Article

Automated Coarse-Grained Mapping Algorithm for the Martini Force Field and Benchmarks for Membrane−Water Partitioning

Thomas D. Potter, Elin L. Barrett, and Mark A. Miller*

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: With a view to high-throughput simulations, we present an automated system for mapping and parameterizing organic molecules for use with the coarse-grained Martini force field. The method scales to larger molecules and a broader chemical space than existing schemes. The core of the mapping process is a graph-based analysis of the molecule's bonding network, which has the advantages of being fast, general, and preserving symmetry. The parameterization process pays special attention to coarse-grained beads in aromatic rings. It also includes a method for building efficient and stable frameworks of constraints for molecules with structural rigidity. The performance of the method is tested on a diverse set of 87 neutral organic molecules and the ability of the resulting models to capture octanol−water and membrane−water partition coefficients. In the latter case, we introduce an adaptive method for extracting partition coefficients from free-energy profiles to take into account the interfacial region of the membrane. We also use the models to probe the response of membrane−water partitioning to the cholesterol content of the membrane.
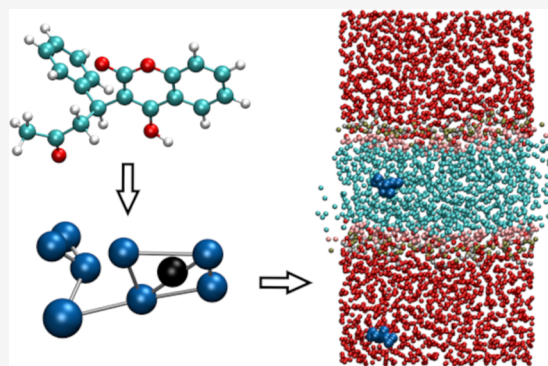
1. INTRODUCTION

Computational screening is important in a variety of fields, from drug discovery[1] to toxicology.[2] The aim of screening is to narrow down a large chemical search space, thereby guiding time-consuming experimental testing toward formulations that are likely to have the desired behavior. Computational methods must generally be capable of high throughput to be viable for screening.

In the field of toxicity and environmental impact, one of the important physical quantities in screening is the membrane−water partition coefficient or its logarithm $\log K_{MW}$, which is a measure of the extent to which a molecule accumulates in biological tissues.[3−8] Computational methods for predicting values of $\log K_{MW}$ vary widely in their sophistication and accuracy. Most simply, an approximate linear relationship between $\log K_{MW}$ and its counterpart for partitioning between *n*-octanol and water, $\log K_{OW}$, can be used and gives acceptable results for many molecules.[9] Multiparameter linear free-energy relationships, based on the correlation of partitioning free energies with various molecular descriptors, have also been used to predict partitioning of solutes between different phases, including between lipid bilayers and water.[10−13] The COSMOmic method, based on the COnductor-like Screening MOdel for Real Solvents (COSMO-RS),[14] explicitly includes information about solute and membrane structures and compositions and so approximates how they interact with each other.[15] However, in more complex cases, particularly the partitioning of ionic molecules, the inherent approximations in

these methods can break down and more accurate methods are needed.[9]

Molecular dynamics simulations include a detailed description of the structure of the system, which is allowed to evolve under the equations of motion rather than relying on assumptions about how the different components affect each other. Molecular dynamics simulations have been used extensively to study the interactions of molecules with membranes,[16] including the calculation of $\log K_{MW}$ directly from simulation trajectories.[17] However, atomistic molecular dynamics simulations are computationally expensive, even when enhanced sampling techniques are used to accelerate the exploration of configurational space.[18,19] Their use in high-throughput screening is therefore unfeasible.

A common strategy to improve the computational efficiency of simulations of biological and soft matter systems, including lipid bilayers, is the use of coarse-grained (CG) models. Coarse-graining involves removing degrees of freedom from an atomistic model by combining groups of atoms into single interaction sites. In such models, force calculations are cheaper because of the reduced number of sites. The reduction in the
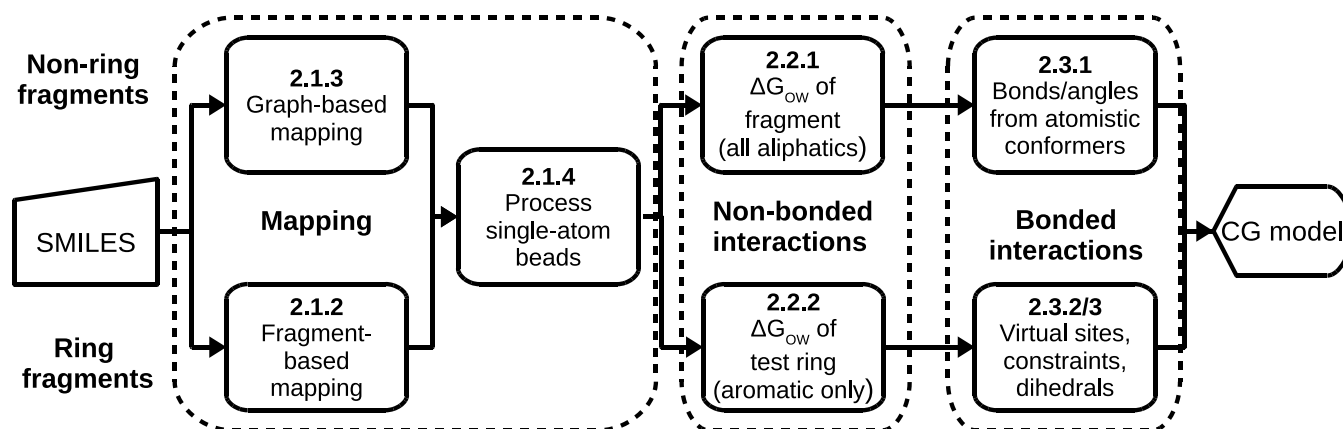
Figure 1. Flowchart outlining the three main stages of the algorithm: mapping of atoms onto beads and parameterization of nonbonded and bonded interactions. In the mapping and bonded interaction stages, the procedure differs for ring and nonring fragments, while for nonbonded interactions, the steps for aliphatic and aromatic fragments are distinct.

number of degrees of freedom also results in a smoother potential energy surface, allowing for longer simulation time steps and therefore faster exploration of configurational space than for a fully atomistic model.

There are two distinct stages to constructing a coarse-grained model: mapping and parameterization. Mapping is the division of the molecule into interaction sites, commonly known as beads. The mapping scheme sets the resolution of a coarse-grained model and thereby determines the speed-up that can be gained. There are many possible mapping schemes for any given molecule, even at a given level of bead resolution, and they vary in how well they capture properties of symmetry and structure.[20,21] Mapping is often done manually but some methods have been developed to automate the process.[22,23]

Methods for the parameterization step are diverse in approach, and the models they produce have distinct strengths and weaknesses. Bottom-up methods fit the parameters of a coarse-grained model to a more detailed atomistic simulation and include structure-based methods,[24,25] force matching,[26,27] and relative entropy.[28] It is possible to achieve an excellent representation of the atomistic system even with a much simpler model. However, some bottom-up methods suffer from poor transferability to conditions that are not represented in the atomistic reference and can be time-consuming to parameterize.[29−32] Nevertheless, coarse-grained models involving more complex interaction types, such as local-density or volume-dependent potentials, have been shown to significantly improve transferability, often with only a small increase in computational cost.[33−36] The transferability of simple pair-potential-based models can also be improved with appropriate parameterization schemes.[37,38] There has also been progress in the use of machine learning methods to parameterize bottom-up coarse-grained force fields, including the automation of coarse-grained mapping.[39−41] While this field is in its early stages, machine learning has the potential to further advance the efficiency and transferability of bottom-up coarse-graining.

Top-down models, in contrast, are fitted to macroscopic experimental thermodynamic data. Such models do not always reproduce fine structural details[32] but are well suited to calculating thermodynamic properties and phase behavior.[42−44] A number of top-down coarse-grained force fields have been developed.[45−48] The most commonly used, particularly in biomolecular simulation, is Martini.[49,50] Version 2 of Martini has been applied to a great variety of systems,

including lipid bilayers,[51] proteins,[52] and polymers.[53,54] Martini 2 consists of a number of predefined coarse-grained bead types, each of which is parameterized to represent specific types of chemical groups according to properties such as polarity and hydrophilicity.[49] Each bead is intended to model around four heavy (nonhydrogen) atoms. The interactions between the beads are defined according to an interaction matrix, with interaction strengths parameterized to match partition coefficients of small molecules between organic solvents. Bonded parameters in Martini are often parameterized to match the underlying atomistic structure and can be generated in a bottom-up fashion from an all-atom simulation.[55] The building-block approach of Martini gives the force field good chemical transferability at a given temperature. This feature is beneficial for high-throughput simulations, as it removes the need to fully reparameterize interactions for each system, as is the case in many other approaches.

The use of Martini in high-throughput screening requires a reliable, and preferably automated, way to generate models of organic solutes. The automartini method of Bereau and Kremer[22] provides a way of automatically mapping and parameterizing small organic molecules within Martini. Automartini models for a set of 653 neutral organic molecules with up to 15 heavy atoms reproduced experimental octanol–water partitioning free energies[22] with a correlation coefficient of 0.91 and a mean absolute error of 3.3 kJ mol^{-1}. The method has also been used to carry out high-throughput studies of membrane interactions for extensive libraries of molecules[56] and to investigate how well coarse-grained models cover chemical space.[57] However, automartini's methods for mapping and assignment of bonded interactions do have limitations. The mapping algorithm becomes impractically slow for molecules with more than about 25 heavy atoms due to the cost of enumerating all possible mapping schemes. The mappings are also not guaranteed to preserve the symmetry of the molecule. Additionally, models generated for extended ring compounds can require manual fine-tuning of the bonded parameters before they are suitable for use in simulations.[22]

In this paper, we present a method for generating coarse-grained Martini models that cover a wider chemical space than other methods. Our mapping approach is scalable to much larger molecules than previously possible, including those with complex ring structures. The models are consistently stable in molecular dynamics algorithms without reduction of the time
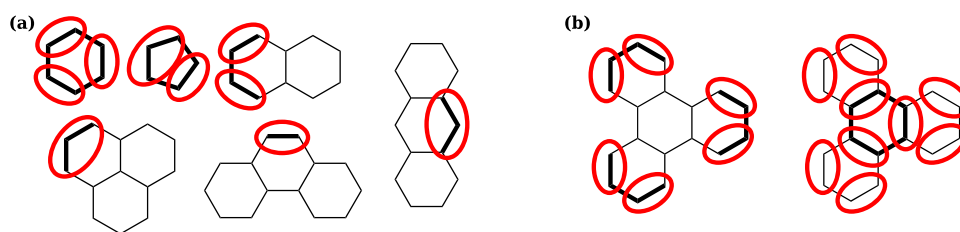
Figure 2. (a) Predefined mappings for rings with examples where these fragments occur. The fragments involved are in bold, with red ellipses indicating the coarse-grained beads. (b) The iterative mapping process for a triphenylene ring fragment.
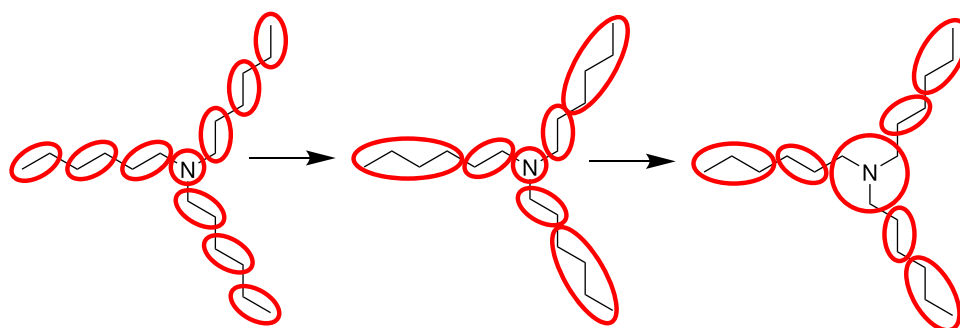

Figure 3. Mapping scheme for trihexylamine, including two iterations of graph-based grouping followed by postprocessing to remove a single-atom bead.

step. This includes ring structures for which models with poorly chosen constraints would be unstable even at low time steps due to the practical numerical difficulty of simultaneously satisfying multiple, interdependent constraints. Our method also guarantees that the symmetry of the molecule is preserved during the coarse-grained mapping. We show that the resulting models reproduce experimental octanol–water and membrane–water partitioning data with good accuracy and devise an adaptive procedure for extracting the latter from the free-energy profile of a given solute. Collectively, these advances open up the use of Martini for high-throughput screening of thermodynamic properties.

