Multiscale modeling of organic materials

Alessandri, Riccardo

DOI:
10.33612/diss.98150035

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2019

[Link to publication in University of Groningen/UMCG research database](#)

6

Martini 3.0 Coarse-Grained Force Field: Ring Structures
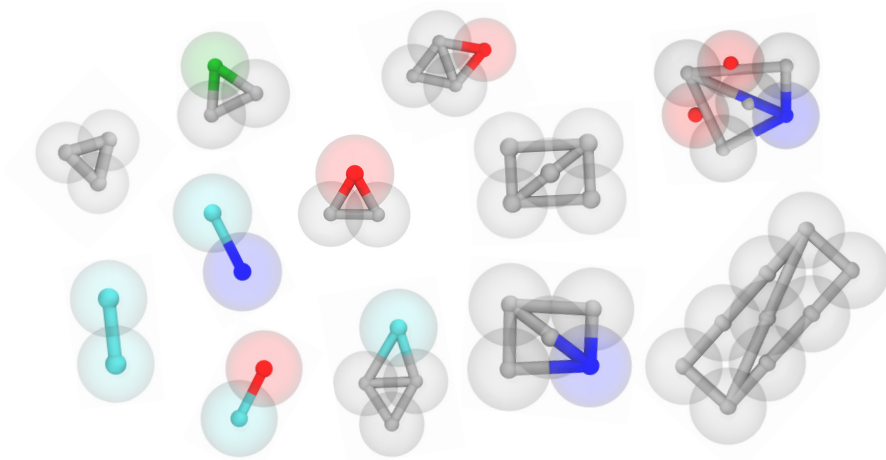
With the release of Martini 3.0, the Martini framework has been refined and expanded after more than 10 years of applications and lessons learned thereof. In this particular account, we present a pool of aliphatic and aromatic ring structures which include solvents and building blocks for macromolecules, such as proteins and (polymeric) organic molecules, relevant for soft materials science. Within Martini 3.0, such structures are described by models which use small and tiny beads. As such, the present pool of rings constitutes one of the cornerstones of the calibration of the new Martini small and tiny beads. The new models show excellent partitioning behavior and solvent properties. Miscibility trends between different bulk phases are also captured, completing the set of "bulk" 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 "local" properties such as stacking distances. We further present guidelines to build Martini 3.0 models for mono- or polycyclic compounds. A collection of all the molecules taken into account is also made available as a database which includes their corresponding Martini model, experimental and computed properties, and links to simulation files. The present parametrization, database, and guidelines, along with the modularity of the Martini force field, are expected to boost the application of the Martini 3.0 CG force field in soft materials science.

6.1. Introduction

6

The use of coarse-grain (CG) models is of paramount importance to explore length and time scales beyond what is feasible with atomistic, or all-atom (AA), models.[35–37,295–297] In CG models, atoms are bundled together into supra-atomic 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 conventional experimental methods to study processes such as spontaneous self-assembly, transformation between phases, domain formation, and host-guest interdiffusion.[35–37,295–298]

CG models can be categorized as mainly following a systematic (also known as hierarchical) or a building block approach.[27,36] 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 thus less sampling. On the other side, building block approaches are often cheaper due to the use of simpler potential forms and only partial parametrization required. Indeed, the parametrization of the building blocks enables their use as part of similar moieties in different molecules, hence the second advantage of such models, that is, their transferability. However, these advantages come at the cost of a more limited structural accuracy, given the necessarily suboptimal representation of the underlying atomistic detail inherent of 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.[36]

The power of building block-based CG models has been harnessed significantly more in the study of biomolecular systems, while many fewer examples are found in the soft materials science literature. Despite pioneering CG modeling works originated from the necessity of modeling polymers,[299] most of the works in soft materials science remains focused on systematic CG modeling approaches which accurately capture the physics of relatively simple systems but are not easily extendable when the systems grow in complexity in terms of the number of different interactions which need to be taken into account. This hampers the broader application of systematic CG models in materials science, as the latter is rich in complex blends of multiple (polymeric) molecular structures which are used in state-of-the-art organic devices.[7,8] Building block CG approaches hold promise to be able to investigate such systems with reasonable accuracy and within time scales which are not only relevant to speed up our understanding of these systems but also to explore new blend compositions *in silico* before going to the experimental trial. Among the building block approaches, notably the Martini CG force field[38,39] has seen wide applications due to successes achieved in the description of several biomolecular systems.[38,40–42] 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 polymer chemistry and organic electronics.[43,45–47,117,197,198,264,266]

With the recent development of Martini 3.0,[210] the Martini framework has been refined and expanded after more than 10 years of applications and lessons learned thereof. In this particular account, we present what has been done concerning the parametrization of ring structures, notably something which was not the main focus of the original Martini work, where the main target of ring particles included only benzene, cyclohexane, and cholesterol molecules.[39] The wide use of the Martini model highlighted certain limitations[278–280,284] (see also chapter 5) in particular in the description of structures which are modeled with more than one bead per 4 atoms, among which we find ring structures. Addressing such limitations was especially important in order to address the growing interest[47,117,197,198,264,266,300] in applying Martini in soft materials science, where (polymeric) conjugated ring systems play an ubiquitous role. In this contribution, we

especially focus on rings in the context of materials science applications. However, the parametrization has also taken into account rings which are relevant for biomolecules like the aromatic amino acids, the DNA bases, and sugars. Those will be described in separate publications dedicated to proteins, DNA, and sugars. In the present work, we describe which properties have been considered during the parametrization of ring structures, we present guidelines to build Martini 3.0 models for mono- or polycyclic compounds, showcase what can be achieved with the new models, and describe expected strengths and potential weaknesses of the new model in soft materials science applications. A collection of all the molecules taken into account is also made available as a database which includes their corresponding Martini model, experimental and computed properties, and links to simulation files.

6.2. Model and Methods

6.2.1. Coarse-Graining Ring Structures within Martini 3.0

Martini 3.0. For a thorough description of the latest version of the Martini force field we refer to the main publication. [210] Briefly, as in the previous versions, the main force field parametrization target remains free energy of transfer data between several pairs of solvents, such as hexadecane/water, octanol/water, and chloroform/water. Molecular fragments are used as building blocks (beads), whose non-bonded interactions are described by modified Lennard-Jones interactions. To describe the chemical nature of the underlying atomistic structure, CG beads with more or less polar character exists. As in version 2, there are four main types of beads: polar (P), non-polar (N), apolar (C), and charged (Q). Moreover, a separate water (W) bead and a set of beads (X) dedicated to groups containing multiple halogen atoms are now included. These types are in turn divided in subtypes based on their hydrogen-bonding capabilities, degree of polarity, or charge softness. [210] To keep the model simple, the sizes of the beads (σ) and the levels of interactions (ε) are discretized. In version 3, the number of interaction levels was extended from 10 to 20, and, besides the three bead sizes (regular (R), small (S), and tiny (T), with σ of 0.47, 0.405, and 0.34 nm, respectively), there are now also cross-interaction sizes. As we will show in the present contribution, this improves, among other aspects, the overall behavior of systems described by more than one bead per 4 atoms and their interaction with 4-to-1 (atoms-to-CG-site) mapped systems. Such finer mappings are usually required in the case of cyclic compounds, [39] or to adhere to the symmetry of the repeat unit of a polymer. [144,145] 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. However, if the chemical group is fully branched, or part of a ring structure, a smaller size than what expected based on the (non-

hydrogen) atom count should be used. In this contribution, we will comprehensively describe the parametrization of low molecular weight cyclic compounds, which forms the basis for the T- and S-bead parameters.

