{vap} = \langle E_{gas} \rangle - \langle E_{liquid} \rangle + RT \qquad (1)$$

In eq 1, the quantities in angled brackets denote the time average of the total energy of the gaseous and liquid systems. R is the gas constant, and T is the absolute temperature in kelvin.

We found relatively good agreement between the GROMOS96 43A1 topologies and experimental data with respect to density (Table 1). Densities were slightly overestimated for hexane and *p*-cresol (by 3.0% and 7.7%, respectively) and slightly underestimated for ethanol (by 3.0%). For PRODRG-produced topologies, the densities of hexane and *p*-cresol were in close agreement with both the GROMOS96 43A1 results and experimental data (overestimated by 2.7% for hexane and 2.6% for *p*-cresol), but that of ethanol was significantly overestimated (by 23.2%).

With respect to ΔH_{vap} values, neither GROMOS96 43A1 nor PRODRG parameters closely reproduced experimental data, with all simulation results producing values that were well below the experimentally determined ΔH_{vap}. The values of ΔH_{vap} generated by GROMOS96 43A1 topologies, however, generally were in somewhat closer agreement with experimental values than were values generated using PRODRG parameters (Table 1).

To explain these results, we examined the structures of the liquids themselves in terms of both hydrogen bonding and radial distribution functions (RDF), which describe the density of chosen particles as a function of distance, a calculation that can be applied to individual atoms or to whole molecules. In an RDF plot, closely interacting groups are separated by shorter distances, such that the resulting density peak is higher, indicating a greater probability of interaction at a given distance.

The decreased value of ΔH_{vap} for ethanol and *p*-cresol resulting from PRODRG parameters can be attributed mainly to a significant decrease in hydrogen-bonding capacity. For ethanol, the use of GROMOS96 43A1 parameters resulted

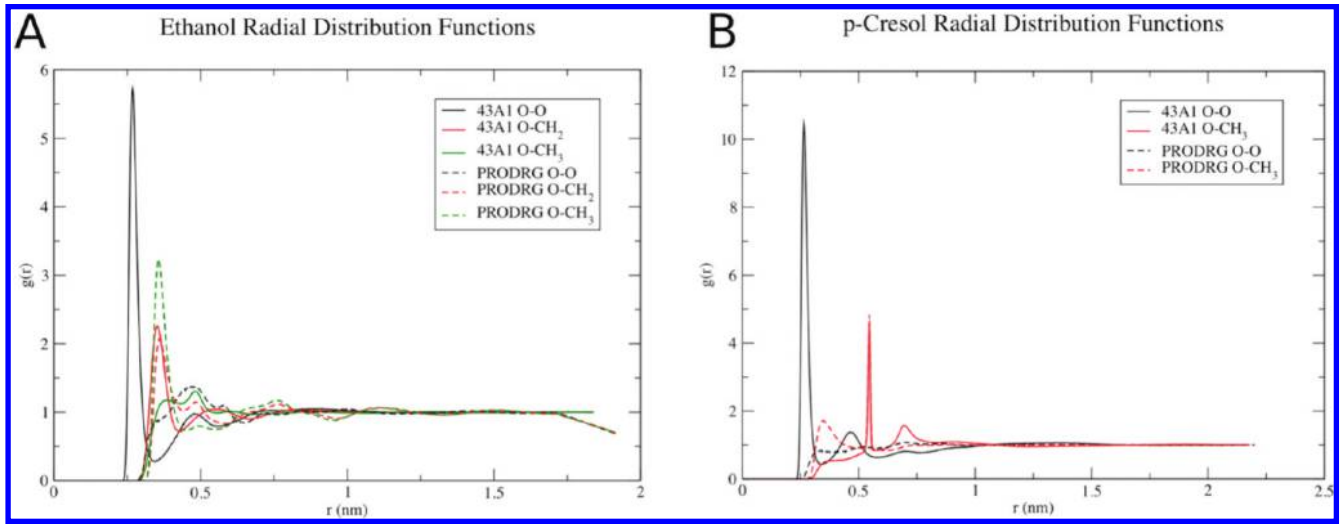

Figure 6. RDF for different atom pairs in (A) ethanol and (B) *p*-cresol. In part B, the intramolecular O−CH_3 peak referred to in the text appears as a single, solid line, since it is present in both curves and is thus overlapping.

in 0.93 ± 0.01 hydrogen bonds per molecule (averaged over time). For *p*-cresol, this value was 0.94 ± 0.01. Use of PRODRG parameters gave substantially lower hydrogen bonding for these two species, 0.12 ± 0.02 hydrogen bonds per ethanol molecule and only 0.05 ± 0.01 per *p*-cresol molecule. The RDF of these two species (Figure 6) provides insight into the reason for this observed behavior.

PRODRG parameters caused ethanol molecules to align head-to-tail, that is, with the hydroxyl group associating most prominently with the methyl group of a nearby molecule, at a distance of approximately 0.3 nm (Figure 6A). Interactions between hydroxyl groups occurred at a much greater distance, nearly 0.5 nm, exceeding the optimal distance for hydrogen bonding. The combination of a relatively large charge (+0.071 *e*) on the terminal methyl group and the comparatively small charge on the hydroxyl hydrogen (+0.037 *e*) leads to this behavior. The association of the hydroxyl oxygen atoms with the terminal methyl groups disrupted hydrogen bonding, decreased ΔH_{vap}, and increased the density relative to the results obtained using GROMOS96 43A1 parameters. Proper hydrogen bonding was observed using GROMOS96 43A1 parameters for ethanol, with hydroxyl−hydroxyl interactions occurring at approximately 0.3 nm, indicating strong hydrogen bonding and comparing closely with the RDF of previous simulations using different parameters.[43]

In the case of *p*-cresol, we also observed a reduction in hydrogen bonding when PRODRG parameters were applied, as well as an increase in hydroxyl−hydroxyl distance relative to the results obtained using GROMOS96 43A1 parameters (Figure 6B). Under PRODRG parameters, there was no distinct hydroxyl−hydroxyl RDF peak, but there was a probable interaction between the hydroxyl group and the *p*-methyl group occurring at approximately 0.35 nm, indicating head-to-tail orientation similar to the case of ethanol. Use of GROMOS96 43A1 parameters showed this behavior as well, but at a much larger distance, approximately 0.7 nm, while hydrogen bonding was clearly occurring due to hydroxyl−hydroxyl interactions at approximately 0.25 nm. A peak at approximately 0.55 nm occurred in the RDF produced by both PRODRG and GROMOS96 43A1 parameters; this peak is the intramolecular distance between the hydroxyl group and the *p*-methyl group.

RDF plots for hexane (data not shown) indicate that there was no distinguishable difference in the liquid structure due to the application of GROMOS96 43A1 or PRODRG parameters. The average end-to-end distance of hexane was 0.607 ± 0.001 nm in both cases, with a radius of gyration of 0.216 ± 0.003 nm. Thus, the presence of small partial charges on hexane assigned by PRODRG did not significantly influence the configurations of hexane or the structure of the bulk liquid.

Interfacial Amino Acid Systems. To assess the quality of PRODRG-generated amino acid topologies relative to those present in the GROMOS96 43A1 force field library, several parameters were analyzed: the position and orientation of the amino acids in the biphasic cyclohexane−water system, the nonbonded energies (Coulombic and Lennard-Jones) between the amino acids and the two solvents, and the ability of relevant functional groups in the amino acids to hydrogen bond with water. The balance of these terms will dictate the behavior of the amino acids in these systems and thus is an important measure of the quality of the small-molecule topologies.