The remainder of this article is organized as follows. Section 2 describes our automated scheme for mapping and parameterizing molecules for use with the Martini 2 force field. In Section 3, we test models from our scheme by using them to calculate octanol–water and membrane–water partition coefficients. Section 4 contains a discussion of the efficiency, strengths, and current limitations of our scheme. Finally, Section 5 provides a summary of the main advances presented in this article.

2. METHODOLOGY

Our automated scheme for producing a coarse-grained model from a simplified molecular input line entry system (SMILES) code is summarized schematically in Figure 1. The three main stages of mapping and parameterization of nonbonded and bonded interactions are described in Sections 2.1–2.3 in turn. Each step in the procedure is labeled with the corresponding subsection number in the flowchart. A self-contained implementation of the full procedure is available to download with the Supporting Information. The code makes use of the open-source RDKit package,[58] version 2020.09.1.

2.1. Mapping Scheme. *2.1.1. General Approach.* Our three priorities for an automated mapping algorithm are that (1) the method must generate mapping schemes compatible with the Martini framework, (2) the mappings must respect

the symmetry of the molecule as much as possible, and (3) the algorithm must readily scale to large molecules.

Our approach, which meets all of these criteria, consists of three steps. First, we isolate ring systems within the molecule and map them using a fragment-matching algorithm. Next, we apply a version of the graph-based method of Webb et al.[23] to systematically map nonring fragments of the molecule. Finally, we apply a postprocessing step to deal with any individual atoms that could not be mapped in the previous two steps. Each of these stages of the algorithm is described in the rest of this section.

2.1.2. Ring Structures. Within the Martini framework, ring compounds are typically modeled by small (S) beads, which represent 2–3 heavy atoms, joined together in a rigid structure by a combination of LINCS constraints and virtual sites.[49,52,59] The structure of such representations makes simulations more prone to instabilities at longer time steps because of the difficulty in satisfying a large number of constraints simultaneously. Hence, it is harder to determine a robust and practical mapping scheme for rigid molecules than for flexible molecules.

Our automated framework for coarse-grained mapping of ring compounds is designed to deal with fragments containing combinations of fused five- and six-membered rings and is illustrated in Figure 2. There are two steps: grouping of the outer parts of the ring system according to predefined patterns, followed by mapping of the remaining inner fragments using graph-based grouping.

The outer parts are mapped by finding continuous chains of atoms that are part of only one ring and mapping them, as shown in Figure 2. These mapped fragments are removed, and the process is repeated for the remaining fragments. If none of the unmapped fragments forms a ring, then the remainder of the ring system is mapped by applying one iteration of the graph-based mapping system described in Section 2.1.3 but with a maximum path length of two bonds.

2.1.3. Graph-Based Grouping. The graph-based mapping method developed by Webb et al. uses only information about the connectivity of the atoms.[23] It is based on representing a molecular structure as a graph in which atoms are the nodes and bonds are the edges. The nodes are ranked and then combined according to some metric. This ranking may be done in a number of ways, and we have adapted the spectral ordering method.[23] In this approach, the whole molecule is analyzed at once, with symmetrically equivalent fragments treated simultaneously, ensuring that the symmetry of the molecule is automatically respected. The whole mapping process for an example nonring compound is shown in Figure 3, including the postprocessing step described in Section 2.1.4.

In spectral mapping, the molecule is represented by an adjacency matrix $\mathbf{A}$. A_{ij} is 1 if nodes i and j are bonded and 0 otherwise. The diagonal terms are used to weight the nodes. The weighting of node i in our scheme is given by

$$W_i = l_i \bar{m}_i \tag{1}$$

where l_i is the maximum number of bonds between any two heavy atoms in the bead and $\bar{m}_i$ is the mean atomic mass of all of the nonhydrogen atoms within the bead. Using l_i in the definition of weight favors the formation of beads with branched internal structures compared to long, linear beads, while the inclusion of $\bar{m}_i$ is a tie-breaker for chemically distinct beads with the same branching structure. Other weighting functions are possible, but this simple product reliably generates Martini-compatible mappings even for highly branched structures. The nodes are then ranked according to their centrality scores,[23] which are obtained by diagonalization of

$$\mathbf{A} = \mathbf{PDP}^{-1} \tag{2}$$

where $\mathbf{P}$ contains the eigenvectors of $\mathbf{A}$, and $\mathbf{D}$ is a diagonal matrix of its eigenvalues. The centrality scores of the nodes in the graph are the components of the eigenvector corresponding to the largest eigenvalue. Starting from the lowest-ranked node, the nodes are grouped with their neighbors. For each node i, all neighbors with equal or higher rank are identified and, of those, the one with the closest ranking is grouped with node i. If multiple nodes of the same rank simultaneously qualify to be combined with the same neighbor, all of these nodes are grouped together. At this stage, all ring fragments generated by the procedure in Section 2.1.2 are excluded. If there is no suitable neighbor, node i remains as its own group.

We have introduced an additional constraint on bead size for compatibility with Martini. For each candidate node i, the maximum path length, l_i, between any two heavy atoms within the bead is measured. If l_i is greater than three bonds, the combination step is rejected and the nodes making up the candidate node remain as their own groups. This ensures that no group grows too large to model using Martini while still allowing smaller nodes to combine in other parts of the molecule.

This procedure can be carried out iteratively until the desired level of coarse-graining is achieved. In practice, we iterate until no new beads are formed.

2.1.4. Postprocessing. In some cases, the mapping scheme results in single atoms that cannot combine with others without exceeding the size limit. There is no appropriate Martini representation for a single heavy atom, so a postprocessing step is needed to take care of these cases. The nodes are ranked by centrality score as before, and the

highest-ranked neighbors of any single-atom nodes are identified, excluding any ring beads. From each of these neighbors, one heavy atom is transferred to the bead of the single atom. If there are multiple neighbors with the same rank, then a heavy atom is removed from each, ensuring that the symmetry of the molecule is retained. If this results in new single-atom beads, the procedure is repeated until all beads have l between 1 and 3. An example of a molecule for which this process is necessary is given in Figure 3.

In some cases, this process can fail to terminate because single atoms are passed back and forth between beads. Two constraints are placed on the procedure to prevent this. First, nodes containing atoms that were either single-atom nodes or bonded to a past single-atom node in previous iterations, are tagged. If a tagged node is the highest-ranked neighbor of a single-atom node, then it is passed over and an atom is taken from the next highest neighbor. Second, in cases where the single-atom node is bonded to a terminal node with path length $l = 1$, the single-atom node and the terminal node are combined to form a new bead. This prevents the formation of terminal single-atom nodes that have no way to combine with any other atoms.

Finally, if the only appropriate neighbor for a single-atom node is a ring node, the two nodes are combined, effectively absorbing the lone atom into the mapping for the ring. In cases where there is a choice between two ring beads, the one with the highest centrality score is chosen.

2.2. Nonbonded Interactions. *2.2.1. Aliphatic Fragments.* Parameterization of nonbonded interactions in the Martini framework consists of selecting from the list of predefined bead types that make up the force field. For parameterizing the subset of beads that represent nonaromatic fragments, we have adopted the bead-selection approach from the automartini method of Bereau and Kremer, which is based on matching to octanol−water partitioning free energies (ΔG_{OW}).[22]

In this procedure, once a mapping scheme has been decided, ΔG_{OW} is predicted for each of the neutral fragments. The prediction is preferentially done using ALOGPS, a neural network trained to predict octanol−water partitioning; it has been shown to produce high-quality results for a diverse range of organic molecules.[60] In cases where ALOGPS is not applicable (e.g., for wholly inorganic fragments such as NO_2), the Wildman−Crippen approach is used; this method determines ΔG_{OW} from contributions by the atoms that make up the fragment.[61] The Martini bead type with the closest ΔG_{OW} to the ALOGPS or Wildmann−Crippen value is chosen to represent that fragment.

In our implementation of this part of the algorithm, we have obtained ΔG_{OW} from ALOGPS for all organic fragments that fit the bead size limits described in Section 2.1.3 and included these values in a data file along with the code (see the Supporting Information (SI)). This self-contained implementation eliminates dependence on the web server hosting the neural network.

Some bead parameterizations also make use of hydrogen-bonding properties. Neutral beads with a predicted ΔG_{OW} within ± 1.0 kJ mol^{-1} of the Nda (nonpolar acceptor and donor) bead type can be assigned as either Na, Nd, or Nda beads if they include hydrogen-bond acceptors, donors, or both. The hydrogen-bonding properties of the atoms are determined using the RDKit library's chemical feature module. Otherwise, the bead type is selected based on ΔG_{OW}, as

described above. Bead types for charged fragments are chosen purely based on hydrogen-bonding properties. Fragments can be assigned as Qa, Qd, or Qda as described above, or as Q0 if no hydrogen-bonding atoms are present.

2.2.2. Aromatic Fragments. Fragments that are part of nonaromatic rings are parameterized in the same way as nonring fragments. For aromatic rings, we have developed a different approach, designed to take account of heteroaromatic fragments and variable-resolution ring mappings (i.e., two or three ring atoms, plus their substituents, per bead). This is illustrated in Figure 4.
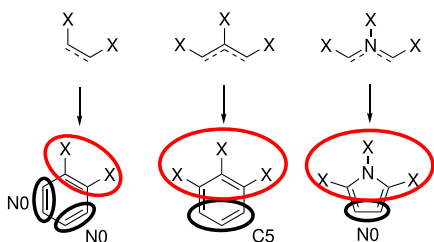

Figure 4. Procedure for constructing full rings from mapped ring fragments.

