JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

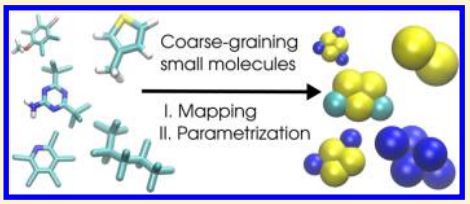
Automated Parametrization of the Coarse-Grained Martini Force Field for Small Organic Molecules

Tristan Bereau* and Kurt Kremer

Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany

Supporting Information

ABSTRACT: The systematic exploration of chemical compound space holds many promises toward structure−function relationships and material design. In the context of computer simulations, progress is hampered by both the sheer number of compounds and the efforts associated with parametrizing a force field for every new molecule. A coarse-grained (CG) representation provides not only a reduced phase space but also a smaller number of compounds, due to the redundancy of CG representations mapping to the same structure. Though many CG models require the explicit force-field parametrization of a molecule with all others, others assume transferability by means of mixing rules, such as the Martini force field. To alleviate the burden associated with tedious parametrizations for each new compound, the present work aims at automating the mapping and parametrization of common small organic molecules for Martini. We test the method by analyzing the water/octanol partitioning of more than 650 neutral molecules, the hydration free energy of 354 others, and the free energies of hydration and solvation in octanol of another 69 compounds. Last, we compare with all-atom simulations the thermodynamics of insertion of four individual solute molecules in a phospholipid membrane. The protocol demonstrates the feasibility of an automated parametrization scheme for Martini and provides prospects for high-throughput simulation methodologies.



1. INTRODUCTION

The systematic exploration of chemical space is a fascinating endeavor that holds many promises toward structure−function relationships and material design.[1−5] Progress in the field is hampered by the daunting number of molecules: the combinatorial explosion of compounds with number of atoms poses serious practical challenges, though they all exert molecular properties governed by few basic physical principles. In classical simulations, the investigation of any new compound requires the parametrization of its force field—a potentially delicate and tedious task. A number of all-atom (AA) force fields now provide the means to help automate the parametrization of new small molecules.[6−11] These protocols often assume transferability between chemically similar compounds and additionally rely on a limited number of quantum calculations to fine-tune certain aspects, for example, dihedrals and partial charges. The state-point dependence and lack of transferability of force fields, as well as the limited number of molecules already parametrized, still call for a certain amount of human intervention.

At the coarse-grained (CG) level,[12,13] the increased state-point dependence of a model further limits its transferability (see, e.g., ref 14). Still, the recent development of a number of algorithms and software solutions has provided systematic routes to parametrize CG force fields (e.g., refs15−25.). These so-called bottom-up techniques rely on structural information (e.g., radial distribution functions and forces) to optimize CG interaction potentials. The reference data almost invariably originates from AA simulations. The availability of an AA force field and the prior—possibly computationally expensive—AA simulation form a prerequisite for the further parametrization of the CG model.

Not all CG models, however, are derived from structure-based, AA information (i.e., structure-based coarse-graining). Larger-scale, phenomenological properties can also be used as target information to optimize CG interaction potentials. Among them, the Martini force field provides a consistent set of mappings—roughly four atoms per CG bead—and bead types (building-block approach) across a wide variety of (bio)molecules, for example, amino acids, lipids, and carbohydrates.[26−32] Martini's thermodynamic-based parametrization assigns a CG bead type based on the water/oil partitioning of the fragment encapsulating the atoms. Further, it relies on mixing rules to generate cross-interactions between molecules, thereby greatly reducing the force-field parametrization effort. The resulting framework can thus dispense from reference AA simulations altogether, relying instead on experimental measurements to optimize both the bonded and nonbonded interactions, as illustrated in Monticelli et al.'s amino acid parametrization.[28] The ability to parametrize a molecule without prior, computationally expensive AA simulations forms the basis of the present study.

Though a number of extensions of the Martini force field rely on different reference observables (e.g., atomistic radial distribution functions or surface tension), we follow the original protocol presented in the original lipid and protein force fields.[26−28] Enforcing consistency between force-field

Received: January 21, 2015
Published: April 28, 2015

DOI: 10.1021/acs.jctc.5b00056
J. Chem. Theory Comput. 2015, 11, 2783−2791

parametrization protocols helps (though in no way guarantees) transferability between compounds.[9]

This work introduces an automated mapping and parametrization procedure to generate a Martini force field for common small organic molecules. The scheme first optimizes the AA to CG mapping (i.e., from atoms to CG beads) in accordance with the design principles governing Martini-described compounds, that is, "mapping rules." Next, it constructs bonded interactions from the equilibrium geometry of the reference molecule. Finally, atom types are assigned based on a fragment's water/octanol partitioning. Rather than extracted directly from experimental measurements—requiring human intervention to find and filter the reference data—the use of recently developed prediction algorithms automates this process.

After presenting in detail how the mapping and force-field parametrization are performed, we test the procedure by measuring the water/octanol partitioning coefficient of 653 molecules of up to 15 heavy (i.e., non-hydrogen) atoms. We then benchmark the hydration free energy of 354 compounds. Finally, we present a comparison with AA simulations of the thermodynamics of insertion of four independent solute molecules in a phospholipid bilayer.

2. ALGORITHM

The following describes the determination of molecular features, the mapping, the parametrization of the bonded interactions, and finally the assignment of atom types for nonbonded interactions. An implementation of the entire protocol is available at https://github.com/tbereau/automartini.

2.1. Molecular Features. The identification of chemical features of a given compound is provided by a number of cheminformatic tools, used, for instance, in the context of virtual screening of drugs. The only input information about the compound that the following protocol requires is either an SDF chemical file or a SMILES string.[33] While the SDF file explicitly includes the chemical element, atomic coordinates, and connectivity of each atom, the SMILES string only includes the chemical element and connectivity of non-hydrogen atoms without any coordinates (e.g., "CCO" for ethanol). From a SMILES string, a three-dimensional structure can be automatically generated (including hydrogens), and further optimized, for example, by using the Universal Force Field.[34]

The input structure is analyzed using RDKit,[35] a collection of cheminformatics software to manipulate chemical structures. RDKit can extract information on bond topology, hydrogen-bond donor/acceptor groups, atoms involved in aromatic rings, and the location of net charges. It will be used in the following to identify the chemical features of individual fragments.

2.2. Mapping. The mapping $\mathbf{M}:\mathbf{r} \rightarrow \mathbf{R} = \mathbf{M}(\mathbf{r})$ defines the CG configuration of CG beads $\mathbf{R}$ given the set of atomic coordinates $\mathbf{r}$.[23] Practically, we aim at finding an optimal set of N CG beads that best encapsulates all n heavy (i.e., non-hydrogen) atoms within the Martini guidelines,[27] which are in a nutshell: (i) Three to four heavy atoms are mapped onto one CG bead; and (ii) Only two atoms per CG bead for ring structures (irrespective of the hybridization state). The bead diameters in Martini are σ = 4.3 Å for aromatic rings and σ = 4.7 Å otherwise (neglecting specific interactions, as described in ref 27).