Configurations and Trajectory Observations. The position of amino acids parametrized under the GROMOS96 43A1 force field was largely consistent over time, with all amino acids remaining at the cyclohexane−water interface. The only exception was one simulation of isoleucine, in which the molecule diffused into the cyclohexane layer to localize at the cyclohexane−water interface that is formed at the periodic boundary of the system in the *z*-dimension (Figure 7). The orientation of the side chains varied depending on the chemical nature of the amino acid. Hydrophobic amino acids (alanine and isoleucine) were oriented such that their side chains were either coincident with the interface or buried in the cyclohexane layer. Polar uncharged amino acids (asparagine and serine) were positioned with their side chains coincident with the interface or directed toward the aqueous solvent. Charged amino acids (aspartate and lysine) consistently oriented their side chains toward the aqueous solvent.

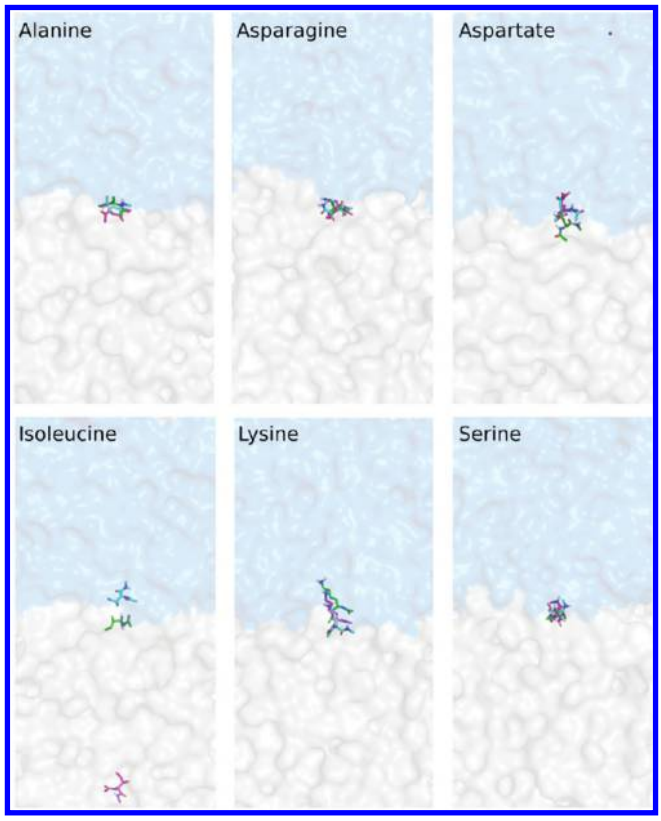

Figure 7. Overlays of amino acid positions generated by GROMOS96 43A1 parameters, taken from the final snapshot (20 ns) of each simulation after fitting for global $x-y$ translation. Amino acids are shown as sticks and colored separately by replicate. Water and cyclohexane are rendered as in Figure 2. The images were generated with PyMOL.[56]

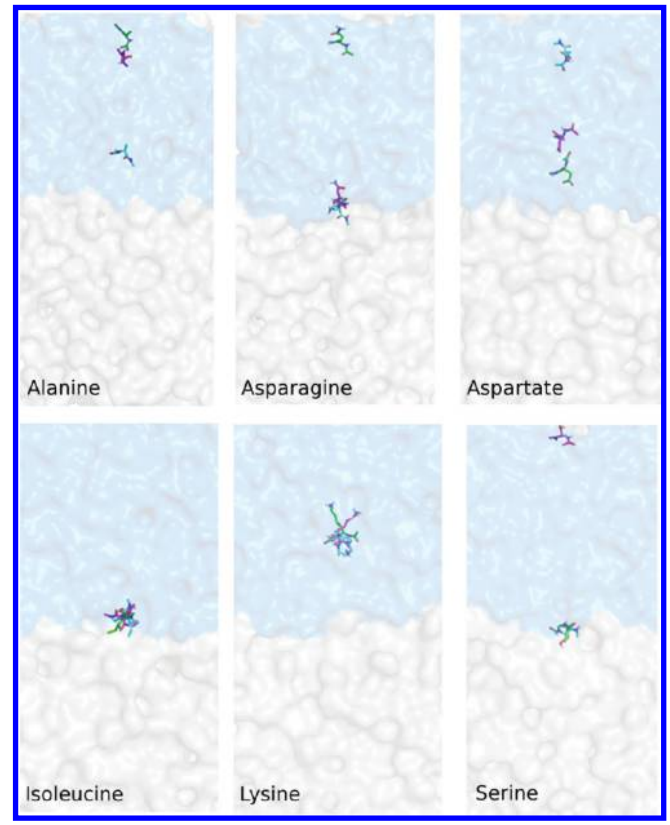

Figure 8. Overlays of amino acid positions generated by PRODRG parameters, taken from the final snapshot (20 ns) of each simulation after fitting for global $x-y$ translation. Amino acids are shown as sticks and colored separately by replicate. Water and cyclohexane are rendered as in Figure 2. The images were generated with PyMOL.[56]

The behavior of the amino acids whose parameters were assigned by the PRODRG server was less consistent (Figure 8). In all three simulations of alanine, the amino acid diffused freely into the water layer. Simulations of isoleucine parametrized by PRODRG performed comparably to the simulations under the GROMOS96 43A1 parameters, with isoleucine remaining at the cyclohexane−water interface and its side chain buried in the cyclohexane layer. The position of asparagine across three simulations was inconsistent; in two cases, asparagine remained largely interfacial, occasionally diffusing as much as 3 nm into the aqueous phase before returning to the interface. In the third simulation of asparagine, the amino acid diffused to the top of the unit cell to associate with the cyclohexane−water interface at the periodic boundary. In the simulations of serine, two systems remained interfacial with their polar side chains buried in the cyclohexane layer. In the third serine simulation, the amino acid diffused to the top of the unit cell to ultimately bury its side chain into the periodic cyclohexane layer. In simulations of both lysine and aspartate, the amino acids diffused freely into the aqueous phase; their positions were variable over time.

Energetics of Interactions with Cosolvents. The interactions of the amino acids with the two cosolvents of the interfacial systems were strongly influenced by the nature of the electrostatic and van der Waals interactions between the small molecules and the aqueous and hydrophobic environments. The balance between these terms dictated the partitioning behavior of the amino acids. Since the cyclohexane topology included no charges on the constituent atoms, the only electrostatic interactions in these systems (aside from water−water interactions) occurred between the amino acid and water. In all cases, the charges assigned in PRODRG-generated topologies led to significantly lower energy (more favorable) electrostatic interactions with water than the results obtained using GROMOS96 43A1 parameters (Figure 9A).

As a consequence of the greater interaction with the aqueous phase in the PRODRG-parametrized systems, atoms were drawn closer together by stronger electrostatic interactions. As a result, van der Waals interactions with the aqueous phase became more repulsive for all amino acids except for lysine, which experienced a greater net Lennard-Jones attraction (Figure 9B). RDF analysis (Figure 10) indicates that more water molecules interacted with the side chain methylene groups when PRODRG charges were applied. These groups are normally uncharged in the GROMOS96 43A1 parameter set, leading to weaker solvent interactions. The oxygen atom in water interacted with each of the side chain methylene groups when small partial charges (assigned by PRODRG) were present such that favorable van der Waals interactions were established, a behavior that was promoted by favorable dipole interactions from the partial charges present.