Aromatic fragments are detected automatically using the aromaticity model in RDKit and are recognized in the fragment's SMILES code. For each aromatic fragment arising from the mapping algorithm, we construct a standardized test molecule containing the aromatic fragment along with the rest of a six-membered ring. In cases where this does not result in a valid structure (most commonly with nitrogen-containing heteroaromatics), we instead form a five-membered ring. The bead type is then chosen in a similar way to nonring fragments, except that the ΔG_{OW} of the whole ring is matched to that of a Martini ring or dimer (depending on the CG resolution). The added carbon atoms are represented by predetermined bead types, which were chosen by finding the closest ΔG_{OW} match for benzene (either an N0 trimer or a C5 dimer). Note that these bead selections do not match the types used in most existing Martini models, but we chose the bead selections that fitted best with ΔG_{OW} for consistency with the approach used for aliphatic fragments. A similar observation was made in a recent paper by Kanekal and Bereau.[57]

2.3. Bonded Interactions. *2.3.1. Bond Lengths and Angles.* Bond lengths and angles are modeled using harmonic potentials. We parameterize these contributions using atomistic conformers of the molecule, generated using the experimental-torsion knowledge distance geometry (ETKDG) method,[62] which is based on information from experimental crystal structures, followed by minimization using the universal force field (UFF).[63] We use 200 such conformers for the parameterization to provide a balance between computational cost and convergence, as discussed in Section 4.1. The set of conformers is mapped to a CG resolution, and bond lengths and angles are taken as the average values across this set. The standard values from the Martini model are used for force constants. Further details on the effect of different bond parameterization schemes can be found in the SI.

The CG beads involved in bond and angle interactions are simply groups of two or three fragments that are bonded together in the atomistic representation of the molecule. This does not apply to pairs or triplets in which all of the beads are part of the same ring. Bonded interactions within rings are

modeled using a network of constraints, dihedrals, and virtual sites, as described in the following sections.

2.3.2. Virtual Sites. Simulations of coarse-grained ring structures often suffer from instabilities at longer time steps because of the complex networks of constraints that must be simultaneously satisfied to enforce the correct molecular geometry. These convergence problems can be mitigated by representing some of the beads in aromatic systems as virtual sites. The virtual sites exert forces on other sites as normal, but any force acting on them is redistributed to the surrounding beads.[64] The position of a virtual site is defined as a weighted combination of the positions of other beads, which also defines how the force is redistributed. Because virtual sites do not interact through normal bonded interactions, they reduce the complexity of the bonded structure of CG ring compounds, improving their stability in simulations. This approach has been successfully deployed, for example, in the treatment of sterols by Melo et al.[59]

We have introduced an automated procedure to include virtual sites in coarse-grained models of large ring systems. Fragments are split into outer and inner sets of beads, which are designated real and virtual sites, respectively. The first stage of this process is to define a two-dimensional coordinate system for each ring system. The axes of this system are found by diagonalization of the inertia tensor, $\mathbf{I}$, of the ring. The eigenvectors associated with the two smallest eigenvalues of $\mathbf{I}$ define an orthogonal basis in the plane of the ring. The coordinates of the ring system are then projected onto this two-dimensional coordinate set. The outer beads are determined by finding the convex hull of the ring system,[65] and these are designated as real sites. The remaining ring beads are designated as virtual sites. Any virtual site that is bonded to a nonring substituent is moved to the list of real sites.

There are many methods for handling virtual sites within simulations. We use a construction in which the position of each virtual site, $\mathbf{v}$, is described by a linear combination of its four nearest neighbors within the same ring system (or three if the system has only three real sites)

$$\mathbf{v} = \sum_{i=1}^{n} w_i \mathbf{r}_i \tag{3}$$

where w_i and $\mathbf{r}_i$ are the weighting and position of site i, respectively, and n is the number of sites used to express $\mathbf{v}$. Choosing appropriate weights for a set of real and virtual sites requires finding a solution to eq 3 for a representative configuration of the real and virtual beads. By definition, rigid fragments have little internal variation between conformers, so any of the conformers generated in Section 2.3.1 can be used. When $n = 3$ or 4, there are infinitely many solutions of eq 3. However, in practice, we would like the solution that gives the most balanced distribution of forces across the constructing sites. This is done with the aid of simple geometric constructions, as described below.

When $n = 4$, there is no unique solution to eq 3. We can choose weights that balance the force redistributed onto each neighbor by placing the virtual site at the intersection of two lines crossing the quadrilateral. The ends of the two lines are defined by the parameters β_h and β_v, which represent fractions of the edges of a quadrilateral, as shown in Figure 5a. The weights are given geometrically by

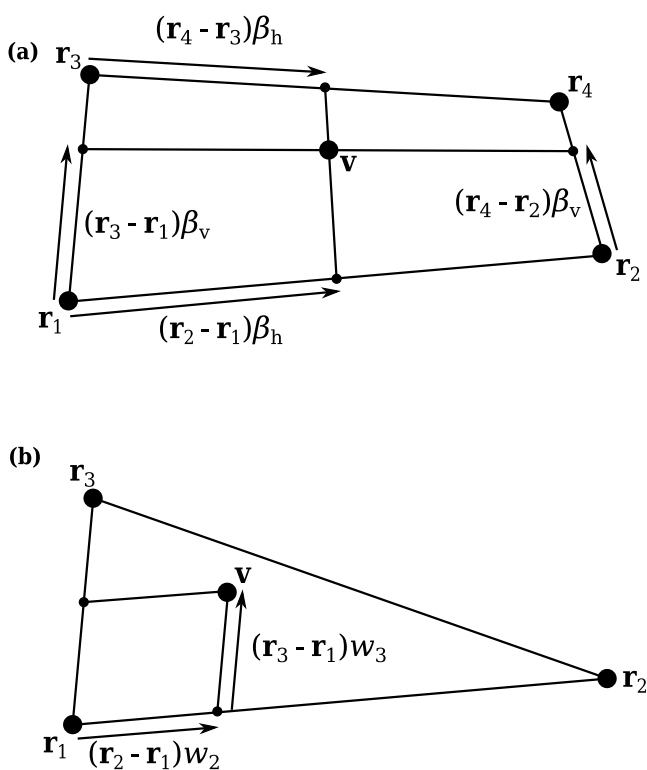**Figure 5.** Schematic of the geometric constructions used to determine weights for virtual sites when (a) $n = 4$ and (b) $n = 3$.

$$w_1 = (1 - \beta_v)(1 - \beta_h)$$
$$w_2 = (1 - \beta_v)\beta_h$$
$$w_3 = \beta_v(1 - \beta_h)$$
$$w_4 = \beta_v\beta_h \tag{4}$$

After shifting the coordinates so that $\mathbf{r}_1$ is at the origin, β_v can be determined by solving the quadratic equation (see the SI for a full derivation)

$$a\beta_h^2 + b\beta_h + c = 0, \quad 0 \leq \beta_h \leq 1 \tag{5}$$

where

$$a = r_{2y}t_x - r_{2x}t_y$$
$$b = r_{2y}r_{3x} - r_{2x}r_{3y} - v_yt_x + v_xt_y$$
$$c = v_xr_{3y} - v_yr_{3x}$$
$$t_k = r_{4k} - r_{3k} - r_{2k} \tag{6}$$

β_v is then calculated from

$$\beta_v = \frac{v_x - \beta_h r_{2x}}{r_{3x} + \beta_h t_x} \tag{7}$$

When $n = 3$, as shown in Figure 5b, the procedure is simpler. Again shifting $\mathbf{r}_1$ to the origin, w_2 and w_3 can be calculated by solving

$$\begin{pmatrix} r_{2x} & r_{3x} \\ r_{2y} & r_{3y} \end{pmatrix}\begin{pmatrix} w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} v_x \\ v_y \end{pmatrix} \tag{8}$$

The weight on $\mathbf{r}_1$ is then calculated from

$$w_1 = 1 - w_3 - w_2 \tag{9}$$

2.3.3. Ring Geometry. The network of constraints and dihedrals in a ring compound is designed to enforce a rigid geometry where required while still allowing for large simulation time steps. We therefore aim to generate a model with the minimum number of interactions necessary to keep the correct molecular geometry based on a hinge construction.[59,66] First, the real sites are connected by a closed ring of distance constraints,[67] giving an outer frame. Further constraints are then placed in a zig-zag pattern across the ring system, splitting it into triangles and creating a series of hinges. Each of these hinges is kept at the correct angle using a single dihedral. The dihedrals are placed to minimize the sharing edges. Pseudocode for the algorithm for adding constraints and dihedrals is included in the SI, along with examples of the results for various ring structures and a discussion on the limits of the dihedral generation algorithm.

The process for generating ring structures, including virtual sites, constraints, and dihedrals, is illustrated in Figure 6 for a warfarin molecule. This molecule includes multiple ring systems that are treated separately from each other by the algorithm before being joined to make the molecule.
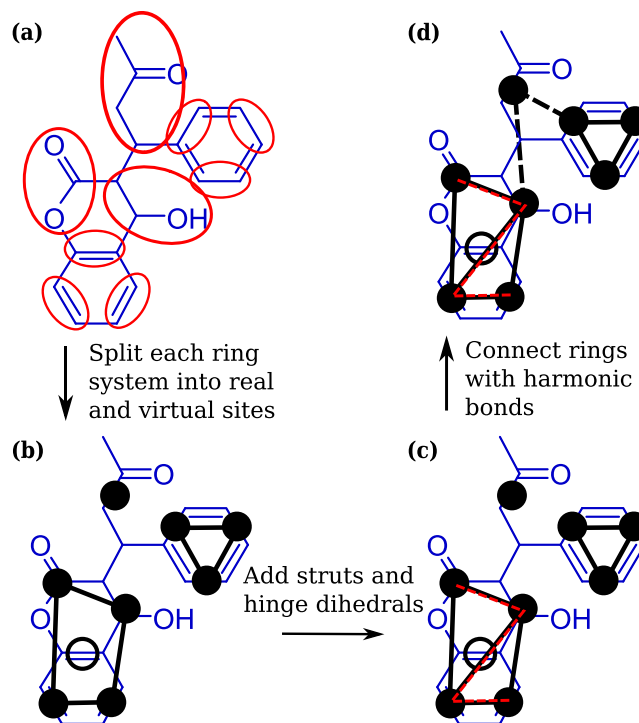

Figure 6. Schematic illustration of the process for generating virtual sites and ring constraints for warfarin. (a) Mapping of atoms onto coarse-grained beads (red ellipses), (b) ring constraints (thick lines), real sites (filled circles), and a virtual site (open circle), (c) dihedral constraint for the double ring (dashed red line), and (d) bonds connecting the ring fragments (thick dashed lines).

3. PARTITION COEFFICIENTS

To test the method presented in Section 2, we use it to derive coarse-grained models for a set of 87 molecules. We benchmark the models by evaluating octanol−water and membrane−water partition coefficients using simulations and comparing them with experimentally measured values.

5782

3.1. General Simulation Parameters.

Coarse-grained simulations were carried out using Gromacs 2018.2.[68] The *new-rf* set of molecular dynamics parameters were used; these were proposed by de Jong et al. as providing the optimal balance between accuracy and computational efficiency.[69] A leapfrog integrator was used, with a time step of 20 fs. Electrostatic interactions were handled using the reaction field method,[70] with a relative permittivity of 15.0. Cutoffs for the van der Waals and electrostatic interactions were set to 1.1 nm. Temperatures were kept constant at 303 K using the velocity rescale method[71] with a coupling constant of 1.0 ps. The Parrinello−Rahman barostat[72] with a coupling constant of 12.0 ps and a reference pressure of 1 bar was used in constant NPT simulations.