For simplicity, this work will consider that a CG bead can only be located on one of the n atoms—this limits the number of possible sets of CG beads for a molecule. The present strategy consists of translating the Martini guidelines, as well as a number of intuitive prescriptions, into an energy function that quantifies the suitability of a given mapping. By evaluating the energy function for all possible sets of CG beads, one can rank the mappings and thereby identify the best candidate.

The energy function E is constructed from the interactions of CG beads, modeled as Gaussians, and atoms. The set of CG beads should encapsulate as many atoms as possible, while keeping a reasonable distance between beads. The mapping attempts to find a compromise between a minimal number of CG beads while ensuring a small distance from each atom to any CG bead. E is composed of four terms (see Figure 1):

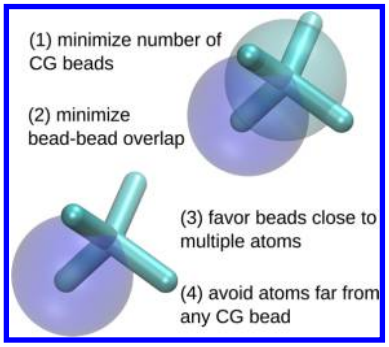


Figure 1. Rules governing the mapping energy function. The four parts illustrate the bead, bead−bead, and bead−atom terms contained in eq 1. See main text for more details. Cartoons rendered with VMD.[36]

1 Penalty for any new CG bead introduced, which limits the total number of CG beads. The penalty amounts to $w_{nr}N_{nr} + w_rN_r$, where w_{nr} and w_r denote the penalty weight coefficients for nonring and ring beads, respectively, and $N_{nr} + N_r = N$ denotes the corresponding numbers of CG beads involved.
2 Repulsive overlap between CG beads, thereby keeping them far apart. The overlap depends on the distance between two CG beads I and J, r_{IJ}, the beads' sizes, $\sigma_{IJ} = (\sigma_I + \sigma_J)/2$, and the weight w_{BB}.
3 Attractive contribution from atom i close to CG bead J. The contribution depends on the atom's mass, m_i, the atom-bead distance r_{iJ}, the size of the CG bead σ_J, and the weight w_{aB}.
4 Penalty for any atom i not located within the radius of any CG bead, with weight w_a. This term aims at penalizing mappings for which certain atoms are far from any CG bead, and thereby ensures a reasonable description of the sterics.

Apart from the bead sizes and atom masses, all weights will require optimization, which will be described below (Sec. 4.1). The objective function for the mapping amounts to

$$E(\mathbf{M}_N) = w_{nr}N_{nr} + w_rN_r + w_{BB}\sum_{I\neq J}\exp\left(-\frac{r_{IJ}^2}{4\sigma_{IJ}^2}\right) - w_{aB}\sum_i\sum_J m_i\exp\left(-\frac{r_{iJ}^2}{2\sigma_J^2}\right) + w_a\sum_i m_i\prod_J\theta(r_{iJ}-\sigma_J), \quad (1)$$

DOI: 10.1021/acs.jctc.5b00056
J. Chem. Theory Comput. 2015, 11, 2783−2791

where $\mathbf{M}_N$ is a mapping of N CG beads, $\theta(x)$ is the Heavyside step function: $\theta(x) = 1$ for $x > 0$ and 0 otherwise, and lowercase and uppercase indices refer to atomistic and CG particles, respectively. The negative sign in front of w_{aB} underlines the favorable contribution to the energy function.

Furthermore, the bead-type assignment (described below) requires a partitioning of the atoms between the CG beads, thereby forming molecular fragments. We apply a Voronoi partitioning scheme,[37,38] which identifies the set of atoms encapsulated in CG bead J, $\mathcal{S}_J$, as those being closest to bead J among all atom-bead distances:

$$\mathcal{S}_J = \{i \in \{1, ..., n\} \mid r_{iJ} \leq r_{iK} \text{ for all } K \neq J\} \quad (2)$$

Though the Voronoi scheme provides a simple method to partition the molecule into fragments, we point out a possible arbitrariness in crowded environments and its significant dependence to the input geometry. Though averaging over conformations may alleviate these issues, this was left out of this study.

We further enforce a number of additional restrictions: (i) two CG beads cannot be located on neighboring (i.e., chemically bonded) atoms; (ii) all atoms within a CG bead must be connected to at least one other atom in that bead; and (iii) a CG bead must encapsulate more than one atom.

The present scheme (i.e., energy function evaluation and assessment of the aforementioned conditions) is applied to the exhaustive set of combinations of CG beads across all atomic locations. The number of CG beads is successively increased until the best-performing mapping (i.e., with the lowest E value) stops improving: min $E(\mathbf{M}_N) \leq$ min $E(\mathbf{M}_{N+1})$. An example in Figure 2 illustrates how different combinations are evaluated and ranked to determine the best mapping.

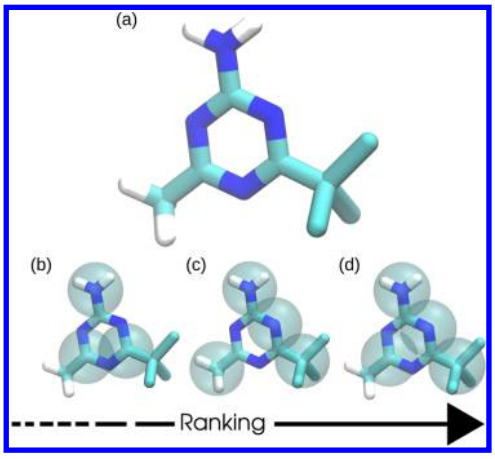


Figure 2. Example: 4-*tert*-butyl-6-methyl-1,3,5-triazinan-2-amine. Atomic structure (a), as well as two putative CG mappings (b) and (c), and the optimal CG mapping (d), given the weights (eq 1) optimized to reproduce the Martini mapping.

2.3. Bonded Interactions. The above-mentioned mapping only provides isolated CG beads that lack any connectivity. The determination of the CG topology is based on a bond recognition algorithm: any CG bead I is connected to J if one atom contained in I is chemically bonded to one atom contained in J, generating a CG bond between the two beads. This definition also applies to rings. The set of CG angles and dihedrals are constructed from the collection of bonds.

Bonded interactions are parametrized from the molecule's equilibrium geometry: the equilibrium distances, angles, and dihedrals are measured from the geometry of the (optimized) CG beads. Following Martini's prescription,[27] bonds are constrained within rings or assigned a harmonic force constant k_{bond} = 1250 kJ/mol in all other cases; angles yield k_{angle} = 25 kJ/mol, except for angles involving a double bond in the center bead, where the force constant is increased to k_{angle} = 45 kJ/mol; and dihedrals: $k_{dihedral}$ = 10 kJ/mol. Given the generic assignment of force constants to all molecules, they are by no means optimized to reproduce a given compound's distribution functions.

