

Martini 3 Coarse-Grained Force Field: Small Molecules

Riccardo Alessandri,^{*,†,§} Jonathan Barnoud,^{†,||} Anders S. Gertsen,[‡] Ilias Patmanidis,[†] Alex H. de Vries,[†] Paulo C. T. Souza,^{*,†,¶} and Siewert J. Marrink^{*,†}

[†]*Groningen Biomolecular Sciences and Biotechnology Institute and Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands*

[‡]*Department of Energy Conversion and Storage, Technical University of Denmark, Fysikvej 310, DK-2800 Kgs. Lyngby, Denmark*

[¶]*Current address: Molecular Microbiology and Structural Biochemistry (MMSB, UMR 5086), CNRS & University of Lyon, 7 Passage du Vercors, 69007, Lyon, France*

[§]*Current address: Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637, United States*

^{||}*Current address: Intangible Realities Laboratory, University of Bristol, Bristol BS8 1UB, U.K.*

E-mail: ric.alessandri@gmail.com; paulocts@gmail.com; s.j.marrink@rug.nl

Abstract

The recent re-parametrization of the Martini coarse-grained force field, Martini 3, improved the accuracy of the model in predicting molecular packing and interactions in molecular dynamics simulations. Here, we describe how small molecules can be accurately parametrized within the Martini 3 framework and present a database of validated small molecule models (available at <https://github.com/ricalessandri/Martini3-small-molecules> and <http://cgmartini.nl>). We pay particular attention to the description of aliphatic and aromatic ring-like structures, which are ubiquitous in small molecules such as solvents and drugs or in building blocks constituting macromolecules such as proteins and synthetic polymers. In Martini 3, ring-like structures are described by models that use higher resolution coarse-grained particles (small and tiny particles). As such, the present database constitutes one of the cornerstones of the calibration of the new Martini 3 small and tiny particle sizes. The models show excellent partitioning behavior and solvent properties. Miscibility trends between different bulk phases are also captured, completing the set of thermodynamic properties considered during the parametrization. We also show how the new bead sizes allow for a good representation of molecular volume, which translates into better structural properties such as stacking distances. We further present design strategies to build Martini 3 models for small molecules of increased complexity. The present database, along with the outlined design strategies and the modularity of the Martini force field, constitute a resource of models that 1) can be used “as is” in (bio)molecular simulations; 2) gives reference points for the construction of other small molecule models; 3) provides guidelines and reference data for automated topology builders; and 4) can be used as building blocks for the construction of more complex (macro)molecules, hence enabling investigations of complex biomolecular and soft material systems.

1 Introduction

The use of coarse-grained (CG) models in molecular dynamics (MD) simulations is of paramount importance to explore length and time scales beyond what is feasible with atomistic, or all-atom (AA), models.^{1–6} In CG models, atoms are bundled together into supratomic particles, or beads. Besides widening the accessible spatiotemporal scales, such models can provide a more insightful picture with respect to atomistic ones, as the reduction in the number of degrees of freedom translates into a simpler representation of the system which highlights its key features. Thus, CG simulations are an indispensable tool to complement experimental methods to study processes such as spontaneous self-assembly, transformation between phases, domain formation, and host-guest binding.^{1–6}

CG models can be categorized as mainly following a systematic (also known as hierarchical) or a building block approach.^{4,7} Models developed on the basis of the former principle reproduce the underlying atomistic structural data accurately, but require reparametrization whenever any condition changes, which makes the parametrization procedure more time-consuming. Also, potential forms required are often complex, which can result in lower performance and therefore less sampling. On the other side, models developed following a building block approach are often cheaper due to the use of simpler potential forms and the only partial parametrization required. Moreover, the parametrization of the building blocks enables their use as part of similar moieties in different molecules, providing a second advantage of the building block approach: the transferability of the models developed within such a framework. However, these advantages come at the cost of a more limited structural accuracy, given the necessarily suboptimal representation of the underlying atomistic detail inherent to building block-based CG models. Most of the CG models, however, do not fit purely in either of the categories but often include features of both CG modeling approaches.⁴

Among the building block approaches to coarse-graining, notably the Martini CG force field⁸ has seen wide applications due to successes achieved in the description of a wide range

of biomolecular systems.⁹ We note that the force field, originally intended for biomolecular simulations, has been successfully applied to describe systems relevant in soft materials science, such as (block co)polymers, self-assembled supramolecular materials, and organic semiconductors.¹⁰ The wide use of the Martini model, however, highlighted a number of shortcomings, in particular in the description of structures which are modeled with a finer resolution than Martini’s standard resolution of 4 non-hydrogen atoms per bead (4-to-1).^{11–15} Finer 3-to-1, or 2-to-1 mappings are required when modeling ring-like structures,⁸ or to adhere to the symmetry of the repeat unit of a polymer.¹⁶ For such finer mappings, “small”⁸ and subsequently even smaller, “tiny” beads¹⁷ were introduced. However, the parametrization of the cross-interactions between small and tiny beads and regular beads was performed on an *ad hoc* basis. As we have shown in previous work,¹¹ this *ad hoc* parametrization in concert with deviations from the standard bonded parameters used in the original Martini parametrization⁸ results in small but systematic deviations in packing and intermolecular interactions.¹¹ Such deviations are one of the main reasons for the observed overestimated interactions between Martini models of proteins and sugars,^{14,18} and the necessity for the development of custom beads for polymers and other molecules which use 3-to-1 or 2-to-1 mappings.^{19,20} Given the ubiquitous role of ring-like structures across the chemical space, addressing such shortcomings was important for applications ranging from protein-ligand binding²¹ to imidazolium-based ionic liquids²² and conjugated molecular materials.²³ To address these shortcomings, a new version of the Martini model has been recently developed, Martini 3: with a re-balanced set of nonbonded interactions and expanded variety of CG particle types, the new version of the force field shows improved accuracy overall in predicting molecular packing and interactions.²⁴

In light of the recent development of Martini 3,²⁴ we here describe how small molecules can be accurately parametrized within the Martini 3 framework and present a database of validated small molecule models. We pay particular attention to the description of aliphatic and aromatic ring-like structures, given their ubiquity in small molecules such as solvents and

drugs or in building blocks constituting macromolecules such as proteins and synthetic polymers. Such ring-like structures are described by models that use small and tiny CG particle types. As such, the present database of small molecules constitutes one of the cornerstones of the calibration of the new Martini 3 small and tiny particle sizes. We comprehensively describe how to parametrize small molecules, describe which properties have been considered during the parametrization, present design strategies to build complex small molecules, show what can be achieved with the new models, and discuss expected strengths and potential weaknesses of the new models. With this work, we initiate a database of Martini 3 models for small molecules (currently 90 entries) that includes Martini 3 models, atomistic reference structures, atomistic-to-CG (and vice versa) mapping files, and reference and computed properties (available at <https://github.com/ricalessandri/Martini3-small-molecules> and <http://cgmartini.nl>).

This paper is structured as follows. In Section 2, we start by summarizing the main features of Martini 3 and we subsequently describe in detail how to model small molecules' aromatic and aliphatic core elements (Section 2.1.1), substituents and functional groups (Section 2.1.2), and molecules of increased complexity (Section 2.1.3); we moreover describe the database (Section 2.2) and methods used (Section 2.3). In Section 3, we describe how the models perform in terms of partitioning behavior between water and apolar phases (Section 3.1), solvent properties such as mass density and heat of vaporization (Section 3.2), miscibility (Section 3.3), and stacking interactions (Section 3.4). Lastly, in Section 3.5, we concisely summarize the key points, limitations, and envisioned applications of the models.

2 Model and Methods

2.1 Coarse-Graining Small Molecules within Martini 3

2.1.1 Basics

Nonbonded Interactions: Martini 3 in a Nutshell. For a thorough description of the Martini 3 force field, we refer to the main publication.²⁴ Briefly, as in the previous version of the force field, the main parametrization target remains the free energy of transfer between several pairs of solvents, such as hexadecane/water, octanol/water, and chloroform/water. Comparison to infinite-dilution properties such as transfer free energies allows to probe mainly solvent-solute cross-interaction, albeit solvent-solvent self-interaction is also probed for the considered solvents. Moreover, in Martini 3, miscibility data constitute a second core parametrization target: depending on the available reference data, miscibility can be checked by either qualitatively looking at the miscibility of two phases or by computing excess free energies of mixing, as it will also be shown in this contribution. Comparison to miscibility data allows to probe at the same time not only the cross-interaction between the two molecular species but also the relative self-interaction of the two species.

Molecular fragments are used as building blocks (beads), whose nonbonded interactions are described by modified Lennard-Jones (LJ) interactions. To describe the chemical nature of the underlying atomistic structure, CG beads with more or less polar character exists. As in Martini 2, there are four main types of beads: polar (P), intermediately polar (N), apolar (C), and charged (Q). Moreover, a separate water (W) bead, a bead for divalent ions (D), and a set of beads (X) dedicated to groups containing halogen atoms are now included. The P-, N-, C-, Q-, and X-bead types are in turn divided in subtypes based on their hydrogen-bonding capabilities, degree of polarity, or charge softness.²⁴ To keep the model simple, the levels of interactions (LJ ε) are discretized. In Martini 3, the number of possible interaction levels was extended from 10 to 22.

Similarly to the LJ ε , also the sizes of the beads (LJ σ) are discretized. There are

three bead sizes—regular (R), small (S), and tiny (T), with σ of 0.47, 0.41, and 0.34 nm, respectively—and cross-interactions across sizes (R-S, R-T, S-T) are specifically parametrized. In particular, the three available bead sizes are now fully balanced, with regular, small, and tiny beads intended to be used for 4-to-1, 3-to-1, and 2-to-1 atoms-to-bead mappings, respectively. As such, small and tiny beads are not only appropriate to represent ring-like fragments, but also suited for cases involving 3-to-1 and 2-to-1 mapping of linear and branched chemical groups. Moreover, when fully branched fragments (*e.g.*, quaternary carbon atoms or tertiary amine groups) need to be described, one should usually use beads of a smaller size, such as an S-bead for a fully branched fragment containing four non-hydrogen atoms. The rationale for this is that the central atom of a fully branched fragment is not exposed to the environment, and this reduces its influence on the environment. A thorough description of Martini 3 is presented in the Supporting Information of Ref. 24.

Bonded Interactions: Center-of-Geometry Mapping. In contrast to Martini 2, which relied on using the center of mass (COM) of the mapped chemical groups to define the CG bonded parameters of a molecule, Martini 3 follows a “size-shape concept” aimed at preserving the volume and shape of molecules with respect to the underlying atomistic reference structures.²⁴ In general, bond lengths obtained from center-of-mass (COM) mapping proved to be unsatisfactory, usually leading to too high packing densities. This is exemplified for benzene in Figure S1 and it is not unexpected: the original Martini 2.0 benzene model already uses bond lengths of 0.27 nm, ~25% longer than bond lengths extracted following COM mapping.⁸ The inadequacy of COM-mapped bond lengths is due to the geometrical mismatch enforced by the lower resolution of the CG representation (*e.g.*, triangular versus hexagonal shape of benzene) and by not taking into account the volume of hydrogen atoms. Instead, center-of-geometry (COG)-based mapping performed *taking into account the hydrogen atoms* leads to better molecular (*e.g.*, solvent accessible surface area) and bulk (*e.g.*, mass densities) properties (Section 3.2). COG-mapping is especially important the more the mapping is close to atomistic resolution, that is, when using T-beads. COM and COG

are indeed almost equivalent in the case of a 4-to-1 mapped alkane chain (see Figure S2). In contrast, they provide considerably different bond lengths in the case of 2-to-1 mapped rings, as just discussed for benzene. COG-based (like COM-based) bond lengths can be directly extracted from AA models: such a direct link to the underlying atomistic structure is important, as it, for example, makes them suitable for automated topology building.^{12,25–27} Hence, using COG-based mapping of atomistic structures constitute the standard procedure for obtaining bonded parameters in Martini 3; such parameters can then, if necessary, be refined, as discussed in more detail in Section 3.2.