For all the other amino acids considered here, the use of PRODRG parameters caused water molecules to be distributed around the solute molecules at either a closer distance or at the same distance but with higher probability than with GROMOS96 43A1 parameters, based on RDF analysis (data

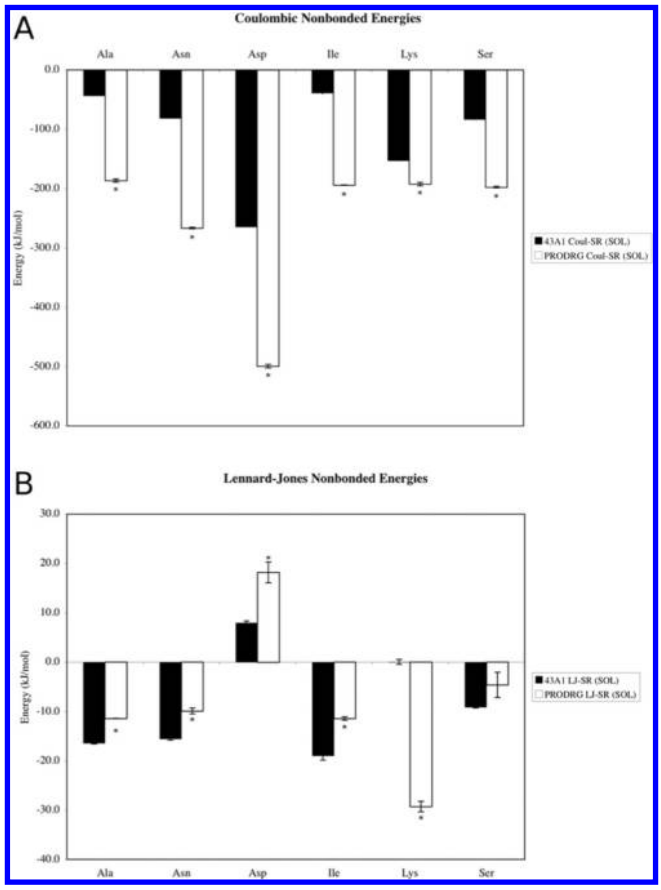

Figure 9. (A) Electrostatic and (B) Lennard-Jones interaction energies between the amino acids (listed on the *x*-axis) and water. Asterisks (*) indicate a significant difference based on a two-tailed Student's *t* test ($p < 0.05$). Error bars represent standard deviations of the three simulations of each system.

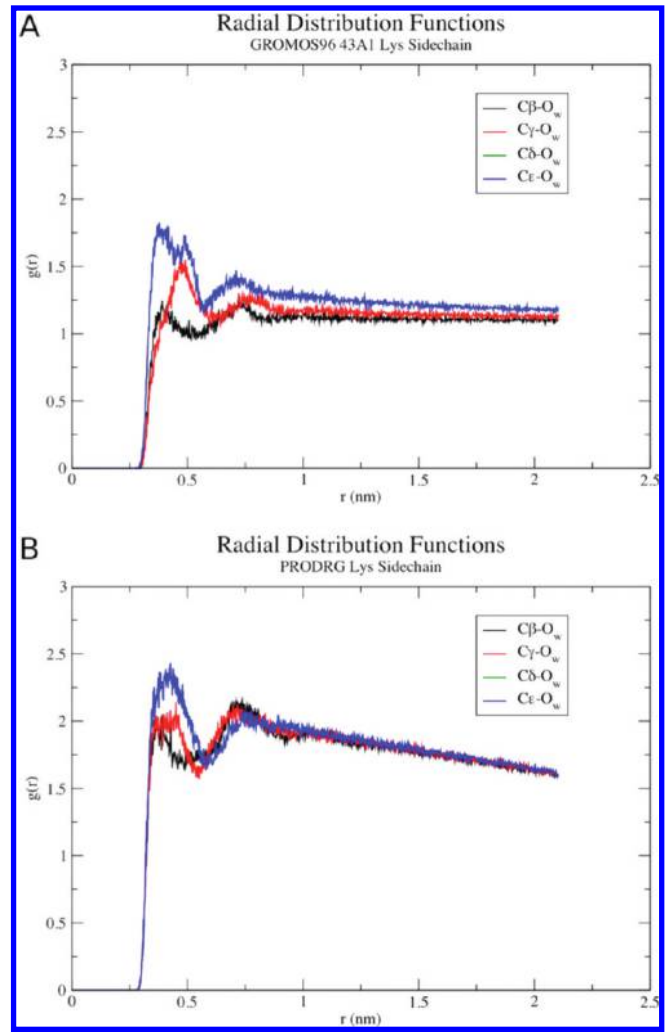

Figure 10. RDF for SPC oxygen (O_w) atoms with respect to lysine side chain atoms using (A) GROMOS96 43A1 and (B) PRODRG parameters. In both panels, the curves for Cδ and Cε atoms are overlapping.

not shown). The attraction of water molecules to these species derived largely from the overpolarization of the backbone and/or side chains of each amino acid in the PRODRG topologies. The result was a net repulsion due to an increased density of water around these amino acids, particularly at distances between atomic pairs above the energy minimum on the Lennard-Jones curve.

Hydrogen Bonding. All amino acids parametrized under GROMOS96 43A1 formed essentially the same number of backbone hydrogen bonds with water (approximately 2), averaged over all simulations and all snapshots in each simulation (Figure 11A). The consistency across simulations is likely due to the fact that the backbone amide groups had identical parameters, in accordance with the transferability of functional groups within the GROMOS parameter sets. That is, in the GROMOS force fields, the same functional groups have the same charges, even if they occur in different species, since charges are derived from model species that represent a small portion of an existing compound, typically an amino acid side chain.[14] For all the amino acids examined here, the PRODRG-produced topologies led to a significant increase in the number of backbone−water hydrogen bonds relative to the results obtained with GROMOS96 43A1 parameters, nearly six in the case of aspartate. This effect can be attributed to the greater separation of charges in the backbone amide groups (Figures 3 and 4). These charges vary somewhat in their magnitude depending on the amino acid, but in all cases the resulting dipoles are greater than those present in the GROMOS96 43A1 topologies.

Side chain hydrogen bonds are another important consideration when examining these systems. Illustrated in Figure 11B are hydrogen bonds between side chain groups (where applicable) and water. In the cases of asparagine and aspartate, significantly more hydrogen bonds formed between side chain groups (amide and carboxylate, respectively) in the case of PRODRG-generated topologies. The principal reason for this increase is the greater magnitude of charge separation on the amide carbonyl in asparagine and carboxylate in aspartate. The stronger dipoles in these groups led to more hydrogen bonding with nearby water. The opposite effect is observed in the case of lysine and serine systems, wherein hydrogen bonds were significantly decreased relative to the system simulated with GROMOS96 43A1 parameters. In the case of lysine, the charges on the ε-amino group hydrogens assigned by PRODRG are −0.002 *e*, opposite in sign and significantly reduced in magnitude compared to the charges that are assigned under GROMOS96 43A1 (+0.248 *e*). The partial negative charges assigned by PRODRG likely inhibit the ability of this ε-amino group to effectively serve as a hydrogen-bond donor. As described above, in the simulations of serine, the PRODRG-derived