2.4. Nonbonded Interactions. The parametrization of the nonbonded interactions is arguably the most difficult aspect of a force-field optimization endeavor. Martini's building block approach reduces this problem to selecting an appropriate bead type among 18 candidates. They model different chemical features of the encapsulated atoms, most notably the fragment's net charge (type "Q"), its ability to donate/accept hydrogen bonds (subscript "d/a"), and finally its thermodynamic properties (polar "P", nonpolar "N", apolar "C").[27] The latter is parametrized from the partitioning free energy between polar and apolar phases, for example, water/octanol coefficient. A characterization of these three properties (i.e., charge, H-bond donor/acceptor, and partitioning free energy) entirely determines the most appropriate bead type. While the first two molecular features—charge and H-bond characteristics—are readily available from the AA molecular conformation itself (e.g., by the chemist's naked eye or suitable cheminformatic tools, here RDKit[35]), an evaluation of the free energy of partitioning requires more consideration.

A fragment's free energy of partitioning between water and octanol, ΔG_{ow}, is directly linked to its partition coefficient. The prevalence of the partition coefficient as a useful molecular descriptor has motivated a wide breadth of research, in particular computational methods to *predict* its value given a molecular structure. Here, we rely on ALOGPS, a neural-network-based prediction algorithm.[39,40] The partition coefficient of the isolated fragment, determined in Mapping Section 2.2, is predicted by an online query to the ALOGPS server. The query amounts to a SMILES string that represents the fragment.

The Martini atom type is then selected according to the following guidelines: (i) a charged fragment will be assigned an atom type Q; (ii) the H-bond analysis will determine the atom type's subscript d, a, or da for donor, acceptor, and donor/acceptor, respectively; and (iii) a noncharged fragment will be assigned the atom type P, N, or C such that its partition coefficient (measured below) best matches ALOGPS' prediction. Of all neutral beads, only type N proposes H-bond substitutes (Table 1). For such H-bond-rich fragments, we account for prediction errors in the partitioning by setting a bead to Nd, Na, or Nda if the mean-absolute error of the predicted fragment with the reference bead type (i.e., N*) is less than 0.7 kcal/mol.

The partitioning of fragments taking part in rings tends to yield more arbitrariness, due to the smaller bead/atom ratio. For those, we predict the partition coefficient of the entire collection of atoms involved in the ring (i.e., typically three beads per ring in Martini), and weigh the contribution of each by an appropriate factor (see below). The three beads may then carry different atom-types depending on the fragments' various H-bond features; they can also vary across one ring depending

DOI: 10.1021/acs.jctc.5b00056
J. Chem. Theory Comput. 2015, 11, 2783−2791

Table 1. Water/Octanol Partitioning Free Energy of Each Martini Bead Type[a]

bead type	ΔG	bead type	ΔG	bead type	ΔG
Qda	−3.6	P3	−2.1	N0	1.0
Qa	−3.6	P2	−0.9	C5	1.7
Qd	−3.6	P1	−0.5	C4	2.4
Q0	−5.4	Nda	0.6	C3	3.0
P5	−2.1	Na	0.6	C2	3.3
P4	−2.2	Nd	0.6	C1	3.4

[a]All free energies in units of kcal/mol. All error bars are within 0.1 kcal/mol.

on the connecting atoms that are part of that fragment—for example, a chlorobenzene will have three beads, one of which has a different bead type than the other two.

In the following, we test and validate the present model against the original Martini water model, rather than more recent "polarizable" models.[41,42] Since we focus on neutral molecules, we do not expect significant changes between standard water and polarizable Martini water. Likewise, the current parametrization has not been tested against the recent implicit-solvent Dry Martini.[43]

3. METHODS

All molecular dynamics simulations were performed in GROMACS 4.5.3[44] using the standard simulation parameters associated with the Martini force field.[27] We applied a 30 fs time step for all systems, except for the solute-in-membrane simulations, where we applied a 10 fs time step. A Langevin thermostat and an Andersen barostat[45] provided control over the temperature (T = 300 K) and pressure (p = 1 bar), respectively. We used coupling constants τ_T = 1.0 ps and τ_p = 12.0 ps for the respective ensembles, and a compressibility of 3 × 10^{-4} bar^{-1}. The LINCS algorithm provided integration of the stiff constraints between ring beads.[46] Any time mentioned in the following simply refers to the aggregate number of integration time steps performed during the simulation—it does not take into account any speedup due to coarse-graining—thus they do not represent realistic time scales.

Solvation free energies and water/octanol partitioning free energies were measured by simulating a given compound in three media: vacuum, water (∼350 CG water molecules were used, including ∼10% of antifreeze particles[27]), and octanol (∼250 octanol molecules). We computed free energies using thermodynamic integration (TI),[47] where solute beads decouple from the solvent, as monitored by the coupling parameter λ. Twenty-one simulations were run for 6 ns each at evenly spaced values of λ between 0 and 1, applying soft-core potentials. The relative coupling of the molecule in water and vacuum provides the hydration free energy, $\Delta G_{hydr} = \Delta G_{wat} - \Delta G_{vac}$, while octanol and water yield the partitioning free energy, $\Delta G_{ow} = \Delta G_{oct} - \Delta G_{wat}$.

The potential of mean force of the insertion of individual solute compounds in a lipid bilayer was computed using umbrella-sampling techniques.[48] Free-energy profiles were further extracted using the weighted histogram analysis method,[49−51] and error bars were evaluated using bootstrapping.[52] Following previous work, we placed two solute molecules in each simulation box to increase sampling and alleviate leaflet area asymmetry.[53−55] A total of 23 umbrellas were placed every 0.1 nm from the bilayer midplane normal to the membrane. A force constant k = 1000 kJ/mol/nm^2 was applied. Membrane patches of (36 nm^2) were generated using the INSANE building tool[56] and subsequently minimized, heated up, and equilibrated. Production runs amounted to 200 ns per umbrella.

4. PARAMETER OPTIMIZATION

Both the mapping procedure and nonbonded parametrization scheme require parameters to be optimized and determined, respectively.

4.1. Mapping. The energy function governing the mapping procedure (eq 1) contains a number of unknown parameters. These weights can be tuned to reproduce a particular mapping scheme—here, Martini—to consistently map atomic structures onto CG representations. The optimization of these coefficients is empirically determined. We rely on existing Martini models as target data, in particular the collection of 22 molecules in the "solvent" force field,[27,31] excluding glycerine trioleate due to its size (i.e., 65 heavy atoms). Free parameters are varied until the mapping procedure best reproduces the *number* of CG beads proposed in the original Martini force field. The present optimization is entirely independent of the determination of the bonded and nonbonded interactions.