Coarse-grained phospholipid parameters were taken from the library of lipids provided on the Martini website.[51] Cholesterol was modeled using the virtual site representation parameterized by Melo et al.[59] Water and ions were modeled using the standard Martini parameters, in which one bead represents four water molecules. Ten percent of the water molecules are the so-called antifreeze particles and prevent the unphysical freezing of water at ambient temperatures.[49]

3.2. Octanol−Water Partitioning.

We start by testing the ability of models generated using our method to reproduce octanol−water partition coefficients (log K_{OW}). Initial structures for these calculations were created by randomly placing 400 water beads or octanol molecules into a cubic box, using the gmx insert-molecule tool included with Gromacs. This was followed by energy minimization and a 10 ns constant NPT run to generate an equilibrated box of solvent.

Solvation free energies in octanol (ΔG_O^{solv}) and water (ΔG_W^{solv}) were calculated using the Bennett acceptance ratio (BAR) method.[73] A solute molecule was placed at a random position in a box of solvent. A series of simulations were carried out in which the interactions between the solute and the solvent were turned off, as defined by a scaling parameter λ. Soft-core potentials were used to prevent singularities when λ is close to 0. The octanol−water partitioning free energy was then calculated directly from

$$\Delta G_{OW}^{part} = \Delta G_W^{solv} - \Delta G_O^{solv} \qquad (10)$$

giving the partition coefficient

$$K_{OW} = \exp(-\Delta G_{OW}^{part}/RT) \qquad (11)$$

Our test set for log K_{OW} is a diverse set of organic molecules, including a variety of functional groups, alkyl chain lengths, and ring structures. A full list is given in the SI. Figure 7 shows the simulation results against experimental measurements. For 16 of the 87 molecules, experimental values were not available, and log K_{OW} predictions from the ALOGPS server were used instead.[60]

Overall, there is a good agreement between the experimental and simulated values, although there are some distinct outliers. The most hydrophilic molecules show more scattered data. Their behavior tends to be dominated by hydrogen-bonding and other directional interactions that are not explicitly included in Martini. These molecules could benefit from the polarizable Martini model[74] or from the recently published Martini 3 force field,[75] and future work will focus on extending the approach to these force fields. There is also a small, systematic overestimation of log K_{OW} in the more lipophilic ring compounds. These molecules are almost exclusively
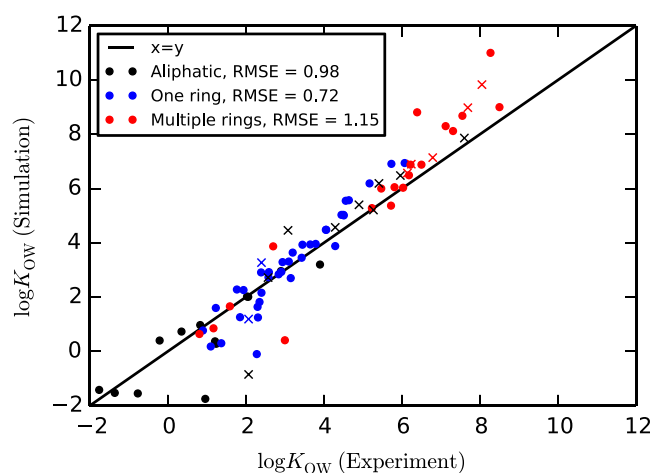

Figure 7. Octanol−water partition coefficients (log K_{OW}) for a set of organic molecules using coarse-grained models compared to experiment (dot markers) or predicted ALOGPS values (cross markers) where experimental data are not available.

multihalogenated aromatics, especially those where a dichloro group is mapped to a single bead.

We also tested our method against a set of existing models from the Martini force field, namely, amino acids. We have compared models for all 20 amino acids generated using our method with those from the Martini 2.2 library.[52] Octanol−water partitioning data for those models are compared to experimental values[76] in Figure 8. Both sets of models
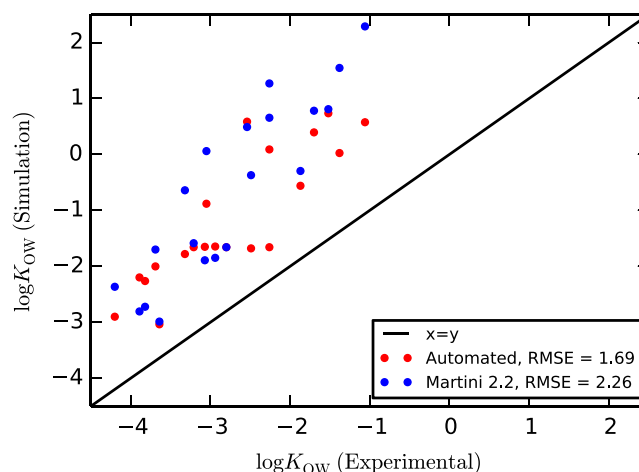

Figure 8. Octanol−water partition coefficients (log K_{OW}) for amino acids calculated using Martini 2.2 and automated models compared to experimental values.

systematically overestimate log K_{OW}, but our automated models (root-mean-square error in log K_{OW}, RMSE = 1.69) outperform the standard Martini 2.2 models (RMSE = 2.26). Of the 20 solutes, 14 have improved results with the automated models. The systematic shift in both sets can be traced to the limitations of the Martini 2 model. It has been shown that using the polarizable Martini model significantly improves the partitioning behavior of highly polar molecules like amino acids, but this is outside the scope of the present study.[74,77] However, it is still encouraging that an automated approach to parameterization results in some improvement

over handmade coarse-grained models within the limitations of the force field itself for this set of compounds.

3.3. Membrane–Water Partitioning. We now test the ability of our automatically generated coarse-grained models to predict membrane–water partition coefficients (log K_{MW}) as a step toward high-throughput screening of organic molecules for bioaccumulation.

3.3.1. Calculating Partition Coefficients. Initial coarse-grained lipid structures were generated using the insane.py tool, which constructs a solvated bilayer by placing lipids and water beads on a cubic lattice with the z-axis normal to the bilayer.[51] All bilayers in this study consist of 128 lipid molecules (two leaflets of 8^2 lipids being a convenient initial construction). A number of water beads were replaced by an equal number of cations and anions to give a 150 mM salt concentration. To facilitate comparison with the experimental data available for each solute, the simulations used either a palmitoyloleoylphosphatidylcholine (POPC) or a dimyristoyl-phosphatidylcholine (DMPC) membrane on a case-by-case basis. Our simulations predict that the differences between log K_{POPC-W} and log K_{DMPC-W} are small but do vary in magnitude between the solutes. For example, in the case of *nnn*-trihexylamine, the mean value across 10 replicas with each membrane was 6.21 for a POPC bilayer and 5.80 for DMPC. For warfarin, the two values were 2.79 and 2.10.

Energy minimization was carried out on the initial structures, followed by a 100 ns constant NPT molecular dynamics run. The average x and z box lengths were calculated from this run, and a single structure with this size was extracted from the trajectory to be used as a starting bilayer configuration.

In this study, umbrella sampling was used to generate probability distributions of solutes across membranes. For each membrane/solute combination, a series of umbrella simulations were run with the distance along the z-axis between the centers of mass of the membrane and the solute restrained at values between 0.0 and 5.0 nm, separated by 0.1 nm. The restraint was achieved using a harmonic potential with a force constant of 1000 kJ mol^{-1} nm^{-2}. The distance was defined using the cylinder method, in which the center of mass of the membrane is calculated using only the lipid molecules that lie within a cylinder of radius 1.5 nm, whose axis is parallel to z and passes through the center of mass of the solute. This reduces the effect of any undulations of the lipid bilayer.[78] Each simulation included two solute molecules being sampled separately across a different leaflet of the membrane to improve sampling without increasing simulation time.[79]

For each umbrella window, the two solute molecules were inserted into the equilibrated bilayer at a fixed z-coordinate and random x and y coordinates. A steepest-descent energy minimization was carried out, followed by a 100 ps equilibration simulation with a 2 fs time step. A 50 ns simulation with a 20 fs time step was then carried out. The first 1 ns of this run was treated as additional equilibration, and the remaining 49 ns were used for analysis. A probability profile was generated using the weighted histogram analysis method (WHAM).[80]

The membrane–water partition coefficient is defined by

$$K_{MW} = \frac{[\text{solute}]_M}{[\text{solute}]_W} \tag{12}$$

where $[\text{solute}]_X$ is the concentration of the solute in the membrane (X = M) or in the aqueous phase (X = W). Extracting K_{MW} from the probability profile requires the profile to be split into parts associated with the solution and with the bilayer. This is complicated by the fact that the interface region is large in relation to the size of the membrane and often has a significantly different interaction with solutes than either the membrane or water. Previous studies have tackled this issue by scaling the contribution to the membrane probability in the interfacial region.[17] Another approach is simply to use a fixed cutoff for a particular lipid type. Both methods work well for most molecules but can neglect important features of the profile when a solute interacts strongly with the interfacial region.

Our approach is based on the assumption that any deviations from the probability in the bulk water region must be due to interactions with the membrane, including the interfacial region. We therefore define a hard cutoff between the two regions but with an adjustable position that depends on the point at which the free-energy profile of the solute starts to deviate from the pure water value. Using this cutoff, we can calculate K_{MW} using

$$K_{MW} = \frac{V(z_n)}{M} \frac{\sum_{i=0}^{R} P_{sol}(z_i)}{P_{sol}(z_n)} \tag{13}$$

Here, the membrane–water system has been divided into notional layers parallel to the xy plane, and z_i refers to the position of layer i, running from $i = 0$ at the center of the membrane to $i = n$ in the outermost water layer (in this work, we have used a resolution of 100 layers). R is the index of the outermost layer designated as part of the membrane region for a given solute. $P_{sol}(z_i)$ is the probability of finding a given solute molecule in layer i. $V(z_n)$ is the volume of the outermost water layer and M is the mass of one leaflet of the lipid bilayer. This gives K_{MW} in units of dm^3 kg^{-1}, in common with many experimental studies. The deviation from bulk behavior at a layer j is estimated using the root-mean-square deviation (RMSD) of the free-energy profile between z_n and z_j. By iterating over j between $j = n$ and 0 and calculating the RMSD at each point, we can define the cutoff as the layer $j = R$, at which the RMSD first increases above 0.1.