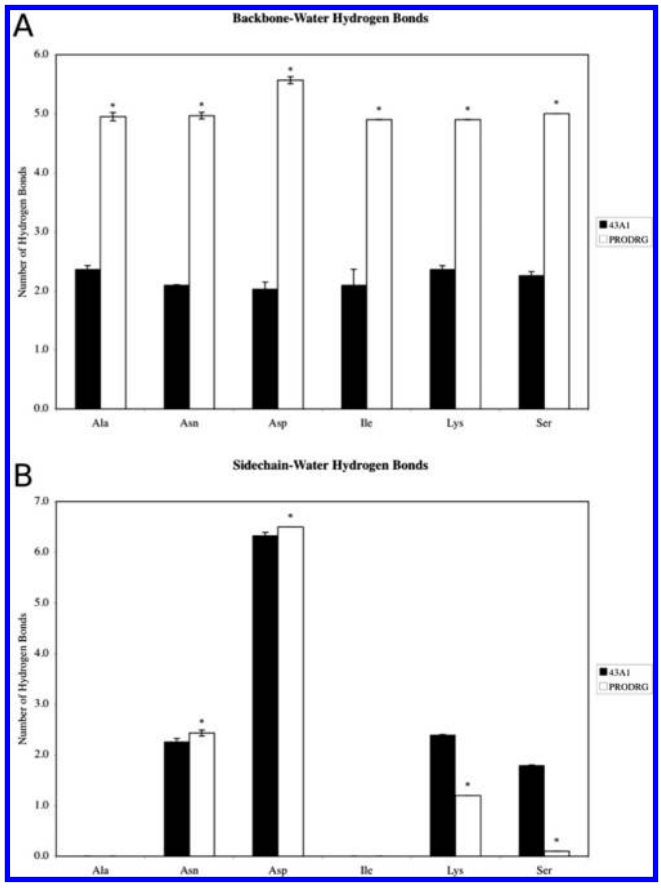

Figure 11. (A) Backbone and (B) side chain hydrogen bonds to water. Asterisks (*) indicate a significant difference based on a two-tailed Student's t test ($p < 0.05$). Error bars represent standard deviations of the three simulations of each system.

charges caused the amino acid to bury its side chain in the cyclohexane layer, rather than in water. The dipole on the side chain hydroxyl group is reduced compared to the GROMOS96 43A1 topology, and thus, hydrogen bonding to water was less of a driving force in the interactions in these systems.

The results from the interfacial amino acid simulations call into question the use of unrefined PRODRG topologies for small molecules. PRODRG was unable to replicate the partial charges of the GROMOS96 43A1 parameter set for the atoms of any of the amino acids, which are the most well-characterized species in most biomolecular force fields. Very little information is publicly available regarding the method by which PRODRG assigns charges to the input molecules, aside from a brief statement in the original paper stating that charges are referenced from a database of common functional groups.[18] In fact, it seems that, in the case of uncharged molecules, PRODRG assigns charges that are a hybrid of polar and nonpolar groups. That is, polar groups (Asn and Ser side chains, for example) are insufficiently polar, while normally uncharged groups (Ala and Ile side chains, methylene groups in Lys) are assigned small partial charges, rendering them more hydrophilic. Some other polar groups, such as backbone amides in all the amino acids, have much larger dipoles assigned by PRODRG than would be expected under the GROMOS96 43A1 parameter set. In the case of charged groups, such as the carboxylate group of the Asp side chain and the amino group of the Lys side chain, the dipole on the carboxylate group is too large, while that of the amino group is too small.

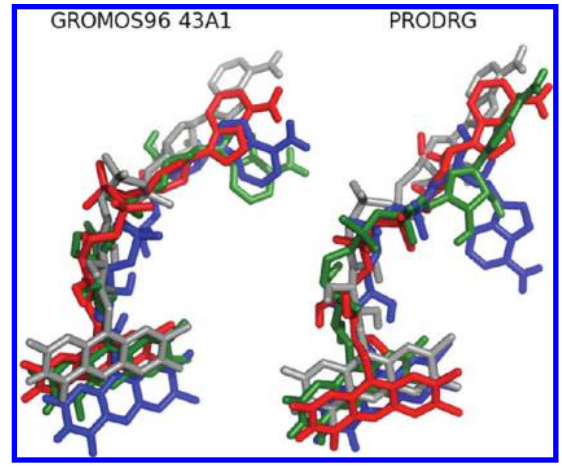

Figure 12. Alignment of FAD structures, with the reference position (the energy-minimized configuration of FAD) for each simulation shown in gray. The final snapshot of each 40-ns simulation is shown and is colored by simulation. Overlays were produced by fitting the structure of the protein to remove global translation and rotation and extracting the resulting FAD coordinates. The images were generated with PyMOL.[56]

UDP-Galactopyranose Mutase. The PRODRG server was designed to produce ligand topologies for use in modeling protein−ligand structures.[18,19] Thus, it is important to assess its applicability to this purpose. To design a test system for which ligand parameters are already present in the GROMOS96 43A1 force field library, we chose to model the UGM enzyme in complex with its FAD cofactor. Modeling other small molecules in protein binding sites would be an interesting exercise, but the lack of a clear method for generating their topologies has limited the focus of this study.

Over the course of triplicate 40-ns trajectories under both GROMOS96 43A1 parameters and those generated by PRODRG, the FAD molecule remained bound to the protein (which, in all cases, was assigned parameters under the GROMOS96 43A1 force field), with its position largely consistent over time. That is, there was no global translation of the cofactor out of its binding site, or rotation therein. However, we did observe a number of important differences with respect to the configuration of FAD and its interactions with the UGM enzyme. Structural overlays, taken from the final snapshot of each trajectory, are shown in Figure 12. While the position of the isoalloxazine ring was largely consistent across the simulations under both GROMOS96 43A1 and PRODRG, the location of the adenine ring was more variable in the simulations utilizing PRODRG parameters.

The configurations generated in these simulations would seem to indicate that the PRODRG parameters assigned to FAD are sufficiently accurate for simulation. However, a conclusion regarding accuracy cannot be made based on structures alone. The nature of the interactions between FAD and UGM is an important criterion for assessing the accuracy of these parameters and is also critically important in simulations of small molecules docked to receptors. MD simulations can be used to assess the stability of docking results and can be used to calculate the energetics of the resulting interactions over time.