Aromatic Rings. Aromatic (*i.e.*, planar, atom-thick) structures are best described by tiny beads. The prototypical aromatic compound, benzene, is described by a three-bead model reminiscent of its Martini 2 model⁸ (see also Figure 1): each of the three beads represents two consecutive carbon atoms and their corresponding hydrogen atoms. However, there are a few important differences. First, the model now uses T-beads rather than S-beads. In this way, the six non-hydrogen atoms which constitute benzene are described by three beads which have been parametrized specifically for 2-to-1 mappings. Moreover, the radius of T-particles is such that stacking distances between Martini models of aromatic structures agree with atomistic and/or experimental stacking distances (see also Section 3.4). For benzene, or more generally non-substituted $-\text{C}=\text{C}-$ groups part of aromatic compounds, the Martini bead type of choice is TC5, that leads to free energies of transfer in excellent agreement with experimental data in the literature (see Section 3.1); therefore benzene is represented by a TC5-TC5-TC5 model. A COG-based bond length of 0.29 nm is used (see also Figure S1), which leads to a mass density for liquid benzene of 0.890 g cm^{-3} , in excellent agreement ($\approx 2\%$ too high) with the experimental value of 0.877 g cm^{-3} .

Aromatic ring models are held together⁸ by constraints.²⁸ Distributions of bond lengths obtained from mapped AA simulations between beads which are part of aromatic structures are indeed very narrow (Figure S1c). As a consequence, they would require very stiff potentials, and such potentials would in turn demand very short time steps. However, at the

Martini scale the fast stretching of these bonds becomes unimportant, and such stiff bonds can be replaced by constraints, *i.e.*, those bond lengths become constants. More extended stiff molecular structures can be conveniently described by using virtual sites, as discussed in detail in Section 2.1.3.

The parameters of benzene form the basis for all the aromatic ring-like structures, some of which are shown in Figure 1. Together with the other topologies presented in this work, this set should guide the development of new Martini 3 models for aromatic molecules and molecular fragments.

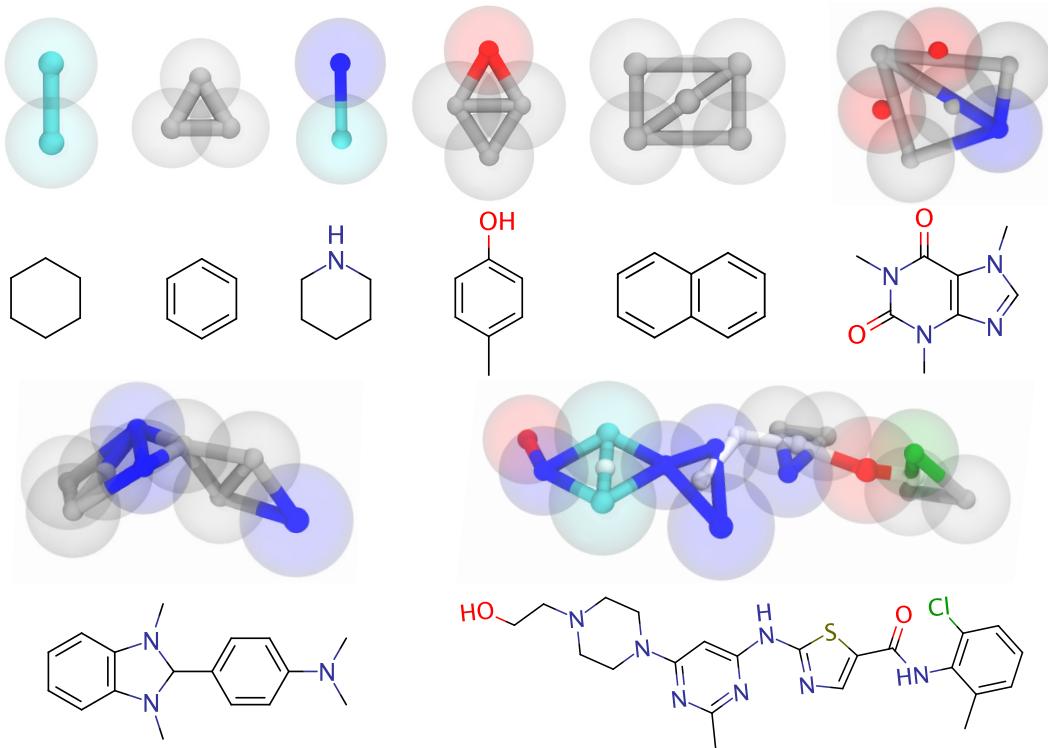


Figure 1: Examples of mappings of small molecules of increasing complexity. 2D chemical structures are shown along with renderings of their corresponding Martini 3 CG models. From left to right: (first row) cyclohexane, benzene, piperidine, *p*-cresol, naphthalene, and caffeine; (second row) the molecular dopant N-DMBI, the drug dasatinib. CG beads are color-coded as follows: apolar aliphatic (C1-C3 type) beads are in cyan, apolar aromatic (C4-C6) beads are in silver, intermediately polar (N1 to N6) beads in blue, polar (P1 to P6) beads in red, halogen-containing (X1 to X4) beads in green, and dummy beads in white.

Aliphatic Rings. Aliphatic ring-like structures are described by small beads. The prototypical aliphatic cyclic compound, cyclohexane, is described by a two-S-bead model,

differentiating it from the Martini 2 model.⁸ The model has been devised to capture the larger bulkiness of cyclohexane, as compared to its aromatic counterpart, *i.e.*, benzene. The beads used to describe cyclohexane are of type SC3 based on partitioning data (see Section 3.1). Bonds keep together aliphatic models, as bond distributions are broader in the case of aliphatic rings (compare Figure S1c and Figure S1f). A bond length of 0.378 nm is used, leading to a mass density for liquid cyclohexane of 0.793 g cm⁻³, in excellent agreement ($\approx 2\%$ too high) with the experimental value of 0.774 g cm⁻³.

Analogously to benzene, cyclohexane constitutes the basis for the aliphatic ring-like structures. Accordingly, the present pool of aliphatic compounds should serve as reference for the development of new Martini 3 models of aliphatic molecules and molecular fragments. Concluding this section we also note that, although the CG models do not preserve the symmetries of both the prototypical aromatic (benzene) and aliphatic (cyclohexane) compounds, they do increasingly capture the size, shape, and symmetries of increasingly larger molecules.

2.1.2 Mapping Small Molecules

Table 1: Small molecule building blocks. CG particle type, corresponding chemical building block, and examples of molecules in which such a block appears. The atoms the CG block is taken to represent are also shown in the 2D chemical structures, with T-beads, S-beads, and R-beads depicted in blue, green, and red, respectively.

[†] Indicates that the (group of) atom(s) is shared with neighboring beads.

^{*} Indicates molecules considered in this work; others are from Ref. 24.

type	chemical building block	examples	
		2D	name (mapping)
P2	CH ₃ CH ₂ —COOH		propanoic acid (P2)
SP2	CH ₃ —COOH		acetic acid (SP2)
P1	(CH ₃) ₂ CH—OH		isopropanol (P1)

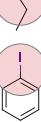
Table 1: (continued)

TP1d	$-\text{CH}_2\text{OH}$ (bonded to $-\text{CF}_3$)		2,2,2-trifluoroethanol (SX4e-TP1d)
SN6	$-\text{CH}^\dagger=\text{C}(-\text{OH})-\text{CH}^\dagger=$ $-\text{CH}_2^\dagger-\text{CH}(-\text{OH})-\text{CH}_2^\dagger-$		*phenol ((TC5) ₂ -SN6) *cyclopentanol (SN6-SC3)
SN6d	$-\text{CH}^\dagger=\text{C}(-\text{NH}_2)-\text{CH}^\dagger=$		*aniline ((TC5) ₂ -SN6d)
TN6d	$=\text{CH}^\dagger-\text{NH}-\text{CH}^\dagger=$		*pyrrole ((TC5) ₂ -TN6d)
TN6	$=\text{C}(-\text{OH})-$		* <i>p</i> -cresol ((TC5) ₂ -TC5-TN6)
N6a	$-(\text{CH}_2)_2-\text{CH}(=\text{O})$		heptanal (C1-N6a)
TN6a	$-\text{CH}=\text{N}-$		*pyridine ((TC5) ₂ -TN6a)
SN5a	$-\text{O}-\text{CH}_2-\text{O}-$		*1,3-dioxolane (SN5a-TC3)
TN5a	$=\text{C}=\text{O}$		*benzoquinone (TN5a-(TC5) ₂ -TN5a)
N4a	$-\text{C}(=\text{O})-\text{O}-\text{CH}_3$ $\text{CH}_3-\text{C}(=\text{O})-\text{O}-\text{CH}_3$		*methyl-benzoate (N4a-(TC5) ₃) methyl-acetate (N4a)
SN4a	$-\text{CH}_2-\text{O}-\text{CH}_2-$ $-\text{C}(=\text{O})-\text{CH}_3$		*tetrahydropyran (SC3-SN4a) *acetophenone (SN4a-(TC5) ₃)
TN4a	$-\text{CH}_2^\dagger-\text{O}-\text{CH}_2^\dagger-$ $-\text{C}\equiv\text{N}$ $-\text{CH}(=\text{O})$		*tetrahydrofuran (SC3-TN4a) *benzonitrile ((TC5) ₃ -TN4a) *benzaldehyde ((TC5) ₃ -TN4a)
SN3a	$-\text{CH}_2-\text{O}-\text{CH}_2-$		*1,4-dioxane (SN3a-SN3a)

Table 1: (continued)

SN3a	$-\text{NO}_2$		*nitrobenzene ((TC5) ₃ -SN3a)
SN2a	$=\text{C}(-\text{O}-\text{CH}_3)-$		* <i>o</i> -methylanisole (SN2a-TC4-(TC5) ₂)
TN2a	$=\text{CH}^\dagger-\text{O}-\text{CH}^\dagger=$ $-\text{O}-\text{CH}_3$		*furan ((TC5) ₂ -TN2a) *methoxybenzene ((TC5) ₃ -TN2a)
SN1	$-\text{N}(-\text{CH}_3)_2$		*N,N-dimethylaniline ((TC5) ₃ -SN1)
TN1	$-\text{N}(-\text{CH}_3)-$		*1-methylimidazole (TC5-TN1-TN6a)
SC6	$-\text{CH}^\dagger=\text{C}(-\text{SH})-\text{CH}^\dagger=$ $-\text{CH}_2-\text{S}-\text{CH}_2-$		*thiophenol ((TC5) ₂ -SC6) *tetrahydrothiophene (SC6-TC3)
TC6	$=\text{CH}^\dagger-\text{S}-\text{CH}^\dagger=$		*thiophene ((TC5) ₂ -TC6)
TC5	$-\text{CH}=\text{CH}-$ $-\text{CH}=\text{CH}_2$		*benzene ((TC5) ₃) *styrene ((TC5) ₄)
TC5e	$(-)_2\text{C}=\text{C}(-)_2$		*naphthalene ((TC5) ₂ -TC5e-(TC5) ₂)
C4	$\text{CH}_3-\text{CH}=\text{CH}-\text{CH}_3$		1-butene (C4)
SC4	$-\text{CH}^\dagger=\text{C}(-\text{CH}_3)-\text{CH}^\dagger=$		*toluene ((TC5) ₂ -SC4)
TC4	$=\text{C}(-\text{CH}_3)-$		* <i>o</i> -methylanisole (SN2a-TC4-(TC5) ₂)
SX4e	$-\text{CF}_3$ $-\text{CF}_3$		*benzotrifluoride ((TC5) ₃ -SX4e) 2,2,2-trifluoroethanol (SX4e-TP1d)
SC3	$-(\text{CH}_2)_3-$ $-\text{CH}_2^\dagger-\text{CH}(-\text{CH}_3)-\text{CH}_2^\dagger-$		*cyclohexane (SC3-SC3) *methylcyclopentane (SC3-SC3)

Table 1: (continued)

TC3	$-\text{CH}_2-\text{CH}_2-$		*tetrahydrothiophene (SC6-TC3)
SX3	$-\text{CH}^\dagger=\text{C}(-\text{Cl})-\text{CH}^\dagger=$		*chlorobenzene ((TC5) ₂ -SX3)
SC2	$-(\text{CH}_2)_2-(\text{CH}_3)$		*3-propyl-thiophene (TC6-(TC5) ₂ -SC2)
X2	CHCl_3		*trichloromethane (X2)
SX2	$-\text{CH}^\dagger=\text{C}(-\text{Br})-\text{CH}^\dagger=$		*bromobenzene ((TC5) ₂ -SX2)
C1	$\text{CH}_3-(\text{CH}_2)_2-\text{CH}_3$		butane (C1)
X1	$-\text{CH}^\dagger=\text{C}(-\text{I})-\text{CH}^\dagger=$		*iodobenzene ((TC5) ₂ -X1)

Bead Type Choice. While benzene and cyclohexane form the basis of the, respectively, aromatic and aliphatic ring-like structures which appear in small molecules, most of the interesting compounds also contain heteroatoms either within the ring structures themselves or as substituents. Table 1 lists a subset of the small and tiny Martini particle types and the corresponding chemical building block which they describe in Martini 3 models, along with examples of small molecules which employ that particle type. Note that in some cases a (group of) atom(s) is shared between two neighboring beads: this is indicated with the superscript \dagger in Table 1 and throughout this manuscript. Atom-sharing between beads is usually necessary in order to keep the symmetry of the underlying atomistic structure—see, for example, phenol, tetrahydrofuran, or toluene in Table 1. Note that atom-sharing is important also because it needs to be taken into account when extracting bonded interactions for the CG models from COG-mapped atomistic simulations.