Center-of-Geometry Mapping. 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 6.8 (Appendix) 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.[39] Instead, center-of-geometry (COG)-based mapping leads to better molecular (*e.g.,* solvent accessible surface area) and bulk (*e.g.,* mass densities) properties (Section 6.3.2). 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 6.9 in the Appendix). 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, for example, it makes them suitable for automated topology building.[279,301,302]

Aromatic Rings. Tiny-beads are especially intended to be used to describe aromatic (*i.e.,* planar, atom-thick) structures. The prototypical aromatic compound, benzene, is described by a three-bead model reminiscent of the Martini 2 model[39] (see also Figure 6.1): each of the three beads represents two consecutive carbon atoms and their associated 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 6.3.4). For benzene, or more generally non-substituted $-C=C-$ groups part of aromatic compounds, the Martini bead type of choice is C5, because it leads to free energies of transfer in excellent agreement with experimental literature data (see Section 6.3.1); therefore benzene is represented by a TC5-TC5-TC5 model. A COG-based bond length of 0.29 nm is used (see also Figure 6.8 in the Appendix), leading to a bulk density for liquid benzene of 0.894 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 together[39] by constraints.[303] Distributions of bond lengths obtained from mapped AA simulations between beads which are part of aromatic structures are indeed very narrow (Figure 6.8c in the Appendix). 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 more detail later in this Section.

The parameters of benzene form the basis for all the aromatic ring structures which have been included in the present parametrization (some of which are shown in Figure 6.1). Together with the other topologies presented in this work, this set should guide the development of new Martini 3 models for aromatic molecules.
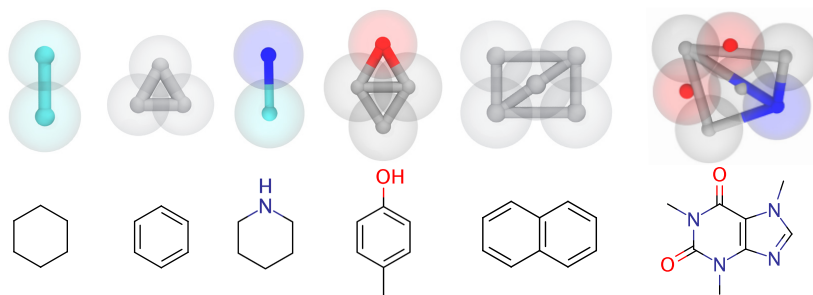

Figure 6.1 | Examples of mappings of aromatic and aliphatic ring structures. 2D chemical structures (bottom) are shown along with renderings of their corresponding CG (top) models. From left to right: cyclohexane, benzene, piperidine, para-cresol, naphthalene, and caffeine. CG-beads are color-coded as follows: apolar aliphatic (C1-C3 type) beads are in cyan, apolar aromatic (C4-C6) beads are in silver, nonpolar (N0 to N5) beads in blue and polar (P1 to P6) beads in red.

Aliphatic Rings. Small-beads are used for aliphatic cyclic compounds. The prototypical aliphatic cyclic compound, cyclohexane, is described by a two-S-bead model, differentiating it from the Martini 2 model.[39] 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 6.3.1). Bonds keep together aliphatic models, as bond distributions are broader in the case of aliphatic rings (compare Figure 6.8c and Figure 6.8f in the Appendix). A bond length of 0.385 nm is used, leading to a bulk density for liquid cyclohexane of 0.81 g cm^{-3}, in good agreement (≈5% too high) with the experimental value of 0.77 g cm^{-3}.

Analogously to benzene, cyclohexane constitutes the basis for the aliphatic rings which have been object of the present parametrization. Accordingly, the present pool of aliphatic compounds should serve as reference for the development of new Martini 3.0 models of aliphatic cyclic compounds.

Heterocyclic Compounds and Substituted Rings. Benzene and cyclohexane form the basis of the aliphatic and aromatic rings. However, most of the interesting compounds contain also heteroatoms either within the ring itself or as substituents. Table 6.1 lists a subset of the small and tiny Martini particle types and the corresponding chemical group which they describe in ring models, along with examples of molecules which employ that particle type. Note that in some cases a (group of) atom(s) is shared between two neighboring beads (indicated with a [†] in Table 6.1). This is usually necessary in order

Table 6.1 | Ring building blocks. CG particle type, corresponding ring 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 and S-beads depicted in blue and green, respectively.
† Indicates that the (group of) atom(s) is shared with neighboring beads.

type	ring building block	examples 2D	name (mapping)
SP1a	$-CH_2^\dagger-C(=O)-CH_2^\dagger-$		cyclopentanone (SC3-SP1a)
TP1	$=C(-OH)-$		p-cresol (TC5-TC5-TC5-TP1)
SN6	$-CH^\dagger=C(-OH)-CH^\dagger=$		phenol (TC5-TC5-SN6)
TN6a	$-CH=N-$		pyridine (TC5-TC5-TN6a)
TN6d	$=CH^\dagger-NH-CH^\dagger=$		pyrrole (TC5-TC5-TN6d)
SN5a	$-CH_2-O-CH_2-$		tetrahydropyran (SC3-SN5a)
TN5a	$-CH_2^\dagger-O-CH_2^\dagger-$		tetrahydrofuran (SC3-TN5a)
TN3a	$=CH^\dagger-O-CH^\dagger=$		furan (TC5-TC5-TN3a)
TN1	$-N(-CH_3)-$		1-methylimidazole (TC5-TN1-TN6a)
SC6	$-CH^\dagger=C(-SH)-CH^\dagger=$		thiophenol (TC5-TC5-SC6)
TC6	$=CH^\dagger-S-CH^\dagger=$		thiophene (TC5-TC5-TC6)
TC5	$-CH=CH-$		benzene (TC5-TC5-TC5)
SC4	$-CH^\dagger=C(-CH_3)-CH^\dagger=$		toluene (TC5-TC5-SC4)
	$-CH^\dagger=C(-Cl)-CH^\dagger=$		chlorobenzene (TC5-TC5-SC4)
SC3	$-(CH_2)_3-$		cyclohexane (SC3-SC3)

6

to keep the symmetry of the underlying atomistic structure—see, for example, phenol, tetrahydrofuran, or toluene. Note that this is important also in view of the extraction of bond and (dihedral) angle distributions from CG-mapped atomistic simulations.