The specific value for the RMSD threshold used here is arbitrary, but we have found in most cases that the results are insensitive to the threshold applied. This is because, in cases where the interface is an important contributor to the membrane partitioning, the change in free energy on approaching the interface from the water phase is usually quite sharp. With the exception of very hydrophilic compounds, the calculated K_{MW} does not significantly depend on the way the probability distribution is split between phases (fixed or variable hard cutoff or scaled cutoff). This is illustrated in Figure 9 for trihexylamine and ethylene glycol. For the former, log K_{MW} is dominated by the well inside the membrane, so the position of the cutoff does not affect the result once the whole tail region is covered. In the case of ethylene glycol, there is only a shallow, local free-energy minimum inside the membrane, making the interfacial region more influential on partitioning, and log K_{MW} does not fully converge with cutoff distance. However, this merely reflects the fact that the free-energy density for this solute is lower in water than anywhere in the membrane, so that artificially associating more water with the membrane will always appear to increase
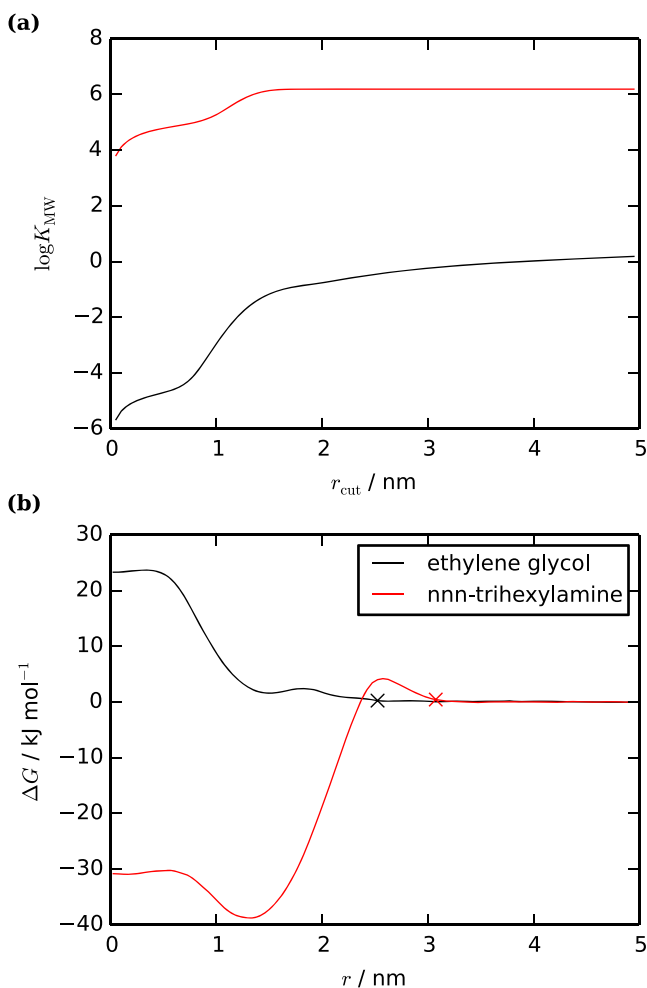
Figure 9. (a) Convergence of $\log K_{MW}$ with cutoff between the membrane and water for ethylene glycol and trihexylamine. (b) Free-energy profiles for the two compounds, with cutoff from an RMSD threshold of 0.1 indicated by crosses. The center of the membrane lies at $r = 0$.

K_{MW}. Nevertheless, as indicated in Figure 9b, using an RMSD threshold of 0.1 does lead to a sensible division between the membrane and water. Overall, for these very hydrophilic compounds, getting a precise estimate of $\log K_{MW}$ is challenging, but, by its very nature, this problem only arises when $\log K_{MW}$ is very small.

As a spot-check on the statistical uncertainty of the $\log K_{MW}$ values calculated by our protocols, we repeated the umbrella sampling procedure 10 times for each of four contrasting molecules using different randomly inserted coordinates of the solute for the starting coordinates. The standard deviation across the 10 calculated values of $\log K_{MW}$ is an estimate of the uncertainty in any one of them (i.e., the typical error present without the repeats). The results from these simulations are shown in Table 1. The estimated uncertainties are all below 0.2 log units. Ethylene glycol has the largest uncertainty, which is due to the lack of a clear-cut division between the membrane and water regions for hydrophilic molecules, as explained above. Experimental measurements of $\log K_{MW}$ are not always published with error estimates, but the statistics given in one recent study[8] (using solid-supported lipid membranes) imply that the uncertainty in an individual measurement was in the range of ±0.03 to ±0.2 with a mean of ±0.13. This level of

Table 1. Mean and Standard Deviation (SD) of $\log K_{MW}$ Calculated from 10 Separate Replicas

solute	mean	SD
trihexylamine	6.21	0.12
warfarin	2.79	0.12
glycerol	0.56	0.04
ethylene glycol	−0.23	0.18

statistical uncertainty is very similar to that of the spot-checks listed in Table 1.

Figure 10a shows $\log K_{MW}$ values for our test set from CG simulations compared to experimental values.[3,7,8,81−95] For
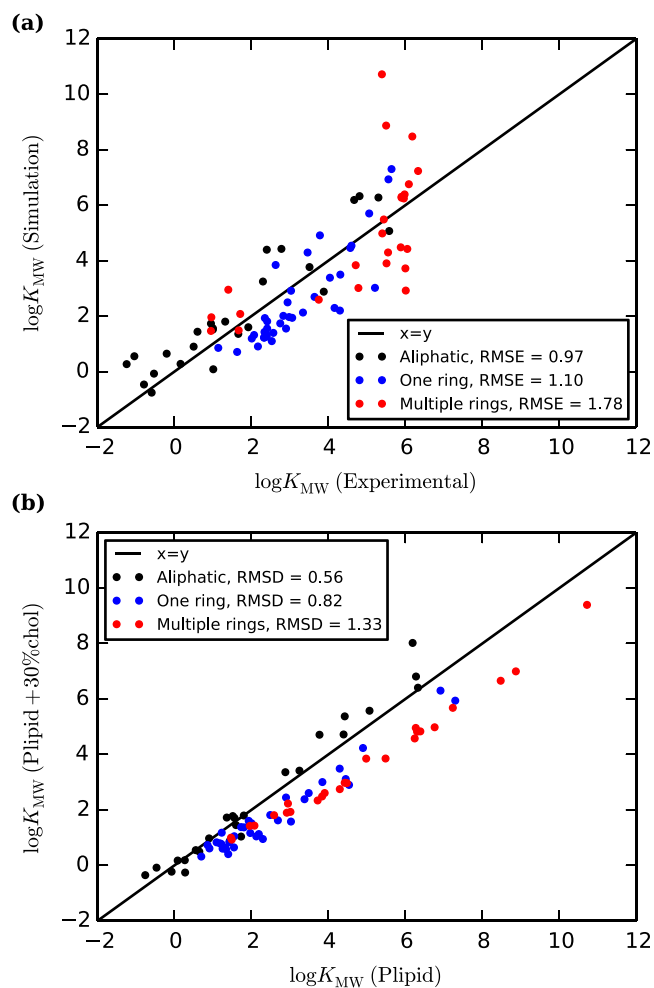

Figure 10. Membrane−water partition coefficients ($\log K_{MW}$) for a range of small organic molecules using coarse-grained models. (a) Coarse-grained vs experimental in a pure phospholipid membrane and (b) $\log K_{MW}$ in a phospholipid + 30% cholesterol vs pure phospholipid membrane, both from the coarse-grained models.

aliphatic compounds, the agreement between the simulation and experiment is essentially the same as for octanol−water partitioning (RMSEs 0.97 and 0.98, respectively), despite the fact that membrane−water partitioning is more challenging both experimentally and computationally. Single-ring compounds also perform well. The least satisfactory results are for a set of halogenated biphenyl compounds. The trend in simulated $\log K_{MW}$ for this set is similar to that in both the simulated and experimental $\log K_{OW}$, but the experimental

$\log K_{MW}$ data have a very small spread. It is therefore not clear whether the discrepancy should be attributed to the models, the experimental data, or a combination of the two. The fact that the octanol–water trend is correct suggests that the models themselves are basically sound, but there may be subtleties in the interactions between halogenated fragments and the membrane, which are not captured well by the bead types in Martini 2.

3.3.2. Membrane Composition. The partitioning of molecules into membranes with realistic compositions is important for the prediction of how compounds interact with real biological systems. A number of experimental and simulation studies have shown that the presence of cholesterol in a lipid bilayer has a significant impact on partitioning.[96−98] Cholesterol is known to have an ordering effect on bilayer systems,[99] and this change is hypothesized to reduce the permeability of the membrane.[96,100−102] However, most studies to date have focused only on small sets of molecules. Automated coarse-graining could be valuable in exploring in depth how different families of molecules interact with different membrane compositions.

As an initial investigation, we have calculated $\log K_{MW}$ for our full test set in lipid bilayers containing 30 mol % of cholesterol and compared this to our simulated values in pure phospholipid bilayers in Figure 10b. The impact of cholesterol in our Martini simulations varies according to the structure of the solute. For molecules containing rigid aromatic groups, cholesterol reduces membrane partitioning, while for small or flexible molecules, it is increased. Figure 11 illustrates the difference in free-energy profiles for tetrachlorocatechol and trihexylamine, which show these contrasting behaviors. Both profiles have a slightly lower free energy in the head-group region of the cholesterol-containing membrane, reflecting the lower concentration of phospholipid heads, but the partition coefficient is dominated by the free energy further into the membrane. Tetrachlorocatechol has a higher free energy in the tail region of the cholesterol membrane. This follows the pattern seen in many studies, where the increase in membrane stiffness on adding cholesterol hinders the reorganization of the lipids, which is required when a solute is added.[96,100−102] In the case of trihexylamine, the free energy is lower in the tail region of the cholesterol membrane. This molecule is more flexible, reducing the lipid reorganization that is required on insertion, and so the favorable interactions with the denser membrane dominate the free-energy profile.

3.3.3. Ionic Molecules. Many of the molecules studied in the previous sections have acidic or basic character, and their protonation state, and therefore charge, will change on moving through a lipid bilayer. We have modeled all of these molecules using neutral Martini beads. These beads are parameterized to match organic solvent–water partitioning data and therefore account for these changes in protonation state in an effective, implicit way.

Strongly ionic molecules are more difficult to model using Martini. The polarizable Martini model, which includes a water model with orientational polarizability, can help with some of the issues. However, the problem remains that there is only a limited selection of ionic bead types within Martini 2, which limits the ionic groups that can be represented. This means our mapping procedure is not guaranteed to work with all ionic molecules.

With prior knowledge of which functional groups can be represented by particular charged bead types, it is possible to
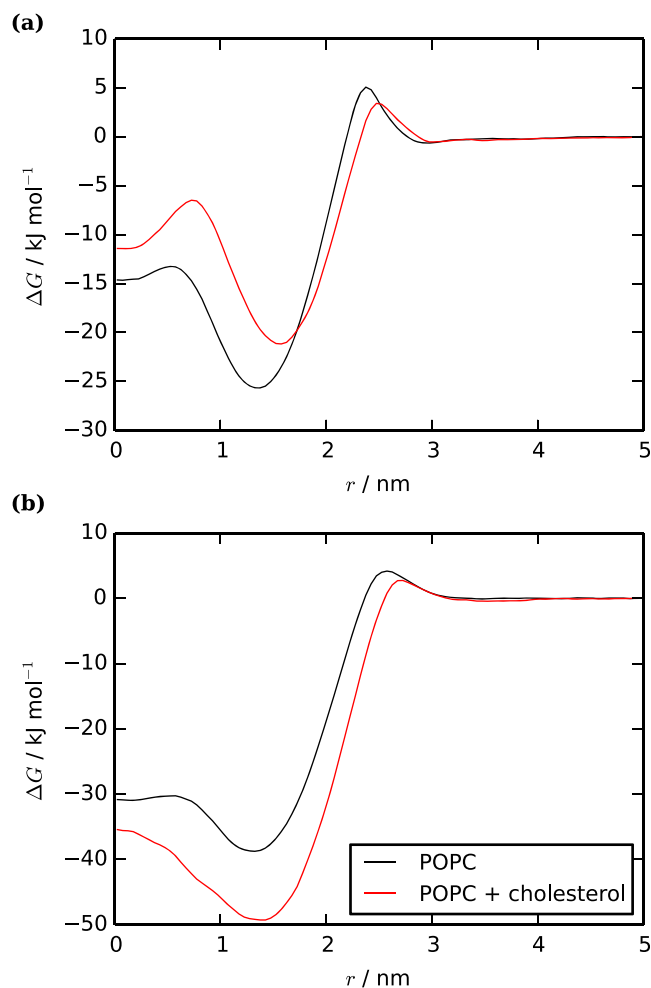
Figure 11. Potentials of mean force between the bilayer center ($r = 0.0$) and bulk water ($r = 5.0$) for (a) tetrachlorocatechol and (b) trihexylamine, in POPC and POPC + 30% cholesterol.

automatically generate models of molecules containing those groups. An additional preprocessing step is added to the mapping scheme where predefined fragments are mapped in a similar way to the aromatic patterns. The bead type for these mappings is also predefined. For example, the SO_4^- and CSO_3^- groups in organic sulfate and sulfonates are represented well by the Qa and Q0 bead types, respectively. This is shown in Figure 12, where these models and models generated without the additional preprocessing are compared to experimental data.[8] Both sets of sulfate models give good results. For the sulfonates, the models based on predefined fragments give significantly improved results over the standard set of models.