Beads are chosen based on partitioning data, the Martini hallmark.⁹ More bead types are now available within Martini 3 with respect to Martini 2, thus, for example, the three groups $-\text{N}(-\text{CH}_3)-$, $=\text{CH}^\dagger-\text{O}-\text{CH}^\dagger=$, and $-\text{CH}_2^\dagger-\text{O}-\text{CH}_2^\dagger-$ (found in 1-methylimidazole, fu-

ran, and tetrahydrofuran, respectively—see also Table 1) described in Martini 2 by a SN0 bead are now described by a TN1, TN2a, and TN4a bead, respectively. Table 1, listing mapping of CG particle types to chemical building blocks, serves as an additional guide towards the assignment of CG particle types. More bead assignments are exemplified in Tables 24 and 25 of the Supporting Information of Ref. 24.

Besides expanding the applicability of the hydrogen bonding “labels”, or interaction modifiers, already available in Martini 2, Martini 3 introduced also *electron polarizability* labels.²⁴ These labels mimic changes to the interactions of a molecular fragment due to inductive/conjugate effects caused by neighboring chemical groups. Accordingly, C- and X-beads can acquire an electron-donor (or ‘enriched’, label ‘e’) or electron-acceptor (‘vacancy’, label ‘v’) character. Such labels were shown to nicely capture preferential orientations in aedamers.²⁴ Moreover, such labels could be used to capture halogen-bonding capabilities of X-beads, as such interactions have been recently shown to be important in, for example, membrane-small molecule interactions.²⁹ The ‘e’ label is also applied to, for example, the central bead describing naphthalene—which is hence a TC5e bead, see Table 1—being that an electron-rich region of the molecule.

Mapping Substituted Ring-like Fragments. Figure 2 shows how to map substituted ring-like fragments as a function of substituent number and size. Such mappings are based on the following two principles: 1) map all the non-hydrogen atoms with the minimum possible number of beads and 2) preserve the symmetry, volume, and shape of the molecule as much as possible, with aromatic rings being best described by T-beads and aliphatic rings by S-beads. Principle 1) alone would, for example, demand a model for benzene made by either 2 S-beads or 1 R-bead and 1 T-bead, as such models would use the minimum possible number of beads (given the 6 non-hydrogen atoms of benzene that need to be described). However, principle 2) tells us to disregard both of these models, since the molecular volume and shape of benzene are best described by T-beads. Hence, a three-T-bead model is the best choice for benzene.

The same principles can guide the mapping of substituted benzene shown in Figure 2. For example, when adding a methyl group to benzene to obtain toluene (top row of Figure 2), one of the three beads used to describe the benzene plane become a larger S-bead to account for the extra carbon atom. When instead adding an ethyl group as substituent on benzene—to obtain ethyl-benzene—an extra T-bead should be used, as shown in Figure 2 (top row). Such a choice allows to keep a three-T-bead model for the benzene moiety, which is the most accurate model for benzene in Martini 3. At the same time, the two extra carbon atoms of the ethyl group can be modelled by a T-bead, which we recall are parametrized specifically for 2-to-1 mappings.

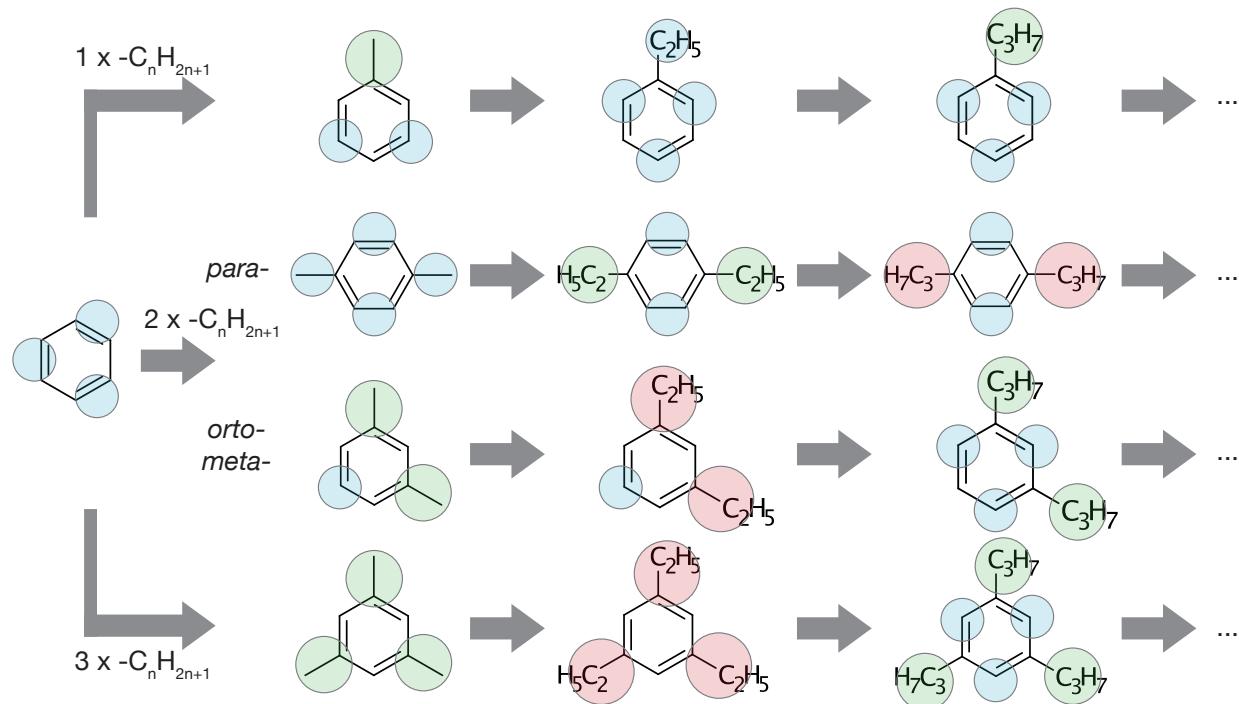


Figure 2: Recommended mappings for substituted aromatic small molecules. Mono- (top), di- (center), and tri-substituted (bottom) rings are shown. In the case of di-substituted rings, we distinguish between *para*- and *meta*- or *ortho*- substitutions, the latter two to be treated identically in terms of mapping. Small differences in bonded parameters may arise when extracting them from COG-mapped atomistic simulations. Regular, small, and tiny beads are indicated in red, green, and blue, respectively, but not drawn to scale.

2.1.3 Advanced Model Design Strategies

Building More Complex Models. We now describe a number of design strategies useful to build more complex small molecule models, that is, models comprising more than one aromatic/aliphatic ring-like structure. Necessarily, the strategies described do not represent a comprehensive list, nor are they the only way to build a model for a certain molecular construction within the Martini framework. However, we found them to be very helpful when building Martini models for more complex small molecules, and we thus describe them here with the aim of enabling a more facile construction of Martini models by the Martini user community. The strategies can be used in combination and can be adapted to different chemistries than the ones described in the examples. Furthermore, they form an initial set of “rules” which may be implemented in the next generation of programs aimed at automated topology building of Martini models.^{12,26}

Many of the model design strategies outlined below leverage virtual (interaction) sites.³⁰ In general, the use of virtual sites improves the numerical stability of Martini models of small molecules which are comprised of multiple, or extended, aromatic structures. As mentioned earlier, Martini aromatic ring structures are held together by constraints. As more extended stiff molecular structures need to be described (*e.g.*, in polycyclic aromatic compounds), the number of constraints that need to be used in their Martini representations grows. However, a highly extended network of constraints leads to numerical instabilities, as it is increasingly complicated to satisfy a growing number of connected constraints (we note that we use GROMACS³¹ and the LINCS²⁸ constraint solver). Saving on the number of constraints by using virtual sites not only increases the numerical stability of the simulations but also improves their performance.

Construction 1: “Hinge” Model. The first design strategy is useful to build models for rigid, fused polycyclic compounds and can be illustrated by the proposed model for the molecule of naphthalene (NAPH)—Figure 3a. The molecule consists of 10 aromatic carbon atoms, which translates into a 5-TC5-bead Martini model. Simply interconnecting

the 5 beads with constraints leads to a network of 8 constraints (see Figure S3a). Such a network of constraints already starts to be hard to satisfy: whereas one molecule in vacuum is still numerically stable with a time step of 20 fs, a condensed phase of such a model (432 molecules in a periodic box) readily leads to numerical instabilities with time steps as low as 10 fs, thus requiring even smaller time steps. This makes large-scale simulations impracticable. The idea to overcome this limitation is inspired by the “hinge” construction used in the latest cholesterol model proposed by Melo and co-workers.¹³ Indeed, another model which could be devised for naphtalene exploits the “hinge” construction for the 4 external beads (Figure 3a and Figure S3b), while the central bead is described as a virtual interaction site, *i.e.*, a particle whose position is completely defined by its constructing particles (in this case, the other four beads). The displacement of the virtual site itself is not calculated by the integrator algorithm, but its position is recalculated from the new positions of the constructing particles after each integration step.³¹ Of course, forces on the virtual interaction site are accounted for—these are distributed to the constructing atoms in each integration step. Virtual sites are also mass-less, thus the mass of the bead they represent must be accounted for; in the NAPH case, it is evenly added to the masses of the constructing particles. Due to the reduction on the number of constraints used (from 8 to 5), the use of the hinge model also accelerates the calculation: in the case of the same NAPH condensed phase, the simulation runs \sim 1.5 times faster. Note also that an improper dihedral is applied in order to keep the hinge model flat (see the comparison of the improper dihedral distribution to AA reference data in Figure S3c).

To get even more speed-up and get rid of the improper dihedral potential, one could think of reducing further the number of particles to three—the minimal amount of particles needed to define the NAPH plane—and then construct the remaining two as virtual interaction sites. However, this would make NAPH an absolutely rigid body. When torque is generated at one end of the model, the consequent rigid-body rotation can cause a very large displacement on the opposite end, possibly resulting in unphysical overlap with other system components.¹³

The hinge model, on the other hand, works as a shock absorber,¹³ preventing such an absolute rigid body behavior. It is thus recommended to opt for such a non-rigid body model. This strategy was used to develop models for other extended ring structures, such as caffeine and polyacenes (see Figure S3d-f). Note that in these other cases, more than one virtual site per hinge is defined.