Beads are chosen based on partitioning data, the Martini hallmark.[38] More bead types are now available within Martini 3 with respect to Martini 2, thus, for example, the three groups $-N(-CH_3)-$, $=CH^{\dagger}-O-CH^{\dagger}=$, and $-CH_2{}^{\dagger}-O-CH_2{}^{\dagger}-$ (found in 1-methylimidazole, furan, and tetrahydrofuran, respectively—see also Table 6.1) once described by a SN0 bead are now described by a TN1, TN3a, and TN5a bead, respectively. Table 6.1, listing mapping of CG particle types to chemical building blocks, serves as a guide towards the assignment of CG particle types.
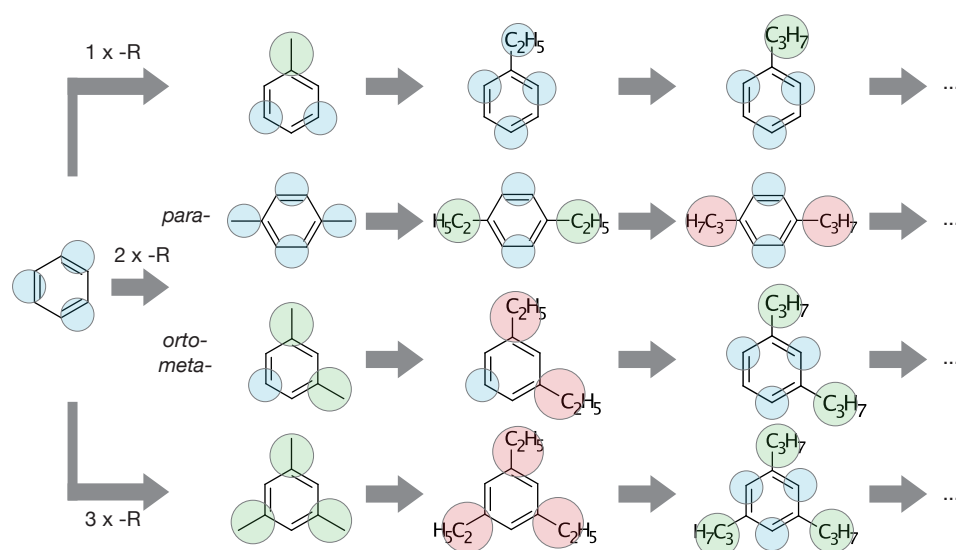
Figure 6.2 | Recommended mappings for substituted rings. Mono- (top), di- (center), and tri-substituted (bottom) rings are shown. In the case of di-substituted rings, we distinguish between *para-* and *meta-* or *orto-* substitutions, the latter two to be treated identically in terms of mapping. Regular, small, and tiny beads are indicated in red, green, and blue, respectively, but not drawn to scale.

Figure 6.2 shows how we mapped substituted rings 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 and shape of the molecule as much as possible.

Leveraging Virtual Sites in Extended Ring Models. For polycyclic compounds that need to be described by more than four particles, virtual interaction sites (VSs)[304] are used to improve the numerical stability of the models. As mentioned earlier, Martini *aromatic* ring structures are held together by constraints. As more extended stiff molecular structures need to be described (as in the case of polycyclic aromatic compounds), the number of constraints which need to be used in their Martini representations grows.

However, a too extended network of constraints leads to numerical instabilities, as it is increasingly complicated to satisfy a growing number of connected constraints. Saving on the number of constraints by using VSs not only increases the numerical stability of the simulations but also improves their performance.

The proposed model for the naphthalene (NAPH) molecule illustrates one way in which VS can be used to improve numerical stability and performance. The molecule consists of 10 aromatic carbon atoms, which translates into a 5-TC5-bead Martini model. Straightforwardly interconnecting the 5 beads with constraints leads to a network of 8 constraints (see Figure 6.3a). Such a network of constraints already starts to be hard to satisfy: while 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.[278] Indeed, another model which could be devised for this molecule exploits the "hinge" construction for the 4 external beads (Figure 6.3b) while the central bead is described as a VS, *i.e.*, a particle whose position is completely defined by its constructing particles (in this case, the other four beads). The displacement of the VS 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.[48] Of course, forces on the VS are accounted for—these are distributed to the constructing atoms in each integration step. VSs 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

<table>
<tr><td>(a)</td><td>(b)</td><td>(c)</td></tr>
<tr><td>5 particles, 8 constraints</td><td>4 particles, 1 virtual site
5 constraints, 1 (improper) dihedral</td><td></td></tr>
</table>
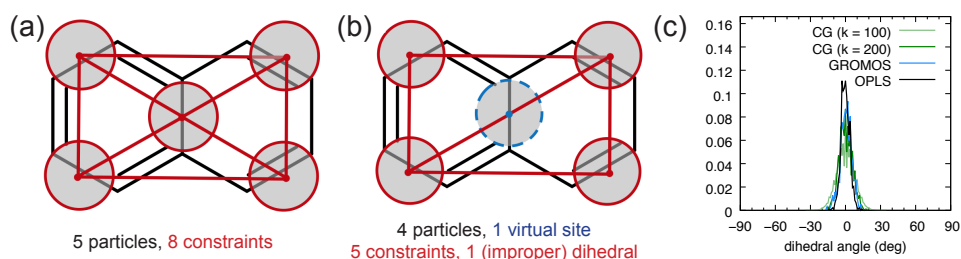
Figure 6.3 | Possible Martini models for naphthalene. A (a) "all-constraints" model which readily leads to numerical instabilities and (b) a "hinge" model which runs with a time step of 20 fs. A schematic representation of the CG model is drawn on top of the chemical structure of naphthalene: the grey circle with solid red contour indicates a CG particle, while a gray circle with a dashed cyan contour represents a virtual site. Solid red lines indicate constraints. The "hinge" model also uses an improper dihedral (between the four vertices of the model) to keep the model flat. A comparison of this dihedral distribution to reference COG-mapped atomistic distributions is shown in (c) for two improper dihedral force constants, 100 kJ mol^{-1} rad^{-2} and 200 kJ mol^{-1} rad^{-2}. Non-bonded interactions between the two CG sites which share no bond in the "hinge" model—top left and bottom right in (b)—are excluded.

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 ~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 6.3c). 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 one would need to define the NAPH plane—and then construct the remaining two as VSs. 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. The hinge model, on the other hand, works as a shock absorber,[278] preventing such an absolute rigid body behavior. It is thus recommended to opt for such a non-rigid-body model. This was used to build models for other extended ring structures, such as caffeine, the DNA bases, and some polyacenes (see Figure 6.10 in the Appendix). Note that in these other cases, more than one VS per hinge is defined.

6.2.2. Database Description

A database containing about 50 molecules, covering most of the chemical functional groups, will be soon made available online on the Martini portal http://cgmartini.nl. The database contains mono- and polycyclic aromatic and aliphatic compounds, 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 polymeric structures. Molecules can be searched via their common name, CAS number, SMILES string, as well as an identifying three to five letter long string. For convenience, the database can be found both as a Microsoft Excel and as a comma-separated value file. Jupyter notebook examples on how to import and query the database for properties, output to file or plot data will be also made available. The database will be also possibly made available via the http://virtualchemistry.org/ portal.[305,306] The database contains also force field topologies (in GROMACS format) which can be used as input for MD simulations.