Care is needed when sampling the free-energy profiles of ionic molecules. The force constant used in the restraint potential of the umbrella sampling for neutral molecules was weak enough that the charged group in the molecule could be pulled away from the desired depth and toward the charged region of the membrane. This behavior prevented convergence of the histograms from the two molecules. The problem was fixed by increasing the force constant from 1000 to 2000 kJ mol nm^{-2}, and reducing the window spacing to 0.05 nm. For neutral molecules, the weaker force constant was sufficient to obtain converged histograms due to the absence of electrostatic interactions between the solute and the membrane. For those molecules, the increased computational expense of
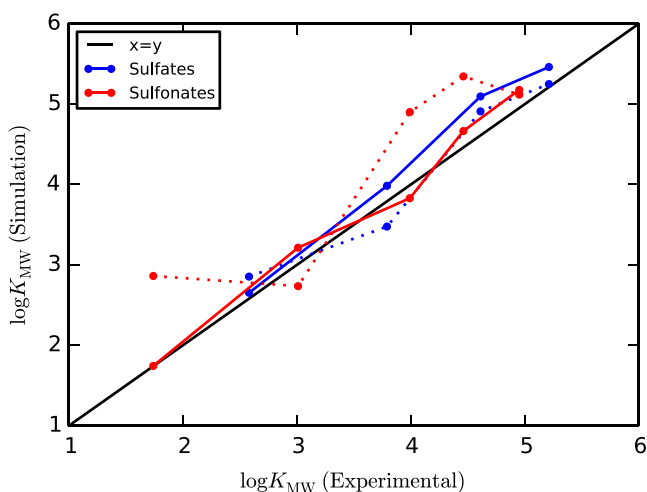
Figure 12. Membrane–water partition coefficients ($\log K_{MW}$) for alkyl sulfates (chain lengths 8, 10, 12, and 13) and sulfonates (chain lengths 8, 10, and 12−14), with models generated with predefined fragment matching (solid lines) and with the standard mapping procedure (dashed lines). In all cases, increasing the chain length increases $\log K_{MW}$.

Figure 13. CPU time of the parameterization for linear alkanes of different chain lengths, with and without UFF minimization of the 200 atomistic conformers. Timings for automartini[22] are also included for reference.

narrower umbrella windows brings no significant improvement in accuracy and so the weaker force constant is preferable.

Predefined mapping for ionic groups can give improved results across specific series of ionic surfactants. However, it is still limited by the resolution of Martini and knowledge of appropriate bead assignments. This is one area where the increased diversity of bead types in the new Martini 3 force field[75] could open up the chemical space that can be covered by our automated approach.

4. DISCUSSION

4.1. Scalability. One advantage of the graph-based mapping algorithm is its computational speed. Even for large molecules, the division into beads takes a matter of seconds and so the approach scales very well with molecule size. The limiting factor for scalability of the coarse-graining procedure as a whole is the conformer generation required for parameterizing bonded interactions. As shown in Figure 13, the entire process for linear alkanes with just one conformer takes only seconds even for C$_{50}$. When 200 conformers (the default in our code) are used, the computational time becomes noticeable. This reflects the cost of generating the conformers using ETKDG and looping over them to calculate their dimensions. Even so, the cost is still small compared to the simulations that are then carried out using the model, so the creation of the CG model would not be the limiting factor in predicting partition coefficients. For reference, on a cluster with 64 GB of RAM, the automartini code[22,57] took 3.7 h to parameterize a C$_{29}$ molecule, and C$_{30}$ could not be parameterized due to memory constraints. With our method, the same molecules took 155 and 167 s, respectively.

While the mapping algorithm is, in principle, scalable to very high molecular weights, the low availability of experimental $\log K_{MW}$ for very large molecules does limit the sizes that can be included in our data set, which range from 2 to 28 heavy atoms. Benchmarking of much larger molecules would therefore require a different application from membrane–water partitioning to be chosen.
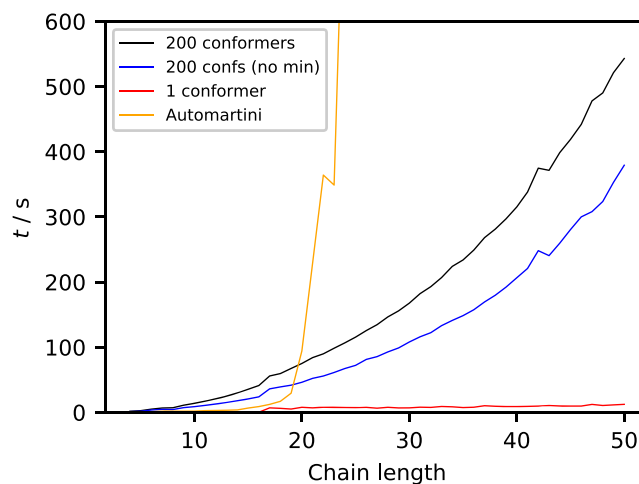
Our coarse-graining method relies on a principle of additivity, i.e., that a molecule can be constructed by joining beads that have been parameterized to represent isolated fragments.[22] It is possible for the most appropriate Martini bead type for a fragment to depend on its chemical surroundings (as demonstrated, for example, in peptides[103]) and the errors associated with this dependence may accumulate as the molecular weight increases. Further work on ways to account for a fragment's environment when choosing a bead type will be extremely valuable and in combination with our mapping approach will further extend the chemical space for which Martini can be used in high-throughput screening studies.

4.2. Nonbonded Parameters. Martini 2 was used in this paper because it is by far the most widely used coarse-grained force field. However, the mapping stage of our scheme is suitable for any force field with the same coarse-grained resolution, such as the recently released Martini 3 model, or any future coarse-grained models. Full automation of the process for new force fields will likely require new approaches to selecting bead types in the parameterization stage. Martini 3, for example, has more bead sizes and more hydrogen-bonding bead types than Martini 2,[66,75] and differentiating between fragments based only on ΔG_{OW} will not be sufficient. Here, we have concentrated on Martini 2 as the simplest version of the force field. However, extension to new force fields can further improve the chemical space coverage of automated coarse-graining, particularly for charged molecules.

The intended bead size in Martini 2 is 4−5 heavy atoms but some of our mappings contain sizes outside this range. The criterion we used for bead size was a compromise between the preferred Martini size and the need for the method to work for any molecule. This includes not only linear or ring fragments but also highly branched structures. In the most branched molecules, such as perfluorinated alkyl chains, beads of 6−8 heavy atoms are possible. In these extreme cases, the larger beads are smaller in radius than a linear bead representing the same number of atoms and help to avoid excessively short bond lengths between beads. Nevertheless, the variation in bead size may cause artifacts in the free-energy profile for molecules where many beads are outside the 4−5 heavy atom

range for which the beads were parameterized,[104] as well as affecting other properties. Martini 2 does have small bead types, but these are identical to the normal bead types aside from their scaled interaction with other small beads.[49] Extension to Martini 3 may alleviate these issues, as the newer force field has more sophisticated treatment of small bead types.[75]

The models generated using our method are intended for use in screening for thermodynamic properties. However, they may also be used as a starting point for a more detailed parameterization. For example, bead types may need to be fine-tuned if the automated assignment is not appropriate. This can occur when the behavior of an individual fragment differs significantly from the fragment within a larger molecule.[103] For example, if an ester linkage (-C($=$O)O-) is assigned to a bead by the mapping algorithm, the addition of the terminal hydrogen atoms will give formic acid, which is chemically very different. In that example, we found that a P3 bead reproduced partition coefficients better than the automatically assigned P1 bead.

4.3. Bonded Parameters. The bonded parameters in our models are sufficient for calculating thermodynamic properties. However, bulk and structural properties require a change of priority when parameterizing intramolecular interactions. Adjusting the automated bonded interactions based on molecule size rather than center-of-mass distances between beads should improve the accuracy of the model for properties such as bulk density, particularly in combination with Martini 3 bead types.[75] However, it is important to note that even with improved bonded interactions, coarse-grained models based on pair potentials will be limited in their ability to correctly represent all of the structural and thermodynamic properties of the underlying atomistic system simultaneously.[30,105] While it is desirable for a model to represent the real system as closely as possible, the simplicity of the model will always come with trade-offs in the properties that can be matched.

All of the automatically generated models shown in this paper have been tested and found to produce a stable simulation with umbrella sampling and a time step of 20 fs, which is a similar level of performance to handmade models of similar systems.[59] For even more complex aromatic molecules with many substituents, a shorter time step may sometimes be necessary since keeping the correct geometry of such molecules will require many bonded parameters acting on a small number of beads.

5. CONCLUSIONS

We have developed an efficient, automated coarse-grained mapping method that is compatible with Martini and other force fields of a similar resolution. The method consists of graph-based mapping for nonring fragments and pattern-matching for ring fragments. Using this approach, we can generate models for a large number of organic molecules. This includes a wide variety of functional groups, as well as linear, branched, and ring compounds. The resulting models reproduce both octanol—water and membrane—water partition coefficients of the solute molecules well. Compared to previous approaches, we have expanded the range of ring-containing molecules that are accessible by introducing a new method of parameterizing virtual sites and constraints. The structures are stable even with a 20 fs time step, which is generally difficult to achieve for this type of molecule.

The new mapping and parameterization approach paves the way for routine use of coarse-grained modeling in high-throughput screening studies involving molecules of industrial interest. Due to the large chemical space that can be covered, it has the potential to be used as an alternative to simpler screening approaches that make more assumptions. We have demonstrated this point for simple lipid bilayers, including preliminary results for membranes that contain cholesterol, but the approach should be applicable to screening for partitioning into other systems for which there is a well-characterized Martini model. Further refinements of the method, including extension to Martini 3, will continue to expand the accessible chemical space and open up further applications.

ASSOCIATED CONTENT

ⓈI Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.1c00322.

> Additional details of the parameterization of virtual sites and constraint networks for ring compounds, and results on the effects of using different bond parameterization schemes (PDF)
>
> Full implementation of the algorithm for generating coarse-grained models described in this study (ZIP)
>
> Gromacs topologies for all of the coarse-grained models generated in the study (ZIP)
>
> Data file containing a full list of compounds studied, along with their experimental and simulated partitioning data (XLSX)

AUTHOR INFORMATION

Corresponding Author

    Mark A. Miller — *Department of Chemistry, Durham University, Durham DH1 3LE, United Kingdom;* ⓞ orcid.org/0000-0002-1403-3343; Email: m.a.miller@durham.ac.uk.