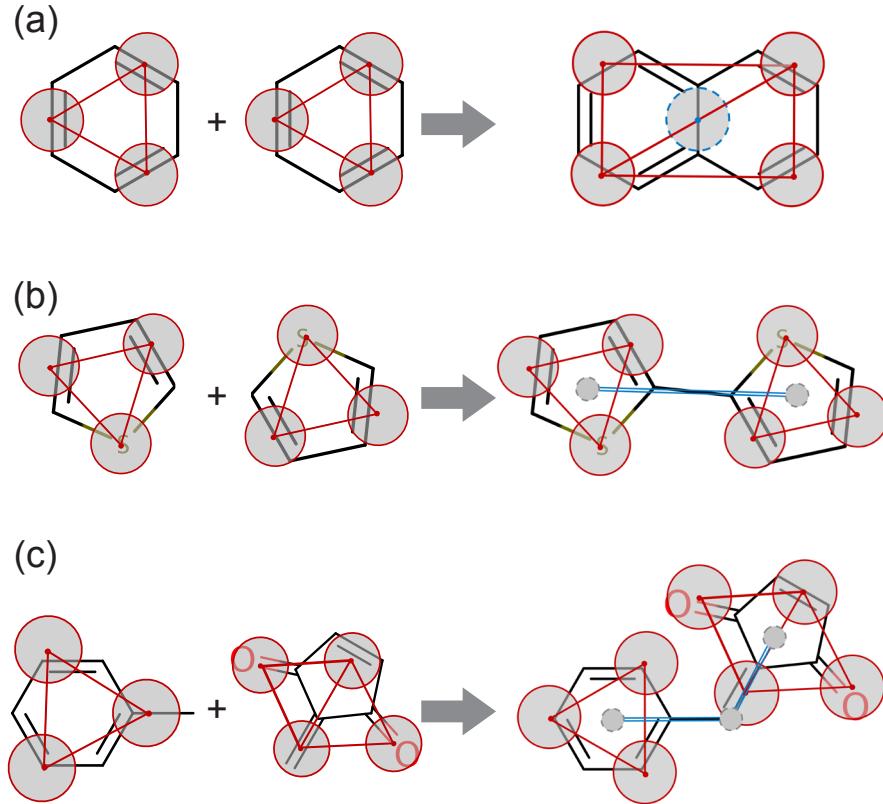


Figure 3: Advanced design strategies to build Martini models for small molecules of increasing complexity. (a) Two fused rings can be modeled with a “hinge” construction, as exemplified for naphthalene. (b) Two (or more) rigid planar fragments, whose relative dihedral angle can be controlled, can be modeled via a “divide and conquer” strategy; this is exemplified for 2,2'-bithiophene. (c) Ring systems linked via a sp^2 hybridized carbon, such as in 2-(phenylmethylidene)cyclopent-4-ene-1,3-dione, represent a “molecular turn” that can be modelled with a specific construction. Schematic representations of the CG models are drawn on top of the chemical structures: a gray circle with solid red contour indicates a CG particle, a gray circle with a dashed cyan contour represents a virtual interaction site, and a smaller gray circle with a dashed gray contour represents a virtual dummy site. Solid red lines indicate constraints, while double blue lines indicate harmonic bonds.

Construction 2: “Divide and Conquer”. A second design strategy is useful to build arbitrarily long chains of rigid planar fragments whose relative dihedral angle can be controlled; it is illustrated by the proposed model for 2,2'-bithiophene (2T)—Figure 3b. This construction is relevant not only for small molecules but also for conjugated polymers.³² Indeed, this second strategy, dubbed “*divide and conquer*”, is inspired by the polymer backbone construction used in Martini polythiophene models.^{32,33} The two thiophene rings are described by three T-beads each. To connect them in 2,2'-bithiophene while allowing for control of the dihedral angle between the two thiophene planes, we use two virtual *dummy* sites at the center of geometry of the thiophenes. Note that by dummy we mean a site that does not interact via nonbonded interactions; this type of virtual site is to be contrasted with the virtual *interaction* site of the previous construction which instead does interact as a standard CG particle. The two dummy sites are then connected by an harmonic bond. In this way, we can apply a dihedral potential $i\text{-}D_i\text{-}D_j\text{-}j$, where D_i and D_j are the dummy sites on the two thiophenes and i and j are two particles on either of the thiophenes; for i and j , it is convenient to take the beads describing the sulphur atoms. Such dihedral represents the torsion between the thiophene rings, and a dihedral potential obtained from quantum mechanical calculation can be directly implemented in such a construction if nonbonded interactions are excluded between the two thiophenes. The definition of this dihedral is also reminiscent of the dummy-assisted dihedral described by Bulacu et al. in the context of increasing the numerical stability of CG models.³⁴ Indeed, by construction this dihedral prevents bead collinearities which lead to numerical instabilities,³⁴ hence providing a numerically robust way of connecting rigid planar fragments whose relative dihedral angle can be controlled.

Construction 3: a “Molecular Turn”. Ring-systems linked through an sp^2 hybridized carbon represent one of the more complicated groups of small molecules to coarse-grain within the Martini framework. As exemplified by the 2-(phenylmethylidene)cyclopent-4-ene-1,3-dione molecule (PMCD, Figure 3c), this type of coupling induces a “molecular turn” in the

link between the two ring-systems, which calls for special attention to preserve the rotation axis for the torsional motion of the two rings relative to each other. Also here, virtual dummy sites come in handy. By defining such a virtual dummy site in the COG of each of the ring-systems and a third virtual dummy site directly on top of the linking sp² hybridized carbon, it is possible to allow the correct torsion around the single bond despite not having an actual CG bead at the rotation axis. The phenyl ring and the dione-substituted cyclopentene are, respectively, mapped using the standard benzene construction and a hinge-type construction similar to the one described above (cf. Figure 3a); the virtual dummy site on the linking sp² hybridized carbon is constructed from the two oxo group beads and the methyldene bead (using [`virtualsites3`] in GROMACS). The connection to the phenyl group is taken care of by using a harmonic bond between the COG virtual dummy site of the phenyl group and the virtual dummy site of the linker, and the angle between the two ring-systems is conserved by a harmonic angle potential between the three virtual sites. This means that the torsion of the phenyl group relative to the rest of the molecule can be described by a proper dihedral potential centered on the COG virtual dummy site of the phenyl group and the virtual dummy site of the linker. Note that it is also necessary to add an improper dihedral potential to keep the phenyl CG beads in plane with the virtual dummy site of the linker as well as one keeping the COG virtual dummy site of the phenyl group in plane with the CG beads of the dione-substituted cyclopentene. With this construction, it is thus possible to maintain the correct rotation axes between ring-systems linked through an sp² hybridized carbon, which is relevant, for example, for several small molecules used in organic electronics.³⁵

2.2 Database Description

A database containing 90 small molecules (see Table S2 for the full list of molecules), is available at <https://github.com/ricalessandri/Martini3-small-molecules> and on the Martini portal <http://cgmartini.nl>. The database comprises small molecules containing

a range of mono- and polycyclic aromatic and aliphatic structures, and covering most of the chemical functional groups, thus including a wide range of (bio)molecular building blocks. However, it is necessarily far from being complete. The models included serve as a guide to build new models, or as building blocks to be used in macromolecules such as proteins or synthetic polymers. For example, the side chains of the four aromatic amino acids—4-methylimidazole (histidine), *p*-cresol (tyrosine), toluene (phenylalanine), and 3-methyl-1H-indole (tryptophan)—have been developed as part of this database and are used in the Martini 3 proteins.²⁴ Similarly, the database contains models for monomers of synthetic polymers, such as 3-hexylthiophene, which when polymerized—which can be done in an automated fashion by the Polyply³⁶ program—constitutes the conjugated polymer poly(3-hexylthiophene).

For every molecule, a two to five letter long string is used as unique identifier (uID) and 8 files are included, namely:

1. `uID.smiles`: the SMILES string representing the molecule;
2. `uID.itp`: the Martini 3 model topology, in GROMACS format;
3. `uID.gro`: the Martini 3 model coordinates, in GROMACS format;
4. `uID_LigParGen.pdb`: atomistic structure in PDB format as obtained from the LigParGen server^{37–39} by entering the corresponding SMILES;
5. `uID_LigParGen.itp`: OPLS-AA/1.14*CM1A-LBCC topology, in GROMACS format, as obtained from the LigParGen server^{37–39} by entering the corresponding SMILES;
6. `uID_oplsaaTOcg_cgbuilder.ndx`: the AA-to-CG index file;
7. `uID_cgTOoplsaa_cgbuilder.map`: the CG-to-AA mapping file;
8. `uID_LigParGen_cgbuilder.png`: a screenshot showing the beads and underlying atomistic structure.

Mapping and index files were generated using CGBuilder (<https://jbarnoud.github.io/cgbuilder/>), and then modified manually if necessary, for example, to introduce atom-sharing which is not handled by CGBuilder. Table S2 lists the uID, common name, and SMILES string for all the molecules in the database.

Reference and computed properties, needed to reproduce the plots shown in this paper, are available at <https://github.com/ricalessandri/Martini3-small-molecules>.

2.3 Simulation Settings and Procedure Details

General Simulation Settings. Settings for the CG simulations adhere to the “new” set of Martini run parameters,⁴⁰ using a time step of 20 fs. Specifically, the Verlet neighbor search algorithm was used to update the neighbor list, with a straight cutoff of 1.1 nm. The Parrinello–Rahman barostat⁴¹ (coupling parameter of 12.0 ps) and the velocity-rescaling thermostat⁴² (coupling parameter of 1.0 ps) were used to maintain pressure and temperature, respectively. CG simulation setting files are available on the Martini portal <http://cgmartini.nl> and at <https://github.com/ricalessandri/Martini3-small-molecules>. A unique set of atomistic run parameters was used for the AA simulations, using a time step of 2 fs. The Verlet neighbor search algorithm was employed to update the neighbor list; a 1.4 nm cutoff for LJ and for Coulomb (reaction-field) interactions was employed. The Parrinello–Rahman barostat⁴¹ (coupling parameter of 5.0 ps) and the Nosé–Hoover thermostat^{43,44} (coupling parameter of 1.0 ps) were used to maintain pressure and temperature, respectively. All simulations were run with GROMACS³¹ 2016.x or 2019.x.

All-Atom Models. For every molecule in the database, a OPLS-AA/1.14*CM1A-LBCC model was obtained from the LigParGen server^{37–39} by entering the SMILES string of the small molecule; for each molecule, a coordinate file (pdb format) and a topology file in GROMACS format (itp) were obtained. Additional all-atom models for a selected number of small molecules (for example the ones for which free energy surfaces of dimerization were computed, see Section 3.4) were obtained with either the GROMOS 54A7⁴⁵ or

the CHARMM General FF (CGenFF)^{46,47} force fields. GROMOS 54A7⁴⁵ all-atom models were obtained, in GROMACS format, from the ATB server.^{48,49} CHARMM General FF (CGenFF)^{46,47} models were obtained by uploading a molecule (in `.mol2` format) to <https://cgenff.umaryland.edu/> to obtain a `.str` file; subsequently, to obtain a topology in GROMACS format, both files were passed to `cgenff_charmm2gmx.py` (using the July2017 version, `charmm36-jul2017.ff.tgz`, both downloaded from http://mackerell.umaryland.edu/charmm_ff.shtml). In the remainder of the article, we will refer to the AA force fiels simply as OPLS, GROMOS, and CHARMM.

Liquid and Gas Phase Simulations. A liquid phase was approximated as an equilibrated box of dimensions of about $5 \times 5 \times 5 \text{ nm}^3$; a gas phase as a single molecule occupying a large ($7 \times 7 \times 7 \text{ nm}^3$) simulation box. Liquid phase simulations were performed in the NPT ensemble at 298 K and 1 bar, while gas phase ones in the NVT ensemble at 298 K. The enthalpy of vaporization (ΔH_{vap}) was computed as:

$$\Delta H_{\text{vap}} \approx U_{\text{gas}} - U_{\text{liq}} + RT \quad (1)$$

where U_{gas} and U_{liq} are the total energies (per mole) of the gas and liquid phase, respectively. Densities were extracted with the GROMACS tool `gmx density`. For this analysis, we used, for each bead, the mass of the constituting atoms rather than the typical bead masses that were used for the simulation. By default Martini 3, like Martini 2, uses standard masses for regular, small, and tiny beads of 72, 54, and 36 Da, respectively.

Free Energies of Transfer Calculations. We used alchemical transformations to compute free energies of solvation $\Delta G_{S \rightarrow \emptyset}$ in a solvent S . ΔG_S were computed between water (W) and a number of organic phases, namely hexadecane (HD), (hydrated) octanol (OCO), and chloroform (CLF). In Martini 3, these are described by a W, C1-C1-C1-C1, SC2-SC2-SP1, and X2 model,²⁴ respectively. Note that we simulate hydrated octanol, that is we add a ≈ 0.3 mole fraction of water according to experimental conditions.⁵⁰ A series of 21

simulations with equally spaced λ points going from 0 to 1 were performed using a stochastic integrator. Simulations were equilibrated for 2 ns and each λ point was run for 10 ns. A soft-core potential (with α of 0.5 and power set to 1) was used to avoid the singularity in the potential when interactions were switched off. The free energies and corresponding errors were finally computed using the Multistate Bennett Acceptance Ratio (MBAR).⁵¹ The free energy associated with transferring a solute from a solvent S_1 to a solvent S_2 ($\Delta G_{S_1 \rightarrow S_2}$) was computed as the difference $\Delta G_{S_1 \rightarrow \emptyset} - \Delta G_{S_2 \rightarrow \emptyset}$.