6.2.3. Simulation Settings and Procedure Details

General Simulation Settings. A unique set of GROMACS atomistic run parameters was used for the AA simulations. 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 barostat[152] (coupling parameter of 5.0 ps) and the Nosé-Hoover thermostat[288,289] (coupling parameter of 1.0 ps) were used to maintain pressure and temperature, respectively. Settings for the CG simulations adhere to the "new"

set of Martini run parameters.[150] Specifically, the Verlet neighbor search algorithm was used to update the neighbor list, with a straight cutoff of 1.1 nm. The Parrinello–Rahman barostat[152] (coupling parameter of 12.0 ps) and the velocity-rescaling thermostat[151] (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.

Atomistic Models. GROMOS 54A7[307] all-atom models were obtained, in GROMACS format, from the ATB server.[205,308] OPLS-AA[53] models were obtained, in GROMACS format, from either Caleman *et al.*[305,306] or the Ligandbook server.[309] CHARMM General FF (CGenFF)[51,310] models were obtained by uploading a molecule (in .mol2 format) to https://cgenff.umaryland.edu/ to obtain a .str file and subsequently pass both files to cgenff_charmm2gmx.py (using the July2017 version, charmm36-jul2017.ff.tgz, both downloaded from http://mackerell.umaryland.edu/charmm_ff.shtml) to obtain a topology in GROMACS format.

Liquid and Gas Phase Simulations. A liquid phase was approximated as an equilibrated box of dimensions of about $5 \times 5 \times 5$ nm^3; a gas phase as a single molecule occupying a large ($7 \times 7 \times 7$ 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:

$$\Delta H_{vap} \approx U_{gas} - U_{liq} + RT \qquad (6.1)$$

where U_{liq} and U_{gas} are the total energies (per mole) of the gas and liquid phase, respectively. Densities were extracted with the GROMACS tool gmx density. Note that, in the case of the CG simulations, this can be done directly only if the actual masses of the molecules—shared evenly among the beads—are used, while by default Martini 3, like Martini 2, uses standard masses for R-, S-, and T-beads of 72, 54, and 36 Da.

Free Energies of Transfer Calculations. We used thermodynamic integration to compute free energies of solvation $\Delta G_{S \to \emptyset}$ in a solvent S. ΔGs 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,[210] respectively. Note that we simulate hydrated octanol, that is we add a 0.26 mole fraction of water according to experimental conditions.[311] 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).[171] The free energy associated with transferring a solute from a solvent S_1 to a solvent S_2 ($\Delta G_{S_1 \to S_2}$) was computed as the difference $\Delta G_{S_1 \to \emptyset} - \Delta G_{S_2 \to \emptyset}$.

6

Dimerization Free Energy Surface Calculations. Free energy surface (FES) profiles for the dimerization of each molecule were computed by either umbrella sampling (US)[312] or (well-tempered[313]) metadynamics (MTD)[314,315] 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)[175] as implemented in the GROMACS tool `gmx wham`. In MTD simulations, the CVs on which the bias was added were the distance between the COGs of the two ring structures and the torsional angle formed by atoms of the rings and their COGs in the case of AA simulations. For 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 μs long in the AA and CG case, respectively. The order parameter S of Eq. 6.4, which describes the relative orientation of the ring planes, was used as a third CV to project the FES by using a reweighting algorithm.[316] 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,[48] patched with PLUMED 2.4[317] 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 $9 \times 9 \times 9$ 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 6.12a in the Appendix.

Vapor–Liquid Equilibrium Simulations. Vapor–Liquid Equilibrium (VLE) simulations were performed by setting up a system of dimensions $7 \times 7 \times 16$ nm^3, where half of the simulation box—$7 \times 7 \times 8$ nm^3—is empty (starting configuration for the vapor phase),

and the other half—$7 \times 7 \times 8$ nm^3—contains a 50:50 mixture of two components (starting configuration for the liquid phase). A rendering of this setup is shown in Figure 6.6d. 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 1 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 Appendix. 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 6.3 as described in the main text.

6

6.3. Results and Discussion

6.3.1. 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 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[172,173,213,318,319] (see Methods for the solvent models). Figure 6.4 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 2.2, 1.8, and 2.4 kJ mol^{-1}, for HD→W, OCO→W, and CLF→W, respectively.
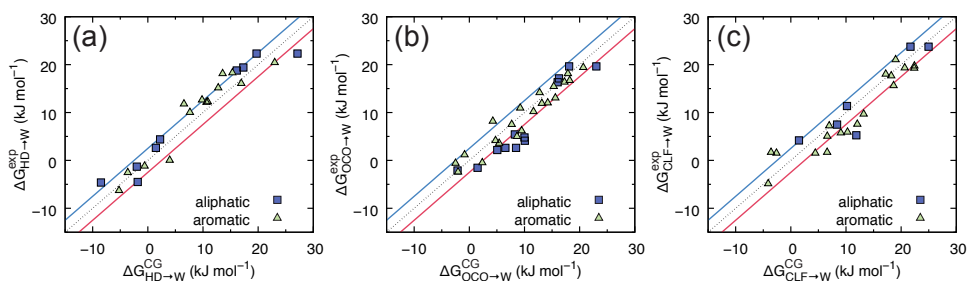

Figure 6.4 | Partitioning behavior of Martini 3.0 cyclic compounds. Computed data are plotted against experimental data in the cases of (a) hexadecane→water, (b) octanol→water, and (c) chloroform→water. The red and blue lines below and above the diagonal represent threshold error lines of 2.5 kJ mol^{-1} (*i.e.*, $k_B T$). Data points corresponding to aromatic and aliphatic rings are indicated by green triangles and blue squares, respectively. Experimental data have been compiled from Refs. 172,173,213,319.

6.3.2. Solvent Properties: Enthalpies of Vaporization and Densities

Many aromatic and aliphatic cyclic compounds are widely used solvents. We computed their mass densities (ρ) and enthalpies of vaporization (ΔH_{vap}) and compared them to experimental values.[271] The results are shown in Figure 6.5, where the CG values were computed as described previously in section 6.2.3. In the case of the enthalpies of vaporization (Figure 6.5a), we also report the values for other Martini 3.0 molecules, in order to show the overall trend of ΔH_{vap} of the Martini 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.[39]

Moreover, we recall (chapter 5) that Martini 2.2 ring models deviate from the overall force field trend and lead to systematically higher heat of vaporizations. This was related to the "overmapping" of such molecules and imperfect S-bead calibration (chapter 5). This is now overcome in Martini 3.0, where all the cyclic compounds lie in the trend of the Martini force field.
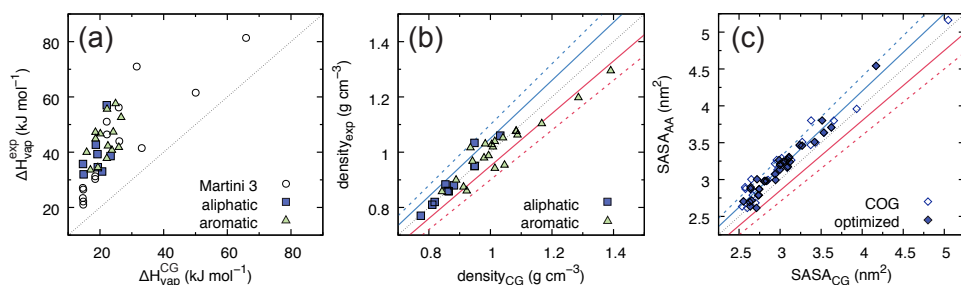