Authors

    Thomas D. Potter — *Department of Chemistry, Durham University, Durham DH1 3LE, United Kingdom;* ⓞ orcid.org/0000-0002-0480-4183

    Elin L. Barrett — *Unilever Safety and Environmental Assurance Centre, Bedfordshire MK44 1LQ, United Kingdom*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jctc.1c00322

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W. Computational Methods in Drug Discovery. *Pharmacol. Rev.* **2014**, *66*, 334–395.

(2) Raies, A. B.; Bajic, V. B. In silico toxicology: computational methods for the prediction of chemical toxicity. *WIREs Comput. Mol. Sci.* **2016**, *6*, 147–172.

(3) Escher, B. I.; Schwarzenbach, R. P.; Westall, J. C. Evaluation of Liposome—Water Partitioning of Organic Acids and Bases. 1. Development of a Sorption Model. *Environ. Sci. Technol.* **2000**, *34*, 3954–3961.

(4) Seydel, J. K. *Drug-Membrane Interactions*; John Wiley & Sons, Ltd., 2003; pp 35−50.

(5) Endo, S. Re-analysis of narcotic critical body residue data using the equilibrium distribution concept and refined partition coefficients. *Environ. Sci.: Processes Impacts* **2016**, *18*, 1024−1029.

(6) Escher, B. I.; Baumer, A.; Bittermann, K.; Henneberger, L.; König, M.; Kühnert, C.; Klüver, N. General baseline toxicity QSAR for nonpolar, polar and ionisable chemicals and their mixtures in the bioluminescence inhibition assay with *Aliivibrio fischeri*. *Environ. Sci.: Processes Impacts* **2017**, *19*, 414−428.

(7) Droge, S. T. J.; Hermens, J. L. M.; Gutsell, S.; Rabone, J.; Hodges, G. Predicting the phospholipophilicity of monoprotic positively charged amines. *Environ. Sci.: Processes Impacts* **2017**, *19*, 307−323.

(8) Droge, S. T. J. Membrane−Water Partition Coefficients to Aid Risk Assessment of Perfluoroalkyl Anions and Alkyl Sulfates. *Environ. Sci. Technol.* **2019**, *53*, 760−770.

(9) Bittermann, K.; Spycher, S.; Goss, K.-U. Comparison of different models predicting the phospholipid-membrane water partition coefficients of charged compounds. *Chemosphere* **2016**, *144*, 382−391.

(10) Platts, J. A.; Abraham, M. H. Partition of Volatile Organic Compounds from Air and from Water into Plant Cuticular Matrix: An LFER Analysis. *Environ. Sci. Technol.* **2000**, *34*, 318−323.

(11) Goss, K.-U. Predicting the equilibrium partitioning of organic compounds using just one Linear Solvation Energy Relationship (LSER). *Fluid Phase Equilib.* **2005**, *233*, 19−22.

(12) Stenzel, A.; Goss, K.-U.; Endo, S. Determination of Polyparameter Linear Free Energy Relationship (pp-LFER) Substance Descriptors for Established and Alternative Flame Retardants. *Environ. Sci. Technol.* **2013**, *47*, 1399−1406.

(13) Hüffer, T.; Endo, S.; Metzelder, F.; Schroth, S.; Schmidt, T. C. Prediction of sorption of aromatic and aliphatic organic compounds by carbon nanotubes using poly-parameter linear free-energy relationships. *Water Res.* **2014**, *59*, 295−303.

(14) Klamt, A. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem. A* **1995**, *99*, 2224−2235.

(15) Klamt, A.; Huniar, U.; Spycher, S.; Keldenich, J. COSMOmic: A Mechanistic Approach to the Calculation of Membrane-Water Partition Coefficients and Internal Distributions within Membranes and Micelles. *J. Phys. Chem. B* **2008**, *112*, 12148−12157.

(16) Di Meo, F.; Fabre, G.; Berka, K.; Ossman, T.; Chantemargue, B.; Paloncýová, M.; Marquet, P.; Otyepka, M.; Trouillas, P. In silico pharmacology: Drug membrane partitioning and crossing. *Pharmacol. Res.* **2016**, *111*, 471−486.

(17) Jakobtorweihen, S.; Zuniga, A. C.; Ingram, T.; Gerlach, T.; Keil, F. J.; Smirnova, I. Predicting solute partitioning in lipid bilayers: Free energies and partition coefficients from molecular dynamics simulations and COSMOmic. *J. Chem. Phys.* **2014**, *141*, No. 045102.

(18) Bernardi, R. C.; Melo, M. C. R.; Schulten, K. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 872−877.

(19) Yang, Y. I.; Shao, Q.; Zhang, J.; Yang, L.; Gao, Y. Q. Enhanced sampling in molecular dynamics. *J. Chem. Phys.* **2019**, *151*, No. 070902.

(20) Rudzinski, J. F.; Noid, W. G. Investigation of Coarse-Grained Mappings via an Iterative Generalized Yvon−Born−Green Method. *J. Phys. Chem. B* **2014**, *118*, 8295−8312.

(21) Chakraborty, M.; Xu, C.; White, A. D. Encoding and selecting coarse-grain mapping operators with hierarchical graphs. *J. Chem. Phys.* **2018**, *149*, No. 134106.

(22) Bereau, T.; Kremer, K. Automated Parametrization of the Coarse-Grained Martini Force Field for Small Organic Molecules. *J. Chem. Theory Comput.* **2015**, *11*, 2783−2791.

(23) Webb, M. A.; Delannoy, J.-Y.; de Pablo, J. J. Graph-Based Approach to Systematic Molecular Coarse-Graining. *J. Chem. Theory Comput.* **2019**, *15*, 1199−1208.

(24) Reith, D.; Pütz, M.; Müller-Plathe, F. Deriving effective mesoscale potentials from atomistic simulations. *J. Comput. Chem.* **2003**, *24*, 1624−1636.

(25) Lyubartsev, A. P.; Naômé, A.; Vercauteren, D. P.; Laaksonen, A. Systematic hierarchical coarse-graining with the inverse Monte Carlo method. *J. Chem. Phys.* **2015**, *143*, No. 243120.

(26) Izvekov, S.; Voth, G. A. Multiscale coarse graining of liquid-state systems. *J. Chem. Phys.* **2005**, *123*, No. 134105.

(27) Izvekov, S.; Voth, G. A. A Multiscale Coarse-Graining Method for Biomolecular Systems. *J. Phys. Chem. B* **2005**, *109*, 2469−2473.

(28) Carmichael, S. P.; Shell, M. S. A New Multiscale Algorithm and Its Application to Coarse-Grained Peptide Models for Self-Assembly. *J. Phys. Chem. B* **2012**, *116*, 8383−8393.

(29) Louis, A. A. Beware of density dependent pair potentials. *J. Phys.: Condens. Matter* **2002**, *14*, 9187−9206.

(30) Johnson, M. E.; Head-Gordon, T.; Louis, A. A. Representability problems for coarse-grained water potentials. *J. Chem. Phys.* **2007**, *126*, No. 144509.

(31) Brini, E.; Marcon, V.; van der Vegt, N. F. A. Conditional reversible work method for molecular coarse graining applications. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10468−10474.

(32) Potter, T. D.; Tasche, J.; Wilson, M. R. Assessing the transferability of common top-down and bottom-up coarse-grained molecular models for molecular mixtures. *Phys. Chem. Chem. Phys.* **2019**, *21*, 1912−1927.

(33) DeLyser, M. R.; Noid, W. G. Extending pressure-matching to inhomogeneous systems via local-density potentials. *J. Chem. Phys.* **2017**, *147*, No. 134111.

(34) Sanyal, T.; Shell, M. S. Transferable Coarse-Grained Models of Liquid−Liquid Equilibrium Using Local Density Potentials Optimized with the Relative Entropy. *J. Phys. Chem. B* **2018**, *122*, 5678−5693.

(35) Jin, J.; Voth, G. A. Ultra-Coarse-Grained Models Allow for an Accurate and Transferable Treatment of Interfacial Systems. *J. Chem. Theory Comput.* **2018**, *14*, 2180−2197.

(36) Rosenberger, D.; van der Vegt, N. F. A. Addressing the temperature transferability of structure based coarse graining models. *Phys. Chem. Chem. Phys.* **2018**, *20*, 6617−6628.

(37) Moore, T. C.; Iacovella, C. R.; McCabe, C. Derivation of coarse-grained potentials via multistate iterative Boltzmann inversion. *J. Chem. Phys.* **2014**, *140*, No. 224104.

(38) Shen, K.; Sherck, N.; Nguyen, M.; Yoo, B.; Köhler, S.; Speros, J.; Delaney, K. T.; Fredrickson, G. H.; Shell, M. S. Learning composition-transferable coarse-grained models: Designing external potential ensembles to maximize thermodynamic information. *J. Chem. Phys.* **2020**, *153*, No. 154116.

(39) Durumeric, A. E. P.; Voth, G. A. Adversarial-residual-coarse-graining: Applying machine learning theory to systematic molecular coarse-graining. *J. Chem. Phys.* **2019**, *151*, No. 124110.

(40) Wang, W.; Gómez-Bombarelli, R. Coarse-graining auto-encoders for molecular dynamics. *npj Comput. Mater.* **2019**, *5*, 1−9.

(41) Wang, J.; Olsson, S.; Wehmeyer, C.; Pérez, A.; Charron, N. E.; de Fabritiis, G.; Noé, F.; Clementi, C. Machine Learning of Coarse-Grained Molecular Dynamics Force Fields. *ACS Cent. Sci.* **2019**, *5*, 755−767.

(42) Lobanova, O.; Avendaño, C.; Lafitte, T.; Müller, E. A.; Jackson, G. SAFT-γ force field for the simulation of molecular fluids: 4. A single-site coarse-grained model of water applicable over a wide temperature range. *Mol. Phys.* **2015**, *113*, 1228−1249.

(43) Lobanova, O.; Mejía, A.; Jackson, G.; Müller, E. A. SAFT-γ force field for the simulation of molecular fluids 6: Binary and ternary mixtures comprising water, carbon dioxide, and n-alkanes. *J. Chem. Thermodyn.* **2016**, *93*, 320−336.

(44) Tasche, J.; Sabattié, E. F. D.; Thompson, R. L.; Campana, M.; Wilson, M. R. Oligomer/Polymer Blend Phase Diagram and Surface Concentration Profiles for Squalane/Polybutadiene: Experimental Measurements and Predictions from SAFT-γ Mie and Molecular Dynamics Simulations. *Macromolecules* **2020**, *53*, 2299−2309.

(45) Marrink, S. J.; de Vries, A. H.; Mark, A. E. Coarse grained model for semiquantitative lipid simulations. *J. Phys. Chem. B* **2004**, *108*, 750−760.

(46) Maerzke, K. A.; Siepmann, J. I. Transferable Potentials for Phase Equilibria−Coarse-Grain Description for Linear Alkanes. *J. Phys. Chem. B* **2011**, *115*, 3452−3465.

(47) Orsi, M.; Essex, J. W. The ELBA Force Field for Coarse-Grain Modeling of Lipid Membranes. *PLoS One* **2011**, *6*, No. e28637.

(48) Wu, Z.; Cui, Q.; Yethiraj, A. A New Coarse-Grained Force Field for Membrane−Peptide Simulations. *J. Chem. Theory Comput.* **2011**, *7*, 3793−3802.

(49) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111*, 7812−7824.