A manual optimization of the parameters yielded the following coefficients: (i) penalty weight for nonring beads, w_{nr} = 50; (ii) penalty weight for ring beads, w_r = 20; (iii) bead−bead overlap penalty, w_{BB} = 9.0; (iv) atom-bead favored encapsulation, w_{aB} = 0.9; and (v) lonely atom penalty, w_a = 0.20. Figure 3 compares the number of CG beads predicted by

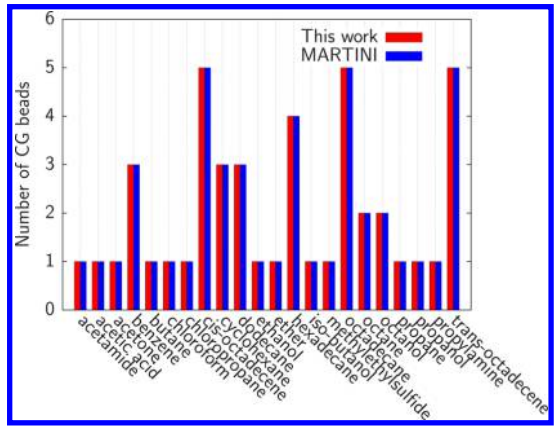


Figure 3. Comparison of the number of CG beads between this work and the original "solvent" Martini force field. The overall agreement between the two sets illustrates the consistency of the mapping procedure with Martini.

the present methodology with the original Martini force field. The protocol prescribes the original number of beads in all 22 cases. The agreement of the present mapping parametrization proves encouraging in providing a faithful depiction of a Martini-type representation.

4.2. Nonbonded Parametrization. Given the use of the partition-coefficient prediction algorithm, the assignment of a bead's water/octanol free energy of partitioning did not require any further parametrization. However, translating this free energy into a Martini bead type required the free energy of partitioning of every one of them. Though Marrink et al. do report these free energies,[27] their results included finite-concentration effects due to the use of several solute molecules in the simulation box. An alternative calculation of hydration

DOI: 10.1021/acs.jctc.5b00056
J. Chem. Theory Comput. 2015, 11, 2783−2791

free energies at infinite dilution, using TI in both vacuum and water, was presented by Yesylevskyy et al.[41] Here, we relied on an equivalent protocol to compute the water/octanol partitioning free energy for all Martini bead types. The results, presented in Table 1, are in good agreement with Marrink et al.,[27] given the spurious finite-concentration effects. Extending the calculations to hydration free energies proves also comparable to the results of Yesylevskyy et al.[41] (data not shown).

As mentioned above, the atom types of CG beads involved in a ring are determined from the *overall* partitioning coefficient, weighed appropriately among the beads involved. Though a 6-atom-to-3-bead mapping would suggest a weight 1/3 (i.e., equal weight on each bead), a comparison with the benzene and cyclohexane parametrizations provided by the Martini force field[27,31] point toward a weight 2/3 to better reproduce their thermodynamic properties. In particular, the naive weight 1/3 leads to a benzene made of SN0 bead types—yielding a polarity that is too pronounced (data not shown)—while the updated weight 2/3 predicts SC5 beads, in agreement with de Jong et al.[31] Likewise, cyclohexane is modeled by SC5 and SC2 bead types when using weights 1/3 and 2/3, respectively—the latter being in reasonable agreement with the suggested SC1 bead-type parametrization.[27] Though the parametrization of rings could benefit from a more extended investigation and parametrization protocol, we chose to keep things as such for the purpose of this study.

5. RESULTS

5.1. Molecular Water/Octanol Partitioning. We first tested the ability of the protocol to reproduce a compound's water/octanol partitioning free energy. Experimental free energies were extracted from a database.[57,58] 653 neutral compounds with up to 15 heavy (i.e., non-hydrogen) atoms were coarse-grained by the present procedure and subject to TI calculations in both water and octanol environments. The resulting free energies are compared with the reference data, as shown on Figure 4. The list of compounds and corresponding

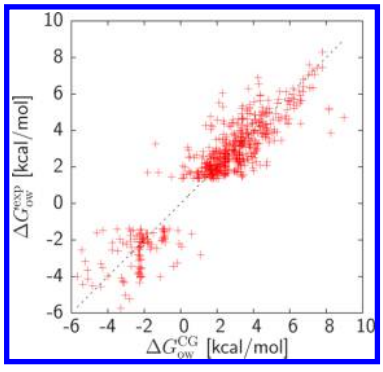


Figure 4. Water/octanol free energy of partitioning for 653 neutral compounds: comparison between CG simulations (ΔG_{ow}^{CG}) and experimental data (ΔG_{ow}^{exp}). Correlation coefficient: R^2 = 0.91; MAE: 0.79 kcal/mol. CG error bars: ≈ 0.1 kcal/mol, not shown for clarity.

experimental and CG free energies are also provided in Table S1 (Supporting Information). We observe a strong correlation between the CG simulations and the experimental data—the correlation coefficient yields R^2 = 0.91. The mean-absolute error (MAE) amounts to 0.79 kcal/mol across all compounds of varying sizes, a very reasonable performance given the presence of experimental error and the lower limit provided by the fragment free-energy prediction algorithm, roughly 0.4 kcal/mol.[39,40] We find a few isolated points at the upper-right corner (i.e., $\Delta G_{ow}^{CG} \approx 9$ kcal/mol) and around the upper-left center (i.e., $\Delta G_{ow}^{CG} \approx -1$ kcal/mol) that are made of strongly halogenated aromatic compounds. Given their presence on both sides of the correlation line, the data point to difficulties in reproducing the thermodynamics of such compounds, rather than a systematic bias.

5.2. Solvation Free Energies. We then explored the model's ability to reproduce experimental hydration free energies for small, neutral compounds. Reference data for 354 compounds were extracted from the work of Mobley et al.[59] TI calculations in vacuum and in water yielded hydration free energies, shown in Figure 5 against experimental reference

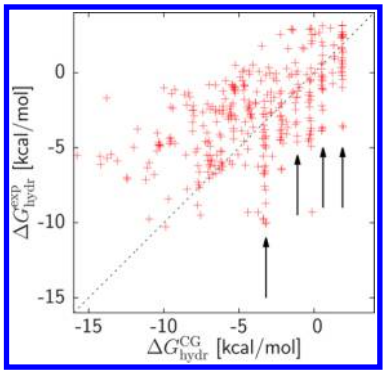


Figure 5. Hydration free energies of 354 neutral compounds:[59] comparison between CG simulations (ΔG_{hydr}^{CG}) and experimental data (ΔG_{hydr}^{exp}). Pearson correlation coefficient: R^2 = 0.57; MAE: 2.7 kcal/mol. Vertical arrows point to stripes of data points, highlighting how multiple compounds map to the same CG molecule. These stripes show slight misalignments due to the statistical uncertainty associated with performing the same simulation multiple times. CG error bars: ≈ 0.1 kcal/mol, not shown for clarity.

values. The list of compounds and corresponding experimental and CG free energies are also provided in Table S2 (Supporting Information). The results yield a correlation coefficient of only R^2 = 0.57 and a significant MAE: 2.7 kcal/mol. The results point to important discrepancies between the CG simulations and experiments. We observe two regimes: (i) At high reference values the CG free energies tend to be too high (i.e., lower-right half of the equality line). This trend was pointed out in the original description of the Martini force field[27] as well as a recent review[32] regarding the contribution of individual beads, which the authors attributed to the limited fluid range of the Lennard-Jones interaction (which also explains why antifreeze particles must be inserted in bulk water). On the other hand, (ii) reference free energies lower than −5 kcal/mol are predicted too *low* (i.e., upper-left half of the equality line). Given that the first regime is primarily represented by single-bead compounds, the second regime immediately raises the issue of additivity: the sum of the beads does not faithfully represent the hydration free energy of the compound. Otherwise, a simple extrapolation of the single-bead behavior would provide higher correlation concentrated in the lower-right panel of the equality line.