Figure 6.5 | Solvent properties of Martini cyclic compounds: (a) ΔH_{vap}, (b) mass density, and (c) solvent accessible surface area (SASA). In (b) and (c), the red and blue solid (dashed) lines indicate threshold error lines of plus or minus 5% (10%), respectively. Data points corresponding to aromatic and aliphatic rings are indicated by green triangles and blue squares, respectively. In (c), results for both models using bond lengths extracted from COG-mapped atomistic models (COG, open diamonds) and optimized bond lengths (optimized, filled diamonds) are shown. Experimental mass densities and enthalpies of vaporization are from Ref. 271. SASA reference values have been obtained with the GROMOS 54A7 [307] or OPLS [53] all-atom force fields (see also Methods). Note that different all-atom force fields show negligible differences in SASA values—see Appendix. For details on the SASA calculations, see the Appendix.

Computed mass densities for the solvents are overall in good agreement with experimental densities: the mean absolute percentage error is 4.3 %. This is shown in Figure 6.5b, where CG *vs* experimental densities are plotted for all the rings in the database. Not surprisingly, not only the LJ parameters but also the bond lengths utilized 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 close reproduction of the solvent accessible surface area (SASA) of the AA model, as shown in Figure 6.5c. This also translates into good agreement in case of mass densities (see Figure 6.11 in the Appendix). However, given the lower resolution of the model, perfectly matching densities cannot be obtained for all cases simply following a COG-based mapping scheme. Thus, bond lengths can be tuned further if higher accuracy is required. This was done 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 6.5b and for the SASA values reported in Figure 6.5c.

6.3.3. Solvent Mixtures: Martini Miscibility Table

We tested the capabilities of the model 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),[54] while established force fields have been tested for their performance in reproducing such equilibria[320] or other related thermodynamic properties of mixtures (*e.g.,* the free energy of mixing).[321] 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 all the aromatic or aliphatic cyclic compounds which usually appear in such tables. The extracted experimental miscibility table is shown in Figure 6.6a, while results for Martini 3.0 are shown in Figure 6.6b, the *Martini miscibility table for cyclic compounds*. 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. Martini fully reproduces the experimental miscibility table.

A more quantitative estimation of how the new Martini model performs with respect to solvent mixtures has been then carried out for the BENZ-CHEX 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:[322]

$$\Delta G_{mix} = \Delta G_{ex} + T\Delta S_{id} = [x_A(\mu_A - \mu_A^*) + x_B(\mu_B - \mu_B^*)] + RT(x_A \ln(x_A) + x_B \ln(x_B)) \quad (6.2)$$

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 6.2 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 6.2 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 6.2, *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,[322] the free energy of solvation *in the mixture* ΔG_{solv} of the ith-component is given by:

$$\Delta G_{\text{solv},i} = RT \ln \left(\frac{\rho_{\text{vap},i}}{\rho_{\text{liq},i}} \right) \tag{6.3}$$

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
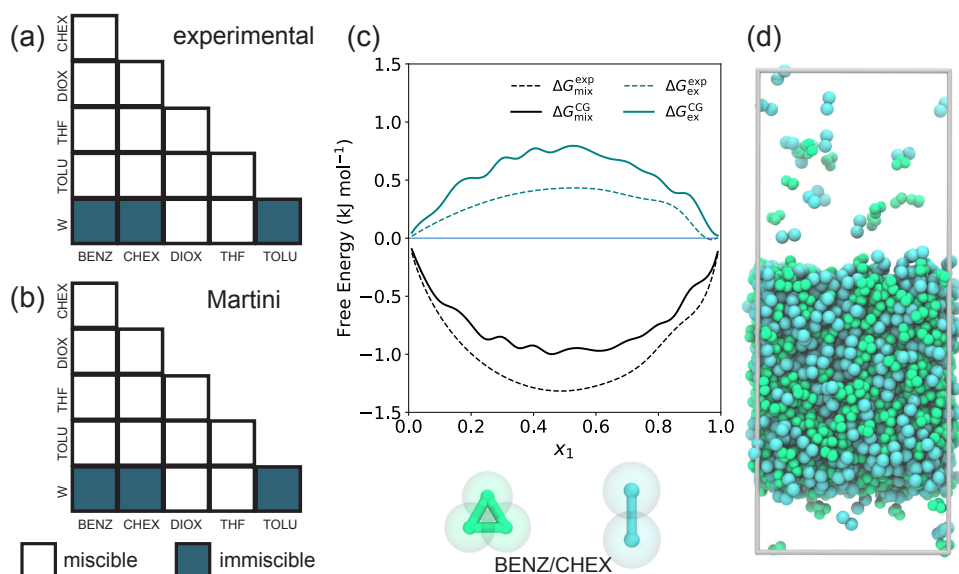

Figure 6.6 | Miscibility of aromatic and aliphatic rings (and water): Martini 3.0 *vs* experiments. The (a) experimental and (b) Martini solvent miscibility tables are shown. Light color means miscible, dark means immiscible. (c) Comparison between Martini and experimental free energy of mixing and excess free energy of mixing for the BENZ-CHEX mixture as a function of the mixture composition. x_1 is the molar fraction of BENZ. Experimental data are obtained from vapor-liquid equilibrium data.[323] Simulation data are obtained from vapor-liquid equilibrium simulations (see Methods). A rendering of the setup in shown in (d); small and tiny bead radii are rendered in scale.

via "direct" simulations, *i.e.*, by simulating equilibrated binary liquid-vapor systems (see Figure 6.6d 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 BENZ-CHEX 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 6.6c shows the ΔG_{mix} and ΔG_{ex} computed according to Equation 6.2. The agreement between simulated and experimental data is good. Experimentally, BENZ-CHEX 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 (BENZ-CLF) mixture has also been investigated with the same approach, and results are shown in Figure 6.12c in the Appendix. 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 VLE data can be used to further validate the miscibility of phases relevant to a certain application of the Martini force field.

