Supporting information for:
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
Contents

1 COG- vs COM-Based Mappings: Benzene and Cyclohexane S3
2 COG- vs COM-Based Mappings: Alkanes S4
3 Advanced Model Design Strategies S5
4 LINCS Settings S6
5 Solvent Accessible Surface Area Calculations S7
6 Solvent Properties: Mass Density with COG-Mapping S8
7 Mixture Simulations: Miscibility Assays S8
8 Vapor–Liquid Equilibrium Simulations S9
9 Stacking Interactions: Dimerization Free Energy Landscapes S10
10 Database: List of Molecules S11

References S14
1 COG- vs COM-Based Mappings: Benzene and Cyclohexane

Figure S1: 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.
Figure S2: Alkane bonded parameters extracted from all-atom (AA) simulations of hexadecane solvated in hexane. CG models are built by mapping 4 C atoms, and associated hydrogens, to each regular bead. Bonded distributions extracted using COM (red) and COG (green) mappings for the (a) central-central and (b) central-terminal 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 small one for (b).
3 Advanced Model Design Strategies

Figure S3: (top) Possible Martini models for naphthalene. (a) An “all-constraints” model that readily leads to numerical instabilities and (b) a “hinge” model that 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$^{-2}$. Nonbonded interactions between the two CG sites which share no bond in the “hinge” model—top left and bottom right in (b)—are excluded. (bottom) Example topologies using the “hinge” model. (d) Rendering of the naphthalene (NAPH) model of panel (b). (e) Rendering of the caffeine model: the model uses the “hinge” construction and additionally three virtual sites are built from the hinge scaffold. (f) Rendering of the tetracene model: the model uses the “hinge” construction and additionally five virtual sites are built from the hinge scaffold. Virtual sites are indicated with the label “VS”.

S5
Figure S4: LINCS settings have no significant impact on mixtures. (a) Benzene-cyclohexane, which mix quite well; (b) benzene-ethanol, mildly phase-separating; (c) benzene-water, strongly phase-separating. \texttt{lincs41} refers to the default LINCS settings (order=4, iter=1); \texttt{lincs82} (order=8, iter=2); \texttt{15fs} uses a time step of 15 fs (instead of the default 20 fs); \texttt{gmx’20} uses Gromacs 2020 (instead of 2016.5). Systems are \( \approx (5.5 \text{ nm})^3 \); contacts and temperatures are averaged over the last 350 ns of 400 ns long MD. In the temperature plots, the gray bar represents the temperature of benzene, the other bar the one of the other component. Error bars are standard deviations (last 350 ns). Default Martini “new” settings.
5 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:

$$r_{vdW}^T = r_m^T = \frac{d_{m-T}^T}{2} = \frac{\sqrt{2} \cdot \sigma_{T-T}}{2} = \frac{\sqrt{2} \cdot 0.34}{2} = 0.191 \text{ nm} \quad (1)$$

where $\sigma_{T-T}$ is the LJ $\sigma$ parameter of T-T interactions, $i.e.$, 0.34 nm. The probe size impacts the absolute SASA values, but not their relative differences (Table S1). 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. 1). Note that the vdW radii database file contains radii associated to atom names (and not atom or bead types). For the atomistic SASA calculations, we used the radii from Rowland and Taylor $^{S1}$ (the file can be found at [https://github.com/ricalessandri/Martini3-small-molecules](https://github.com/ricalessandri/Martini3-small-molecules)).

Table S1 shows also that there is negligible difference between different all-atom force fields.
Table S1: 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 \texttt{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 | 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 Solvent Properties: Mass Density with COG-Mapping

![Graph](a)

Figure S5: Mass densities and SASA values obtained used COG-based mappings. Same as Figure 5c-d but using exclusively COG-based bond lengths for the Martini models.

7 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 \texttt{index.ndx} contains two groups, one containing all the the beads of component \(A\), the other all the beads of component \(B\); \texttt{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 S6a: 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 S6: (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 vapor–liquid equilibrium simulations; the inset shows the difference in densities of the vapor phases. (c) Martini 3 vs experimental benzene-chloroform (BENZ-CLF) $\Delta G_{ex}$ curves as a function of mixture composition. $x_1$ is the BENZ molar fraction. Experimental data are from Ref. S2.

8 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 \n        -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 S6b.

**Benzene-Chloroform Mixture.** Chloroform (and other chlorinated solvents such as chlorobenzene) is a widely used solvent in organic electronics. 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 (\(\Delta G_{ex}\)) as a function of the composition of the mixture. The computed values, obtained as described in Section 2.3, are compared to experimental data in Figure S6c.