Given the lack of systematic, large-scale study of the hydration properties of Martini-based models, it is difficult to evaluate to what extent the discrepancy arises from the parametrization protocol alone, rather than from inherent

DOI: 10.1021/acs.jctc.5b00056
J. Chem. Theory Comput. 2015, 11, 2783−2791

issues associated with Martini. To a large extent, the faithful reproduction of the molecular water/octanol partitioning free energies presented in Section 5.1 arises by construction of the model. Though we do not hold experimental water/octanol free-energy data for the 354 neutral compounds, we report in Figure S1 of the Supporting Material the correlation between CG results and predictions from ALOGPS. The high correlation suggests that the CG simulations also reproduce the water/octanol partitioning of those compounds. The results clearly demonstrate that the ability to reproduce water/octanol partitioning with high fidelity does not guarantee a satisfactory reproduction of the hydration properties. The implications of Martini's Lennard-Jones interactions on its gas/fluid and solid/fluid properties have been discussed in a recent review.[32]

To better understand the qualitative difference between the accuracy of water/octanol partitioning (Figure 4) and of hydration free energies (Figure 5), we also attempted to probe the quality of free energies of solvation in octanol: A simple thermodynamic cycle links the water/octanol partitioning to the sum of a hydration free energy (i.e., gas to water) and solvation free energy in octanol (i.e., gas to octanol). Duffy and Jorgensen have gathered experimental data for both terms—free energy of hydration and solvation in octanol—for a set of small molecules.[60] Figure 6a shows the correlation for each

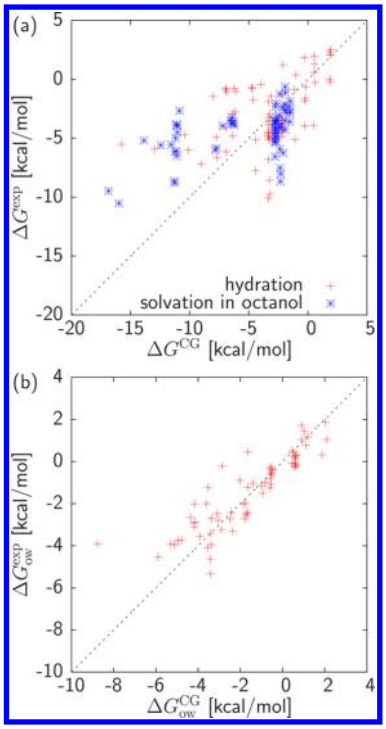


Figure 6. (a) Free energies of hydration and solvation in octanol of 69 neutral compounds:[60] comparison between CG simulations (ΔG^{CG}) and experimental data (ΔG^{exp}_{hydr}). (b) Correlation plot for water/octanol transfer free energy, computed by summing the free energy of hydration and solvation in octanol shown in (a). CG error bars ≈ 0.1 kcal/mol, not shown for clarity.

term across 69 compounds. The list of compounds and corresponding experimental and CG free energies are also provided in Table S3 (Supporting Information). We encounter very similar trends as before: mean absolute errors of 2.6 and 2.9 kcal/mol, as well as correlation coefficients of 46% and 51%, respectively. These numbers are in reasonable agreement with the above-mentioned data set of 354 neutral compounds (Figure 5a). Summing up the two contributions provides the water/octanol partitioning free energy, for which we plot a similar correlation plot in Figure 6b. The improved accuracy (mean absolute error of 0.78 kcal/mol and correlation coefficient of 87%) compared to the individual terms strongly points at the role of the gas/fluid transfer free energy.

Interestingly, the hydration results of Figure 5 emphasize how multiple compounds map to the same CG molecule, as seen from the vertical stripes in the inset: a number of molecules with different hydration free energies reduce to the same thermodynamic description at the CG level (up to statistical uncertainty of the simulation), thereby reducing chemical space.

5.3. Solute Insertion in Lipid Membrane. Jakobtorweihen et al. recently computed the potential of mean force (PMF) for the insertion of individual solute compounds in lipid bilayers from AA simulations.[55] They focused on four systems: (i) 4-ethylphenol in stearoyloleoyl phosphatidylcholine (SOPC)/water, (ii) propanol in dimethyl pyrocarbonate (DMPC)/water, (iii) 5-phenylvaleric acid in (2*R*)-2,7-dimethyloctyl](methyl)amine (DOPC)/water, and (iv) dibenz-[*a*,*h*]anthracene in 1-palmitoyl-2-oleoyl-*sn*-glycero-3 phosphocholine (POPC)/water. We used the present methodology to derive CG models of these solutes. These compounds were then embedded in a 128-lipid bilayer to assess the performance of the parametrization by comparing the PMFs with the reference AA data. The insertion was monitored by means of the normal coordinate to the bilayer midplane, *z*. Note that the Martini force field does not provide (as of yet) a parametrization for SOPC lipids. Given the observed similarities between insertion profiles in SOPC and DOPC,[55] we use the latter as replacement. Finally, small (i.e., up to 0.3 nm) deviations in membrane thickness between the CG and AA curves were compensated by shifting the profiles horizontally.

Figure 7 compares the CG and AA PMF curves for the insertion of the four solute compounds in their respective lipid bilayer. The parametrization protocol maps the first compound to three ring-forming beads: SC3−SNda−SC5, where the SC3 and the SNda fragments encapsulate the ethyl and hydroxyl groups, respectively. We get excellent agreement for the PMF (Figure 7a) throughout the entire range, except close to the bilayer midplane, where the compound shows slightly too hydrophobic. In the range of $z \leq 0.5$ nm, the CG model fails to describe the sharp increase, that is, the CG representation shows too hydrophobic around the bilayer midplane. This issue arises from the limited number of bead types: the coarse fragmentation of water/octanol partitioning free energies among bead types (Table 1) limits the thermodynamic resolution of the CG compound. We note, however, that the global minimum around the interfacial lipid region is faithfully reproduced.

Further, the second compound shows excellent agreement in the entire range of interest—the CG curve falls within the AA profile (Figure 7b). This may result from the comparatively smaller size of the compound, which was mapped to a single CG bead type Nda, thereby avoiding additivity issues upon combining fragments.