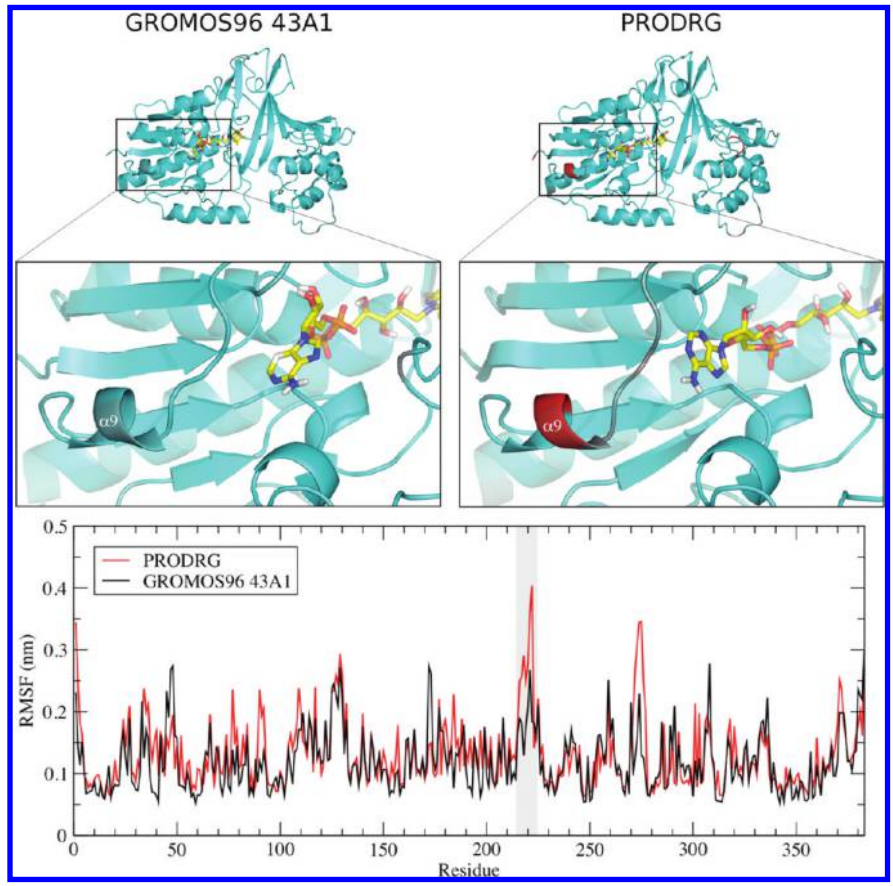

Figure 13. RMSF of the UGM backbone. Only the charges on the FAD molecule differ between the two models. The protein (cartoon) is colored according to the backbone RMSF, using a color gradient from blue to red, with blue regions indicating less fluctuation and red regions indicating more fluctuation. The FAD cofactor is shown as sticks and colored by element. Structures are taken from representative snapshots of each simulation set (GROMOS96 43A1 and PRODRG FAD parameters). The images were generated with PyMOL.[57] The RMSF of residues referred to in the main text is indicated in the shaded box on the RMSF plot.

We found that the parameters assigned to FAD by PRODRG led the cofactor to participate in significantly fewer hydrogen bonds to UGM. In the original crystal structure, eight hydrogen bonds between FAD and UGM were detected, using a cutoff angle of 30° and a cutoff radius of 3.5 Å. Under GROMOS96 43A1, 10.3 ± 0.4 hydrogen bonds were formed (averaged over time and over three replicate simulations). Using PRODRG parameters resulted in the formation of just 5.0 ± 0.9 hydrogen bonds. Several of these missing hydrogen bonds can likely be attributed to the assignment of zero charge to the atoms of the C^6-amino group of the adenine ring (Figure 5). The absence of hydrogen bonding between this moiety and UGM also explains the variability of the position of the adenine ring across the three simulations.

MD simulations are also commonly used to predict binding energies between small molecules and their target proteins.[44] By measuring nonbonded interaction energies between FAD and UGM, the nature of the interactions between the protein and its cofactor can be assessed. When GROMOS96 43A1 charges were assigned to FAD, the average total interaction energy between UGM and FAD was −866 ± 52 kJ mol^{-1}. When PRODRG-derived charges were assigned to FAD, the average interaction energy was −588 ± 30 kJ mol^{-1}. While the Lennard-Jones energies were similar in both simulation sets, the Coulombic energies differed drastically. With GROMOS96 43A1 charges, the magnitude of the intermolecular Coulombic interactions was, on average, −427 ± 36 kJ mol^{-1}, but with PRODRG charges, this term was −158 ± 22 kJ mol^{-1}. Thus, despite the apparent similarity in the positions of FAD within the UGM enzyme, the energetics of this interaction were remarkably different, particularly with respect to the Coulombic interactions.

It is apparent that the inconsistent charge and charge group assignments generated by PRODRG had an adverse effect on the dynamics and behavior of the FAD cofactor in its binding site. However, it is even more of a concern that the PRODRG-assigned FAD charges might influence the dynamics of the protein itself. A measure of the root-mean-square fluctuation (RMSF) in UGM indicated that helix α9 and the loop following it (residues 215−224) significantly increased in fluctuation ($p < 0.05$) in one simulation in which the FAD molecule was assigned PRODRG charges (Figure 13). The average RMSF over the last 20 ns of simulation using GROMOS96 43A1 parameters for FAD was 0.18 ± 0.04 nm, while that of the simulation using PRODRG parameters was 0.26 ± 0.08. This loop is important in binding the adenine moiety of the FAD, and we attribute the increased fluctuation to a lack of hydrogen bonding between UGM and FAD in this region. The two other trajectories using PRODRG charges for FAD resulted in reasonable configurations for the cofactor, similar to those obtained by assigning charges from GROMOS96 43A1. Nonetheless, we have demonstrated that the energetics of these interactions deviate substantially from the expected results using GROMOS96 43A1. The implication that

PRODRG topology charges extend beyond the small molecule for which they were originally intended raises concerns about the validity of any simulations that have used unrefined PRODRG topologies.

Strategies for Charge Assignment in GROMOS Topologies. The empirical charge assignment applied in the GROMOS96 force field parameter sets presents one of the principal challenges in deriving small-molecule topologies. There is no complete description in the literature regarding the proper derivation of charges for molecules within the GROMOS96 parameter sets, aside from empirical fitting. It is for this reason that automated methods of charge calculation and assignment are so attractive. Having concluded that the charges assigned by PRODRG did not reproduce the expected behavior for even the most well-defined molecules and common functional groups, we offer strategies for initial charge assignment that are relatively fast and easily applied to GROMOS topologies. For application to small molecules in MD simulations, these initial charges should be evaluated and refined as necessary to reproduce proper condensed-phase behavior using the methodology described by the authors of the GROMOS96 parameter sets[10−14] and perhaps the strategies employed here. The charge-derivation strategies we discuss here should be considered a starting point, not an ending point, for this process.

Quantum mechanical calculations are not explicitly stated as a source of GROMOS96 charges, although they present one method for deriving these parameters. We utilized the Antechamber program within AmberTools[45] (version 1.0) to calculate charges for all-atom versions of several of the small molecules studied here by applying several different semiempirical charge calculation methods. The results of these calculations are summarized in Table 2. Overall, the charge calculation method that best reproduced the GROMOS96 43A1 partial charges is AM1-BCC.[46,47] Charges on O and H atoms of hydroxyl groups were in good agreement with expected force field values for ethanol, *p*-cresol, and serine. For primary alcohol functional groups in these molecules, if the AM1-BCC partial charges of the C and H atoms in the CH_2 group are combined, the resulting charge was similar to the force field value of +0.150 *e*, +0.1677 *e* in the case of ethanol and +0.1919 *e* in the case of serine.

Aromatic C and H atoms had charges of comparable magnitude as well, a result that was also produced by Mulliken charges.[48] The Mulliken charges for amide groups were in closest agreement with the expected values in the GROMOS96 43A1 parameter set. The AM1-BCC method tended to overpolarize amide functional groups, but in all cases the direction of the dipole was correct. That is, none of the N atoms were assigned partial positive charges, as was the case in all of the PRODRG amino acid topologies. All of the methods utilized here also had a tendency to place small partial charges on nonpolar groups, an inconsistency with the GROMOS96 43A1 force field shared by PRODRG. The Gasteiger method[49] produced charges that were in poorest agreement with expected GROMOS96 43A1 parameters.

Antechamber is freely available within the AmberTools distribution, but it is limited to semiempirical calculations. Spartan '04 (Wave function, Inc., Irvine, CA) is a commercially available quantum mechanical program that can perform some higher level of theory calculations, and we sought to determine what level of theory would be necessary to best reproduce GROMOS96 43A1 charges. We employed three methods: the semiempirical AM1 method,[50] the ab initio Hartree−Fock (HF) method[51] with the 6-31G** basis set, and a common ab initio density functional theory (DFT) hybrid method (B3LYP),[52] also with the 6-31G** basis set. Charges were derived using the CHELP algorithm[53] within Spartan, which is a type of electrostatic charge (ESP) method suitable for describing intermolecular interactions.