6.3.4. Martini Stacking Interactions: Dimerization Free Energy Landscapes

We now turn to describe the performance of the model in testing more "local" 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 6.7 shows potentials of mean force (PMFs) of dimerization in water for several representative aromatic compounds—at the Martini 3.0, Martini 2.2, 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 the Methods section. In particular, we analyze the following aromatic rings: (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 mimina in Martini 3.0 is in good agreement with the atomistic reference data, showing improvements with respect to Martini 2.2—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 3.0 as compared to Martini 2.2. This is more in line with the atomistic reference data, and such an improvement is not unexpected but it is due to the introduction of LJ cross-interaction sizes, as opposed to Martini 2.x were these were absent (see also chapter 5). Note that, in most of the cases, we compare to reference AA PMFs obtained with more than one atomistic force field. A second look a the data shows that, going from Figure 6.7a to Figure 6.7c, that is, upon increasing the size of the aromatic system, the discrepancy between the overall shape of the AA and Martini 3.0 PMFs decreases. This can be explained by comparing the two-dimensional free energy profiles of Figure 6.7g–6.7h (AA) and Figure 6.7j–6.7k (Martini 3.0). The free energy is now plotted on the two-dimensional coordinate space formed by the distance between the
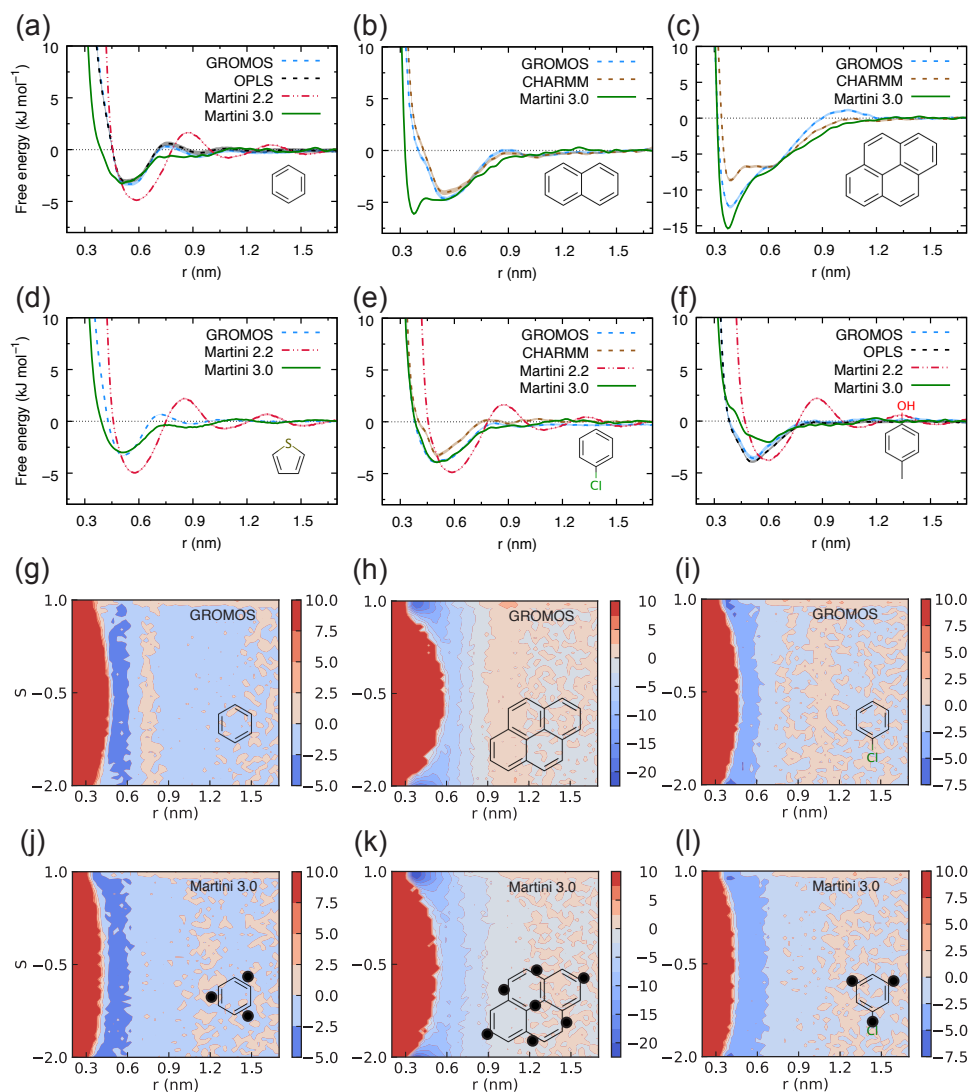
Figure 6.7 | One- and two-dimensional free energy profiles of dimerization for several aromatic compounds in water. (a)–(f) The free energy is plotted along the distance between the COGs of the two molecules. The profiles obtained with the Martini 3.0 force field (solid lines) are shown along with the Martini 2.2 (dash-dotted lines) and the atomistic (GROMOS 54A7,[307] OPLS-AA,[53] or CHARMM (CGenFF),[51,310] dashed lines) ones. (a)–(c) show how the dimerization profiles change upon increasing of the molecular size in the (poly)cyclic aromatic hydrocarbons benzene (a), naphthalene (b), and pyrene (c); (d) shows the profile for the 5-membered heterocyclic aromatic compound thiophene; (e) and (f) show profiles for substituted aromatic compounds which possess a permanent dipole moment: (e) chlorobenzene and (f) *p*-cresol. (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. 6.4. (g)–(i) are AA, and (j)–(l) are Martini 3.0 surfaces obtained adding the orientation parameter *S* (Eq. 6.4) as the second coordinate. Molecular structures (and mappings) are shown as figure insets.

COGs of the two molecules and the order parameter indicating the relative orientation of the planes of the two aromatic molecules:

$$S = \frac{(3\cos\theta - 1)}{2} \qquad (6.4)$$

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 ($\theta = 0°$) or -2 ($\theta = 180°$) 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 ($\theta = 90°$) denote a *T-shaped* conformation.

Figure 6.7g and 6.7h 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 6.13 in the Appendix) 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 6.13b). However, as the size of the molecule increases, the "hydrophobic effect" prevails over the quadrupolar interaction, leading to a sandwich interaction which minimizes the solute-water contacts. Standard Martini 3.0 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 6.7c for pyrene—such as the one of semi-crystalline polymers for organic electronics which were previously systematically off due to the larger size of S-beads.[117]

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 6.7e and 6.7f. Indeed, the 2D free energy landscape of Figure 6.7i (CLBZ) show the dominance of the sandwich conformation in molecules with a permanent dipole moment such as CLBZ (and PCRE, Figure 6.13a Appendix). This is again captured by Martini 3 (Figure 6.7k and 6.13c).

6.4. Conclusion

We presented the new parametrization of aromatic and aliphatic ring structures within the Martini 3.0 framework. The presented pool of aliphatic and aromatic ring structures was one of the cornerstones of the calibration of the new Martini small and tiny beads. The models show excellent partitioning behavior and very good solvent properties. Miscibility trends between different bulk phases are also captured, completing the set of "bulk" thermodynamic properties considered during the parametrization. We also showed how the new bead sizes allow for a good representation of molecular volume, which translates into better "local" properties such as stacking distances. We presented guidelines to build Martini 3.0 models for mono- or polycyclic compounds. Topologies which were developed within this work, which include solvents and building blocks for macromolecules such as proteins and (polymeric) organic molecules relevant for soft materials science, will be soon made available online on the Martini portal http://cgmartini.nl.