(50) Marrink, S. J.; Tieleman, D. P. Perspective on the Martini model. *Chem. Soc. Rev.* **2013**, *42*, 6801−6822.

(51) Wassenaar, T. A.; Ingólfsson, H. I.; Böckmann, R. A.; Tieleman, D. P.; Marrink, S. J. Computational Lipidomics with insane: A Versatile Tool for Generating Custom Membranes for Molecular Simulations. *J. Chem. Theory Comput.* **2015**, *11*, 2144−2155.

(52) de Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schäfer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J. Chem. Theory Comput.* **2013**, *9*, 687−697.

(53) Lee, H.; de Vries, A. H.; Marrink, S.-J.; Pastor, R. W. A Coarse-Grained Model for Polyethylene Oxide and Polyethylene Glycol: Conformation and Hydrodynamics. *J. Phys. Chem. B* **2009**, *113*, 13186−13194.

(54) Rossi, G.; Monticelli, L.; Puisto, S. R.; Vattulainen, I.; Ala-Nissila, T. Coarse-graining polymers with the MARTINI force-field: polystyrene as a benchmark case. *Soft Matter* **2011**, *7*, 698−708.

(55) Graham, J. A.; Essex, J. W.; Khalid, S. PyCGTOOL: Automated Generation of Coarse-Grained Molecular Dynamics Models from Atomistic Trajectories. *J. Chem. Inf. Model.* **2017**, *57*, 650−656.

(56) Menichetti, R.; Kanekal, K. H.; Kremer, K.; Bereau, T. In silico screening of drug-membrane thermodynamics reveals linear relations between bulk partitioning and the potential of mean force. *J. Chem. Phys.* **2017**, *147*, No. 125101.

(57) Kanekal, K. H.; Bereau, T. Resolution limit of data-driven coarse-grained models spanning chemical space. *J. Chem. Phys.* **2019**, *151*, No. 164106.

(58) *RDKit: Open-source Cheminformatics.* https://www.rdkit.org,10.5281/zenodo.4107869.

(59) Melo, M. N.; Ingólfsson, H. I.; Marrink, S. J. Parameters for Martini sterols and hopanoids based on a virtual-site description. *J. Chem. Phys.* **2015**, *143*, No. 243152.

(60) Tetko, I. V.; Tanchuk, V. Y. Application of Associative Neural Networks for Prediction of Lipophilicity in ALOGPS 2.1 Program. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1136−1145.

(61) Wildman, S. A.; Crippen, G. M. Prediction of Physicochemical Parameters by Atomic Contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 868−873.

(62) Riniker, S.; Landrum, G. A. Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation. *J. Chem. Inf. Model.* **2015**, *55*, 2562−2574.

(63) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024−10035.

(64) Feenstra, K. A.; Hess, B.; Berendsen, H. J. C. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comput. Chem.* **1999**, *20*, 786−798.

(65) Barber, C. B.; Dobkin, D. P.; Huhdanpaa, H. The quickhull algorithm for convex hulls. *ACM Trans. Math. Softw.* **1996**, *22*, 469−483.

(66) Souza, P. C. T.; Thallmair, S.; Conflitti, P.; Ramírez-Palacios, C.; Alessandri, R.; Raniolo, S.; Limongelli, V.; Marrink, S. J. Protein−ligand binding with the coarse-grained Martini model. *Nat. Commun.* **2020**, *11*, No. 3714.

(67) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.

(68) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1−2*, 19−25.

(69) de Jong, D. H.; Baoukina, S.; Ingólfsson, H. I.; Marrink, S. J. Martini straight: Boosting performance using a shorter cutoff and GPUs. *Comput. Phys. Commun.* **2016**, *199*, 1−7.

(70) Barker, J. A.; Watts, R. O. Monte Carlo studies of the dielectric properties of water-like models. *Mol. Phys.* **1973**, *26*, 789−792.

(71) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, No. 014101.

(72) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.

(73) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245−268.

(74) Michalowsky, J.; Schäfer, L. V.; Holm, C.; Smiatek, J. A refined polarizable water model for the coarse-grained MARTINI force field with long-range electrostatic interactions. *J. Chem. Phys.* **2017**, *146*, No. 054501.

(75) Souza, P. C. T.; et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* **2021**, 1−7.

(76) Howard, P.; Meylan, W. *Physical/Chemical Property Database (PHYSPROP)*, 1999.

(77) Yesylevskyy, S. O.; Schäfer, L. V.; Sengupta, D.; Marrink, S. J. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLoS Comput. Biol.* **2010**, *6*, No. e1000810.

(78) Nitschke, N.; Atkovska, K.; Hub, J. S. Accelerating potential of mean force calculations for lipid membrane permeation: System size, reaction coordinate, solute-solute distance, and cutoffs. *J. Chem. Phys.* **2016**, *145*, No. 125101.

(79) Bochicchio, D.; Panizon, E.; Ferrando, R.; Monticelli, L.; Rossi, G. Calculating the free energy of transfer of small solutes into a model lipid membrane: Comparison between metadynamics and umbrella sampling. *J. Chem. Phys.* **2015**, *143*, No. 144108.

(80) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.

(81) Katz, Y.; Diamond, J. M. A method for measuring non-electrolyte partition coefficients between liposomes and water. *J. Membr. Biol.* **1974**, *17*, 69−86.

(82) Rogers, J. A.; Davis, S. S. Functional group contributions to the partitioning of phenols between liposomes and water. *Biochim. Biophys. Acta* **1980**, *598*, 392−404.

(83) Betageri, G. V.; Rogers, J. A. Thermodynamics of partitioning of β-blockers in the n-octanol- buffer and liposome systems. *Int. J. Pharm.* **1987**, *36*, 165−173.

(84) Gobas, F. A. P. C.; Lahittete, J. M.; Garofalo, G.; Shiu, W. Y.; Mackay, D. A Novel Method for Measuring Membrane-Water Partition Coefficients of Hydrophobic Organic Chemicals: Comparison with 1-Octanol−Water Partitioning. *J. Pharm. Sci.* **1988**, *77*, 265−272.

(85) De Young, L. R.; Dill, K. A. Partitioning of nonpolar solutes into bilayers and amorphous n-alkanes. *J. Phys. Chem. A* **1990**, *94*, 801−809.

(86) Janes, N.; Hsu, J. W.; Rubin, E.; Taraschi, T. F. Nature of alcohol and anesthetic action on cooperative membrane equilibria. *Biochemistry* **1992**, *31*, 9467−9472.

(87) Austin, R. P.; Davis, A. M.; Manners, C. N. Partitioning of ionizing molecules between aqueous buffers and phospholipid vesicles. *J. Pharm. Sci.* **1995**, *84*, 1180−1183.

(88) Dulfer, W. J.; Govers, H. A. J. Membrane-Water Partitioning of Polychlorinated Biphenyls in Small Unilamellar Vesicles of Four Saturated Phosphatidylcholines. *Environ. Sci. Technol.* **1995**, *29*, 2548−2554.

(89) Ottiger, C.; Wunderli-Allenspach, H. Partition behaviour of acids and bases in a phosphatidylcholine liposome−buffer equilibrium dialysis system. *Eur. J. Pharm. Sci.* **1997**, *5*, 223−231.

(90) Vaes, W. H.; Ramos, E. U.; Hamwijk, C.; van Holsteijn, I.; Blaauboer, B. J.; Seinen, W.; Verhaar, H. J.; Hermens, J. L. Solid phase microextraction as a tool to determine membrane/water partition coefficients and bioavailable concentrations in in vitro systems. *Chem. Res. Toxicol.* **1997**, *10*, 1067−1072.

(91) Vaes, W. H. J.; Ramos, E. U.; Verhaar, H. J. M.; Hermens, J. L. M. Acute toxicity of nonpolar versus polar narcosis: Is there a difference? *Environ. Toxicol. Chem.* **1998**, *17*, 1380−1384.

(92) Schweigert, N.; Zehnder, A. J. B.; Eggen, R. I. L. Chemical properties of catechols and their molecular modes of toxic action in cells, from microorganisms to mammals. *Environ. Microbiol.* **2001**, *3*, 81−91.

(93) van der Heijden, S. A.; Jonker, M. T. O. Evaluation of Liposome−Water Partitioning for Predicting Bioaccumulation Potential of Hydrophobic Organic Chemicals. *Environ. Sci. Technol.* **2009**, *43*, 8854−8859.

(94) Droge, S. T. J.; Hermens, J. L. M.; Rabone, J.; Gutsell, S.; Hodges, G. Phospholipophilicity of CxHyN+ amines: chromatographic descriptors and molecular simulations for understanding partitioning into membranes. *Environ. Sci.: Processes Impacts* **2016**, *18*, 1011−1023.

(95) Timmer, N.; Droge, S. T. J. Sorption of Cationic Surfactants to Artificial Cell Membranes: Comparing Phospholipid Bilayers with Monolayer Coatings and Molecular Simulations. *Environ. Sci. Technol.* **2017**, *51*, 2890−2898.

(96) Wennberg, C. L.; van der Spoel, D.; Hub, J. S. Large Influence of Cholesterol on Solute Partitioning into Lipid Membranes. *J. Am. Chem. Soc.* **2012**, *134*, 5351−5361.

(97) Zocher, F.; van der Spoel, D.; Pohl, P.; Hub, J. S. Local Partition Coefficients Govern Solute Permeability of Cholesterol-Containing Membranes. *Biophys. J.* **2013**, *105*, 2760−2770.

(98) Díaz-Tejada, C.; Ariz-Extreme, I.; Awasthi, N.; Hub, J. S. Quantifying Lateral Inhomogeneity of Cholesterol-Containing Membranes. *J. Phys. Chem. Lett.* **2015**, *6*, 4799−4803.

(99) Róg, T.; Pasenkiewicz-Gierula, M.; Vattulainen, I.; Karttunen, M. Ordering effects of cholesterol and its analogues. *Biochim. Biophys. Acta, Biomembr.* **2009**, *1788*, 97−121.

(100) Bennett, W. F. D.; MacCallum, J. L.; Tieleman, D. P. Thermodynamic Analysis of the Effect of Cholesterol on Dipalmitoylphosphatidylcholine Lipid Membranes. *J. Am. Chem. Soc.* **2009**, *131*, 1972−1978.

(101) Yacoub, T. J.; Reddy, A. S.; Szleifer, I. Structural effects and translocation of doxorubicin in a DPPC/Chol bilayer: the role of cholesterol. *Biophys. J.* **2011**, *101*, 378−385.

(102) Pham, V. T.; Nguyen, T. Q.; Dao, U. P. N.; Nguyen, T. T. On the interaction between fluoxetine and lipid membranes: Effect of the lipid composition. *Spectrochim. Acta, Part A* **2018**, *191*, 50−61.

(103) Bereau, T.; Kremer, K. Protein-Backbone Thermodynamics across the Membrane Interface. *J. Phys. Chem. B* **2016**, *120*, 6391−6400.

(104) Alessandri, R.; Souza, P. C. T.; Thallmair, S.; Melo, M. N.; de Vries, A. H.; Marrink, S. J. Pitfalls of the Martini Model. *J. Chem. Theory Comput.* **2019**, *15*, 5448−5460.

(105) Wagner, J. W.; Dama, J. F.; Durumeric, A. E. P.; Voth, G. A. On the representability problem and the physical meaning of coarse-grained models. *J. Chem. Phys.* **2016**, *145*, No. 044108.