The charges calculated by the Spartan methods (Table 3) were, in general, all in good agreement with those charges found in the GROMOS96 43A1 force field. In the case of ethanol, the AM1 calculations yielded a near-neutral charge on the CH_3 group (−0.003 *e*), whereas the HF- and DFT-derived charges were somewhat more negative (−0.075 *e* and −0.057 *e*, respectively). The net charge in the CH_2 group calculated by the AM1 method (+0.240 *e*) was again closest

Table 2. Charges Calculated by Antechamber Semiempirical Methods

			partial charges (*e*)			
molecule	group	atom	AM1-BCC	Mulliken	Gasteiger	GROMOS96
ethanol	CH_3	C	−0.0971	−0.2150	−0.041839	
		H1	+0.0380	+0.0773	+0.025375	
		H2	+0.0481	+0.0874	+0.025375	
		H3	+0.0482	+0.0875	+0.025375	
		sum	+0.0372	+0.0372	+0.034286	0.000
	CH_2	C	+0.1312	−0.0192	+0.040211	
		H1	+0.0183	+0.0576	+0.056075	
		H2	+0.0182	+0.0575	+0.056075	
		sum	+0.1677	+0.0959	+0.152361	+0.150
	OH	O	−0.6024	−0.3296	−0.396675	−0.548
		H	+0.3976	+0.1996	+0.210027	+0.398
p-cresol	CH_3	C	−0.0487	−0.1739	−0.039771	
		H1	+0.0399	+0.0792	+0.027965	
		H2	+0.0437	+0.0830	+0.027965	
		H3	+0.0437	+0.0830	+0.027965	
		sum	+0.0786	+0.0713	+0.044214	0.000
	C	C	−0.1123	−0.1050	−0.050945	0.000
	CH[a]	C	−0.1381	−0.1381	−0.044274	−0.100
		H	+0.1371	+0.1371	+0.063842	+0.100
	COH	C	+0.1187	+0.0736	+0.071255	+0.150
		O	−0.4987	−0.2526	−0.360943	−0.548
		H	+0.4179	+0.2169	+0.218239	+0.398
serine	acetyl CH_3	C	−0.1766	−0.2445	+0.012705	
		H1	+0.0829	+0.1222	+0.032806	
		H2	+0.0783	+0.1176	+0.032806	
		H3	+0.0674	+0.1067	+0.032806	
		sum	+0.0520	+0.1029	+0.111123	0.000
	N-terminal amide	C	+0.6657	+0.3096	+0.209208	+0.380
		O	−0.6022	−0.3631	−0.278059	−0.380
		N	−0.5854	−0.3945	−0.301933	−0.280
		H	+0.3153	+0.2288	+0.150258	+0.280
	CαH	C	+0.0447	+0.0180	+0.122426	
		H	+0.1033	+0.1426	+0.062991	
		sum	+0.1480	+0.1606	+0.185417	0.000
	$C\beta H_2$	C	+0.1238	−0.0266	+0.070315	
		H1	+0.0283	+0.0676	+0.058897	
		H2	+0.0398	+0.0791	+0.058897	
		sum	+0.1919	+0.1201	+0.188109	+0.150
	OH	O	−0.5855	−0.3127	−0.393908	−0.548
		H	+0.4101	+0.2091	+0.210159	+0.398
	C-terminal amide	C	+0.6454	+0.2893	+0.234251	+0.380
		O	−0.6069	−0.3678	−0.275489	−0.380
		N	−0.5679	−0.3770	−0.316830	−0.280
		H	+0.3207	+0.2342	+0.149048	+0.280
	C-terminal N−CH_3	C	+0.0873	−0.0680	−0.000395	
		H1	+0.0491	+0.0884	+0.043014	
		H2	+0.0269	+0.0662	+0.043014	
		H3	+0.0357	+0.0750	+0.043014	
		sum	+0.1990	+0.1616	+0.128647	0.000

[a] Averaged over all equivalent groups in the ring.

Table 3. Charges Calculated by Methods Available in Spartan '04

			partial charges (*e*)			
molecule	group	atom	semiempirical AM1	Hartree−Fock 6-31G**	DFT (B3LYP) 6-31G**	GROMOS96
ethanol	CH_3	C	−0.437	−0.502	−0.504	
		H1	+0.154	+0.158	+0.162	
		H2	+0.154	+0.156	+0.161	
		H3	+0.126	+0.113	+0.124	
		sum	−0.003	−0.075	−0.057	0.000
	CH_2	C	+0.135	+0.344	+0.219	
		H1	+0.052	−0.001	+0.022	
		H2	+0.053	−0.002	+0.022	
		sum	+0.240	+0.341	+0.263	+0.150
	OH	O	−0.518	−0.668	−0.572	−0.548
		H	+0.318	+0.402	+0.367	+0.398
p-cresol	CH_3	C	−0.547	−0.758	−0.778	
		H1	+0.163	+0.205	+0.208	
		H2	+0.166	+0.211	+0.217	
		H3	+0.167	+0.213	+0.217	
		sum	−0.051	−0.129	−0.136	0.000
	C	C	+0.075	+0.307	+0.304	0.000
	CH[a]	C	−0.211	−0.324	−0.277	−0.100
		H	+0.144	+0.211	+0.178	+0.100
	COH	C	+0.417	+0.479	+0.386	+0.150
		O	−0.524	−0.648	−0.566	−0.548
		H	+0.354	+0.442	+0.412	+0.398
serine	acetyl CH_3	C	−0.547	−0.623	−0.600	
		H1	+0.181	+0.205	+0.173	
		H2	+0.162	+0.165	+0.199	
		H3	+0.163	+0.178	+0.150	
		sum	−0.041	−0.075	−0.078	0.000
	N-terminal amide	C	+0.716	+0.875	+0.718	+0.380
		O	−0.556	−0.660	−0.562	−0.380
		N	−0.554	−0.852	−0.723	−0.280
		H	+0.291	+0.364	+0.331	+0.280
	CαH	C	−0.065	+0.542	+0.500	
		H	+0.139	+0.028	+0.034	
		sum	+0.074	+0.570	+0.534	0.000
	CβH_2	C	+0.042	−0.092	−0.161	
		H1	+0.078	+0.094	+0.102	
		H2	+0.088	+0.097	+0.100	
		sum	+0.208	+0.099	+0.041	+0.150
	OH	O	−0.537	−0.689	−0.607	−0.548
		H	+0.349	+0.453	+0.422	+0.398
	C-terminal amide	C	+0.637	+0.598	+0.438	+0.380
		O	−0.553	−0.646	−0.544	−0.380
		N	−0.443	−0.490	−0.360	−0.280
		H	+0.299	+0.353	+0.320	+0.280
	C-terminal N−CH_3	C	−0.323	−0.410	−0.447	
		H1	+0.157	+0.171	+0.194	
		H2	+0.130	+0.194	+0.151	
		H3	+0.146	+0.146	+0.173	
		sum	+0.110	+0.101	+0.071	0.000

[a] Averaged over all equivalent groups in the ring.

to the GROMOS96 43A1 parameters (+0.150 *e*), and the HF- and DFT-derived charges were slightly greater (+0.341 *e* and +0.263 *e*, respectively). The magnitudes of the charges on the oxygen and hydrogen of the hydroxyl groups of all Spartan methods were in close agreement with the GROMOS96 43A1 charges, with the closest approximation coming from the DFT calculations. The important result is that the dipoles derived by these methods were greater in the hydroxyl group than in the methyl group, likely resulting in a correct hydrogen-bonding behavior, which was not observed in the case of PRODRG-derived charges.