Dimerization Free Energy Surface Calculations. Free energy surface (FES) profiles for the dimerization of each molecule were computed by either umbrella sampling (US)⁵² or (well-tempered⁵³) metadynamics (MTD)⁵⁴ simulations. In US simulations, the two solute molecules were placed in a box and solvated. Umbrella windows were spaced 0.1 nm apart along the collective variable (CV), this being the distance between the COGs of the two solute molecules. For each window, this distance was kept fixed by an umbrella potential with a force constant of 1500 kJ mol⁻¹ nm⁻². Each window was simulated for at least 150 and 500 ns in the case of AA and CG systems, respectively. The FES profiles were calculated using the weighted histogram analysis method (WHAM)⁵⁵ as implemented in the GROMACS tool `gmx wham`. In MTD simulations, the bias was added on both the distance between the COGs of the two small molecules and the torsional angle formed by atoms of the small molecules and their COGs in the case of AA simulations. In the CG simulations, the efficiency of Martini allowed us to use only the distance between the COG of the molecules as a CV and obtain converged profiles. Simulations were 100-150 ns or 1 μ s long in the AA and CG case, respectively. The order parameter S of Eq. 4, which describes the relative orientation of the ring planes, was used as a third CV to project the FES by using a reweighting algorithm.⁵⁶ The height of the deposited Gaussians was set to 1.0 kJ mol⁻¹ and the width to 0.05 nm and 0.2 rad for the distance and torsion, respectively. Gaussians were deposited every 500 steps and the bias factor was set to 5. A wall, in the form of a harmonic potential with a force constant of 200 kJ mol⁻¹, was added at a distance

beyond 2 nm to prevent the molecules from exploring conformations at distances irrelevant to binding. Block analysis was used to estimate the error from the free energy calculations (<https://plumed.github.io/doc-v2.4/user-doc/html/trieste-4.html>). Both US and MTD FES need to be entropy corrected and shifted so that the free energy at large distances (1.7-2.0 nm) is very close to zero. A stochastic integrator was employed in both US and MTD simulations. All simulations were performed with GROMACS 2016.x,³¹ patched with PLUMED 2.4⁵⁷ in the case of the MTD simulations.

Binary Mixture Simulations. An equal number, namely 500, of molecules of species *A* and *B* was placed randomly in a simulation box of dimensions $9 \times 9 \times 9 \text{ nm}^3$. The box was then energy minimized, simulated at a higher pressure (100 bar) for 1 ns and then subsequently relaxed for 1 ns in the NPT ensemble at 1 bar and 300 K. Simulations were then run in the same ensemble for at least 400 ns. Miscibility was monitored following the number of *A*-*B* contacts with the GROMACS tool `gmx mindist`, with a cutoff distance of 0.6 nm, which comprises the nearest neighbor CG sites around a CG particle. Typical evolutions of the number of *A*-*B* contacts as a function of simulation time for mixtures with miscible and immiscible components are shown in Figure S6a.

Vapor–Liquid Equilibrium Simulations. Vapor–liquid equilibrium simulations were performed by setting up systems of dimensions $9 \times 9 \times Z \text{ nm}^3$ where *Z* ranges from ≈ 20 to ≈ 30 nm depending on the mixture; at least 7 nm of the *Z* dimension is empty—constituting the starting configuration for the vapor phase. The remaining of the box contains a 50:50 mixture of two components—the starting configuration for the liquid phase. A rendering of this setup is shown in Figure 6c. Such starting simulation boxes were set up for all the compositions possible from $x_A = 0$ to $x_A = 1$ in steps of 0.05 where x_A is the molar fraction of component *A* in the mixture. The vapor–liquid binary systems were then simulated for at least 2 μs at each x_A in the NVT ensemble at 300 K. This leads to statistically reliable vapor and liquid densities which can be extracted with the GROMACS tool `gmx density`, as described more in detail in the Supporting Information. The equilibrium densities thus

obtained were then used to compute the $\Delta G_{\text{solv},i}$ of the i -th component in the mixture according to Equation 3 as described in Section 3.3.

3 Results and Discussion

3.1 Partitioning Behavior: Free Energies of Transfer

We first examine the performance of the new models in reproducing experimental free energies of transfer, as partitioning between different solvents constitutes the main target of the Martini force field parametrization. Transfer free energies of all the molecules in the database have been thus computed and compared to available experimental data^{58–63} in order to settle on consistent mappings for all the molecules in the database. Experimental data are mainly available for hexadecane→water (HD→W), (hydrated) octanol→water (OCO→W) and chloroform→water (CLF→W) partitioning^{58–63} (see Section 2.3 for the solvent models). Figure 4 displays the performance of the final models with respect to partitioning by showing computed *vs* experimental free energies in these three cases. The agreement is excellent in all three cases, with a mean absolute error of 1.8, 1.8, and 2.2 kJ mol⁻¹, for HD→W, OCO→W, and CLF→W, respectively.

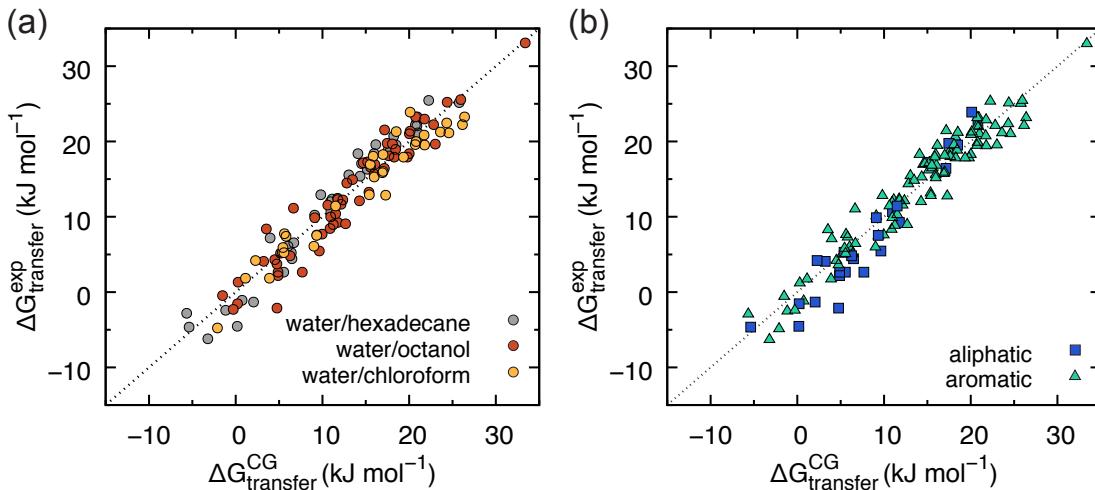


Figure 4: Partitioning behavior of Martini 3 small molecules. In (a), computed data are plotted against experimental data in the cases of hexadecane→water (gray circles), octanol→water (red circles), and chloroform→water (gold circles). In (b), the same data are plotted by discriminating aliphatic (blue squares) and aromatic (green triangles) small molecules. Experimental data have been compiled from Refs. 58–63. Statistical uncertainty for the computed ΔG is below 0.3 kJ mol⁻¹ and thus not visible in the plots.

3.2 Solvent Properties: Enthalpies of Vaporization and Densities

We now examine the performance of the small molecule models in reproducing mass densities and enthalpies of vaporization (ΔH_{vap}), a secondary parametrization target of the Martini 3 force field. We computed mass density and ΔH_{vap} for the CG models presented here, and compared them to experimental values,⁶⁴ where available; the results are shown in Figure 5, where the CG values are computed as described in Section 2.3. In the case of the enthalpies of vaporization (Figure 5a-b), we also report the values for simpler, linear Martini 3 molecules,²⁴ in order to show the overall trend of ΔH_{vap} of the Martini 3 force field. Indeed, we remark that Martini enthalpies of vaporization are notoriously systematically underestimated due to the limited fluid range of the employed 12-6 Lennard-Jones potential form.⁸ Normalizing the data by dividing by the ΔH_{vap} of water leads to the correlation plot presented in Figure 5b. Moreover, we recall¹¹ that Martini 2 ring-like models deviate from the overall force field trend and lead to systematically higher heat of vaporizations. This deviation was related to the “overmapping” of such molecules and the imperfect S-bead calibration.¹¹ This is now overcome in Martini 3, where all the small-molecules containing ring-like structures lie in the general trend of the Martini force field.

Computed mass densities for the solvents are overall in good agreement with experimental densities: the mean absolute percentage error is 5.8 %. This is shown in Figure 5c, where CG *vs* experimental densities are plotted. Unsurprisingly, not only the LJ parameters but also the bond lengths used in the models have a large impact on the final densities of the simulated solvents. As anticipated earlier, bond lengths obtained from COG-mapped structures in general lead to good molecular and bulk properties. For example, such bond lengths allow for a good reproduction of the solvent accessible surface area (SASA) of the AA model, as shown in Figure S5b. This also translates into good agreement in case of mass densities (see Figure S5a). However, given the lower resolution of the model, perfectly matching densities cannot be obtained for all cases simply by following a COG-based mapping scheme. Thus, bond lengths can be refined further if higher accuracy is required. This fine-tuning was done

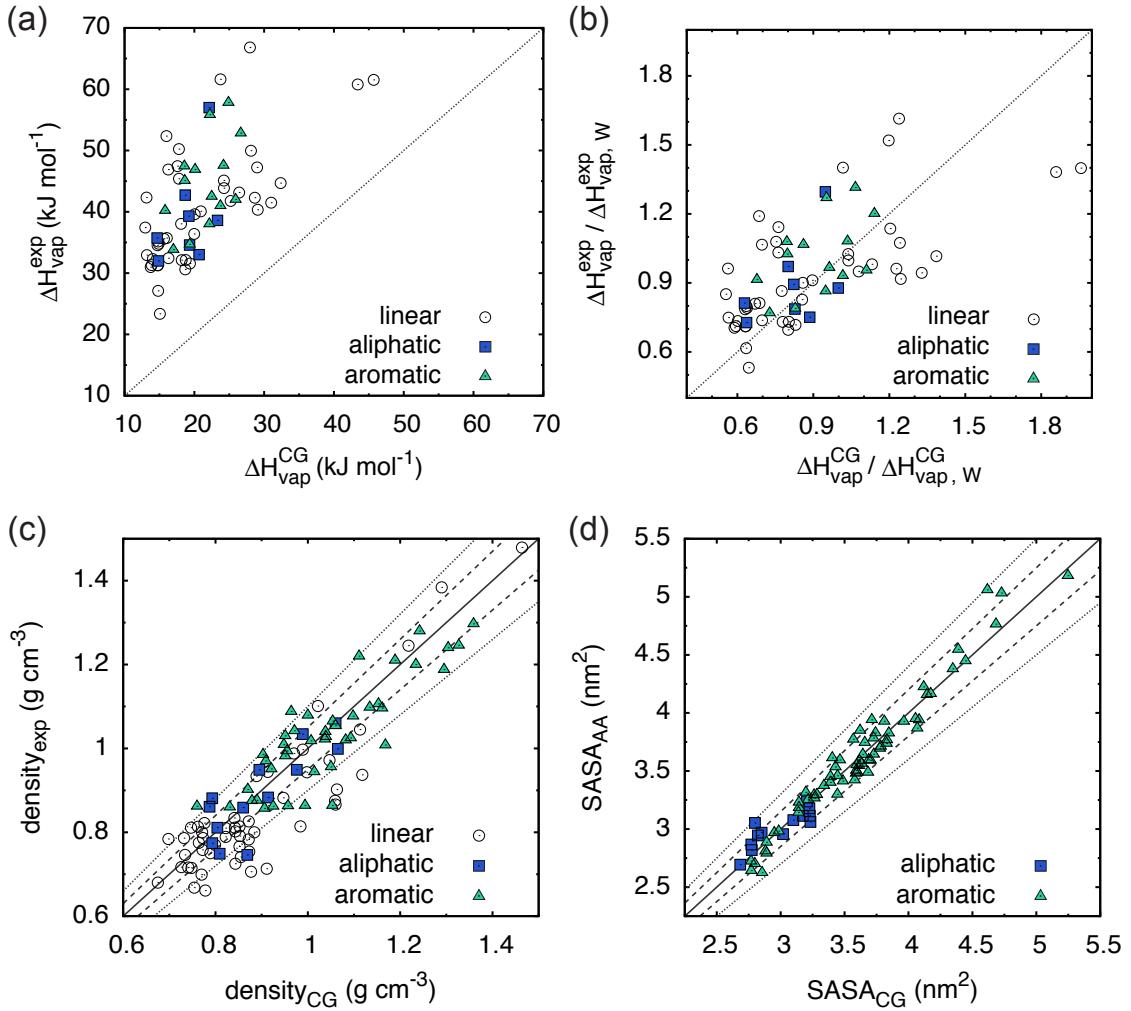


Figure 5: Solvent properties of Martini 3 small molecules: (a)-(b) enthalpy of vaporization (ΔH_{vap}), (c) mass density, and (d) solvent accessible surface area (SASA). Data points corresponding to aromatic and aliphatic compounds are indicated by green triangles and blue squares, respectively. In (a)-(c), also linear molecules from Ref. 24 are reported to show the general trend of the Martini 3 force field. In (c) and (d), the dashed and the dotted lines indicated threshold error lines of $\pm 5\%$ and $\pm 10\%$, respectively. Experimental mass densities and enthalpies of vaporization are from Ref. 64. SASA reference values have been obtained with the OPLS all-atom force field. For details on the SASA calculations, see the Supporting Information.

for the models where the mismatch between the experimental and simulated densities was the highest. Thus, some of the topologies contained in the database contain “optimized” bond lengths, which lead to correlation plot for the mass densities reported in Figure 5c—mean absolute percentage error of 7.2 %—and for the SASA values reported in Figure 5d.