Despite widening even more the already broad range of applications for the Martini force field, limitations must be kept in mind. Certain limitations apply generally to the Martini force field,[39,210] 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. This is particularly true for systems described by regular beads for which softer potential forms would be more appropriate. Other limitations pertain more specifically to the systems subject of the present study, namely ring structures described by finer mappings. 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 options are currently being explored in our group. Finally, despite some design principles have been formulated in the present work, building models for extended planar (polymeric) systems can be challenging as the extended networks of constraints often required by such structures easily lead to numerical instabilities. We are currently working on a more extensive set of guiding principles which make it easier to build topologies for such systems.

While keeping in mind the limitations, the present parametrization, database, and guidelines, along with the modularity of the Martini force field, are expected to boost the application of the Martini 3.0 CG force field in soft materials science.

Acknowledgments

6.5. Appendix: Method Details

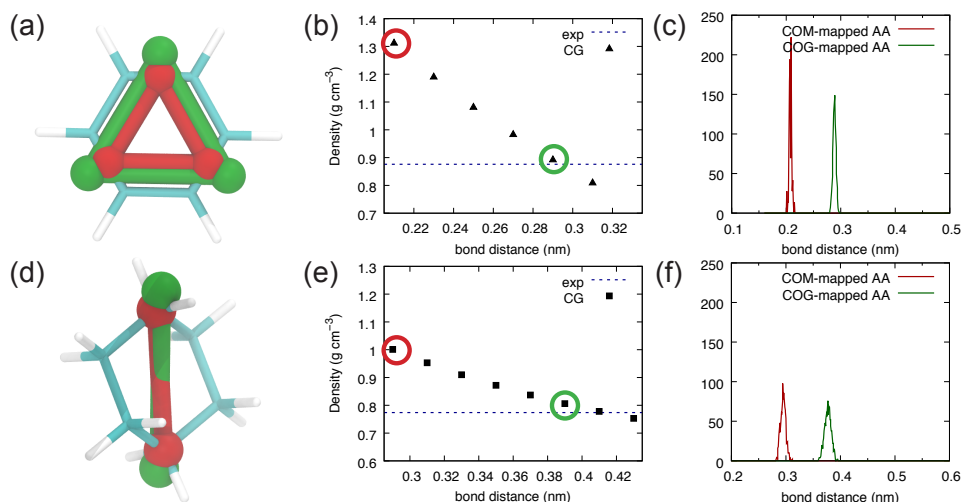### 6.5.1. COG- *vs* COM-Based Mappings: Benzene and Cyclohexane

Figure 6.8 | Benzene and cyclohexane bonded parameters. Bond lengths based on center-of-mass (COM) mapping—(a), (d), red—proved to be unsatisfactory, leading to too high packing densities (b), (e). Instead, center-of-geometry (COG)-based bond lengths–(a), (d), green—lead to densities close to experimental values (b), (e). The bond distributions for COM- and COG-mapped benzene (c) and cyclohexane (f) are also shown.

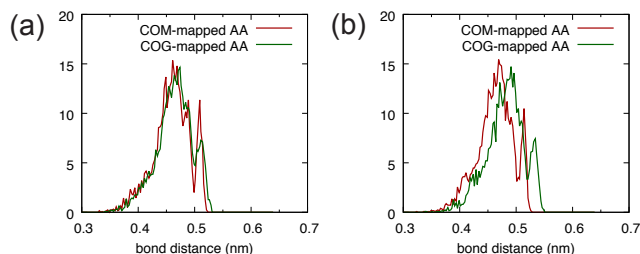### 6.5.2. COG- *vs* COM-Based Mappings: Alkanes

Figure 6.9 | Alkane bonded parameters extracted from hexadecane atomistic simulations. 4 atoms, including hydrogens, are mapped onto one CG bead, giving rise to a standard 4 R-bead model for hexadecane. Bonded distributions for an atomistic hexadecane chain in hexane extracted using COM (red) and COG (green) mappings for the (a) *central-central* and (b) *central-termini* bead. The average distance is 0.461 nm (COM) and 0.466 nm (COG) in (a) and 0.463 nm (COM) and 0.481 nm (COG) in (b), indicating a negligible difference for (a) and a very small one for (b).

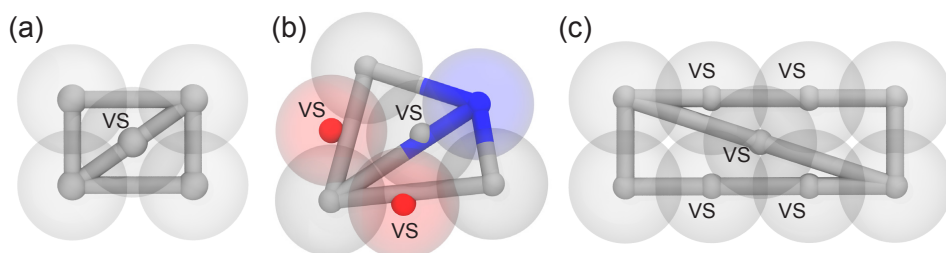### 6.5.3. Leveraging Virtual Sites in Extended Ring Models

Figure 6.10 | Example topologies using virtual sites. (a) Rendering of the naphthalene (NAPH) of Figure 6.3b. (b) Rendering of the caffeine model: the model uses the "hinge" construction, as described for NAPH, and additionally three virtual sites are built from the hinge scaffold. (c) Rendering of the tetracene model: the model uses the "hinge" construction, as described for NAPH, and additionally five virtual sites are built from the hinge scaffold. Virtual sites are indicated with the label "VS".

6.5.4. Solvent Accessible Surface Area Calculations

SASA Calculation Details. Solvent Accessible Surface Area (SASA) values have been computed using the GROMACS tool `gmx sasa` on energy-minimized AA or CG geometries. At both levels, we use the command:

```
gmx sasa -s benzene.gro -o sasa.xvg -probe 0.191 -ndots 4800
```

where the `-ndots` flag, which specifies the number of grid points used to calculate the SASA, should be set at least to 4800 for accurate values, and the probe size of 0.191 nm corresponds to the van der Waals (vdW) radius of a T-bead, r_{vdW}^T, that is:

$$r_{vdW}^T = r_m^T = \frac{d_m^{T-T}}{2} = \frac{\sqrt[6]{2} \cdot \sigma_{T-T}}{2} = \frac{\sqrt[6]{2} \cdot 0.34}{2} = 0.191 \, \text{nm} \tag{6.5}$$

where σ_{T-T} is the LJ σ parameter of T-T interactions, *i.e.*, 0.34 nm. The probe size impacts the absolute SASA values, but not their relative differences (Table 6.2). For the CG calculations, the file `vdwradii.dat` from

```
/usr/local/gromacs-XXX/share/gromacs/top/vdwradii.dat
```

(where `XX` could be any gromacs version) should be copied to the folder where the `gmx sasa` command is executed. The file, containing default vdW radii for atomistic force fields, should be modified so as to contain the vdW radii of Martini beads (computed as done for a T-bead in Eq. 6.5). Note that the vdW radii database file contains radii associated to *atom names* (and *not* atom or bead *types*).