In the case of *p*-cresol, the *p*-methyl group was assigned a net negative charge by all methods, which was offset by the net positive charge on the carbon in the 4-position. The four equivalent carbons (at the 2-, 3-, 5-, and 6-positions) were all slightly overpolarized with the closest approximation to the GROMOS96 43A1 parameters coming from the AM1 calculations. The dipole on the hydroxyl group, however, was similar to the magnitude of this same group in the GROMOS96 43A1 parameter set in all cases, with the closest approximation derived by the DFT method. It appears that the magnitude of the dipole on the hydroxyl group compared to that of the *p*-methyl group was large enough such that a reasonable pattern of hydrogen-bonding behavior would result.

In the case of serine, many of the same patterns are observed. The net charges on the N- and C-terminal methyl groups were in close agreement with the accepted GROMOS96 43A1 charges, whereas both the N- and C-terminal amide atoms were calculated to have the correct sign, but

with a slight overpolarization by all methods. The largest disparity among these methods was the charge calculated for the α-carbon. The AM1 method predicted a small negative charge (−0.074 *e*), while the HF and DFT methods both calculated large net positive charges (+0.570 *e* and +0.534 *e*, respectively). The dipole on the hydroxyl group, however, was in close agreement with GROMOS96 43A1 parameters when calculated by all methods, with the closest being the AM1 method.

It is imperative to note that simple calculation of charges by any of the methods here would be insufficient for parametrizing new functional groups that are compatible with the GROMOS96 force fields. Topologies still must be validated using thermodynamic integration and an assessment of other condensed-phase criteria.[14] We do believe, however, that the application of any of the aforementioned methods, especially the semiempirical AM1 methods discussed here, provides a reasonable starting point for deriving charges for novel functional groups, taking much of the guesswork out of the empirical derivation strategy. Assignment of known functional group charges to small molecules is appropriate in the case of GROMOS96 parameter sets, as they have been designed to contain transferable functional groups that have the same parameters (with respect to partial charges), regardless of the remainder of the molecule.

CONCLUSIONS

We have demonstrated the effects of incorrect small-molecule topologies in a variety of simulated systems. We conclude that the automated PRODRG 2.5 server is a valuable tool for the preparation of small-molecule topologies for molecular simulation. It is fast, versatile, and easy to use. Bonded parameters and atom types assigned by PRODRG are correctly assigned in all cases. Unfortunately, the resulting topologies often suffer from deficiencies in the charges and charge groups that are assigned. These parameters often deviate substantially from the functional groups defined in the GROMOS96 43A1 parameter set, with the same functional groups in different molecules having different charges, leading to incorrect behavior in a variety of well-characterized systems. In general, we find that hydrophobic groups parametrized by PRODRG are inappropriately assigned partial charges, making them overly hydrophilic. As a consequence, the amino acid alanine inappropriately diffused into the aqueous solvent of our biphasic water−cyclohexane systems and the aliphatic portion of the lysine side chain experienced overly attractive van der Waals interactions with water. Dipoles in polar uncharged and charged species were somewhat unpredictable; as in the case of asparagine and aspartate, the side chain dipole was too large, resulting in increased hydrogen bonding, while for lysine and serine, the dipole of the relevant polar group was too small, decreasing the hydrogen-bonding capacity of these molecules. Application of small partial charges to hydrophobic moieties of *p*-cresol and ethanol led to substantially reduced hydrogen-bonding capacity and altered liquid structural properties.

Perhaps the most troubling result of all comes from the simulations of the UGM−FAD complex. The PRODRG topology for FAD appeared to give rise to reasonable configurations for FAD, despite some increased flexibility in the adenine moiety. However, the nature and strength of the interactions between FAD and UGM were significantly different between the GROMOS96 43A1-generated FAD topology and the topology from the PRODRG server. Hydrogen bonds were significantly reduced, due in large part to a completely uncharged amine group, and the value of nonbonded interaction energies was drastically different. Further, the negative effects of the PRODRG charges and charge groups assignments extended to the dynamics of UGM itself, conferring greater flexibility to nearby residues relative to what should be expected with GROMOS96 43A1 parameters applied to FAD. Thus, the applicability of unrefined PRODRG topologies to high-throughput screening methods is questionable, at best.

The validity of any program that purports to produce small-molecule topologies deserves some scrutiny, and the resulting topologies should always be rigorously validated before using these parameters in any simulation study. We caution all investigators to properly calculate partial charges for the molecules in their simulations, first by analogy to similar chemical entities already in the force field library and then by semiempirical quantum mechanical calculations for any group not present in the force field library, and to always validate their small-molecule topologies in accordance with the prescribed force field methodology. Topologies generated by PRODRG are increasingly used in diverse applications in the literature, but the results obtained may not be consistent with the parent force field, GROMOS96 43A1. That is, the results obtained may suffer from many of the limitations and inaccuracies we have reported here. This fact raises significant concerns about the widespread use of this automated tool in the absence of any further topology modification or validation. The original intent for the PRODRG server was to be used in conjunction with crystal structure refinement and energy minimization, and extrapolation to more complicated systems may not be appropriate in the absence of substantial topology refinement.

The group concept employed in the GROMOS force fields allows for versatile transferability of chemical moieties between molecules, that is, parameters of known functional groups can be applied to different molecules, independent of other structural differences between these species. Thus, it is reasonable to construct a topology for an arbitrary small molecule under the GROMOS force fields in a way that is consistent with the original force field derivation, with minimal effort, by piecing together existing building blocks. For functional groups that are not yet defined, fast charge calculation methods like AM1-BCC, for which extensive geometry optimization is not required, should suffice as a starting point for further topology refinement, but we emphasize that partial charges for any new chemical functional groups should be iteratively refined as necessary and rigorously validated before use in production simulations.

ACKNOWLEDGMENT

The authors thank Michelle Oppenheimer for useful discussions regarding UGM, Prof. J. M. Tanko for allowing us the use of Spartan '04, and the Virginia Tech Terascale Computing Facility for computing time on the SystemX supercomputer. The material is based upon work supported by the Macromolecular Interfaces with Life Sciences (MILES) Integrative Graduate Education and Research Traineeship

(IGERT) of the National Science Foundation under Agreement No. DGE-0333378 and by the Institute for Critical Technology and Applied Science (ICTAS) at Virginia Tech.

Supporting Information Available: Results of the Antechamber and Spartan '04 charge calculations for the five remaining amino acids. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz Jr, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(2) Wang, J.; Cieplak, P.; Kollman, P. A. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules. *J. Comput. Chem.* **2000**, *21*, 1049–1074.

(3) Sorin, E. J.; Pande, V. S. Exploring the Helix−Coil Transition via All-Atom Equilibrium Ensemble Simulations. *Biophys. J.* **2005**, *88*, 2472–2493.

(4) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

(5) Neria, E.; Fischer, S.; Karplus, M. Simulation of Activation Free Energies in Molecular Systems. *J. Chem. Phys.* **1996**, *105*, 1902.

(6) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(7) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J. Comput. Chem.* **2004**, *25*, 1400–1415.

(8) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(9) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.

(10) Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. The GROMOS Biomolecular Simulation Program Package. *J. Phys. Chem. A* **1999**, *103*, 3596–3607.

(11) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberer, P. H.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G., *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; Verlag de Fachvereine: Zürich, Switzerland, 1996.

(12) Daura, X.; Mark, A. E.; van Gunsteren, W. F. Parametrization of Aliphatic CH_n United Atoms of GROMOS96 Force Field. *J. Comput. Chem.* **1997**, *19*, 535–547.