3.3 Solvent Mixtures: Martini Miscibility Table

We tested the capabilities of the models further by examining its performance in reproducing mixing behaviors of binary mixtures. Some force fields have been parametrized based on experimental data on binary mixtures, such as vapor-liquid equilibria (*e.g.*, TraPPE),⁶⁵ while established force fields have been tested for their performance in reproducing such equilibria⁶⁶ or other related thermodynamic properties of mixtures (*e.g.*, the free energy of mixing).⁶⁷ We performed mixing assays to estimate whether two Martini solvent models mixed or not, and compared the observed behavior to reference experimental data. Data regarding the most common solvents are commonly available as solvent miscibility tables. We selected the main aromatic or aliphatic ring-like small molecules which usually appear in such tables. The extracted experimental miscibility table is shown in Figure 6a; results for Martini 3 are in agreement with experimental data, so Figure 6a also represent the Martini miscibility table for ring-like small molecules (and water). In the simulations, the miscibility was monitored following the number of contacts between molecules *A* and molecules *B* of a equimolar *A:B* mixture. Starting from a random mixture of the two, where each species makes a certain number of contacts with the other, monitoring the number of contacts allows for detection of phase separation (see also Methods). Immiscible phases readily demix (sharp decrease in the number of *A-B* contacts), while for miscible phases the average number of contacts remains constant (see Figure S6a for typical evolution of number of *A-B* contacts for mixing and demixing binary mixtures).

A more quantitative estimation of how the new Martini model performs with respect to solvent mixtures has then been carried out for the benzene-cyclohexane mixture. We

computed the free energy of mixing (ΔG_{mix}) as a function of the composition of the mixture. Considering a mixture of A and B , let x_A (x_B) be the mole fraction of A (B) in the system, the (molar) ΔG_{mix} is given by:⁶⁸

$$\Delta G_{\text{mix}} = \Delta G_{\text{ex}} + T\Delta S_{\text{id}} = [x_A(\mu_A - \mu_A^*) + x_B(\mu_B - \mu_B^*)] + RT(x_A \ln(x_A) + x_B \ln(x_B)) \quad (2)$$

where μ_A (μ_B) is the chemical potential of component A (B) in the mixture, and μ_A^* (μ_B^*) the chemical potential of component A (B) in its pure state. In this expression, the chemical potentials represent the reversible work to solvate the two components (A or B) in the mixture (μ_A or μ_B) or in their own liquid state (μ_A^* or μ_B^*). As such, they correspond to the solvation free energies of the two components in the mixture and in their own liquid state. The first term in Equation 2 can be referred to as the excess free energy of mixing (ΔG_{ex}), and has to do with the change in interactions experienced by A and B in the mixture with respect to their pure states. In the case of an ideal mixture, where the interactions between A (B) molecules are as strong as the A - B interactions, this term vanishes (and the experimental manifestation of it is Raoult's law). The second term in Equation 2 is an entropic contribution (S_{id}) due to the increased number of states available upon mixing (sometimes referred to as the *ideal* mixing free energy given the fact that it is the only term which remains in the case of ideal mixtures).

The chemical potentials which appear in Equation 2, *i.e.*, the solvation free energies of the two components in the mixture and in their own liquid state, can be computed directly from the equilibrium vapor and liquid densities of a binary liquid-gas system. Following the scheme of Ben-Naim,⁶⁸ the free energy of solvation *in the mixture* ΔG_{solv} of the i -th component is given by:

$$\Delta G_{\text{solv},i} = RT \ln \left(\frac{\rho_{\text{vap},i}}{\rho_{\text{liq},i}} \right) \quad (3)$$

where $\rho_{\text{vap},i}$ and $\rho_{\text{liq},i}$ are the vapor and liquid densities, respectively, of the i -th component in the mixture at the given composition. We obtained the equilibrium densities *via* "direct"

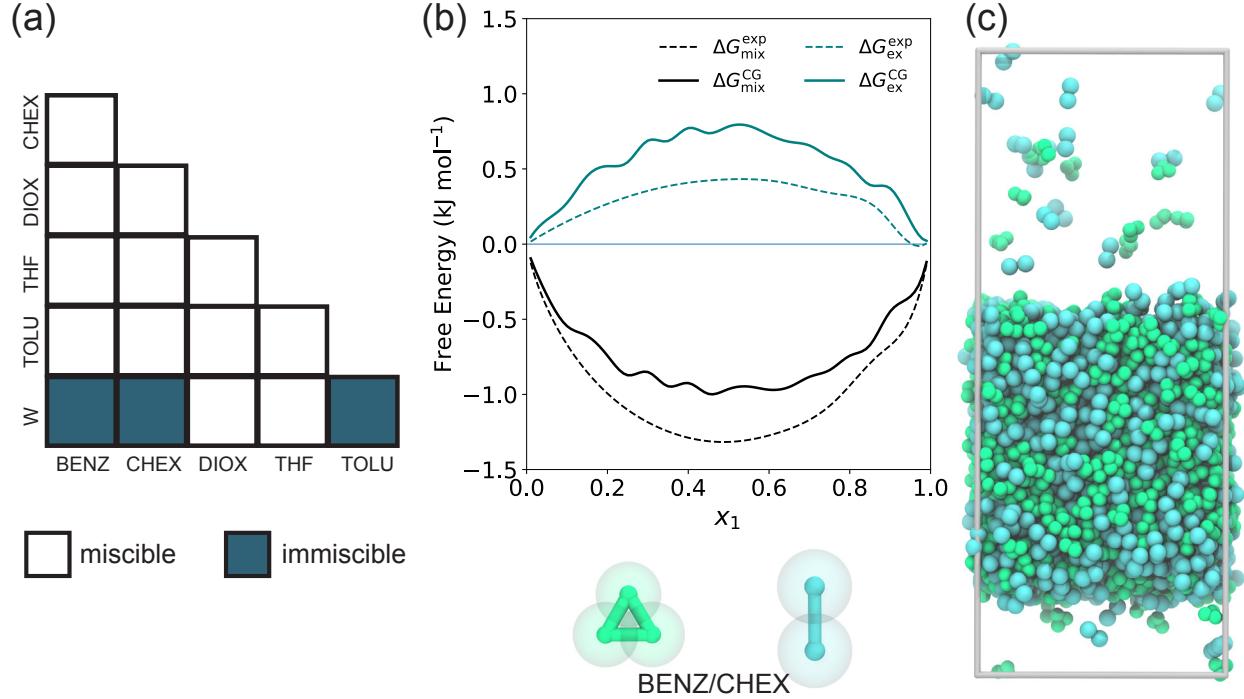


Figure 6: Miscibility of aromatic or aliphatic small molecules (and water): Martini 3 *vs* experiments. (a) The experimental solvent miscibility table is shown. Light color means miscible, dark means immiscible. The same mixing behavior is observed for Martini 3, so Martini fully reproduces the experimental miscibility table. Molecules are indicated by their uID: BENZ (benzene), CHEX (cyclohexane), DIOX (1,4-dioxane), THF (tetrahydrofuran), TOLU (toluene), and W (water). (b) Comparison between Martini and experimental free energy of mixing (ΔG_{mix}) and excess free energy of mixing (ΔG_{ex}) for the benzene-cyclohexane mixture as a function of the mixture composition. x_1 is the molar fraction of benzene. Experimental data are obtained from vapor-liquid equilibrium data.⁶⁹ Simulation data are obtained from vapor-liquid equilibrium simulations (see Methods). A rendering of the setup is shown in (c); small and tiny bead radii are rendered in scale.

simulations, *i.e.*, by simulating equilibrated binary liquid-vapor systems (see Figure 6c and Methods). Such simulations can be very easily extended in the microsecond range with the Martini model, enough to obtain statistically reliable results on the equilibrium vapor and liquid densities. We did this for the benzene-cyclohexane mixture along the whole range of compositions—from pure cyclohexane ($x_1 = 0$) to pure benzene ($x_1 = 1$) in steps of 0.05 in x_1 . Figure 6b shows the ΔG_{mix} and ΔG_{ex} computed according to Equation 2. The agreement between simulated and experimental data is good. Experimentally, benzene-cyclohexane deviates positively from Raoult’s law, showing less than ideal mixing. This feature is captured at the Martini level. The simulated ΔG_{ex} is somewhat more positive than the experimental one over the whole composition range. It measures 0.78 kJ mol⁻¹ for an equimolar mixture ($x_A = 0.5$), to be compared to the experimental 0.43 kJ mol⁻¹. The benzene-chloroform mixture has also been investigated with the same approach, and results are shown in Figure S6c. Again, the agreement between experiments and CG is good, when taking into account that the difference between the ΔG_{ex} for the equimolar mixture is around 0.3 kJ mol⁻¹. We propose that vapor–liquid equilibrium data can be used to further validate the miscibility of phases relevant to a certain application of the Martini force field.

Given that evidence of constraints leading to temperature gradient and miscibility artifacts in ternary lipid mixtures has been recently presented,⁷⁰ and that many of the (aromatic) small molecules rely heavily on constraints, we tested the impact of the LINCS²⁸ constraint settings on mixtures. Namely, we investigated the impact on mixtures of small molecules by comparing the default LINCS settings (`lincs-order = 4`; `lincs-iter = 1`; time step of 20 fs) versus more stringent LINCS settings (`lincs-order = 8`; `lincs-iter = 2`) or a smaller time step (15 fs). As shown in Figure S4, we found no significant impact on three representative mixtures—benzene-cyclohexane, benzene-ethanol, and benzene-water—neither on the mixing (as quantified by the number of contacts) nor on eventual differences in temperatures between the molecules involved. However, for larger molecules with multiple constraints (such as cholesterol), we recommend the user to test these settings.

3.4 Stacking Interactions: Dimerization Free Energy Landscapes

We now turn to describe the performance of the model by investigating local structural properties, *i.e.*, properties which have to do with the “local” structure, such as molecular stacking or packing. In particular, we analyzed stacking behaviors for several aromatic rings, given the ubiquitous role played by aromatic structures in the self-assembly of soft matter. Figure 7 shows potentials of mean force (PMFs) of dimerization in water for several representative aromatic compounds—at the Martini 3, Martini 2, and atomistic levels—along the distance between the COGs of the two molecules. PMFs have been obtained either *via* Umbrella Sampling, or Metadynamics, as described in detail in Methods. In particular, we analyze the following aromatic compounds: (a)–(c) (poly)cyclic aromatic hydrocarbons, (d) a heterocyclic 5-membered ring, and (e) mono-, and (f) di-substituted aromatic compounds.