Table 6.2 shows also that there is negligible difference between different all-atom force fields.

Table 6.2 | Solvent Accessible Surface Area (SASA) values in nm^2 for several rings with different all-atom force fields. Values in parentheses are computed using a T-bead sized probe, while the other values use the default probe of `gmx sasa`. Relative errors (%) of the Martini values (which are based on COG-mapping) with respect to the two all-atom force fields are also shown. BENZ = benzene; NAPH = naphthalene; CYPO = cyclopentanone.

molecule	GROMOS	OPLS	Martini 3.0	Err. % (GROMOS)	Err. % (OPLS)
BENZ	2.433 (2.976)	2.437 (2.986)	2.293 (2.807)	−6% (−6%)	−6% (−6%)
NAPH	3.108 (3.707)	3.106 (3.703)	3.035 (3.624)	−2% (−2%)	−2% (−2%)
CYPO	2.538 (3.073)	2.523 (3.063)	2.406 (2.935)	−5% (−5%)	−5% (−5%)

6.5.5. Solvent Properties: Mass Density with COG-Mapping
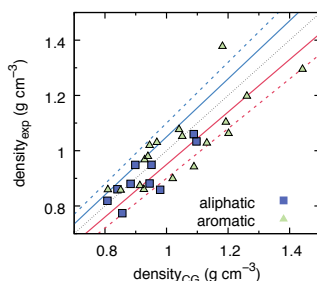

Figure 6.11 | Same as Figure 6.5 but using exclusively COG-based bond lengths for the Martini models.

6.5.6. Mixture Simulations: Miscibility Assays

We compute the contacts between the two components *A* and *B* with the command:

```
echo 0 1 | gmx mindist -f run.xtc  -d 0.6 -n index.ndx -on contacts.xvg
```

where `index.ndx` contains two groups, one containing all the the beads of component *A*, the other all the beads of component *B*; `run.xtc` is the trajectory of the production phase (at least 400 ns long), and `contacts.xvg` contains the number of *A-B* contacts as a function of simulation time. Two examples are plotted in Figure 6.12a: binary mixtures either readily demix and thus show a low number of *A-B* contacts, *e.g.*, benzene-water (BENZ-W) mixture, or stay mixed, as in the benzene-toluene (BENZ-TOLU) case.
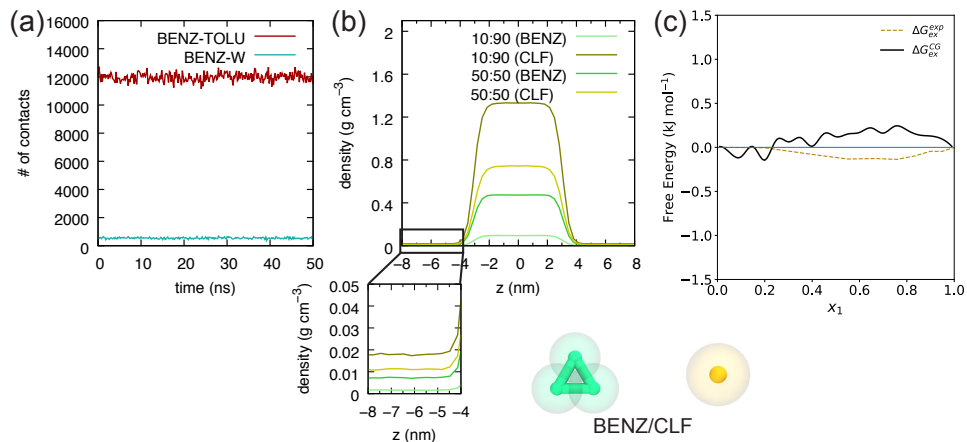

Figure 6.12 | (a) Typical evolutions of number of *A-B* contacts for a binary mixture. A low number of contacts—benzene-water (BENZ-W) mixture—indicate a phase separated systems; a high number of contacts—benzene-toluene (BENZ-TOLU) mixture—indicate that the two components are miscible. (b) Typical density profiles (centered around the liquid phase) obtained from VLE simulations; the inset shows the difference in densities of the vapor phases. (c) Martini 3.0 *vs* experimental benzene-chloroform (BENZ-CLF) ΔG_{ex} curves as a function of mixture composition. x_1 is the BENZ molar fraction. Experimental data are from Ref. 324.

6.5.7. Vapor–Liquid Equilibrium Simulations

Method Details. Vapor and liquid densities were extracted with the GROMACS tool `gmx density` using the following command:

```
echo A A | gmx density -f mix.xtc -s mix.tpr -d Z \
                       -o A-density-Z.xvg -symm -center
```

where "A" is the label of component *A*, *e.g.*, "BENZ" or "CLF". With the first "A" we select the group around which the density profile should be centered; with the second "A" we select the group for which we calculate the density. Typical profiles are shown in Figure 6.12b.

   Benzene-Chloroform Mixture. Chloroform (and other chlorinated solvents such as chlorobenzene) is a widely used solvent in organic electronics.[325] Aromatic systems, especially if functionalized with alkyl side chains, show good solubilities in such a solvent. We therefore also estimated quantitatively how the new Martini model performs with respect to the miscibility of chloroform and benzene by computing excess free energy of mixing (ΔG_{ex}) as a function of the composition of the mixture. The computed values, obtained as described in Section 6.2.3, are compared to experimental data[324] in Figure 6.12c.

6.5.8. Martini Stacking Interactions: Dimerization Free Energy Landscapes
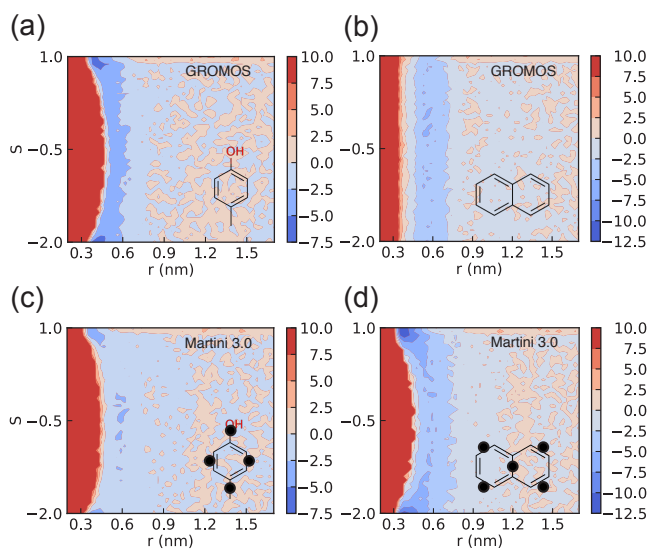

Figure 6.13 | Two-dimensional free energy profiles of dimerization for several aromatic compounds in water. The free energy surface is plotted on the 2D coordinate space formed by the distance between the COGs of the molecules and the order parameter of Eq. 6.4. (a)–(b) are AA, and (c)–(d) are Martini 3.0 surfaces. Molecular structures (and mappings) are shown as figure insets.