(13) Schuler, L. D.; Daura, X.; van Gunsteren, W. F. An Improved GROMOS96 Force Field for Aliphatic Hydrocarbons in the Condensed Phase. *J. Comput. Chem.* **2001**, *22*, 1205–1218.

(14) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. 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.* **2004**, *25*, 1656–1676.

(15) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(16) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell, A. D., Jr. CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2010**, *31*, 671–690.

(17) Kandt, C.; Ash, W. L.; Tieleman, D. P. Setting Up and Running Molecular Dynamics Simulations of Membrane Proteins. *Methods* **2007**, *41*, 475–488.

(18) van Aalten, D. M. F.; Bywater, R.; Findlay, J. B. C.; Hendlich, M.; Hooft, R. W. W. PRODRG, a Program for Generating Molecular Topologies and Unique Molecular Descriptors from Coordinates of Small Molecules. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 255–265.

(19) Schüttelkopf, A. W.; van Aalten, D. M. F. PRODRG: a Tool for High-Throughput Crystallography of Protein-Ligand Complexes. *Acta Crystallogr., Sect D: Biol. Crystallogr.* **2004**, *D60*, 1355–1363.

(20) Yang, C.; Zhu, X.; Li, J.; Shi, R. Exploration of the Mechanism for LPFFD Inhibiting the Formation of B-Sheet Conformation of Aβ(1−42) in Water. *J. Mol. Model.* **2010**, *16*, 813–821.

(21) Li, J.; Zhu, X.; Yang, C.; Shi, R. Characterization of the Binding of Angiotensin II Receptor Blockers to Human Serum Albumin Using Docking and Molecular Dynamics Simulation. *J. Mol. Model.* **2010**, *16*, 789–798.

(22) Sharma, M.; Khanna, S.; Bulusu, G.; Mitra, A. Comparative Modeling of Thioredoxin Glutathione Reductase from *Schistosoma mansoni*: A Multifunctional Target for Antischistosomal Therapy. *J. Mol. Graphics Model.* **2009**, *27*, 665–675.

(23) Levin, L. B.-A.; Ganoth, A.; Amram, S.; Nachliel, E.; Gutman, M.; Tsfadia, Y. Insight into the interaction sites between fatty acid binding proteins and their ligands. *J. Mol. Model.* **2010**, *16*, 929–938.

(24) Boggara, M. B.; Faraone, A.; Krishnamoorti, R. Effect of pH and Ibuprofen on the Phospholipid Bilayer Bending Modulus. *J. Phys. Chem. B* **2010**, *114*, 8061–8066.

(25) Cordomí, A.; Prades, J.; Frau, J.; Vögler, O.; Funari, S. S.; Perez, J. J.; Esribá, P. V.; Barceló, F. Interactions of Fatty Acids with Phosphatidylethanolamine Membranes: X-Ray Diffraction and Molecular Dynamics Studies. *J. Lipid Res.* **2010**, *51*, 1113–1124.

(26) Boggara, M. B.; Krishnamoorti, R. Partitioning of Nonsteroidal Antiinflammatory Drugs in Lipid Membranes: A Molecular Dynamics Simulation Study. *Biophys. J.* **2010**, *98*, 586–595.

(27) Liu, F.-F.; Ji, L.; Dong, X.-Y.; Sun, Y. Molecular Insight into the Inhibition Effect of Trehalose on the Nucleation and Elongation of Amyloid β-Peptide Oligomers. *J. Phys. Chem. B* **2009**, *113*, 11320–11329.

(28) Turner, D. C.; Yin, F.; Kindt, J. T.; Zhang, H. Molecular Dynamics Simulations of Glycocholate−Oleic Acid Mixed Micelle Assembly. *Langmuir* **2010**, *26*, 4687–4692.

(29) Cilpa, G.; Hyvönen, M. T.; Koivuniemi, A.; Riekkola, M.-L. Atomistic Insight into Chondroitin-6-sulfate Glycosaminoglycan Chain Through Quantum Mechanics Calculations and Molecular Dynamics Simulations. *J. Comput. Chem.* **2010**, *31*, 1670–1680.

(30) Gandhi, N. S.; Mancera, R. L. Can Current Force Fields Reproduce Ring Puckering in 2-*O*-Sulfo-α-L-iduronic Acid? A Molecular Dynamics Simulation Study. *Carbohydr. Res.* **2010**, *345*, 689–695.

(31) Sanjeewa, R.; Weerasinghe, S. Development of a Molecular Mechanics Force Field for Caffeine to Investigate the Interactions of Caffeine in Different Solvent Media. *J. Mol. Struct. THEOCHEM* **2010**, *944*, 116–123.

(32) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(33) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116–122.

(34) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N•log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(35) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(36) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(37) Nosé, S. A Unified Formulation of the Constant Temperature Molecular Dynamics Methods. *J. Chem. Phys.* **1984**, *81*, 511–519.

(38) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A: At. Mol. Opt. Phys.* **1985**, *31*, 1695–1697.

(39) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(40) Nosé, S.; Klein, M. L. Constant Pressure Molecular Dynamics for Molecular Systems. *Mol. Phys.* **1983**, *50*, 1055–1076.

(41) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction Models for Water in Relation to Protein Hydration. In *Intermolecular Forces*, Pullman, B., Ed.; Reidel: Dordrecht, 1981.

(42) Beis, K.; Srikannathasan, V.; Liu, H.; Fullerton, S. W. B.; Bamford, V. A.; Sanders, D. A. R.; Whitfield, C.; McNeil, M. R.; Naismith, J. H. Crystal Structures of *Mycobacteria tuberculosis* and *Klebsiella*

pneumoniae UDP-Galactopyranose Mutase in the Oxidised State and *Klebsiella pneumoniae* UDP-Galactopyranose Mutase in the (Active) Reduced State. *J. Mol. Biol.* **2005**, *348*, 971–982.

(43) Zhang, C.; Yang, X. Molecular Dynamics Simulation of Ethanol/Water Mixtures for Structure and Diffusion Properties. *Fluid Phase Equilib.* **2005**, *231*, 1–10.

(44) Åqvist, J.; Luzhkov, V. B.; Brandsal, B. O. Ligand Binding Affinities from MD Simulations. *Acc. Chem. Res.* **2002**, *35*, 358–365.

(45) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Jr, K. M. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.

(46) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132–146.

(47) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II Parameterization and Validation. *J. Comput. Chem.* **2002**, *23*, 1623–1641.

(48) Mulliken, R. S. A New Electroaffinity Scale; Together with Data on Valence States and on Valence Ionization Potentials and Electron Affinities. *J. Chem. Phys.* **1934**, *2*, 782–792.

(49) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity—A Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219–3228.

(50) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.

(51) Szabo, A.; Ostlund, N. S., *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover Publications: New York, 1996.

(52) Becke, A. D. The Role of Exact Change. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(53) Chirlian, L. E.; Francl, M. M. Atomic Charges Derived from Electrostatic Potentials: A Detailed Study. *J. Comput. Chem.* **1987**, *8*, 894–905.

(54) Lide, D. R. *CRC Handbook of Chemistry and Physics*, 84th ed.; CRC Press LLC: Boca Raton, FL, 2004.

(55) Richard, L. S.; Bernardes, C. E. S.; Diogo, H. P.; Leal, J. P.; Minas da Piedade, M. E. Energetics of Cresols and of Methylphenoxyl Radicals. *J. Phys. Chem. A* **2007**, *111*, 8741–8748.

(56) DeLano, W. L., *The PyMOL Molecular Graphics System*; DeLano Scientific LLC: Palo Alto, CA, 2009.

CI100335W