### 9 Stacking Interactions: Dimerization Free Energy Landscapes

Figure S7: 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. 4. (a)–(b) are AA, and (c)–(d) are Martini 3 surfaces. Molecular structures (and mappings) are shown as figure insets.
## 10 Database: List of Molecules

Table S2: Molecules in the database. The unique identifier (uID), common name, and SMILES string for all the molecules in the database.

| uID | name                        | SMILES                                   |
|-----|-----------------------------|------------------------------------------|
| 1MIMI | 1-METHYLIMIDAZOLE            | Cn1ccnc1                                 |
| 2MPYR  | 2-METHYL-PYRIDINE           | Cc1cccn1                                 |
| 2NIMX  | 2-NITRO-m-XYLENE            | Cc1c(c(cc1)C)[N+](=O)[O-]               |
| 2NITL  | 2-NITROTOLUENE              | Cc1cccccc1[N+](=O)[O-]                  |
| 2T     | 2,2'-BITHIOPHENE            | c1cc(sc1)c2cccs2                        |
| 3HT    | 3-HEXYL-THIOPHENE           | CCCCCCc1ccsc1                            |
| 3PT    | 3-PROPYL-THIOPHENE          | CCCc1ccsc1                               |
| 4MIMI  | 4-METHYLIMIDAZOLE           | [nH]1cc(nc1)C                           |
| 4NIAN  | 4-NITROANISOLE              | COc1ccc(cc1)[N+](=O)[O-]                |
| ACPH   | ACETOPHENONE                | CC(=O)c1cccccc1                         |
| ANIL   | ANILINE                    | Nc1cccccc1                               |
| ANTH   | ANTHRACENE                  | c1ccc2cc3cccccc3cc2c1                   |
| BEAL   | BENZALDEHYDE                | O=Cc1cccccc1                            |
| BENZ   | BENZENE                     | c1cccccc1                                |
| BRA    | 4-BROMOANISOLE              | COc1ccc(cc1)Br                          |
| BRBZ   | BROMOBENZENE                | c1ccc(cc1)Br                            |
| BZDOL  | 1,3-BENZENEDIOL             | Oc1cc(cc1)O                             |
| BZIM   | BENZIMIDAZOLE               | c1ccc2c(c1)[nH]cn2                     |
| BZNI   | BENZONITRILE                | N#Cc1cccccc1                             |
| BZQU   | para-BENZOQUINONE           | c1cc(=O)ccc1=O                         |
| BZTA   | BENZOTHIAZOLE               | n1c2cccccc2sc1                          |
| BZTF   | BENZOTRIFLUORIDE            | c1ccc(cc1)C(F)(F)F                     |
| BZTH   | BENZOTHIOPHENE              | c1c2cccccc2sc1                          |
| uID   | name                        | SMILES                                                                 |
|-------|-----------------------------|------------------------------------------------------------------------|
| CAFF  | CAFFEINE                    | Cn1cnc2n(C)c(=O)n(C)c(=O)c12                                          |
| CHEX  | CYCLOHEXANE                 | C1CCCCC1                                                               |
| CHEXE | CYCLOHEXENE                 | C1CCceC1                                                               |
| CLBZ  | CHLOROBENZENE               | Clc1cccccc1                                                            |
| CLPR  | CHLORPROPHAM                | CC(C)OC(=O)Ne1cc(ccc1)Cl                                                |
| CLTL  | 2-CHLOROTOLUENE             | Cc1c(cccc1)Cl                                                          |
| CNAP  | 1-CHLORO-NAPHTHALENE        | Clc2ccccc1ccccc12                                                       |
| CPR   | CYCLOPROPANE                | C1CC1                                                                  |
| CUME  | CUMENE                      | CC(C)c1cccccc1                                                         |
| CYPE  | CYCLOPENTANE                | C1CCCCC1                                                               |
| CYPO  | CYCLOPENTANONE              | C1CCC(=O)C1                                                            |
| CYPOL | CYCLOPENTANOL               | C1CCC(C1)O                                                             |
| DBRBZ | 1,2-DIBROMOBENZENE          | c1ccc(c(c1)Br)Br                                                        |
| DCLBZ | 1,2-DICHLOROBENZENE         | Clc1cccc(Cl)c1                                                         |
| DIOX  | 1,4-DIOXANE                 | O1CCOCC1                                                               |
| DMAN  | N,N-DIMETHYLANILINE         | CN(C)e1ccccce1                                                         |
| DMBZQ | 2,5-DIMETHYL-1,4-BENZOQUINONE| Cc1cc(=O)c(cc1=O)C                                                       |
| DXLA  | 1,3-DIOXOLANE               | O1CCOC1                                                                |
| EBEN  | ETHER-BENZENE               | CCc1cccccl                                                             |
| ENAPH | 1-ETHYL-NAPHTHALENE         | CCc1cccc2cccccc21                                                      |
| FURA  | FURAN                       | c1ccoc1                                                                |
| IMID  | IMIDAZOLE                   | c1cnc[nH]l                                                              |
| INDA  | INDAZOLE                    | c2ccc1[nH]nc1c2                                                         |
| INDO  | INDOLE                      | c12c(ccn2)cccc1                                                        |
| IOBZ  | IODOBENZENE                 | c1ccc(cc1)I                                                             |
| uID  | name                      | SMILES                                      |
|------|---------------------------|---------------------------------------------|
| IOPHE| 2-iodophenol              | c1ccc(c(c1)O)I                             |
| MBZOA| methyl-benzoate           | COc1ccc1c1ccccc1                           |
| MCYPE| methycyclopentane         | CC1CCCC1                                   |
| MESI | mesitylene                | Cc1cc(c(c1)C)C                             |
| MIND | 3-methyl-1H-indole        | c1ccccc2c1c[c(nH)2]C                       |
| MINDA| 1-methyl-indazole         | Cn1c2cccccc2cn1                            |
| MNAP | 1-methyl-naphthalene      | Ce2cccc1ccccccc12                         |
| MXYLE| m-xylene                 | Cc1ccccc(c1)C                              |
| NAPH | naphthalene               | c1cc2cccccc2c1                             |
| NAPY | 1,5-naphthyridine         | c1cc2ncccc2n1                              |
| NBAPH| N-BOC-2-aminophenol       | Oc1ccccc1NC(=O)OC(C)(C)(C)                 |
| NDMBI| N-DMBI                   | CN(C)e1cc(c1)C2N(C)e3cccccc3N2C           |
| NIBZ | nitrobenzene              | c1ccc(c1)[N+](=O)[O-]                      |
| OMA  | ortho-methylanisole       | COc1ccccc1c1                              |
| OXYLE| o-xylene                  | Cc1c(C)cccc1                               |
| PBEN | propyl-benzene            | CCCc1cccc1                                 |
| PBZOA| propyl-benzoate           | CCCOC(=O)c1ccccc1                         |
| PCRE | para-cresol               | Cc1cc(O)c1c1                              |
| PCYM | p-cymene                  | c1cc(ccc1C(C)C)C                          |
| PHEN | phenol                    | c1cc(c1)(CC)O                             |
| PIPER| piperidine                | C1CCNCC1                                   |
| PXYLE| p-xylene                  | Cc1cc(CC)(cc1)                             |
| PYAZ | pyridazine                | c1ccntn1                                   |
| PYLI | pyrrolidine               | C1CCNCl                                    |
| PYMI | pyrimidine                | c1cnccc1                                   |
Table S2: (continued)