Overall, the depth of the mimima in Martini 3 is in good agreement with the atomistic reference data, showing improvements with respect to Martini 2—in particular for the more hydrophobic compounds, such as benzene and thiophene. Moreover, desolvation barriers—the free energy barriers located between the first and the second minimum of each PMF—are reduced in Martini 3 as compared to Martini 2. This reduction is more in line with the atomistic force fields. The improvement is due to the introduction of LJ cross-interaction sizes, as opposed to Martini 2 where these were absent.^{8,11} Note that, in most of the cases, we compare to reference AA PMFs obtained with more than one atomistic force field. A second look at the data shows that, going from Figure 7a to Figure 7c, that is, upon increasing the size of the aromatic system, the discrepancy between the overall shape of the AA and Martini 3 PMFs decreases. This can be explained by comparing the two-dimensional free energy profiles of Figure 7g–7h (AA) and Figure 7j–7k (Martini 3). The free energy is now plotted on the two-dimensional coordinate space formed by the distance between the COGs of the two molecules and the order parameter indicating the relative orientation of the planes of the two aromatic molecules:

$$S = \frac{(3 \cos^2 \theta - 1)}{2} \quad (4)$$

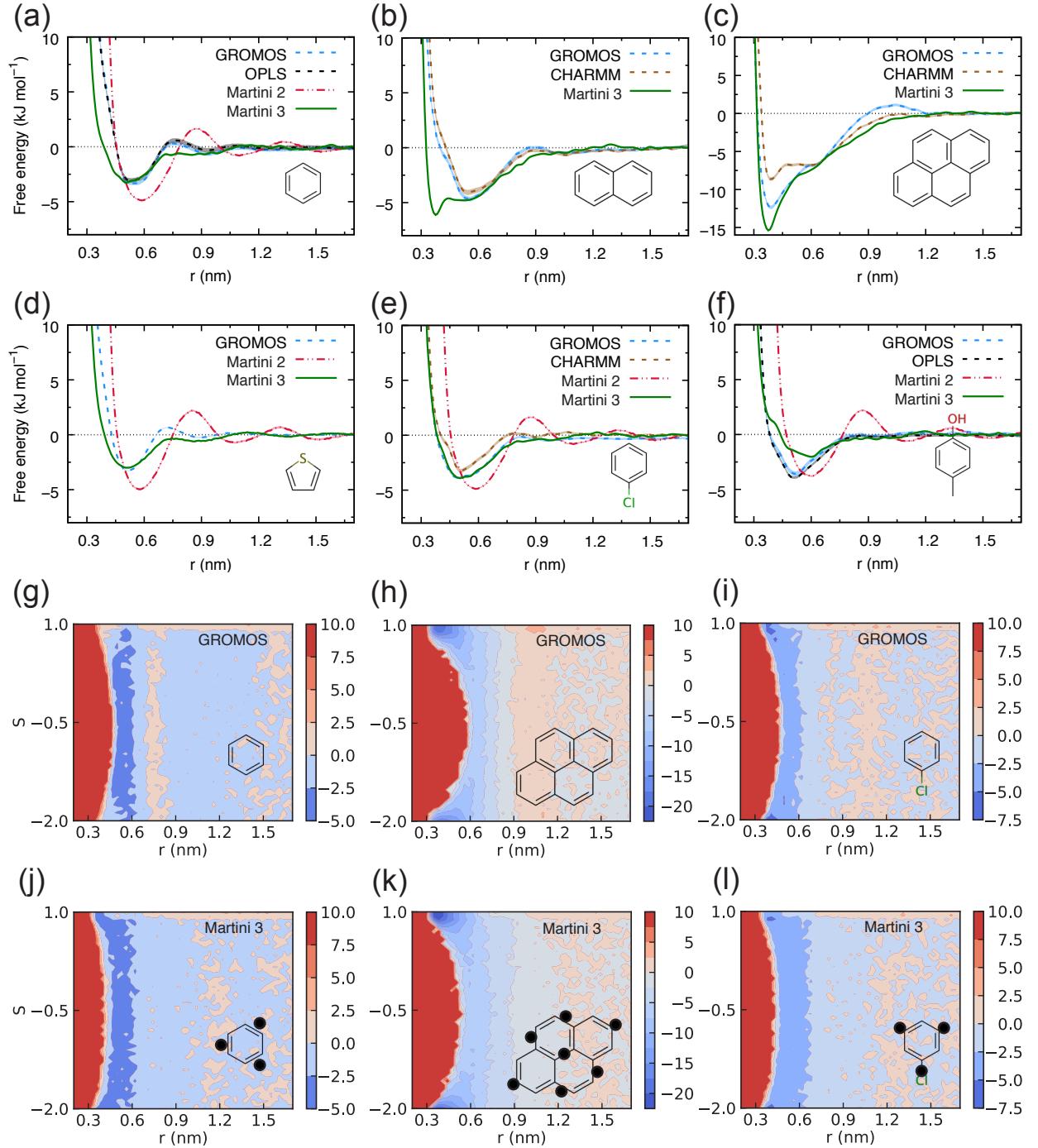


Figure 7: One- and two-dimensional free energy surfaces of dimerization for several aromatic small molecules in water. (a)–(f) The free energy is plotted along the distance between the COGs of the two molecules. The profiles obtained with Martini 3 (green solid lines) are shown along with the Martini 2 (red dash-dotted lines) and the atomistic (GROMOS,⁴⁵ OPLS,³⁷ or CHARMM,^{46,47} dashed lines) ones; (a)–(c) show how the dimerization profiles change upon increasing the molecular size of (poly)cyclic aromatic hydrocarbons (benzene (a), naphthalene (b), and pyrene (c)); (d) shows the profile for the 5-membered heterocyclic aromatic compound thiophene; (e)–(f) show profiles for substituted aromatic compounds which possess a permanent dipole moment (chlorobenzene (e), and *p*-cresol (f)). (g)–(l) The free energy surface is plotted on the 2D coordinate space formed by the distance used in (a)–(f) and the order parameter of Eq. 4; (g)–(i) are all-atom, and (j)–(l) are Martini 3 surfaces. Molecular structures and mappings are shown as figure insets.

In this equation, θ is the angle between the two vectors perpendicular to the planes of the two aromatic molecules. A value of S close to 1 ($\theta = 0^\circ$) or -2 ($\theta = 180^\circ$) indicates that the two aromatic molecules are perfectly stacked, *i.e.*, they are in a *sandwich* conformation. In contrast, S values close to -0.5 ($\theta = 90^\circ$) denote a *T-shaped* conformation.

Figure 7g and 7h show such 2D surfaces for benzene and pyrene, respectively. It can be seen that in the case of the smaller benzene, a T-shaped dimer conformation is preferred at the AA level over the sandwich conformation. As the size of the molecule increases, there is a shift from the T-shaped minimum of benzene (and naphthalene, see Figure S7) to the sandwich minimum of pyrene. This behavior is not captured by the Martini models; instead, Martini predicts the two conformations to be practically equivalent in terms of free energy for benzene, while it predicts decidedly the sandwich conformation to be the global minimum for the larger naphthalene and pyrene molecules. The T-shaped minimum of benzene is known to be a consequence of the quadrupole moment driving the interaction between the two molecules (given the null dipole moment). The quadrupole interaction appears to be the dominant driving force also in the case of naphthalene, as the sandwich conformation does not appear to be a minimum of the free energy landscape (Figure S7b). However, as the size of the molecule increases, the “cavity cost” prevails over the quadrupolar interaction, leading to a sandwich interaction which maximizes the water-water contacts. Standard Martini 3 models for benzene and naphthalene will not be able to capture this effect, which is driven by the specific distribution of charges which is absent in the models, and only accounted for effectively within the LJ parameters. In contrast, the “hydrophobic effect” is intrinsically captured by Martini. However, when the dominant interaction is not a quadrupolar one, Martini 3 models are now expected to reproduce experimental stacking distances—as can be seen in Figure 7c for pyrene—such as the one of conjugated polymers which were previously systematically off due to the larger size of S-beads.³²

The fact that Martini favors sandwich conformations with respect to T-shaped ones also explains the better agreement of the free energy profiles of chlorobenzene (CLBZ) and

p-cresol (PCRE) of Figure 7e and 7f. Indeed, the 2D free energy landscape of Figure 7i (CLBZ) show the dominance of the sandwich conformation in molecules with a permanent dipole moment such as CLBZ (and PCRE, Figure S7a). This is again captured by Martini 3 (Figure 7k and S7c).

3.5 Possible Applications and Limitations

The present database, along with the outlined model design strategies and the modularity of the Martini force field, open up possibilities to more easily model complex (bio)molecular systems. On the biomolecular side, a natural application of the small molecules presented here combined with the Martini protein models²⁴ is to investigate protein-ligand interactions in unbiased or biased MD simulations.^{21,71} The parametrization guidelines and strategies are expected to facilitate the building of databases with accurate Martini 3 models of small molecules; this possibly will require the help of the next-generation of programs aimed at automated topology building of Martini models,^{12,26} and automated parametrization of bonded parameters.^{25,27} On the materials science side, Martini 3 and the presented parametrization guidelines are expected to open up the possibility of more easily model organic functional materials in general. For example, modeling of organic semiconductors—the key components of organic solar cells, organic transistors, and organic bioelectronic devices to name but a few—and their morphology is an area of particular richness and in need of computational insights at the length scales which can be reached with Martini.^{10,72} Organic semiconductors are rich in complex conjugated structures which can now be accurately modeled with Martini 3 and the advanced model design strategies described here.³⁵

Despite the numerous improvements and even wider applications opened up by Martini 3, limitations of this coarse-graining approach must be kept in mind. Certain limitations apply generally to the Martini force field,^{8,24,73} and include a non-quantitative agreement with experimental free energies of solvation and narrower fluid ranges due to the use of the 12-6 Lennard-Jones potential, and an entropy-enthalpy imbalance—due to enthalpy com-

pensating for the loss of degrees of freedom upon coarse-graining—which is known to affect the temperature dependence of several properties.^{10,24} Other limitations pertain more specifically to the systems subject of the present study, namely small-molecules containing ring-like structures described by finer mappings. As shown in Section 3.4, conformations driven by quadrupolar interactions, such as T-shaped stacking, are not captured by standard Martini models due to the lack of specific electrostatic interactions. Care must be taken thus if such quadrupolar interactions are expected to be driving a particular self-assembly process. The inclusion of partial charges may remedy to this, and this is now possible in Martini 3 thanks to the “*q*” labels.²⁴ Exploring such an avenue is a natural possibility now opened up by the increased flexibility of Martini 3. Additional model refinement is also made possible in Martini 3 by the introduction of the so-called *self-interaction* labels. These more generic labels decrease or increase the self-interaction of a certain bead type without changing its free energy of transfer. Accordingly, they allow for quick model refinement and, together with the wider bead type selection and the other labels, are expected to greatly reduce the need for the development of ad-hoc beads. Finally, although some model design strategies have been formulated in the present work, building models for extended planar (polymeric) systems can be challenging and time-consuming, as the extended networks of constraints often required by such structures easily lead to numerical instabilities. To fully harness the capabilities of Martini 3 CG modeling of (small) molecules, such strategies need to be implemented in the next-generation of programs aimed at automated topology building of Martini models.^{12,26}

In conclusion, we presented how small molecules can be accurately parametrized within the Martini 3 framework, and initiated a database of validated small molecule models. The present database, along with the outlined model design strategies and the modularity of the Martini force field, constitute a resource of models that 1) can be used “as is” in (bio)molecular simulations; 2) gives reference points for the construction of other small molecule models; 3) provides guidelines and reference data for automated topology builders;

and 4) can be used as building blocks for the construction of more complex (macro)molecules, hence enabling investigations of complex biomolecular^{21,71} and soft material systems.^{10,72}

Acknowledgement

R.A. thanks the Dutch Research Council (NWO) (Graduate Programme Advanced Materials, No. 022.005.006) for financial support, and Maria Tsanai for preliminary tests on Martini 3 numerically unstable, T-bead-rich systems. J.B. acknowledges funding from the TOP grant from S.J.M. (NWO) and the EPSRC program grant no. EP/P021123/1. This work was partly carried out on the Dutch national e-infrastructure with the support of SURF Cooperative. A.S.G.'s work has been performed under the Project HPC-EUROPA3 (INFRAIA-2016-1-730897), with the support of the EC Research Innovation Action under the H2020 Programme; in particular, A.S.G. gratefully acknowledges the support of S.J.M. at the University of Groningen and the computer resources and technical support provided by SURFsara.

Supporting Information Available

Comparison between COG- and COM-based mappings; more examples and further details on the advanced design strategies; tests of LINCS constraint settings on mixtures; further details on procedures and extended results; list of the uID and corresponding common name and SMILES string for all the molecules in the database.

This material is available free of charge via the Internet at <http://pubs.acs.org/>.

References

- (1) Klein, M. L.; Shinoda, W. Large-Scale Molecular Dynamics Simulations of Self-Assembling Systems. *Science* **2008**, *321*, 798–800.