Next, the third compound, 5-phenylvaleric acid, was mapped to five rings: P1−C5−SC4−SC5−SC5. It showed an overall faithful CG PMF (Figure 7c), though the slope at the lipid/water interface is somewhat too large, resulting in slightly exaggerated features at the interface region—*a* ≈ 1 kcal/mol

DOI: 10.1021/acs.jctc.5b00056
J. Chem. Theory Comput. 2015, 11, 2783−2791

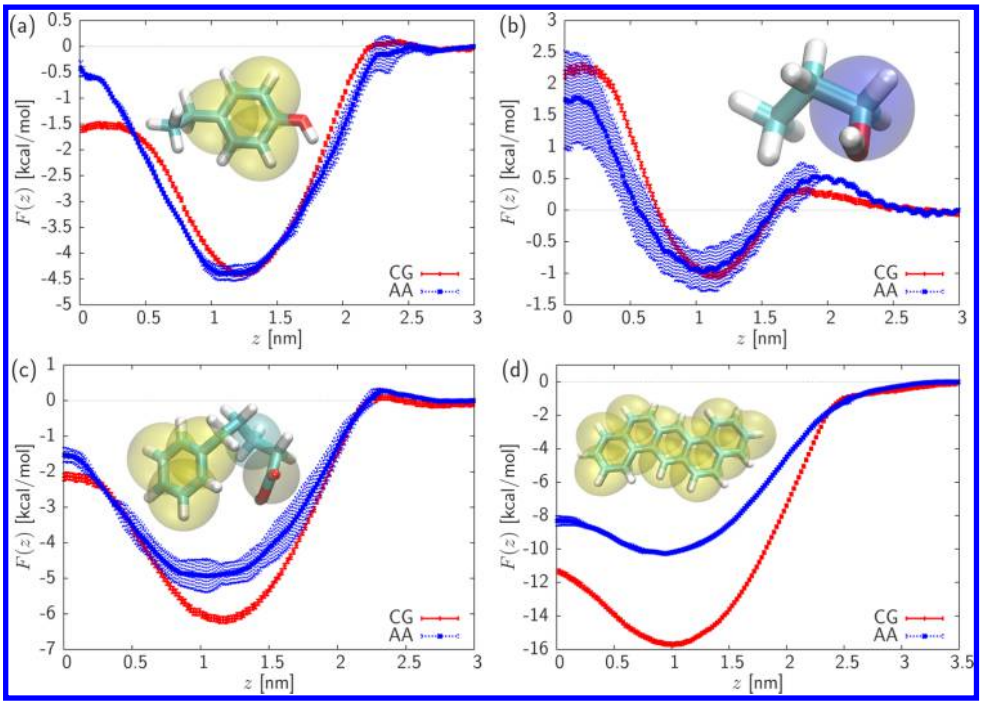


Figure 7. Potential of mean force curves of the insertion of solute molecules in a lipid membrane from CG (red, solid lines) and AA (blue, dashed lines) simulations: (a) 4-ethylphenol in SOPC/water (DOPC/water in the CG simulations); (b) propanol in DMPC/water; (c) 5-phenylvaleric acid in DOPC/water; and (d) dibenz[*a*,*h*]anthracene in POPC/water. Atomistic data from Jakobtorweihen et al.[55] Cartoon representations depict both the atomistic structure and CG mapping. Color-coding of the CG bead types: yellow (S*), cyan (C*), dark blue (N*), and amber (P*). The CG bead sizes are not to scale.

discrepancy. Otherwise, the profile between the bilayer midplane (z = 0) and the interfacial region ($z \approx$ 1.5 nm) is correctly reproduced. Overall, the CG model reliably reproduces the AA PMF curve semiquantitatively.

Finally, the parametrization of dibenz[*a*,*h*]anthracene amounted to 11 SC*beads—most of them SC5 (like benzene in the Martini force field), though two of them were assigned SC4 types, due to slight differences in the partitioning of atoms between fragments. Although the protocol, in line with the Martini guidelines, assigns rigid constraints between all ring-type beads, we found that the network of constraints made for unstable simulations—we thus replaced all constraints by harmonic bonds and weakened the dihedral force constants to k_{dihedral} = 2 kJ/mol. We emphasize that this modification both preserves the overall geometry of the compound and is the only occasion throughout this work that a force field was modified after its automatic generation. Figure 7d features an exaggerated free energy of transfer between bulk water and the interfacial lipid region. The collection of ringlike beads enhances the effect to yield a 3 kcal/mol offset at z = 0. We note, here again, that the overall shape of the profile is reproduced. These results point to a discrepancy of the transfer free energy of benzene-like beads between water and the interfacial region of the membrane. Though the present parametrization (i.e., collection of SC5 beads) supplies a fragment that is too hydrophobic, we note that the closest bead that exhibits lower hydrophobicity, namely, SN0, produces strongly polar thermodynamics, in *qualitative* disagreement with the reference data of the compound. Though Figure 7b displays the thermodynamics of an Nda bead, the overall shape of the PMF is similar. Akin to the first compound, we thus rationalize the present discrepancy by a lack of resolution in the choice of bead types between C5 and N0. This effect is here compounded by the sheer number of beads.

6. DISCUSSION

This protocol makes a number of assumptions upon segmenting a molecule into fragments. First, the method assumes *linearity* of the partitioning free energies, that is, the molecular free energy of partitioning is the sum of the individual bead contributions. To the extent of the system sizes studied here (i.e., up to ∼20 non-hydrogen atoms), this assumption seems to hold, as illustrated in Figure 4. We expect this assumption to break down for larger molecules—the threshold size remains unknown.

Evidently, the assignment of a partitioning free energy from the content of an *isolated* fragment ignores the chemistry of the neighboring atoms and, most importantly, the associated connectivity. One would expect the sheer error made on the boundary atoms to distort the partition-coefficient assignment. Interestingly, the resolution of the CG model does not allow to resolve such level of detail. Case in point, butane and octane in the standard Martini force field are both built using the same bead type, C1—one and two such beads, respectively—though clearly octane is not the concatenation of two butane molecules. The resolution of the CG model thus makes a fragment-based prediction of the beads at all possible.

The prediction algorithm used here to assign bead types, ALOGPS, is a neural network based on molecular fragments. In turn, the algorithm is not capable of predicting any arbitrary fragment but rather the most commonly found. In case ALOGPS fails to predict the partitioning of a fragment, we propose to fall back on an atom-based method, such as that of Wildman−Crippen.[61] Though the method is not as accurate (the error is roughly twice as large as for ALOGPS[39,40,61]), it is

DOI: 10.1021/acs.jctc.5b00056
J. Chem. Theory Comput. 2015, 11, 2783−2791

implemented as an optional fall-back option in the script. We note, however, that in light of homogeneity, all of the results presented in this work relied solely on ALOGPS.