| uID  | name                          | SMILES                                                |
|------|-------------------------------|-------------------------------------------------------|
| PYRI | PYRIDINE                      | c1ccncnc1                                             |
| PYRR | PYRROLE                       | [nH]1ccnc1                                            |
| QUIN | QUINOLINE                     | n1ccnc2ccnc12                                         |
| STYR | STYRENE                       | c1ccnc1C=C                                           |
| TCLBZ| 1,2,4-TRICHLOROBENZENE        | Cclccc(Cl)c(Cl)c1                                     |
| TDMBI| TEG-DMBI                      | COCCOCCOCCOCCOc1cccc(cc1)⋯                             |
|      |                               | ⋮                                                       |
| TECE | TETRACENE                     | c34cc2cc1ccccc1cc2cc3ccccc4                           |
| THAZ | THIAZOLE                      | n1ccsc1                                               |
| THF  | TETRAHYDROFURAN               | C1CCOC1                                               |
| THIO | THIOPHENE                     | c1ccsc1                                               |
| THP  | TETRAHYDROPYRAN               | O1CCCCC1                                              |
| TPH  | THIOPHENOL                    | Sc1ccccc1                                             |
| THPY | THIENO[2,3-c]PYRIDINE         | c1cncc2cc1ccs2                                        |
| THT  | TETRAHYDROTHIOPHENE           | S1CCCC1                                               |
| TOLU | TOLUENE                       | Cc1ccccc1                                             |
| XBZ  | METHOXYBENZENE                | COc1ccccc1                                            |
| XNAPH| 1-METHOXYNAPHTHALENE          | COc1ccccc2ccccc21                                     |

References

(S1) Rowland, R. S.; Taylor, R. J. Phys. Chem. 1996, 100, 7384–7391.

(S2) Campbell, A.; Kartzmark, E. M.; Chatterjee, R. Can. J. Chem. 1966, 44, 1183–1189.

(S3) Zhang, S.; Ye, L.; Zhang, H.; Hou, J. Mater. Today 2016, 19, 533–543.