- (2) Peter, C.; Kremer, K. Multiscale simulation of soft matter systems – from the atomistic to the coarse-grained level and back. *Soft Matter* **2009**, *5*, 4357–4366.
- (3) de Pablo, J. J. Coarse-Grained Simulations of Macromolecules: From DNA to Nanocomposites. *Annu. Rev. Phys. Chem.* **2011**, *62*, 555–574.
- (4) Noid, W. G. Perspective: coarse-grained models for biomolecular systems. *J. Chem. Phys.* **2013**, *139*, 090901.
- (5) Saunders, M. G.; Voth, G. A. Coarse-Graining Methods for Computational Biology. *Ann. Rev. Biophys.* **2013**, *42*, 73–93.
- (6) Ingólfsson, H. I.; López, C. A.; Uusitalo, J. J.; de Jong, D. H.; Gopal, S. M.; Periole, X.; Marrink, S. J. The Power of Coarse Graining in Biomolecular Simulations. *WIREs Comput. Mol. Sci.* **2014**, *4*, 225–248.
- (7) Voth, G. A. *Coarse-graining of condensed phase and biomolecular systems*; CRC press, 2008.
- (8) 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.
- (9) Marrink, S. J.; Tieleman, D. P. Perspective on the Martini Model. *Chem. Soc. Rev.* **2013**, *42*, 6801–6822.
- (10) Alessandri, R.; Grünwald, F.; Marrink, S. J. The Martini Model in Materials Science. *Adv. Mater.* **2021**, *33*, 2008635.
- (11) 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.

- (12) 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.
- (13) 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*, 243152.
- (14) Schmalhorst, P. S.; Deluweit, F.; Scherrers, R.; Heisenberg, C.-P.; Sikora, M. Overcoming the Limitations of the MARTINI Force Field in Simulations of Polysaccharides. *J. Chem. Theory Comput.* **2017**, *13*, 5039–5053.
- (15) Genheden, S. Solvation free energies and partition coefficients with the coarse-grained and hybrid all-atom/coarse-grained MARTINI models. *J. Comput. Aided Mol. Des.* **2017**, *31*, 867–876.
- (16) 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.
- (17) Uusitalo, J. J.; Ingólfsson, H. I.; Akhshi, P.; Tielemans, D. P.; Marrink, S. J. Martini Coarse-Grained Force Field: extension to DNA. *J. Chem. Theory Comput.* **2015**, *11*, 3932–3945.
- (18) Stark, A. C.; Andrews, C. T.; Elcock, A. H. Toward Optimized Potential Functions for Protein-Protein Interactions in Aqueous Solutions: osmotic Second Virial Coefficient Calculations Using the MARTINI Coarse-Grained Force Field. *J. Chem. Theory Comput.* **2013**, *9*, 4176–4185.
- (19) 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.

- (20) Monticelli, L. On Atomistic and Coarse-Grained Models for C₆₀ Fullerene. *J. Chem. Theory Comput.* **2012**, *8*, 1370–1378.
- (21) 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*, 3714.
- (22) Vazquez-Salazar, L. I.; Selle, M.; de Vries, A. H.; Marrink, S. J.; Souza, P. C. T. Martini coarse-grained models of imidazolium-based ionic liquids: from nanostructural organization to liquid-liquid extraction. *Green Chem.* **2020**, *22*, 7376–7386.
- (23) Liu, J.; Qiu, L.; Alessandri, R.; Qiu, X.; Portale, G.; Dong, J.; Talsma, W.; Ye, G.; Sengrian, A. A.; Souza, P. C. T.; Loi, M. A.; Chiechi, R. C.; Marrink, S. J.; Hummelen, J. C.; Koster, L. J. A. Enhancing Molecular n-Type Doping of Donor-Acceptor Copolymers by Tailoring Side Chains. *Adv. Mater.* **2018**, *30*, 1704630.
- (24) Souza, P. C. T.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grünewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B. M. H.; Wassenaar, T. A.; Kroon, P. C.; Melcr, J.; Nieto, V.; Corradi, V.; Khan, H. M.; Domański, J.; Javanainen, M.; Martinez-Seara, H.; Reuter, N.; Best, R. B.; Vattulainen, I.; Monticelli, L.; Periole, X.; Tielemans, D. P.; de Vries, A. H.; Marrink, S. J. Martini 3: a General Purpose Force Field for Coarse-Grained Molecular Dynamics. *Nat. Methods* **2021**, *18*, 382–388.
- (25) 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.
- (26) Kroon, P. C.; co-workers, Cartographer: an automated topology builder for Martini. <https://github.com/marrink-lab/cartographer>.
- (27) Empereur-mot, C.; Pesce, L.; Doni, G.; Bochicchio, D.; Capelli, R.; Perego, C.; Pavani, G. M. Swarm-CG: Automatic Parametrization of Bonded Terms in MARTINI-

Based Coarse-Grained Models of Simple to Complex Molecules via Fuzzy Self-Tuning Particle Swarm Optimization. *ACS Omega* **2020**, *5*, 32823–32843.

- (28) Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. E. M. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (29) Nunes, R. S.; Vila-Viosa, D.; Costa, P. J. Halogen Bonding: An Underestimated Player in Membrane–Ligand Interactions. *J. Am. Chem. Soc.* **2021**, *143*, 4253–4267.
- (30) 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.
- (31) 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*, 19–25.
- (32) Alessandri, R.; Uusitalo, J. J.; de Vries, A. H.; Havenith, R. W. A.; Marrink, S. J. Bulk Heterojunction Morphologies with Atomistic Resolution from Coarse-Grain Solvent Evaporation Simulations. *J. Am. Chem. Soc.* **2017**, *139*, 3697–3705.
- (33) Janeliunas, D. *Self-Assembly of Facial Oligothiophene Amphiphiles*; Ph.D. Dissertation, TU Delft, Delft University of Technology, 2014.
- (34) Bulacu, M.; Goga, N.; Zhao, W.; Rossi, G.; Monticelli, L.; Periole, X.; Tieleman, D. P.; Marrink, S. J. Improved Angle Potentials for Coarse-Grained Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2013**, *9*, 3282–3292.
- (35) Gertsen, A. S. *Multiscale Modelling of Organic Solar Cell Materials—morphology evolution in solution processed bulk heterojunctions*; Ph.D. Dissertation, Technical University of Denmark, 2020.

- (36) Grünwald, F.; Alessandri, R.; Kroon, P. C.; Monticelli, L.; Souza, P. C. T.; Marrink, S. J. Polyppy: a python suite for facilitating simulations of (bio-)macromolecules and nanomaterials. **2021**, preprint arXiv:2105.05890.
- (37) Jorgensen, W. L.; Tirado-Rives, J. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl. Acad. Sci. U.S.A* **2005**, *102*, 6665–6670.
- (38) Dodda, L. S.; Cabeza de Vaca, I.; Tirado-Rives, J.; Jorgensen, W. L. LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. *Nucleic Acids Res.* **2017**, *45*, W331–W336.
- (39) Dodda, L. S.; Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L. 1.14*CM1A-LBCC: Localized Bond-Charge Corrected CM1A Charges for Condensed-Phase Simulations. *J. Phys. Chem. B* **2017**, *121*, 3864–3870.
- (40) 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.
- (41) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: a New Molecular Dynamics Method. *J. App. Phys.* **1981**, *52*, 7182–7190.
- (42) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
- (43) Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **1984**, *52*, 255–268.
- (44) Hoover, W. G. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.

- (45) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40*, 843.
- (46) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darrian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell Jr., A. D. CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671–690.
- (47) Yu, W.; He, X.; Vanommeslaeghe, K.; MacKerell Jr., A. D. Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *J. Comput. Chem.* **2012**, *33*, 2451–2468.
- (48) Koziara, K. B.; Stroet, M.; Malde, A. K.; Mark, A. E. Testing and validation of the Automated Topology Builder (ATB) version 2.0: prediction of hydration free enthalpies. *J. Comput. Aided Mol. Des.* **2014**, *28*, 221–233.
- (49) Stroet, M.; Caron, B.; Visscher, K. M.; Geerke, D. P.; Malde, A. K.; Mark, A. E. Automated Topology Builder Version 3.0: prediction of Solvation Free Enthalpies in Water and Hexane. *J. Chem. Theory Comput.* **2018**, *14*, 5834–5845.
- (50) Bernazzani, L.; Cabani, S.; Conti, G.; Mollica, V. Thermodynamic study of the partitioning of organic compounds between water and octan-1-ol. Effects of water as cosolvent in the organic phase. *J. Chem. Soc., Faraday Trans.* **1995**, *91*, 649–655.
- (51) Shirts, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys.* **2008**, *129*, 124105.
- (52) Kastner, J. Umbrella sampling. *WIREs Comput. Mol. Sci.* **2011**, *1*, 932–942.
- (53) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: a Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, 020603.

- (54) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.
- (55) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The Weighted histogram Analysis Method for Free-Energy Calculationss on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
- (56) Branduardi, D.; Bussi, G.; Parrinello, M. Metadynamics with Adaptive Gaussians. *J. Chem. Theory Comput.* **2012**, *8*, 2247–2254.
- (57) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: new feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604–613.
- (58) Sangster, J. Octanol-Water Partition Coefficients of Simple Organic Compounds. *J. Phys. Chem. Ref. Data* **1989**, *18*, 1111–1229.
- (59) Hansch, C.; Leo, A.; Hoekman, D.; Livingstone, D. *Exploring QSAR: hydrophobic, electronic, and steric constants*; American Chemical Society Washington, DC, 1995; Vol. 48.
- (60) Abraham, M. H.; Chadha, H. S.; Whiting, G. S.; Mitchell, R. C. Hydrogen Bonding. 32. An Analysis of Water-Octanol and Water-Alkane Partitioning and the $\Delta\log P$ Parameter of Seiler. *J. Pharm. Sci.* **1994**, *83*, 1085–1100.
- (61) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W. Refinement and Parametrization of COSMO-RS. *J. Phys. Chem. A* **1998**, *102*, 5074–5085.
- (62) Abraham, M. H.; Platts, J. A.; Hersey, A.; Leo, A. J.; Taft, R. W. Correlation and estimation of gas-chloroform and water-chloroform partition coefficients by a linear free energy relationship method. *J. Pharm. Sci.* **1999**, *88*, 670–679.
- (63) Natesan, S.; Wang, Z.; Lukacova, V.; Peng, M.; Subramaniam, R.; Lynch, S.; Balaz, S.

- Structural Determinants of Drug Partitioning in n-Hexadecane/Water System. *J. Chem. Inf. Model.* **2013**, *53*, 1424–1435.
- (64) Haynes, W. M. *CRC handbook of chemistry and physics*; CRC press, 2014.
- (65) Martin, M. G.; Siepmann, J. I. Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes. *J. Phys. Chem. B* **1998**, *102*, 2569–2577.
- (66) Martin, M. G. Comparison of the AMBER, CHARMM, COMPASS, GROMOS, OPLS, TraPPE and UFF force fields for prediction of vapor–liquid coexistence curves and liquid densities. *Fluid Ph. Equilibria* **2006**, *248*, 50–55.
- (67) Sweere, A. J.; Fraaije, J. G. E. M. Accuracy test of the OPLS-AA force field for calculating free energies of mixing and comparison with PAC-MAC. *J. Chem. Theory Comput.* **2017**, *13*, 1911–1923.
- (68) Ben-Naim, A. Y. *Solvation thermodynamics*; Springer, 2013.
- (69) Scatchard, G.; Wood, S. E.; Mochel, J. M. Vapor-Liquid Equilibrium. III. Benzene-Cyclohexane Mixtures. *J. Phys. Chem.* **1939**, *43*, 119–130.
- (70) Javanainen, M.; Fabian, B.; Martinez-Seara, H. Comment on “Capturing Phase Behavior of Ternary Lipid Mixtures with a Refined Martini Coarse-Grained Force Field”. **2020**, preprint arXiv:2009.07767.
- (71) Souza, P. C. T.; Limongelli, V.; Wu, S.; Marrink, S. J.; Monticelli, L. Perspectives on High-Throughput Ligand/Protein Docking With Martini MD Simulations. *Front. Mol. Biosci.* **2021**, *8*, 199.
- (72) Jackson, N. E. Coarse-Graining Organic Semiconductors: The Path to Multiscale Design. *J. Phys. Chem. B* **2021**, *125*, 485–496.

- (73) Jarin, Z.; Newhouse, J.; Voth, G. A. Coarse-Grained Force Fields from the Perspective of Statistical Mechanics: Better Understanding of the Origins of a MARTINI Hangover. *J. Chem. Theory Comput.* **2021**, *17*, 1170–1180.