None of the tests and applications presented here involved charged compounds. Compared to other bead types, the assignment of charged fragments is straightforward: RDKit detects whether the fragment of interest bears a net charge. If so, it also looks for possible acceptor/donor sites in the fragment to assign the bead type to either Q0, Qa, Qd, or Qda. Preliminary tests on charged compounds have shown promise but certainly call for an investigation of its own.

The ability to capture conformational and structural properties were left out of this study but would likely benefit from a number of improvements. First, the Voronoi partitioning scheme, which assigns atoms to CG beads, entails a strong dependence to the input structure. Averaging over a representative set of conformations may well diminish this bias. Further, the parametrization of the bonded interactions is currently modeled by generic force-constant assignments commonly applied in the Martini force field. A more careful parametrization that better captures the underlying probability distributions would evidently benefit the method. Such refinements would likely show increasingly useful for larger molecules.

In terms of computational investment, the time required to parametrize one compound depends strongly on its number of heavy (non-hydrogen) atoms. While the force field of a molecule up to ∼15 heavy atoms takes less than one second on a desktop computer, larger molecules quickly prove challenging. For example, the 22-atom dibenz[*a*,*h*]anthracene (Figure 7) required 25 Gb of memory to enumerate all possible mappings (up to the point that extra beads did not further improve the energy function, eq 1) and required ∼1 h of CPU time. Computational improvements for the enumeration scheme can be achieved and will be presented separately. Furthermore, the parametrization of larger molecules are expected to be made possible by other optimization algorithms, such as a Monte Carlo search through mapping space. For the purpose of the present work, however, we deem enumeration an appropriate tool for a proof-of-principle study.

7. CONCLUSIONS

In this work, we introduce a fully automated protocol for the generation of CG Martini force fields applicable to a wide variety of small, organic molecules. The protocol firsts maps atoms into a (reduced) number of CG beads optimized by means of a heuristic energy function that mimics the essential principles behind the Martini force field. The exhaustive enumeration of mappings limits the compound size to roughly 20−25 non-hydrogen atoms on current desktop-computer architectures. Bonded interactions are further assigned from standard interaction forms and strengths based on the all-atom (AA) input geometry and commonly used Martini parameters. Lastly, the protocol assigns CG bead types that define the nonbonded interactions, obtained from partition-coefficient predictions.

The method was shown capable of reproducing the molecular water/octanol partitioning free energy of more than 650 neutral compounds with an MAE value of only 0.8 kcal/mol—slightly more than thermal energy at room temperature. This shows satisfying given the errors inherent to experimental measurements and the neural-network-based prediction of partition coefficient for each fragment, as well as the inherent additivity assumption. The description of strongly halogenated aromatic compounds shows most challenging in terms of reproducing the underlying thermodynamics.

The hydration free energies of a collection of 354 neutral molecules and free energies of hydration and solvation in octanol for 69 other compounds confirmed the discrepancy arising from single-bead molecules, owing at least partially to the limited fluid range of the Lennard-Jones potential.[27,32] Results on a wide variety of larger molecules point to difficulties in reproducing the free energies of hydration and solvation in octanol—systematically too low—and an important scatter in the correlation plot (Figure 5). These findings ultimately suggest a deficiency in the additivity of CG beads regarding solvation properties and caution regarding the use of the model for gas/fluid properties.

Lastly, we compared the thermodynamics of small molecules inserted in a lipid membrane with AA reference results. We found strong agreement between the two resolutions, which demonstrates that the CG systems can reproduce the characteristic thermodynamic features of the original system. The limited number of bead types can cap the overall accuracy of the CG model, as illustrated from the PMF of dibenz[*a*,*h*]-anthracene (Figure 7d), where a slight misfit of the bead type gets compounded due to the many aromatic rings involved.

Though tailored to Martini, the method can be extended to other CG models: the mapping procedure is easily transferable and can encompass different coarse-graining resolutions by varying the free parameters of eq 1. As for the nonbonded parametrizations, the current methodology leverages the ability to *predict* partition coefficients based on molecular fragments—a protocol that could be extended to other thermodynamic properties, for example, density and interfacial tension.[62]

With ∼1000 small molecules parametrized in this work, the method provides the means to study a large number of compounds at the CG level. It offers a number of features, such as the possibility to evaluate the quality of the CG model over *ensembles* of molecules (e.g., hydration properties) and the rapid, systematic—and, most importantly, automatic—parametrization of additional compounds. Though not designed to compete with cheminformatic tools that accurately predict partitioning coefficients of molecules in simple environments, the method provides the means to estimate thermodynamic properties in more complex environments, for example, a solute in a membrane with a mixture of lipids. The most intriguing aspect of a CG representation to study small molecules is the redundancy of mappings across several compounds, in such a way that coarse-graining effectively reduces chemical compound space. The systematic evaluation of molecular properties across a large number of molecules is thereby facilitated, leading the way toward structure−function relationships.

■ ASSOCIATED CONTENT

Supporting Information

List of chemical compounds and associated experimental and coarse-grained free energies of water/octanol partitioning and hydration. This material is available free of charge via the Internet at http://pubs.acs.org/.The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00056.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: bereau@mpip-mainz.mpg.de.

DOI: 10.1021/acs.jctc.5b00056
J. Chem. Theory Comput. 2015, 11, 2783−2791

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank M. Jochum for fruitful discussions and advice, R. Potestio and S.-J. Marrink for a critical reading of the manuscript, and S. Jakobtorweihen for providing the atomistic free-energy profiles of the solute-in-membrane systems.

■ REFERENCES

(1) Fried, L. E.; Manaa, M. R.; Pagoria, P. F.; Simpson, R. L. *Annu. Rev. Mater. Res.* **2001**, *31*, 291−321.
(2) Yaghi, O. M.; O'Keeffe, M.; Ockwig, N. W.; Chae, H. K.; Eddaoudi, M.; Kim, J. *Nature* **2003**, *423*, 705−714.
(3) Dobson, C. M. *Nature* **2004**, *432*, 824−828.
(4) Lipinski, C.; Hopkins, A. *Nature* **2004**, *432*, 855−861.
(5) Farha, O. K.; Hupp, J. T. *Acc. Chem. Res.* **2010**, *43*, 1166−1175.
(6) Wang, J.; Kollman, P. A. *J. Comput. Chem.* **2001**, *22*, 1219−1228.
(7) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graphics Modell.* **2006**, *25*, 247−260.
(8) Ribeiro, A. A.; Horta, B. A.; Alencastro, R. B. d. *J. Braz. Chem. Soc.* **2008**, *19*, 1433−1435.
(9) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr *J. Comput. Chem.* **2010**, *31*, 671−690.
(10) Vanommeslaeghe, K.; MacKerell, A. D., Jr *J. Chem. Inf. Model.* **2012**, *52*, 3144−3154.
(11) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D., Jr *J. Chem. Inf. Model* **2012**, *52*, 3155−3168.
(12) Voth, G. A. *Coarse-Graining of Condensed Phase and Biomolecular Systems;* CRC Press: Boca Raton, FL, 2008.
(13) Noid, W. *J. Chem. Phys.* **2013**, *139*, 090901.
(14) Mullinax, J.; Noid, W. *J. Chem. Phys.* **2009**, *131*, 104110.
(15) Lyubartsev, A. P.; Laaksonen, A. *Phys. Rev. E* **1995**, *52*, 3730.
(16) Tschöp, W.; Kremer, K.; Batoulis, J.; Bürger, T.; Hahn, O. *Acta Polym.* **1998**, *49*, 61−74.
(17) Tschöp, W.; Kremer, K.; Hahn, O.; Batoulis, J.; Bürger, T. *Acta Polym.* **1998**, *49*, 75−79.
(18) Eilhard, J.; Zirkel, A.; Tschöp, W.; Hahn, O.; Kremer, K.; Schärpf, O.; Richter, D.; Buchenau, U. *J. Chem. Phys.* **1999**, *110*, 1819−1830.
(19) Reith, D.; Meyer, H.; Müller-Plathe, F. *Macromolecules* **2001**, *34*, 2335−2345.
(20) Reith, D.; Meyer, H.; Müller-Plathe, F. *Comput. Phys. Commun.* **2002**, *148*, 299−313.
(21) Reith, D.; Pütz, M.; Müller-Plathe, F. *J. Comput. Chem.* **2003**, *24*, 1624−1636.
(22) Faller, R. *Polymer* **2004**, *45*, 3869−3876.
(23) Ayton, G. S.; Noid, W. G.; Voth, G. A. *Curr. Opin. Struct. Biol.* **2007**, *17*, 192−198.
(24) Rühle, V.; Junghans, C.; Lukyanov, A.; Kremer, K.; Andrienko, D. *J. Chem. Theory Comput.* **2009**, *5*, 3211−3223.
(25) Jochum, M.; Andrienko, D.; Kremer, K.; Peter, C. *J. Chem. Phys.* **2012**, *137*, 064102.
(26) Marrink, S. J.; de Vries, A. H.; Mark, A. E. *J. Phys. Chem. B* **2004**, *108*, 750−760.
(27) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812−7824.
(28) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. *J. Chem. Theory Comput.* **2008**, *4*, 819−834.
(29) Marrink, S. J.; de Vries, A. H.; Harroun, T. A.; Katsaras, J.; Wassall, S. R. *J. Am. Chem. Soc.* **2008**, *130*, 10−11.
(30) López, C. A.; Rzepiela, A. J.; De Vries, A. H.; Dijkhuizen, L.; Hunenberger, P. H.; Marrink, S. J. *J. Chem. Theory Comput.* **2009**, *5*, 3195−3210.
(31) de Jong, D. H.; Singh, G.; Bennett, W. D.; Arnarez, C.; Wassenaar, T. A.; Schafer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. *J. Chem. Theory Comput.* **2012**, *9*, 687−697.
(32) Marrink, S. J.; Tieleman, D. P. *Chem. Soc. Rev.* **2013**, *42*, 6801−6822.
(33) Weininger, D. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31−36.
(34) Rappé, A. K.; Casewit, C. J.; Colwell, K.; Goddard Iii, W.; Skiff, W. *J. Am. Chem. Soc.* **1992**, *114*, 10024−10035.
(35) *RDKit*: Open-source cheminformatics, version09.1. 2014. Online: http://www.rdkit.org (accessed: July 1, 2014).
(36) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.
(37) Okabe, A., Boots, B., Sugihara, K., Chiu, S. N. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams;* Wiley: Hoboken, NJ, 2009; Vol. *501*.
(38) Bereau, T.; von Lilienfeld, O. A. *J. Chem. Phys.* **2014**, *141*, 034101.
(39) Tetko, I. V.; Tanchuk, V. Y.; Villa, A. E. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1407−1421.
(40) Tetko, I. V.; Tanchuk, V. Y. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1136−1145.
(41) Yesylevskyy, S. O.; Schäfer, L. V.; Sengupta, D.; Marrink, S. J. *PLoS Comput. Biol.* **2010**, *6*, e1000810.
(42) Wu, Z.; Cui, Q.; Yethiraj, A. *J. Phys. Chem. B* **2010**, *114*, 10524−10529.
(43) Arnarez, C.; Uusitalo, J. J.; Masman, M. F.; Ingólfsson, H. I.; de Jong, D. H.; Melo, M. N.; Periole, X.; De Vries, A. H.; Marrink, S. J. *J. Chem. Theory Comput.* **2014**, DOI: 10.1021/ct500477k.
(44) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. *Bioinformatics* **2013**, btt055.
(45) Andersen, H. C. *J. Chem. Phys.* **1980**, *72*, 2384−2393.
(46) Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(47) Van Gunsteren, W.; Berendsen, H. *J. Comput.-Aided Mol. Des.* **1987**, *1*, 171−176.
(48) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187−199.
(49) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(50) Bereau, T.; Swendsen, R. H. *J. Comput. Phys.* **2009**, *228*, 6119−6129.
(51) Hub, J. S.; De Groot, B. L.; Van Der Spoel, D. *J. Chem. Theory Comput.* **2010**, *6*, 3713−3720.
(52) Mooney, C. Z., Duval, R. D. *Bootstrapping: A Nonparametric Approach to Statistical Inference;* Sage:Los Angeles, 1993.
(53) MacCallum, J. L.; Bennett, W.; Tieleman, D. P. *Biophys. J.* **2008**, *94*, 3393−3404.
(54) Bereau, T.; Wang, Z.-J.; Deserno, M. *J. Chem. Phys.* **2014**, *140*, 115101.
(55) Jakobtorweihen, S.; Zuniga, A. C.; Ingram, T.; Gerlach, T.; Keil, F.; Smirnova, I. *J. Chem. Phys.* **2014**, *141*, 045102.
(56) Ingólfsson, H. I.; Melo, M. N.; van Eerden, F. J.; Arnarez, C.; López, C. A.; Wassenaar, T. A.; Periole, X.; De Vries, A. H.; Tieleman, D. P.; Marrink, S. J. *J. Am. Chem. Soc.* **2014**, *136*, 14554−14559.
(57) Sangster, J. *J. Phys. Chem. Ref. Data* **1989**, *18*, 1111−1229.
(58) Sangster, J. *LOGKOW Databank;* Sangster Research Laboratories: Montreal, Canada, 1993
(59) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. *J. Chem. Theory Comput.* **2009**, *5*, 350−358.
(60) Duffy, E. M.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2000**, *122*, 2878−2888.
(61) Wildman, S. A.; Crippen, G. M. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 868−873.
(62) Shinoda, W.; Devane, R.; Klein, M. *Mol. Simul.* **2007**, *33*, 27−36.

DOI: 10.1021/acs.jctc.5b00056
J. Chem. Theory Comput. 2015, 11, 2783−2791